

Запорізький національний технічний університет



Радіоелектроніка Інформатика Управління

2(12)'2004

Науковий журнал

Видається двічі на рік

Зареєстрований **29 січня 2003 року**
Державним комітетом інформаційної політики,
телебачення та радіомовлення України.

Свідоцтво – серія **КВ № 6904**.

Засновник і видавник – *Запорізький національний технічний університет*

Запоріжжя, ЗНТУ
2004

Передплатний індекс – 22914

Журнал “Радіоелектроніка. Інформатика. Управління” видається з 1999 року.

Журнал входить до переліку наукових фахових видань України, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора і кандидата технічних наук та фізико-математичних наук за спеціальністю “Радіофізика”. Статті, що публікуються в журналі, реферуються в базах даних та РЖ ВІНІТІ (Росія) і “Джерело” (Україна)

РЕДАКЦІЙНА КОЛЕГІЯ

Головний редактор – д.т.н. Потапенко Є. М.

Заст. головного редактора – к.т.н. Дубровін В. І.

Члени редколегії :

д.ф.-м.н. Ахметшин А. М.

д.т.н. Волков О. В.

д.ф.-м.н. Горбань О. М.

д.ф.-м.н. Горр Г. В.

д.т.н. Гостєв В. І.

д.ф.-м.н. Дробахин О. О.

д.ф.-м.н. Ковальов О. М.

д.ф.-м.н. Корніч Г. В.

д.т.н. Павлов О. А.

д.т.н. Піза Д. М.

д.ф.-м.н. Погосов В. В.

д.т.н. Труфанов І. Д.

д.ф.-м.н. Чумаченко В. П.

Редакційно-видавнича рада:

Беліков С. Б., Внуков Ю. М., Піза Д. М., Яримбаш С. Т., Потапенко Є. М., Дубровін В. І., Петруша Ю. П., Флора В. Д., Соболев Ю. О., Дедков М. В., Рибіна Ю. А., Табунщик Г. В., М'ясников М. Л., Богданова Л. Ф., Кобзар Н. М., Федоренко Е. С.

Рекомендовано до видання вченою радою Запорізького національного технічного університету, протокол № 4 від 01.12.2004 р.

Рукописи проходять незалежне рецензування з залученням провідних фахівців, за результатами якого редакційна колегія приймає рішення про опублікування

Журнал зверстаний редакційно-видавничим відділом
Запорізького національного технічного університету

Комп'ютерне моделювання та верстання

Новикова К. О.

Плюйко І. І.

Адреса редакції: 69063, м. Запоріжжя, вул. Жуковського, 64, ЗНТУ,
редакція журналу "PIU"

Тел: (0612)63-09-99 – редакційно-видавничий відділ

Факс: (0612)64-21-41

E-mail: rvv@zntu.edu.ua

TQM@zntu.edu.ua

Web-сайт журналу розташований в Internet за адресою: <http://www.zntu.edu.ua/RIC>

© 2004 Запорізький національний технічний університет

ЗМІСТ

РАДІОФІЗИКА 6

В. П. Бондарев, С. С. Самойлик
ЭЛЕКТРОМАГНИТНОЕ ПОЛЕ ПРЯМОУГОЛЬНОГО
РЕЗОНАТОРА С НЕСТАЦИОНАРНОЙ
ЦИЛИНДРИЧЕСКОЙ НЕОДНОРОДНОСТЬЮ 6

И. И. Зиненко, В. П. Пьянков, В. П. Чумаченко
СОГЛАСОВАНИЕ ПЯТИПЛЕЧЕГО
Е-ПЛОСКОСТНОГО СОЕДИНЕНИЯ
ПРЯМОУГОЛЬНЫХ ВОЛНОВОДОВ10

Л. М. Карпуков, Р. Д. Пулов, С. Н. Романенко
МОДЕЛИРОВАНИЕ МНОГОСЛОЙНЫХ
МИКРОПОЛОСКОВЫХ СТРУКТУР В КВАЗИ-
ДИНАМИЧЕСКОМ ПРИБЛИЖЕНИИ14

В. М. Онуфрієнко, Т. О. Штефан
ІНТЕГРОДИФЕРЕНЦІАЛЬНЕ МОДЕЛЮВАННЯ
НЕЛІНІЙНОЇ ПОЛЯРИЗАЦІЇ ТА СТРУМУ
У ФРАКТАЛЬНО НЕОДНОРІДНИХ
ПЛАЗМОЇДАХ 17

В. П. Пьянков
ЧИСЛЕННЫЙ АНАЛИЗ СЛОЖНЫХ
ВОЛНОВОДНЫХ ИЗЛУЧАТЕЛЕЙ
С ФЛАНЦАМИ НА ОСНОВЕ ОБОБЩЕННЫХ
МАТРИЦ РАССЕЯНИЯ И МЕТОДА
ПРОИЗВЕДЕНИЯ ОБЛАСТЕЙ 21

РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ26

В. И. Гостев
СИСТЕМА РЕГУЛИРОВАНИЯ ТЕМПЕРАТУРЫ
ТЕПЛОНОСИТЕЛЯ НА ВЫХОДЕ СМЕСИТЕЛЯ
С НЕЧЕТКИМИ РЕГУЛЯТОРАМИ26

В. Н. Журавлев, В. С. Кабак, В. О. Рыбин
АНАЛИЗ ПАРАМЕТРОВ АВТОКОРРЕЛЯЦИОННОЙ
ФУНКЦИИ ФРАГМЕНТОВ ФОНЕМ В ЗАДАЧАХ
ИДЕНТИФИКАЦИИ АБОНЕНТА29

М. М. Касьян, К. М. Касьян
СИНТЕЗ ВІДБРАКУВАЛЬНИХ ПОТОЧНИХ ДО-
ПУСКІВ НА ПАРАМЕТРИ ЕЛЕМЕНТІВ У ЗАЛЕЖ-
НОСТІ ВІД ЧАСУ ТА УМОВ ЕКСПЛУАТАЦІЇ33

*В. М. Крищук, Г. М. Шило, А. О. Намлинський,
М. П. Гапоненко*
ВИБІР ЕЛЕМЕНТІВ ПРИ КОМПЕНСАЦІЇ
ЗОВНІШНІХ ВПЛИВІВ36

С. И. Прохорец, М. А. Хажмурадов, В. П. Лукьянова
АНАЛИЗ ЕМКОСТНЫХ ПАРАМЕТРОВ СТРИП-
ДЕТЕКТОРА 41

О. Ю. Фарафонов, С. О. Самоквіт, О. Ю. Воропай
ПРОЕКТУВАННЯ МІКРОСМУЖКОВИХ
НАПРАВЛЕНИХ ВІДГАЛУЖУВАЧІВ
З ОПТИМАЛЬНИМИ ДОПУСКАМИ 44

В. Д. Флора
ОСОБЛИВОСТІ ЖИВЛЕННЯ ДВИГУНА
ПОСТІЙНОГО СТРУМУ ЧЕРЕЗ ІМПУЛЬСНИЙ
ПЕРЕТВОРЮВАЧ 47

Н. П. Чернобородова, М. П. Чернобородов
ЦИФРОВАЯ СИСТЕМА КОМПЕНСАЦИИ
ИМПУЛЬСНЫХ ПОМЕХ 49

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ55

Н. В. Алипов, И. Н. Алипов, М. И. Хиль, В. Н. Сидоров
МЕТОДЫ ЗАЩИТЫ ИНФОРМАЦИИ В ДИСКРЕТ-
НОМ КАНАЛЕ НА ОСНОВЕ ПОМЕХОУСТОЙЧИ-
ВЫХ К СИММЕТРИЧНЫМ РЕГУЛЯРНЫМ
ВОЗДЕЙСТВИЯМ АЛГОРИТМОВ ПОИСКА
ТОЧКИ С ХАРАКТЕРНЫМ ПРИЗНАКОМ55

Ю. С. Афонин, В. И. Дубровин
МОДЕЛИРОВАНИЕ СОСТАВОВ СМЕСЕЙ
МЕТОДОМ СИМПЛЕКСНЫХ РЕШЕТОК60

А. И. Вершина, Г. Г. Киричек, Д. М. Пиза
МОДЕЛЬ ИНФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ
УЧЕБНОГО ПРОЦЕССА УНИВЕРСИТЕТА64

И. В. Гребенник, Т. Е. Романова, Л. Г. Евсеева
ОСНОВНАЯ ОПТИМИЗАЦИОННАЯ ЗАДАЧА
ГЕОМЕТРИЧЕСКОГО ПРОЕКТИРОВАНИЯ
В ИНТЕРВАЛЬНОМ ВИДЕ68

В. И. Долгов, А. В. Нелася
МЕТОДЫ УВЕЛИЧЕНИЯ СКОРОСТИ
КРИПТОГРАФИЧЕСКИХ ПРЕОБРАЗОВАНИЙ
НА ЭЛЛИПТИЧЕСКИХ КРИВЫХ 72

В. И. Дубровин
МЕТОДЫ СОКРАЩЕНИЯ ОБЪЕМА ДАННЫХ ..78

И. Ш. Невлюдов, Е. П. Второв, Б. А. Шостак
ОПТИМИЗАЦИЯ ПРОВЕРКИ ЦИФРОВЫХ
МОДУЛЕЙ ТЕХНОЛОГИЧЕСКОГО
ОБОРУДОВАНИЯ В УСЛОВИЯХ ЧАСТИЧНОЙ
НЕОПРЕДЕЛЕННОСТИ 84

В. В. Орлов
БЛОК КЛАВИАТУРЫ ИНФОРМАЦИОННО-
ИЗМЕРИТЕЛЬНОЙ СИСТЕМЫ НА БАЗЕ
МИКРОКОНТРОЛЛЕРА 87

<i>М. Ю. Сенашова</i> ОЦЕНКИ ПОГРЕШНОСТЕЙ ВЫЧИСЛЕНИЯ СЛОЖНОЙ ФУНКЦИИ МНОГИХ ПЕРЕМЕННЫХ И ЕЕ ГРАДИЕНТА	<i>А. М. Чугай</i> МЕТОД ПОИСКА ЛОКАЛЬНЫХ ЭКСТРЕМУМОВ В ЗАДАЧЕ УПАКОВКИ ОДИНАКОВЫХ ЦИЛИНДРОВ В МНОГОУГОЛЬНОЙ ПРИЗМЕ С ЗОНАМИ ЗАПРЕТА	90	94
НЕЙРОИНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ		101	
<i>В. П. Авраменко, В. В. Калачева, А. В. Цурихин</i> АДАПТИВНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ СОСТОЯНИЙ ПРОИЗВОДСТВЕННЫХ СИСТЕМ	<i>М. А. Новотарский</i> КЛЕТОЧНЫЕ НЕЙРОННЫЕ СЕТИ С ТРАНЗИТНЫМИ ПЕРЕСЫЛКАМИ	101	118
<i>Е. В. Бодянский, А. Н. Слипченко</i> АНСАМБЛЬ НЕЙРОПРЕДИКТОРОВ С ПЕРЕМЕННЫМ ЧИСЛОМ УЗЛОВ	<i>Е. А. Самойлин</i> КОМПЛЕКСНАЯ ОЦЕНКА АЛГОРИТМОВ ФИЛЬТРАЦИИ-СЕКМЕНТАЦИИ ИЗОБРАЖЕНИЙ	107	122
<i>Л. А. Жуков, Н. В. Решетникова</i> ИССЛЕДОВАНИЕ РАЗЛИЧНЫХ ВАРИАНТОВ КОНТРАСТИРОВАНИЯ И БИНАРИЗАЦИИ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОБРАБОТКИ МЕДИЦИНСКИХ ДАННЫХ		112	
ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ		127	
<i>А. А. Завалин, А. А. Зенькович</i> АВТОМАТИЗИРОВАННОЕ ФОРМИРОВАНИЕ ОБЪЕКТНОГО СЛОВАРЯ РЕЛЯЦИОННОЙ БАЗЫ ДАННЫХ	<i>В. М. Опанасенко, І. Г. Тимошенко</i> АРХІТЕКТУРНА ОРГАНІЗАЦІЯ РЕКОНФІГУРОВНИХ КОМП'ЮТЕРІВ НА БАЗІ ПЛІС	127	139
<i>А. А. Каргин, А. В. Григорьев</i> НЕЧЕТКИЙ МЕТОД К ИДЕНТИФИКАЦИИ ЧЕЛОВЕКА ПО ФОТОПОРТРЕТУ		131	
<i>А. Б. Кунгурцев</i> ИССЛЕДОВАНИЕ ИНФОРМАЦИОННЫХ СИСТЕМ ДЛЯ ВНЕДРЕНИЯ МЕХАНИЗМОВ МАТЕРИАЛИЗОВАННЫХ ПРЕДСТАВЛЕНИЙ И ШАБЛОНОВ		135	
ТЕОРІЯ І МЕТОДИ АВТОМАТИЧНОГО УПРАВЛІННЯ		145	
<i>И. А. Орловский</i> НАСТРОЙКА МАСШТАБИРУЮЩИХ КОЭФФИЦИЕНТОВ ФАЗЗИ-РЕГУЛЯТОРА	<i>Е. А. Шушлятин, А. Е. Работнов</i> УПРАВЛЕНИЕ ТЕРМИНАЛЬНЫМИ НЕЛИНЕЙНЫМИ МНОГОИНДЕКСНЫМИ ДИСКРЕТНЫМИ СИСТЕМАМИ МЕТОДОМ КОНЕЧНОГО СОСТОЯНИЯ	145	152
КЕРУВАННЯ У ТЕХНІЧНИХ СИСТЕМАХ		157	
<i>В. В. Осадчий</i> ОСОБЕННОСТИ АДАПТИВНОГО ПРОГРАММНОГО УПРАВЛЕНИЯ ТЕХНОЛОГИЧЕСКИМИ ПРОЦЕССАМИ ПРОИЗВОДСТВА ОГНЕУПОРНЫХ МАТЕРИАЛОВ	<i>Е. М. Потапенко, А. В. Соломаха, Е. Е. Потапенко</i> КАЛИБРОВКА ДАТЧИКОВ ОДНОФАЗНЫХ ГАРМОНИЧЕСКИХ СИГНАЛОВ С ПОМОЩЬЮ ДИНАМИЧЕСКИХ ФИЛЬТРОВ	157	164
АВТОРЫ НОМЕРА		168	



БОГУСЛАЄВ
Олександр В'ячеславович



СУББОТІН
Сергій Олександрович

ЩИРО ВІТАЄМО

наших авторів провідного інженера ВАТ "Мотор Січ", кандидата технічних наук **Богуслаєва Олександра В'ячеславовича** і старшого викладача кафедри програмних засобів Запорізького національного технічного університету **Субботіна Сергія Олександровича з присудженням** їм Указом Президента України Л. Кучми № 1451/2004 від 09.12.04 р. на основі представлення Комітету з Державних премій України в галузі науки і техніки **премії Президента України для молодих вчених** за роботу "Методи та алгоритми синтезу нейромережєвих діагностичних і прогнозуючих моделей"!

Редакція журналу "Радіоелектроніка. Інформатика. Управління"



УКАЗ ПРЕЗИДЕНТА УКРАЇНИ

Про присудження щорічних премій Президента України для молодих вчених 2004 року

На підставі подання Комітету з Державних премій України в галузі науки і техніки **постановляю:**

Присудити щорічні премії Президента України для молодих вчених 2004 року:

...

– за роботу "Методи та алгоритми синтезу нейромережєвих діагностичних і прогнозуючих моделей":

БОГУСЛАЄВУ Олександрю В'ячеславовичу – кандидатові технічних наук, провідному інженерові відкритого акціонерного товариства "Мотор Січ",

СУББОТІНУ Сергію Олександровичу – старшому викладачеві Запорізького національного технічного університету.

...

Президент України

Л. КУЧМА

м. Київ

09 грудня 2004

№ 1451/2004

РАДІОФІЗИКА

РАДИОФИЗИКА

RADIOPHYSICS

УДК 621.372.81

В. П. Бондарев, С. С. Самойлик

ЭЛЕКТРОМАГНИТНОЕ ПОЛЕ ПРЯМОУГОЛЬНОГО РЕЗОНАТОРА С НЕСТАЦИОНАРНОЙ ЦИЛИНДРИЧЕСКОЙ НЕОДНОРОДНОСТЬЮ

Краевая задача для определения электромагнитного поля в резонансной системе, включающей цилиндрическую неоднородность с изменяющейся во времени диэлектрической проницаемостью и удельной проводимостью сведена к интегральному уравнению. Для малых амплитуд модуляции получено характеристическое уравнение для ТЕ колебаний, позволяющее определить поле вблизи и в области основного параметрического резонанса. При численном анализе решений найдены зоны неустойчивости и значения характеристического показателя в широкой области изменения частоты модуляции.

ВВЕДЕНИЕ

Большой интерес для возбуждения электромагнитных волн в резонансных системах представляют элементы, параметры которых могут изменяться во времени под действием волны накачки [1–3]. Строго говоря, эти явления происходят при взаимодействии электромагнитной волны с нелинейным элементом. Однако, когда распространяющаяся волна гораздо слабее мощной волны накачки, то допустима линеаризация задачи. На основе таких структур возможно создание эффективных параметрических усилителей и генераторов, а также преобразователей спектра электромагнитных волн.

В настоящее время отсутствуют электродинамические методы расчета подобных структур, а существующие аналоги используют для расчета только радиотехнические модели с сосредоточенными параметрами, не позволяющими учитывать геометрические размеры нестационарной неоднородности. Поэтому решение таких задач требует применения строгих методов решения и более подробного анализа и обоснования [4, 5].

ПОСТАНОВКА ЗАДАЧИ

В качестве математической модели параметрических устройств выбран резонатор, содержащий объемный

твердотельный элемент в виде стержня цилиндрической формы, диэлектрическая проницаемость и удельная проводимость которого изменяются во времени по гармоническому закону. При выполнении определенных соотношений между частотой накачки и одной из резонансных частот резонатора в устройстве возможно возникновение параметрического резонанса.

Рассмотрим прямоугольный резонатор (рис. 1). Предположим, что в отсутствие возмущения в резонаторе существовали свободные гармонические колебания типа H_{m0n} с частотой ω_{mn} . В момент времени $t = 0$ в резонатор помещают стержень с диэлектрической проницаемостью $\epsilon(t)$ и удельной проводимостью $\sigma(t)$.

При $t > 0$ электромагнитное поле в резонаторе удовлетворяет интегральному уравнению [4, 5]:

$$\begin{aligned} \vec{E}(\vec{r}, t) = & \vec{E}_0(\vec{r}, t) - \epsilon_0 \mu_0 \frac{\partial^2}{\partial t^2} \int_0^t dt \int_V G(\vec{r}, t, \vec{r}', t') \times \\ & \times [\epsilon(t') - \epsilon_1] \cdot \vec{E}(\vec{r}', t') d\vec{r}' - \frac{\partial}{\partial t} \mu_0 \int_0^t dt' \int_V G(\vec{r}, t, \vec{r}', t') \times \\ & \times \sigma(t') \cdot \vec{E}(\vec{r}', t') d\vec{r}', \end{aligned} \quad (1)$$

где $\vec{E}_0(\vec{r}, t)$ – начальное поле в резонаторе, $G(\vec{r}, t, \vec{r}', t')$ – скалярная функция Грина прямоугольного резонатора, ϵ_1 – диэлектрическая проницаемость свободного пространства.

Будем считать, что диэлектрическая проницаемость и удельная проводимость стержня изменяются по гармоническому закону:

$$\epsilon(t) = \epsilon + \Delta\epsilon \cdot \cos(\Omega t + \varphi_0), \quad (2)$$

$$\sigma(t) = \sigma + \Delta\sigma \cdot \cos(\Omega t + \varphi_0), \quad (3)$$

где Ω – частота накачки, φ_0 – начальная фаза, ε – диэлектрическая проницаемость стержня в отсутствие накачки, $\Delta\varepsilon$ – амплитуда модуляции, σ – начальная проводимость среды, $\Delta\sigma$ – амплитуда модуляции потерь.

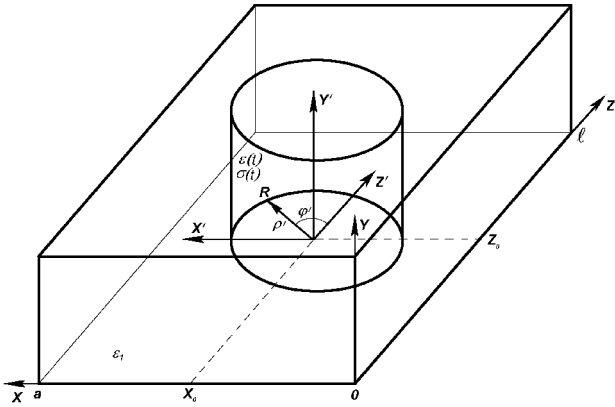


Рисунок 1 – Резонатор, частично заполненный нестационарной средой

РЕШЕНИЕ

Будем считать, что диэлектрический стержень однороден вдоль оси y . Тогда возбуждаемое внутри резонатора поле можно выразить через y составляющую $E_y = E$ и представить в виде разложения по собственным функциям прямоугольного резонатора:

$$E(\vec{r}, t) = \sum_{\alpha\beta=1}^{\infty} E_{\alpha\beta}(t) \cdot \sin\left(\frac{\alpha\pi}{a}x\right) \cdot \sin\left(\frac{\beta\pi}{l}z\right). \quad (4)$$

Для рассматриваемой структуры функция Грина, в общем случае тензорная, имеет только диагональные элементы, которые должны удовлетворять граничным условиям на стенках резонатора, накладываемым на соответствующую компоненту поля. Поэтому функцию Грина $G(\vec{r}, t, \vec{r}', t')$ с условием ортогональности и нормировки в области $0 \leq x \leq a$, $0 \leq z \leq l$ можно представить в виде:

$$G(\vec{r}, t, \vec{r}', t') = \frac{4}{\varepsilon \cdot a \cdot l \cdot \varepsilon_0 \mu_0} \sum_{m,n=1}^{\infty} \frac{\sqrt{\varepsilon_1}}{\omega_{mn}} \sin\left(\frac{\omega_{mn}}{\sqrt{\varepsilon_1}}(t-t')\right) \times \Phi_{mn}(x, z) \cdot \Phi_{mn}(x', z'), \quad (5)$$

где a, l – размеры прямоугольного резонатора, $\omega_{mn} = c \cdot \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{l}\right)^2}$ – резонансная частота, $\Phi_{mn} = \sin\left(\frac{m\pi}{a}x\right) \cdot \sin\left(\frac{n\pi}{l}z\right)$.

Подставим выражения для напряженности электромагнитного поля (4) и функции Грина (5) в интегральное уравнение (1). Тогда получим следующее выражение:

$$\begin{aligned} \sum_{m,n=1}^{\infty} E_{mn}(t) \cdot \Phi_{mn}(x, z) = E_0(t, x, z) - \\ - \varepsilon \mu \frac{\partial^2}{\partial t^2} \int_0^t dt' \int_V \frac{4}{\varepsilon_0 \mu_0 \varepsilon_1 a l} \sum_{m,n=1}^{\infty} \frac{\sin(\omega(t-t'))}{\omega_{mn}} \times \\ \times \Phi_{mn}(x, z) \cdot \Phi_{mn}(x', z') \cdot (\varepsilon(t') - \varepsilon_1) \sum_{\alpha\beta=1}^{\infty} E(t') \times \\ \times \Phi_{\alpha\beta}(x', z') dx' dz' - \frac{\partial}{\partial t} \mu_0 \int_0^t dt' \int_V \frac{4}{\varepsilon_0 \mu_0 \varepsilon_1 a l} \times \\ \times \sum_{m,n=1}^{\infty} \frac{\sin(\omega_{mn}(t-t'))}{\omega_{mn}} \cdot \Phi_{mn}(x, z) \cdot \Phi_{mn}(x', z') \times \\ \times \sigma(t') \sum_{\alpha\beta=1}^{\infty} E_{\alpha\beta}(t') \cdot \Phi_{\alpha\beta}(x', z') dx' dz', \quad (6) \end{aligned}$$

где введено обозначение $\omega_{mn} = \frac{\omega_{mn}}{\sqrt{\varepsilon_1}}$.

После вычисления интегралов по пространственным координатам и некоторых преобразований получаем:

$$\begin{aligned} \sum_{m,n=1}^{\infty} E_{mn}(t) \cdot \sin\left(\frac{m\pi}{a}x\right) \cdot \sin\left(\frac{n\pi}{l}z\right) = \\ = E_0(t) \cdot \sin\left(\frac{m_0\pi}{a}x\right) \cdot \sin\left(\frac{n_0\pi}{l}z\right) - \\ - \frac{1}{\varepsilon_1} \frac{\partial^2}{\partial t^2} \int_0^t \sum_{0m,n=1}^{\infty} \sum_{\alpha,\beta=1}^{\infty} A_{mn,\alpha\beta} \frac{\sin(\omega_{mn}(t-t'))}{\omega_{mn}} \times \\ \times \sin\left(\frac{m\pi}{a}x\right) \cdot \sin\left(\frac{n\pi}{l}z\right) \cdot (\varepsilon(t') - \varepsilon_1) \cdot E_{\alpha\beta}(t') dt' - \\ - \frac{1}{\varepsilon_0 \varepsilon_1} \frac{\partial}{\partial t} \int_0^t \sum_{0m,n=1}^{\infty} \sum_{\alpha,\beta=1}^{\infty} A_{mn,\alpha\beta} \frac{\sin(\omega_{mn}(t-t'))}{\omega_{mn}} \times \\ \times \sin\left(\frac{m\pi}{a}x\right) \cdot \sin\left(\frac{n\pi}{l}z\right) \cdot \sigma(t') \cdot E_{\alpha\beta}(t') dt', \quad (7) \end{aligned}$$

где $A_{mn,\alpha\beta}$ – коэффициент, учитывающий геометрию неоднородности.

Учитывая ортогональность собственных функций резонатора по координатам x и z на соответствующих отрезках $0 \leq x \leq a$, $0 \leq z \leq l$, из (7) получим систему интегральных уравнений относительно неизвестных функций $E_{mn}(t)$, которые можно переписать в следующем виде:

$$E_p(t) = E_{p_0}(t) \cdot \delta_{pp_0} - \frac{1}{\varepsilon_1 \overline{\omega}_p} \sum_{q=1}^{\infty} A_{pq} \left\{ \frac{\partial^2}{\partial t^2} \int_0^t \sin(\overline{\omega}_{mn}(t-t')) \times \right. \\ \left. \times (\varepsilon(t') - \varepsilon_1) \cdot E_q(t') dt' - \frac{1}{\varepsilon_0} \frac{\partial}{\partial t} \int_0^t \sin(\overline{\omega}_{mn}(t-t')) \times \right. \\ \left. \times \mu(t') \cdot E_q(t') dt' \right\}, \quad (8)$$

где $p = 1, 2, 3, \dots$, под индексами p, p_0, q необходимо понимать совокупность индексов $m, n; m', n_0; a, b$ соответственно, δ_{pp_0} – символ Кронекера.

Линейно независимые решения системы (8) представим в виде ряда Флоке:

$$E_p(t) = \sum_{r,s=-\infty}^{\infty} \sum_{rp} \left(U_{rp}^{+(s)} \cdot e^{i\omega_{rs}t} + U_{rp}^{-(s)} \cdot e^{-i\omega_{rs}t} \right), \quad (9)$$

где $\omega_{rs} = \omega_r + s\Omega$, $s = 1, 2, 3, \dots$ – частоты s -х гармоник для каждого из r -типов колебаний, ω_r – частота основной гармоники r -го типа колебаний, Ω – частота накачки, U_{rq}^{\pm} – амплитуды соответствующих гармоник r -х типов колебаний, подлежащие определению.

Вычисляя интегралы, входящие в (8) и производя некоторые преобразования для слагаемых пропорциональных $e^{i\overline{\omega}_p t}$ получаем следующие соотношения:

$$E_p \delta_{pp_0} e^{i\overline{\omega}_{p_0} t} - \sum_{s=-\infty}^{\infty} \sum_{q=1}^{\infty} \left[U_{rq}^{+(s)} \cdot e^{i\overline{\omega}_p t} \times \right. \\ \left. \times \left(-\frac{(\varepsilon - \varepsilon_1) A_{pq} \overline{\omega}_p}{2\varepsilon_1(\omega_{rs} - \overline{\omega}_p)} - \frac{\Delta \varepsilon A_{pq} \overline{\omega}_p e^{i\varphi_0}}{4\varepsilon_1((\omega_{r,s+1}) - \overline{\omega}_p)} - \right. \right. \\ \left. \left. - \frac{\Delta \varepsilon A_{pq} \overline{\omega}_p e^{-i\varphi_0}}{4\varepsilon_1((\omega_{r,s-1}) - \overline{\omega}_p)} + \frac{i \cdot \sigma A_{pq}}{2\varepsilon_0 \varepsilon_1 (\omega_{rs} - \overline{\omega}_p)} + \right. \right. \\ \left. \left. + \frac{i \cdot \Delta \sigma A_{pq} e^{i\varphi_0}}{4\varepsilon_0 \varepsilon_1 (\omega_{r,s+1} - \overline{\omega}_p)} + \frac{i \cdot \Delta \sigma A_{pq} e^{-i\varphi_0}}{4\varepsilon_0 \varepsilon_1 (\omega_{r,s-1} - \overline{\omega}_p)} \right) + \right. \\ \left. + U_{rq}^{-s} \cdot e^{i\overline{\omega}_p t} \frac{\varepsilon - \varepsilon_1 A_{pq} \overline{\omega}_p}{2\varepsilon_1 \omega_{rs} + \overline{\omega}_p} + \frac{\Delta \varepsilon A_{pq} \overline{\omega}_p e^{i\varphi_0}}{4\varepsilon_1 \omega_{r,s-1} - \overline{\omega}_p} + \right. \\ \left. + \frac{\Delta \varepsilon A_{pq} \overline{\omega}_p e^{-i\varphi_0}}{4\varepsilon_1 \omega_{r,s+1} + \overline{\omega}_p} - \frac{i \cdot \sigma A_{pq}}{2\varepsilon_0 \varepsilon_1 \omega_{rs} + \overline{\omega}_p} - \right. \\ \left. - \frac{i \cdot \Delta \sigma A_{pq} e^{i\varphi_0}}{4\varepsilon_0 \varepsilon_1 \omega_{r,s-1} + \overline{\omega}_p} - \frac{i \cdot \Delta \sigma A_{pq} e^{-i\varphi_0}}{4\varepsilon_0 \varepsilon_1 \omega_{r,s+1} + \overline{\omega}_p} \right] = 0. \quad (10)$$

Для слагаемых пропорциональных $e^{-i\overline{\omega}_p t}$ из (9) получаем:

$$\sum_{s=-\infty}^{\infty} \sum_{q=1}^{\infty} \left[U_{rq}^{+(s)} \cdot e^{-i\overline{\omega}_p t} \times \right. \\ \left. \times \left(\frac{(\varepsilon - \varepsilon_1) A_{pq} \overline{\omega}_p}{2\varepsilon_1(\omega_{r,s} + \overline{\omega}_p)} - \frac{\Delta \varepsilon A_{pq} \overline{\omega}_p e^{i\varphi_0}}{4\varepsilon_1((\omega_{r,s+1}) + \overline{\omega}_p)} - \right. \right. \\ \left. \left. - \frac{\Delta \varepsilon A_{pq} \overline{\omega}_p e^{-i\varphi_0}}{4\varepsilon_1((\omega_{r,s-1}) + \overline{\omega}_p)} + \frac{i \cdot \sigma A_{pq}}{2\varepsilon_0 \varepsilon_1 (\omega_{rs} + \overline{\omega}_p)} + \right. \right. \\ \left. \left. + \frac{i \cdot \Delta \sigma A_{pq} e^{i\varphi_0}}{4\varepsilon_0 \varepsilon_1 (\omega_{r,s+1} + \overline{\omega}_p)} + \frac{i \cdot \Delta \sigma A_{pq} e^{-i\varphi_0}}{4\varepsilon_0 \varepsilon_1 (\omega_{r,s-1} + \overline{\omega}_p)} \right) + \right. \\ \left. + U_{rq}^{-(s)} \cdot e^{-i\overline{\omega}_p t} \left(-\frac{(\varepsilon - \varepsilon_1) A_{pq} \overline{\omega}_p}{2\varepsilon_1(\omega_{rs} - \overline{\omega}_p)} - \frac{\Delta \varepsilon A_{pq} \overline{\omega}_p e^{i\varphi_0}}{4\varepsilon_1((\omega_{r,s-1}) - \overline{\omega}_p)} - \right. \right. \\ \left. \left. - \frac{\Delta \varepsilon A_{pq} \overline{\omega}_p e^{-i\varphi_0}}{4\varepsilon_1((\omega_{r,s+1}) - \overline{\omega}_p)} - \frac{i \cdot \sigma A_{pq}}{2\varepsilon_0 \varepsilon_1 (\omega_{rs} - \overline{\omega}_p)} - \right. \right. \\ \left. \left. - \frac{i \cdot \Delta \sigma A_{pq} e^{i\varphi_0}}{4\varepsilon_0 \varepsilon_1 (\omega_{r,s-1} - \overline{\omega}_p)} - \frac{i \cdot \Delta \sigma A_{pq} e^{-i\varphi_0}}{4\varepsilon_0 \varepsilon_1 (\omega_{r,s+1} - \overline{\omega}_p)} \right) \right] = 0. \quad (11)$$

Выражения (10) и (11) связывают амплитуды поля в невозмущенном резонаторе с амплитудами гармоник в возмущенном.

Из (8) для коэффициентах при $e^{\pm(i\omega_{r,s}t)}$ получаем следующую систему алгебраических уравнений относительно неизвестных амплитуд $U_{r,q}^{\pm(s)}$:

$$\sum_{q=1}^{\infty} \left(\frac{A_{pq}(\varepsilon - \varepsilon_1)}{\varepsilon_1} \cdot \frac{\omega_{r,s}^2}{(\omega_{r,s}^2 - \overline{\omega}_p^2)} - \right. \\ \left. - i \cdot \frac{A_{pq}}{\varepsilon_0 \varepsilon_1} \cdot \frac{\sigma \omega_{r,s}}{(\omega_{r,s}^2 - \overline{\omega}_p^2)} + \delta_{pq} \right) U_{r,q}^{\pm(s)} + \\ + \left(\frac{A_{pq} \Delta \varepsilon}{2\varepsilon_1} \cdot \frac{(\omega_{r,s})^2}{(\omega_{r,s})^2 - \overline{\omega}_p^2} - i \cdot \frac{A_{pq} \Delta \sigma}{2\varepsilon_0 \varepsilon_1} \cdot \frac{(\omega_{r,s})}{(\omega_{r,s})^2 - \overline{\omega}_p^2} \right) \times \\ \times (U_{r,q}^{\pm(s-1)} e^{i\varphi_0} + U_{r,q}^{\pm(s+1)} e^{-i\varphi_0}) = 0. \quad (12)$$

где $p = 1, 2, 3, \dots; s = 0, \pm 1, \pm 2, \dots$

Рассмотрим случай, малого заполнения резонатора, когда радиус стержня $R \ll a$ и $R \ll l$ (рис. 1) и будем считать, что для значений амплитуд модуляции выполняются условия $\Delta \varepsilon \ll \varepsilon$, $\Delta \sigma \ll \sigma$.

Индекс r в (12) не влияет на вид матричных коэффициентов и поэтому при решении системы его можно опустить. Для колебания типа квази TE_{101} система (12) принимает вид:

$$\begin{cases}
 \left(\frac{A_{11}\omega_{-1}}{\varepsilon_1(\omega_{-1}^2 - \omega_{11}^2)} \left[(\varepsilon - \varepsilon_1)\omega_{-1} - i\frac{\sigma}{\varepsilon_0} \right] + 1 \right) \cdot U^{(-1)} + \\
 + \frac{A_{11}\omega_{-1}}{2\varepsilon_1(\omega_{-1}^2 - \omega_{11}^2)} \left[\Delta\varepsilon\omega_{-1} - i\frac{\Delta\sigma}{\varepsilon_0} \right] \cdot U^{(0)} e^{-i\varphi_0} = 0, \\
 \left(\frac{A_{11}\omega_0}{\varepsilon_1(\omega_0^2 - \omega_{11}^2)} \left[(\varepsilon - \varepsilon_1)\omega_0 - i\frac{\sigma}{\varepsilon_0} \right] + 1 \right) \cdot U^{(0)} + \\
 + \frac{A_{11}\omega_0}{2\varepsilon_1(\omega_0^2 - \omega_{11}^2)} \left[\Delta\varepsilon\omega_0 - i\frac{\Delta\sigma}{\varepsilon_0} \right] \times \\
 \times (U^{(-1)} e^{i\varphi_0} + U^{(+1)} e^{-i\varphi_0}) = 0, \\
 \left(\frac{A_{11}\omega_1}{\varepsilon_1(\omega_1^2 - \omega_{11}^2)} \left[(\varepsilon - \varepsilon_1)\omega_1 - i\frac{\sigma}{\varepsilon_0} \right] + 1 \right) \cdot U^{(1)} + \\
 + \frac{A_{11}\omega_1}{2\varepsilon_1(\omega_1^2 - \omega_{11}^2)} \left[\Delta\varepsilon\omega_1 - i\frac{\Delta\sigma}{\varepsilon_0} \right] \cdot U^{(0)} e^{i\varphi_0} = 0.
 \end{cases} \quad (13)$$

Нетривиальные решения (13) получим при равенстве нулю определителя, составленного из коэффициентов при неизвестных (см. (14) в конце страницы).

ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Для резонатора сантиметрового диапазона, геометрические размеры (a, b, l) которого $23 \times 10 \times 25$ мм, с расположенным в центре резонатора цилиндрическим стержнем, найдем решение уравнения (14) относительно ω .

На рисунке 2 и 3 показана зависимость частоты основной гармоники колебания H_{101} от частоты модуляции для резонатора заполненного средой с параметрами $\varepsilon = 0,9$; $\varepsilon_1 = 1$; $\frac{\Delta\varepsilon}{\varepsilon} = 0,01$; $0,02$; $0,04$ без учета удельной проводимости с $R = 1$ мм.

На рисунках 2 и 3 частота основной гармоники нормирована к соответствующей собственной частоте резонатора в отсутствии модуляции.

Из графика на рисунке 2, видно, что модуляция диэлектрической проницаемости стержня приводит к изменению частоты основной гармоники. Причем, в области $\Omega > 2\omega_p$ наблюдается сдвиг частоты в сторону увеличения. Максимальное значение ω достигает на границе зоны неустойчивости и затем с дальнейшим

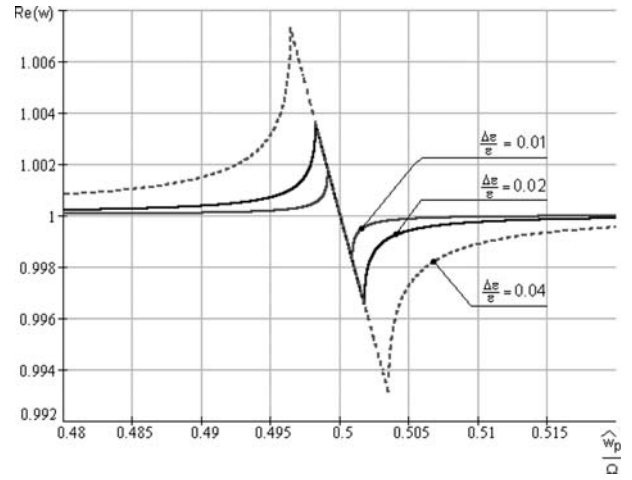


Рисунок 2 – Зависимость реальной части характеристического показателя от частоты модуляции

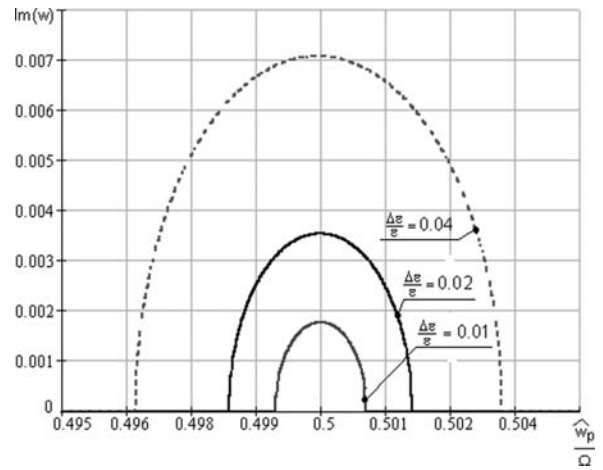


Рисунок 3 – Зависимость мнимой части характеристического показателя от частоты модуляции

уменьшением Ω также уменьшается, достигая минимума на другой границе зоны, но уже при $\Omega < 2\omega_p$. График на рисунке 3, где изображена мнимая часть частоты основной гармоники позволяет более наглядно определить границы зоны неустойчивости, так как именно по появлению комплексной составляющей, можно судить о том что наступает явление параметрического резонанса.

$$\begin{vmatrix}
 \frac{A_{11}\omega_{-1}}{\varepsilon_1(\omega_{-1}^2 - \omega_{11}^2)} \left[(\varepsilon - \varepsilon_1)\omega_{-1} - i\frac{\sigma}{\varepsilon_0} \right] + 1 & \frac{A_{11}\omega_{-1}e^{-i\varphi_0}}{2\varepsilon_1(\omega_{-1}^2 - \omega_{11}^2)} \left[\Delta\varepsilon\omega_{-1} - i\frac{\Delta\sigma}{\varepsilon_0} \right] & 0 \\
 \frac{A_{11}\omega_0e^{i\varphi_0}}{2\varepsilon_1(\omega_0^2 - \omega_{11}^2)} \left[\Delta\varepsilon\omega_0 - i\frac{\Delta\sigma}{\varepsilon_0} \right] & \frac{A_{11}\omega_0}{\varepsilon_1(\omega_0^2 - \omega_{11}^2)} \left[(\varepsilon - \varepsilon_1)\omega_0 - i\frac{\sigma}{\varepsilon_0} \right] + 1 & \frac{A_{11}\omega_0e^{-i\varphi_0}}{2\varepsilon_1(\omega_0^2 - \omega_{11}^2)} \left[\Delta\varepsilon\omega_0 - i\frac{\Delta\sigma}{\varepsilon_0} \right] \\
 0 & \frac{A_{11}\omega_1e^{i\varphi_0}}{2\varepsilon_1(\omega_1^2 - \omega_{11}^2)} \left[\Delta\varepsilon\omega_1 - i\frac{\Delta\sigma}{\varepsilon_0} \right] & \frac{A_{11}\omega_1}{\varepsilon_1(\omega_1^2 - \omega_{11}^2)} \left[(\varepsilon - \varepsilon_1)\omega_1 - i\frac{\sigma}{\varepsilon_0} \right] + 1
 \end{vmatrix} = 0. \quad (14)$$

На рисунке 4 показана зависимость частоты основной гармоники колебания H_{101} от частоты модуляции для резонатора с параметрами стержня $\varepsilon = 10,9$; $\varepsilon_1 = 1$; $\frac{\Delta\varepsilon}{\varepsilon} = 0,05$ с учетом удельной проводимости для $\sigma = 0,4$; $0,5$; $0,6$ см/м. Частота основной гармоники нормирована к соответствующей собственной частоте резонатора в отсутствие модуляции.

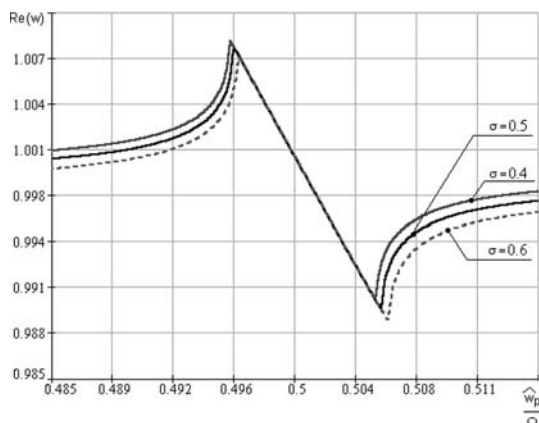


Рисунок 4 – Зависимость реальной части характеристического показателя от частоты модуляции

Графики рисунка 4 показывают, что изменение удельной проводимости стержня приводит к сдвигу частоты основной гармоники. Причем сдвиг происходит в область более высоких значений отношения ω_p/Ω .

ВЫВОДЫ

Краевая задача для определения электромагнитных полей в резонаторных системах, заполненных средой с изменяющейся во времени диэлектрической проницаемостью и проводимостью сведена к интегральному уравнению Фредгольма второго рода позволяющему произвести алгебраизацию задачи.

В предположении малого заполнения резонатора, когда радиус стержня $R \ll a$ и $R \ll l$. При выполнении условий $\Delta\varepsilon \ll \varepsilon$, $\Delta\sigma \ll \sigma$ получен определитель третьего порядка относительно квадрата характеристического

показателя, позволяющий исследовать свойства такой структуры.

Результаты численного анализа показали, что поле в резонаторе, частично заполненном нестационарной средой, в диапазоне частот накачки близкой к удвоенной частоте стационарного резонатора нарастает во времени (характеристический показатель ω является комплексным). При изменении параметров цилиндрического стержня определены его значения, представленные в виде графиков.

Рассмотренный алгоритм решения задачи позволяет численно исследовать зоны неустойчивости основного резонанса более высокого порядка в широкой области изменения частоты модуляции.

ПЕРЕЧЕНЬ ССЫЛОК

1. Мандельштам Л.И. Полное собрание трудов: В5–ИТ. – М.: Изд-во АН СССР, 1955. – Т. 3.
2. Люиссел У. Связанные и параметрические колебания в электронике. – М.: ИЛ, 1963. – 351 с.
3. Ярив А. Квантовая электроника и нелинейная оптика. – М.: Сов. радио, 1973. – 241 с.
4. Хижняк Н.А. Инженерные уравнения макроскопической электродинамики. – К.: Наукова думка, 1986. – 279 с.
5. Бондарев В.П. К устойчивости электромагнитных колебаний в резонаторе, заполненном нестационарной средой // Известия высших учебных заведений. Радиофизика. – 1986. – Т. 29. – № 4, с. 470–476.

Надійшла 02.04.2004
Після доробки 29.10.2004

Країнова задача для визначення електромагнітного поля у резонансній системі, котра включає циліндричну неоднорідність з змінною у часі діелектричною проникністю і питомою провідністю зведена до інтегрального рівняння. Для малих амплітуд модуляції отримане характеристичне рівняння для ТЕ коливань, з котрого можливо визначити поле поблизу й в області основного параметричного резонансу. При чисельному аналізі рішень знайдені зони нестійкості та значення характеристичного показнику для резонансу в широких межах зміни частоти модуляції.

The boundary problem level for determination of electromagnetic field in resonance system that includes cylinder inhomogeneity with the changing in time dielectric permittivity and specific conductivity is reduced to an integral equation. For small amplitudes of modulation the characteristic equation for TE oscillations was received and it enables to determine the field near and in the zone of the main parametric resonance. At numerical analysis of solutions the zones of instability and values of characteristic indexes in the wide zone of the modulation frequency alteration were found.

УДК 621.396.96

И. И. Зиненко, В. П. Пьянков, В. П. Чумаченко

СОГЛАСОВАНИЕ ПЯТИПЛЕЧЕГО Е-ПЛОСКОСТНОГО СОЕДИНЕНИЯ ПРЯМОУГОЛЬНЫХ ВОЛНОВОДОВ

Представлены результаты исследования возможности согласования пятиплечевого соединения прямоугольных волноводов путем увеличения расстояния между волноводами, а также внесением кругового цилиндра в область связи.

Для нахождения элементов матрицы рассеяния применен модифицированный метод произведения областей. Приведены и проанализированы графики для коэффициентов матрицы рассеяния рассматриваемого устройства.

ВВЕДЕНИЕ

В измерительной технике СВЧ находят широкое применение различного рода волноводные соединения. Для точности измерений и уменьшения потерь очень важно иметь устройства, которые согласованы по питающему волноводу в довольно широкой полосе рабочего диапазона. В настоящее время накоплен определенный теоретический и экспериментальный материал в сфере анализа свойств координатных волноводных n -разветвлений. Эти работы проведены различными методами, при различном уровне строгости и точности. В работах [1–3] выполнен анализ E -плоскостных пятиплечих ответвителей, применяемых в измерительных схемах миллиметрового диапазона. В частности, исследованы ответвители с цилиндрическими соединительными полостями разных диаметров.

Целью данной работы является изучение возможности согласования E -плоскостного пятиплечего ответвителя путем расширения расстояния между волноводами за счет изменения параметра d , а также внесением кругового цилиндра радиуса R в область связи, как показано на рис. 1. Возбуждение устройства производится со стороны волновода 1 волной типа H_{10} .

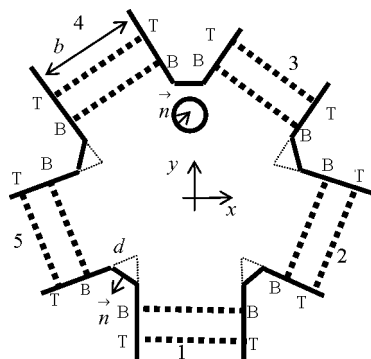


Рисунок 1

Результаты, опубликованные в данной работе, будут полезны при разработке как реальных узлов на основе прямоугольных волноводов, так и соединений других линий передачи, для анализа которых применяются волноводные модели, например, полосковых линий.

Численный анализ выполнен на основе метода произведения областей (ПО) [4]. Погрешность вычислений коэффициентов матрицы рассеяния S_{ik} не превышает 0,005.

ПОСТАНОВКА ЗАДАЧИ

Рассмотрим E -плоскостной волноводный трансформатор, соединительная полость которого имеет кусочно-линейный граничный контур и может иметь многоугольные и круговые вставки (рис. 1). Все поверхности полагаем идеально проводящими. В регулярных частях волноводов на некотором расстоянии от их концов введем отсчетные плоскости T – T . Положим, что по одно-

му из волноводов (питающему) к соединительной полости набегают возбуждающая волна типа H_{10} , амплитуда и фаза которой в плоскости T – T известны. При $a = \infty$, где a – размер широкой стенки волновода, имеет место волна типа TEM . Задача сводится к нахождению компоненты u напряженности магнитного поля, перпендикулярной плоскости рис. 1, удовлетворяющей уравнению Гельмгольца

$$\Delta u + k^2 u = 0, \quad (1)$$

где $k = 2\pi \sqrt{\frac{1}{\lambda^2} - \left(\frac{1}{2a}\right)^2}$, λ – длина волны в свободном пространстве, а также однородному граничному условию Неймана на проводящих поверхностях, условию интегрируемости плотности энергии поля в любой ограниченной области и условиям излучения на бесконечности.

Весь граничный контур соединительной полости с прилегающими к ней участками волноводов до плоскостей T – T разобьем на элементы $\{S_i\}$: прямолинейные отрезки и окружности. В пространстве эти элементы представляют собой полосы и круговые цилиндры. Для их обозначения используем сквозную нумерацию. Пусть I_P – множество номеров полос проводящих поверхностей; I_H – множество номеров поперечных сечений волноводов (плоскостей T – T); I_C – множество номеров круговых цилиндров; $I_R = I_P \cup I_H$; $I = I_R \cup I_C$; \vec{n} – нормаль, направление которой указано на рис. 1; S_i^+ , S_i^- – внешняя и внутренняя стороны элемента S_i .

На указанных элементах $\{S_i, i \in I\}$ неизвестная компонента u должна удовлетворять следующим граничным условиям:

$$\left. \frac{\partial u}{\partial n} \right|_{S_i^+} = 0, i \in I_P \cup I_C; \quad \left. \frac{\partial u}{\partial n} \right|_{S_i^+} = \left. \frac{\partial U_i}{\partial n} \right|_{S_i}, i \in I_H, \quad (2)$$

где U_i – искомая компонента поля в регулярной части соответствующего волновода.

Согласно методу ПО функцию $u(r)$ будем искать в виде

$$u(r) = \sum_{i \in I} u_i(r). \quad (3)$$

Здесь каждая функция $u_i(r)$ удовлетворяет однородному уравнению Гельмгольца всюду вне своего i -го элемента и условию излучения на бесконечности. В силу наложенных на $u_i(r)$ требований справедливы интегральные равенства:

$$u_i(r) = \int_{S_i^+} \frac{\partial u_i(r')}{\partial n'} G_i(r, r') dS' \forall r \notin S_i, i \in I_C; \quad (4)$$

$$u_i(r) = \int_{S_i^+} \frac{\partial u_i(r')}{\partial n'} G_i(r, r') dS' - \\ - u_i(r) = \int_{S_i^-} \frac{\partial u_i(r')}{\partial n'} G_i(r, r') dS' \forall r \notin S_i, i \in I_R; \quad (5)$$

где $G_i(r, r')$ – функция Грина, удовлетворяющая однородному граничному условию Неймана на соответствующей полосе или круговом цилиндре. При $i \in I_C$ функция $G_i(r, r')$ – известный ряд по функциям Ханкеля. При $i \in I_R$ функция $G_i(r, r')$ представляет собой ряд по функциям Матье [5].

Так как нас интересует поле внутри соединительной полости, мы имеем некоторую свободу выбора граничных условий на сторонах $\{S_i^-, i \in I_R\}$. Доопределим граничные условия на сторонах S_i^- ($i \in I_R$) следующим образом:

$$\left. \frac{\partial u}{\partial n} \right|_{S_i^-} = - \left. \frac{\partial u}{\partial n} \right|_{S_i^+}, i \in I_R. \quad (6)$$

Тогда уравнение (5) примет вид

$$u_i(r) = 2 \int_{S_i^+} \frac{\partial u_j(r')}{\partial n'} G_i(r, r') dS' \forall r \notin S_i, i \in I_R, \quad (7)$$

и выражение для $G_i(r, r')$ ($i \in I_R$) будет содержать только ряд по четным функциям Матье.

Для каждого волновода в силу непрерывности электромагнитного поля в однородной среде можно записать:

$$U_i(r) = u(r) \forall r \in B-B, i \in I_H, \quad (8)$$

где $B-B$ – некоторая секущая плоскость, расположенная в регулярной части волновода между плоскостью $T-T$ и плоскостью раскрытия волновода в соединительную полость (см. рис. 1).

Выразив $\left. \frac{\partial u}{\partial n} \right|_{S_i^+}$ согласно граничным условиям (2) и подставив в (4), (7), получим:

$$u_i(r) = \int_{S_i^+} \sum_{j \neq i} \frac{\partial u_j(r')}{\partial n'} G_i(r, r') dS' \forall r \notin S_i, i \in I_C; \quad (9)$$

$$u_i(r) = -2 \int_{S_i^+} \sum_{j \neq i} \frac{\partial u_j(r')}{\partial n'} G_i(r, r') dS' \forall r \notin S_i, i \in I_P; \quad (10)$$

$$u_i(r) = -2 \int \left(\frac{\partial U_i(r')}{\partial n'} - \sum_{j \neq i} \frac{\partial u_j(r')}{\partial n'} \right) \times \\ \times G_i(r, r') dS' \forall r \notin S_i, i \in I_H; \quad (11)$$

где в подынтегральных выражениях суммирование ведется по всем $j \neq I$, кроме $j \neq i$. Исходя из (3) и (8), получим

$$U_i(r) = \sum_{j \in I} u_j(r) \forall r \in B-B, i \in I_H. \quad (12)$$

Уравнения (9)–(12) представляют собой систему интегро-дифференциальных уравнений относительно функций u_i ($i \in I$) и U_i ($i \in I_H$). Функцию U_i ищем в виде разложения по собственным волнам волновода. Функ-

цию u_i при $i \in I_C$ ищем в виде разложения по функциям Ханкеля, а при $i \in I_R$ – в виде разложения по четным функциям Матье. Исходя из системы уравнений (9)–(12) получим бесконечную систему линейных алгебраических уравнений (СЛАУ) относительно коэффициентов упомянутых разложений, которая может быть решена методом редукции [6]. Выбор четных функций Матье на практике обеспечивает хорошую сходимость метода (относительно небольшой порядок усечения соответствующих рядов).

РЕЗУЛЬТАТЫ ЧИСЛЕННОГО АНАЛИЗА

Описанный алгоритм реализован в виде программы на языке Фортран. Программа предназначена для расчета матрицы рассеяния и других электродинамических параметров.

Вначале рассмотрим модифицированное пятиплечее соединение волноводов без кругового цилиндра в области связи. На рис. 2 показаны зависимости модуля коэффициента отражения $|S_{11}|$ от частотного параметра $\chi^* = \frac{bk}{2\pi}$, где b – размер узкой стенки волноводов. Штриховой линией изображена зависимость немодифицированного пятиплечевого соединения ($d = 0$). Принята следующая нумерация кривых: 1 – $d = 0,65b$; 2 – $d = 0,55b$; 3 – $d = 0,5b$; 4 – $d = 0,45b$.

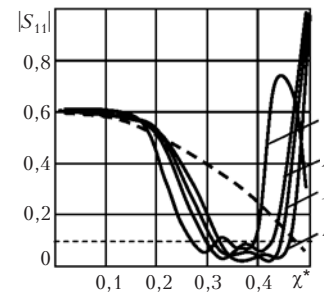


Рисунок 2

Из рис. 2 видно, что расширение области связи (увеличение параметра d) приводит к значительному снижению уровня отраженной волны и позволяет достичь высокой степени согласования рассматриваемого устройства в достаточно широком частотном интервале. Так при $d = 0,45b$ полоса согласования по уровню $|S_{11}| = 0,1$ составляет 27%, а по уровню $|S_{11}| = 0,2$ – 33%. При увеличении параметра d до значения $0,55b$ ширина полосы согласования остается неизменной, а ее среднее значение смещается к центру рабочего диапазона. При дальнейшем увеличении параметра d до $0,65b$ уровень отраженной волны при среднем значении полосы согласования несколько увеличивается и начинает превышать уровень $|S_{11}| = 0,1$. Дальнейшее увеличение параметра d приводит к ухудшению согласования модифицированного пятиплечевого соединения.

На рис. 3, а, б показаны зависимости модулей коэффициентов передачи $|S_{12}|$ (рис. 3, а) и $|S_{13}|$ (рис. 3, б) от частотного параметра χ^* . Штриховой линией изображены зависимости без расширения области связи ($d=0$). Видно, что в полосе согласования коэффициенты передачи близки к значению 0,5. Это важно для применения рассматриваемого устройства в различных схемах СВЧ, в том числе измерительных, так как узел работает как идеальный делитель.

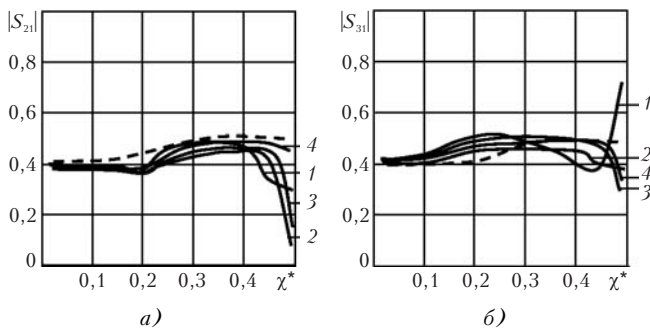


Рисунок 3

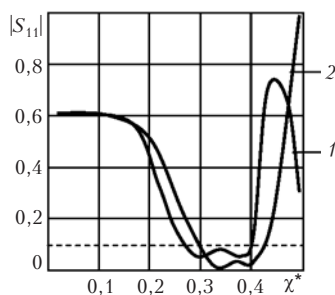


Рисунок 4

На рис. 4 изображены зависимости модуля коэффициента отражения $|S_{11}|$ от частотного параметра χ^* при наличии кругового цилиндра в области связи волноводов, как показано на рис. 1. Внесение кругового цилиндра в область связи позволяет снизить уровень коэффициента отражения $|S_{11}|$. Необходимо отметить, что наилучшего согласования удастся достичь при оптимальных значениях геометрических параметров: радиуса R и координат центра кругового цилиндра (X, Y). При помощи численного анализа такие оптимальные параметры были получены. На рис. 4 представлены кривые при наилучшем согласовании рассматриваемого модифицированного устройства: 1 – $d = 0,65b, R = 0,12b, Y = 0,9b, X = 0$; 2 – $d = 0,55b, R = 0,12b, Y = 0,96b, X = 0$. Так при $d = 0,65b$ без цилиндра уровень коэффициента отражения в середине полосы согласования превосходит 0,1 (кривая 1 рис. 2), с круговым же цилиндром этот уровень удастся снизить до 0,095 (кривая 1 рис. 4), а при $d = 0,55b$ без цилиндра уровень коэффициента

отражения в середине полосы согласования достигает 0,1 (кривая 2 рис. 2), с круговым же цилиндром этот уровень снижается до 0,055 (кривая 2 рис. 4).

ВЫВОДЫ

Приведенные в данной работе результаты исследований зависимостей коэффициентов отражения и прохождения от геометрического параметра d (см. рис. 1) свидетельствуют, что для согласования пятиплечего соединения в E -плоскости целесообразно использовать расширение области связи, изменяя параметр d в пределах $0,4b - 0,6b$. При этом полоса согласования по уровню $|S_{11}| = 0,1$ в среднем достигает 30%. Дальнейшее увеличение параметра d приводит к ухудшению согласования пятиплечего соединения. Внесение в область связи кругового цилиндра позволяет улучшить согласование при изменении d в указанных пределах, более того, расширить эти пределы: $0,4b - 0,7b$. Такое увеличение значений параметра d позволяет смещать полосу согласования в центр рабочего диапазона частот.

ПЕРЕЧЕНЬ ССЫЛОК

1. E. R. B. Hansson, G. P. Riblet, "An ideal 6-port network consisting of a matched reciprocal lossless five-port and a perfect directional coupler", IEEE Trans. Microwave Theory Tech., Vol. MTT-31, pp. 284-288, Mar., 1983.
2. G. P. Riblet, E. R. B. Hansson, "Aspects of the Calibration of a Single Six-Port Using a Load and Offset Reflection Standards", IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-30, pp. 2120-2125, Dec. 1982.
3. G. P. Riblet, "An E-plane coupled matched waveguide 5-port for making 6-port measurements over a full waveguide band", in IEEE Int. Microwave Symp. Dig., pp. 372-374, 1983.
4. Чумаченко В. П. Решение задач дифракции на многоугольной цилиндрической поверхности методом произведения областей // Докл. АН УССР. Сер. А. 1989. № 7. С. 73-76.
5. Засовенко В. Г., Чумаченко В. П. Функция Грина уравнения Гельмгольца для идеально проводящего эллиптического цилиндра // Электродинамика и физика СВЧ. Днепропетровск, 1983. С. 52-56.
6. Чумаченко В. П. К обоснованию одного метода решения двумерных задач дифракции электромагнитных волн на многоугольных идеально проводящих структурах. // Радиотехника и электроника. 1988. Т. 33. № 8. С. 1600-1609.

Надійшла 17.05.2004

Наведено результати дослідження можливості узгодження п'ятиплечого з'єднання прямокутних хвильоводів шляхом збільшення відстані між хвильоводами, а також внесення кругового цилиндра до з'єднувальної порожнини. Елементи матриці розсіювання були розраховані за допомогою модифікованого методу добутку областей. Наведені та проаналізовані графіки для коефіцієнтів матриці розсіювання пристрою, що розглядається.

Matching properties of the connecting cavity of an E-plane waveguide 5-port are studied. The cavity contains a capacitive post and is bounded by the waveguide apertures and by the walls of the truncated reentering angles. The analysis is performed on the basis of the modified domain-product technique. Graphs of the respective frequency dependencies of scattering characteristics are presented and discussed.

Л. М. Карпуков, Р. Д. Пулов, С. Н. Романенко

МОДЕЛИРОВАНИЕ МНОГОСЛОЙНЫХ МИКРОПОЛОСКОВЫХ СТРУКТУР В КВАЗИДИНАМИЧЕСКОМ ПРИБЛИЖЕНИИ

Предложен метод составления декомпозиционных схем многослойных плоскостойких структур при расчете в квазидинамическом приближении функций Грина краевых задач электростатики. Разработан метод аналитического решения дисперсионного уравнения для вычисления динамической поправки к квазистатическим значениям эффективной диэлектрической проницаемости микрополосковых линий. Представлены результаты моделирования дисперсии в двухслойных конструкциях микрополосковых линий.

ВВЕДЕНИЕ

Основой инженерного проектирования микрополосковых устройств СВЧ являются квазистатические модели микрополосковых передающих линий [1]. Необходимость повышения качества проектных работ на этапе многовариантного анализа при синтезе структур проектируемых микрополосковых устройств требует привлечения более строгих электродинамических моделей и методов с возможностью анализа дисперсионных характеристик микрополосковых линий. Однако в строгой электродинамической постановке решение задачи моделирования отличается высокой трудоемкостью и повышенными требованиями к вычислительным ресурсам. Сложность и трудоемкость задач электродинамического моделирования еще более возрастают при разработке многослойных конструкций микрополосковых устройств.

Повысить достоверность результатов квазистатического моделирования при незначительном увеличении объема вычислений можно путем введения динамических поправок к вычисляемым на нулевой частоте значениям эффективной диэлектрической проницаемости исследуемых конструкций микрополосковых линий.

Эффективным способом нахождения динамических поправок является составление и решение дисперсионного уравнения микрополосковой линии в квазидинамическом приближении [2, 3, 4]. Квазидинамическое моделирование микрополосковых линий, реализуемых на однослойной подложке, проведено в [4, 5], где получены простые аналитические соотношения, обеспечивающие расчет эффективной диэлектрической проницаемости в широком диапазоне частот. Целью настоящей работы является обобщение методики квазидинамического моделирования, разработанной в [4, 5] для однослойной подложки, и ее использование для разработки дисперсионных моделей многослойных конструкций микрополосковых линий.

МЕТОДИКА КВАЗИДИНАМИЧЕСКОГО МОДЕЛИРОВАНИЯ МНОГОСЛОЙНЫХ СТРУКТУР

На рис. 1 приведена многослойная структура подложки микрополосковой линии. Структура состоит из n диэлектрических слоев с различными толщинами h_i и относительными диэлектрическими проницаемостями ϵ_i . Вдоль осей x , y слои имеют бесконечные размеры. Сверху и снизу структура может быть ограничена металлическими экранами. Проводники микрополосковых линий, шириной w_i , располагаются на границах раздела диэлектрических сред.

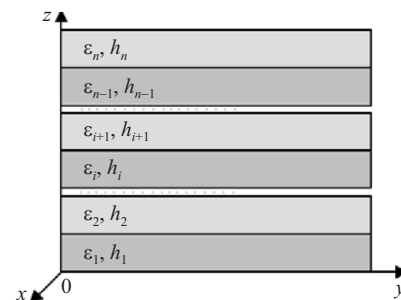


Рисунок 1 – Структура поперечного сечения n -слойной подложки

Для упрощения анализа пренебрежем поперечной составляющей поверхностной плотности тока на микрополосковой линии. Продольную составляющую тока представим в виде

$$J_x(x, y) = f(y) e^{-jk_x x}, \quad (1)$$

где $k_x = k_0 \sqrt{\epsilon_{\text{эф}}}$ – постоянная распространения волны вдоль продольной оси линии, k_0 – волновое число свободного пространства, $\epsilon_{\text{эф}}$ – эффективная диэлектрическая проницаемость линии.

Приняв распределение $f(y)$ равномерным, составим дисперсионное уравнение линии:

$$\int_{w_i} dy \int_{w_i} Z_{xx}(y, y_0) dy_0 = 0, \quad (2)$$

где $Z_{xx} = -j\omega G_{xx} + \frac{-jk_x}{j\omega\mu_0\epsilon_0\epsilon_r} \left(-jk_x G_{xx} + \frac{\partial G_{zx}}{\partial z} \right)$.

Здесь G_{xx} , G_{zx} – компоненты тензорной функции Грина слоистой структуры.

Зависимости для компонент тензора Грина могут быть получены по их спектральным представлениям [5]. Для этого выразим в спектральной области функцию $\partial \hat{G}_{zx} / \partial z$ через вспомогательную функцию \hat{G}_{vx} [6]:

$$\frac{\partial \hat{G}_{zx}}{\partial z} = \frac{-jk_x}{\alpha} \left(\frac{\partial \hat{G}_{vx}}{\partial z} - k_{z_r}^2 \hat{G}_{vx} \right), \quad (3)$$

где $\alpha = \sqrt{k_x^2 + k_y^2}$, $k_{z_r} = \sqrt{\alpha^2 - k_r^2}$.

С учетом (3) спектральная функция поверхностного импеданса из (1) приобретет следующий вид:

$$\hat{Z}_{xx} = -j\omega \hat{G}_{xx} + \frac{(-jk_x)^2}{j\omega\mu_0\epsilon_0\epsilon_r} \left[\left(1 - \frac{k_z^2}{\alpha^2} \right) \hat{G}_{xx} + \frac{1}{\alpha^2} \frac{\partial \hat{G}_{vx}}{\partial z} \right]. \quad (4)$$

В квазидинамическом приближении принимается [2, 3, 4]:

$$\alpha \approx k_{z_r}, \quad r = 0, 1, 2, \dots, l. \quad (5)$$

Поэтому функция (4) существенно упрощается:

$$\hat{Z}_{xx} = -j\omega \hat{G}_{xx} + \frac{(-jk_x)^2}{j\omega\mu_0\epsilon_0\epsilon_r} \frac{1}{k_z^2} \frac{\partial \hat{G}_{vx}}{\partial z}. \quad (6)$$

Для нахождения компонент \hat{G}_{xx} , \hat{G}_{vx} воспользуемся декомпозиционными схемами многослойной структуры, которые представим ориентированными графами, моделирующими распределения касательной A_r и нормальной A_z составляющих векторного электродинамического потенциала, возбуждаемых касательной составляющей тока [6, 7]. При построении схем коэффициенты отражения от разделов сред определим для полей A_r , A_z в соответствии с приближением (5):

$$\Gamma_{\tau_i} = \frac{k_{z_i} - k_{z_{i+1}}}{k_{z_i} + k_{z_{i+1}}} \approx 0, \quad (7)$$

$$\Gamma_{z_i} = \frac{k_{z_i}/\epsilon_i - k_{z_{i+1}}/\epsilon_{i+1}}{k_{z_i}/\epsilon_i + k_{z_{i+1}}/\epsilon_{i+1}} \approx -\Gamma_{v_i} = \frac{\epsilon_{i+1} - \epsilon_i}{\epsilon_{i+1} + \epsilon_i}. \quad (8)$$

Здесь Γ_v – коэффициент отражения для электростатического потенциала.

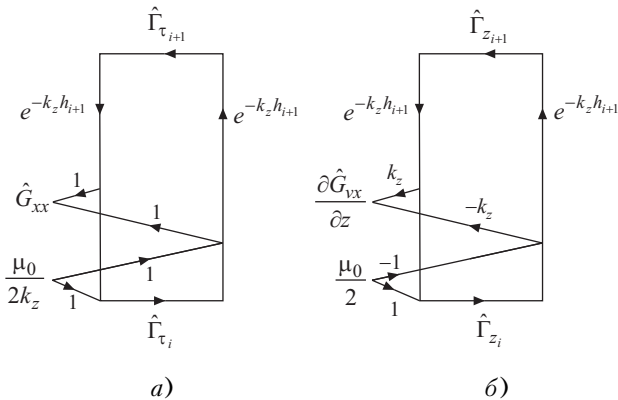


Рисунок 2 – Ориентированные графы для функций \hat{G}_{xx} (а), и $\partial \hat{G}_{vx} / \partial z$ (б)

На рис. 2 приведены графы, эквивалентные рассматриваемой многослойной структуре, моделирующие в квазидинамическом приближении составляющие функции поверхностного импеданса при расположении точечного источника тока на границе раздела i -й и $i+1$ -й сред. Символами $\hat{\Gamma}_{\tau_i}$, $\hat{\Gamma}_{z_i}$ и $\hat{\Gamma}_{\tau_{i+1}}$, $\hat{\Gamma}_{z_{i+1}}$ на рисунке обозначены коэффициенты отражений, эквивалентные слоям, прилегающим к $i+1$ -му слою.

Используя передаточные функции графов, получим

$$\hat{Z}_{xx} = -j\omega\mu_0 \frac{\hat{K}_{xx}}{2k_z} + \frac{(-jk_x)^2}{j\omega\epsilon_0\epsilon_{i+1}} \frac{\hat{K}_{vx}}{2k_z}. \quad (9)$$

Здесь

$$\hat{K}_{xx} = \frac{1 + \hat{\Gamma}_{\tau_i} + \hat{\Gamma}_{\tau_{i+1}} e^{-2k_z h_{i+1}} + \hat{\Gamma}_{\tau_i} \hat{\Gamma}_{\tau_{i+1}} e^{-2k_z h_{i+1}}}{1 - \hat{\Gamma}_{\tau_i} \hat{\Gamma}_{\tau_{i+1}} e^{-2k_z h_{i+1}}}, \quad (10)$$

$$\hat{K}_{vx} = \frac{1 - \hat{\Gamma}_{z_i} - \hat{\Gamma}_{z_{i+1}} e^{-2k_z h_{i+1}} + \hat{\Gamma}_{z_i} \hat{\Gamma}_{z_{i+1}} e^{-2k_z h_{i+1}}}{1 - \hat{\Gamma}_{z_i} \hat{\Gamma}_{z_{i+1}} e^{-2k_z h_{i+1}}}. \quad (11)$$

На основании приближения (7) коэффициенты отражения в (10) будут иметь следующий вид:

$$\left. \begin{aligned} \hat{\Gamma}_{\tau_i} &= \Gamma_{\tau_1} e^{-2k_z (h_1 + h_2 + \dots + h_i)}, \\ \hat{\Gamma}_{\tau_{i+1}} &= \Gamma_{\tau_{n+1}} e^{-2k_z (h_{i+2} + h_{i+3} + \dots + h_n)}, \end{aligned} \right\} \quad (12)$$

где Γ_{τ_1} , $\Gamma_{\tau_{n+1}}$ равны нулю для свободного пространства, или минус единице – для металлического экрана.

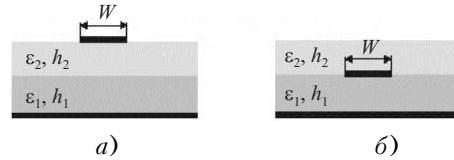


Рисунок 3 – Примеры двухслойных структур

Нахождение коэффициентов отражения в (11) может быть выполнено с помощью метода, разработанного для расчета в численно-символьном виде функций Грина многослойных диэлектрических сред [8]. Например, для двухслойной структуры на рис. 3, ограниченной снизу металлическим экраном, эквивалентные коэффициенты отражения будут составлены в следующем виде:

$$\left. \begin{aligned} \hat{\Gamma}_{z_2} &= \frac{-\Gamma_{z_3} + \Gamma_{z_1} \Gamma_{z_2} \Gamma_{z_3} e^{-2k_z h_1} - \Gamma_{z_2} e^{-2k_z h_2} + \Gamma_{z_1} e^{-2k_z (h_1 + h_2)}}{1 - \Gamma_{z_1} \Gamma_{z_2} e^{-2k_z h_1} + \Gamma_{z_2} \Gamma_{z_3} e^{-2k_z h_2} - \Gamma_{z_1} \Gamma_{z_3} e^{-2k_z (h_1 + h_2)}}, \\ \hat{\Gamma}_{z_3} &= 0 \end{aligned} \right\} \quad (13)$$

– для структуры на рис. 3, а;

$$\left. \begin{aligned} \hat{\Gamma}_{z_2} &= \frac{-\Gamma_{z_2} + \Gamma_{z_1} e^{-2k_z h_1}}{1 - \Gamma_{z_1} \Gamma_{z_2} e^{-2k_z h_1}}, \\ \hat{\Gamma}_{z_3} &= \Gamma_{z_3} \end{aligned} \right\} \quad (14)$$

– для структуры на рис. 3, б.

Здесь $\Gamma_{z_3} = \frac{1-\varepsilon_2}{1+\varepsilon_2}$, $\Gamma_{z_2} = \frac{\varepsilon_2-\varepsilon_1}{\varepsilon_2+\varepsilon_1}$, $\Gamma_{z_1} = 1$.

На основании (9)–(14) функции поверхностного импеданса двухслойной структуры запишутся в спектральной области следующим образом:

$$\hat{Z}_{xx2} = \frac{-j\omega\mu_0}{2k_z} \left(1 - e^{-2k_z(h_1+h_2)}\right) + \frac{k_x^2}{j\omega\varepsilon_0\varepsilon_2 2k_z} (1 + \Gamma_{z_2}) \times \frac{1 - e^{-2k_z h_1} - \Gamma_{z_3} e^{-2k_z h_2} + \Gamma_{z_3} e^{-2k_z(h_1+h_2)}}{1 - \Gamma_{z_2} e^{-2k_z h_1} + \Gamma_{z_2} \Gamma_{z_3} e^{-2k_z h_2} - \Gamma_{z_3} e^{-2k_z(h_1+h_2)}} \quad (15)$$

– при расположении точечного источника тока в сечении $z = h_1$;

$$\hat{Z}_{xx2} = \frac{-j\omega\mu_0}{2k_z} \left(1 - e^{-2k_z(h_1+h_2)}\right) + \frac{k_x^2}{j\omega\varepsilon_0 2k_z} (1 + \Gamma_{z_3}) \times \frac{1 - \Gamma_{z_2} e^{-2k_z h_1} + \Gamma_{z_2} e^{-2k_z h_2} - e^{-2k_z(h_1+h_2)}}{1 - \Gamma_{z_2} e^{-2k_z h_1} + \Gamma_{z_2} \Gamma_{z_3} e^{-2k_z h_2} - \Gamma_{z_3} e^{-2k_z(h_1+h_2)}} \quad (16)$$

– при расположении точечного источника тока в сечении $z = h_1 + h_2$.

По этим спектральным представлениям составим по методу, изложенному в [8], оригиналы искомых функций:

$$Z_{xx1}(y, y_0) = \frac{-\omega\mu_0}{4} \left[H_0^{(2)}(\gamma r_0) - H_0^{(2)}(\gamma r_{2n_1}) \right] + \frac{k_x^2}{4\omega\varepsilon_0} \frac{2}{\varepsilon_1 + \varepsilon_2} K_{vx1}(y, y_0), \quad (17)$$

$$Z_{xx2}(y, y_0) = \frac{-\omega\mu_0}{4} \left[H_0^{(2)}(\gamma r_0) - H_0^{(2)}(\gamma r_{2n_1+2n_2}) \right] + \frac{k_x^2}{4\omega\varepsilon_0} \frac{2}{1 + \varepsilon_2} K_{vx2}(y, y_0). \quad (18)$$

Здесь $\gamma = \sqrt{k_0^2 - k_x^2}$, $r_m = \sqrt{(y - y_0)^2 + (2mh)^2}$, $H_0^{(2)}$ – функция Ханкеля.

Функции K_{vx} вычисляются по рекуррентным формулам [8]:

$$K_{vx1}(mh) = \left[H_0^{(2)}(\gamma r_m) - H_0^{(2)}(\gamma r_{m+2n_1}) - \Gamma_{z_3} H_0^{(2)}(\gamma r_{m+2n_2}) \right] + \Gamma_{z_3} H_0^{(2)}(\gamma r_{m+2n_1+2n_2}) + \Gamma_{z_2} K_{vx1}(2n_1h + mh) - \Gamma_{z_2} \Gamma_{z_3} K_{vx1}(2n_2h + mh) + \Gamma_{z_3} K_{vx1}(2n_1h + 2n_2h + mh),$$

$$K_{vx2}(mh) = \left[H_0^{(2)}(\gamma r_m) - \Gamma_{z_2} H_0^{(2)}(\gamma r_{m+2n_1}) + \Gamma_{z_2} H_0^{(2)}(\gamma r_{m+2n_2}) \right] - H_0^{(2)}(\gamma r_{m+2n_1+2n_2}) + \Gamma_{z_2} K_{vx2}(2n_1h + mh) - \Gamma_{z_2} \Gamma_{z_3} K_{vx2}(2n_2h + mh) + \Gamma_{z_3} K_{vx2}(2n_1h + 2n_2h + mh),$$

$$n_1 = h_1/h, \quad n_2 = h_2/h, \quad m = \dots 3, 2, 1, 0.$$

МЕТОДИКА РЕШЕНИЯ ДИСПЕРСИОННОГО УРАВНЕНИЯ И ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Представим функцию Ханкеля первыми членами разложения в ряд [9]:

$$H_0^{(2)}(x) = 1 - j2/\pi \cdot [c + \ln(x/2) + x^2/4], \quad (19)$$

где c – постоянная Эйлера.

Используя (21) и (17), (18), выполним интегрирование в соответствии с формулой (2). В результате получим искомые выражения для частотной зависимости эффективной диэлектрической проницаемости $\varepsilon_{\text{эф}}(\omega) = k_x^2/k_0^2$, представленные в виде динамической поправки к ее значению $\varepsilon_{\text{эф}0}$ на нулевой частоте:

$$\varepsilon_{\text{эф}i} = \frac{2\varepsilon_{\text{эф}0i}}{1 - \frac{B_{0,n_1+n_2}}{C_i} - \sqrt{\left(1 - \frac{B_{0,n_1+n_2}}{C_i}\right)^2 + 4 \frac{D_i}{C_i} A_{0,n_1+n_2}}},$$

где $\varepsilon_{\text{эф}0_1} = \frac{A_{0,n_1}}{C_1}$, $\varepsilon_{\text{эф}0_2} = \frac{A_{0,n_1+n_2}}{C_2}$, $C_i = \varphi_i(A_{0,n})$, $D_i = \varphi_i(B_{0,n})$,

$$A_{m+p,m+q} = \frac{1}{2} (F_{m+p} - F_{m+q}) + \frac{k_0^2 w^2}{4} (\chi_{m+p}^2 - \chi_{m+q}^2),$$

$$B_{m+p,m+q} = \frac{-k_0^2 w^2}{4} (\chi_{m+p}^2 - \chi_{m+q}^2),$$

$$F_r = (w^2 + \chi_r^2) \ln(w^2 + \chi_r^2) - \chi_r^2 \ln(\chi_r^2) - 3w^2 - 4\chi_r [w \cdot \arctg(w/\chi_r) - \frac{\chi_r}{2} \ln(w^2 + \chi_r^2) - \frac{\chi_r}{2} \ln(\chi_r^2)], \quad \chi_r = 2rh.$$

Коэффициенты C_i , D_i определяются по рекуррентным формулам в соответствии с (19), (20):

$$\varphi_1(f_m) = \frac{2}{\varepsilon_2 + 1} (A_{m,m+n_1+n_2} - \Gamma_{z_2} A_{m+n_1,m+n_2}) + \Gamma_{z_2} \varphi_1(f_{m+n_1}) - \Gamma_{z_2} \Gamma_{z_3} \varphi_1(f_{m+n_2}) + \Gamma_{z_3} \varphi_1(f_{m+n_1+n_2}),$$

$$\varphi_2(f_m) = \frac{2}{\varepsilon_1 + \varepsilon_2} (A_{m,m+n_1} - \Gamma_{z_3} A_{m+n_2,m+n_1+n_2}) + \Gamma_{z_2} \varphi_2(f_{m+n_1}) - \Gamma_{z_2} \Gamma_{z_3} \varphi_2(f_{m+n_2}) + \Gamma_{z_3} \varphi_2(f_{m+n_1+n_2}).$$

На рис. 4, 5 представлены примеры расчета дисперсии эффективной диэлектрической проницаемости основного типа волны в двухслойных конструкциях микрополосковых линий.

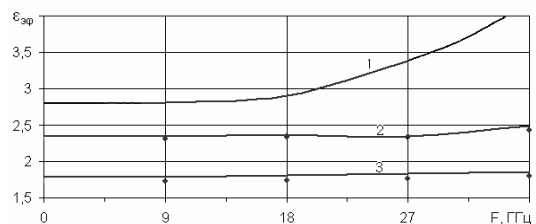


Рисунок 4 – Дисперсия $\varepsilon_{\text{эф}}$ для структуры на рис. 3, а с параметрами: $h_1 = 1$ мм, $\varepsilon_1 = 2, 1$, $\varepsilon_2 = 12, 95$, $w = 1$ мм; кривая 1 – $h_2/h_1 = 0, 2$, 2 – $h_2/h_1 = 0, 1$, 3 – $h_2/h_1 = 0$

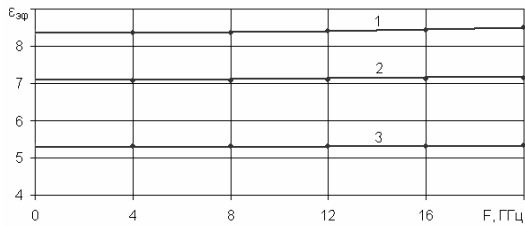


Рисунок 5 – Дисперсія $\epsilon_{эф}$ для структури на рис. 3, б с параметрами: $h_1 = 0.2$ мм, $h_2 = 0.003$ мм, $\epsilon_1 = 12.95$, $\epsilon_2 = 3.5$; кривая 1 – $w = 0.32$ мм; 2 – $w = 0.08$ мм; 3 – $w = 0.01$ мм

Точками на рисунках отмечены данные, полученные строгим электродинамическим расчетом методом моментов [10, 11].

ЗАКЛЮЧЕНИЕ

Разработана методика составления функций Грина многослойных структур подложек микрополосковых линий в квазидинамическом приближении. Проведен анализ микрополосковых линий на двухслойной подложке и получены простые аналитические формулы для расчета дисперсии. Представленные результаты расчетов дисперсии достаточно хорошо согласуются с данными численных расчетов по строгим электродинамическим методикам. Повысить точность моделирования можно путем уточнения величины $\epsilon_{эф0}$, входящей в формулы для дисперсии, в результате более строгого решения краевых задач электростатики, формулируемых для исследуемых структур.

ПЕРЕЧЕНЬ ССЫЛОК

1. Проектирование интегральных устройств СВЧ: Справочник / Ю.Г. Ефремов, В.В. Конин, Б.Д. Солганик и др. – К.: Техніка, 1990. – 159 с.
2. Das N.K., Pozar D.M. A generalized spectral-domain Green's function for multilayer Dielectric substrates with application

- to multilayer transmission lines // IEEE Trans. Microwave Theory Tech. – 1987. – Vol. 35. – № 3. – P. 326–335.
3. Arabi T.R., Murphy A.T., Sarkar T.K., Harrington R.F., Djordjevic A.R. Analysis of arbitrarily oriented microstrip lines a quasi-dynamic approach // IEEE Trans. Microwave Theory Tech. – 1991. – Vol. 39. – № 1. – P. 75–82.
4. Карпуков Л.М., Романенко С.Н. Упрощенный расчет дисперсии в микрополосковой линии // Радиотехника. – 1991. – №5. – С. 97–98.
5. Карпуков Л.М., Романенко С.Н., Пулов Р.Д. Аналітичний розрахунок дисперсії у багатопровідних мікросмужкових лініях на основі квазидинамічного наближення // Вісник Національного університету "Львівська політехніка". Радіоелектроніка та телекомунікації. – Львів: Вид-во Націон. університету "Львівська політехніка". – 2002. – №440. – С. 212–219.
6. Карпуков Л.М. Алгоритм расчета тензоров Грина для полосково-щелевых структур в слоистой среде // Радіоелектроніка. Інформатика. Управління. – 1999. – №1. – С. 11–15.
7. Карпуков Л.М., Пиза Д.М. Метод составления функций Грина для моделирования микрополосковых конструкций // Радіоелектроніка. Інформатика. Управління. – 2002. – №2. – С. 20–25.
8. Карпуков Л.М. Алгоритм моделирования функций Грина многослойных диэлектрических структур // Радіоелектроніка. Інформатика. Управління. – 2001. – №1. – С. 87–89.
9. Градштейн И.С., Рыжик И.М. Таблицы интегралов, сумм, рядов и произведений. – М.: Наука, 1971. – 1100 с.
10. Verma A.K., Hassani Sadr G. Unified Dispersion Model for Multilayer Microstrip Line. // IEEE Trans. Microwave Theory Tech. – 1992. – Vol. 40. – № 7. – P. 1587 – 1592.
11. Jackson D.R., Alexopoulos N.G. Analysis of Planar Strip Geometries in a Substrate-Superstrate Configurations. // IEEE Trans. on Antennas and Propag. – 1986. – Vol. 18. – № 6. – P. 1430–1438.

Надійшла 16.04.2004
Після доробки 05.10.2004

Запропоновано метод складання декомпозиційних схем багатослойних плоскошаруватих структур при розрахунку у квазидинамічному наближенні функцій Грина крайових задач електростатики. Розроблено метод аналітичного рішення дисперсійного рівняння для обчислення динамічної поправки до квазістатичних значень ефективної діелектричної проникності мікросмужкових ліній. Наведено результати моделювання дисперсії у двошарових конструкціях мікросмужкових ліній.

Method of compiling of multilayered structures decomposition schemes at calculation in quasi-dynamic approach of Green's functions for electrostatic boundary problems is proposed. Method of analytical solution of dispersion equation for calculation of dynamical correction to quasi-static value of effective dielectric permittivity of microstrip lines is developed. Results of dispersion modeling in two-layer constructions of microstrip lines are presented.

УДК 537.876.23

В. М. Онуфрієнко, Т. О. Штефан

ІНТЕГРОДИФЕРЕНЦІАЛЬНЕ МОДЕЛЮВАННЯ НЕЛІНІЙНОЇ ПОЛЯРИЗАЦІЇ ТА СТРУМУ У ФРАКТАЛЬНО НЕОДНОРІДНИХ ПЛАЗМОЇДАХ

Пропонується модель нелінійної поляризації та струму в фрактально неоднорідних плазмоїдах (неоднорідного фрактального заповнення деякої області згустками неоднорідної плазми). За допомогою інтегродиференціального числення задача зводиться до класичного розгляду однорідного

заповнення області речовиною (плазмою), але у термінах α -характеристик радіус-вектора положення для однорідної множини. Виявлено вплив негетогенної структури плазмоїда на величину діелектричної проникності та можливість керування нею.

ВСТУП

Основною проблемою застосування електромагнітної теорії до заповнених матеріальним середовищем тіл та областей виступає розробка методу для прогнозування результатів експериментальних спостережень та вимірювань. Для цього на основі теоретичної моделі складається математична модель середовища, що достовірно описує всі спостережувані електричні та механічні характеристики, але, водночас, простіша від тієї, що надає атомна теорія. Необхідну модель можна отримати шляхом прийняття допустимих спрощень атомної картини середовища, що складається з атомів і молекул, побудованих з електронів, протонів і нейтронів. Для більшості макроскопічних явищ подробиці форми та будови цих частинок не мають важливого значення і можна обмежуватися уявленнями про них у вигляді невеликих мас і зарядів, що зосереджуються у деяких фізичних дискретних точках (геометрично-фрактальних) (див., наприклад, [1]).

Необхідне геометричне згладжування контурів та поверхонь, що виникає у класичному підході до вимірювання протяжності гладких або кусково-гладких ліній та поверхонь, призводить до алгебраїчного визначення границь довжин вписаних у контур прямолінійних відрізків, що дає можливість визначити диференціал гладкої лінії та її довжину застосуванням формул інтегрування. Спроби запровадження такого підходу у задачах вимірювання протяжності точкових множин, "шорстких", сильно порізаних, пористих контурів призводить до наступного: довжина таких топологічно одновимірних множин дорівнює нескінченності, а площа – нулю, для інших множин площа дорівнює нескінченності, об'єм – нулю.

Звичайно, ця проблема виникає і у спробах застосування класичного аналізу для визначення протяжності множин зарядів і струмів у сильно структурованих середовищах, таких як плазма у нескінченному просторі або у областях згущення (плазмоїдах). Такі негомогенні структури мають включення у вигляді великої кількості частинок речовини, розділених просторово на відстані, порівнювані з діаметром виділеної неоднорідності.

Відому методику наближення некоординатних меж областей визначення поля за допомогою покриття простими компактами (прямокутниками, кругами, еліпсами і таке інше) [2] будемо використовувати далі для дослідження одного з основних математичних аспектів теорії фракталів, яким є питання про збіжність до фракталу утвореної послідовності множин. Для цього необхідне вимірювання відстані між компактними множинами, тобто введення відповідної метрики.

Розгляд фрактальних множин у метриці Хаусдорфа дозволяє порівнювати величину хаусдорфової розмірності з показником порядку α дробового інтегродиференціала, конструкція якого виникає у задачі про визначення протяжності фрактальної множини, що є моделлю для опису просторової неоднорідності (див., наприклад [3]).

Виявленні таким чином зв'язки дають можливість означити поняття дифереінтегральної α -форми довжини (площі та об'єму) та виявити спектр її застосування для розв'язування задач фрактальної теорії поля та електродинаміки [3].

ПОСТАНОВКА ЗАДАЧІ

Розглянемо окремий випадок неоднорідного заповнення області простору речовиною у вигляді системи заряджених частинок (електронів, іонів та нейтральних атомів) – плазми (частково або повністю іонізованого газу за наявності електромагнітного поля, що гармонійно коливається).

Відомо (див., наприклад, [4]), якщо зовнішнє поле прикладається до матеріалу, що складається з полярних множин з випадковою орієнтацією, то на множини діє механічний момент, що прагне направити осі поляризації множини паралельно застосованому електричному полю. Сили, що повертають множини до випадкових орієнтацій, у першу чергу є теплового походження і тому сильно залежать від температури.

Основною мікроскопічною характеристикою поляризованої множини незалежно від механізму поляризації (поляризація постійна, обумовлена внутрішньою будовою, або тимчасова, індукована зовнішнім електричним полем) є α -польний момент, що виникає внаслідок мікроскопічного розділення центрів додатних та від'ємних зарядів і розміщенням їх вздовж деякої осі.

Якщо діелектрична $\epsilon^{(\alpha)}$ та магнітна $\mu^{(\alpha)}$ проникності не залежать від поля, то співвідношення між індукціями та напруженностями є лінійними, а фрактальне середовище можна назвати лінійним у смислі процесів поляризації.

Лінеаризація залежностей між напруженістю електричного вектора та поляризацією й індукцією здійснюється, коли у відповідних розвиненнях у ряд Тейлора за степенями E

$$(D^{(\alpha)} f)(E) = \sum_{j=0} \frac{(D_E^j f)(0)}{\Gamma(1+j)} E^j, \quad \beta = \alpha + j, \quad \text{Re } \alpha > 0$$

відкинуто дуже малі нелінійні члени.

Для фрактального матеріалу середовища запишемо розвинення в ряд скалярних компонент поляризації $P^{(\alpha)}(E)$ та струму $J^{(\alpha)}(E)$:

$$P^{(\alpha)}(E) = \epsilon_0 \chi_e^{(\alpha)}(E) \cdot E^\alpha, \quad (1)$$

$$J^{(\alpha)}(E) = \sigma^{(\alpha)}(E) \cdot E^\alpha, \quad (2)$$

де

$$\chi_e^{(\alpha)}(E) = \chi_{e0}^{(\alpha)} + \chi_{e1}^{(\alpha)} E + \chi_{e2}^{(\alpha)} E^2 + \dots + \chi_{ei}^{(\alpha)} E^i + \dots,$$

$$\sigma^{(\alpha)}(E) = \sigma_0^{(\alpha)} + \sigma_1^{(\alpha)} E + \sigma_2^{(\alpha)} E^2 + \dots + \sigma_i^{(\alpha)} E^i + \dots,$$

$$\chi_{e0}^{(\alpha)} = \frac{(D_E^\alpha P)(0)}{\Gamma(1+\alpha)}, \chi_{e1}^{(\alpha)} = \frac{(D_E^{1+\alpha} P)(0)}{\Gamma(2+\alpha)}, \dots, \chi_{ej}^{(\alpha)} = \frac{(D_E^{j+\alpha} P)(0)}{\Gamma(j+1+\alpha)},$$

$$\sigma_0^{(\alpha)} = \frac{(D_E^\alpha J)(0)}{\Gamma(1+\alpha)}, \sigma_1^{(\alpha)} = \frac{(D_E^{1+\alpha} J)(0)}{\Gamma(2+\alpha)}, \dots, \sigma_j^{(\alpha)} = \frac{(D_E^{j+\alpha} J)(0)}{\Gamma(j+1+\alpha)}.$$

Формули (1)–(2) описують класичне нефрактальне нелінійне середовище. У цьому випадку перші доданки \vec{P}_0 та \vec{J}_0 таких розвинень є сталими векторами, що не залежать від електромагнітного поля і є власними для матеріалу розглядуваного середовища.

Сталий вектор \vec{P}_0 описує так звану спонтану поляризацію. Як відомо, у деяких анізотропних кристалах (електретах) існує власний електричний момент, що "застигає" у присутності електричного поля, а у надпрвідниках струми \vec{J}_0 з великою точністю являють собою електричні струми, що збігаються.

Урахування відсутності спонтанної поляризації $P_0 = \epsilon_0 \chi_{e0}^{(0)} E^0 = 0$ та струму $J_0 = \sigma_0 E^0 = 0$ з нехтовно малими значеннями сум $(\chi_{e2}^{(0)} E^2 + \chi_{e3}^{(0)} E^3 + \dots)$, $(\sigma_2^{(0)} E^2 + \sigma_3^{(0)} E^3 + \dots)$, приводить до задачі з лінеаризованою проникностями.

Для фрактального середовища з (1)–(2) впливає, що скалярні компоненти перших доданків у розвиненнях в ряд Тейлора поляризації та струму можна представити у вигляді

$$\begin{aligned} \epsilon_0 \chi_{e0}^{(\alpha)} E^\alpha &= \epsilon_0 \frac{D_E^\alpha P(0)}{\Gamma(1+\alpha)} E^\alpha; \\ \sigma_0^{(\alpha)} E^\alpha &= \frac{D_E^\alpha J(0)}{\Gamma(1+\alpha)} E^\alpha. \end{aligned} \quad (3)$$

Якщо застосувати положення про відсутність спонтанної поляризації для фрактального середовища, тобто покласти (3) таким, що дорівнює нулю, то формально це стає можливим, коли $E=0$, $\alpha > 0$, або $(D_E^\alpha P)(0)=0$. Якщо у першому випадку, як і у класичному, нульове значення $P^{(\alpha)}(0)=0$ відповідає відсутності спонтанної поляризації, то у другому випадку, що відповідає наявності фрактально структурованого поля, з $(D_E^\alpha P)(0)=0$ впливає можливість представлення першого доданку поляризації (1) у явному вигляді степеневої функції $P(E, \alpha) = \frac{1}{\Gamma(1+\alpha)} \frac{C}{E^{1-\alpha}}$, де C – будь-яка константа. Таким чином, можна записати формальну рівність $P(E, \alpha) = a_\alpha D^\alpha \delta(E)$.

Звідси впливає висновок про те, що впровадженням у середовище сторонніх джерел енергії з фрактальним дельта-подібним по полю E розподілом зарядів (струмів) можна створити штучну "спонтанну" нелінійну електричну поляризацію.

Аналогічний розгляд для струму дає представлення $J(E, \alpha) = c_\alpha D^\alpha \delta(E)$.

ФРАКТАЛЬНА НЕОДНОРІДНІСТЬ ПЛАЗМОЇДІВ

Рух у плазмоїді кожного згустку заряду Q і маси m під дією електричного поля $E = E_m \cos(\omega t + \varphi)$ описується рівнянням

$$\frac{Q}{m} E_m \cos(\omega t + \varphi) = \frac{d^2}{dt^2} (r_0 D_r^\alpha r^{(\alpha)}) + \nu \frac{d}{dt} (r_0 D_r^\alpha r^{(\alpha)}) \quad (4)$$

Розв'язок цього рівняння знаходимо у термінах α -характеристик радіус-вектора положення заряду

$$r_0 D_r^\alpha (r^{(\alpha)} - r_0^{(\alpha)}) = v_0^{(\alpha)}(t) - \frac{Q E_m \cos(\omega t + \varphi)}{m \omega (\omega - i\nu)},$$

причому функція $r_0 D_r^\alpha (r^{(\alpha)} - r_0^{(\alpha)}) = v_0^{(\alpha)}(t)$ характеризує положення згустку ("товстої" фрактальної точки) у плазмоїді за відсутності поля.

З урахуванням властивостей інтегродиференціалів знаходимо

$$r^{(\alpha)} - r_0^{(\alpha)} = \frac{1}{\Gamma(1+\alpha)} \left[v_0^{(\alpha)}(t) - \frac{Q E_m \cos(\omega t + \varphi)}{m \omega (\omega - i\nu)} \right] (r^{(\alpha)} - r_0^{(\alpha)})^\alpha,$$

і відтворюємо α -характеристику радіус-вектора

$$r^{(\alpha)} - r_0^{(\alpha)} = \left(\frac{1}{\Gamma(1+\alpha)} \left[v_0^{(\alpha)}(t) - \frac{Q E_m \cos(\omega t + \varphi)}{m \omega (\omega - i\nu)} \right] \right)^{\frac{1}{1-\alpha}}. \quad (5)$$

Помічаємо, що значення скейлінгового параметра $\alpha=0$ відповідає класичній моделі однорідної плазми (див., наприклад, [4]), коли $r^{(\alpha)} = r$, $r_0^{(\alpha)} = r_0$, гама-функція Ейлера $\Gamma(1)=1$.

Для згустків плазмоїда у вигляді нейтральної системи зарядів (число електронів та від'ємних іонів дорівнює числу позитивних іонів) з урахуванням (5) обчислюємо електричний момент довільного елемента об'єму $\Delta V^{(\alpha)}$ плазмоїда

$$p^{(\alpha)} \Delta V^{(\alpha)} = \sum_i^{(\Delta V^{(\alpha)})} Q_i \left(\frac{1}{\Gamma(1+\alpha)} \left[v_0^{(\alpha)}(t) - \frac{Q E_m \cos(\omega t + \varphi)}{m \omega (\omega - i\nu)} \right] \right)^{\frac{1}{1-\alpha}},$$

де індекс i означає підсумовування по всіх частинках з зарядами Q_i всередині $\Delta V^{(\alpha)}$. Після переходу до комплексних амплітуд з урахуванням того, що всі заряди однакові за абсолютною величиною і дорівнюють заряду Q_i , а іони – одновалентні, маємо

$$\dot{p}_m^{(\alpha)} \Delta V^{(\alpha)} = \sum_i^{(\Delta V^{(\alpha)})} \left(-\frac{Q_i^{2-\alpha} \dot{E}_m}{m_i \omega (\omega - i\nu) \Gamma(1+\alpha)} \right)^{\frac{1}{1-\alpha}}.$$

Якщо не брати до уваги вплив іонів з їх масами, що дуже великі у порівнянні з масою електрона, то для

плазмоїда з зарядом Q і масою m маємо

$$\dot{P}_m^{(\alpha)} \Delta V^{(\alpha)} = \left(-\frac{Q^{2-\alpha} N_{\Delta V^{(\alpha)}} \dot{E}_m}{m\omega(\omega - i\nu)\Gamma(1+\alpha)} \right)^{\frac{1}{1-\alpha}}, \quad (6)$$

де $N_{\Delta V^{(\alpha)}}$ – число електронів в об'ємі $\Delta V^{(\alpha)}$.

Граничним переходом в (6) визначається комплексна амплітуда вектора поляризації середовища плазмоїда

$$\dot{P}_m^{(\alpha)} = \left(\frac{(-Q)^{2-\alpha} \tilde{N} \dot{E}_m}{m\omega(\omega - i\nu)\Gamma(1+\alpha)} \right)^{\frac{1}{1-\alpha}} \quad (7)$$

(\tilde{N} – число електронів у відношенні до одиниці об'єму).

Одержаний результат надає можливість характеризувати плазмоїд за допомогою електричної сприйнятливості $\dot{\chi}^{(\alpha)}$ та діелектричної проникності $\dot{\varepsilon}^{(\alpha)}$, які вводяться за загальною схемою (див., напр., [4]) і обчислюються з урахуванням (7) за формулою

$$\dot{\varepsilon}^{(\alpha)} = \varepsilon_0 \left(1 + \dot{\chi}^{(\alpha)} \right) = \varepsilon_0 - \left(\frac{(-Q)^{2-\alpha} \tilde{N}}{m\omega(\omega - i\nu)\Gamma(1+\alpha)} \right)^{\frac{1}{1-\alpha}},$$

звідки відносна проникність

$$\dot{\varepsilon}_r^{(\alpha)} = 1 - \frac{1}{\varepsilon_0} \left(\frac{(-Q)^{2-\alpha} \tilde{N}}{m\omega(\omega - i\nu)\Gamma(1+\alpha)} \right)^{\frac{1}{1-\alpha}}.$$

За аналогією з класичним випадком суцільної плазми одержуємо для плазмоїда, неоднорідність якого характеризується скейлінговим показником α , формулу для відносної діелектричної проникності

$$\dot{\varepsilon}_r^{(\alpha)} = 1 - \frac{\omega_p^2}{[\omega(\omega - i\nu)]^{2/(1-\alpha)}},$$

де $\omega_p^2 = \frac{1}{\varepsilon_0} \left(\frac{(-Q)^{2-\alpha} \tilde{N}}{m\Gamma(1+\alpha)} \right)^{\frac{1}{1-\alpha}}$, а параметр ω_p^2 можна називати плазмоїдною частотою.

На рис. 1 представлено залежність від частоти ω радіус-вектора, що визначає просторову неоднорідність плазмоїда і обчислюється за формулою (5).

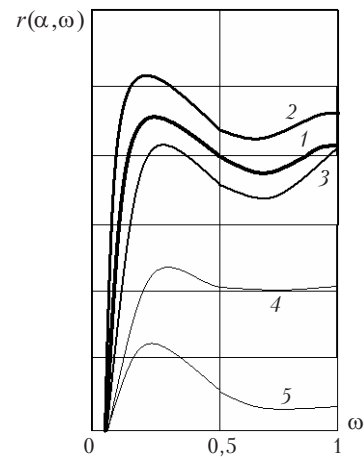


Рисунок 1 – Радіус-вектор характеристики просторової неоднорідності плазмоїда у залежності від частоти

Варіювання скейлінгового показника надає можливість управління просторовою неоднорідністю плазмоїда (крива 2 відповідає значенню $\alpha = 0,1$; крива 3 – значенню $\alpha = -0,2$; крива 4 – значенню $\alpha = -0,5$; крива 5 – значенню $\alpha = -1$).

Можливість варіювання у широких межах просторової неоднорідності радіус-вектора за допомоги зміни скейлінгового показника α показана на рисунку 2.

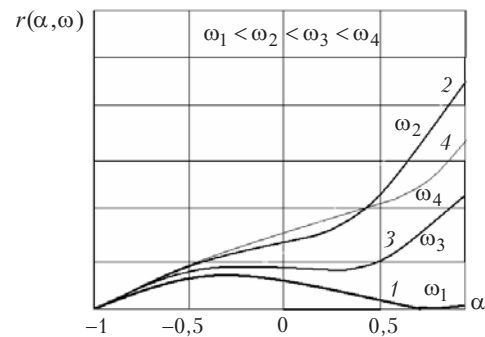


Рисунок 2 – Скейлінгові властивості радіус-вектора просторової неоднорідності плазмоїда

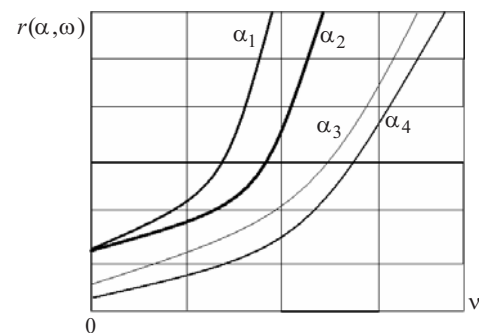


Рисунок 3 – Радіус-вектор характеристики просторової неоднорідності плазмоїда у залежності від значення ν

ВИСНОВКИ

Отже, наявність у середовищі сторонніх джерел енергії з фрактальним дельта-подібним по полю E розподілом зарядів (струмів) створює штучну "спонтанну" нелінійну електричну поляризацію.

Практикою підтверджується використання розвинення у ряд з розглядом лінійних членів для аналізу більшості матеріалів (діелектриків, діаманетиків, парамагнетиків та металевих провідників). Але тільки лінійні члени не можуть виражати поляризацію та струми у матеріалах, що набувають фрактальних змін у структурі за рахунок впливу інтенсивного лазерного опромінювання, у феромагнітних матеріалах з доменною структурою, у неметалевих пористих провідниках типу вугілля, у сильно структурованих металах.

Для середовища з відсутніми спонтанними поляризаціями та струмом урахування у наведених розвиненнях (1)–(2) одного доданку, що за значень $\alpha = 0$ стає лінійним, для значень $-1 < \alpha < 1$ надає можливість урахування "долінійних" та "післялінійних" (до другого порядку) випадків нелінійності, що характеризує фрактально середовище. Це дозволяє говорити про те, що фрактально структуроване середовище завжди є слабо нелінійним.

З появою потужних лазерів стали доступними впровадження сильних електромагнітних полів, у зв'язку з чим значно розширилось коло спостережуваних нелінійних явищ, пояснення перебігу яких можливе за рахунок розгляду введеної інтегродиференціальної моделі [8].

ПЕРЕЛІК ПОСИЛАНЬ

1. V.M. Onufrienko. The Differintegral Model for Describing Fractal Coupling Between Waveguide Surfaces // Telecommunications and Radio Engineering. V. 57, N1. – 2002. – PP. 30–36.

2. Onufrienko V.M., Прохода И.Г., Чумаченко В.П. Численное решение задачи о волноводном трансформаторе с соединительной полостью сложной формы // Изв. вузов. Радиофизика. – 1975. – 18. – №4. – С. 584–587.
3. В.М. Онуфрієнко. Диферінтегральні альфа-форми у хаусдорфовій метриці на фрактальних множинах // Радіоелектроніка. Інформатика. Управління. – 2002. – №2(8). – С. 31–39.
4. В.В. Никольский. Электродинамика и распространение радиоволн. – М.: Наука, 1973. – 608 с.
5. K.J. Falconer. The Geometry of Fractal Sets. Cambr. Univ. Press, Cambridge, 1985. – 268 p.
6. Л. Левин. Теория волноводов. Методы решения волноводных задач: Пер. с англ. – М.: Радио и связь, 1981. – 312 с.
7. Онуфрієнко В.М., Штефан Т.О. Інтегродиференціальне моделювання просторової фрактальної неуніформності плазмодів // Радіоелектроніка. Інформатика. Управління. – 2003. – №2(10). – С. 38–42.
8. Onufrienko V.M., Onufrienko L.M. Field of the Pulsed Space-time source in simulated medium // Proc. 2004 intern. Workshop Ultrawideband and Ultrashort Impulse Signals (UWBUSIS' 04), Sevastopol, Ukraine. 2004. – P. 179 – 181.

Надійшла 28.05.2004
Після доробки 02.11.2004

Предлагается модель нелинейной поляризации и тока в фрактально неоднородных плазмоидах (неоднородного фрактального заполнения некоторой области сгустками однородной плазмы). С помощью интегродифференциального исчисления задача сводится к классическому рассмотрению однородного заполнения области веществом (плазмой), но в терминах α -характеристик радиус-вектора положения для однородного множества. Выявлено влияние неомогенной структуры плазмоида на величину диэлектрической проницаемости и возможность управления ею.

The models of nonlinear polarization and current in fractal nonuniform plasmoids (the nonuniform fractal filling of some region by clots of the homogeneous plasma) are proposed. Using the integro-differential calculus the problem is reduced to a classical consideration of the region filled with a homogeneous substance (plasma), but in terms α -characteristics of the position vector standing for a homogeneous set. Influence of the non-homogeneous structure of a plasmoid on the value of inductivity and an opportunity of its operation are detected.

УДК 621.372

В. П. Пьянков

ЧИСЛЕННЫЙ АНАЛИЗ СЛОЖНЫХ ВОЛНОВОДНЫХ ИЗЛУЧАТЕЛЕЙ С ФЛАНЦАМИ НА ОСНОВЕ ОБОБЩЕННЫХ МАТРИЦ РАССЕЙНИЯ И МЕТОДА ПРОИЗВЕДЕНИЯ ОБЛАСТЕЙ

Объекты исследования – возбуждаемые волноводами плоскостные излучатели с фланцами, имеющие многоугольный контур проводящих поверхностей и кусочно-однородное диэлектрическое заполнение. Цель – создание единой математической модели электромагнитных полей указанных излучателей. Метод – совместное применение обобщенных матриц рассеяния и метода произведения областей. Приведены результаты численных экспериментов.

ВВЕДЕНИЕ

В данной работе математический аппарат [1, 2] анализа волноводных трансформаторов, основанный на

совместном применении метода произведения областей (ПО) и метода обобщенных матриц рассеяния (ОМР), развивается на задачи излучения из плоских структур. Излучающая структура представляет собой щель с бесконечным фланцем в многоугольной полости с кусочно-однородным заполнением, к которой подведены питающие волноводы (наиболее простые из таких структур – открытый конец волновода с фланцем и рупор с фланцем). Указанные структуры являются двухмерными моделями волноводных и рупорных антенн разнообразной формы, широко применяемых для работы в СВЧ и КВЧ диапазонах, например, на борту летательных аппаратов.

В работе приведены результаты расчетов диаграмм направленности (ДН) излучения, которые хорошо совпадают с известными данными, полученными строгими методами, разработанными для различных структур определенной конфигурации. С целью демонстрации возможностей также выполнены расчеты для других сложных структур.

ФОРМУЛИРОВКА ЗАДАЧИ И ОПИСАНИЕ СПОСОБА РЕШЕНИЯ

Рассмотрим цилиндрические излучающие структуры (рисунок 1). Полагаем, что цилиндрическая структура зажата вместе с областью излучения между двумя идеально проводящими плоскостями, перпендикулярными образующей и расположенными друг от друга на расстоянии b , или бесконечна вдоль образующей ($b = \infty$). Регулярные волноводы возбуждаются модами типа LM_{pl} (H -плоскостная задача) или LE_{pl} (E -плоскостная), где l – количество вариаций поля вдоль образующей. Причем $l = 0$, если $b = \infty$. Отметим, что возбуждение LE_{pl} , где $l \neq 0$, рассматривается только в случае, если внутри и вне структуры вакуум (воздух). Под волной LE_{00} надо понимать волну ТЕМ. H_{pl} – другое обозначение волн LM_{pl} . При возбуждении типа LM (LE) решение задачи сводится к нахождению компоненты U электрического (магнитного) поля, параллельной образующей.

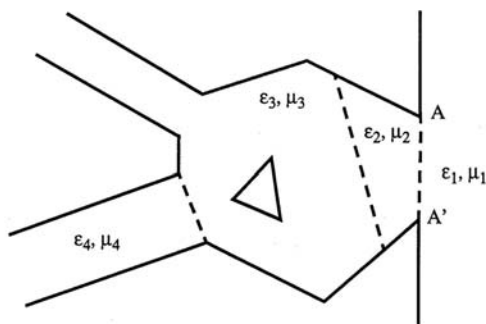


Рисунок 1

Для общности будем рассматривать идеально проводящие поверхности (сплошные линии на рисунке 1), как электрические, так и магнитные. В $H(E)$ -плоскостной задаче компонента U на электрической поверхности удовлетворяет однородному граничному условию Дирихле (Неймана), а на магнитной – условию Неймана (Дирихле).

Полость может иметь кусочно-однородное беспотерное магнитоэлектрическое заполнение. Такое допущение позволяет, например, учесть наличие защитных радиопрозрачных (в некоторой степени) кожухов, закрывающих отверстия излучателей. Ограничимся рассмотрением случаев, когда границы раздела сред заполнения (штриховые линии на рисунке 1) представляют собой отрезки прямых линий, концы которых лежат на контуре проводящих стенок. Внутри любой среды заполнения могут быть идеально проводящие

многоугольные включения, не соприкасающиеся с границами раздела. Такие проводящие вставки могут иметь различное назначение: быть крепежным элементом, улучшать согласование с питающим волноводом, формировать ДН специального вида и др.

Произведем декомпозицию (рисунок 2) излучателя согласно методике [1, 2]. Отсечем регулярные полубесконечные участки волноводов плоскостями, перпендикулярными стенкам волноводов. Сложную соединительную полость (при необходимости) также разделим плоскостями на блоки. При этом полагаем, что открытые концы волноводов и окна блоков, полученные при расчленении, имеют бесконечные фланцы, которые сливаются при $d \rightarrow 0$. На сливающиеся фланцы накладываем одинаковые однородные граничные условия либо Неймана, либо Дирихле. Промежуточная среда между фланцами может быть взята любой беспотерной магнитоэлектрической, что не влияет на результаты расчетов. В качестве посредников взаимодействия между блоками будем рассматривать эллиптические волны, излучаемые окнами. Этот выбор связан с тем, что в методе ПО вклад каждого прямолинейного звена граничного контура в описываемое поле представляется в виде разложения по эллиптическим волнам. Таким образом, появляется возможность единообразно описывать внутренние и внешние поля каждого блока, без излишних преобразований, приводящих к снижению точности и увеличению времени счета.

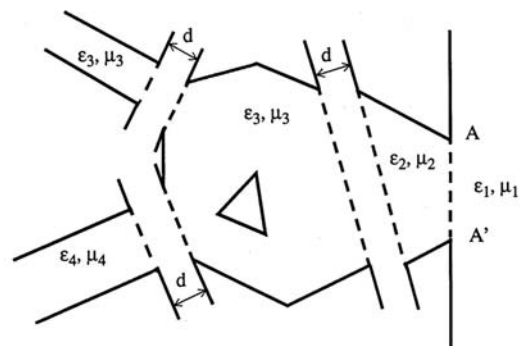


Рисунок 2

Согласно [1, 2] по методу ПО для каждого промежуточного блока автономно, отдельно от других блоков, находим обобщенные матрицы рассеяния, которые состоят из коэффициентов разложений по эллиптическим волнам полей излучения из окон. Так столбец с номером p матрицы рассеяния $N S_{ij}^p$ N -го блока состоит из коэффициентов разложения по эллиптическим волнам поля излучения из i -го окна при возбуждении эллиптической модой p -го порядка с единичным коэффициентом, излучаемой апертурой соседнего блока, которая сливается с j -м окном рассматриваемого блока при $d \rightarrow 0$ (рисунок 3). Аналогично рассчитываем ОМР конечного блока излучателя, содержащего апертуру AA' . При этом нет отличий в описании излучения из какого-либо промежуточного окна и конечной апертуры AA'

(кроме одного: апертура AA' не возбуждается со стороны полупространства излучения).

Расчет ОМР открытых концов волноводов (ОКВ) с бесконечными фланцами производим по алгоритму [1, 2]. Две матрицы рассеяния (соответствующие двум способам возбуждения) ОКВ состоят из коэффициентов разложения полей излучения по эллиптическим волнам, две другие – из коэффициентов разложения полей внутри волновода по собственным волнам волновода. Возбуждение со стороны волновода производится поочередно собственными волнами волновода (как распространяющимися, так и затухающими), набегающими на раскрыв и имеющими в его плоскости единичную амплитуду (рисунок 4, а), а со стороны полупространства – поочередно единичными по норме эллиптическими модами щели в бесконечном экране (окна соседнего блока), сливающейся с раскрывом волновода при $d \rightarrow 0$ (рисунок 4, б). ОКВ с фланцем является блоком, преобразующим собственные волны волновода в эллиптические волны и наоборот.

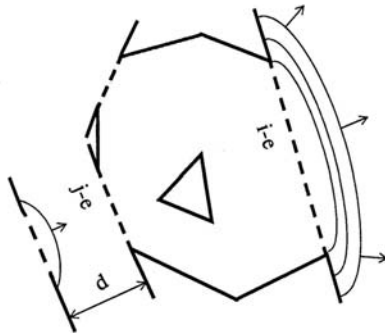


Рисунок 3

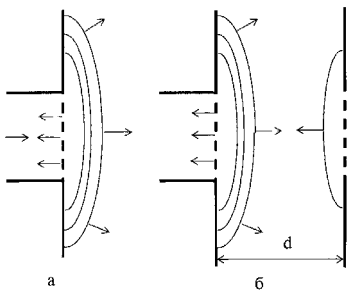


Рисунок 4

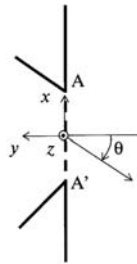


Рисунок 5

Обобщенные матрицы рассеяния всей излучающей структуры находим по известным рекомпозиционным формулам [3]. Далее эти итоговые матрицы будем обозначать через S^{ij} , где j может пробегать номера волноводов, а i – номера этих же каналов и еще номер выходного канала-апертуры AA' , который будем обозначать буквой A .

Допустим, что волновод с номером K возбужден собственной волной с номером p . Тогда поле излучения из апертуры AA' описывается рядом

$$U = \sum_{n=1}^{\infty} S_{np}^{AK} Ma_n(\xi) ma_n(\eta), \quad (1)$$

где S_{np}^{AK} – элементы матрицы S^{AK} , (ξ, η) – эллиптические координаты, связанные с декартовыми (рисунок 5) формулами

$$x = f \operatorname{ch} \xi \cos \eta, \quad y = f \operatorname{sh} \xi \sin \eta, \quad (2)$$

где f – половина длины апертуры, начало отсчета декартовой системы – в середине апертуры, ось Oy направлена внутрь излучающей структуры. В полупространстве излучения $\eta \in [-\pi, 0]$, $\xi \in [0, \infty)$.

Если на фланце апертуры AA' наложено однородное граничное условие Неймана, то специальные функции, используемые в (1), равны

$$ma_n(\eta) = ce(\eta, q), \quad Ma_n(\xi) = \frac{Me_{n-1}^{(2)}(\xi, q)}{Me_{n-1}^{(2)}(0, q)}. \quad (3)$$

Если условие Дирихле, то

$$ma_n(\eta) = se(\eta, q), \quad Ma_n(\xi) = \frac{Ne_n^{(2)}(\xi, q)}{Ne_n^{(2)}(0, q)}, \quad (4)$$

где $ce_n(\eta, q)$, $se_n(\eta, q)$, $Me_n^{(2)}(\xi, q)$, $Ne_n^{(2)}(\xi, q)$ – функции Матье [4],

$$q = (\chi f / 2)^2, \quad (5)$$

$$\chi = 2\pi \sqrt{\frac{\epsilon \mu}{\lambda^2} - \left(\frac{l}{2b}\right)^2}, \quad (6)$$

λ – длина волны в свободном пространстве, ϵ, μ – относительные диэлектрическая и магнитная проницаемости среды полупространства излучения. Число χ – параметр уравнения Гельмгольца

$$\Delta U + \chi^2 U = 0, \quad (7)$$

которому должна удовлетворять компонента U в полупространстве излучения.

Диаграмма направленности (ДН) излучения в дальней зоне вычисляется согласно выражению

$$F(\theta) = \lim_{\rho \rightarrow \infty} U(\rho, \alpha) \sqrt{\pi \chi \rho / 2} e^{i^* \chi \rho} \quad (8)$$

где θ – угол направления излучения, отсчитываемый как показано на рисунок 5, (ρ, α) – полярная система координат, связанная с декартовыми соотношениями

$$x = \rho \cos \alpha, \quad y = \rho \sin \alpha, \quad (9)$$

i^* – мнимая единица. Угол α связан с углом θ равенством $\alpha = -\theta - \pi/2$. (Как правило, для полярного угла θ , используемого при описании двумерных ДН, положительным направлением отсчета берется направление по часовой стрелке).

Порядок редукции ряда (1) выбирался в зависимости от длины $2f$ апертуры обычно согласно рекомендациям, выработанным для E -плоскостных задач [5].

РЕЗУЛЬТАТЫ ЧИСЛЕННЫХ ЭКСПЕРИМЕНТОВ

Предлагаемым способом были рассчитаны ДН излучения из открытых концов плоских волноводов с бесконечными фланцами, перпендикулярными стенкам волноводов при возбуждении волнами TEM и H_{10} . При этом волноводы были либо пусты, либо заполнены различными диэлектрическими средами. Рассчитанные ДН с графической точностью совпали с ДН [6–8], полученными строгим методом факторизации. Кроме того, были выполнены расчеты, результаты которых изображены на следующих рисунках.

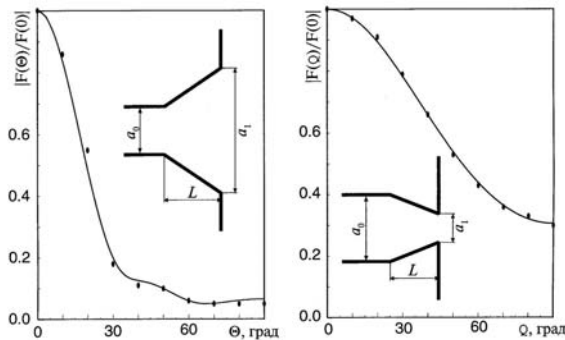


Рисунок 6

Рисунок 7

На рисунке 6 показана нормированная ДН по полю питаемого волноводом рупора с фланцем, а на рисунок 7 – ДН по полю сужающегося по линейному закону конца волновода с фланцем. Волноводы возбуждались волнами типа TEM , среда везде – вакуум, фланцы – бесконечные, все поверхности – идеально проводящие электрические (однородное граничное условие Неймана). Размеры структур: $a_0/\lambda = 4,5/\pi$; $a_1 = 1,5a_0$; $L = 0,5a_0$ (рисунок 6); $a_0/\lambda = 4,5/\pi$; $a_1 = 0,5a_0$; $L = a_0$ (рисунок 7). Сплошные линии – ДН, рассчитанные по разработанному алгоритму, "вытянутые точки" – данные [9]. Результаты практически совпали.

Более сложной, чем рассмотренные ранее, излучающей системой является структура [10], изображенная на рисунке 8, а. Это ОКВ шириной a_0 с фланцем, перед которым на расстоянии δ_1 находится металлический экран толщиной δ_2 со щелью шириной a_1 . Щель заполнена радиопрозрачным однородным веществом, диэлектриком с относительной проницаемостью ϵ_2 . Фланец и экран имеют бесконечную длину. Все металлические элементы полагаются идеально проводящими. У всех неметаллических сред относительная магнитная проницаемость равна единице. Волновод (шириной a_0) возбуждается TEM -волной. Щель в экране располагается симметрично относительно оси волновода или смещена в сторону.

Промежуток между фланцем и экраном можно рассматривать как два волновода, которые отходят в стороны и забирают часть подводимой энергии. Понимание этого является принципиальным для анализа описанной структуры предлагаемым методом. На рисунок 8, б показана схема разбиения структуры на блоки. Заглавными одинаковыми буквами обозначены узловые точки, которые сливаются в одну (см. рисунок 8, а) при соединении блоков. На рисунке 8, в изображен один из возможных промежуточных этапов рекомпозиции, когда к блоку BCDEFG (рисунок 8, б) прямоугольной формы с четырьмя окнами подсоединены боковые отводящие волноводы с апертурами BG и EF. В результате получается блок, описывающий целиком свойства промежутка между раскрытием с фланцем питающего волновода и экраном с диэлектрической щелью.

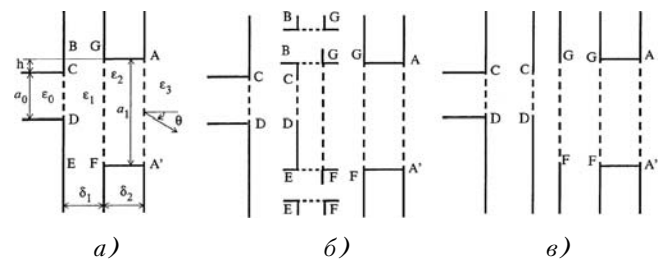


Рисунок 8

На рисунках 9–11 приведены ДН такой излучающей структуры при $\epsilon_2 = 2,56$; $\epsilon_0 = \epsilon_1 = \epsilon_3 = 1$; $a_0 = 0,284\lambda$; $a_1 = 6,064\lambda$ и различных значениях расстояния δ_1 и смещения $\Delta h = h - (a_1 - a_0)/2$ осей волновода и щели. Обозначение осей ординат – $F_0(\theta) = \left| \frac{F(\theta)}{\max F(\theta)} \right|$. Сплошные линии и кружки – соответственно теоретические и экспериментальные ДН, взятые из [10]. Пунктирные линии – ДН, полученные по предлагаемому алгоритму. При этом в рядах (1), соответствующих окнам GF и AA' большой электрической длины, учитывалось до 20 членов. Из рисунков видно, что совпадение результатов удовлетворительное.

На рисунок 12, а изображен контур H -плоскостной излучающей системы [11], представляющей собой открытый конец двухмодового волновода с фланцем, питаемого уголковым элементом, возбуждающим смесь волн H_{10} и H_{20} .

Как указывается в [11], при определенных размерах уголкового элемента осуществляет высокоэффективное преобразование набегающей по питающему волноводу шириной a_0 H_{10} -волны в H_{20} -волну двухмодового выходного волновода шириной a_1 . При этом формируется ДН, имеющая два главных лепестка с глубоким минимумом, расположенным между ними практически на линии симметрии выходного волновода [11]. Это подтверждает рисунок 12, б, на котором изображена ДН, рассчитанная предлагаемым способом при частоте $f = 67,5$ ГГц, возбуждении типа LM_{01} , $\epsilon = \mu = 1$ и сле-

дующих размерах: $a_0 = 3,6$ мм; $a_1 = 6,11$ мм; $L = 22,79$ мм; $c_1 = 1,58$ мм; $c_2 = 5,15$ мм; $\alpha = 80^\circ 30'$.

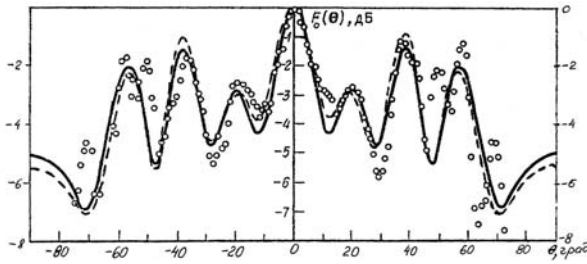


Рисунок 9
 $\Delta_1 = 0,888\lambda$; $\Delta h = 0$

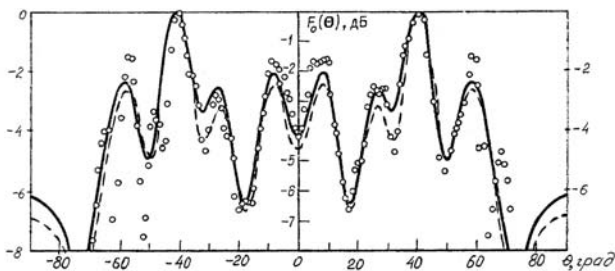


Рисунок 10
 $\Delta_1 = 1,360\lambda$; $\Delta h = 0$

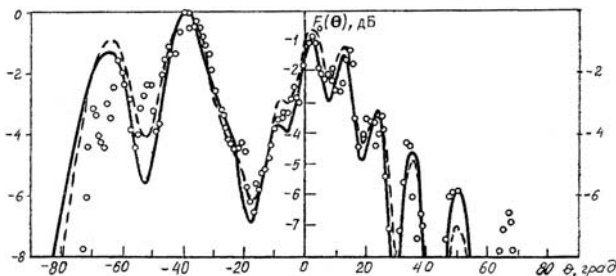


Рисунок 11
 $\Delta_1 = 0,888\lambda$; $\Delta h = 2\lambda$

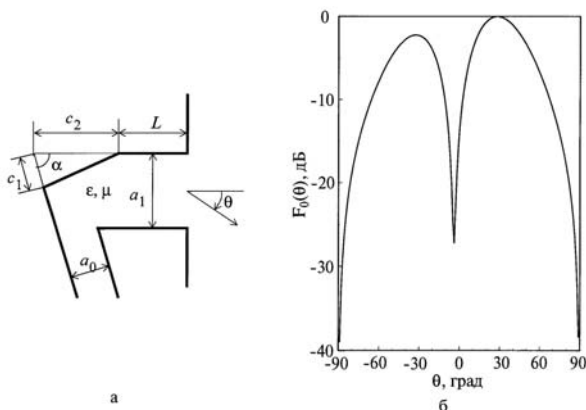


Рисунок 12

ЗАКЛЮЧЕНИЕ

Метод ПО является универсальным и эффективным в классе многоугольных цилиндрических структур. Привлечение метода ОМР делает разработки еще более эффективными, универсальными и гибкими. Предлагаемый алгоритм расчета ДН излучения и параметров согласования с питающими волноводами позволяет произвольно варьировать геометрией устройств в классе плоских структур с кусочно-линейным контуром. Поэтому он может представлять интерес для разработчиков САПР техники связи СВЧ и КВЧ диапазонов.

ПЕРЕЧЕНЬ ССЫЛОК

1. Чумаченко В.П., Пьянков В.П. Расчет сложных плоскостных волноводных структур при помощи аппарата обобщенных матриц рассеяния и метода произведения областей. // Материалы 5-й Крымской конференции и выставки "СВЧ-техника и спутниковый прием" (Сева-стополь, Украина, 25-27 сентября, 1995), Т.1, С. 107-110.
2. Chumachenko, V.P., and Pyankov, V.P. Numerical analysis of complicated waveguide circuits on the basis of generalized scattering matrices and domain product technique. // IEEE Transactions on Microwave Theory and Techniques. - 2000. - Vol. 48, no 2. P. 305-308.
3. Митра Р., Ли С. Аналитические методы теории волноводов. - М.: Мир, 1974. - 323 с.
4. Мак-Лахлан Н.В. Теория и приложения функций Матье. - М.: ИЛ, 1953. - 475 с.
5. Чумаченко В.П. Решение задач дифракции волн на многоугольной цилиндрической поверхности методом произведения областей // Докл. АН УССР. Сер. А. - 1989. - №7. - С.73-76.
6. Воскресенский Г.В., Журав С.М. Излучение из плоского волновода с фланцем // Радиотехника и электроника. - 1976. - Т.21, № 7. - С. 1390-1395.
7. Воскресенский Г.В., Журав С.М. Излучение из открытого конца плоского волновода с диэлектрическим заполнением // Радиотехника и электроника. - 1976. - Т.21, №12. - С. 2608-2614.
8. Воскресенский Г.В., Журав С.М., Коршунков С.Ю. Излучение магнитных волн из плоского волновода с фланцем // Радиотехника и электроника. - 1986. - Т.31, № 7. - С. 1441-1444.
9. Слепян А.Я. Дифракция электромагнитных волн на открытом конце плоского нерегулярного волновода с фланцем // Радиотехника и электроника. - 1990. - Т.33, №3. - С. 334-341.
10. Мануйлов Б.Д., Мелешин А.С. Излучение из плоского волновода с фланцем через щель в металлическом экране конечной толщины // Радиотехника и электроника. - 1984. - Т.29, №5. - С. 886-893.
11. Сверделенко Е.А., Ткаченко В.И., Кириленко А.А. Излучение смеси H20 и H10 волн из двухмодового волновода // Рассеяние электромагнитных волн: Междунедом. тематич. научн. сб. - Таганрог, 1989. - Вып. 7. - С. 72-74.

Надійшла 26.04.2004
Після доробки 26.10.2004

Об'єкти дослідження - площинні випромінювачі з фланцями, що збуджуються хвилювачами і мають багатокутний контур провідних поверхонь та кусково-однорідне діелектричне заповнення. Мета - створення єдиної математичної моделі електромагнітних полів вказаних випромінювачів. Метод - спільне застосування узагальнених матриць розсіяння і методу добутку областей. Приведені результати чисельних експериментів.

The objects under study are two-dimensional waveguide-fed radiating structures having multiangular conducting boundaries with flanges and a piecewise homogeneous filling. The goal is to develop a unified mathematical model of the above radiators. The method used is a hybrid approach based on generalized scattering matrices and the domain-product technique. Results of the numerical experiments are given.

РАДИОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ

РАДИОЕЛЕКТРОНИКА И ТЕЛЕКОММУНИКАЦИИ

RADIO ELECTRONICS AND TELECOMMUNICATIONS

УДК 62-55:681.515

В. И. Гостев

СИСТЕМА РЕГУЛИРОВАНИЯ ТЕМПЕРАТУРЫ ТЕПЛОНОСИТЕЛЯ НА ВЫХОДЕ СМЕСИТЕЛЯ С НЕЧЕТКИМИ РЕГУЛЯТОРАМИ

Исследована система регулирования температуры теплоносителя на выходе смесителя с нечеткими регуляторами, которые кроме температуры регулируют также объемные расходы исходных теплоносителей. Получены показатели качества системы в переходном и установившемся режимах.

ВВЕДЕНИЕ

Системы регулирования температуры теплоносителя на выходе смесителя, основанные на смешении двух исходных компонент, широко используются в химической и пищевой промышленности. К таким системам предъявляются достаточно высокие требования. Например, отклонение регулируемой температуры теплоносителя от заданного значения в установившемся режиме должно быть не более 0,5 °С, время регулирования не более 120 с, перерегулирование не более 3% [1]. Целью управления является обеспечение требуемого расхода и температуры теплоносителя на выходе удаленной магистральной при изменениях температуры и объема исходных смешиваемых носителей. В данной работе методом математического моделирования в интерактивной системе MATLAB исследована система регулирования температуры теплоносителя на выходе смесителя с нечеткими регуляторами (регуляторами, работающими на базе нечеткой логики) и получены показатели качества системы в переходном и установившемся режимах.

РЕШЕНИЕ ЗАДАЧИ

Заданная температура теплоносителя θ_3 поддерживается за счет смешения двух исходных теплоносителей (например, холодной и горячей воды или пара). В процессе функционирования смесителя регулируются объемные расходы горячей V_r и холодной V_x воды при поддержании заданного расхода теплоносителя на выходе смесителя, равного $V = V_r + V_x$.

При смешении без отвода тепла двух жидкостей с разными температурами в установившемся режиме значения температуры θ и объема V смеси связаны следующими уравнениями:

$$V_r(\theta - \theta_r) + V_x(\theta - \theta_x) = 0; \quad V = V_r + V_x, \quad (1)$$

$$\theta = (V_r\theta_r + V_x\theta_x) / V. \quad (2)$$

При требуемых (заданных) температуре θ_3 и объеме V_3 теплоносителя на выходе смесителя заданные объемы горячей и холодной воды вычисляются из уравнений (1) по формулам:

$$V_r = \frac{V_3(\theta_3 - \theta_x)}{\theta_r - \theta_x}; \quad V_x = \frac{V_3(\theta_r - \theta_3)}{\theta_r - \theta_x}.$$

Структурная схема многосвязной системы управления температурой и расходом теплоносителя представлена на рис. 1. Пунктирный квадрат 1 в объекте управления составлен на основе формулы (2). Пунктирный квадрат 2 на входе регуляторов составлен на основе формулы (3).

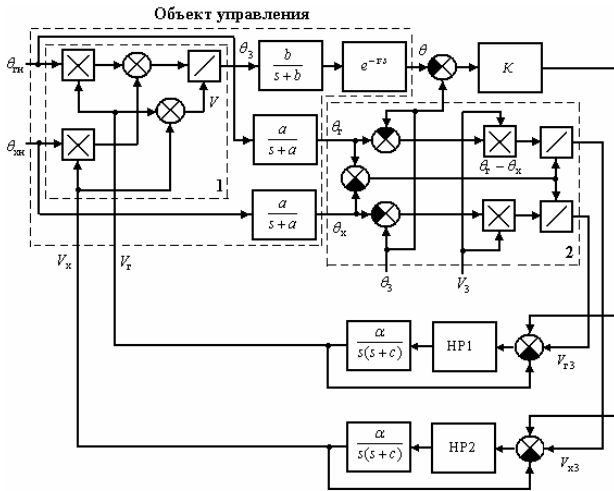


Рисунок 1

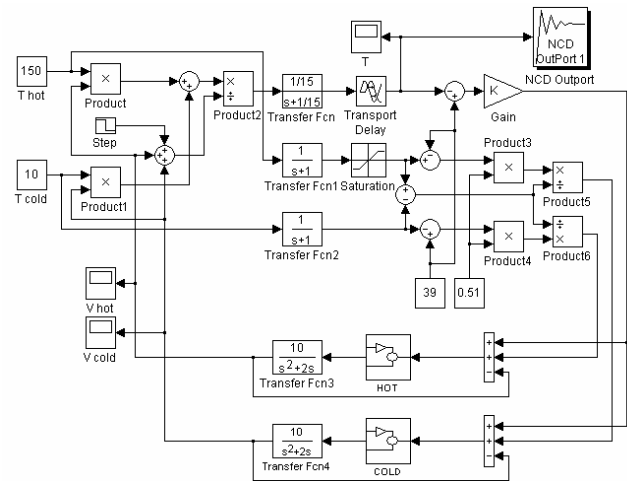


Рисунок 2

Температуры горячего и холодного исходных теплоносителей обозначены соответственно $\theta_{\text{ГН}}$ и $\theta_{\text{ХН}}$, а измеренные – $\theta_{\text{Г}}$ и $\theta_{\text{Х}}$.

Динамические свойства датчиков температуры горячего и холодного теплоносителей описываются передаточной функцией

$$G_{\text{Г}}(s) = G_{\text{Х}}(s) = \frac{a}{s+a}. \quad (3)$$

В установившемся режиме по объемным расходам горячей (пара) и холодной воды при $\theta_{\text{Г}} = \theta_{\text{ГН}}$ и $\theta_{\text{Х}} = \theta_{\text{ХН}}$ имеем:

$$V_{\text{Г3}} = \frac{V_3(\theta_3 - \theta_{\text{ХН}})}{\theta_{\text{ГН}} - \theta_{\text{ХН}}}; \quad V_{\text{Х3}} = \frac{V_3(\theta_{\text{ГН}} - \theta_3)}{\theta_{\text{ГН}} - \theta_{\text{ХН}}} \quad (4)$$

и

$$\theta_3 = (V_{\text{Г3}}\theta_{\text{ГН}} + V_{\text{Х3}}\theta_{\text{ХН}}) / V. \quad (5)$$

Инерционные свойства датчика температуры смеси и динамика установления температуры на выходе смесителя с учетом времени прохождения жидкости от исполнительных механизмов до датчика температуры смеси характеризуется передаточной функцией

$$G_0(s) = \frac{b}{s+b} e^{-\tau \cdot s}. \quad (6)$$

Динамические свойства исполнительных механизмов, в качестве которых используются двигатели, управляющие заслонками, описываются передаточной функцией

$$G(s) = \frac{\alpha}{s(s+c)}. \quad (7)$$

Глубина коррекции температуры смеси задается коэффициентом K .

Математическая модель системы регулирования температуры теплоносителя с нечеткими регуляторами, составленная в интерактивной системе MATLAB, представлена на рис. 2.

При моделировании приняты следующие значения величин и параметров передаточных функций, изображенных на рис. 1:

$$\theta_{\text{ГН}} = 150^\circ \text{C}; \quad \theta_{\text{ХН}} = 10^\circ \text{C}; \quad \theta_3 = 39^\circ \text{C}; \quad V_3 = 0,51;$$

$$b = 1/15c^{-1}; \quad a = 1c^{-1}; \quad c = 2c^{-1}; \quad \tau = 5.$$

При заданных значениях величин согласно формуле (5) в установившемся режиме значения объемных расходов горячей (пара) и холодной воды определяются:

$$V_{\text{Г3}} = \frac{0,51(39-10)}{150-10} \cong 0,106;$$

$$V_{\text{Х3}} = \frac{0,51(150-39)}{150-10} \cong 0,404; \quad V_{\text{Г3}} + V_{\text{Х3}} = V_3 = 0,51.$$

На рис. 2 параметры $\theta_{\text{ГН}}$ и $\theta_{\text{ХН}}$ обозначены соответственно $T \text{ hot}$ и $T \text{ cold}$, $V_{\text{Г}}$ и $V_{\text{Х}}$ соответственно $V \text{ hot}$ и $V \text{ cold}$, а параметр θ на выходе объекта управления обозначен T .

Нечеткие регуляторы HOT и COLD выполнены по идентичным схемам. Полная схема нечеткого регулятора HOT представлена на рис. 3.

На входе регулятора непрерывный сигнал преобразуется в цифровой аналого-цифровым преобразователем АЦП Zero-Order Hold, а на выходе цифровой сигнал преобразуется в аналоговый цифроаналоговым преобразователем ЦАП Zero-Order Hold1 (см. рис. 3, а).

Синтез нечетких регуляторов HOT и COLD выполнен по формулам (3.12)–(3.27) из работы [2] для треугольных функций принадлежности с шагом квантования (шагом поступления данных в нечеткий регулятор) $h = 0,01 \text{ с}$.

В каждом нечетком регуляторе настраиваются диапазоны изменения входных и выходной переменных $[\theta_{\text{min}}, \theta_{\text{max}}]$, $[\dot{\theta}_{\text{min}}, \dot{\theta}_{\text{max}}]$, $[\ddot{\theta}_{\text{min}}, \ddot{\theta}_{\text{max}}]$, $[m_{\text{min}}, m_{\text{max}}]$ треугольных функций принадлежности: $\mu^1(u) = 1 - u$, $\mu^2(u) = u$, где u – параметр (элемент) единого универсального множества $U = [0,1]$.

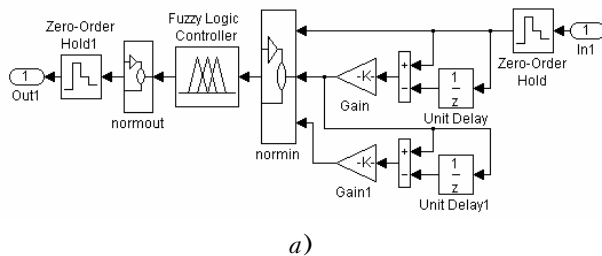
Для упрощения нормировки (пересчета значений сигналов в значения элементов единого универсального множества) и уменьшения числа параметров настройки диапазоны изменения входных и выходного сигналов (параметров нечеткого регулятора) принимаем симметричными:

$$\theta_{\max} = -\theta_{\min}; \dot{\theta}_{\max} = -\dot{\theta}_{\min}; \ddot{\theta}_{\max} = -\ddot{\theta}_{\min}; m_{\max} = -m_{\min}.$$

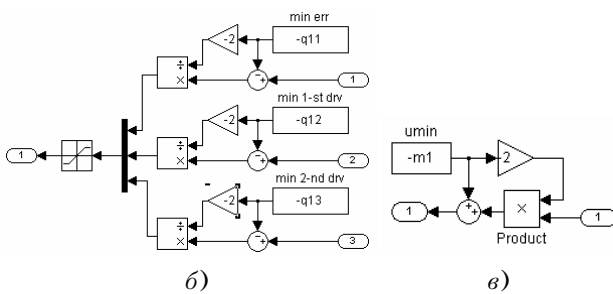
При этом пересчет значений сигналов в значения элементов единого универсального множества производится по формулам [2]:

$$u_1^* = -(\theta^* - \theta_{\min}) / (2\theta_{\min}); \quad u_2^* = -(\dot{\theta}^* - \dot{\theta}_{\min}) / (2\dot{\theta}_{\min});$$

$$u_3^* = -(\ddot{\theta}^* - \ddot{\theta}_{\min}) / (2\ddot{\theta}_{\min}); \quad m^* = m_{\min}(1 - 2u_c^*).$$



а)



б)

в)

Рисунок 3

На основании этих формул построены структурные схемы блоков нормировки входных (normin) и выходного (normout) параметров регулятора (см. рис. 3, б, в). Значения диапазонов q_{1i} , q_{2i} , $i=1,2,3$, $m1$ и $m2$ при настройке (оптимизации параметров) каждого нечеткого регулятора подбираются либо вручную, либо автоматически путем решения оптимизационной задачи. При настройке уточняется также значение коэффициента K .

Отметим, что при настройке нечетких регуляторов в интерактивной системе MATLAB целесообразно использовать блок NCD (Nonlinear Control Design), который реализует метод динамической оптимизации для проектирования систем управления. Этот инструмент, разработанный для использования с Simulink, автоматически настраивает системные параметры, основываясь на определенных ограничениях на временные характеристики (например, время регулирования и перерегулирование для реакции на ступенчатое воздействие или пределы для текущей ошибки рассогласования).

Ошибка на выходе АЦП $\theta(k)$ в каждом канале управления, ее первая $\dot{\theta}(k) = [\theta(k) - \theta(k-1)]/h$ и вторая $\ddot{\theta}(k) = [\dot{\theta}(k) - \dot{\theta}(k-1)]/h$ разности подаются на вход соответствующего блока нормировки входных (normin) параметров. В центральном блоке нечеткого регулятора Fuzzy Logic Controller (см. рис. 3, а) выбираются функции принадлежности membership functions и задается база правил rules. Сигнал с этого блока подается на вход блока нормировки выходного (normout) параметра. Сигнал с выхода каждого блока нормировки выходного параметра нечеткого регулятора поступает на ЦАП (фиксатор нулевого порядка с передаточной функцией $H(s) = (1 - e^{-hs})/s$) и далее на вход соответствующего исполнительного механизма.

В результате настройки получены следующие оптимальные параметры нечетких регуляторов и коэффициента K :

$$q_{11}=0.024; \quad q_{12}=0.3293; \quad q_{13} = 55.3936; \quad m1=27.4152;$$

$$q_{21}=0.9675; \quad q_{22}=0.799; \quad q_{23}=58.2343; \quad m2=28.0226;$$

$$K=0.0127.$$

На рис. 4 приведены осциллограммы установления температуры теплоносителя на выходе смесителя при подаче на вход смесителя исходных теплоносителей (пара с температурой $\theta_{\text{TH}} = 150^\circ \text{C}$ и холодной воды с температурой $\theta_{\text{XH}} = 10^\circ \text{C}$). При перерегулировании в 1°C , что соответствует перерегулированию 2,5% (участок кривой с перерегулированием отдельно приведен на рис. 4, б), требуемая температура на выходе смесителя $\theta_3 = 39^\circ \text{C}$ устанавливается с учетом чистого запаздывания за 22 секунды и после окончания переходного процесса отклонение регулируемой температуры от заданного значения в установившемся режиме равно нулю.

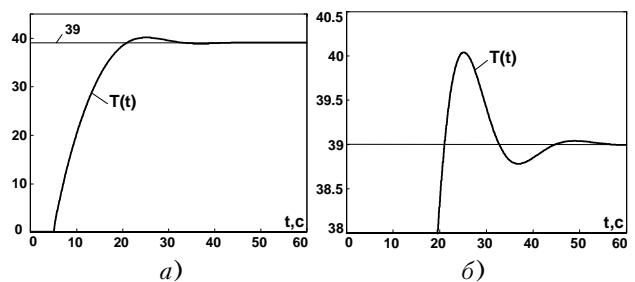


Рисунок 4

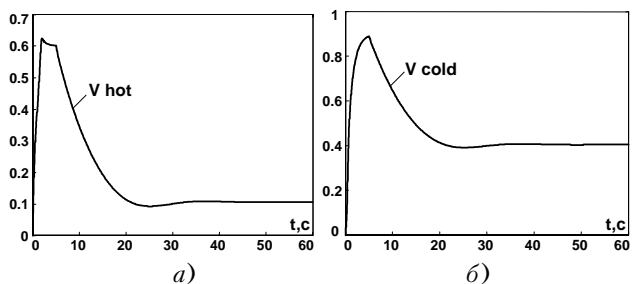


Рисунок 5

На рис. 5 приведены осциллограммы изменения объемных расходов пара и холодной воды при подаче на вход смесителя исходных теплоносителей. После окончания переходных процессов за 42 секунды устанавливаются заданные значения объемных расходов пара и холодной воды: $V_{r3} = 0,106$; $V_{x3} = 0,404$. Исследование системы методом математического моделирования показывает, что применение нечетких регуляторов позволяет спроектировать систему регулирования температуры теплоносителя весьма высокого качества: система обладает достаточным быстродействием (время установления заданной температуры на выходе смесителя с учетом чистого запаздывания составляет 22 с, перерегулирование не превышает 2,5%) и нулевой ошибкой в установившемся режиме. Поэтому применение нечетких регуляторов для таких систем целесообразно и перспективно.

ПЕРЕЧЕНЬ ССЫЛОК

1. Интеллектуальные системы автоматического управления / Под ред. И.М. Макарова, В.М. Лохина. – М.: ФИЗМАТЛИТ, 2001. – 576 с.
2. Гостев В.И. Синтез нечетких регуляторов систем автоматического управления. – К.: Издательство "Радиоаматор", 2003. – 512 с.

Надійшла 22.11.2004

Досліджена система регулювання температури теплоносія на виході змішувача з нечіткими регуляторами, які крім температури регулюють також об'ємні витрати початкових теплоносіїв. Отримані показники якості системи в перехідному та сталому режимах.

The regulating system of temperature of the heat-carrier on an output of the mixer with the fuzzy controllers, which regulate not only temperature but also volume flow rate of the initial heat-carriers, is probed. The indexes of quality of a system in transient state and steady regime are obtained.

УДК 681.391

В. Н. Журавлев, В. С. Кабак, В. О. Рыбин

АНАЛИЗ ПАРАМЕТРОВ АВТОКОРРЕЛЯЦИОННОЙ ФУНКЦИИ ФРАГМЕНТОВ ФОНЕМ В ЗАДАЧАХ ИДЕНТИФИКАЦИИ АБОНЕНТА

В статье рассмотрен алгоритм цифровой обработки идентификационного фрагмента фонемы, позволяющий на основе анализа интервала корреляции автокорреляционной функции фонемы повысить информационную эффективность матрицы идентификации.

ВВЕДЕНИЕ

Основой обеспечения безопасности и живучести телекоммуникационных систем и сетей передачи данных является проблема идентификации абонентов, имеющих доступ к конфиденциальной информации. Отождествление абонента точки доступа – комплекс задач, решение которых позволяет организовать весь процесс управления правами доступа, а также реализовать ряд других вспомогательных решений, имеющих самостоятельное прикладное значение.

Рассматриваемая в статье проблема биометрической идентификации статических и динамических параметров и характеристик речеобразующего тракта абонента нашла свое отражение в трудах лингвистов [1] и психоакустиков [2]. Однако ввиду различных подходов к определению параметров модели речеобразующего аппарата, большим разнообразием параметров речевых сигналов, которые, в большинстве, носят квазистационарный и субъективный характер, данная проблема является предметом исследований многих научно-технических коллективов, как в нашей стране, так и за границей [3–5].

Обычно в качестве фрагмента речевого сигнала принимают фонему, как наименьшую часть устной осмыс-

ленной речи, которая не может быть разложена на меньшие речевые элементы [6]. Объективные предпосылки идентификации абонента по индивидуальным особенностям в фонамах русского языка исследованы в [7].

Для расчета параметров и характеристик, идентифицирующих физиологические особенности речеобразующего тракта, наиболее часто применяются классические методы спектрально-временного анализа [8]. Такие методы анализа речевого сигнала адекватны природному механизму восприятия речи [9], что позволяет отождествлять индивидуальные особенности при анализе интегральных спектральных распределений отдельных фонем и анализе распределений текущего спектра фрагмента речевого сигнала.

Функциональная схема системы преобразования, обработки речевого сигнала, а также синтеза матрицы динамических спектральных состояний его параметров и временного образа сигнала идентификации [10] представлена на рисунке 1.



Рисунок 1 – Функциональная схема алгоритма обработки речевого сигнала

В блоке 1 производится сегментация, взвешивание и нормализация фрагмента речевого сигнала $s(t)$ во временном окне выборки-наблюдения $W(T)$ с постоянной

времени слуха T [1]. Результатом данного преобразования является сигнал $s_1(t)$

$$s_1(t) = W [s(t), t \in T]. \quad (1)$$

В блоке 2 осуществляется ортогональное преобразование $F(\omega)$ сигнала $s_1(t)$ с переносом его в частотную область. Результатом данного преобразования является сигнал $s_2(\omega)$

$$s_2(\omega) = F[s_1(t)]. \quad (2)$$

Результатом работы блока 3, который, анализируя сигнал $s_2(\omega)$, выделяет индивидуальные особенности спектральных составляющих динамики работы речеобразующего тракта абонента, является матрица динамических спектральных составляющих $\mathbf{SC}[s_2(\omega)]$ размерности N .

Блок 4 синтезирует (SNT) значения образа идентифицирующего сигнала $s'(t)$, производя обратные ортогональные преобразования $F^{-1}(SC)$ над определенными значениями матрицы $\mathbf{SC}[s_2(\omega)]$ и их постобработку по специальному алгоритму Pr

$$s'(t) = \text{SNT}[F^{-1}(SC), \text{Pr}]. \quad (3)$$

Рассмотренный выше алгоритм позволяет анализировать фрагменты речевых сигналов на временном отрезке T , определяемым, с одной стороны, частотным разрешением алгоритмов спектральных преобразований $F(\omega)$, и, с другой стороны, постоянной времени квазидетерминированного состояния речевого сигнала T .

Вышеуказанные ограничения не позволяют учитывать девиацию спектральных составляющих фонемы внутри интервала времени T , которые зависят не только от физиологических параметров речеобразующего тракта, но и от темпа речи, а также эмоционального состояния абонента.

По нашему мнению, повышение эффективности существующего алгоритма идентификации можно достигнуть, введя в матрицу динамических спектральных составляющих дополнительный параметр, в качестве которого можно предложить интервал корреляции автокорреляционной функции (АКФ) фрагмента фонемы.

В данной статье предлагается метод повышения эффективности матрицы динамических спектральных составляющих сегмента речевого сигнала на основе дополнения идентификационной базы данных значением интервала корреляции фрагмента фонемы, текущее значение которого позволит учитывать изменение параметров спектральных составляющих фонемы внутри интервала идентификации.

ПОСТАНОВКА ЗАДАЧИ

Будем считать, что фрагмент фонемы речевого сигнала $s(t)$ телекоммуникационной системы связи, идентифицируемый на промежутке времени T , соответствует стыку тональных частот (ТЧ) и представляет сумму следующих сигналов:

- импульсов сигнала основного тона $s_{от}(t)$;

- суммы формант $\Sigma s_i(t)$, $i \in 1, n$, где n – количество частотных полос, в которых идентифицируются форманты исследуемой фонемы;

- стохастических шумов $s_{ш}(t)$, которые некоррелированы с сигналами $s_{от}(t)$ и $\Sigma s_i(t)$

$$s(t) = s_{от}(t) + \Sigma s_i(t) + s_{ш}(t), t \in T, i \in 1, n. \quad (4)$$

При произношении слитной речи артикуляционные органы речеобразующего аппарата человека находятся в постоянном движении, параметры которого определяются информационным составом сообщения.

Как известно, фонетические признаки речи можно охарактеризовать функцией, описывающей изменение спектров во времени [6], которая определена как логарифм отношения спектральных интенсивностей S_ω и $S_\omega(t-\tau)$ на частоте ω , при этом спектры рассчитываются для сигналов, отстоящих на интервал времени $\tau \in 0, T$. Однако, в [6] не приведены критерии выбора и определения параметра τ .

Сигнал $s(t)$ на промежутке времени идентификации T является сигналом с конечной энергией, не содержит дельта-функций и разрывов. Исходя из этого, можно предположить, что исследование АКФ $R_{ss}(\tau)$ фрагмента фонемы на промежутке времени идентификации T позволит уточнить параметры изменения сигнала $s(t)$

$$R_{ss}(t) = \int_0^T s(t)s(t-\tau)dt. \quad (5)$$

При некотором значении $\tau = \tau_0$ интеграл (5) будет иметь первый локальный минимум

$$\frac{d[R_{ss}(\tau_0)]}{dt} = 0, \quad (6)$$

а значение τ_0 определяется в [11] как интервал корреляции. Для данной задачи предлагается рассматривать τ_0 как идентификационный параметр, характеризующий фонетические признаки речи.

Представим (4) как:

$$s(T) = s_d(T) + s_{ш}(T), \quad (7)$$

где $s_d(T)$ – детерминированная составляющая сигнала $s(t)$ на интервале времени T и равная, в соответствии с (4),

$$s_d(T) = s_{от}(T) + \Sigma s_i(T), i \in 1, n). \quad (8)$$

Так как стохастическая составляющая сигнала $s_{ш}(T)$ некоррелирована с сигналом $s_d(T)$, то она не содержится в АКФ $R_{ss}(\tau)$.

Введем функцию $F_n(s)$, которая определяет алгоритм расчета значений строк i и столбцов j матрицы идентификационных параметров \mathbf{M}_n сигнала $s(t)$

$$\mathbf{M}_n[i, j; i \in 1, n; j \in 1, m] = F_n [R_{ss}(t)], \quad (9)$$

где n – номер частотной полосы, m – номер временного интервала τ_0 в частотной полосе.

Как известно [11], АКФ $R_{ss}(t)$ сигнала $s(t)$ обладает следующими свойствами:

- несет идентификационную информацию о минимальной постоянной времени (интервал корреляции τ_0) сигнала $s(t)$;
- содержит детерминированные (на интервале идентификации T) спектральные составляющие речевого сигнала $s(t)$.

Таким образом, исходя из вышеизложенного, эффективность идентификации сигнала фонемы внутри интервала идентификации T можно повысить, проведя анализ спектральных составляющих на интервале корреляции τ_0 АКФ $R_{ss}(t)$ фрагмента речевого сигнала $s(t)$.

ПРЕДЛАГАЕМЫЕ АЛГОРИТМЫ И РЕШЕНИЯ

Сигнал $s(t)$ преобразовывается в блоке 1 (рис. 2) в цифровую форму $s(\Delta t)$ аналого-цифровым преобразователем АЦП на основании обобщенной теоремы отсчетов [11] с частотой дискретизации f_s , которая должна быть не менее удвоенной верхней частоты $f_B^{TЧ}$ канала ТЧ [12]

$$f_s = f_B^{TЧ}. \quad (10)$$

Для нормирования частотного диапазона сигнал $s(\Delta t)$ обрабатывается в блоке 2 фильтром с передаточной характеристикой $H_{TЧ}(f_{TЧ})$, соответствующей каналу ТЧ

$$s_{TЧ}(\Delta t) = s(\Delta t)H_{TЧ}(f_{TЧ}). \quad (11)$$

Блок 3 осуществляет оконное $W(d, k)$ сегментирование сигнала $s_{TЧ}(\Delta t)$ с параметрами:

d – определяет длительность сегмента идентификации T , исходя из количества анализируемых отсчетов сиг-

нала $s_{TЧ}(\Delta t)$ на основании нижней частоты $f_B^{TЧ}$ канала ТЧ и частотного разрешения Δf_{FFT} алгоритмов спектральных преобразований

$$d = f(f_B^{TЧ}, \Delta f_{FFT}), \quad (12)$$

k – определяет количество сегментов идентификации и является функцией от параметра d , длительности фонемы T_Φ , а также коэффициента перекрытия сегментов идентификации $K_{пер}$, который может изменяться от нуля до 0,5 [6]

$$k = f(d, T_\Phi, K_{пер} \in 0, 0,5). \quad (13)$$

Идентификацию изменения формантных частот [13] по i полосам (со средними частотами f_i) осуществляют блоки фильтров 4...6, с передаточными характеристиками $H_i(f_i)$, на выходах которых образуются сигналы

$$S_i(f_i) = s_{TЧ}(\Delta t, T)H_i(f_i). \quad (14)$$

Огибающая амплитуды $S_i(f_0)$ сигналов $S_i(f_i)$ рассчитывается в блоках 7...9, имеющих передаточную характеристику $H_d(f_i)$

$$S_i(f_0) = S_i(f_i)H_d(f_i). \quad (15)$$

В блоке 13 рассчитывается (5) корреляционная функция $R_{ss}(\Delta t)$. В соответствии с (6) рассчитываются параметры интервала корреляции τ_0 . Сигналы $S_i(f_0)$ статистически обрабатываются в блоках 10...12 с учетом интервала корреляции τ_0 идентифицируемого сигнала $s_{TЧ}(\Delta t)$. Математическое ожидание M_i и дисперсия σ_i^2 сигнала $s_{TЧ}(\Delta t)$ на интервале τ_0 записываются блоком 14 в матрицу идентификации $\mathbf{M}_n(i, j)$.

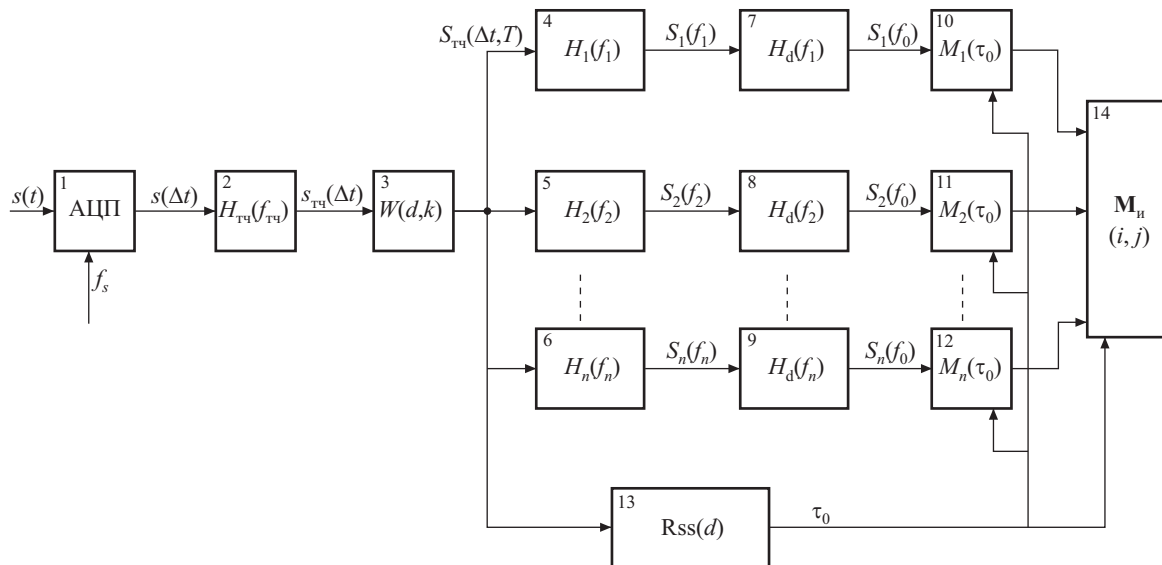


Рисунок 2 – Функциональная схема алгоритма идентификации параметров фрагментов речевого сигнала

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

По предлагаемому алгоритму был произведен расчет параметров фонемы «О» слова «ЛЕТО» в программной среде MatLab 6.5 и получены следующие графики (рис. 3–7).

Речевой сигнал записывался на жесткий диск ПЭВМ с частотой дискретизации 44,1 кГц и 16-разрядным квантованием по уровню, затем сегментировался на фонемы длительностью T_{Φ} (рис. 3).

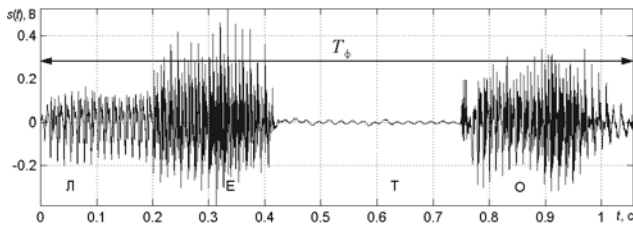


Рисунок 3 – График фонемы «О» на интервале $T_{\Phi} = 0,092$ с

Сигнал фонемы $s(T_{\Phi})$ дискретно сканировался временным окном, определяемым постоянной времени слуха T , с коэффициентом перекрытия (14) $K_{\text{пер}} = 0,5$. В качестве параметра стабильности характеристик сигнала $s(T)$ на интервале времени T_{Φ} была принята минимальная разница между коэффициентами ковариации, которые рассчитывались для окон, примыкающих к идентифицируемому сегменту. На сегменте времени $T = 0,046 \dots 0,069$ с сигнал фонемы $s(T_{\Phi})$ удовлетворяет выбранному параметру стабильности. На нем была рассчитана спектральная плотность $S(\omega)$ в полосе частот ТЧ (рис. 4).

Как видно из рисунка 4, спектральная плотность фонемы на взятом интервале идентификации сосредоточена

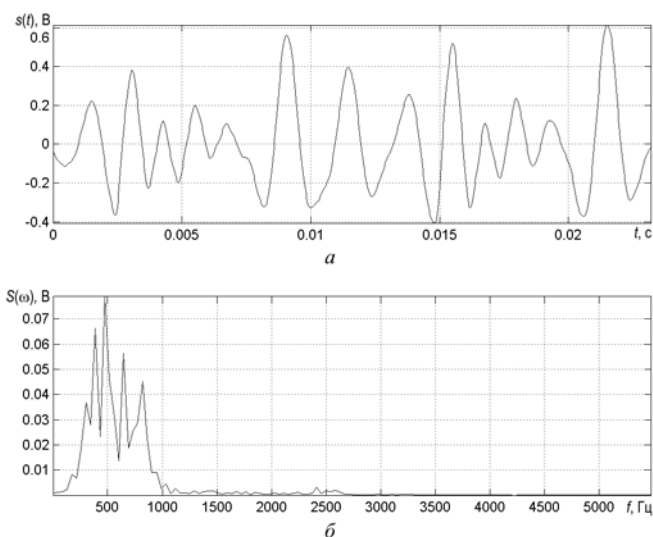


Рисунок 4 – Графики сегмента фонемы «О» на интервале идентификации $T = 23$ мс (а) и спектра сегмента фонемы «О» в полосе частот 300–3400 Гц (б)

(по уровню 0,05) в полосе частот 300–1000 Гц и содержит спектральные составляющие на частотах: 300, 400, 490, 600 и 740 Гц.

В соответствии с выражением (5) рассчитана АКФ сегмента фонемы (рис. 5) и определен интервал корреляции, который равен $\tau_0 = 0,920$ мс.

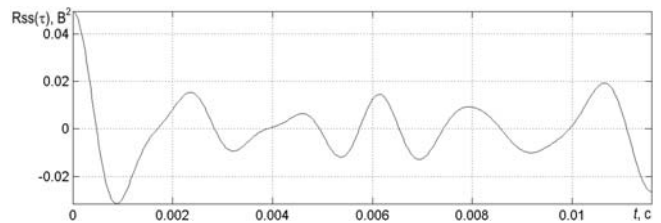


Рисунок 5 – График АКФ $R_{ss}(\tau) = f(t)$ сегмента фонемы «О» на половине времени идентификации

Затем была проведена фильтрация выбранного сигнала сегмента идентификации блоком из n полосовых фильтров с граничными частотами, которые соответствуют равноартикуляционным полосам [13]: 300–400; 400–510; 510–630; 630–700; 770–915; 915–1250; 1250–1475; 1475–1710; 1710–1990; 1990–2310; 2310–2690; 2690–3125 Гц.

Исследованы сигналы сегмента идентификации во 2-й ($n = 2$, 400–510 Гц) и 4-й ($n = 4$, 630–700 Гц) частотных полосах (рисунки 6, 7), в которых сосредоточены основные спектральные составляющие фонемы.

При расчете огибающей сигнала фонемы после полосового фильтра данные разбивались на участки длительностью $\tau_0 = 0,920$ мс, в которых рассчитывались математическое ожидание $M_n[S_n(t)]$ сигнала и его дисперсия.

Как видно из рисунков 5–7, интервал корреляции τ_0 является информационным параметром сигнала фонемы,

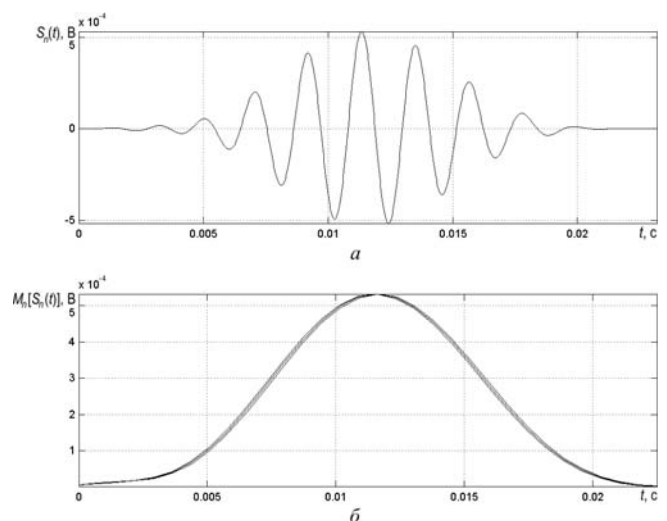


Рисунок 6 – Графики сигнала $S_2(t)$ ($n = 2$, 400–510 Гц) фонемы «О» на интервале идентификации 23 мс (а) и математического ожидания его огибающей $M_n[S_2(t)]$ во второй полосе частот (б)

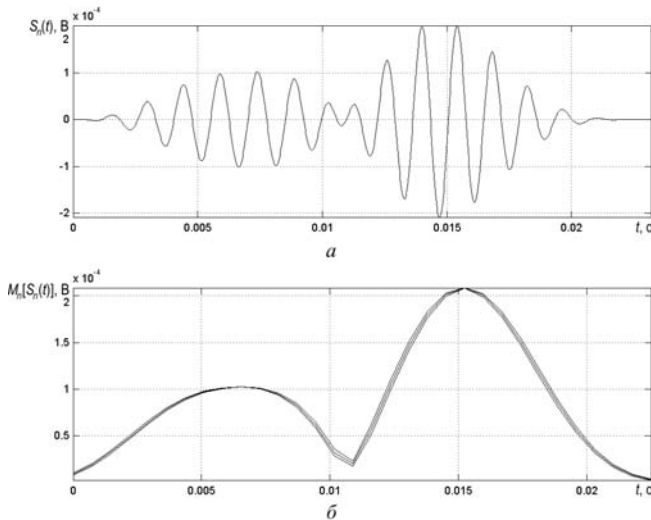


Рисунок 7 – Графики сигнала $S_4(t)$ ($n=4$, 630–700 Гц) фонемы «О» на интервале идентификации 23 мс (а) и математического ожидания его огибающей $M_n[S_4(t)]$ в четвертой полосе частот (б)

позволяющим проводить статистическую обработку результатов идентификации сегмента речевого сигнала. Включение данного параметра в матрицу идентификации позволит повысить ее информативность.

ВЫВОДЫ

Полученные результаты позволяют сделать вывод о целесообразности применения предложенного алгоритма при идентификации абонента по фрагментам речевого сигнала.

Повышение эффективности известных алгоритмов идентификации достигается дополнением матрицы динамических спектральных составляющих параметрами статистического анализа спектрального состава формант речевого сигнала $s(t)$ на отрезках времени, равных интервалу корреляции τ_0 его АКФ.

При расчете линии регрессии реализаций сигнала $s(t)$ можно, в дальнейшем, применять математический аспа-

рат множественного корреляционного анализа на интервале корреляции τ_0 .

ПЕРЕЧЕНЬ ССЫЛОК

1. Михайлов В.Г., Златоустова Л.В. Измерение параметров речи / Под ред. М.А. Сапожкова. – М.: Радио и связь, 1987. – 168 с.
2. Алдошина И.А. Основы психоакустики // Звукорежиссер. – 2000. – №6. – С. 36–40.
3. Кобиелус Джеймс. Информационная безопасность: идентификация и аутентификация. – М.: Связь, 1997. – 252 с.
4. Белінський В. та ін. Апаратно-програмний комплекс визначення розбірливості українського мовлення та апаратні засоби контролю службових приміщень від витoku мовної інформації в електроакустичних та віброакустичних каналах передавання інформації // Матеріали 3-ї науково-технічної конференції «Правове, нормативне та метрологічне забезпечення систем захисту інформації в Україні». – К.: – 2001. – С. 109–110.
5. Аграновский А.В., Леднов Д.А., Репалов С.А. Исследование проблемы верификации дикторов, с использованием формантных характеристик // Материалы III Всероссийской конференции «Теория и практика речевых исследований» (АРСО–2003). – М.: – 2003. – С. 3–7.
6. Вокодерная телефония. Методы и проблемы. Под ред. А.А. Пирогова – М.: Связь, 1974. – 536 с.
7. Рамишвили Г.С. Автоматическое опознавание говорящего по голосу. – М.: Радио и связь, 1981. – 224 с.
8. Рабинер Л., Шафер Р. Цифровая обработка речевых сигналов. – М.: Радио и связь, 1981. – 496 с.
9. Цвикер Э., Фельдкеллер Р. Ухо как приемник информации. Пер. с нем. под ред. Б.Г. Белкина. – М.: Связь, 1971. – 225 с.
10. Петраков А.В., Лагутин В.С. Защита абонентского телеграфика. – М.: Радио и связь, 2001. – 504 с.
11. Гоноровский И.С. Радиотехнические цепи и сигналы: Учебник для вузов. – М.: Радио и связь, 1986. – 512 с.
12. ГОСТ 21655–87. Каналы и тракты магистральной первичной сети единой автоматизированной системы связи. Электрические параметры и методы измерений. – М.: Издательство стандартов, 1989.
13. Вемян Г.В. Передача речи по сетям электросвязи. – М.: Радио и связь, 1985. – 272 с.

Надійшла 12.01.2004
Після доробки 07.10.2004

В статті розглянутий алгоритм цифрової обробки ідентифікаційного фрагменту фонем, що дозволяє на основі аналізу інтервалу кореляції автокореляційної функції фонем підвищити інформаційну ефективність матриці ідентифікації.

The algorithm of phoneme identification fragment digital processing, which allows to increase the informational effectiveness of identification matrix on the base of phoneme autocorrelation function correlation window analysis is under review.

УДК 621.396.6.004

М. М. Касьян, К. М. Касьян

СИНТЕЗ ВІДБРАКУВАЛЬНИХ ПОТОЧНИХ ДОПУСКІВ НА ПАРАМЕТРИ ЕЛЕМЕНТІВ У ЗАЛЕЖНОСТІ ВІД ЧАСУ ТА УМОВ ЕКСПЛУАТАЦІЇ

Пропонується метод синтезу поточних відбракувальних допусків на параметри елементів, необхідних для проведення діагностування пристроїв перетворення сигналів в процесі експлуатації. Математична база методу забезпечує компактність програмної реалізації при здійсненні автоматизації досліджень за допомогою ЕОМ.

ВСТУП

При розробці принципової електричної схеми пристроїв перетворення сигналів (ППС), за заданими у технічному завданні (ТЗ) допусками на вихідні харак-

теристики пристрою та ймовірними характеристиками, інженер-проектувальник розраховує загальні (інтегральні) допуски на електричні параметри електрорадіоелементів (ЕРЕ). При цьому він може використовувати різні методи синтезу допусків: метод рівних допусків, метод найгіршого випадку, метод синтезу через аналіз. Обраний для цього метод синтезу залежить від специфіки розглянутої схеми, від наявної елементної бази, від технологічного процесу виробництва обраної елементної бази, від виду конструкторського виконання (який залежить від умов експлуатації ППС) [1]. Розрахований інтегральний допуск на кожен ЕРЕ повинен враховувати усі впливи, зазначені в ТЗ на ППС: температуру, вологість, тиск, ядерну радіацію й інші фактори, а також заданий час збереження й експлуатації.

ПОСТАНОВКА ЗАДАЧІ

Перед проведенням діагностування на етапі експлуатації ППС необхідно розрахувати поточні (відбракувальні) допуски на параметри елементів для адекватного визначення технічного стану ППС на даний момент. При цьому потрібно визначити сумарні відбракувальні допуски на параметри ЕРЕ при їхньому діагностуванні в різних теплових режимах: 1) у момент включення, коли температури ЕРЕ практично дорівнюють температурі навколишнього середовища; 2) під час розігріву ППС; 3) після виходу ППС на стаціонарний тепловий режим, а також при різних значеннях часу їхньої роботи на момент діагностування ППС.

МЕТОДИ ВИРІШЕННЯ ТА РЕЗУЛЬТАТИ

При проектуванні схеми перед застосуванням конкретних ЕРЕ необхідно із синтезованого інтегрального допуску $\Delta q^{\text{інт}}$ виділити кожен складову: допуски на температуру $\Delta q^{\text{тем}}$ та на старіння $\Delta q^{\text{стар}}$. Залишок $\Delta q^{\text{зал}}$ враховує всі інші фактори (технологію виготовлення ЕРЕ, вологість, тиск, радіаційний вплив). З усіх факторів залишку значимою є технологія виготовлення ЕРЕ, якщо в ТЗ з усіх зовнішніх впливів указуються фактори: механічний, температурний, вологістний і тиск навколишнього середовища. Отже, температурна складова допуску і складова на старіння визначаються розрахунковим шляхом, а інші складові допуску визначаються як різниця

$$(\Delta q^{\text{зал}})^2 = (\Delta q^{\text{інт}})^2 - (\Delta q^{\text{тем}})^2 - (\Delta q^{\text{стар}})^2.$$

Складову допуску $\Delta q^{\text{зал}}$ можна представити

$$(\Delta q^{\text{зал}})^2 = (\Delta q^{\text{техн}})^2 + (\Delta q^{\text{інш}})^2.$$

Як правило, $\Delta q^{\text{техн}} \gg \Delta q^{\text{інш}}$. І тоді $\Delta q^{\text{зал}} \approx \Delta q^{\text{техн}}$.

Якщо технологічний допуск на даний тип елемента (здається заводом-виробником елементів) $\Delta q^{\text{техн}}$ не збі-

гається з $\Delta q^{\text{зал}}$, то необхідно застосовувати елементи з меншими технологічними допусками, тобто $\Delta q^{\text{техн}} < \Delta q^{\text{зал}}$. Однак може виявитися так, що в даного елемента схеми розраховані дві складові перевищать інтегральний допуск

$$(\Delta q^{\text{тем}})^2 = (\Delta q^{\text{стар}})^2 + (\Delta q^{\text{інт}})^2.$$

У цьому випадку перед застосуванням даного типу елемента необхідно передбачити в схемі його термостабілізацію, чи ввести термостабілізуючий електричний зворотній зв'язок, або застосовувати елементи з малими значеннями температурних коефіцієнтів параметрів (прецизійні елементи), а також з малими коефіцієнтами старіння (що веде до різкого подорожчання усього ППС), домагаючись щоб сума $(\Delta q^{\text{стар}})^2 + (\Delta q^{\text{тем}})^2$ прийняла значення, при якому

$$(\Delta q^{\text{тем}})^2 + (\Delta q^{\text{стар}})^2 + (\Delta q^{\text{техн}})^2 < (\Delta q^{\text{інт}})^2.$$

У цій нерівності значення $\Delta q^{\text{інт}}$ трохи більше суми трьох складових на деяку величину $\Delta q^{\text{інш}}$, що враховує усі фактори, що залишилися.

При проектуванні ППС неможливо врахувати весь комплекс зовнішніх факторів, що впливають, тому в ТЗ виділені основні зовнішні фактори, які впливають. Однак не можна виключити, що на практиці не буде інших, навіть у малих дозах і величинах. Їхній вплив на елементи відбито в значенні $\Delta q^{\text{інш}}$, на яке відрізняється $\Delta q^{\text{інт}}$ і три складові допуски. Тобто в даному випадку інженер-розроблювач крім технології виготовлення ЕРЕ виділяє з усієї сукупності факторів, що впливають, тільки тепловий режим ЕРЕ і можливе старіння матеріалів ЕРЕ, як найбільш впливові на відхід електричних параметрів. Перед застосуванням конкретного типу елемента інженер домагається виконання нерівності (1), не розраховуючи наскільки ліві і праві частини цієї нерівності відрізняються.

Таким чином, отримані інтегральні допуски $\Delta q^{\text{інт}}$ на кожен ЕРЕ є тією різницею, при якій, якщо значення їхніх параметрів знаходяться усередині цієї границі (допуску), то і будь-яка вихідна характеристика ППС буде залишатися усередині свого заданого допуску незалежно від того, з яких причин відбулися відхилення електричних параметрів ЕРЕ від своїх номінальних значень [2].

Розрахунок поточних відбракувальних допусків на параметр ЕРЕ здійснюється в наступній послідовності:

1. Визначаються математичні очікування і середньоквадратичні відхилення температурного коефіцієнта а і коефіцієнтів старіння параметра при збереженні β_{36} та при експлуатації β_e . Якщо відповідні коефіцієнти задані нижніми і верхніми граничними значеннями і задані закони їхнього розподілу (у більшості випадків – нормальні), то математичні очікування і середньоквадратичні відхилення розраховуються за наступними формулами:

$$\left. \begin{aligned} m(\alpha) &= (\alpha^H + \alpha^B)/2; & \sigma(\alpha) &= [m(\alpha) - \alpha^H]/\chi; \\ m(\beta_{зб}) &= (\beta_{зб}^B + \beta_{зб}^H)/2; & \sigma(\beta_{зб}) &= [m(\beta_{зб}) - \beta_{зб}^H]/\chi; \\ m(\beta_e) &= (\beta_e^H + \beta_e^B)/2; & \sigma(\beta_e) &= [m(\beta_e) + \beta_e^H]/\chi. \end{aligned} \right\} (14)$$

де χ – коефіцієнт пропорційності (для нормального закону розподілу параметра q та ймовірності перебування параметра в межах допусків, рівної 0,9973, $\chi = 3$).

Якщо в довідковій літературі відсутні дані за коефіцієнтом $\beta_{зб}$ і приводяться тільки значення коефіцієнта старіння β_e при роботі ЕРЕ в межах номінального навантаження і при впливі температури, вологи і механічних навантажень припустимих ТУ, то можна прийняти $m(\beta_{зб}) = m(\beta_e)$ і $\sigma(\beta_{зб}) = \sigma(\beta_e)$.

2. Визначаються математичне очікування $m(q_{відх})$ відхиленого фактичного значення параметра $q_{відх}$ і середньоквадратичне відхилення $\sigma_{темп}(q_{відх})$, $\sigma_{зб}(q_{відх})$, $\sigma_e(q_{відх})$, обумовлених впливом температури, часу збереження і часу експлуатації. Якщо причини відмовлень елементів вивчені, технологічні процеси виконуються точно і матеріали мають високу чистоту, то деградаційні процеси можна вважати близькими до детермінованих

$$m(q_{відх}) = q_{ном} \cdot [1 + m(\alpha) \cdot \Delta T + m(\beta_{зб}) \cdot t_{зб} + m(\beta_e) \cdot t_e], \quad (15)$$

де $q_{ном}$ – номінальне значення параметра; $\Delta T = T_{роб} - 20$ °С – різниця між фактичною температурою елемента $T_{роб}$ і нормальною температурою; $T_{зб}$ – час збереження ЕРЕ; T_e – час експлуатації ЕРЕ;

$$\begin{aligned} \sigma_{темп}(q_{відх}) &= \Delta T \cdot \sigma(\alpha) \cdot q_{ном}; \\ \sigma_{зб}(q_{відх}) &= t_{зб} \cdot \sigma(\beta_{зб}) \cdot q_{ном}; \\ \sigma_e(q_{відх}) &= t_e \cdot \sigma(\beta_e) \cdot q_{ном}. \end{aligned} \quad (16)$$

3. Визначається сумарне середньоквадратичне відхилення (розкид) параметра $q_{відх}$ обумовлене впливом температури і фактора старіння. У цьому випадку здійснюється підсумовування всіх середньоквадратичних відхилень: технологічного, температурного й обумовленого старінням при збереженні й експлуатації. Визначення засноване на підсумовуванні дисперсій зазначених відхилень

$$\sigma(q_{відх}) = \sqrt{\sigma_{техн}^2(q_{відх}) + \sigma_{темп}^2(q_{відх}) + \sigma_{зб}^2(q_{відх}) + \sigma_e^2(q_{відх})}. \quad (17)$$

Таким чином, закон нормального розподілу не змінюється, а змінюються його параметри: математичне очікування і дисперсія. Це відбито на рисунку 1.

4. Визначаються границі перебування фактичного значення параметра q з урахуванням усіх зазначених вище факторів

$$\begin{aligned} q_B &= m(q_{відх}) + \chi_B \cdot \sigma(q_{відх}); \\ q_H &= m(q_{відх}) + \chi_H \cdot \sigma(q_{відх}). \end{aligned} \quad (18)$$

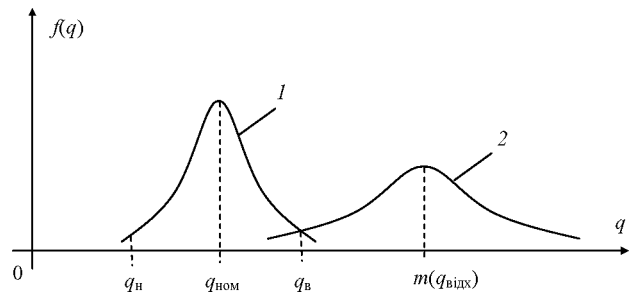


Рисунок 1 – Зміна параметрів нормального закону розподілу параметрів ЕРЕ з урахуванням технології виготовлення, температурного і часового факторів:

1 – відповідає технологічним розкидам параметрів ЕРЕ; 2 – відповідає відхиленням значенням параметрів ЕРЕ, обумовлене розкидами температурного коефіцієнта й коефіцієнтів старіння

А значення верхнього δ_q^B і нижнього δ_q^H відносних допусків на параметр ЕРЕ дорівнює

$$\delta_q^B = \frac{q^B}{q_{ном}} - 1; \quad \delta_q^H = 1 - \frac{q^H}{q_{ном}}.$$

При діагностуванні ППС у різних експлуатаційних умовах і на різних життєвих циклах його роботи необхідно знати поточний відбракувальний допуск, а кожен параметр елемента – критерій відбракування. Поточний відбракувальний допуск (максимальне відхилення) $\Delta q_{відб}$ варіюється в межах від $\Delta q_{техн}$ на елемент до

$$\Delta q_{відб} = \Delta q_{техн} + \Delta q_{темп} + \Delta q_{стар}, \quad (19)$$

у залежності від того, у якому тепловому режимі знаходиться кожен елемент і який час він пропрацював і знаходився у виключеному стані до моменту діагностування, у т. ч. і час збереження ЕРЕ до його установки в ППС.

Якщо діагностування ППС проводиться до виходу виробу на свій стаціонарний тепловий режим, кожен елемент буде мати відхилення свого параметра тільки за рахунок процесу старіння

$$\Delta q_{відб} = \Delta q_{техн} + \Delta q_{стар}.$$

При цьому

$$\Delta q_{стар} = (\beta_{зб} \cdot t_1 + \beta_{роб} \cdot t_2) \cdot q_{ном},$$

де $\beta_{зб}$ – коефіцієнт старіння елемента при обумовлених у ТУ на ЕРЕ умовах його збереження; $\beta_{роб}$ – коефіцієнт старіння елемента при фактичній температурі; t_1 – час збереження елемента на складі перед установкою його в апаратуру і час неробочого стану ППС; t_2 – час робочого стану ППС.

Якщо діагностування проводиться після виходу ППС на стаціонарний тепловий режим, то тоді складова відбракувального допуску $\Delta q_{темп}$ приймає своє максимальне значення. Аналогічно, якщо проводиться діагностування ППС на останньому році його експлуатації, то складова $\Delta q_{стар}$ приймає максимальне значення, визна-

чаючи тим самим, відповідний відбракувальний допуск на параметри ЕРЕ при відповідному значенні $\Delta q^{\text{тем}}$, у залежності від того, яка температура на момент діагностування ППС є присутня на елементах.

ВИСНОВКИ

Інженер-проектувальник у процесі проектування ППС одержує інтегральні допуски на ЕРЕ за різними методами синтезу допусків. Ці інтегральні допуски є постійними величинами у відмінності від відбракувальних допусків, які необхідні при діагностуванні ППС у процесі експлуатації. Відбракувальні допуски на параметри ЕРЕ розраховуються при безпосередній діагностиці ППС і приймають мінімальне значення, якщо діагностування проводиться відразу після етапу виготовлення ППС і до його прогріву (вони дорівнюють технологічним допускам на елемент), далі, по мірі експлуатації ППС, вони приймають проміжне значення, поступово наближаючись до максимального значення, що відповідає виходу ППС на свій стаціонарний режим та моменту перед утилізацією. Відбракувальний допуск адекватно варіюється, з огляду на два найбільш суттєві експлуатаційні фактори.

Запропонований метод, завдяки своїй універсальності, може бути застосований для перевірки чи оцінки

технічного стану ППС у будь-яких умовах і на різних етапах його експлуатації.

ПЕРЕЛІК ПОСИЛАНЬ

1. Кофанов Ю.Н. Теоретические основы конструирования, технологии и надежности радиоэлектронных средств: Учебник для вузов. М.: Радио и связь, 1991. – 360 с.: ил.
2. Касьян Н.Н. Методика диагностирования аналоговых устройств функционального преобразования на этапе эксплуатации. Системные проблемы качества, математического моделирования и информационных технологий // Материалы Международной конференции и Российской научной школы. Часть 6. – Москва: НИИ «Автоэлектроника». – 2000. – С. 58–59.

Надійшла 31.05.2004

Після доробки 29.11.2004

Предлагается метод синтеза текущих отбраковочных допусков на параметры элементов, необходимых для проведения диагностирования устройств преобразования сигналов в процессе эксплуатации. Математическая база метода обеспечивает компактность программной реализации при осуществлении автоматизации исследований с помощью ЭВМ.

The method of synthesis of the current rejection admissions on parameters of elements necessary for realization of diagnosing of devices of transformation of signals while in service is offered. The mathematical base of a method provides compactness of program realization at realization of automation of researches with the help of the computer.

УДК 621.396.6.004:004.942

В. М. Крищук, Г. М. Шило, А. О. Намлинський, М. П. Гапоненко

ВИБІР ЕЛЕМЕНТІВ ПРИ КОМПЕНСАЦІЇ ЗОВНІШНІХ ВПЛИВІВ

Запропоновано метод призначення номінальних допусків і коефіцієнтів зовнішніх впливів з урахуванням компенсації дії зовнішніх факторів. Враховано обмеження елементної бази та нелінійність вихідної функції. Забезпечено оптимальний вибір параметрів. Розглянуто стратегії максимальних допусків і максимальних коефіцієнтів зовнішніх впливів.

1 ПОСТАНОВКА ЗАДАЧІ

Вибір елементів є кінцевим етапом проектування електронних схем. На цьому етапі з урахуванням обмежень елементної бази призначаються усі параметри електрорадіоелементів, зокрема, номінальні допуски і коефіцієнти зовнішніх впливів. Результат вибору елементів безпосередньо відбивається на стабільності параметрів електронного приладу та його собівартості.

Існуючі методи вибору елементів не дозволяють враховувати усю множину параметрів елементів, які формують вихідні параметри пристрою. У більшості випадків призначення допусків відбувається без урахування

дії зовнішніх чинників і нелінійності вихідних функцій [1, 2].

Метою роботи є розробка методу вибору елементів і призначення допусків з урахуванням нелінійності вихідної функції, обмежень елементної бази і впливу зовнішніх факторів. Для вирішення цієї задачі необхідно:

- отримати співвідношення для відхилення вихідних параметрів з урахуванням нелінійних властивостей вихідних функцій електронних апаратів;
- врахувати обмеження елементної бази при виборі параметрів елементів;
- розробити алгоритми для різних стратегій вибору елементів.

При вирішенні поставленої задачі використовуються спрощені інтервальні моделі вихідних функцій [3], що дозволяє врахувати нелінійність цих функцій. Стратегії вибору елементів забезпечують оптимальний вибір параметрів завдяки оптимальному призначенню допусків [4] і коефіцієнтів зовнішніх впливів [5].

2 ОБМЕЖЕННЯ ЕЛЕМЕНТНОЇ БАЗИ І СТРАТЕГІЇ ВИБОРУ ЕЛЕМЕНТІВ

Компенсаційні властивості електронних апаратів характеризуються рівнями компенсації [5]:

$$\underline{\lambda} = \frac{w_n}{w_e} = \frac{y_n - y_n}{y_e - y_n}; \quad \bar{\lambda} = \frac{\bar{w}_n}{\bar{w}_e} = \frac{\bar{y}_n - y_n}{\bar{y}_e - y_n},$$

де $\underline{\lambda}$ і $\bar{\lambda}$ – рівні компенсації для нижньої і верхньої меж вихідної функції; w_n, \bar{w}_n, w_e та \bar{w}_e – нижні і верхні межі номінальних та експлуатаційних відхилень вихідної функції; y_n, \bar{y}_n, y_e та \bar{y}_e – номінальне значення вихідної функції та її значення при нижніх і верхніх номінальних та експлуатаційних відхиленнях.

Номінальні і межові експлуатаційні значення вихідної функції призначаються при розробці електричних схем. В залежності від бажаного рівня компенсації визначаються номінальні відхилення і відповідні їм значення вихідної функції. При заданих номінальних відхиленнях вихідної функції може бути утворена її інтервальна модель і визначені оптимальні номінальні відхилення вхідних параметрів.

Під дією зовнішніх факторів значення параметрів елементів і вихідної функції змінюються. Лінійній залежності параметрів пристроїв відповідають співвідношення:

$$y_e = y_n(1 + \alpha_y \cdot \vartheta); \quad x_e = x_n(1 + \alpha_x \cdot \vartheta), \quad (3)$$

де α_y і α_x – приведені до температури коефіцієнти зовнішніх впливів вихідної функції і параметрів елементів; ϑ – діапазон зміни температури навколишнього середовища; x_e, x_n – експлуатаційне і номінальне значення вхідного параметру.

Інтервальна модель вихідної функції при нижньому експлуатаційному відхиленні записується у вигляді:

$$y_e = a_{e0} + \sum_{i=1}^n a_{ei} x_{ei} + \sum_{i=1}^n \bar{a}_{ei} \bar{x}_{ei}, \quad (4)$$

де $a_{ei}, \bar{a}_{ei}, x_{ei}$ та \bar{x}_{ei} – нижні та верхні межі коефіцієнтів інтервальної моделі та параметрів елементів при експлуатаційному відхиленні вихідної функції.

Підстановка співвідношень (1) в рівняння (2) визначає умови компенсації зовнішніх впливів:

$$\begin{aligned} s_y + \frac{1}{\vartheta} \left(y_n - a_{e0} + \sum_{i=1}^n a_{ei} x_{ni} - \sum_{i=1}^n \bar{a}_{ei} \bar{x}_{ni} \right) = \\ = \sum_{i=1}^n a_{ei} x_{ni} \alpha_{xi} + \sum_{i=1}^n \bar{a}_{ei} \bar{x}_{ni} \alpha_{xi}, \end{aligned} \quad (5)$$

де $s_y = y_n \alpha_y$ – фактор компенсації вихідної функції; x_{ni} і \bar{x}_{ni} – нижня і верхня межі параметрів елементів при їх номінальних відхиленнях.

Умова (3) може бути записана у вигляді:

$$s_{ey} = \sum_{i=1}^p s_i^+ + \sum_{j=1}^m s_j^-, \quad (6)$$

де $s_i = a_{ei} x_{ni} \alpha_{xi}$ – фактор компенсації елементу; s_i^+ і s_j^- – додатні і від'ємні фактори компенсації елементів; p і m – кількість додатних і від'ємних факторів компенсації елементів; s_{ey} – еквівалентний фактор компенсації вихідної функції, який записується у вигляді:

$$s_{ey} = s_y + \frac{1}{\vartheta} \left(y_n - a_{e0} + \sum_{i=1}^n a_{ei} x_{ni} - \sum_{i=1}^n \bar{a}_{ei} \bar{x}_{ni} \right). \quad (7)$$

Еквівалентний фактор компенсації вихідної функції враховує її нелінійні властивості. Значення цього фактору залежить від меж допусків і діапазонів зміни зовнішніх впливів.

Вибір елементів виконується з урахуванням обмежень елементної бази на коефіцієнти зовнішніх впливів:

$$\alpha_{xi \min} \leq \alpha_{xi} \leq \alpha_{xi \max},$$

де $\alpha_{xi \min}$ і $\alpha_{xi \max}$ – мінімальні і максимальні коефіцієнти зовнішніх впливів.

Цим обмеженням відповідають обмеження для факторів компенсації елементів:

$$s_{i \min} \leq s_i \leq s_{i \max}. \quad (8)$$

Підстановка обмежень (6) в (4) показує, що існує шість випадків взаємного розташування межових сум, які надаються на рис. 1, де позначено:

$$S^+ = \sum_{i=1}^p s_i^+; \quad S^- = \left| \sum_{j=1}^m s_j^- \right|. \quad (9)$$

Індекси \min і \max відповідають сумам з мінімальними і максимальними коефіцієнтами зовнішніх впливів.

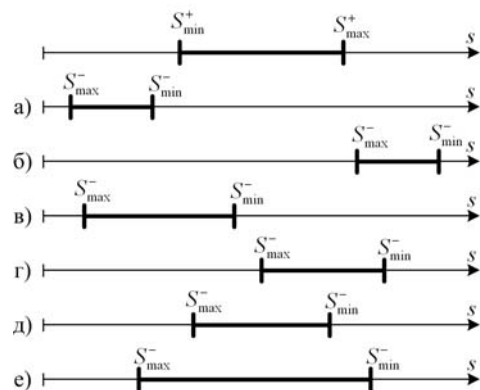


Рисунок 1 – Випадки розташування областей факторів компенсації

Із рис. 1, а і рис. 1, б видно, що повної компенсації дії зовнішніх впливів не відбувається, якщо виконується одна з умов:

$$S_{\min}^+ > S_{\min}^-; S_{\max}^+ < S_{\max}^- \quad (10)$$

В інших випадках (рис. 1, в – рис. 1, е) відбувається перекриття областей існування сум (7). Виникає можливість вибору елементів при $s_{ey} = 0$ з оптимізацією їх параметрів. При виборі стратегії оптимізації слід враховувати, що для нелінійних вихідних функцій умова $s_{ey} = 0$ не означає досягнення повної компенсації дії зовнішніх чинників на вихідну функцію [5]. У зв'язку з цим визначаються дві граничні стратегії вибору елементів, які дозволяють забезпечити максимальний об'єм простору коефіцієнтів зовнішніх впливів (МКЗВ) або номінальних допусків (МНД).

3 ВИБІР ЕЛЕМЕНТІВ ПРИ ОБМЕЖЕННЯХ КОМПЕНСАЦІЇ

Умови (8) означають, що при будь-яких сполученнях сум (4) не відбувається повної компенсації дії зовнішніх чинників на будь яку вихідну функцію. Максимальний рівень компенсації досягається коли ефективний фактор компенсації вихідної функції досягає значень:

$$s_{ey} = \begin{cases} S_{\min}^+ - S_{\min}^-, & (S_{\min}^+ > S_{\min}^-) \\ S_{\max}^+ - S_{\max}^-, & (S_{\max}^+ < S_{\max}^-) \end{cases} \quad (11)$$

При обчисленні (9) використовуються межові значення коефіцієнтів зовнішніх впливів. Необхідно також знати інтервальні експлуатаційні коефіцієнти вихідної функції і номінальні межі параметрів елементів. Оскільки на початковому етапі вибору елементів їх номінальні межі параметрів не визначено, то необхідно використовувати ітераційну процедуру.

Для початкового наближення доцільно використовувати співвідношення:

$$\underline{y}_{-n}^{(0)} = \underline{y}_{-e}; \quad s_y^{(0)} = 0. \quad (12)$$

Номінальні допуски на параметри елементів, обчислені для цього значення вихідної функції, використовуються, у відповідності із співвідношенням (1), для обчислення експлуатаційних значень вхідних параметрів. Ці значення використовуються для обчислення значення вихідної функції при різних температурах. Серед обчислених значень вихідної функції вибирається таке, що найбільше виходить за межі експлуатаційних відхилень. Вибране значення вихідної функції використовується для обчислення необхідної зміни її фактора компенсації. Це дозволяє визначити нові межі номінального значення вихідної функції. Таким чином ітераційний процес використовує співвідношення:

$$\underline{y}_{-t}^{(k)} = \min \{ \underline{y}(\vartheta_m), \underline{y}(\vartheta_p) \}; \quad (13)$$

$$\Delta s_y^{(k)} = (\underline{y}_{-t}^{(k)} - \underline{y}_{-e}) / \vartheta_t; \quad (14)$$

$$\underline{s}_y^{(k)} = \underline{s}_y^{(k-1)} + \Delta s_y^{(k)}; \quad (15)$$

$$\underline{y}_{-n}^{(k)} = \underline{y}_{-e} - \underline{s}_y^{(k)} \vartheta_t, \quad (16)$$

де ϑ_m та ϑ_p – перепади температур в нижньому та верхньому температурному діапазонах; ϑ_t – перепад температури, який відповідає значенню вихідної функції $\underline{y}_{-t}^{(k)}$.

Ітераційний процес завершується якщо:

$$\left| \underline{\lambda}_n^{(k)} - \underline{\lambda}_n^{(k-1)} \right| \leq \varepsilon, \quad (17)$$

де ε – похибка обчислення.

Верхні номінальні відхилення обчислюються з використанням межової функції:

$$\bar{y}_t^{(k)} = \max \{ \bar{y}(\vartheta_m), \bar{y}(\vartheta_p) \}.$$

Ітераційний процес обчислення допусків на верхній межі є аналогічним і використовує співвідношення подібні (11)–(15).

4 СТРАТЕГІЯ МАКСИМАЛЬНИХ КОЕФІЦІЄНТІВ ЗОВНІШНІХ ВПЛИВІВ

Існування області перекриття сум (7) призводить до нетривіального вирішення задачі вибору елементів. Випадки розподілу параметрів, що відповідають межам інтервалу перекриття, реалізуються стратегіями МКЗВ і МНД.

Стратегія МКЗВ дозволяє послабити вимоги до стабільності параметрів елементів. В цій стратегії вибір однієї з множин факторів компенсації відбувається за допомогою критерію:

$$S_c = \sum_{i=1}^p s_{i\max}^+ + \sum_{j=1}^m s_{j\min}^- - s_y,$$

де p і m – кількість елементів з додатними і від'ємними значеннями фактора компенсації відповідно.

Обчислення критерію S_c проводиться для кожної з меж вихідної функції і для кожного з діапазонів впливу зовнішніх чинників. Знак критерію дозволяє прийняти рішення щодо корегування додатної або від'ємної множин параметрів. Необхідний рівень компенсації зовнішніх впливів забезпечується, якщо $S_c = 0$. В цьому випадку приймається:

$$s_j^- = s_{j\min}^-; \quad s_i^+ = s_{i\max}^+, \quad (S_c = 0). \quad (18)$$

При додатних значеннях критерію S_c об'єм факторів компенсації елементів обмежується множиною від'ємних факторів. Це призводить до співвідношень:

$$s_j^- = s_{j\min}^-; \quad p \cdot s^+ = s_y - \sum_{j=1}^m s_{j\min}^-, \quad (S_c > 0). \quad (19)$$

Якщо оптимальні значення факторів компенсації для деяких елементів не реалізуються, то для них приймається:

$$s_i^+ = s_{i\max}^+, \quad (s_{i\max}^+ < s^+).$$

Додатні фактори компенсації, що залишилися, використовують оптимальні співвідношення:

$$(p-k) \cdot s_{p-k}^+ = s_y - \sum_{j=1}^m s_{j\min}^- - \sum_{\substack{i=1 \\ s_{i\max}^+ < s^+}}^p s_{i\max}^+, \quad (20)$$

де k – кількість елементів з обмеженнями $s_{i\max}^+ < s^+$; s_{p-k}^+ – оптимальні фактори компенсації для $p-k$ елементів, що залишилися.

Від'ємним значенням критерію S_c відповідають обмеження за множиною додатних факторів компенсації елементів. Цей випадок приводить до виразів:

$$s_i^+ = s_{i\max}^+; \quad m \cdot s^- = s_y - \sum_{i=1}^p s_{i\max}^+, \quad (S_c < 0). \quad (21)$$

Для факторів компенсації, що не реалізуються, приймається:

$$s_j^- = s_{j\min}^-; \quad (s_{j\min}^- > s^-).$$

Від'ємні фактори, що залишилися, використовують оптимальний розподіл:

$$(m-l) \cdot s_{m-l}^- = s_y - \sum_{j=1}^m s_{j\min}^- - \sum_{i=1}^p s_{i\max}^+, \quad (22)$$

де l – кількість елементів з обмеженнями $s_{j\min}^- > s^-$; s_{m-l}^- – оптимальні фактори компенсації для $m-l$ елементів, що залишилися.

Обчислення параметрів елементів відбувається в ітераційному процесі. Спочатку для кожної з меж використовується початкове наближення (10). Далі обчислюються межові номінальні значення основних параметрів елементів та коефіцієнти інтервальних моделей. Це дозволяє визначити коефіцієнти зовнішніх впливів із співвідношень (16)–(20) і застосувати ітераційну процедуру (11)–(15).

Номінальні межові відхилення елементів і коефіцієнти зовнішніх впливів визначаються для однієї з меж вихідного параметру. Номінальні межові відхилення елементів для протилежної межі враховують прийняті значення коефіцієнтів зовнішніх впливів. Така сама процедура використовується для іншої межі вихідної функції. Як остаточне вибирається рішення, що забезпечує максимальну ширину інтервалу номінальних відхилень вихідного параметру.

5 СТРАТЕГІЯ МАКСИМАЛЬНИХ НОМІНАЛЬНИХ ДОПУСКІВ

В стратегії максимальних номінальних допусків вибір факторів компенсації здійснюється за допомогою критерію:

$$S_t = \sum_{j=1}^m s_{j\max}^- + \sum_{i=1}^p s_{i\min}^+ - s_y.$$

Якщо $S_t = 0$, то граничні фактори $s_{i\min}^+$ і $s_{j\max}^-$ забезпечують необхідний рівень компенсації. При від'ємних значеннях критерію S_t об'єм факторів компенсації елементів обмежується множиною від'ємних факторів. Це призводить до співвідношень:

$$s_j^- = s_{j\max}^-; \quad p \cdot s^+ = s_y - \sum_{j=1}^m s_{j\max}^-, \quad (S_t < 0).$$

Якщо оптимальні значення факторів компенсації для деяких елементів не реалізуються, то для них приймається:

$$s_i^+ = s_{i\max}^+, \quad (s_{i\max}^+ < s^+).$$

Додатні фактори компенсації, що залишилися, використовують оптимальні співвідношення:

$$(p-k) \cdot s_{p-k}^+ = s_y - \sum_{j=1}^m s_{j\max}^- - \sum_{\substack{i=1 \\ s_{i\max}^+ < s^+}}^p s_{i\min}^+,$$

де k – кількість елементів з обмеженнями $s_{i\min}^+ > s^+$; s_{p-k}^+ – оптимальні фактори компенсації для елементів, що залишилися.

Додатним значенням критерію S_t відповідають обмеження за множиною додатних факторів компенсації елементів. Цей випадок приводить до виразів:

$$s_i^+ = s_{i\min}^+; \quad m \cdot s^- = s_y - \sum_{i=1}^p s_{i\min}^+, \quad (S_t > 0).$$

Для факторів компенсації, що не реалізуються, приймається:

$$s_j^- = s_{j\max}^-; \quad (s_{j\max}^- < s^-).$$

Від'ємні фактори, що залишилися, використовують оптимальний розподіл:

$$(m-l) \cdot s_{m-l}^- = s_y - \sum_{j=1}^m s_{j\max}^- - \sum_{i=1}^p s_{i\min}^+,$$

де l – кількість елементів з обмеженнями $s_{j\max}^- < s^-$; s_{m-l}^- – оптимальні фактори компенсації для $m-l$ елементів, що залишилися.

Організація обчислень в стратегії МНД подібна стратегії МКЗВ.

6 ПРИКЛАД ЗАСТОСУВАННЯ

За запропонованою методикою проведено вибір елементів для фільтра нижніх частот, схема якого наведена на рис. 2.

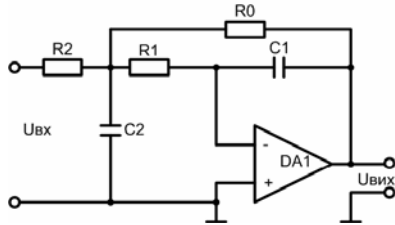


Рисунок 2 – Схема фільтра нижніх частот

Коефіцієнт загасання фільтра обчислювався за співвідношенням:

$$a(w) = \frac{R_2}{R_0} \sqrt{\Theta_1 + \Theta_2},$$

$$\Theta_1 = \left(1 - w^2 \cdot C_1 \cdot C_2 \cdot R_1 \cdot R_0\right)^2;$$

$$\Theta_2 = \left(w \cdot C_1 \cdot R_1 \cdot \left(1 + \frac{R_0}{R_1} + \frac{R_0}{R_2}\right)\right)^2;$$

де $\omega = 2\pi f$ – кутова частота; R_i, C_i – параметри елементів.

Номинальні параметри елементів і результати розрахунків надаються в табл. 1, де $\underline{\delta}_n$ і $\bar{\delta}_n$ – нижнє і верхнє номінальні відхилення параметрів елементів; $\alpha = [\alpha_{\min}; \alpha_{\max}]$ – інтервал коефіцієнтів зовнішніх впливів. Температура змінювалась у діапазоні від -40 до 60°C . Обчислення проводилися для частоти 12 кГц, на якій коефіцієнт загасання $a = 6,1697$. Експлуатаційний допуск на вихідну функцію дорівнював $\pm 10\%$. Для зазначених у табл. 1 обмежень на коефіцієнти зовнішніх впливів у верхньому температурному діапазоні існують обмеження на компенсацію, які відповідають рис. 1, б. Задавалась точність обчислення $\varepsilon = 0,001$, при якій процес завершувався за 5 ітерацій. Визначені номінальні відхилення вихідного параметру $\delta_{\text{ун}} = [5,60\%; 3,13\%]$.

Таблиця 1 – Вихідні дані і параметри елементів у випадку неповної компенсації

Параметр	x_{ni}	$\alpha \cdot 10^6, 1/^\circ\text{C}$	$\underline{\delta}_n, \%$	$\bar{\delta}_n, \%$	$\alpha \cdot 10^6, 1/^\circ\text{C}$
R_0	100 кОм	[55;500]	20,00	17,12	500
R_1	91 кОм	[55;500]	1,18	0,61	500
R_2	100 кОм	[55;500]	1,39	0,68	500
C_1	1,0 нФ	[-1750;-1000]	1,06	0,55	-1000
C_2	4,7 нФ	[-1250;-1200]	1,54	0,73	-1200

Якщо змінити межові значення коефіцієнтів зовнішніх впливів, як наведено у табл. 2 і табл. 3, то виникає можливість реалізувати стратегії МКЗВ і МНД. В стратегії МКЗВ номінальний допуск вихідного параметра при стабілізації нижньої та верхньої меж мав значення $\delta_{\text{ун}} = [9,405\%; 10,000\%]$ і $\delta_{\text{ун}} = [9,001\%; 10,000\%]$. Рішення, що стабілізує нижню межу, дає більше значення рівня компенсації і обирається як оптимальне. Параметри елементів при цьому мають значення наведені в табл. 2.

Таблиця 2 – Вихідні дані та параметри елементів при застосуванні стратегії МКЗВ

Параметр	$\alpha \cdot 10^6, 1/^\circ\text{C}$	$\underline{\delta}_n, \%$	$\bar{\delta}_n, \%$	$\alpha \cdot 10^6, 1/^\circ\text{C}$
R_0	[55;1200]	20,00	20,00	1200
R_1	[55;1200]	2,06	2,19	1200
R_2	[55;1200]	2,42	2,44	1200
C_1	[-1750;-120]	1,84	2,01	-1000
C_2	[-1750;-120]	2,70	2,60	-1490

Два варіанти рішення в стратегії МНД характеризуються значеннями параметрів $\delta_{\text{ун}} = [9,991\%; 9,971\%]$ і $\delta_{\text{ун}} = [9,991\%; 9,996\%]$ при стабілізації нижньої і верхньої меж, відповідно. Стабілізація верхньої межі забезпечує більше значення рівня компенсації вихідного параметру. Цьому випадку відповідають параметри елементів наведені у табл. 3.

Таблиця 3 – Вихідні дані та параметри елементів при застосуванні стратегії МНД

Параметр	$\alpha \cdot 10^6, 1/^\circ\text{C}$	$\underline{\delta}_n, \%$	$\bar{\delta}_n, \%$	$\alpha \cdot 10^6, 1/^\circ\text{C}$
R_0	[55;1200]	20,00	20,00	1200
R_1	[55;1200]	2,29	2,19	92
R_2	[55;1200]	2,70	2,44	102
C_1	[-1750;-120]	2,05	2,00	-120
C_2	[-1750;-120]	3,01	2,60	-120

Порівняння результатів вибору елементів для компенсаційних режимів фільтра нижніх частот за різними стратегіями показує, що використання стратегії МНД не дає суттєвого збільшення допусків на параметри елементів. В той же час вимоги до коефіцієнтів зовнішніх впливів значно зростають. Тому при зазначених обмеженнях на параметри елементів доцільно використовувати стратегію МКЗВ.

ВИСНОВКИ

Аналіз компенсаційних режимів електронних апаратів з використанням інтервальних моделей вихідних функцій показав, що елементна база може привести до обмеження

рівня компенсації вихідного параметра. При відсутності обмежень виникає можливість вибору елементів із максимальними значеннями коефіцієнтів зовнішніх впливів або максимальними номінальними допусками. Розроблені процедури вибору елементів враховують нелінійні властивості вихідних функцій електронних пристроїв і відбуваються в ітераційному режимі. Вибрані параметри оптимізуються за відносними об'ємами.

Розроблені процедури вибору елементів гарантують знаходження вихідного параметру у заданих межах при заданій зміні зовнішніх факторів, якщо при виготовленні апаратури значення основного параметру елементів не виходить за межі призначених номінальних допусків. Методика забезпечує вибір елементів при номінальних значеннях коефіцієнтів зовнішніх впливів.

ПЕРЕЛІК ПОСИЛАНЬ

1. Михайлов А.В., Савин К.С. Точность радиоэлектронных устройств. – М.: Машиностроение, 1976. – 214 с.
2. Фридендер И.Г. Расчеты точности машин при проектировании. – Киев–Донецк: Вища школа, 1980. – 184 с.

УДК 621.382.002

С. И. Прохорец, М. А. Хажмурадов, В. П. Лукьянова

АНАЛИЗ ЕМКОСТНЫХ ПАРАМЕТРОВ СТРИП-ДЕТЕКТОРА

Рассмотрена математическая модель кремниевого стрип-детектора, представленного в виде набора из обратно смещенных p-n переходов, емкостей и резисторов. Для полностью обедненного детектора каждый стрип представлен как линия определенной длины, имеющая свое сопротивление и емкостные связи с соседним стрипом и противоположной плоскостью. На основе теории графов написано математическое уравнение стрип-детектора для моделирования его емкостных параметров.

ВВЕДЕНИЕ

Стрип-детекторы за последние годы находят широкое применение в экспериментальной физике, астрофизике, медицине и различных отраслях техники. Их планируется использовать в уникальных физических установках [1, 2], создаваемых на большом адронном коллайдере (ЛНС) в Европейском центре ядерных исследований (ЦЕРН, Швейцария). Эти установки должны работать длительный срок (не менее десяти лет) без замены детектирующих элементов, которые проходят длительное тестирование всех параметров как предприятиями-изготовителями, так и непосредственно в физических центрах потребителями. Кроме кремния, в качестве материала для изготовления стрип-детекторов начали применять такие полупроводниковые материалы как GaAs и CdZnTe [3, 4], что обещает в будущем расширить сферу применения позиционно-чувствитель-

3. Шило Г.М. Формування інтервальних моделей для обчислення допусків // Радіоелектроніка. Інформатика. Управління. – 2002. – №1. – С.90–95.
4. Шило Г.М., Воропай О.Ю., Гапоненко М.П. Інтервальні методи призначення експлуатаційних допусків // Радіоелектроніка. Інформатика. Управління. 2003. №2. С. 78–82.
5. Шило Г.Н., Намлинский А.А., Гапоненко Н.П. Компенсация и оптимизация при назначении номинальных допусков // Радиоелектроника и информатика. 2004. №2.

Надійшла 13.05.2004
Після доробки 25.11.2004

Предложен метод назначения номинальных допусков и коэффициентов внешних воздействий с учетом компенсации действий внешних факторов. Учитываются ограничения элементной базы и нелинейность выходной функции. Обеспечивается оптимальный выбор параметров. Рассмотрены стратегии максимальных допусков и коэффициентов внешних воздействий.

The method of assigning nominal tolerances and external action coefficient selection is suggested. The compensation of external action influences is held in this method. The limitations to the element list and non-linearity of output characteristics are taken into account. Optimal parameter selection is ensuring. The maximal tolerance and maximal external action coefficient strategies are considered.

ных детекторов. На всех этапах от проектирования и до исследования параметров таких детекторов в условиях эксперимента возникает потребность в математических программах и алгоритмах, основанных на модельных представлениях о проектируемом или исследуемом объекте [5, 6]. Целью настоящей работы является создание математической модели стрип-детектора, позволяющей на ее основе создавать специализированные программные средства для системы автоматизированного проектирования (САПР) и исследования детекторов.

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПОСТРОЕНИЯ МОДЕЛИ СТРИП-ДЕТЕКТОРА

В научной литературе стрип-детекторы принято разделять на односторонние и двухсторонние, в которых измеряется одна или две координаты прошедшей через них заряженной частицы соответственно [5, 6]. Конструктивно они состоят из отдельных обратно смещенных p-n переходов, емкостей и резисторов. Поэтому для построения математической модели устройства из таких базовых элементов можно использовать направленные графы $G(V, S)$ [7–10]. В направленном графе каждая ветвь V_{ij} соответствует двухполюсному базовому элементу модели, а ее направление совпадает с направлением действия на этом элементе фазовой переменной –

тока или напряжения. Каждая вершина графа S_{ij} является точкой соединения двух или более базовых элементов и соответствует узлу модели. Ветвь называют инцидентной вершине, если она начинается или заканчивается в этой вершине. Используя инцидентность, информацию о графе в вычислительной машине можно представлять в виде матрицы инцидентий $A_{ij}=[a_{ij}]_{(m+1)n}$ с $S=(m+1)$ вершинами и $V=n$ ветвями, в которой элемент $a_{ij}=1$, если ветвь инцидентна вершине и направлена от нее, $a_{ij}=-1$, если ветвь инцидентна вершине и направлена к ней, $a_{ij}=0$, если ветвь не инцидентна вершине.

Матрица инцидентий позволяет записать законы Кирхгофа для токов и напряжений в электрической цепи. Если выделить дерево графа, то законы Кирхгофа можно записать с помощью матрицы главных сечений D и матрицы главных контуров B [9], между которыми и матрицей инцидентий существуют фундаментальные соотношения $BD^t=0$, $DB^t=0$ и $AB^t=0$. Верхний индекс t означает транспонированную матрицу. Если электрическая цепь состоит из n ветвей и $m+1$ узлов, то ее граф содержит m ребер и $n-m$ хорд. Для связного графа с $m+1$ узлами и n ветвями максимальное количество линейно-независимых строк M матрицы B равно $M=n-m$ [10].

Любая матрица главных контуров может быть разделена на две подматрицы $B=[B_1]$ [10], где B_1 соответствует ветвям дерева, а единичная матрица порядка $n-m$ соответствует хордам. Если в цепи есть независимые источники напряжения, то их также включают в матрицу главных контуров B , которую в этом случае называют дополненной и обозначают B_D . Матрицу B_D также можно разделить на следующие подматрицы $B_D=[B_E B]$, где B_E соответствует независимым источникам напряжения.

Упомянутые выше матрицы представляют собой топологические уравнения электрической модели изучаемого устройства. Следовательно, написав одну из этих матриц для стрип-детектора или для отдельной схемы измерения его параметров, получим его или ее уравнение как электрического устройства.

МОДЕЛИРОВАНИЕ ЕМКостей В СТРИП-ДЕТЕКТОРЕ

Каждый стрип в детекторе можно представить как линию определенной длины, имеющую свое сопротивление и емкостные связи с противоположной плоскостью C_g и соседним стрипом через межстриповую емкость C_{is} (рис. 1). Емкостью центрального стрипа относительно второго соседнего можно пренебречь [11], поэтому она не показана на рисунке. К основным параметрам стрип-детектора также относятся сопротивление обратно смещенного $p-n$ перехода индивидуального стрипа R_g , которое определяет ток утечки детектора, межстриповое сопротивление R_{is} и сопротивление стрипа R_s , которое характеризует параметры напыленной проводящей полоски из алюминия.

В качестве исходной предпосылки для моделирования емкостных параметров одностороннего стрип-детектора предположим, что его схему можно представить в виде

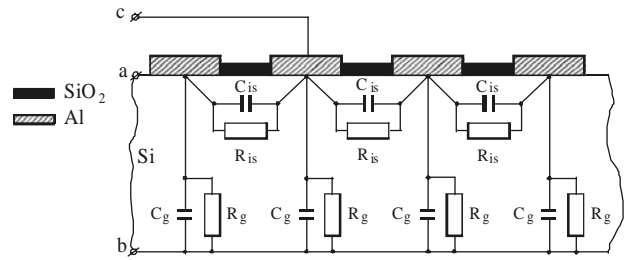


Рисунок 1 – Упрощенная схема одностороннего стрип-детектора:

a, b – точки для измерения емкости стрипа относительно противоположной плоскости; a, c – точки для измерения межстриповой емкости

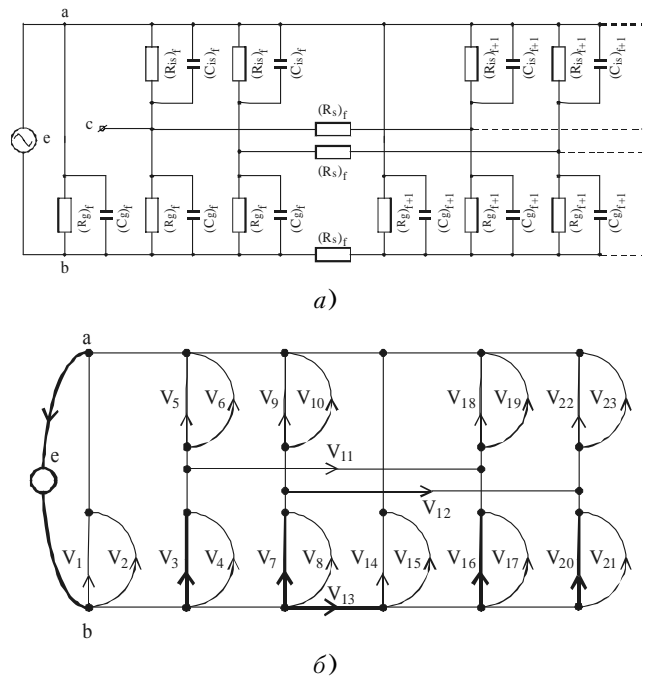


Рисунок 2 – Эквивалентная схема двух ячеек стрип-детектора (а) и ее граф (б) с выделенным деревом:

e – источник переменного напряжения

элементарных N ячеек, в каждую из которых входят межстриповая емкость $(C_{is})_f=C_{is}/N$ и сопротивление $(R_{is})_f=N \cdot R_{is}$, емкость стрипа относительно противоположной плоскости $(C_g)_f=C_g/N$ и сопротивление обратно смещенного $p-n$ перехода $(R_g)_f=N \cdot R_g$. Каждая из ячеек f связана с ячейкой $f+1$ через сопротивления $(R_s)_f=(N-1) \cdot R_s$, $(R_{is})_f$, $(R_g)_f$, $(R_s)_f$ принимаем как чисто активные, а сопротивления емкостей – зависящими по обычному закону от частоты. Две ячейки схемы для измерения емкости стрипа приведены на рис. 2, а, а ее граф – на рис. 2, б. Он содержит 7 узлов ($m+1=7$) и 24 ветви. Дерево графа – это ветви $e, V_3, V_7, V_{13}, V_{17}, V_{20}$.

Подматрица главных контуров, соответствующая ребрам графа, при таком выборе дерева имеет вид:

Матрица	Контур	Ветви										
		e	V ₃	V ₇	V ₁₃	V ₁₆	V ₂₀	V ₂₆	V ₂₉	V ₃₃	V ₃₉	
[B _e B _v]	1	1	0	0	0	0	0	0	0	0	0	0
	2	1	0	0	0	0	0	0	0	0	0	0
	3	0	-1	0	0	0	0	0	0	0	0	0
	4	1	1	0	0	0	0	0	0	0	0	0
	5	1	1	0	0	0	0	0	0	0	0	0
	6	0	0	-1	0	0	0	0	0	0	0	0
	7	1	0	1	0	0	0	0	0	0	0	0
	8	1	0	1	0	0	0	0	0	0	0	0
	9	0	1	0	-1	-1	0	0	0	0	0	0
	10	0	0	1	-1	0	-1	0	0	0	0	0
	11	1	0	0	1	0	0	0	0	0	0	0
	12	1	0	0	1	0	0	0	0	0	0	0
	13	0	0	0	0	-1	0	0	0	0	0	0
	14	1	0	0	1	1	0	0	0	0	0	0
	15	1	0	0	1	1	0	0	0	0	0	0
	16	0	0	0	0	0	-1	0	0	0	0	0
	17	1	0	0	1	0	1	0	0	0	0	0
	18	1	0	0	1	0	1	0	0	0	0	0

Отметим, что разбиение эквивалентной схемы детектора на две ячейки носит чисто иллюстративный характер. При исследовании частотной зависимости емкостей детектора количество ячеек значительно больше двух и, по имеющимся в литературе сведениям [5], может достигать 100 на длине 1 см, что приводит к существенному увеличению количества узлов графа и линейно независимых строк матрицы B .

Для отыскания дерева графа в сложной схеме и составления матрицы главных контуров обычно необходима специальная программа на ЭВМ. При написании матрицы главных контуров для стрип-детектора удается обойтись без такой программы, воспользовавшись периодичностью в эквивалентной схеме детектора. Эта особенность приводит к тому, что дополненная матрица главных контуров B_D , соответствующая исследуемой схеме с N ячейками, имеет размерность $(10N - 2) \times (13N - 2)$ и описывается следующими соотношениями:

$$B_D(10k - 9, 3k - 2) = B_D(10k - 8, 3k - 2) = 1;$$

$$B_D(10k - 7, 3k - 1) = -1;$$

$$B_D(10k - 6, 3k - 2) = B_D(10k - 6, 3k - 2) = 1;$$

$$B_D(10k - 5, 3k - 2) = B_D(10k - 5, 3k - 2) = 1;$$

$$B_D(10k - 4, 3k) = -1;$$

$$B_D(10k - 3, 3k - 2) = B_D(10k - 3, 3k) = 1;$$

$$B_D(10k - 2, 3k - 2) = B_D(10k - 2, 3k) = 1;$$

$$B_D(10k - 9, 3j + 1) = B_D(10k - 8, 3j + 1) = 1;$$

$$B_D(10k - 6, 3j + 1) = B_D(10k - 5, 3j + 1) = 1;$$

$$B_D(10k - 3, 3j + 1) = B_D(10k - 2, 3j + 1) = 1, \quad k = 1 \dots N, \quad j = 0 \dots k - 2;$$

$$B_D(10k - 11, 3k - 4) = 1;$$

$$B_D(10k - 11, 3k - 2) = B_D(10k - 11, 3k - 1) = -1;$$

$$B_D(10k - 10, 3k - 3) = 1;$$

$$B_D(10k - 10, 3k - 2) = B_D(10k - 10, 3k) = -1, \quad k = 2 \dots N;$$

$$B_D(k, 3N + k) = 1, \quad k = 1 \dots 10N - 2.$$

Это свойство графа позволяет записать в ЭВМ информацию о топологии схемы стрип-детектора при исследовании как емкости стрипа относительно противоположной плоскости, так и межстриповой емкости.

ЗАКЛЮЧЕНИЕ

В выполненной работе показано, что эквивалентную схему полоскового полупроводникового детектора можно представить в виде набора элементов с распределенными параметрами – сопротивлений и емкостей. Получена система топологических уравнений, позволяющая определить частотные зависимости емкости стрипа относительно противоположной плоскости и межстриповой емкости. Для написания матрицы главных контуров, описывающей топологию исследуемой схемы, получено аналитическое выражение, позволяющее обойтись без специальной программы для отыскания дерева графа исследуемого объекта.

ПЕРЕЧЕНЬ ССЫЛОК

1. ALICE Collaboration. Technical Proposal for A Large Ion Collider Experiment at the CERN LHC. CERN/LHCC/95-71 LHCC/P3, 1995, 237 p.
2. Krammer M. The Silicon Sensors for the Inner Tracker of the Compact Muon Solenoid Experiment. 5th International Workshop on Radiation Imaging Detectors. – Abstract Book, Riga, Latvia, 2003, 99 p.
3. Mocri J.R., Apotovsky B.F., Butler J.F. et al. Development of an Orthogonal-Stripe CdZnTe Gamma Radiation Imaging Spectrometer // IEEE Transactions on Nuclear Science, vol. 43, N 3, 1996, p. 1458-1461.
4. Chen J., Geppert R., Irsigler R. et al. Beam tests of GaAs strip detectors // Nucl. Instr. and Meth. A 369 (1996) p. 62-68.
5. Angarano M.M., Bader A., Greanza D. et al. Characterization and simulation of a singlesided, n+ on a n silicon microstrip detector before and after neutron irradiation // Nucl. Instr. and Meth. A 428 (1999), p. 336-347.
6. Fautchi M.A., Hoeferkamp H.R., Seidel S.C. Capacitance Measurements of Double-Sided Silicon Microstrip Detectors. CDF/DOC VTX/ CDF/2546. – The New Mexico Center for Particle Physics, Albuquerque, 1995, 62 p.
7. Атабеков Г.И. Основы теории цепей. – М.: Энергия, 1969, 424 с.
8. Влах И., Сингал К. Машинные методы анализа и проектирования электронных схем. – М.: Радио и связь, 1988, 560 с.
9. Чуа Л.О., Лин Пен-Мин. Машинный анализ электронных схем. – М.: Энергия, 1980, 640 с.
10. Алексеев О.В., Головкин А.А., Пивоваров И.Ю., Чавка Г.Г. Автоматизация проектирования радиоэлектронных средств – М.: Высш. шк., 2000, 480 с.
11. Прохорец С.И. Расчет параметров полосковых элементов модулей стрип-детекторов // АСУ и приборы автоматики, 2002, вып. 120, с. 131-137.

Надійшла 09.08.2004

Розглянуто математичну модель кремнієвого стрип-детектору, поданого у вигляді набору зворотно зміщених p-n переходів, ємностей та резисторів. Для цілком збудненого детектора кожен стрип представлено як лінію визначеної довжини, що має свій опір і ємнісні зв'язки із сусіднім стрипом і протилежною площиною. На основі теорії графів написано математичне рівняння стрип-детектора для моделювання його ємнісних параметрів.

A mathematical model of the silicon strip-detector which is considered as set of individual p-n junctions, capacitances and resistors is submitted. For fully depleted junction every strip is represented as line with definite length with its impedance and capacitive connection with neighboring strips and opposite plane. On the basis of graph theory mathematical equations for simulation of capacitance measurements in strip detector have been developed.

О. Ю. Фарафонов, С. О. Самоквіт, О. Ю. Воропай

ПРОЕКТУВАННЯ МІКРОСМУЖКОВИХ НАПРАВЛЕНИХ ВІДГАЛУЖУВАЧІВ З ОПТИМАЛЬНИМИ ДОПУСКАМИ

Проведено дослідження впливу допусків елементів конструкції на параметри направлених відгалужувачів: на зв'язаних лініях, шлейфного та Ланге. Використовуються інтервальні моделі. Визначені найбільш критичні за точністю конструктивні елементи. Це дозволило надати рекомендації по вибору мікросмужкових направлених відгалужувачів.

ВСТУП

Широке поширення у надвисокочастотному діапазоні набули направлені відгалужувачі (НВ), які використовуються для побудови смужкових подільників потужності, змішувачів, модуляторів, дискримінаторів, суматорів потужності та інших пристроїв. Характеристики таких пристроїв у значній мірі обумовлюються параметрами НВ. Основними характеристиками НВ є: перехідне послаблення, напрямленість, узгодження плеч відгалужувача з підведеними лініями, фазові співвідношення для хвиль напруг в вихідних плечах, робоче загасання в первісній лінії. Крім цих параметрів, в технічному завданні на НВ зазвичай вказуються: нерівномірність характеристики перехідного послаблення і широкосмуговість.

В залежності від чисельного значення перехідного послаблення відгалужувачі поділяються на пристрої з сильним зв'язком ($C \leq 10$ дБ) та слабким ($C > 10$ дБ) [1–3]. Крім того, НВ також класифікуються в залежності від ширини робочої смуги частот [3, 4]. При проектуванні НВ необхідно проводити вибір конструкції НВ відповідно до вимог по конструктивному виконанню, тобто враховуючи габаритні розміри, складність конструкції та точність виготовлення окремих елементів. Це призводить до необхідності визначення найбільш критичних допусків конструктивних елементів. Однак існуючі рекомендації вибору НВ не враховують вимоги до забезпечення точності.

Метою роботи є надання рекомендацій по вибору найбільш оптимальної конструкції мікросмужкових НВ при заданих вихідних характеристиках за допомогою порівняння допусків на конструкторсько-технологічні параметри. При обчисленні необхідно використовувати сучасні досягнення допускового аналізу. Зокрема обчислення допусків можуть проводитись за допомогою програмного комплексу INTOL [5], який в алгоритмі синтезу допусків використовує інтервальні моделі. Це дозволяє визначити вплив номінальних відхилень параметрів елементів конструкції на його основну харак-

теристику – коефіцієнт зв'язку з урахуванням нелінійних властивостей вихідних функцій НВ. При синтезі забезпечується максимальний об'єм простору допусків.

ОБЧИСЛЕННЯ НАПРАВЛЕНИХ ВІДГАЛУЖУВАЧІВ

Найбільше поширення на практиці отримали НВ на зв'язаних лініях, шлейфний та Ланге. При проектуванні мікросмужкових НВ зі слабким зв'язком часто використовуються НВ на зв'язаних лініях [1, 3]. Найбільш зручними для технологічної реалізації є мікросмужкові НВ на зв'язаних лініях з боковим зв'язком. Складність реалізації НВ з сильним зв'язком обумовлюється труднощами утворення малих зазорів між мікросмужковими провідниками на діелектричній підкладці та відкритим характером лінії. Фазові швидкості парної та непарної мод відрізняються, що призводить до звуження робочої смуги відгалужувача [2]. Конструкція НВ на зв'язаних лініях являє собою дві паралельно розташовані металічні смужки (рис. 1). Максимум зв'язку таких відгалужувачів досягається при довжині області зв'язку, яка дорівнює чверті довжини хвилі в лінії [1–3].

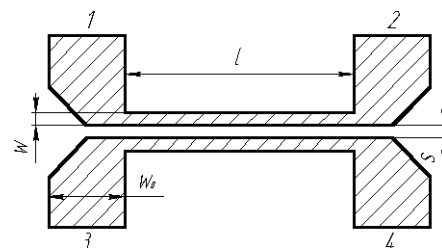


Рисунок 1 – Топологія направленої відгалужувача на зв'язаних лініях

Шлейфні НВ використовуються, як правило, для отримання сильного зв'язку, оскільки забезпечення слабого зв'язку пов'язано з конструктивними та технологічними труднощами утворення досить малого поперечного перерізу шлейфів (рис. 2). Довжина шлейфів та відстань між ними дорівнюють чверті довжини хвилі в лінії. Направленість та діапазонні характеристики покращуються із зростанням кількості шлейфів, але при великій кількості шлейфів ($n > 3$) зростають хвильові опори зовнішніх шлейфів [1].

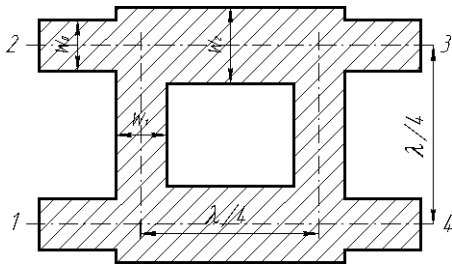


Рисунок 2 – Топологія мікросмужкових провідників шлейфного направлено відгалужувача

Недоліки притаманні НВ на зв'язаних лініях частково усунені в відгалужувачах Ланге (рис. 3). Фазові швидкості парної та непарної мод в зв'язаних лініях вирівнюються завдяки симетрії схеми, що і дозволяє досягти приблизно октавної робочої смуги в такому відгалужувачі [2]. В направлено відгалужувачі Ланге використовується структура на зустрічних стрижнях з перемичками. Необхідно прагнути до мінімальної довжини перемичок, але слідкувати, щоб введення перемичок не призвело до утворення гальванічного зв'язку між сусідніми провідниками. Ширина всіх провідників в структурі однакова, як і зазори між ними.

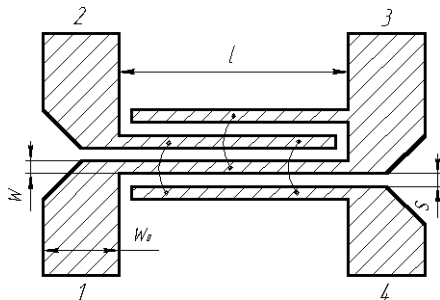


Рисунок 3 – Топологія смужкових провідників симетричного відгалужувача Ланге

Проектування НВ на зв'язаних лініях та Ланге враховує різницю фазових швидкостей парної та непарної мод. Для розрахунку параметрів зв'язаних ліній використовуються методика [6], точність якої досліджувалась в [7]. При синтезі НВ на зв'язаних лініях спочатку визначається коефіцієнт зв'язку

$$C_0 = 10^{-\frac{C}{20}}, \quad (1)$$

де C – коефіцієнт зв'язку по напрузі, дБ.

Далі обчислюються опори Z_{BE} , Z_{BO} для парної та непарної мод відповідно

$$Z_{BE} = Z_B \cdot \sqrt{\frac{1+C_0}{1-C_0}}, \quad (2)$$

$$Z_{BO} = Z_B \cdot \sqrt{\frac{1-C_0}{1+C_0}}, \quad (3)$$

де Z_B – хвильовий опір, Ом.

Завершується процедура синтезу визначенням геометричних розмірів та довжини зв'язку.

Аналогічно розраховуються параметри конструктивних елементів відгалужувача Ланге. Відмінність полягає лише у визначенні опорів парного та непарного типів коливань

$$Z_{BE,O} = \frac{2Z_B(n-1)\sqrt{R_1R_2(1-C^2)}}{2Z_B(1\mp C) + (n-2)\sqrt{R_1R_2(1-C^2)}}, \quad (4)$$

де n – кількість ліній в багатопрвідній структурі, R_1 , R_2 – опори навантаження.

При розрахунках геометричних розмірів шлейфного НВ перш за все визначаються провідності кожного шлейфу, використовуючи еквівалентну схему шлейфного НВ [2]. Далі отримані провідності перераховуються в хвильові опори для кожного шлейфу. Геометричні розміри відрізків ліній, що відповідають отриманим величинам хвильових опорів, визначаються за допомогою співвідношень для розрахунку одиночної мікросмужкової лінії передачі [6]. Похибка таких розрахунків становить не більше 2% [6].

ДОСЛІДЖЕННЯ ТОЧНОСТІ МІКРОСМУЖКОВИХ НАПРАВЛЕНИХ ВІДГАЛУЖУВАЧІВ НА ЗВ'ЯЗАНИХ ЛІНІЯХ ТА ЛАНГЕ

Для мікросмужкових ліній точний теоретичний аналіз параметрів провести дуже важко, оскільки в лінії цього типу окрім основної TEM -хвилі присутні паразитні поверхневі хвилі [8]. Це призводить до необхідності спрощень математичної моделі пристрою і, як наслідок, до зменшення її точності. В результаті точність визначення параметрів елементів конструкції обумовлюється точністю математичної моделі. Серед існуючих методів обчислення пристроїв на зв'язаних лініях найбільшу точність забезпечує наступна методика [6]. В цій методиці залежність коефіцієнту зв'язку на зв'язаних лініях від параметрів елементів визначається співвідношенням

$$C_0 = \frac{Z_{BE} - Z_{BO}}{Z_{BE} + Z_{BO}}. \quad (5)$$

При обчисленні коефіцієнту зв'язку для НВ Ланге враховується кількість провідників

$$C_0 = \frac{(Z_{BE} - Z_{BO})(n-1)}{(Z_{BE} + Z_{BO})(n-1) - \frac{Z_{BE}Z_{BO}(n-1)}{z_0}}, \quad (6)$$

де

$$z_0 = \frac{(Z_{BE} + Z_{BO})}{2},$$

де Z_{BE} та Z_{BO} – опори парного та непарного типів коливань відповідно

$$Z_{BE} = 2Z_0 \text{ при } (w_{E\Phi}/h)_{BE}; \quad (7)$$

$$Z_{BO} = 2Z_0 \text{ при } (w_{E\Phi}/h)_{BO}. \quad (8)$$

Хвильовий опір Z_0 при $(w_{E\Phi}/h)_{BE}$ та $(w_{E\Phi}/h)_{BO}$ розраховується із співвідношення

$$Z_0 = \frac{42,4}{\sqrt{\varepsilon+1}} \ln \left\{ 1 + \frac{4h}{w} \left[\left(\frac{14 + \frac{8}{\varepsilon}}{11} \right) \frac{4h}{w} + \sqrt{\left(\frac{14 + \frac{8}{\varepsilon}}{11} \right)^2 \left(\frac{4h}{w} \right)^2 + \frac{1 + \frac{1}{\varepsilon}}{2} \pi^2} \right] \right\}, \quad (9)$$

де ε – діелектрична проникливість підкладки.

Ефективна ширина смужки, яка відповідає парному та непарному типам коливань визначається виразами:

$$\left(\frac{w_{E\Phi}}{h} \right)_{BE} = \frac{2}{h} \operatorname{arch} \left[\frac{(2H - g + 1)}{(g + 1)} \right], \quad (10)$$

$$\left(\frac{w_{E\Phi}}{h} \right)_{BO} = \frac{2}{h} \operatorname{arch} \left[\frac{2H - g - 1}{g - 1} \right] + \frac{\operatorname{arch} \left[1 + \frac{2(w + \Delta w)}{h} \right]}{\pi}, \quad (11)$$

де $g = \operatorname{ch} \left[\frac{\pi}{2} \left(\frac{s - \Delta w}{h} \right) \right]$; $H = \operatorname{ch} \left[\pi \frac{w + \Delta w}{h} + \frac{\pi}{2} \left(\frac{s - \Delta w}{h} \right) \right]$; Δw – поправка за рахунок товщини смужки, мм;

$$\frac{\Delta w}{h} = \frac{t}{\pi h} \ln \frac{10,873}{\sqrt{\left(\frac{t}{h} \right)^2 + \frac{1/\pi}{w/t + 1,1}}}; \quad (12)$$

h – товщина підкладки, мм; w – ширина смужки, мм; s – зазор, мм; t – товщина смужки, мм.

Порівняння вимог до точності елементів конструкції проводились для НВ з коефіцієнтом зв'язку по напрузі $C = -6$ дБ; хвильовому опору мікросмушкової лінії передачі $Z_B = 50$ Ом; опору навантаження (відгалужувач Ланге) $R_1 = R_2 = 50$ Ом; діелектричної проникності матеріалу підкладки $\varepsilon = 9,6$; товщина підкладки $h = 1$ мм; товщині смужки $t = 0,01$ мм; кількості провідників (відгалужувач Ланге) $n = 4$.

Відхилення параметрів матеріалу і геометричних розмірів елементів конструкції обчислювались при заданому відхиленні коефіцієнту зв'язку – 5%. Результати розрахунків для НВ надаються в табл. 1.

Як витікає з табл. 1, найбільший вплив на коефіцієнт зв'язку НВ на зв'язаних лініях має зазор між провідниками, а в НВ Ланге – ширина провідника. Товщина підкладки суттєво впливає на коефіцієнт зв'язку в обох конструкціях. Однак НВ Ланге більш критичні до допусків на товщину підкладки.

Таблиця 1 – Результати розрахунку допусків для аналізуємих НВ

Тип НВ	Параметри	Номинальне значення	Нижня межа	Верхня межа	Нижня межа, %	Верхня межа, %
НВ на зв'язаних лініях	ε	9,600	9,3510	9,8490	2,59	2,59
	h , мм	1,000	0,9850	1,0146	1,50	1,46
	t , мм	0,010	0,0095	0,0106	5,20	5,50
	w , мм	0,610	0,5820	0,6390	4,65	4,91
	s , мм	0,072	0,0660	0,0760	8,00	6,00
НВ Ланге	ε	9,600	9,3950	9,8540	2,66	2,65
	h , мм	1,000	0,9810	0,0197	1,90	1,97
	t , мм	0,010	0,0093	0,0107	6,70	6,80
	w , мм	0,099	0,0940	0,1040	4,20	4,20
	s , мм	0,292	0,2790	0,3060	4,40	4,80

ДОСЛІДЖЕННЯ ТОЧНОСТІ МІКРОСМУЖКОВИХ ШЛЕЙФНИХ НАПРАВЛЕНИХ ВІДГАЛУЖУВАЧІВ

При обчисленні шлейфних направлених відгалужувачів складається з наступних етапів. Коефіцієнт зв'язку по напрузі розраховується за співвідношенням:

$$C_0 = 10^{-\lg \left(\frac{Y_2}{Y_1} \right)}, \quad (13)$$

де Y_1 та Y_2 – провідності шлейфів, $Y_i = Z_B/Z_i$.

Хвильові опори шлейфів Z_B та Z_i обчислюються за співвідношенням:

$$Z_{1,2} = \frac{42,4}{\sqrt{\varepsilon+1}} \cdot \ln \left\{ 1 + \frac{4 \cdot h}{(w_{E\Phi})_{1,2}} \cdot \left[\left(\frac{14 + \frac{8}{\varepsilon}}{11} \right) \cdot \frac{4 \cdot h}{(w_{E\Phi})_{1,2}} + \sqrt{\left(\frac{14 + \frac{8}{\varepsilon}}{11} \right)^2 \cdot \left(\frac{4 \cdot h}{(w_{E\Phi})_{1,2}} \right)^2 + \frac{1 + \frac{1}{\varepsilon}}{2} \cdot \pi^2} \right] \right\}. \quad (14)$$

Відносна ефективна ширина шлейфа обчислювалась за виразом:

$$\left(\frac{w_{E\Phi}}{h} \right)_{1,2} = \left(\frac{w}{h} \right)_{1,2} + \left(\frac{\Delta w}{h} \right)_{1,2}, \quad (15)$$

де Δw розраховується за формулою (12).

Результати розрахунку допусків на конструкторсько-технологічні параметри шлейфного НВ наведені в табл. 2.

Таблиця 2 – Результати розрахунку допусків шлейфового НВ

Параметри	Номинальне значення	Нижня межа	Верхня межа	Нижня межа, %	Верхня межа, %
ε	9,600	9,2500	9,9300	3,60	3,40
h , мм	1,000	0,9720	1,0300	2,80	3,00
t , мм	0,010	0,0093	0,0107	6,80	7,10
w_1 , мм	0,203	0,1970	0,2090	3,00	2,80
w_2 , мм	1,279	1,2040	1,3440	5,90	5,10

Як витікає з табл. 2, у шлейфних НВ найбільш критичним до допусків виявляється ширина зовнішніх шлейфів. Порівняння відхилень у конструкціях НВ показує, що в шлейфному НВ допуски на елементи конструкції менш жорсткі, але габаритні розміри найбільші.

ВИСНОВКИ

Дослідження точності на конструктивні параметри мікросмужкових НВ дає можливість порівняти НВ не тільки за габаритними розмірами, складністю конструкції, але і за жорсткістю допусків на конструктивні елементи НВ. При однакових вихідних даних шлейфний НВ має найбільші габарити. Номінальні розміри конструкції та їх допуски вище ніж в НВ на зв'язаних лініях та Ланге. Габаритні розміри НВ Ланге ненабагато перевищують габаритні розміри НВ на зв'язаних лініях. При цьому номінальні значення розмірів елементів конструкції в НВ Ланге більше ніж в НВ на зв'язаних лініях. Крім цього робоча смуга пропускання частот при використанні НВ Ланге розширюється.

Проведене дослідження показує, що НВ на зв'язаних лініях доцільно обирати, якщо необхідно забезпечити мінімальні габаритні розміри, простоту конструкції та відсутність жорстких допусків на елементи конструкції. Направлений відгалужувач Ланге доцільно обирати, якщо основними вимогами є мінімізація габаритних розмірів та розширення смуги пропускання. Шлейфний НВ доцільно обирати, якщо мають місце вимоги зменшення точності механічного виконання, стабілізація вихідної характеристики та відсутність умови мінімальних габаритних розмірів.

Запропоновані математичні моделі та приведені рекомендації по вибору мікросмужкових НВ можна рекомендувати для виконання проектних робіт на виробництві

та в навчальному процесі. Розглянуті математичні моделі синтезу допусків створюють передумови для утворення програмних комплексів дослідження пристроїв НВЧ за точністю.

ПЕРЕЛІК ПОСИЛАНЬ

1. Конструирование и расчет полосковых устройств. Учебное пособие для вузов. Под редакцией чл.-корр. Академии наук БССР проф. И.С. Ковалева, М., "Сов. радио", 1974.
2. Фуско В. СВЧ цепи: Анализ и автоматизированное проектирование: Пер. с англ. – М.: Радио и связь, 1990. – 288 с.: ил.
3. Малорацкий Л.Г., Явич Л.Р. Проектирование и расчет СВЧ элементов на полосковых линиях. – М.: "Сов. радио" 1972 – 232 с.
4. Справочник по расчету и конструированию СВЧ полосковых устройств / С.И. Бахарев, В.И. Вольман, Ю.Н. Либ и др. / Под ред. В. И. Вольмана. – М.: Радио и связь, 1982. – 328 с., ил.
5. Шилов Г. М., Воропай О. Ю., Гапоненко М. П. Интервальные методы назначения эксплуатационных допусков // Радиоэлектроника. Информатика. Управління. – 2003. – №1.
6. Проектирование интегральных СВЧ устройств. Справочник / Ю. Г. Ефремов, В. В. Кочин, Солганик и др. – К.: Техника, 1990. – 159 с.
7. Vladimir Krischuk, Alexey Farafonov, Sergey Romanenko "Analysis of projection methods of microstrip coupled lines filters".
8. Воробьев Е.А. Расчет производственных допусков устройств СВЧ. Л. "Судостроение" 1980. – 147 с.

Надійшла 08.04.2004
Після доробки 25.10.2004

Проведено исследование влияния допусков элементов конструкции на параметры направленных ответвителей: на связанных линиях, шлейфного и Ланге. Используются интервальные модели. Определены наиболее критические по точности конструктивные элементы. Это позволило предоставить рекомендации по выбору микрополосковых направленных ответвителей.

Research of influence of tolerances of elements a design on parameters directional couplers carried out: on the coupled lines, branch and Lange. Interval models are used. The most critical on accuracy constructive elements are determined. It has allowed to give recommendations at the choice of microstrip directional couplers.

УДК 621.335.2.024:621.314.1

В. Д. Флора

ОСОБЛИВОСТІ ЖИВЛЕННЯ ДВИГУНА ПОСТІЙНОГО СТРУМУ ЧЕРЕЗ ІМПУЛЬСНИЙ ПЕРЕТВОРЮВАЧ

Для імпульсного перетворювача, який знижує напругу нерегульованого джерела постійного струму, одержані аналітичні вирази загальних вольт-амперних, вхідних та зовнішніх характеристик перетворювача при живленні від різних джерел.

При побудові теорії силових напівпровідникових імпульсних перетворювачів постійного струму необхідно розглядати особливості їх живлення від різних джерел без фільтрів та з фільтрами, в широкому діапазоні зміни струмів навантаження.

Найчастіше при імпульсному регулюванні електричної енергії застосовуються схеми перетворювачів для зниження напруги нерегульованого джерела постійного

струму [1]. У роботі [2] для випадку живлення двигуна постійного струму через імпульсний перетворювач без фільтру для середньої напруги на виході при безперервному струмі навантаження одержане співвідношення:

$$U_{\text{ср}} = \frac{U - I_{\text{нсп}} \left[\alpha^2 R_{\text{дж}} + R_{\text{н}} \left(1 + \frac{1-\alpha}{\alpha} \cdot \frac{L_{\text{дж}} + L_{\text{н}}}{L_{\text{н}}} \right) \right]}{1 + \frac{1-\alpha}{\alpha} \cdot \frac{L_{\text{дж}} + L_{\text{н}}}{L_{\text{н}}}}, \quad (1)$$

де U – напруга холостого ходу джерела; $I_{\text{нсп}}$ – неперервний середній струм навантаження, який в усталеному

режимі для якірного ланцюга двигуна постійного струму

$$I_{\text{нсп}} = \frac{M_{\text{с}}}{C_{\text{м}} \Phi}; \quad (2)$$

$M_{\text{с}}$ – статичний момент навантаження на валу двигуна; $C_{\text{м}}$ – машинна стала; Φ – магнітний потік збудження двигуна; α – коефіцієнт заповнення імпульсу

$$\alpha = \frac{T_{\text{нк}}}{T}; \quad (3)$$

$T_{\text{нк}}$ – проміжок часу накопичення енергії в індуктивностях навантаження; T – період комутації; $R_{\text{дж}}$, $R_{\text{н}}$ – внутрішні активні опори джерела та навантаження; $L_{\text{дж}}$, $L_{\text{н}}$ – внутрішні індуктивності джерела та навантаження.

Характеристика (1) по суті є загальною вихідною вольт-амперною характеристикою перетворювача без фільтрів. Тобто, якщо навантаженням є двигун постійного струму, то в даному разі $U_{\text{ср}}$ є середньою електро рушійною силою якірного ланцюга двигуна, яка складається з ЕРС обертання та ЕРС самоіндукції:

$$U_{\text{ср}} = E_{\text{ср}}. \quad (4)$$

Тобто (1) – це залежність $E_{\text{ср}}$ від інших величин, які входять в цю формулу.

Зовнішня характеристика перетворювача може бути одержана з (1), якщо не враховувати втрати енергії на активному опорі навантаження ($R_{\text{н}} = 0$), тобто вважати всю енергію на виході перетворювача, яка подається до навантаження, корисною. При цьому при значній (у порівнянні з індуктивністю навантаження $L_{\text{н}}$) індуктивності джерела $L_{\text{дж}}$ потрібно враховувати $L_{\text{н}} \neq 0$, оскільки вона створює ЕРС самоіндукції, яка в цьому разі суттєво впливає на величину середньої напруги на виході перетворювача $U_{\text{срвих}}$:

$$U_{\text{срвих}} = \frac{U - I_{\text{нсп}} R_{\text{дж}} \alpha^2}{1 + \frac{(1 - \alpha) \cdot (L_{\text{дж}} + L_{\text{н}})}{\alpha L_{\text{н}}}}. \quad (5)$$

Якщо джерелом енергії при двигуновому навантаженні є акумуляторна батарея без вхідного фільтра перетворювача, то $L_{\text{дж}} \ll L_{\text{н}}$. Тобто наближено можна вважати $L_{\text{дж}} \approx 0$. Тоді зі співвідношення (1) з урахуванням (4):

$$E_{\text{ср}} = U \alpha - I_{\text{нсп}} (\alpha^3 R_{\text{дж}} + R_{\text{н}}), \quad (6)$$

а зовнішня характеристика перетворювача з формули (6):

$$U_{\text{срвих}} = U \alpha - I_{\text{нсп}} \alpha^3 R_{\text{дж}}. \quad (7)$$

Якщо на вході перетворювача встановлено Г-подібний LC-фільтр, то у формулі (1) потрібно замість напруги U враховувати середню напругу на вході фільтра $U_{\text{ф1ср}}$ [3]. Крім того, не потрібно враховувати $L_{\text{дж}}$ та $R_{\text{дж}}$. Тоді зі співвідношення (1):

$$E_{\text{ср}} = U_{\text{ср}} = U_{\text{ф1ср}} \alpha - I_{\text{нсп}} R_{\text{н}}. \quad (8)$$

В роботі [3] одержано для вхідного фільтру:

$$U_{\text{ф1ср}} = U - \alpha I_{\text{нсп}} (R_{\text{дж}} + R_{\text{ф1}}), \quad (9)$$

де $R_{\text{ф1}}$ – активний опір дроселя вхідного фільтру.

Підставляючи з (9) у (8), одержимо:

$$U_{\text{ср}} = E_{\text{ср}} = [U - \alpha I_{\text{нсп}} (R_{\text{дж}} + R_{\text{ф1}})] \alpha - I_{\text{нсп}} R_{\text{н}}. \quad (10)$$

Тоді на виході перетворювача при $R_{\text{н}} = 0$:

$$U_{\text{срвих}} = U \alpha - I_{\text{нсп}} \alpha^2 (R_{\text{дж}} + R_{\text{ф1}}). \quad (11)$$

У випадку, коли живлення здійснюється, наприклад, від батареї ультраконденсаторів [4], потужної акумуляторної батареї і т. і., можливо знехтувати опором джерела й дроселя фільтру. Тоді з (11) одержимо відому формулу

$$U_{\text{срвих}} = U \alpha. \quad (12)$$

Для того, щоб одержати напругу на вході перетворювача U , потрібно за формулою (12) поділити $U_{\text{срвих}}$ на α :

$$U_{\text{вх}} = U = \frac{U_{\text{срвих}}}{\alpha}. \quad (13)$$

Враховуючи $R_{\text{дж}}$ та $R_{\text{ф1}}$, з (11):

$$U_{\text{вх}} = U - I_{\text{нсп}} \alpha (R_{\text{дж}} + R_{\text{ф1}}). \quad (14)$$

Зі співвідношення (6):

$$U_{\text{вх}} = U - I_{\text{нсп}} \alpha^2 R_{\text{дж}}, \quad (15)$$

а з формули (5):

$$U_{\text{вх}} = \frac{U - I_{\text{нсп}} R_{\text{дж}} \alpha^2}{\alpha + (1 - \alpha) \cdot \frac{(L_{\text{дж}} + L_{\text{н}})}{L_{\text{н}}}}. \quad (16)$$

Досить поширеним випадком є живлення навантаження через імпульсний перетворювач від діодного випрямляча. Для цього у всіх одержаних формулах для різних випадків вольтамперної характеристики імпульсного перетворювача слід враховувати комутацію діодного випрямляча. Найчастіше живлячою мережею є трифазна система змінного струму. При роботі трифазного мостового діодного випрямляча виникають режими, коли послідовно увімкнені дві фази, а третя вимкнена, або паралельно з'єднані дві фази і цей ланцюг увімкнений послідовно з третьою фазою. Для цих випадків у трифазних вентильних електричних машинах використовується формула для еквівалентного опору обмотки трифазного якоря [5]:

$$r_{\text{я}} = \frac{2r \left(\frac{\pi}{3} - \gamma \right) + 1,5r\gamma}{\frac{\pi}{3}} = \left(2 - 1,5 \frac{\gamma}{\pi} \right) r, \quad (17)$$

де r – активний опір фази обмотки; γ – кут комутації діодного випрямляча.

Якщо у робочому режимі $\gamma = 15 \div 20^\circ$, то

$$r_{\text{я}} \approx 1,85r. \quad (18)$$

Аналогічно можливо підрахувати й еквівалентну індуктивність якоря:

$$L_{\text{я}} \approx (2 - 1,5 \frac{\gamma}{\pi}) L_{\text{ф}} \approx 1,85 L_{\text{ф}}, \quad (19)$$

де $L_{\text{ф}}$ – індуктивність однієї фазної обмотки.

Отже, у випадку живлення імпульсного перетворювача від трифазної мережі змінного струму через діодний випрямляч можливо вважати, що за формулами (17), (19):

$$R_{\text{дж}} = r_{\text{я}}, \quad (20)$$

$$L_{\text{дж}} = L_{\text{я}}. \quad (21)$$

З урахуванням формул (20), (21) для цього випадку живлення можливо розрахувати потрібні вольт-амперні характеристики перетворювача за вищенаведеними співвідношеннями.

Якщо імпульсний перетворювач живиться через діодний випрямляч від трифазної мережі змінного струму, то в результаті періодичних комутацій в діодному випрямлячі, які відбуваються з деякою частотою, періодично змінюються $R_{\text{дж}}$, $L_{\text{дж}}$. В одних проміжках часу це $2r$, $2L_{\text{ф}}$, а в інших – $1,5r$; $1,5L_{\text{ф}}$. Тому виникають періодичні (інфранизкочастотні) коливання напруги на виході імпульсного перетворювача, частота яких залежить від співвідношення частот виникнення комутаційного процесу у випрямлячі та комутації імпульсного перетворювача.

Іншою причиною появи низькочастотних коливань вихідної напруги перетворювача є така. Фільтровий конденсатор з ємністю $C_{\text{ф1}}$ та індуктивністю $L_{\text{дж}} + L_{\text{ф1}}$, де $L_{\text{ф1}}$ – індуктивність вхідного фільтрового дроселя, створює коливальний контур. При періодичному (з частотою комутації випрямляча) змінні $L_{\text{дж}}$ виникають інфранизкочастотні коливання, частота яких залежить від співвідношення власної частоти вхідного фільтру та частоти комутації діодного випрямляча.

При живленні імпульсного перетворювача з вхідним фільтром від генератора постійного струму на виході

перетворювача виникають низькочастотні коливання енергії між фільтром та генератором.

З виконаних досліджень можливо зробити такі висновки.

1. Одержані загальні співвідношення вихідної вольт-амперної характеристики імпульсного перетворювача для зниження напруги нерегульованого джерела, з якої можливо одержати вхідну та зовнішню характеристики з вхідним фільтром та без нього.

2. При живленні імпульсного перетворювача від випрямляча або генератора виникають низькочастотні коливання вихідної напруги, які залежать від співвідношення частот комутації випрямляча та комутатора, а також – від автоколивальних процесів обміну енергії між конденсатором фільтру та індуктивністю джерела енергії.

ПЕРЕЛІК ПОСИЛАНЬ

1. Флора В.Д., Коробков Ю.С. Полупроводниковые устройства: Методическое пособие по курсу "Электрические и электронные аппараты" для студентов, обучающихся по направлению "Электротехника, электромеханика и электротехнологии". М.: Издательство МЭИ, 2002 – 64 с.
2. Флора В.Д. Влияние параметров джерела та навантаження на характеристики імпульсного перетворювача в режимі зниження напруги. // Радіоелектроніка, інформатика, управління. – 2000. – №2. – С. 26 – 30.
3. Флора В.Д. Визначення параметрів вхідного та вихідного фільтрів знижуючого напругу імпульсного перетворювача для електротранспорту. // Електротехніка та електроенергетика. – 2003. – № 1. – С. 18–24.
4. Винников Д., Бойко В., Лаугис Ю. Исследование возможностей применения ультраконденсаторов в электросистеме автомобиля. // Техническая электродинамика. Тематичний випуск: Силова електроніка та енергоефективність. Ч. 3. – 2003. – С. 27–30.
5. Захарченко Д.Д., Ротанов Н.А. Тяговые электрические машины. – М.: Транспорт, 1991. – 343 с.

Надійшла 26.03.2004
Після доробки 05.10.2004

Для імпульсного преобразователя, понижающего напряжение нерегулируемого источника постоянного тока, получены аналитические выражения общих вольт-амперных, входных и внешних характеристик преобразователя при питании от различных источников.

The general dependence of volt-ampere characterization for the pulse inverter getting different power supply is obtained. This dependence allows to get an external, regulating and other characterizations.

УДК 621.396

Н. П. Чернобородова, М. П. Чернобородов

ЦИФРОВАЯ СИСТЕМА КОМПЕНСАЦИИ ИМПУЛЬСНЫХ ПОМЕХ

Проведено математическое моделирование линейного одноканального алгоритма работы цифровой системы компенсации несинхронных импульсных помех в когерентно-импульсном радиолокаторе. Показана высокая эффективность предложенной системы компенсации в сложной сигнально-помеховой обстановке при использовании простых зондирующих сигналов.

ВВЕДЕНИЕ

К современным радиолокационным станциям (РЛС) предъявляются высокие требования к помехозащищенности в условиях воздействия комбинированных помех. Существующие технические решения при воздействии

несинхронных импульсных помех предполагают включение в состав системы помехозащиты РЛС известной [1] схемы ШОУ (широкая полоса – ограничение – узкая полоса), предложенной в 1943 г. Щусевым. Однако применение нелинейных алгоритмов обработки сигналов в приемном тракте РЛС приводит к снижению эффективности доплеровской фильтрации сигналов. Обусловлено это возникновением дополнительных спектральных составляющих вследствие нелинейного преобразования принимаемой смеси сигналов и помех. Поэтому использование алгоритма ШОУ существенно ограничивает возможности когерентно-импульсных РЛС в части обнаружения полезных сигналов на фоне пассивной компоненты комбинированной (пассивная + несинхронная импульсная) помехи.

ПОСТАНОВКА ЗАДАЧИ

В [2] и [3] рассмотрена перспективная линейная одноканальная цифровая система компенсации (ЦСК) несинхронных импульсных помех (НИП), предназначенная для устранения влияния действующей помехи на процедуру обнаружения целей в цифровых системах первичной обработки радиолокационных станций. При этом исследование проводилось в предположении, что несинхронная импульсная помеха присутствовала среди откликов только от одного зондирующего импульса частотной пачки. Т. е. длительность такой одиночной несинхронной импульсной помехи не превышала периода повторения зондирующих импульсов анализируемой частотной пачки, и такая помеха попадала в анализируемый дискрет дальности только один раз за одну частотную пачку зондирующих импульсов (более подробно см. [2]).

Однако, в реальной сигнально-помеховой обстановке нередко встречаются ситуации, когда несинхронные импульсные помехи попадают в обрабатываемый дискрет дальности чаще, чем один раз за пачку зондирующих импульсов. При этом амплитудный портрет спектра таких импульсных помех перестает быть равномерным, что нередко усложняет процедуру их обнаружения и/или процедуру обнаружения целей на их фоне. В этой связи возникает необходимость рассмотреть более общий случай, когда в обрабатываемом дискрете дальности частотной пачки, содержащей N эхо-импульсов, присутствует несколько несинхронных импульсных помех.

Пусть на вход цифровой системы первичной обработки РЛС поступают последовательности откликов N зондирующих импульсов, составляющих аддитивную смесь сигналов от целей, активных и пассивных помех, принятых антенной радиолокатора:

$$\hat{S}(t) = \hat{S}_H(t) + \hat{S}_{ш.п.}(t) + \hat{S}_Ц(t) + \hat{S}_{п.п.}(t), \quad (1)$$

где $\hat{S}_H(t)$ – несинхронная импульсная помеха; $\hat{S}_{ш.п.}(t)$ – активная шумовая помеха; $\hat{S}_Ц(t)$ – отраженный от цели сигнал; $\hat{S}_{п.п.}(t)$ – пассивная помеха.

Предположим, что с выхода устройства формирования квадратурных составляющих на вход ЦСК НИП

поступают 3 импульсные помехи в разные моменты времени (среди откликов от разных зондирующих импульсов анализируемого дискрета дальности). Рассмотрим случай прохождения только импульсных помех, т. е. $\hat{S}(t) = \hat{S}_H(t)$.

АЛГОРИТМ РЕШЕНИЯ ПРИ ОДИНОЧНОЙ ПОМЕХЕ

В [2] было показано, что одиночная НИП после расфилтровки присутствует во всех фильтрах с модулем амплитуды $|\hat{G}_{нип}|$, равным модулю амплитуды НИП на входе устройства ДПФ $|\hat{S}_{нип}|$, а сдвиг фаз между спектральными составляющими НИП соседних фильтров ДПФ – постоянный и определяется временем поступления одиночной помехи на вход ЦСК. Т. е., одиночная НИП имеет равномерный амплитудный и линейно изменяющийся фазовый спектры. Поскольку дискретное преобразование Фурье производит когерентное накопление обрабатываемых входных данных с одинаковыми модулями амплитуд и линейно изменяющимися (нарастающими или убывающими) фазами, то спектральные составляющие НИП могут быть обработаны по алгоритму преобразования Фурье с целью когерентного накопления амплитуды помехи в одном из фильтров. В случае когерентного накопления импульсной помехи в одном из фильтров становится возможным ее обнаружение. Т. о., обработка спектральных составляющих $\hat{G}(n)$, где n – номер фильтра, по алгоритму преобразования Фурье (повторно) позволяет определить момент поступления в цифровую систему обработки одиночной импульсной помехи. Эта функция $\hat{X}(i)$ позволяет когерентно накопить амплитуду НИП в том фильтре повторного преобразования Фурье i , который соответствует номеру зондирующего импульса l , среди откликов от которого присутствует одиночная импульсная помеха, т. е. определить момент (время) воздействия НИП. Таким образом, функция $\hat{X}(i)$ для рассматриваемого дискрета дальности будет иметь унимодовую характеристику.

АЛГОРИТМ РЕШЕНИЯ В СЛОЖНОЙ ПОМЕХОВОЙ ОБСТАНОВКЕ

В случае воздействия нескольких импульсных помех в одном дискрете дальности, функция $\hat{X}(i)$ будет иметь многомодовую характеристику с числом мод, равным количеству импульсных помех в текущем дискрете частотной пачки. Номера фильтров повторного преобразования Фурье с когерентно накопленными амплитудами помех содержат информацию о времени (моментах) воздействия каждой из помех, которые соответствуют тем периодам повторения зондирующих импульсов, среди откликов от которых в рассматриваемом дискрете дальности присутствовали импульсные помехи.

Таким образом, как в случае одиночной импульсной помехи, так и в случае нескольких импульсных помех в дискрете, определив номера фазовых фильтров повторного преобразования Фурье, в которых произошло коге-

рентное накопление дискретизированных функций времени $\hat{X}_i(l)$, можно вычислить номера зондирующих импульсов l , среди откликов от которых присутствовали импульсные помехи (более подробно см. [2]).

Анализ прохождения импульсных помех через устройство вычисления преобразования Фурье позволяет сформулировать алгоритмы обнаружения, определения параметров и компенсации импульсных помех. На рис. 1 представлена соответствующая им структурная схема ЦСК НИП, реализованная по циклическому алгоритму компенсации импульсных помех во временной области [3].

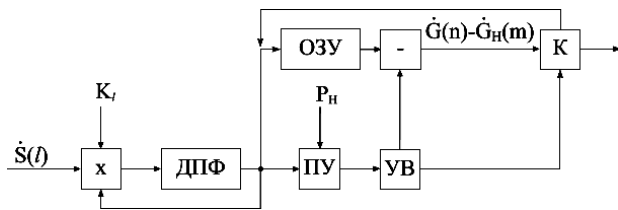


Рисунок 1 – Структурная схема ЦСК НИП:

ОЗУ – оперативное запоминающее устройство; «x» – умножитель; K_l – коэффициенты применяемого весового окна; ДПФ – устройство вычисления дискретного преобразования Фурье; ПУ – пороговое устройство; УВ – устройство вычисления параметров несинхронной импульсной помехи; «-» – устройство вычитания; К – коммутатор

ЦСК НИП работает следующим образом. На вход системы поступают квадратурные составляющие $\hat{S}(l)$, соответствующие данным текущего дискрета некоторой частотной пачки принятой радиолокационной информации ($l = 1 \dots N$, где N – количество зондирующих импульсов). Эта входная информация поступает на вход устройства вычисления дискретного преобразования Фурье (ДПФ) для получения спектра входного сигнала. Далее эти спектральные составляющие задерживаются в ОЗУ на время, необходимое для проведения компенсации всех возможно присутствующих импульсных помех. Одновременно, спектральные составляющие входного сигнала повторно поступают на вход устройства ДПФ для определения момента поступления НИП на вход ЦСК. При использовании весовой обработки с целью минимизации уровня боковых лепестков соседних (перекрывающихся) фильтров ДПФ входные данные домножаются на коэффициенты применяемого весового окна K_l .

Повторно обработанные спектральные составляющие входного сигнала по алгоритму преобразования Фурье поступают на вход порогового устройства ПУ, где производится сравнение их амплитуд с порогом обнаружения импульсной помехи P_n . В случае обнаружения каждой очередной импульсной помехи она поступает на вход устройства вычисления (УВ) параметров импульсных помех (начальной фазы, амплитуды и номера зондирующего импульса m , среди откликов от которого присутствует очередная импульсная помеха).

По известным параметрам обнаруженной импульсной помехи в устройстве вычитания производится ее компенсация в спектральной области: из каждой задержанной линией задержки ОЗУ спектральной составляющей ана-

лизируемого дискрета $\hat{G}(n)$ вычитается комплексная амплитуда помехи $\hat{G}_n(m)$, вычисленная в устройстве УВ. Скомпенсированный таким образом сигнал $\hat{G}_k(n) = \hat{G}(n) - \hat{G}_n(m) \forall n = 1 \dots N$, поступает на вход коммутатора.

В случае обнаружения хотя бы одной импульсной помехи в рассматриваемом дискрете дальности скомпенсированный сигнал через коммутатор снова задерживается в ОЗУ и одновременно подается на вход устройства ДПФ с целью обнаружения и последующей компенсации возможно присутствующих оставшихся импульсных помех. Необходимость в такой обратной связи обуславливается применением адаптивного порога обнаружения НИП P_n .

С целью иллюстрации работы алгоритма ЦСК НИП, ограничимся рассмотрением случая воздействия на РЛС, излучающей частотные пачки из 8 простых зондирующих импульсов, только импульсных помех, т. е. $\hat{S}(t) = \hat{S}_n(t)$. Будем считать, что НИП воздействует на РЛС в 3-м, 5-м и 6-м периодах повторения зондирующих импульсов. Модули амплитуды и фазы НИП рассматриваемого дискрета дальности текущей частотной пачки на входе устройства ДПФ в зависимости от номера зондирующего импульса l представлены на рис. 2 и 3, соответственно.

Приведенные на рис. 2 и 3 данные, обработанные по алгоритму 8-точечного ДПФ без весового окна, представлены на рис. 4 и 5, соответственно, в зависимости от номера фильтра n . Из рис. 4 и 5 следует, что при воздействии нескольких импульсных помех в одном дискрете дальности разных периодов повторения их амплитудный спектр (см. рис. 4) не будет равномерным, а фазовый (см. рис. 5) – линейно изменяющимся. Но при этом амплитудно-фазовые соотношения между спектральными составляющими соседних фильтров ДПФ такой «суммарной» помехи позволят произвести когерентное накопление присутствующих НИП путем (повторной) обработки по алгоритму ДПФ спектра входного сигнала. На рис. 6 и 7 представлены модули амплитуд и фазы НИП после повторной обработки по алгоритму ДПФ в зависимости от номера зондирующего импульса l : все НИП когерентно накопились в фазовых фильтрах ДПФ, соответствующих времени их поступления на вход ЦСК (см. рис. 2 и 3). При этом модули амплитуд когерентно накопленных НИП $|\hat{X}_{\text{НИП}}|$ (см. рис. 6) с точностью до коэффициента когерентного накопления ДПФ (равному числу импульсов или фильтров) $N = 8$ соответствует модулям амплитуд НИП, присутствующих на входе ЦСК (см. рис. 2). Фазы когерентно накопленных НИП (см. рис. 7) соответствуют фазам НИП, присутствовавших на входе системы компенсации (см. рис. 3).

С целью обнаружения НИП, возможно присутствующих среди откликов в рассматриваемом дискрете дальности, необходимо последовательно сравнить амплитуду каждой составляющей $|\hat{X}(l)|$ с величиной порога обнаружения импульсной помехи P_n . При выполнении условия $|\hat{X}(l)| \geq P_n$ становится возможным определение момента поступления одиночной импульсной помехи [4].

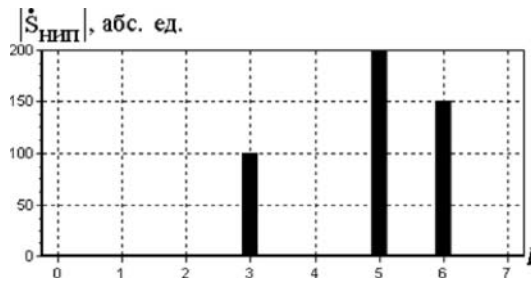


Рисунок 2 – Модули амплитуд НИП на входе ДПФ

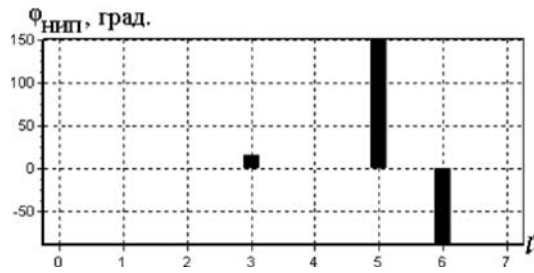


Рисунок 3 – Фазы НИП на входе ДПФ

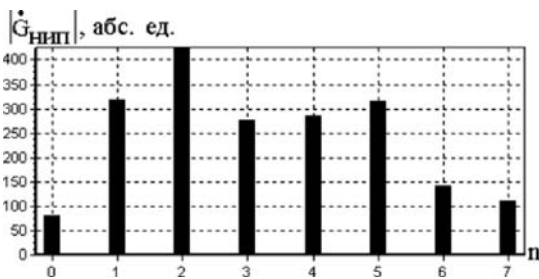


Рисунок 4 – Модули амплитуд спектра НИП после обработки по алгоритму ДПФ

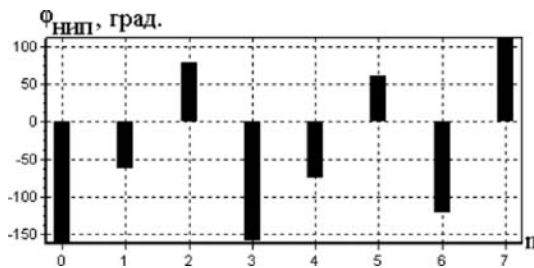


Рисунок 5 – Фазы спектра НИП после обработки по алгоритму ДПФ

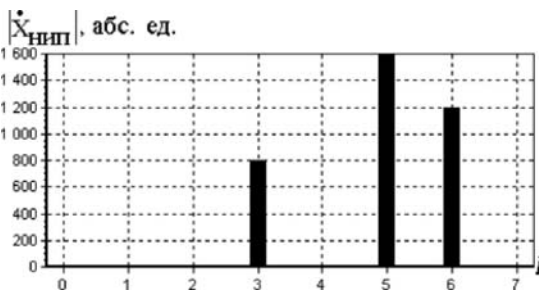


Рисунок 6 – Модули амплитуд НИП после повторной обработки по алгоритму ДПФ

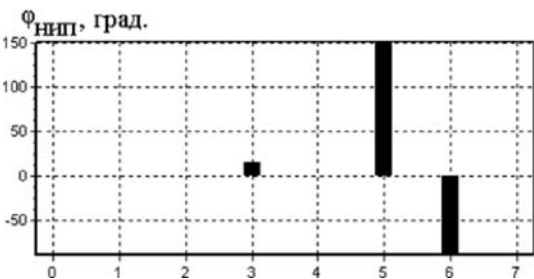


Рисунок 7 – Фазы НИП после повторной обработки по алгоритму ДПФ

Таким образом, обработка спектральных составляющих $\dot{G}(n)$ входного сигнала по алгоритму преобразования Фурье позволяет произвести обнаружение всех присутствующих в анализируемом дискрете дальности НИП, вычислить их параметры (определить моменты поступления каждой помехи, ее модуль амплитуды и начальную фазу) и произвести их компенсацию.

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Для оценки эффективности ЦСК НИП, представленной на рис. 1, было проведено математическое моделирование. С выхода устройства формирования квадратурных составляющих на вход цифровой системы компенсации поступала аддитивная смесь НИП, шумовой помехи, отраженного от цели сигнала и пассивной помехи со следующими параметрами для рассматриваемого дискрета дальности:

НИП № 1: амплитуда – 9 разрядов АЦП (150 абс. ед.); начальная фаза – 15°; время – присутствует среди откликов только от зондирующего импульса № 2 частотной пачки;

НИП № 2: амплитуда – 8 разрядов АЦП (100 абс. ед.); начальная фаза – 100; время – присутствует среди откликов только от зондирующего импульса № 3;

НИП № 3: амплитуда – 9 разрядов АЦП (200 абс. ед.); начальная фаза – 70°; время – присутствует среди откликов только от зондирующего импульса № 7; дисперсия собственных шумов – 3 разряда АЦП при нулевом математическом ожидании; цель: амплитуда – 4 разряда АЦП ($|\dot{S}_ц(t)| = 10$ абс. ед.), начальная фаза – 0, доплеровский сдвиг фазы – 90; пассивная помеха: амплитуда – 4 разряда АЦП ($|\dot{S}_{п.п.}(t)| = 10$ абс. ед.), начальная фаза – 15, доплеровский сдвиг фазы пассивной помехи – 10.

Сформированные квадратурные составляющие аддитивной смеси подавались на вход ЦСК НИП с целью ком-

пенсации присутствующих импульсных помех. При вычислении 8-точечного ДПФ весовое окно не применялось.

На рис. 8...12 представлены результаты работы ЦСК НИП, позволяющие оценить ее эффективность в условиях сложной сигнально-помеховой обстановки.

На рис. 8 приведены модули амплитуды спектра входного сигнала без компенсации. Если не учитывать шумовую составляющую входной смеси сигналов, то, согласно заданным параметрам, модули амплитуд спектральных составляющих цели и пассивной помехи можно рассчитать исходя из следующих соображений. Для применяемого 8-точечного ДПФ без весовой обработки входных данных, отклики с постоянным модулем амплитуды $|\dot{A}|$ и доплеровскими сдвигами фазы, кратными $360/N=360/8=45$ (N – количество фильтров ДПФ), будут попадать строго в центр фильтра ДПФ и когерентно накапливаться в нем с амплитудой, равной $|\dot{A}| \cdot N$. Поскольку доплеровский сдвиг фазы цели задавался 90, то отклик от цели должен накопиться в фильтре ДПФ № 2 с амплитудой, равной $|\dot{S}_{ц}(t)| \cdot N = 10 \cdot 8 = 80$ абс. ед. Отклик от пассивной помехи, должен накопиться в фильтре ДПФ № 0 с амплитудой, приблизительно (из-за флуктуации положения местного предмета, поскольку доплеровский сдвиг фазы пассивной помехи задавался не нулевым) равной $|\dot{S}_{п.п}(t)| \cdot N = 10 \cdot 8 = 80$ абс. ед.

В результате моделирования установлено (см. рис. 8), что вследствие воздействия НИП модуль спектральной составляющей пассивной помехи увеличился в 4 раза (до 260 абс. ед.), а модуль спектральной составляющей цели уменьшился в 4 раза (до 20 абс. ед.) относительно заданных значений. Резкое увеличение амплитуды пассивной

помехи может вызвать появление ложного обнаружения, особенно в случае применения в устройствах обнаружения, входящих в состав ЦСК РЛС, алгоритмов выявления целей, движущихся с нулевой радиальной скоростью относительно РЛС, например, алгоритмов с использованием карты местных предметов [5].

Уменьшение когерентно накопленной амплитуды цели до 20 абс. ед. из-за воздействия НИП приводит к существенному усложнению обнаружения цели из-за того, что ее амплитуда стала очень близкой к значению модулей амплитуд спектра шумовой составляющей входного сигнала (см. рис. 9). При этом обнаружение не возможно ни для простейшего порогового обнаружителя, для которого значение порога определяется фиксированным уровнем шумов (т. е. с постоянным порогом обнаружения по дальности – см., например, [5]), ни для систем стабилизации ложных тревог с обучающей выборкой (т. е. с адаптивным порогом обнаружения по дальности – см., например, [6]). В последнем случае порог обнаружения формируется для каждого дискрета дальности анализируемой частотной пачки в зависимости от уровня шумов в скользящем «окне» по дальности, а анализируемый дискрет дальности находится в середине такого «окна».

Существенное возрастание уровня шумов вследствие присутствия НИП (см. рис. 8) по сравнению с их действительным уровнем (см. рис. 9), приводит к тому, что описанные обнаружители сформируют сигналы ложного обнаружения по фильтрам № 1, 3...7.

На рис. 10 представлены модули амплитуды спектра сигнала $\dot{S}(t) = \dot{S}_{ш.п.}(t) + \dot{S}_{ц}(t) + \dot{S}_{п.п.}(t)$ с указанными выше параметрами в отсутствие НИП (что эквивалентно

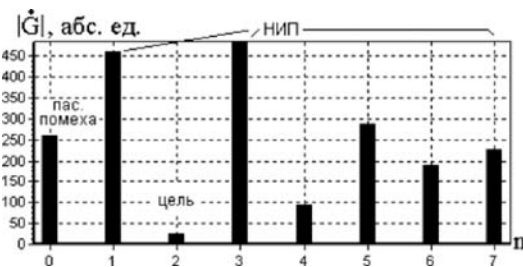


Рисунок 8 – Модуль амплитуды сигнала $\dot{S}(t)$ без компенсации НИП

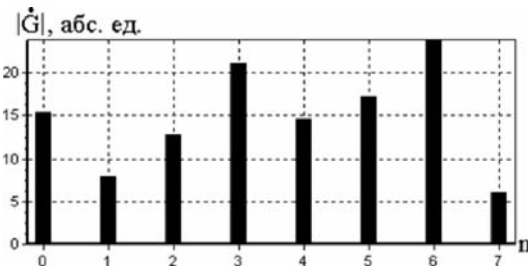


Рисунок 9 – Модуль амплитуды шумовой составляющей сигнала $\dot{S}(t)$

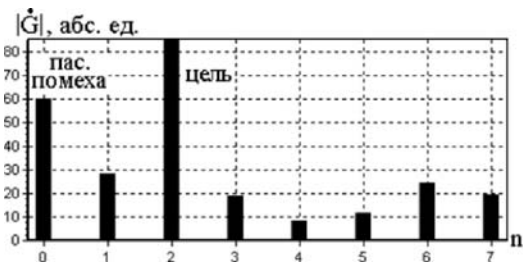


Рисунок 10 – Модуль амплитуды смеси $\dot{S}(t)$ при идеальной компенсации НИП

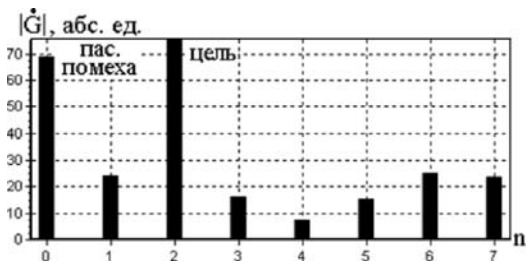


Рисунок 11 – Модуль амплитуды $\dot{S}(t)$ на выходе цифровой системы компенсации НИП (при реальной компенсации НИП)

случаю идеальной компенсации) как эталонные для сравнения с представленными на рис. 11 модулями амплитуды спектра сигнала $\hat{S}(t) = \hat{S}_n(t) + \hat{S}_{ш.п.}(t) + \hat{S}_ц(t) + \hat{S}_{п.п.}(t)$ с выхода ЦСК (что соответствует случаю реальной компенсации).

Представленные на рис. 11 модули амплитуды сигнала $\hat{S}(t)$ свидетельствуют о высокой эффективности работы ЦСК НИП: все 3 присутствующие импульсные помехи подавлены полностью, а значения когерентно накопленных амплитуд пассивной помехи и цели в соответствующих фильтрах ДПФ после компенсации помех изменились в пределах 10%. Уровень шумов в фильтрах № 1, 3, 4 несущественно уменьшился, а в фильтрах № 5...7 – несущественно возрос. Кроме того, сравнение спектральных составляющих рис. 10 и 11 свидетельствует о том, что ЦСК НИП не вносит нелинейных искажений в скомпенсированный сигнал.

ВЫВОДЫ

В результате математического моделирования установлено, что, в общем случае, в условиях сложной сигнально-помеховой обстановки предложенный линейный алгоритм компенсации несинхронной импульсной помехи имеет достаточно высокую эффективность. В дальнейшем необходимо провести дополнительные исследования работоспособности предложенного алгоритма в условиях использования сложных сигналов.

ПЕРЕЧЕНЬ ССЫЛОК

1. Гоноровский И. С. Радиотехнические цепи и сигналы, часть II, М., Сов. радио, 1967. – 328 с.
2. Часовский В. А., Чернобородова Н. П., Чернобородов М. П., Пиза Д. М. Цифровая система компенсации несинхронных импульсных помех // Радиотехника. Информатика. Управління. – 2002. – №2. – С. 41–44.
3. Чернобородова Н. П., Чернобородов М. П. Моделирование цифровой системы компенсации несинхронных импульсных помех // Радиотехника. Информатика. Управління. – 2003. – №2. – С. 74–78.
4. Цифровая система компенсации несинхронных импульсных помех: Пат. 58183 А Україна: МКИ³ G 01 S 7/36/ Часовський В. О., Чернобородова Н. П., Чернобородов М. П.; №2002108319; Заявлено 21.11.2002; Опубл. 15.07.2003, Бюл. №7.
5. Бакулев П. А., Степин В. М. Методы и устройства селекции движущихся целей. – М.: Радио и связь, 1986. – 288 с., ил.
6. Пристрій для визначення середнього рівня шуму за його відліками: Пат. 56651 А Україна: МКИ³ G 01 S 7/36, H 04 B 15/00/ Чернобородова Н. П., Чернобородов М. П.; №2002086587; Заявл. 07.08.2002; Опубл. 15.05.2003, Бюл. №5.

Надійшла 17.05.2004
Після доробки 20.11.2004

Проведено математичне моделювання лінійного одноканального алгоритму роботи цифрової системи компенсації несинхронних імпульсних завад у когерентно-імпульсному радіолокаторі. Показана висока ефективність запропонованої системи компенсації у складному сигнально-завадовому оточенні за використання простих імпульсів зондування.

There has been mathematically modelled the linear single-channel algorithm of the digital system operation for cancelling asynchronous pulse interface in the coherent pulse radar. A high efficiency of the offered cancellation system is shown for complicated signal interference environment when using simple transmitting signals.

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL AND COMPUTER MODELLING

УДК 681.3+681.5:007

Н. В. Алипов, И. Н. Алипов, М. И. Хиль, В. Н. Сидоров

МЕТОДЫ ЗАЩИТЫ ИНФОРМАЦИИ В ДИСКРЕТНОМ КАНАЛЕ НА ОСНОВЕ ПОМЕХОУСТОЙЧИВЫХ К СИММЕТРИЧНЫМ РЕГУЛЯРНЫМ ВОЗДЕЙСТВИЯМ АЛГОРИТМОВ ПОИСКА ТОЧКИ С ХАРАКТЕРНЫМ ПРИЗНАКОМ

Строятся помехоустойчивые к симметричным регулярным виртуальным последовательностям алгоритмы поиска точки с характерным признаком. Такие алгоритмы задают функционирование конечных автоматов с псевдослучайными переходами из одного состояния в другое. Подобные дискретные автоматы используются в системах защиты информации для генерации шифра защиты.

ВВЕДЕНИЕ

Новым направлением в развитии криптографических методов защиты информации является направление, связанное с применением конечных автоматов с псевдослучайными переходами из начального состояния в одно и то же конечное состояние [1]. Такие автоматы являются генераторами шифров замены. Причем для одного того же символа входного алфавита формируется некоторое множество кодовых комбинаций неравномерного кода. Выбор той или иной кодовой комбинации (шифра замены) осуществляется конечным автоматом [1] псевдослучайным образом. Для организации подобного блуждания используют специальные алгоритмы их функционирования. Основу таких алгоритмов составляют помехоустойчивые алгоритмы поиска точки с характерным признаком в условиях воздействия на процесс

поиска различного рода возмущений [2]. В работе [3] приведены помехоустойчивые алгоритмы поиска точки с характерным признаком для симметричного регулярного воздействия, у которого пауза между отрицательным и положительным выбросами отсутствует ($H = 0$), длительность импульса равнялась шагу алгоритма ($\ell = 1$), амплитуда выброса постоянна и равна величине $a \cdot \delta$ (δ – дискретность преобразования). На основе таких алгоритмов поиска были разработаны методы защиты информации.

Цель исследования – синтез алгоритмов поиска точки с характерным признаком, помехоустойчивых к регулярным симметричным воздействиям, для которых интервал времени между двумя соседними выбросами $H > 0$, длительность выброса $\ell \geq 1$, амплитуда выброса $a > 0$.

ПОМЕХОУСТОЙЧИВЫЕ АЛГОРИТМЫ ПОИСКА ТОЧКИ С ХАРАКТЕРНЫМ ПРИЗНАКОМ

Первоначально приведем решение задачи синтеза алгоритмов поиска для случая, когда $H = \ell = 1$, количество шагов эксперимента $k = 1$.

Предположим, что оптимальный алгоритм существует, и он, за i шагов, разбивает исходный интервал неопределенности $(0,1)$ на $\Psi_{2,8}^{a,\ell,H}(i,1)$ равных частей. Пусть некоторым образом выбрана точка эксперимента x_1^1 и выполнен первый шаг. Тогда возможен один из исходов:

$$a) x(t_1) \in [0, x_1^1]; \quad b) x(t_1) \in [x_1^1, 1).$$

На втором шаге алгоритма применяем принцип повторных сравнений: $(x_1^2 = x_1^1)$ и получаем один из исходов:

$$a_1) x(t_1 + \Delta t) \in [0, x_1^2]; \quad a_2) x(t_1 + \Delta t) \in [x_1^2, 1);$$

$$b_1) x(t_1 + \Delta t) \in [x_1^2, 1); \quad b_2) x(t_1 + \Delta t) \in [0, x_1^2).$$

Для исходов $a_1)$ и $b_1)$, исходя из того, что длительность помехи равна длительности шага алгоритма, соответственно устанавливаем:

$$x \in [0, x_1^2]; \quad x \in [x_1^2, 1).$$

По предположению, оптимальный алгоритм существует, в его распоряжении осталось $(i-2)$ шага. Посредством этого алгоритма указанные полукрытые интервалы неопределенности будут разбиты соответственно на $\Psi_{2,8}^{a,1,1}(i-2,1)$ частей равных частей. Откуда имеем:

$$\Psi_{2,8}^{a,\ell,H}(i,1) = 2 \cdot \Psi_{2,8}^{a,\ell,H}(i-2,1). \quad (1)$$

Для исходов $a_2)$ и $b_2)$ возникло противоречие, которое свидетельствует о действии помехи на первом либо на втором шагах алгоритма. Для устранения этой неопределенности на третьем шаге снова применяем принцип повторных сравнений $(x_1^3 = x_1^2)$ и получаем исходы

$$a_{21}) x(t_1 + 2\Delta t) \in [x_1^3, 1); \quad a_{22}) x(t_1 + 2\Delta t) \in [0, x_1^3];$$

$$b_{21}) x(t_1 + 2\Delta t) \in [0, x_1^3]; \quad b_{22}) x(t_1 + 2\Delta t) \in [x_1^3, 1).$$

Для исхода $a_{21})$ характерно то, что на первом такте действовала помеха отрицательной полярности, на втором ее проявление отсутствовало, на третьем такте действовала помеха положительной полярности.

Итак, проявление помехи обнаружено (помеха действует на нечетных шагах алгоритма).

Действуя только на четных шагах алгоритма непомехоустойчивым алгоритмом, разобьем полукрытый интервал неопределенности $[x_1^3, 1)$ на $\Phi_{2,8}(\alpha, 1)$ равных частей, где $\Phi_{2,8}(\alpha, 1) = (k+1)^\alpha$;

$$\alpha = \begin{cases} \frac{i-3}{2}, & \text{если } (i-3) \bmod 2 = 0; \\ \lfloor \frac{i-3}{2} \rfloor + 1, & \text{если } (i-3) \bmod 2 \neq 0. \end{cases} \quad (2)$$

Для исхода $b_{21})$ характерно то, что на первом шаге действовала помеха положительной полярности, на втором шаге проявление помехи отсутствовало, на третьем шаге алгоритма действовала помеха отрицательной полярности (помеха также действует на нечетных шагах

алгоритма). Действуя непомехоустойчивым алгоритмом на четных шагах алгоритма, разобьем полукрытый интервал неопределенности $[0, x_1^3)$ на $\Phi_{2,8}(\alpha, 1)$ равных частей. Поскольку

$$l([0, 1]) = l([0, x_1^3]) + l([x_1^3, 1]), \quad (3)$$

где $l([a, b])$ – длина отрезка $[a, b]$, то для функции $\Psi_{2,8}^{a,1,1}(i,1)$ будет справедливо соотношение:

$$\Psi_{2,8}^{a,1,1}(i,1) = 2\Phi_{2,8}(\alpha, 1). \quad (4)$$

Для исходов $a_{22})$ и $b_{22})$ характерно то, что помеха действует на четных шагах алгоритма (в первом случае на втором шаге действует помеха положительной полярности, во втором случае – помеха отрицательной полярности). Действуя на последующих нечетных шагах непомехоустойчивым алгоритмом, разобьем соответственно полукрытые интервалы неопределенности $[0, x_1^3)$, $[x_1^3, 1)$ на $\Phi_{2,8}(\alpha_1, 1)$, где $\Phi_{2,8}(\alpha_1, 1) = (k+1)^{\alpha_1}$;

$$\alpha_1 = \begin{cases} \frac{i-3}{2}, & \text{если } (i-3) \bmod 2 = 0; \\ \lfloor \frac{i-3}{2} \rfloor, & \text{если } (i-3) \bmod 2 \neq 0. \end{cases} \quad (5)$$

С учетом соотношения (3) будем иметь:

$$\Psi_{2,8}^{a,1,1}(i,1) = 2\Phi_{2,8}(\alpha_1, 1). \quad (6)$$

Итак, для функции $\Psi_{2,8}^{a,1,1}(i,1)$ справедливы соотношения (1), (4), (6).

Исходя из минимаксного критерия, устанавливаем:

$$\Psi_{2,8}^{a,1,1}(i,1) = 2 \min \{ \Psi_{2,8}^{a,1,1}(i-2,1), \Phi_{2,8}(\alpha, 1), \Phi_{2,8}(\alpha_1, 1) \}.$$

Проведенный логический анализ позволяет сформулировать следующий алгоритм поиска:

1-й шаг: первый эксперимент совершить в точке

$$x_1^1 = \delta \cdot \min \{ \Psi_{2,8}^{a,1,1}(i-2,1), \Phi_{2,8}(\alpha, 1), \Phi_{2,8}(\alpha_1, 1) \}.$$

2-й шаг: применить принцип «повторных сравнений»: $x_1^2 = x_1^1$. Если при этом возникает исход $a_1)$ либо $b_1)$, то соответственно на отрезках $[0, x_1^2]$, $[x_1^2, 1]$; действует оптимальный $(i-2)$ -шаговый алгоритм, если возникает исход $a_2)$ либо $b_2)$, выполнить третий шаг.

3-й шаг: применить принцип «повторных сравнений»: $x_1^3 = x_1^2$, если при этом возникает исход $a_{21})$ либо $b_{21})$, то применить соответственно в полукрытых интервалах $[0, x_1^3]$, $[x_1^3, 1]$; на четных шагах непомехоустойчивый алгоритм, посредством которого они будут разбиты на $\Phi_{2,8}(\alpha, 1)$ равные части;

если возникает исход $a_{22})$ либо $b_{22})$, то применить соответственно в полукрытых интервалах $[0, x_1^3]$, $[x_1^3, 1)$ на нечетных шагах алгоритма непомехоустойчивый алгоритм, посредством которого они будут разбиты на $\Phi_{2,8}(\alpha_1, 1)$ равные части.

Рассмотрим другой характерный случай, для которого $H > 1, \ell = 1, k = 1$.

В результате выполнения первого шага алгоритма может возникнуть один из исходов $a)$ либо $b)$. Как и ранее, на втором шаге алгоритма применяем принцип «повторных сравнений». При этом может возникнуть один из исходов: $a_1)$ либо $a_2)$, $b_1)$ либо $b_2)$. Для исходов $a_1)$ и $b_1)$ устанавливаем истинность соотношения (1). Для исходов $a_2)$ и $b_2)$ на третьем шаге применяем принцип «повторных сравнений». В результате совершения третьего шага алгоритма может возникнуть один из исходов $a_{21})$ либо $a_{22})$, $b_{21})$ либо $b_{22})$. Если возникает исход $a_{21})$ либо $b_{21})$, то для них характерно действие помехи на первом шаге алгоритма, на последующих H шагах такого действия не будет наблюдаться. Следующее ее проявление будет на $(H+2)$ -м шаге алгоритма. При дальнейшем планировании шагов алгоритма необходимо после выполнения первых трех шагов алгоритма применить на последующих $(H-2)$ шагах алгоритма непомехоустойчивый алгоритм; затем на последующей серии шагов, состоящей из $(H+1)$ шага, пропустить один шаг, затем снова на последующих H шагах применить непомехоустойчивый алгоритм и т. д.

Следовательно, первоначально за $(H-2)$ шага исходный интервал будет разбит на $(k+1)^{H-2}$ равные части. Затем каждый из этих отрезков на последующих $(H+1)$ шагах будет разбит на $(k+1)^H$ равные части. Таких совокупностей шагов, состоящих из $(H+1)$ шага, будет α_1 , где

$$\alpha_1 = \begin{cases} \frac{i-H-1}{H+1}, & \text{если } (i-H-1) \bmod (H+1) = 0; \\ \lfloor \frac{i-H-1}{H+1} \rfloor, & \text{если } (i-H-1) \bmod (H+1) \neq 0. \end{cases}$$

В том случае, когда величина $(i-H-1)$ делится на $(H+1)$ с остатком, на α_2 шагах алгоритма снова применяют непомехоустойчивый алгоритм, где

$$\alpha_2 = \begin{cases} 0, & \text{если } (i-H-1) \bmod (H+1) = 0; \\ ((i-H-1) \bmod (H+1)) - 1, & \text{если } (i-H-1) \bmod (H+1) \neq 0. \end{cases}$$

С учетом сказанного выделенные интервалы неопределенности $[x_1^3, 1]$, $[0, x_1^3]$ будут соответственно разбиты на $\Phi_{2,8}(\alpha_1, \alpha_2, 1)$ равные части, где

$$\Phi_{2,8}(\alpha_1, \alpha_2, 1) = (k+1)^{H-2} \cdot (k+1)^{H \cdot \alpha_1} \cdot (k+1)^{\alpha_2}. \quad (7)$$

Для исхода $a_{22})$ либо $b_{22})$ характерно действие помехи на втором шаге алгоритма, на третьем, четвертом и т. д. на $(H+2)$ -м шаге помеха будет отсутствовать. Ее проявление начнется снова на $(H+3)$ -м шаге алгоритма. Затем, начиная с $(H+4)$ -го шага и кончая $(2 \cdot H+3)$ -м помехи не будет. Новое ее проявление начнется с $(2 \cdot H+4)$ -го шага и т. д. Действуя на тех шагах, на которых отсутствует проявление помехи непомехоустойчивым алгоритмом, разобьем полукрытый интервал неопределенности $[0, x_1^3]$, $[x_1^3, 1]$ на $\Phi_{2,8}(\alpha_1, \alpha_2, k)$ равные части, где

$$\begin{aligned} \Phi_{2,8}(\alpha_3, \alpha_4, k) &= (k+1)^{H-1} \cdot (k+1)^{H \cdot \alpha_3} \cdot (k+1)^{\alpha_4} = \\ &= (k+1)^{H(1+\alpha_3)+\alpha_4-1}, \end{aligned} \quad (8)$$

$$\alpha_3 = \begin{cases} \frac{i-H-2}{H+1}, & \text{если } (i-H-2) \bmod (H+1) = 0; \\ \lfloor \frac{i-H-2}{H+1} \rfloor, & \text{если } (i-H-2) \bmod (H+1) \neq 0, \end{cases}$$

$$\alpha_4 = \begin{cases} 0, & \text{если } (i-H-2) \bmod (H+1) = 0; \\ ((i-H-2) \bmod (H+1)) - 1, & \text{если } (i-H-2) \bmod (H+1) \neq 0. \end{cases}$$

С учетом соотношений (1), (7), (8) для целевой функции алгоритма будет справедливо равенство:

$$\begin{aligned} \Psi_{2,8}^{a,1,H}(i, 1) &= \\ &= 2 \min \{ \Psi_{2,8}^{a,1,H}(i-2, 1), \Phi_{2,8}(\alpha_1, \alpha_2, 1), \Phi_{2,8}(\alpha_3, \alpha_4, 1) \}. \end{aligned} \quad (9)$$

Проведенный анализ исходов, возникающих при поиске точки с характерным признаком, позволяет установить такой алгоритм:

1-й шаг: эксперимент совершить в точке:

$$x_1^1 = \delta \cdot \min \{ \Psi_{2,8}^{a,1,H}(i-2, 1), \Phi_{2,8}(\alpha_1, \alpha_2, 1), \Phi_{2,8}(\alpha_3, \alpha_4, 1) \}.$$

2-й шаг: применить принцип «повторных сравнений»: $x_1^2 = x_1^1$.

Если при этом возникает исход $a_1)$ либо $b_1)$, то соответственно на отрезках $[0, x_1^2]$, $[x_1^2, 1]$ действует оптимальный $(i-2)$ -шаговый алгоритм.

Если возникает $a_2)$ либо $b_2)$, то выполняем третий шаг.

3-й шаг: применяем принцип «повторных сравнений»: $x_1^3 = x_1^2$.

Если при этом возникает исход $a_{21})$ либо $b_{21})$, то на отрезках $[x_1^3, 1]$, $[0, x_1^3]$ применим следующий непомехоустойчивый алгоритм: на последующих $(H-2)$ шагах алгоритма применяется непомехоустойчивый алгоритм, затем шаг пропускается и на следующих H шагах алгоритма снова применяем непомехоустойчивый алгоритм и т. д. до последнего шага алгоритма; если возникает исход $a_{21})$ либо $b_{21})$, то применяют такой алгоритм: на последующих $(H-1)$ шагах алгоритма применяют непомехоустойчивый алгоритм, затем шаг пропускается, затем на последующих H шагах снова применяют непомехоустойчивый алгоритм и т. д. до последнего, i -го шага алгоритма.

Рассмотрим более общую ситуацию, для которой характерно такое сочетание параметров алгоритма: $\ell > 1, H > 1, H > \ell, k = 1$.

Для этого случая, как и для ранее рассмотренных, после совершения первого шага алгоритма возникают исходы $a)$ и $b)$. На последующих ℓ шагах алгоритма применяется принцип «повторных сравнений». Самым худшим случаем будет тот случай, когда на $(\ell+1)$ -м первом шаге появляется исход типа $a_2)$ и $b_2)$. Проанализируем исход типа $b_2)$, для которого характерно то, что на первых ℓ шагах алгоритма смесь сигнала и помехи превосходит

координату точки x_1^1 , а на $(\ell + 1)$ -м шаге они по своей величине уменьшались. В дальнейшем на последующем $(\ell + 1)$ -м шаге применяют принцип «повторных сравнений». При этом на $2(\ell + 1)$ -м шаге могут возникнуть исходы:

$$x(t + 2\ell\Delta t) \in [0, x_1^{2\ell+1}); \quad x(t + 2\ell\Delta t) \in [x_1^{2\ell+1}, 1).$$

Для первого исхода характерно то, что помеха действовала на первых ℓ шагах алгоритма, а на последующих $(\ell + 1)$ -м, $(\ell + 2)$ -м, ..., $(H - \ell - 1)$ -м ее по условию не будет. Поскольку проявление помехи обнаружено, то необходимо применить такую комбинацию алгоритмов: на последующих $(H - \ell - 1)$ шагах алгоритма применить непомехоустойчивый алгоритм, затем пропустить последующие ℓ шагов алгоритма, потом снова применить непомехоустойчивый алгоритм и т. д. Такое планирование шагов алгоритма приведет к тому, что полуоткрытый интервал неопределенности $[0, x_1^1)$ будет разбит на $\varphi_{2,8}(\alpha_5, \alpha_6, k)$ равных частей, где $k = 1$;

$$\varphi_{2,8}(\alpha_5, \alpha_6, 1) = (k + 1)^{H-\ell-1} \cdot (k + 1)^{H \cdot \alpha_5} \cdot (k + 1)^{\alpha_6}; \quad (10)$$

$$\alpha_5 = \begin{cases} \frac{i-H-\ell}{H+\ell}, & \text{если } (i-H-\ell) \bmod (H+\ell) = 0; \\ \left] \frac{i-H-\ell}{H+\ell} \right[, & \text{если } (i-H-\ell) \bmod (H+\ell) \neq 0; \end{cases}$$

$$\alpha_6 = \begin{cases} 0, & \text{если } (i-H-\ell) \bmod (H+\ell) = 0 \\ \text{либо } (i-H-\ell) \bmod (H+\ell) \leq \ell; \\ ((i-H-\ell) \bmod (H+\ell)) - \ell, & \\ \text{если } (i-H-\ell) \bmod (H+\ell) > \ell. \end{cases}$$

Для второго исхода характерно то, что помеха действовала на $(\ell + 1)$, $(\ell + 2)$, ..., 2ℓ -м шагах алгоритма, а на $2(\ell + 1)$ -м шаге алгоритма помеха отсутствовала. Применяем в этом случае такую стратегию: на последующих $(H - 1)$ шагах используем непомехоустойчивый алгоритм, затем пропускаем ℓ шагов, потом снова на H шагах алгоритма применяем непомехоустойчивый алгоритм и т. д. Посредством такой процедуры разобьем полуоткрытый интервал неопределенности на $\varphi_{2,8}(\alpha_7, \alpha_8, k)$ равные части, где $k = 1$;

$$\varphi_{2,8}(\alpha_7, \alpha_8, 1) = (k + 1)^{H-1} \cdot (k + 1)^{H \cdot \alpha_7} \cdot (k + 1)^{\alpha_8}; \quad (11)$$

$$\alpha_7 = \begin{cases} \frac{i-H-2\ell}{H+\ell}, & \text{если } (i-H-2\ell) \bmod (H+\ell) = 0; \\ \left] \frac{i-H-2\ell}{H+\ell} \right[, & \text{если } (i-H-2\ell) \bmod (H+\ell) \neq 0; \end{cases}$$

$$\alpha_8 = \begin{cases} 0, & \text{если } (i-H-2\ell) \bmod (H+\ell) = 0 \\ \text{либо } (i-H-2\ell) \bmod (H+\ell) \leq \ell; \\ ((i-H-2\ell) \bmod (H+\ell)) - \ell, & \\ \text{если } (i-H-2\ell) \bmod (H+\ell) > \ell. \end{cases}$$

Аналогично можно показать, что для исхода типа $a_2)$ в результате совершения первых $(2\ell + 1)$ шагов алгоритма интервал неопределенности $[x_1^{2\ell+1}, 1)$, $[0, x_1^{2\ell+1}]$ будет со-

ответственно разбит на $\varphi_{2,8}(\alpha_5, \alpha_6, 1)$, $\varphi_{2,8}(\alpha_7, \alpha_8, 1)$ равные части.

Если в результате выполнения первых ℓ шагов алгоритма результаты эксперимента совпадают, то для исхода типа $a)$ и для исхода типа $b)$ на последующих шагах алгоритма применить $(i - \ell - 1)$ -шаговый помехоустойчивый алгоритм. Каждый из этих полуоткрытых интервалов неопределенности будет разбит на $\Psi_{2,8}^{a,\ell,H}(i - \ell - 1, 1)$ равные части.

Проведенный анализ возможных исходов, возникающих для рассмотренного случая, позволяет записать соотношение для целевой функции алгоритма:

$$\Psi_{2,8}^{a,\ell,H}(i, 1) = 2 \min \{ \Psi_{2,8}^{a,\ell,H}(i - \ell - 1, 1), \varphi_{2,8}(\alpha_5, \alpha_6, 1), \varphi_{2,8}(\alpha_7, \alpha_8, 1) \}. \quad (12)$$

Обобщим соотношения (10), (11) для любого количества точек эксперимента. Пусть некоторым образом выбраны k точек первого эксперимента и выполнен первый шаг алгоритма. Тогда возможен один из исходов:

$$c) \quad x(t_1) \in [x_q^1, x_{q+1}^1], \quad q = \overline{0, k}, \quad x_0^1 = 0; \quad x_{k+1}^1 = 1.$$

На втором шаге алгоритма применяем смешанную стратегию алгоритма $x_1^2 = x_q^1$; $x_{q_2}^2 \in (x_q^1, x_{q+1}^1)$; $x_k^2 = x_{q+1}^1$; $q_2 = \overline{2, k-1}$. В результате совершения второго шага алгоритма возможен один из исходов:

$$c_1) \quad x(t + \Delta t) \in [x_q^{1,1}, x_q^{1,1}];$$

$$c_2) \quad x(t + \Delta t) \in [x_{q_1}^2, x_{q_1+1}^2];$$

$$c_3) \quad x(t_1 + \Delta t) \in [x_k^2, x_{q+1}^{1,2}], \quad q_1 = \overline{1, k-1}.$$

Для исхода $c_1)$ возникшее противоречие свидетельствует о действии помехи на первом либо втором шаге. В этой ситуации точки эксперимента размещают на третьем шаге следующим образом: все точки эксперимента размещают равномерно в интервале $(x_q^{1,1}, x_q^{1,1})$. Заметим, что если помеха действовала на первом шаге, то новое её проявление будет через H шагов алгоритма; если помеха начала действовать на втором шаге алгоритма, то её действие продолжается ℓ шагов алгоритма, только на $(\ell + 2)$ -м шаге выяснится, на каком шаге действовала помеха. По этой причине четвертый, пятый, ... $(\ell + 1)$ -й шаг алгоритма выполняют таким же образом как и третий шаг: равномерно размещают точки эксперимента в выделенном интервале неопределенности относительно точки с характерным признаком. Следовательно, интервал неопределенности $(x_q^{1,1}, x_q^{1,1})$ за $(\ell - 1)$ -й шаг разобьется на $(k + 1)^{\ell-1}$ равных частей. Следующий $(\ell + 2)$ -й шаг необходимо планировать таким образом: $(k - 1)$ точку разместить равномерно в выделенном на $(\ell + 1)$ -м шаге алгоритма интервале неопределенности; последнюю точку – в точке x_q^1 .

При выполнении $(\ell + 2)$ -го шага алгоритма важную информацию несет последняя точка эксперимента. Причем, если $x(t + (\ell + 1)\Delta t) < x_q^1$, то помеха действовала на первом шаге; если $x(t + (\ell + 1)\Delta t) > x_q^1$, то помеха дей-

ствовала начиная со второго шага до $(\ell + 1)$ -го. Для первого случая, как это известно из ранее рассмотренных алгоритмов, помеха будет отсутствовать $(H - \ell - 1)$ шаг, затем проявится снова на следующих ℓ шагах алгоритма и т. д.

Для второго случая помеха будет отсутствовать $(H - 1)$ шаг, затем снова проявится на последующих ℓ шагах алгоритма, затем на последующих H шагах снова будет отсутствовать и т. д.

В первом случае интервал неопределенности $(x_q^{1,1}, x_q^1)$ с учетом соотношения (10) будет разбит на

$$\Phi_{2,8}^*(i-1, k) = k(k+1)^{\ell-1} \cdot \Phi_{2,8}(\alpha_5, \alpha_6, k) \quad (13)$$

равные части. Во втором случае исходный интервал неопределенности (x_q^1, x_{q+1}^1) – на

$$\Phi_{2,8}^{**}(i-1, k) = k(k+1)^{\ell-1} \cdot \Phi_{2,8}(\alpha_7, \alpha_8, k). \quad (14)$$

Если возникает исход c_3 , то поступают таким же образом, как и для исхода c_1 с тем отличием, что интервалом неопределенности является интервал $(x_{q+1}^1, x_{q+1}^{1,2})$: на последующих $(\ell - 1)$ шагах равномерно размещают точки эксперимента в выделенном на предыдущем шаге интервале неопределенности; $(\ell + 2)$ -й шаг планируют таким образом: первую точку размещают в точке x_{q+1}^1 , а остальные – равномерно в интервале неопределенности выделенном на $(\ell + 2)$ -м шаге алгоритма.

При выполнении $(\ell + 2)$ -го шага может возникнуть один из исходов:

$$x(t_1 + (\ell + 1)\Delta t) > x_{q+1}^1, \quad x(t_1 + (\ell + 1)\Delta t) < x_{q+1}^1.$$

Для первого исхода характерно действие помехи на первом шаге и по этой причине интервал неопределенности $(x_{q+1}^1, x_{q+1}^{1,2})$ разбивается на $\Phi_{2,8}^*(i-1, k)$ равные части (см. соотношение (13)).

Для второго исхода характерно действие помехи на втором, третьем, $(\ell + 1)$ -м шагах алгоритма, поэтому интервал неопределенности $(x_{q+1}^1, x_{q+1}^{1,2})$ будет разбит (см. решение исхода c_1) на $\Phi_{2,8}^{**}(i-1, k)$ равные части, а исходный интервал – на $(k + 1) \cdot \Phi_{2,8}^{**}(i-1, k)$ равные части.

Если возникает исход c_2 , то его решают таким же образом, как и исход c .

Пусть исход c_2 появляется на первом, втором, ..., j -м шагах алгоритма, в результате которого выделяется полуоткрытый интервал неопределенности $[x_{q_3}^j, x_{q_3+1}^j)$, $q_3 = \overline{0, k}$. При планировании $(j + 1)$ -го шага применяется стратегия исхода c , в результате ее выполнения возникают исходы

$$d_1) \quad x(t + j\Delta t) \in [x_{q_2}^{j,1}, x_{q_2+1}^j);$$

$$d_2) \quad x(t + j\Delta t) \in [x_{q_4}^{j+1}, x_{q_4+1}^{j+1});$$

$$d_3) \quad x(t + j\Delta t) \in [x_{q_2+1}^j, x_{q_2+1}^{j,2}), \quad q_4 = \overline{2, k-1}.$$

Исход d_1 разрешается таким же образом, как исход c_1 . При этом, если при выполнении $((j + 1) + \ell)$ -го шага

алгоритма возникает исход $x(t + (j + \ell)\Delta t) < x_{q_2}^j$, то на первых j шагах алгоритма действовала помеха. Поскольку помеха обнаружена, то полуоткрытый интервал неопределенности $[x_{q_2}^{j,1}, x_{q_2}^j)$ будет разбит на $\Phi_{2,8}(i-j, k)$ равных частей, где

$$\begin{aligned} \bar{\Phi}_{2,8}(i-j, k) &= \\ &= (k+1)^{\ell-1} (k+1)^{H-\ell-1} \cdot (k+1)^{H \cdot \bar{\alpha}_1} \cdot (k+1)^{H \cdot \bar{\alpha}_2}; \quad (15) \end{aligned}$$

$$\bar{\alpha}_1 = \begin{cases} \frac{i-j-H}{H+\ell}, & \text{если } (i-H-j) \bmod (H+\ell) = 0; \\ \left] \frac{i-j-H}{H+\ell} \right[, & \text{если } (i-H-j) \bmod (H+\ell) \neq 0; \end{cases}$$

$$\bar{\alpha}_2 = \begin{cases} 0, & \text{если } (i-H-j) \bmod (H+\ell) = 0 \\ \text{либо } (i-H-j) \bmod (H+\ell) \leq \ell; \\ ((i-H-j) \bmod (H+\ell)) - \ell, & \\ \text{если } (i-H-j) \bmod (H+\ell) > \ell. \end{cases}$$

Если же при выполнении $(j + \ell + 1)$ -го шага алгоритма возникает исход $x(t_1 + (j + \ell)\Delta t) < x_{q_3}^1$, то, как это уже было показано, помеха действовала на $(j + 1) \dots (j + \ell)$ -м шаге алгоритма и исходный интервал неопределенности $[0, 1]$ будет разбит на $\bar{\Phi}_{2,8}(i-j, k)$ равные части, где

$$\begin{aligned} \bar{\Phi}_{2,8}(i-j, k) &= (k+1) \cdot (k+1)^{j-1} (k+1)^{H-1} \times \\ &\times (k+1)^{H \cdot \bar{\alpha}_3} \cdot (k+1)^{\bar{\alpha}_4}; \quad (16) \end{aligned}$$

$$\bar{\alpha}_3 = \begin{cases} \frac{i-j-\ell-H}{H+\ell}, & \text{если } (i-j-\ell-H) \bmod (H+\ell) = 0; \\ \left] \frac{i-j-\ell-H}{H+\ell} \right[, & \text{если } (i-j-\ell-H) \bmod (H+\ell) \neq 0; \end{cases}$$

$$\bar{\alpha}_4 = \begin{cases} 0, & \text{если } (i-H-j-\ell) \bmod (H+\ell) = 0 \\ \text{либо } (i-H-j-\ell) \bmod (H+\ell) \leq \ell; \\ ((i-H-j-\ell) \bmod (H+\ell)) - \ell, & \\ \text{если } (i-j-H-\ell) \bmod (H+\ell) > \ell. \end{cases}$$

В том случае, когда $(j + 1) = (\ell + 1)$ и возникает исход d_2 , который будет свидетельствовать истинность соотношения $x \in [x_q^1, x_{q+1}^1)$. В этом случае на последующих шагах алгоритма применяют смешанную стратегию исхода c .

Для целевой функции алгоритма поиска с характерным признаком справедливо такое соотношение:

$$1/\Psi_{2,8}^{a,\ell,H}(i, k) = \left\{ \frac{I[x_q^{1,1}, x_q^1]}{\Phi_{2,8}^*(i-1, k)}, \frac{I[x_{q_2}^{j,1}, x_{q_2}^1]}{\bar{\Phi}_{2,8}(i-j, k)}, \frac{1}{\bar{\Phi}_{2,8}(i-j, k)} \right\}. \quad (17)$$

Проведенный анализ исходов, возникающих при поиске точки с характерным признаком, позволяет сформулировать следующий алгоритм:

1-й шаг: точки эксперимента выбрать исходя из соотношения:

$$x_q^1 = \delta q \Phi_{2,8}(\alpha_7, \alpha_8, k), \quad q = \overline{1, k}.$$

2-й шаг: применить смешанную стратегию c).

3-й шаг: если на втором шаге алгоритма возникает исход c_1) либо c_3), то применить к выделенным полуоткрытым интервалам ранее рассмотренные стратегии непомехоустойчивых алгоритмов; если возник исход c_2), то переходим на четвертый шаг.

4-й шаг: если на третьем шаге возникает исход типа c_1) либо c_3), то применить описанные стратегии непомехоустойчивых алгоритмов. Если возник исход c_2), то применить смешанную стратегию к выделенному интервалу неопределенности и перейти к пятому шагу и т. д.

ЗАКЛЮЧЕНИЕ

Итак, нами показано, что схемы алгоритмов помехоустойчивого поиска точки с характерным признаком, правила выбора стратегии поиска (распределения точек эксперимента) и выделения нового интервала неопределенности позволяют методом индукции решить задачу синтеза помехоустойчивых алгоритмов поиска для любых значений параметров регулярного воздействия и параметров алгоритма и, тем самым, задать функционирование нового класса конечных автоматов с псевдослучайными переходами, являющихся генераторами шифров замены для символов входного алфавита.

УДК 519.242

Ю. С. Афонин, В. И. Дубровин

МОДЕЛИРОВАНИЕ СОСТАВОВ СМЕСЕЙ МЕТОДОМ СИМПЛЕКСНЫХ РЕШЕТОК

Рассмотрена проблема проектирования и оптимизации многокомпонентных смесей. Исследована возможность применения для решения данной задачи метода симплексных решеток.

ВВЕДЕНИЕ

Одной из актуальных технологических проблем является проектирование составов различных смесей, а также необходимость предсказания их характеристик и свойств, получаемых в результате смешения различных составляющих. Как правило, измеряемые характеристики зависят не от количества смеси, а исключительно от пропорций содержащихся в них ингредиентов [1–4]. При этом объектом исследования могут являться не только количественные характеристики, но и качественные. В качестве примера можно рассмотреть задачу о проектировании бетонных смесей.

МЕТОД СИМПЛЕКСНЫХ РЕШЕТОК

В данной работе рассматривается возможность компьютерного моделирования в задачах оптимизации составов смесей с применением симплекс-решетчатых планов.

ПЕРЕЧЕНЬ ССЫЛОК

1. Алипов Н.В. Дискретные автоматы с псевдослучайными переходами и подстановочные методы защиты информации на их основе // Радиоэлектроника и информатика. 2001. № 4. С. 95–98
2. Алипов Н.В. Помехоустойчивый поиск точки с характерным признаком и кодирование информации // Радиоэлектроника и информатика. 2000. № 4. С. 82–86.
3. Алипов Н.В. Охалкин А.А., Ребезюк Л.Н. Защита информации в дискретном канале на основе устойчивых к периодическим помехам алгоритмов поиска точки с характерным признаком // АСУ и приборы автоматики. 1999. Вып. 109. С. 108–115.

Надійшла 03.05.2004
Після доробки 19.11.2004

Будуються завадостійкі до симетричних регулярних віртуальних послідовних перешкод алгоритми пошуку точки з характерною ознакою. Такі алгоритми задають функціонування дискретних автоматів із псевдовипадковими переходами з одного стану в інше. Подібні дискретні автомати використовуються в системах захисту інформації для генерації шифру захисту.

Built antinoise to symmetrical regular virtual consequent hindrances algorithms of searching spots with the distinctive sign. Such algorithms will assign an operation of discrete automats with pseudorandom transition from one condition in the another. Similar discrete automats are used in systems of protection information for generations of cipher of protection.

Компьютерное моделирование свойств веществ позволяет решить большой комплекс задач по планированию эксперимента, получению и обработке опытных данных, а также провести оптимизацию по составу многокомпонентных смесей, что и является основной целью этой работы.

Преимущество симплекс-решетчатых планов состоит в том, что, располагая результатами эксперимента, можно предсказать значение свойства для многокомпонентной смеси любого состава.

Основным условием для применения симплексного метода в задачах оптимизации составов смесей является то, что содержание всех компонентов в любой смеси составляет 100%, что можно продемонстрировать в виде [5]:

$$\sum_{i=1}^p x_i = 1. \quad (1)$$

Из ограничения (1) следует, что ковариационная матрица оказывается вырожденной, если в матрицу независимых переменных включить столбец, состоящий из единиц (свободный член модели). Эта же матрица является вырожденной, если в матрицу факторов включить квадраты и парные произведения факторов ($x_i x_j$, x_i^2).

С другой стороны, поверхность отклика в подобных задачах имеет сложную конфигурацию, поэтому для ее описания является целесообразным использование моделей высоких порядков.

Однако если прибегнуть к некоторым преобразованиям независимых переменных, то ограничение (1) можно превратить в преимущество. Рассмотрим это на примере квадратичной формы. Для трехкомпонентной задачи полная квадратичная форма имеет вид:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2. \quad (2)$$

Произведя преобразования, получим следующий вид модели (2):

$$y = b''_1x_1 + b''_2x_2 + b''_3x_3 + b''_{12}x_1x_2 + b''_{13}x_1x_3 + b''_{23}x_2x_3, \quad (3)$$

где:

$$\begin{aligned} b''_1 &= b_0 + b_1 + b_{11}; & b''_2 &= b_0 + b_2 + b_{22}; \\ b''_3 &= b_0 + b_3 + b_{33}; & b''_{12} &= b_{12} - b_{11} - b_{22}; \\ b''_{13} &= b_{13} - b_{11} - b_{33}; & b''_{23} &= b_{23} - b_{22} - b_{33}. \end{aligned}$$

Для модели (3) ковариационная матрица не является вырожденной.

В общем случае для p -компонентов модели выглядят следующим образом:

1. Квадратичная модель

$$y = \sum_{1 \leq i \leq p} \beta_i x_i + \sum_{1 \leq i < j \leq p} \beta_{ij} x_i x_j; \quad (4)$$

2. Неполная кубическая модель

$$y = \sum_{1 \leq i \leq p} \beta_i x_i + \sum_{1 \leq i < j \leq p} \beta_{ij} x_i x_j + \sum_{1 \leq i < j < k \leq p} \beta_{ijk} x_i x_j x_k. \quad (5)$$

Для трехкомпонентных смесей диаграммы "состав-свойство" [6–7] представляют собой сеть изолиний на треугольнике концентраций. Для построения диаграмм требуется выполнить большой объем экспериментальных исследований. Например, для шага 5% при изучении трехкомпонентной смеси требуется провести 210 опытов. С целью уменьшения числа опытов при построении диаграмм было применено планирование эксперимента в виде симплекс-решетчатых планов, которые сокращают число опытов до семи на каждый план.

Экспериментальные точки располагаются на симплексах или симплексных решетках. В работе [8] предлагается восемь различных моделей: от линейной до модели четвертой степени. Выбор каждой модели определяется ее сложностью (наличием экстремумов функции – минимумов или максимумов и перегибов). Модели представлены на рис. 1. Модель 1 описывает плоскость, модели 2 и 3 позволяют учесть взаимное влияние факторов и получить неполные модели второго порядка с перегибами. Модели второго порядка, модель 4, позволяют описывать функции с одним экстремумом, модели

5 и 6 являются переходными к моделям третьего порядка и описывают функции с перегибами и несколькими экстремумами. Модель третьего порядка, модель 7, способна описать поверхность с помощью полинома третьего порядка, а модель четвертого порядка, модель 8, соответственно, полинома четвертого порядка, которая включает в себя экстремумы, седла и перегибы.

В таблице 1 представлено количество опытов в зависимости от числа компонентов смеси и степени полинома, описывающего поверхность отклика. Коэффициенты уравнений регрессии вычисляются по результатам экспериментов.

Таблица 1 – Количество опытов для построения моделей разных степеней

Число компонентов	Степень полинома			
	вторая	третья неполная	третья	четвертая
2	2	3	4	5
3	6	7	10	15
4	10	14	20	35
5	15	25	35	70
6	21	41	56	126
8	36	92	120	330
1	55	175	220	715

ПРОВЕДЕНИЕ ЭКСПЕРИМЕНТА И ПОСТРОЕНИЕ МОДЕЛИ

В работе была поставлена задача решить проблему проектирования составов бетона с заданными свойствами, определяющими как качество бетона и конструкций на его основе, так и технико-экономическую эффективность их изготовления.

Одним из важных свойств бетона в момент его производства является подвижность. Подвижность бетонной смеси определяет степень его удобоукладываемости [9], т. е. насколько полно будет заполнена та или иная форма, при помощи которой задаются геометрические параметры изделия.

Получение желаемой подвижности бетона при его производстве можно добиться разными путями: изменением соотношения песок/щебень/"цементное молоко" (под "цементным молоком" понимается смесь цемента и воды), добавлением различных добавок, изменением фракционности щебня или использование щебней разных фракций.

В данной работе рассматривается лишь один способ прогнозирования изменения подвижности, а именно изменение соотношения песок/щебень/"цементное молоко".

Выбор исследования именно этого фактора, влияющего на подвижность цементной смеси, обусловлен тем, что эти три составляющие (песок, щебень и "цементное молоко") являются основой любого бетона. Прежде чем искать более дорогие методики улучшения данной характеристики, наиболее целесообразно будет попытаться добиться максимального эффекта, используя только стандартные ингредиенты, а также их оптимальный состав без применения дорогостоящих добавок и использования различных щебней по фракциям.


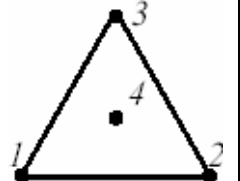

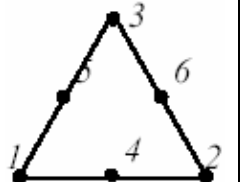
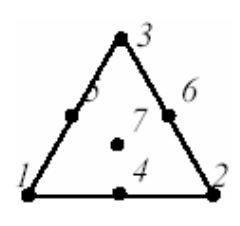
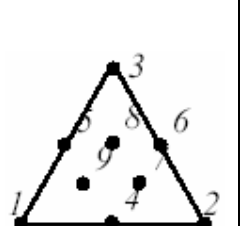
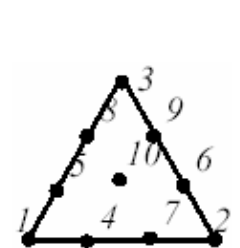
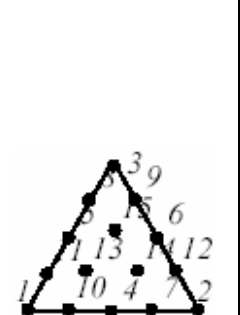
Модель 1 – модель первого порядка				Модель 2 – модель первого порядка с центральной точкой внутри решетки					
№	Доли компонентов				№	Доли компонентов			
1	1	0	0		1	1	0	0	
2	0	1	0		2	0	1	0	
3	0	0	1		3	0	0	1	
$Y = \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3$				$Y = \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \beta_{123} \cdot X_1 \cdot X_2 \cdot X_3$					
Модель 3 – модель первого порядка с тремя точками внутри решетки				Модель 4 – модель второго порядка					
№	Доли компонентов				№	Доли компонентов			
1	1	0	0		1	1	0	0	
2	0	1	0		2	0	1	0	
3	0	0	1		3	0	0	1	
4	0,5	0,25	0,25		4	0,5	0,5	0	
5	0,25	0,5	0,25		5	0,5	0	0,5	
6	0,25	0,25	0,5	6	0	0,5	0,5		
$Y = \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \beta_{123} \cdot X_1^2 \cdot X_2 \cdot X_3 + \beta_{1223} \cdot X_1 \cdot X_2^2 \cdot X_3 + \beta_{1233} \cdot X_1 \cdot X_2 \cdot X_3^2$				$Y = \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \beta_{12} \cdot X_1 \cdot X_2 + \beta_{13} \cdot X_1 \cdot X_3 + \beta_{23} \cdot X_2 \cdot X_3$					
Модель 5 – модель второго порядка с центральной точкой внутри решетки				Модель 6 – модель второго порядка с тремя точками внутри решетки					
№	Доли компонентов				№	Доли компонентов			
1	1	0	0		1	1	0	0	
2	0	1	0		2	0	1	0	
3	0	0	1		3	0	0	1	
4	0,5	0,5	0		4	0,5	0,5	0	
5	0,5	0	0,5		5	0,5	0	0,5	
6	0	0,5	0,5		6	0	0,5	0,5	
7	0,333	0,333	0,334	7	0,5	0,25	0,25		
$Y = \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \beta_{12} \cdot X_1 \cdot X_2 + \beta_{13} \cdot X_1 \cdot X_3 + \beta_{23} \cdot X_2 \cdot X_3 + \beta_{123} \cdot X_1 \cdot X_2 \cdot X_3$				$Y = \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \beta_{12} \cdot X_1 \cdot X_2 + \beta_{13} \cdot X_1 \cdot X_3 + \beta_{23} \cdot X_2 \cdot X_3 + \beta_{1123} \cdot X_1^2 \cdot X_2 \cdot X_3 + \beta_{1223} \cdot X_1 \cdot X_2^2 \cdot X_3 + \beta_{1233} \cdot X_1 \cdot X_2 \cdot X_3^2$					
Модель 7 – модель третьего порядка				Модель 8 – модель четвертого порядка					
№	Доли компонентов				№	Доли компонентов			
1	1	0	0		1	1	0	0	
2	0	1	0		2	0	1	0	
3	0	0	1		3	0	0	1	
4	0,333	0,667	0		4	0,5	0,5	0	
5	0,333	0	0,667		5	0,5	0	0,5	
6	0	0,333	0,667		6	0	0,5	0,5	
7	0,666	0,334	0		7	0,25	0,75	0	
8	0,666	0	0,334		8	0,25	0	0,75	
9	0	0,666	0,334		9	0	0,25	0,75	
10	0,333	0,333	0,334	10	0,75	0,25	0		
$Y = \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \beta_{12} \cdot X_1 \cdot X_2 + \beta_{13} \cdot X_1 \cdot X_3 + \beta_{23} \cdot X_2 \cdot X_3 + \gamma_{12} \cdot X_1 \cdot X_2 \cdot (X_1 - X_2) + \gamma_{13} \cdot X_1 \cdot X_3 \cdot (X_1 - X_3) + \gamma_{23} \cdot X_2 \cdot X_3 \cdot (X_2 - X_3) + \beta_{123} \cdot X_1 \cdot X_2 \cdot X_3$				$Y = \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \beta_{12} \cdot X_1 \cdot X_2 + \beta_{13} \cdot X_1 \cdot X_3 + \beta_{23} \cdot X_2 \cdot X_3 + \gamma_{12} \cdot X_1 \cdot X_2 \cdot (X_1 - X_2) + \gamma_{13} \cdot X_1 \cdot X_3 \cdot (X_1 - X_3) + \gamma_{23} \cdot X_2 \cdot X_3 \cdot (X_2 - X_3) + \delta_{12} \cdot X_1 \cdot X_2 \cdot (X_1 - X_2)^2 + \delta_{13} \cdot X_1 \cdot X_3 \cdot (X_1 - X_3)^2 + \delta_{23} \cdot X_2 \cdot X_3 \cdot (X_2 - X_3)^2 + \beta_{1123} \cdot X_1^2 \cdot X_2 \cdot X_3 + \beta_{1223} \cdot X_1 \cdot X_2^2 \cdot X_3 + \beta_{1233} \cdot X_1 \cdot X_2 \cdot X_3^2$					

Рисунок 1 – Планы экспериментов

Для исследования были выбраны такие составляющие: цементно-водная смесь – "цементное молоко", песок речной и щебень одной фракционности. Поскольку в данной работе рассматривается только подвижность бетона и не рассматриваются его прочностные характеристики, на которые большое влияние оказывает цементно-водное соотношение, то при проведении экспериментов количество и соотношение сухого цемента и воды были выбраны таким образом, чтобы использование минимального количества "цементного молока" не приводило к потере прочностных характеристик получаемого в результате бетона.

В результате такого выбора ингредиентов смеси получаем необходимое для использования метода симплексных решеток условие, что исследуемое свойство зависит только от состава смеси.

Для проведения опытов и построения модели было принято решение об использовании пятой модели, с помощью которой можно описать функцию, имеющую перегибы, а также несколько экстремумов.

Результаты проведения экспериментов представлены в таблице 2.

Таблица 2 – Матрица планирования эксперимента

Матрица планирования				Результаты
№	Z ₁	Z ₂	Z ₃	Y _{эксп}
1	1	0	0	16,32
2	0	1	0	5,1
3	0	0	1	2,23
4	0,5	0,5	0	20,05
5	0,5	0	0,5	7,1
6	0	0,5	0,5	3,31
7	0,333	0,333	0,334	10,2

В таблице 2 Z₁ – доли компонентов количества "цементного молока", Z₂ – доли компонентов количества песка, Z₃ – доли компонентов количества щебня, Y_{эксп} – подвижность экспериментально полученного бетона.

По результатам проведенных экспериментов была построена модель, коэффициенты уравнения которой представлены в таблице 3.

Таблица 3 – Расчетные данные (коэффициенты уравнения)

β_1	β_2	β_3	β_{12}	β_{13}	β_{23}	β_{123}
-0.9330	-1.1014	-0.0052	0.0027	5.8086e-004	7.1566e-004	-1.5932e-006

Все эксперименты, поставленные в рамках данной работы, проводились сертификационной строительной лабораторией ООО "Будиндустрия ЛТД".

Для проверки адекватности полученной модели были проведены дополнительные четыре эксперимента в проверочных точках, результат приведен в таблице 4.

Проведенные эксперименты позволили провести проверку адекватности полученной модели. Погрешность модели составила 2,9%, что является удовлетворительным показателем для данной сферы строительной индустрии.

Таблица 4 – Проведение проверочных экспериментов

Матрица планирования				Результаты	
№	Z ₁	Z ₂	Z ₃	Y _{эксп}	Y _{расч}
1	0,25	0,75	0	15,1	14,8
2	0,75	0,25	0	19,7	20,3
3	0,25	0	0,75	3,12	2,5
4	0,75	0	0,25	8,8	9,5

ВЫВОДЫ

Основным преимуществом метода, использованного в данной работе, является то, что для построения приемлемой модели некоторой смеси необходимо провести, сравнительно с другими методами, небольшое количество экспериментов, что значительно ускоряет построение самой модели, а также существенно снижает стоимость ее получения.

Располагая результатами эксперимента, спланированного с применением симплекс-решетчатых планов для трехкомпонентной системы, можно предсказать значение свойства для смеси любого состава с хорошей точностью.

Полученная математическая модель дает возможность не только прогнозировать свойства смеси, но и оптимизировать ее состав по желанию пользователя.

ПЕРЕЧЕНЬ ССЫЛОК

1. Дубровин В.И. Идентификация и оптимизация сложных технических процессов и объектов. – Запорожье: ЗГТУ, 1997. – 92 с.
2. Дубровин В.И., Зонов А.В., Харитонов Ф.Я. Оптимизация состава стеатитового материала. // Стекло и керамика, 1984. № 8. С. 23.
3. Дубровин В.И., Зонов А.В. Оптимизация технологического процесса изготовления установочной радиокерамики. // Теория и практика конструирования и обеспечения надежности и качества электронной аппаратуры и приборов: Тезисы докладов Всесоюзной научно-технической конференции. М.: Радио и связь, 1984. С. 72–73.
4. Дубровин В.И., Зонов А.В., Харитонов Ф.Я. Планирование эксперимента при разработке новых стеатитовых материалов // Научно-технический процесс в разработке и применении новых керамических материалов и изделий для электроники, в механизации и автоматизации технологических процессов и оборудования: Тезисы докладов Всесоюзного научно-технического совещания "Керамика-86". – М: Информэлектро, 1986. С. 1–2.
5. Львовский Е.Н. Статистические методы построения эмпирических формул: Учеб. пособие. – М.: Высшая школа, 1982. – 224 с.
6. Ахназарова С.Л., Кафаров В.В. Оптимизация эксперимента в химии и химической технологии. – М.: Высшая школа, 1978. – 319 с.
7. Курнаков Н.С. Собрание избранных работ. Т. 1. Л.: Химтеоретиздат, 1938. – 560 с.
8. Бондарь А.Г., Статюха Г.А. Планирование эксперимента в химической технологии (основные положения примеры и задачи). – Киев: Вища школа, 1976. – 184 с.
9. Дворкин О.Л. Многопараметрическое проектирование составов бетонов: Монография: – Ровно: РГТУ, 2001. – 121 с.

Надійшла 03.05.2004
Після доробки 19.11.2004

Розглянуто проблему проектування і оптимізації багатоконпонентних сумішей. Досліджено можливість застосування для вирішення даної задачі методу симплексних решіток.

The article considers design problem and optimization of multi-component composition. Investigate a question of use simplex lattice method as applied to the given problem.

А. И. Вершина, Г. Г. Киричек, Д. М. Пиза

МОДЕЛЬ ИНФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ УЧЕБНОГО ПРОЦЕССА УНИВЕРСИТЕТА

Построение модели учебного процесса университета тесно связано с информационным обеспечением. В статье рассматриваются возможность и особенности решения задачи построения модели информационно-образовательной системы университета, основанной на предположении, что вероятность получения некоторого объема информации в бесконечно малом промежутке времени из различных мест хранения пропорциональна размеру этого промежутка. Такой подход хорошо согласуется с моделью учебного процесса, основанной на этих же принципах.

ВВЕДЕНИЕ

В настоящее время уровень информатизации учебной и научной деятельности в университетах еще весьма низок, хотя практически все они имеют выход в мировую информационную сеть. Поэтому создание условий для перехода к новому уровню образования на основе информационных технологий является актуальной задачей дальнейшей информатизации университета [1].

На пути компьютеризации и внедрения новых технологий библиотека университета предоставляет условия для реализации права на информацию всем пользователям, формируя современную библиотечно-информационную базу, обеспечивающую доступ профессорско-преподавательскому составу, сотрудникам и студентам к библиотечным фондам и базам данных. Создается полнотекстовая учебно-методическая база данных: методические пособия и указания, литература и рекомендации, базы данных реферативных и электронных журналов по всем дисциплинам и направлениям научных исследований.

Необходимость обеспечения учебного процесса и научных исследований выше перечисленными информационными ресурсами определяет целесообразность внедрения в кратчайшие сроки новых методов предоставления информации всем участникам учебного процесса. В статье рассматривается одна из возможностей построения нижнего уровня информационно-образовательной системы университета, основанной на ее математической модели. В дальнейшем возможно использовать решение данной задачи для построения более высоких уровней модели информационной системы университета с интеграцией полученных наработок в единую систему.

ПОСТАНОВКА ЗАДАЧИ

В связи с тем, что создание и хранение информационных ресурсов дополняется обеспечением к ним удаленного доступа и, обращаясь к системе, пользователи получают не только ссылку на имеющийся доку-

мент, но и сам документ, быстрота доставки необходимой информации является важным элементом данного исследования [2].

Время получения и передачи определенного объема информации каждому из участников учебного процесса зависит от места хранения и скорости передачи этой информации, а также факторов, носящих случайный характер.

Построение современной эффективной системы информационного обеспечения решает задачи объединения традиционной и электронной форм предоставления информации, создания больших возможностей оперативно и полно удовлетворять потребности в ней, обеспечения принципиально нового уровня получения и обобщения знаний, их распространения и использования.

Первоочередной становится проблема адаптации участников учебного процесса к всевозрастающему потоку информации, свободному ориентированию в информационных массивах и умению быстро находить и использовать в своей работе все имеющиеся информационные ресурсы [3].

Решение затронутых выше проблем является трудоемкой задачей, требующей значительных затрат. Вопрос об очередности и целесообразности проведения тех или иных мероприятий требует создания такой математической модели, которая позволила бы количественно оценить влияние информационного обеспечения на учебный процесс. Попытки создания математической модели учебного процесса изложены в работах [4, 5]. Модель обучения основана на понятиях "интенсивности" обучения, условий, при которых качество обучения остается постоянным, что позволяет рассматривать обучение с позиции цепей Маркова. В рамках такой модели влияние систем информационного обеспечения проявляется в значениях параметров.

РЕШЕНИЕ ЗАДАЧИ

Простейшая модель получения информации основана на предположении, что вероятность $\Delta i(t)$ получения некоторого объема информации в бесконечно малом промежутке времени Δt будет пропорциональна величине этого промежутка. Это позволяет записать

$$\Delta i(t) = i(t + \Delta t) - i(t) = [1 - i(t)]\nu \Delta t, \quad (10)$$

где $i(t)$ – вероятность того, что за время t получен определенный объем информации; ν – коэффициент пропорциональности, который отражает интенсивность по-

лучения информации и определяет уровень эффективности систем информационного обеспечения.

Переходя к пределу $\Delta t \rightarrow 0$, получим дифференциальное уравнение

$$di(t) = [1 - i(t)]\nu dt, \quad (11)$$

решение которого имеет вид

$$i(t) = 1 - e^{-\nu t}. \quad (12)$$

Плотность распределения времени получения элемента нужной информации определяется выражением

$$p_i(t) = \frac{di(t)}{dt} = \nu e^{-\nu t}, \quad (13)$$

то есть подчиняется экспоненциальному закону.

Если получение определенного объема информации описывается экспоненциальным законом, то получение больших объемов информации, представляющих собой совокупность α объемов, описываемых экспоненциальным законом, приводит к гамма-распределению [6], то есть

$$p(t) = \frac{\nu^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\nu t}, \quad (14)$$

где $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ – гамма-функция Эйлера.

Для произвольного времени T , затраченного на получение определенного объема информации, вероятность $I(T)$ получения нужной информации определяется выражением

$$I(T) = \int_0^T p(t) dt = \frac{\nu^\alpha}{\Gamma(\alpha)} \int_0^T t^{\alpha-1} e^{-\nu t} dt. \quad (15)$$

Ожидаемое время получения информации равно

$$\bar{T} = \int_0^\infty t p(t) dt = \frac{\nu^\alpha}{\Gamma(\alpha)} \int_0^\infty t^\alpha e^{-\nu t} dt = \frac{\alpha}{\nu}. \quad (16)$$

Как правило, время на получение информации T_0 ограничено и оно пропорционально ожидаемому времени \bar{T} :

$$T_0 = \eta \bar{T} = \eta \frac{\alpha}{\nu}, \quad (17)$$

где η – коэффициент пропорциональности.

В этом случае имеем

$$I(T_0) = \int_0^{\frac{\eta \alpha}{\nu}} p(t) dt = \frac{\nu^\alpha}{\Gamma(\alpha)} \int_0^{\frac{\eta \alpha}{\nu}} t^{\alpha-1} e^{-\nu t} dt. \quad (18)$$

После внесения ν^α под знак интеграла, замены переменной $z = \nu * t$ и изменения пределов интегрирования получим

$$I(T_0) = \frac{1}{\Gamma(\alpha)} \int_0^{\eta \alpha} z^{\alpha-1} e^{-z} dz = I(\eta, \alpha). \quad (19)$$

Из этого следует, что вероятность получения информации в случае выделения времени пропорционально ожидаемому \bar{T} не зависит от значения ν .

Если за заданное время T_0 объем информации не получен, то необходимо дополнительное время T_d . Это время разумно также выделять пропорционально ожидаемому времени на получение информации. Если считать, что объем дополнительной информации в одинаковой мере затрагивает каждую порцию, то количество элементов α останется прежним, а изменится значение ν . В соответствии с полученным выражением (10), вероятность получения информации за дополнительное время не изменится.

Проведя выкладки аналогично выкладкам, приведенным в работе [4], получим выражение

$$T_d = \frac{\nu^\alpha}{\Gamma(\alpha)} T_0^\alpha \frac{1}{\nu} * e^{-\nu T_0} + \left(\frac{\alpha}{\nu} - T_0\right) [1 - I(T_0)]. \quad (20)$$

После подстановки T_0 из формулы (8), с учетом (10) получим

$$T_d = \frac{1}{\nu \Gamma(\alpha)} (\eta \alpha)^\alpha e^{-\eta \alpha} + (1 - \eta) \frac{\alpha}{\nu} [1 - I(\eta, \alpha)]. \quad (21)$$

Если считать, что дополнительное время выделяется также пропорционально ожидаемому времени, то отношение δ дополнительного времени на получение материала к времени, выделенному на первоначальный объем информации, равно

$$\delta = \frac{(\eta \alpha)^\alpha}{\Gamma(\alpha + 1)} e^{-\eta \alpha} + (1 - \eta) [1 - I(\eta, \alpha)]. \quad (22)$$

Из полученного выражения следует, что отношение дополнительного времени на обучение к первоначально выделяемому времени будет постоянной величиной, если считать, что $\eta = \text{const}$ и $\alpha = \text{const}$. В этих условиях процесс получения информации может описываться цепями Маркова.

В основе описанной модели лежит гипотеза о гамма-распределении времени получения информации. Проверка этой гипотезы в реальных условиях встречает определенные трудности.

Рассмотрим простейшие процессы получения информации участниками учебного процесса из двух возможных мест ее хранения (сеть Интернет, файл-сервер библиотеки). За основу возьмем поиск определенного объема информации при одинаковых условиях доступа к ней 60 участниками эксперимента с регистрацией времени ее получения. K -й участник данного эксперимента при получении информации поднимал руку, фиксировал время на компьютере и сообщал его наблюдателю за экспериментом. Такая реализация процесса поиска информации прошла без особых трудностей и позволила получить определенный статистический материал по времени получения необходимых для процесса образования информационных материалов определенного объема из двух различных мест хранения. Полученные результаты зафиксированы в приведенных ниже таблицах 1 и 4, и по их

данным нами составлены итоговые таблицы 2 и 5, которые содержат частоты попадания n_i в соответствующие временные интервалы. На рисунках 1 и 2 представлены, соответственно, гистограмма частот времени получения информации из сети Интернет и гистограмма частот времени получения информации из базы библиотеки, представляющие собой кривые гамма-распределения.

Таблица 1 – Время получения определенного объема информации из Интернет

k	t _k , мин	k	t _k , мин	k	t _k , мин
1	2.8	21	2.0	41	3.4
2	3.4	22	3.2	42	4.8
3	3.3	23	3.5	43	1.9
4	3.8	24	1.5	44	5.6
5	2.2	25	4.8	45	3.4
6	3.4	26	2.8	46	4.9
7	1.9	27	5.6	47	2.1
8	5.6	28	2.0	48	4.3
9	2.7	29	4.5	49	7.1
10	1.5	30	4.2	50	6.4
11	6.4	31	3.2	51	4.0
12	3.3	32	4.1	52	1.5
13	7.1	33	3.4	53	3.6
14	3.5	34	6.3	54	2.8
15	2.8	35	2.6	55	2.5
16	4.8	36	5.1	56	4.8
17	3.6	37	5.6	57	4.5
18	3.6	38	3.7	58	2.4
19	4.8	39	3.5	59	4.4
20	1.5	40	1.9	60	2.8

Таблица 2 – Частота попадания во временные интервалы для таблицы 1

$t_i < t \leq t_{i+1}$	$0 < t \leq 1.5$	$1.5 < t \leq 3.0$	$3.0 < t \leq 4.5$	$4.5 < t \leq 6.0$	$6.0 < t \leq 7.5$
n_i	4	16	24	11	5

Для значений

$$\nu = 1.28 \text{ и } \alpha = 5.0 \tag{23}$$

предполагаемая плотность распределения равна

$$p(t) = \frac{1.28^{5.0}}{\Gamma(5)} t^{4.0} e^{-1.28t} \tag{24}$$

Для сравнения теоретического и экспериментального распределений воспользуемся критерием Пирсона [5].

Определим вероятности попадания значений времени в границы выделенных интервалов:

$$P(t_i < t \leq t_{i+1}) = \int_{t_i}^{t_{i+1}} p(t) dt \tag{25}$$

Находим ожидаемое количество событий n_i' , приходящихся на каждый интервал

$$n_i' = N * P(t_i < t \leq t_{i+1}), \tag{26}$$

где N – количество проведенных наблюдений.

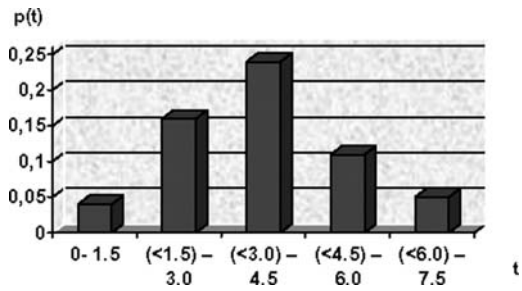


Рисунок 1 – Гистограмма частот времени получения информации из сети Интернет

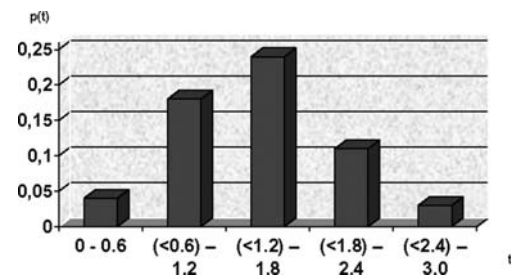


Рисунок 2 – Гистограмма частот времени получения информации из базы библиотеки

Определим значение $\chi^2_{набл}$, которое находится по формуле

$$\chi^2_{набл} = \sum_i \frac{(n_i - n_i')^2}{n_i} \tag{27}$$

Результаты расчетов сведем в таблицу 3.

Таблица 3 – Оценка наблюдаемого $\chi^2_{набл}$

$t_i < t \leq t_{i+1}$	n_i	$P(t_i < t \leq t_{i+1})$	n_i'	$\frac{(n_i - n_i')^2}{n_i}$
$0 < t \leq 1.5$	4	0.046	2.74	0.575
$1.5 < t \leq 3.0$	16	0.294	17.65	0.155
$3.0 < t \leq 4.5$	24	0.342	20.50	0.599
$4.5 < t \leq 6.0$	11	0.199	11.94	0.074
$6.0 < t \leq 7.5$	5	0.082	4.90	0.002

$$\chi^2_{набл} = 1.405$$

Число степеней свободы определяется выражением

$$i - r - 1 = 2, \tag{28}$$

где $i = 5$ – количество интервалов после объединения; $r = 2$ – число параметров, оцениваемых при выборке (14).

Из таблицы критических точек распределения $\chi^2_{кр}$ по уровню значимости 0.05 и числу степеней свободы, равном 2, находим критическую точку правосторонней критической области:

$$\chi^2_{кр}(0.05, 2) = 6.0. \tag{29}$$

Так как $\chi_{\text{набл}}^2 < \chi_{\text{кр}}^2$, то нет оснований отвергать гипотезу о гамма-распределении и данные наблюдений согласуются с этой гипотезой.

Исследуем время получения определенного объема информации из базы библиотеки.

Полученные данные сведем в таблицу 4.

Таблица 4 – Время получения определенного объема информации из базы библиотеки

k	t _k , мин	k	t _k , мин	k	t _k , мин
1	1.1	21	0.8	41	1.4
2	1.4	22	1.2	42	1.9
3	1.3	23	1.5	43	0.7
4	1.9	24	0.5	44	2.2
5	0.9	25	1.8	45	1.4
6	1.4	26	1.0	46	1.9
7	0.7	27	2.2	47	0.7
8	2.5	28	0.8	48	1.7
9	1.0	29	1.7	49	2.9
10	0.5	30	1.6	50	2.5
11	2.4	31	1.2	51	1.5
12	1.3	32	1.5	52	0.6
13	1.9	33	1.4	53	1.4
14	1.4	34	2.0	54	1.0
15	1.0	35	1.0	55	1.3
16	1.8	36	1.9	56	1.9
17	1.4	37	2.1	57	1.7
18	1.4	38	1.4	58	0.9
19	1.7	39	1.4	59	1.8
20	0.5	40	0.7	60	1.0

Таблица 5 – Частота попадания во временные интервалы для таблицы 4

$t_i < t \leq t_{i+1}$	$0 < t \leq 0.6$	$0.6 < t \leq 1.2$	$1.2 < t \leq 1.8$	$1.8 < t \leq 2.4$	$2.4 < t \leq 3.0$
n_i	4	18	24	11	3

Для данных, приведенных в таблице 4, используем значения

$$\nu = 0.344 \text{ и } \alpha = 3.921. \quad (30)$$

Отсюда предполагаемая плотность распределения равна

$$p(t) = \frac{0.344^{3.921}}{\Gamma(3.921)} t^{2.921} e^{-0.344t}. \quad (31)$$

Данные для сравнения теоретического и экспериментального распределений по критерию Пирсона [5] сведем в таблицу 6.

Таблица 6 – Оценка наблюдаемого $\chi_{\text{набл}}^2$

$t_i < t \leq t_{i+1}$	n_i	$P(t_i < t \leq t_{i+1})$	n_i^*	$\frac{(n_i - n_i^*)^2}{n_i}$
$0 < t \leq 0.6$	4	0.097	5.80	0.557
$0.6 < t \leq 1.2$	18	0.303	18.19	0.078
$1.2 < t \leq 1.8$	24	0.282	16.92	3.854
$1.8 < t \leq 2.4$	11	0.172	10.32	0.045
$2.4 < t \leq 3.0$	3	0.085	5.11	0.870

$$\chi_{\text{набл}}^2 = 5.404$$

Число степеней свободы равно

$$i - r - 1 = 2. \quad (32)$$

В соответствии со значением критической точки правосторонней критической области (20) также имеем $\chi_{\text{набл}}^2 < \chi_{\text{кр}}^2$, то есть нет оснований отвергать и в этом случае гипотезу о гамма-распределении и данные наблюдений также согласуются с этой гипотезой.

Модель процесса получения знаний, описанная в работе [4], основана также на том, что вероятность получения знаний в бесконечно малом интервале времени пропорциональна этому интервалу. При этом предполагалось, что объем информации предлагаемый для усвоения получен. Так как процесс получения информации является неотъемлемой частью обучения, то обучение можно рассматривать как произведение двух событий: "получения нужной информации" и "усвоение информации". Вероятность обучения $P_{\text{об}}$ равна произведению вероятности получения нужной информации $P_{\text{ин}}$ на вероятность усвоения полученной информации $P_{\text{ус}}$:

$$P_{\text{об}} = P_{\text{ин}} P_{\text{ус}}. \quad (33)$$

Данное выражение можно использовать в случае, когда время, выделяемое на поиск информации и время на ее усвоение, задаются независимо друг от друга. Причем, если выделение времени осуществляется пропорционально ожидаемому, все составляющие выражения (24) будут принимать постоянные значения.

В случае, если выделяется общее время на поиск и усвоение, плотность распределения времени на обучение $p_{\text{об}}(t)$ представляет собой композицию плотности распределения времени на поиск информации $p_i(t)$ и плотности распределения времени на усвоение $p_{\text{ус}}(t)$:

$$p_{\text{об}}(t) = p_i(t) * p_{\text{ус}}(t), \quad (34)$$

где * – символ композиции.

Плотность распределения времени на усвоение материала, в соответствии с работой [4], имеет вид

$$p_{\text{ус}}(t) = \lambda e^{-\lambda t}. \quad (35)$$

Плотность распределения величины определяется выражением

$$p_{\text{об}}(t) = \int_{-\infty}^{\infty} p_i(x) p_{\text{ус}}(t-x) dx. \quad (36)$$

Подстановка плотностей распределения (4) и (26) в (27) с учетом того, что при $\tau < 0$ имеем $p_i(\tau) = 0$ и $p_{\text{ус}}(t-\tau) = 0$ при $\tau > t$ приводит к обобщенному закону Эрланга 1-го порядка [6]:

$$p_{\text{об}}(t) = \frac{\nu \lambda (e^{-\nu t} - e^{-\lambda t})}{\lambda - \nu}. \quad (37)$$

При $\nu = \lambda$, после раскрытия неопределенности имеем

$$p_{\text{об}}(t) = \lambda^2 t e^{-\lambda t}. \quad (38)$$

Условия, при которых процесс обучения можно будет описывать цепями Маркова требуют дальнейших исслед-

дований. Если считать, что усвоение знаний и получение информации не зависят друг от друга, то можно воспользоваться выражением (25), а каждый из процессов описывать цепями Маркова.

РЕЗУЛЬТАТЫ

Рассмотрены законы распределения времени получения информации участниками учебного процесса. Показано, что получение информации достаточно хорошо описывается экспоненциальным законом для малых объемов информации и гамма-распределением для больших объемов. На примерах продемонстрировано подтверждение теоретических предположений экспериментальными данными. Намечены пути объединения модели учебного процесса с процессом его информационного обеспечения.

Как мы видим из сравнения таблиц 1 и 3, наличие хорошо систематизированных материалов в собственной информационной системе университета обеспечит уменьшение материальных и временных затрат на поиск и получение необходимых информационных ресурсов всеми участниками учебного процесса.

ЗАКЛЮЧЕНИЕ

Ориентация библиотеки на обеспечение образовательной деятельности университета и современные условия подготовки специалистов предъявляют высокие требования к актуальности информации, своевременности ее получения [7].

Предлагаемая модель информационного обеспечения как основа образовательного пространства создает предпосылки для получения эффективного сотрудничества между различными участниками образовательного процесса, целью которого является создание единой информационно-образовательной среды университета.

Достигнутые результаты планируется использовать в построении более высоких уровней иерархии информационной модели вуза, а именно: при построении иерархической структуры информационного обеспечения процесса обучения; при исследовании качества взаимодействия системы обучения с системой ее информационного обеспечения; при построении алгоритма анализа качества

информационного обеспечения процесса обучения; при создании модели управления информационными потоками для процессов образования в университете; при построении системы информационного обеспечения на всех уровнях обучения.

ПЕРЕЧЕНЬ ССЫЛОК

1. A Survey of Distance Education Challenges and Technologies. International Journal of Distance Education Technologies, Jan-Mar 2003, Vol. 1. Issue 1, p1, 21p; (AN 10166088).
2. Юдин В.В. Библиотечный комплекс вуза для дистанционного образования // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества: Тема 2003 года: Библиотека и доступность информации в современном мире: электронные ресурсы науке, культуре и образованию: Тр.конф./ 10-я юбил.междунар.конф. "Крым-2003", - М.: ГПНТБ России, 2003. - Т. 3. - С. 906-908.
3. Немцев О.В. Информационная среда вуза // Проектирование образовательных информационных ресурсов, систем и технологий. Сб. докладов и сообщений. - М.: ИЦПКПС, 1998. - С. 55-60.
4. Вершина А.И., Солдатов Б.Т. Моделирование процесса обучения // "Радиоелектроніка. Інформатика. Управління". - Запоріжжя: ЗНТУ, 2003. - №1. - С. 65-72.
5. Вершина А.И., Солдатов Б.Т. Ермоленко А.А. Учебный процесс как иерархическая система // "Радиоелектроніка. Інформатика. Управління". - Запоріжжя: ЗНТУ, 2004. - №1. - С. 54-62.
6. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. - М.: Высшая школа, 1975. - 336 с.
7. Киричек Г.Г. Использование электронных ресурсов библиотеки университета в учебном процессе // Електронні ресурси в бібліотеках - 2004: Матеріали конференції / Науково-практична конференція, 17-21 травня 2004 р., м. Миколаїв. - К.: ТОВ "ІММ "Фраксім", 2004. - 92 с.

Надійшла 19.04.2004
Після доробки 29.10.2004

Побудова моделі навчального процесу університету тісно зв'язана з інформаційним забезпеченням. У статті розглядаються можливість і особливості рішення задачі побудови моделі інформаційно-освітньої системи університету, заснованої на припущенні, що імовірність одержання деякого обсягу інформації в нескінченно малому проміжку часу з різних місць збереження пропорційна розміру цього проміжку. Такий підхід добре узгоджується з моделлю навчального процесу, яка заснована на цих же принципах.

Modelling a university educational process is closely related to information supply. The article deals with the way to model the university information and educational system, based on the assumption that the probability of retrieving a certain amount of information from different storage locations within an infinitesimal time interval is proportional to the length of this interval. Such approach is well in line with the educational process model, based on the same principles.

УДК 519.6:514.1

И. В. Гребенник, Т. Е. Романова, Л. Г. Евсева

ОСНОВНАЯ ОПТИМИЗАЦИОННАЯ ЗАДАЧА ГЕОМЕТРИЧЕСКОГО ПРОЕКТИРОВАНИЯ В ИНТЕРВАЛЬНОМ ВИДЕ

Строится математическая модель оптимизационной задачи геометрического проектирования в интервальном пространстве. Выделяются множества базовых двумерных и трехмерных интервальных геометрических объектов. Вводится

понятие геометрической информации о базовых и составных интервальных геометрических объектах. Формируется область допустимых решений оптимизационной интервальной задачи размещения с использованием метода Ф-функций.

АКТУАЛЬНОСТЬ

Многие прикладные и практические задачи размещения, упаковки, раскроя и покрытия, модели которых используются в различных предметных областях, относятся к классу задач геометрического проектирования [1]. Традиционно математические модели и методы решения задач геометрического проектирования строились и применялись без учета погрешностей исходных данных, то есть в идеализированной форме. Развитие геометрического проектирования как научного направления вызвало необходимость учета погрешностей метрических характеристик и параметров размещения геометрических объектов в математических моделях задач и методах их решения.

Важный класс задач геометрического проектирования составляют оптимизационные задачи, решение которых связано с поиском экстремального значения некоторой целевой функции. Математические модели оптимизационных задач размещения (упаковки, раскроя), покрытия, в том числе некоторые комбинаторные задачи, учитывающие погрешности исходных данных, а также подходы к их решению представлены, например, в [2–5].

Основным способом учета погрешностей исходных данных и метрических характеристик в этих работах является применение элементов теории интервального анализа [6] в геометрическом проектировании [7, 8].

Однако, к настоящему времени не построена математическая модель основной задачи геометрического проектирования [1], реализации которой покрывали бы множество математических моделей задач рассматриваемого класса в интервальном виде.

Целью настоящей работы является построение математической модели основной оптимизационной интервальной задачи геометрического проектирования, которая, с одной стороны, рациональным образом учитывала бы погрешности исходных данных, а с другой стороны была представлена в таком виде, который бы позволял реализовать ее современными эффективными методами оптимизации.

ПОСТРОЕНИЕ МАТЕМАТИЧЕСКОЙ МОДЕЛИ

В качестве математических моделей реальных объектов в пределах данного исследования рассматриваются точечные множества двумерных и трехмерных интервальных пространств $I_s^m \mathbf{R}$, $m = 2, 3$ [7–9].

Как известно [1, 10], основой построения математических моделей задач геометрического проектирования является моделирование отношений между геометрическими объектами.

Под *отношениями* интервальных геометрических объектов будем понимать отношения пересечения, касания, непересечения в обычном теоретико-множественном смысле.

Поскольку мощность множества пространственных форм объектов интервальных пространств, также как и евклидовых, – континуум, то, очевидно, что описание в аналитическом виде отношений между всеми парами интервальных объектов является задачей неразрешимой.

Вследствие этого, по аналогии с идеализированным случаем [11, 12], возникла идея выделения класса так называемых *базовых* интервальных объектов.

В данном исследовании в качестве *базовых* двумерных интервальных объектов \mathbf{T} выбираются: интервальные круги \mathbf{C} , интервальные прямоугольники \mathbf{R} , правильные интервальные многоугольники \mathbf{H} , выпуклые интервальные многоугольники \mathbf{K} и замыкания их дополнений до всего двумерного интервального пространства, т. е. $\mathbf{T}^* = \text{frT} \cup (I_s^2 \mathbf{R} \setminus \text{clT})$, где frT , clT – интервальная граница и топологическое замыкание множества \mathbf{T} [9]. В качестве *базовых* трехмерных интервальных объектов рассматриваются: интервальные шары \mathbf{S} , интервальные параллелепипеды \mathbf{P} , интервальные цилиндры \mathbf{C}_y , интервальные конусы \mathbf{C}_o , выпуклые интервальные многогранники \mathbf{K}_p , а также замыкания их дополнений до всего трехмерного интервального пространства, т. е. $\mathbf{T}^* = \text{frT} \cup (I_s^3 \mathbf{R} \setminus \text{clT})$.

Обозначим множество базовых интервальных объектов $\mathfrak{X} = \mathfrak{S} \cup \mathfrak{S}^*$, где в двумерном случае $\mathfrak{S} = \{\mathbf{C}, \mathbf{R}, \mathbf{H}, \mathbf{K}\}$, $\mathfrak{S}^* = \{\mathbf{C}^*, \mathbf{R}^*, \mathbf{H}^*, \mathbf{K}^*\}$, в трехмерном – $\mathfrak{S} = \{\mathbf{S}, \mathbf{P}, \mathbf{C}_y, \mathbf{C}_o, \mathbf{K}_p\}$, $\mathfrak{S}^* = \{\mathbf{S}^*, \mathbf{P}^*, \mathbf{C}_y^*, \mathbf{C}_o^*, \mathbf{K}_p^*\}$.

Используя различные комбинации операций пересечения и объединения базовых интервальных объектов, можно генерировать интервальные объекты более сложной пространственной формы. Аналогично идеализированному случаю будем называть такие интервальные объекты *составными* [13].

Определим кортежи геометрической информации, следуя [1], индуцирующие в интервальном пространстве *базовые* интервальные объекты.

Пусть $\mathbf{g} = (\Upsilon, M, U)$ – кортеж геометрической информации, индуцирующий однозначно произвольный интервальный объект $\mathbf{T} \in \mathfrak{X}$ в пространстве $I_s^m \mathbf{R}$, $m = 2, 3$, где Υ, M, U – пространственная форма, метрические характеристики и параметры размещения объекта $\mathbf{T} \in \mathfrak{X}$ в соответствующем пространстве $I_s^m \mathbf{R}$, $m = 2, 3$.

Геометрическая информация об интервальных базовых 2D объектах задается следующим образом: $\mathbf{g}_{\mathbf{C}} = (\mathbf{C}, \langle R \rangle, U)$, $\mathbf{g}_{\mathbf{C}^*} = (\mathbf{C}^*, \langle R \rangle, U)$, где $\langle R \rangle = \langle r, v_r \rangle$ – метрическая характеристика, характеризующая радиус интервального круга \mathbf{C} , здесь и далее $U = (\langle x, v_x \rangle, \langle y, v_y \rangle)$ – параметры размещения объекта $\mathbf{T} \in I_s^2 \mathbf{R}$; $\mathbf{g}_{\mathbf{R}} = (\mathbf{R}, (\langle A \rangle, \langle B \rangle), U)$, $\mathbf{g}_{\mathbf{R}^*} = (\mathbf{R}^*, (\langle A \rangle, \langle B \rangle), U)$, где $\langle A \rangle = \langle a, v_a \rangle$, $\langle B \rangle = \langle b, v_b \rangle$ – параметры, характеризующие соответственно длину и ширину интервального прямоугольника \mathbf{R} ; $\mathbf{g}_{\mathbf{H}} = (\mathbf{H}, (n, \langle R \rangle, \langle \Theta \rangle), U)$, $\mathbf{g}_{\mathbf{H}^*} = (\mathbf{H}^*, (n, \langle R \rangle, \langle \Theta \rangle), U)$, где n – число сторон многоугольника \mathbf{H} , $\langle R \rangle = \langle r, v_r \rangle$ – метрическая характеристика, характеризующая радиус описанной около \mathbf{H} интервальной окружности, и $\langle \Theta \rangle = \langle \theta, v_\theta \rangle$ – интервальный угол поворота многоугольника \mathbf{H} в собственной системе координат \mathbf{H} , при условии $0 \leq \theta < 2\pi/n$; $\mathbf{g}_{\mathbf{K}} = (\mathbf{K}, (V_1, V_2, \dots, V_n), U)$, $\mathbf{g}_{\mathbf{K}^*} = (\mathbf{K}^*, (V_1, V_2, \dots, V_n), U)$, где $V_i = (\langle x_i, v_{x_i} \rangle, \langle y_i, v_{y_i} \rangle)$, $\langle x_i, v_{x_i} \rangle, \langle y_i, v_{y_i} \rangle$ – координаты i -й вершины многоугольника \mathbf{K} , $i = 1, \dots, n$, заданные против часовой стрелки в его собственной системе интервальных координат.

Геометрическая информация о базовых интервальных 3D объектах задается следующим образом: $\mathbf{g}_S = (S, \langle R \rangle, U)$, $\mathbf{g}_{S^*} = (S^*, \langle R \rangle, U)$, где $\langle R \rangle$ – метрическая характеристика, характеризующая интервальный радиус интервального шара S , здесь и далее $U = (\langle x, v_x \rangle, \langle y, v_y \rangle, \langle z, v_z \rangle)$ – параметры размещения объекта $\mathbf{T} \in \mathbf{I}_s^3 \mathbf{R}$; $\mathbf{g}_P = (P, (\langle A \rangle, \langle B \rangle, \langle H \rangle), U)$, $\mathbf{g}_{P^*} = (P^*, (\langle A \rangle, \langle B \rangle, \langle H \rangle), U)$, где $\langle A \rangle = \langle a, v_a \rangle$, $\langle B \rangle = \langle b, v_b \rangle$, $\langle H \rangle = \langle h, v_h \rangle$ – метрические характеристики, характеризующие соответственно длину, ширину и высоту интервального параллелепипеда P ; $\mathbf{g}_{C_y} = (C_y, (\langle H \rangle, \langle R \rangle), U)$, $\mathbf{g}_{C_y^*} = (C_y^*, (\langle H \rangle, \langle R \rangle), U)$, где $\langle H \rangle$, $\langle R \rangle$ – метрические характеристики, характеризующие соответственно интервальную высоту и радиус основания цилиндра C_y ; $\mathbf{g}_{C_o} = (C_o, (\langle H \rangle, \langle R \rangle), U)$, $\mathbf{g}_{C_o^*} = (C_o^*, (\langle H \rangle, \langle R \rangle), U)$, где $\langle H \rangle$, $\langle R \rangle$ – метрические характеристики, характеризующие соответственно высоту и радиус основания конуса C_o ; $\mathbf{g}_{K_p} = (K_p, (V_1, V_2, \dots, V_n), U)$, $\mathbf{g}_{K_p^*} = (K_p^*, (V_1, V_2, \dots, V_n), U)$, где $V_i = (\langle X_i \rangle, \langle Y_i \rangle, \langle Z_i \rangle)$, $\langle X_i \rangle, \langle Y_i \rangle, \langle Z_i \rangle$ – координаты i -й вершины многогранника K_p , $i = 1, \dots, n$, в его собственной системе интервальных координат.

Полагаем, что полюс базового интервального объекта $\mathbf{T} \in \mathbf{I}_s^m \mathbf{R}$, $m = 2, 3$, совпадает с началом его собственной системы координат.

Определим кортеж геометрической информации о составном интервальном объекте.

Пусть $\mathbf{T} = \mathbf{T}_1 \circ \mathbf{T}_2 \circ \dots \circ \mathbf{T}_i \circ \mathbf{T}_{i+1} \circ \dots \circ \mathbf{T}_{k-1} \circ \mathbf{T}_k$, где $\circ \in \{\cup, \cap\}$; \mathbf{T}_i – базовый интервальный объект.

Тогда геометрическую информацию о составном интервальном объекте можно представить кортежем вида

$$\mathbf{g}_T = (\mathbf{g}_1 \circ \mathbf{g}_2 \circ \dots \circ \mathbf{g}_i \circ \mathbf{g}_{i+1} \circ \dots \circ \mathbf{g}_{k-1} \circ \mathbf{g}_k, U),$$

где U – параметры размещения объекта \mathbf{T} , $\circ \in \{\cup, \cap\}$ – символ композиции в таком смысле: $\mathbf{g}_1 \cap \mathbf{g}_2$ значит, что рассматривается пересечение объектов $\mathbf{T}_1 \cap \mathbf{T}_2$, а $\mathbf{g}_1 \cap \dots \cap \mathbf{g}_j \cup \mathbf{g}_{j+1} \cap \dots \cap \mathbf{g}_k$ – объединение объектов $\mathbf{T}_1 \cap \dots \cap \mathbf{T}_j$ и $\mathbf{T}_{j+1} \cap \dots \cap \mathbf{T}_k$, \mathbf{g}_i – геометрическая информация о базовом интервальном объекте \mathbf{T}_i , $i = 1, 2, \dots, n$.

Заметим, что в пределах данного исследования допускаются только собственные конгруэнтные преобразования интервальных объектов типа трансляции. Объект \mathbf{T} , транслированный на интервальный вектор U в пространстве $\mathbf{I}_s^m \mathbf{R}$, $m = 2, 3$, обозначим как $\mathbf{T}(U) = \{X \in \mathbf{I}_s^m \mathbf{R} | X = U + Y, Y \in \mathbf{T}\}$.

Для моделирования отношений между интервальными геометрическими объектами, следуя идеализированному случаю [13], будем использовать интервальные Φ -функции [14]. Φ -функция зависит от взаимного положения пары интервальных объектов. При этом, значения этой функции дают интервальную оценку (в зависимости от выбранной в интервальном пространстве метрики) следующих ситуаций: два интервальных объекта пересекаются (имеют общие внутренние точки); касаются

(имеют общие только граничные точки) или не имеют общих точек вообще. Более того, значения Φ -функции являются некоторой оценкой пересечения пары объектов в первом случае и, по крайней мере, могут быть оценкой интервального расстояния между парой интервальных объектов в остальных случаях.

Основываясь на методе, приведенном в [15], можно построить Φ -функции всех пар базовых интервальных двумерных и трехмерных объектов.

Интервальные поверхности 0- и ρ -уровней интервальных Φ -функций и нормализованных интервальных Φ -функций базовых двумерных и трехмерных интервальных объектов зададим так:

$$\gamma(U) = \{U \in \mathbf{I}_s^{2m} \mathbf{R} : (\text{int } \mathbf{T}_1(U_1) \cap \text{int } \mathbf{T}_2(U_2) = \emptyset) \wedge (\text{fr } \mathbf{T}_1(0) \cap \text{fr } \mathbf{T}_2(U_2) \neq \emptyset)\},$$

$$\gamma(U) = \{(U_1, U_2) \in \mathbf{I}_s^{2m} \mathbf{R} : \Phi(U_1, U_2) = \langle 0 \rangle\},$$

$$\gamma^\rho(U_1, U_2) = \{(U_1, U_2) \in \mathbf{I}_s^{2m} \mathbf{R} : \tilde{\Phi}(U_1, U_2) = \rho\} \rho \in \mathbf{I}_s \mathbf{R}.$$

Для построения интервальных Φ -функций пары составных интервальных объектов могут быть естественным образом использованы Φ -функции базовых интервальных объектов.

Пусть информация $\mathbf{g} = (Y, M, U)$ индуцирует точечное множество \mathbf{T} в пространстве $\mathbf{I}_s^m \mathbf{R}$, а информация $\hat{\mathbf{g}} = (\hat{Y}, \hat{M}, \hat{U})$ – точечное множество $\hat{\mathbf{T}}$ в пространстве $\mathbf{I}_s^m \mathbf{R}$. Отображение \mathbf{P} вида $\hat{\mathbf{g}} = \mathbf{P}\mathbf{g}$ называется интервальной задачей геометрического проектирования.

Нефиксированные элементы компонент информации \mathbf{g} называются переменными этой информации.

Например, задача поиска покрытия интервальной области $\mathbf{T}_0(U_0)$ интервальными объектами $\mathbf{T}_1(U_1)$, $\mathbf{T}_2(U_2)$, ..., $\mathbf{T}_n(U_n)$ состоит в задании такого отображения \mathbf{P} , при котором $\hat{\mathbf{g}}$ индуцирует множество $\mathbf{T}_0(U_0) \cap \bigcup_{i=1}^n \mathbf{T}_i(U_i)$, и

выполняется условие $\mathbf{T}_0(U_0) \cap \bigcup_{i=1}^n \mathbf{T}_i(U_i) = \mathbf{T}_0(U_0)$, а задача размещения интервальных объектов $\mathbf{T}_1(U_1)$, $\mathbf{T}_2(U_2)$, ..., $\mathbf{T}_n(U_n)$ в интервальной области $\mathbf{T}_0(U_0)$ состоит в поиске отображения \mathbf{P} , при котором $\hat{\mathbf{g}}$ индуцирует множество $\mathbf{T}_0(U_0) \cap \bigcup_{i=1}^n \mathbf{T}_i(U_i)$ и выполняется условие

$$\left(\mathbf{T}_0(U_0) \cap \bigcup_{i=1}^n \mathbf{T}_i(U_i) = \bigcup_{i=1}^n \mathbf{T}_i(U_i) \right) \wedge (\text{int } \mathbf{T}_i(U_i) \cap \text{int } \mathbf{T}_j(U_j) = \emptyset, i \neq j = 1, 2, \dots, n).$$

Здесь переменными информации \mathbf{g} являются параметры размещения U_1, U_2, \dots, U_n интервальных объектов $\mathbf{T}_1(U_1), \mathbf{T}_2(U_2), \dots, \mathbf{T}_n(U_n)$.

Интервальные задачи геометрического проектирования, в которых переменные $U = U_1, U_2, \dots, U_n$ информации \mathbf{g} определяются в результате решения оптимизационной задачи, называются оптимизационными интервальными задачами геометрического проектирования.

В таких задачах осуществляется отображение \mathbf{P} исходной информации \mathbf{g} , при котором элементы компонент этой информации преобразуются таким образом, чтобы заданная на некотором множестве $\mathbf{W} \subset \mathbf{I}_s^k \mathbf{R}$, $k = mn$, интервальная функция $\mathbf{k}(U)$ достигала своего наибольшего или наименьшего значения. Здесь $\mathbf{I}_s^k \mathbf{R} = \underbrace{\mathbf{I}_s^m \mathbf{R} \times \mathbf{I}_s^m \mathbf{R} \times \dots \times \mathbf{I}_s^m \mathbf{R}}_n$ [16].

Введенные понятия позволяют сформулировать основную оптимизационную интервальную задачу геометрического проектирования.

Требуется определить такое значение \hat{U} на области допустимых решений $\mathbf{W} \subset \mathbf{I}_s^{nm} \mathbf{R}$ и соответствующий образ $\hat{\mathbf{T}}_U$, индуцируемый в пространстве $\mathbf{I}_s^k \mathbf{R}$ информацией $\hat{\mathbf{g}} = \mathbf{P}\mathbf{g}$, чтобы интервальная функция цели $\mathbf{k}(U)$ на области допустимых решений \mathbf{W} достигала наименьшего или наибольшего значения:

$$\mathbf{k}(\hat{U}) = \inf_{U \in \mathbf{W}} \mathbf{k}(U) (\mathbf{k}(\hat{U}) = \sup_{U \in \mathbf{W}} \mathbf{k}(U)) . \quad (1)$$

Если в основной оптимизационной интервальной задаче (1) \mathbf{W} является некоторым интервальным комбинаторным множеством \mathbf{IE} [17], то такая задача называется основной комбинаторной интервальной задачей геометрического проектирования.

Рассмотрим пример реализации математической модели (1) основной интервальной задачи геометрического проектирования.

Задача. Пусть имеется конечный набор геометрических объектов $\mathbf{T}_i(U_i) \subset \mathbf{I}_s^m \mathbf{R}$, $i \in \{1, 2, \dots, n\} = J_n$, и область размещения $\mathbf{T}_0(U_0) \subset \mathbf{I}_s^m \mathbf{R}$, $m \in \{2, 3\}$, метрические характеристики M_i , $i \in J_n$, и параметры размещения U_i , $i \in J_n$, которых заданы с некоторыми погрешностями. Заданы минимально ρ_{0i}^- , ρ_{ij}^- и максимально ρ_{0i}^+ , ρ_{ij}^+ допустимые интервальные расстояния между объектами $\mathbf{T}_0^*(U_0)$, $\mathbf{T}_i(U_i)$ и $\mathbf{T}_j(U_j)$.

Необходимо, учитывая заданные погрешности, разместить данный набор геометрических объектов в области таким образом, чтобы некоторый критерий качества размещения достигал своего экстремума.

Интервальное расстояние $\rho(\mathbf{T}_1, \mathbf{T}_2)$ между интервальными объектами \mathbf{T}_1 и \mathbf{T}_2 , при условии $\mathbf{T}_1 \cap \mathbf{T}_2 = \emptyset$, определяется так [9]:

$$\rho(\mathbf{T}_1, \mathbf{T}_2) = \min_{Z_1 \in \mathbf{T}_1, Z_2 \in \mathbf{T}_2} \rho(Z_1, Z_2).$$

Условие размещения объекта $\mathbf{T}_i(U_i) \subset \mathbf{I}_s^m \mathbf{R}$ в области $\mathbf{T}_0(U_0) \subset \mathbf{I}_s^m \mathbf{R}$:

$$\Phi_{0i}(U_0, U_i) \geq \langle 0 \rangle.$$

Ограничения на минимально и максимально допустимые расстояния между объектами $\mathbf{T}_0^*(U_0) = (\mathbf{I}_s^m \mathbf{R} / \text{cl} \mathbf{T}_0(U_0)) \cup \text{fr} \mathbf{T}_0(U_0)$ и $\mathbf{T}_i(U_i)$:

$$\rho_{0i}^- \leq \check{\Phi}_{0i}(U_0, U_i) \leq \rho_{0i}^+,$$

где ρ_{0i}^- и ρ_{0i}^+ – интервальное расстояние между объектами $\mathbf{T}_i(U_i)$ и $\mathbf{T}_0^*(U_0)$, $\check{\Phi}_{0i}(U_0, U_i)$ – нормализованная Φ -функция объектов $\mathbf{T}_i(U_i)$ и $\mathbf{T}_0(U_0)$.

Условие взаимного непересечения объектов $\mathbf{T}_i(U_i)$ и $\mathbf{T}_j(U_j)$, $i \neq j$:

$$\Phi_{ij}(U_i, U_j) \geq \langle 0 \rangle.$$

Ограничения на минимально и максимально допустимые расстояния между объектами $\mathbf{T}_i(U_i)$ и $\mathbf{T}_j(U_j)$

$$\rho_{ij}^- \leq \check{\Phi}_{ij}(U_i, U_j) \leq \rho_{ij}^+,$$

где ρ_{ij}^- и ρ_{ij}^+ – интервальное расстояние между интервальными объектами $\mathbf{T}_i(U_i)$ и $\mathbf{T}_j(U_j)$, $\check{\Phi}_{ij}(U_i, U_j)$ – нормализованная Φ -функция объектов $\mathbf{T}_i(U_i)$ и $\mathbf{T}_j(U_j)$.

Тогда интервальная математическая модель задачи размещения примет вид:

$$\inf_{U \in \mathbf{D} \subset \mathbf{I}_s^k \mathbf{R}} \chi(U), \quad (2)$$

$$U = (\langle U_1 \rangle, \langle U_2 \rangle, \dots, \langle U_n \rangle) \in \mathbf{I}_s^k \mathbf{R}, \quad k = mn, \quad m \in \{2, 3\},$$

\mathbf{D} – область допустимых решений, которая описывается системой интервальных неравенств

$$\begin{cases} \Phi_{ij}(U_i, U_j) - \rho_{ij}^- \geq \langle 0 \rangle, \\ -\Phi_{ij}(U_i, U_j) + \rho_{ij}^+ \geq \langle 0 \rangle, \\ \Phi_{0i}(U_0, U_i) - \rho_{0i}^- \geq \langle 0 \rangle, \\ -\Phi_{0i}(U_0, U_i) + \rho_{0i}^+ \geq \langle 0 \rangle, \end{cases} \quad (3)$$

$$i, j = 1, 2, \dots, n, \quad i \neq j.$$

Таким образом, математическая модель (2) является реализацией математической модели (1). При этом, интервальная область допустимых решений \mathbf{D} описывается системой неравенств с использованием интервальных нормализованных Φ -функций. Легко видеть, что в случае, когда ρ_{0i}^- , ρ_{ij}^- , ρ_{0i}^+ , ρ_{ij}^+ равны нулю, достаточно использовать не нормализованные интервальные Φ -функции соответствующих пар объектов. Тогда система ограничений (3) примет следующий вид:

$$\begin{cases} \Phi_{ij}(U_i, U_j) \geq \langle 0 \rangle, \\ \Phi_{0i}(U_0, U_i) \geq \langle 0 \rangle, \end{cases} \quad (4)$$

$$i, j = 1, 2, \dots, n, \quad i \neq j.$$

Если положить все погрешности исходных данных равными нулю, то модели (2), (3) и (2), (4) описывают идеализированные задачи размещения геометрических объектов евклидовых пространств.

ВЫВОДЫ

Построенная в работе математическая модель (1) основной оптимизационной задачи геометрического проектирования в интервальном пространстве позволит строить адекватные математические модели задач геометрического проектирования (упаковки, раскрыя, покрытия) с учетом погрешностей исходных данных. Она может быть использована при моделировании задач, связанных

с преобразованием геометрической информации, для решения которых применяются методы евклидовой комбинаторной оптимизации [18].

Однако, открытой проблемой является расширение классов двумерных и трехмерных базовых интервальных объектов, а также построение полного класса Φ -функций интервальных базовых неориентированных геометрических объектов.

ПЕРЕЧЕНЬ ССЫЛОК

1. Стоян Ю.Г. Математические модели и оптимизационные методы геометрического проектирования / Стоян Ю.Г., Яковлев С.В. – Киев: Наук. думка, 1986. – 268 с.
2. Стоян Ю.Г. Комбинаторная оптимизационная задача размещения прямоугольников с учетом погрешностей исходных данных / Стоян Ю.Г., Романова Т.Е., Евсеева Л.Г. // Докл. НАН Украины. Сер. А. – 1997. – № 7 – С. 56–60.
3. Емец О.А. Интервальная математическая модель комбинаторной задачи цветной упаковки прямоугольников / Емец О.А., Евсеева Л.Г., Романова Т.Е. // Кибернетика и системный анализ. – 2001. – № 3 – С. 131–138.
4. Стоян Ю.Г. Оптимизационная задача размещения правильных интервальных многоугольников / Стоян Ю.Г., Романова Т.Е., Сысоева Ю.А. // Докл. НАН Украины. Сер. А. – 1998. – № 9 – С. 114–120.
5. Гребенник И.В. Учет погрешностей при построении математических моделей оптимизационных комбинаторных задач / Гребенник И.В., Романова Т.Е. // АСУ и приборы автоматизации. – 2002. Вып. 119. – С. 64–69.
6. Kaucher E. Interval Analysis in the Extended Interval Space IR // Comp. Suppl. – 1980. – № 2. – P. 33–49.
7. Stoyan Yu. G. The extended interval space and elementary mappings // Proc. of the IMACS–GAMM Intern. Symp. on Numerical Methods and Error Bounds. (Oldenburg, Germany). – Oldenburg, 1995. – P. 270–279.
8. Стоян Ю.Г. Метрическое пространство централизованных интервалов // Докл. НАН Украины. Сер. А. – 1996. – № 7. – С. 23–25.
9. Стоян Ю.Г. Выпуклые интервальные многоугольники // Доп. НАН України. – 2000. – № 5. – С. 33–39.
10. Stoyan Yu. G. Analytical description of interactions of point sets // Journal of mechanical engineering – 2001. – № 1–2, v. 4 – С. 77–88.
11. Stoyan Y. Φ -function for 2D primary objects // Stoyan Y., Terno J., Scheithauer G., Gil N., Romanova T. // Studia Informatica, Paris, University. – 2002. – Vol. 2, № 1. – P. 1–32.
12. Stoyan Yu. Φ -function for primary 3D objects/ Stoyan Yu., Scheithauer G., Pridatko D., Romanova T.: Prepr./ Technische Universitat Dresden; MATH-NM-15-2002. – Dresden. 2002. – 27 p.
13. Stoyan Y. Φ -function for complex 2D objects/ Stoyan Y., Scheithauer G., Gil M., Romanova T. // 40R Quarterly Journal of the Belgian, French and Italian Operations Research Societies. Volume 2, Number 1, 2004 – P. 69–84.
14. Stoyan Yu. G. Φ -function and its basic properties // Докл. НАН України. Сер. А. – 2001. – № 8. – С. 112–117.
15. Stoyan Y. Construction of a Φ -function for two convex polytopes / Stoyan Y., Terno J., Gil M., Romanova T., Scheithauer G. // Applicationes Mathematicae. – 2002. – Vol. 2, № 29. – P. 199–218.
16. Романова Т.Е. Интервальное пространство I_s^n // Докл. НАН України. Сер. А. – 2000. – № 9. – С. 36–41.
17. Гребенник И.В. Интервальные модели комбинаторной оптимизации квазилинейных функций в пространстве I_s^n // Докл. НАН України. Сер. А. – 2004. – № 9 – С. 60–64.
18. Стоян Ю.Г. Теорія і методи евклідової комбинаторної оптимізації / Стоян Ю.Г., Емец О.О. – Київ: Інститут системних досліджень освіти, 1993. – 188 с.

Надійшла 18.05.2004
Після доробки 27.11.2004

Будується математична модель оптимізаційної задачі геометричного проектування в інтервальному просторі. Виділяються множини базових двовимірних та тривимірних інтервальних геометричних об'єктів. Введено поняття геометричної інформації щодо базових та складених інтервальних об'єктів. Формується область припустимих рішень оптимізаційної інтервальної задачі розміщення з використанням методу Φ -функцій.

Mathematical model of a basic optimization problem of geometric design is constructed in the interval space. A set of primary two and three dimensional interval geometric objects are given. A concept of geometric information on primary and complex interval objects is introduced. A feasible region of the basic optimization interval placement problem is formed by means of Φ -functions.

УДК 681.3.06

В. И. Долгов, А. В. Неласая

МЕТОДЫ УВЕЛИЧЕНИЯ СКОРОСТИ КРИПТОГРАФИЧЕСКИХ ПРЕОБРАЗОВАНИЙ НА ЭЛЛИПТИЧЕСКИХ КРИВЫХ

В статье проводится обзор методов оптимизации скорости вычислений при криптографических преобразованиях в группе точек эллиптической кривой. Предложено использовать методы параллельного программирования для повышения скорости прямых преобразований.

ВВЕДЕНИЕ

Бурное развитие информационных технологий, в том числе систем электронного документооборота, Internet-технологий наряду с огромными преимуществами порождает и определенные проблемы. Информация должна быть доступна только тем пользователям, для кото-

рых она предназначена. Разрушение информационного ресурса, его временная недоступность или несанкционированное использование могут нанести компании значительный материальный ущерб. Внедрение банками систем электронных платежей также предъявляет повышенные требования к целостности и подлинности электронных сообщений.

Одним из направлений защиты информации являются методы аутентификации сообщений. Информация считается аутентичной, когда потребитель имеет гарантии ее целостности и авторства. В процессе аутентификации объекта проверяется подлинность идентификатора, пред-

ставленного с некоторыми данными. Для аутентификации объекта применяется шифрование симметричным алгоритмом, выработка имитовставки или электронной подписи. Первые два способа применяются в том случае, когда отправитель и верификатор доверяют друг другу. Если необходимо иметь возможность доказательства подлинности идентификатора третьей стороне, например необходима юридическая значимость пересылаемых электронных документов, требуется цифровая подпись (ЦП).

Цифровая подпись основана на применении асимметричных алгоритмов и представляет собой добавленные к информации данные, вычисленные посредством криптографического преобразования на секретном ключе отправителя защищаемой информации и параметров, наличие которых позволяет удостовериться в целостности информации и подлинности ее источника, а также обеспечить защиту от подлога со стороны получателя. Проверка целостности и подлинности производится с использованием открытого ключа, причем знание открытого ключа не позволяет подделать ЦП с заданной вероятностью.

В основу алгоритмов цифровой подписи положены асимметричные криптографические преобразования, стойкость которых основана на сложности решения некоторых математических задач. Такими задачами, используемыми в настоящее время в криптографии, являются факторизация больших чисел, нахождение дискретного логарифма в кольцах и полях Галуа, а также проблема вычисления дискретного логарифма на эллиптической кривой.

Еще одной областью применения асимметричных алгоритмов является направленное шифрование, суть которого состоит в том, что информация шифруется на открытом ключе получателя, либо на ключе, полученном с помощью открытого ключа, а расшифровывается на секретном ключе получателя, либо на ключе, полученном с помощью секретного ключа.

Криптографические преобразования на эллиптических кривых на сегодня считаются наиболее стойкими среди всех известных асимметричных преобразований. Задача логарифмирования на кубической кривой является более сложной, чем задача дискретного логарифмирования в простом поле, так как вторая является частным случаем первой. Наилучший из известных на сегодня методов криптоанализа распределенный на r процессоров оптимальный алгоритм ρ -Полларда имеет сложность

$$I = \sqrt{\frac{\pi n}{4r^2}},$$

где n – порядок циклической подгруппы группы точек эллиптической кривой.

При правильном выборе кривой задача логарифмирования в группе простого порядка, выражаемом числом длиной 160 бит примерно соответствует сложности логарифмирования в мультипликативной группе простого поля, порядок которого выражается числом длиной 1024 бита [1]. Это позволяет уменьшить длину ключа и

соответственно увеличить скорость вычислений по сравнению с криптоалгоритмами RSA той же стойкости. Однако в условиях интенсивного развития вычислительной техники, а также совершенствования средств и методов криптоанализа для обеспечения требуемого уровня стойкости приходится непрерывно увеличивать длину модуля преобразований и ключей, что приводит к снижению производительности систем шифрования. Поэтому актуальной является задача оптимизации скорости прямых криптографических преобразований на эллиптических кривых.

1 ЭЛЛИПТИЧЕСКИЕ КРИВЫЕ В КРИПТОГРАФИИ. ОБЩИЕ СВЕДЕНИЯ

В криптографии применяется редуцированная форма эллиптической кривой, определенной над конечным полем Галуа $GF(q)$, где модуль q является либо простым числом, либо степенью простого числа. В настоящее время наиболее часто употребляются кривые над простым полем $GF(p)$, где p – большое простое число [2, 5] и кривые над полем $GF(2^m)$ [3, 5].

Эллиптическая кривая (ЭК) над простым полем $GF(p)$ в аффинных координатах задается уравнением:

$$y^2 = (x^3 + ax + b) \pmod{p}, \quad (1)$$

где p – простое число, являющееся модулем преобразований; (x_i, y_i) – точки ЭК. Точка принадлежит ЭК, если удовлетворяет условию (1); a, b – параметры ЭК, причем должно выполняться условие

$$4a^3 + 27b^2 \neq 0 \pmod{p}, \quad (2)$$

которое обеспечивает гладкость кривой.

Эллиптическая кривая над расширенным полем $GF(2^m)$ в аффинных координатах имеет вид:

$$y^2 + xy = (x^3 + ax^2 + b) \pmod{f(x), 2}, \quad (3)$$

где координаты точек ЭК x_i, y_i и коэффициенты $a, b \in GF(2^m), f(x)$ – примитивный полином над полем $GF(2^m)$.

Во втором случае приведение осуществляется по двойному модулю $(f(x), 2)$.

Кроме точек (x_i, y_i) , удовлетворяющих соответственно уравнениям (1) или (3), в множество точек ЭК также включается точка на бесконечности, обозначаемая как O . На таком множестве можно определить операцию сложения [5], обладающую следующими свойствами:

1. $P + O = O + P = P$ для любой точки P , принадлежащей ЭК.

2. Для любой точки P существует такая точка Q , что $P + Q = O$. Точка Q является обратным элементом по отношению к P . Для простого поля (1) взаимно обратными являются точки с координатами (x, y) и $(x, -y)$, для расширенного поля (3) – точки с координатами (x, y) и $(x, x+y)$.

3. Для любых точек P и Q существует точка $R = P + Q = Q + P$, принадлежащая ЭК.

В криптографии используется тот факт, что множество точек ЭК с операцией сложения, удовлетворяющей отмеченным выше свойствам, образует абелеву группу.

Основой криптографических преобразований является выполнение операции скалярного умножения в группе точек эллиптической кривой. Для ЭК над простым полем [2] сумма двух различных точек $P_1 = (x_1, y_1)$ и $P_2 = (x_2, y_2)$ в аффинных координатах определяется следующим образом:

$$P_3 = P_1 + P_2 = (x_1, y_1) + (x_2, y_2) = (x_3, y_3); \quad (4)$$

$$x_3 = (\lambda^2 - x_1 - x_2) \bmod p; \quad (5)$$

$$y_3 = (\lambda(x_1 - x_3) - y_1) \bmod p; \quad (6)$$

$$\lambda = \frac{y_2 - y_1}{x_2 - x_1} \pmod{p}. \quad (7)$$

Сумма двух одинаковых точек (удвоенная точка $P_1 = (x_1, y_1)$) для $x_1 \neq 0$ равна:

$$P_3 = 2P_1 = (x_3, y_3). \quad (8)$$

В этом случае $x_1 = x_2$, $\lambda = \frac{3x_1^2 + a}{2y_1} \pmod{p}$, а x_3 и y_3 вычисляются по формулам (5) и (6) соответственно.

Для ЭК над расширенным полем [3], координаты суммы двух различных точек $P_1 = (x_1, y_1)$ и $P_2 = (x_2, y_2)$ в аффинных координатах определяются следующим образом:

$$x_3 = \left(\frac{y_1 + y_2}{x_1 + x_2} \right)^2 + \frac{y_1 + y_2}{x_1 + x_2} + x_1 + x_2 + a \pmod{f(x, 2)}; \quad (9)$$

$$y_3 = \left(\frac{y_1 + y_2}{x_1 + x_2} \right) (x_1 + x_3) + x_3 + y_1 \pmod{f(x, 2)}; \quad (10)$$

Координаты суммы двух одинаковых точек (удвоенная точка $P_1 = (x_1, y_1)$) для $x_1 \neq 0$ равны:

$$x_3 = x_1^2 + \frac{b}{x_1}; \quad (11)$$

$$y_3 = x_1^2 + \left(x_1 + \frac{y_1}{x_1} \right) x_3 + x_3. \quad (12)$$

При этом в обоих случаях, если $P_1 = -P_2$, то $P_3 = 0$. При удвоении точки с нулевой x координатой результатом будет точка O .

Порядком группы точек ЭК называется число точек ЭК, включая бесконечно удаленную точку. Если n – простое нечетное число, которое делит порядок ЭК, $nP = 0$ и $kP \neq 0$ для $0 < k < n$, то точка P порождает циклическую подгруппу порядка n и называется базовой точкой порядка n . При этом любую точку этой подгруппы можно представить как кратную базовой точке:

$$Q = d \times P = \underbrace{P + P + \dots + P}_d. \quad (13)$$

В криптографии Q считается открытым ключом, d – секретным. Основой стойкости криптографических пре-

образований на эллиптических кривых является большая вычислительная сложность задачи нахождения числа d для известных P и Q при $n \rightarrow \infty$ в циклическом модуле простого порядка, которая называется проблемой дискретного логарифма для ЭК. Если порядок циклического модуля составной, то сложность логарифмирования определяется максимальным простым делителем порядка модуля.

Стандартным является следующий алгоритм для вычисления открытого ключа (скалярного произведения целого числа на базовую точку):

Алгоритм. Бинарное скалярное умножение на эллиптической кривой [3, 8].

Исходные данные: число $d \neq 0$, точка P , эллиптическая кривая $E = \langle a, b, p \rangle$.

Результат: точка $Q = d \times P$.

1. Если $d = 1$, то $Q := P$; закончить работу алгоритма.
2. $k := l_d - 2$; $Q := P$, где l_d – длина числа d в битах;
3. Для i , принимающего значения от k до 0, выполнить шаги 4–5.

4. $Q := Q + Q$.

5. Если i -й бит d равен 1, то $Q := Q + P$.

В работе [8] приведены также алгоритм блочного скалярного умножения на эллиптической кривой и параллельного блочного умножения на эллиптической кривой и дан анализ их вычислительной сложности. Показано, что арифметические операции на эллиптической кривой выполняются медленнее, чем соответствующие операции над $GF(p)$ при одинаковой длине p . Однако алгоритмы на эллиптических кривых обладают большей стойкостью, что приводит к значительному уменьшению длины ключевой информации и следовательно к повышению производительности систем защиты.

6. Закончить работу алгоритма.

2 МЕТОДЫ ОПТИМИЗАЦИИ ВЫЧИСЛЕНИЙ НА ЭЛЛИПТИЧЕСКИХ КРИВЫХ

В протоколах цифровой подписи [2–4] и направленного шифрования на эллиптических кривых основной является операция умножения точки на скаляр, реализуемая посредством последовательных сложений у удвоенных точек ЭК. Кратко рассмотрим сущность основных подходов к оптимизации этих операций и оценим их возможности.

При самом общем рассмотрении можно отметить, что реализация криптопротоколов на эллиптических кривых, предусматривает:

1. Выполнение модульных операций над длинными целыми числами для кривых над простым полем.
2. Выполнение операций над элементами основного поля (модульных операций над многочленами) для расширенного поля.
3. Выполнение операций над точками эллиптической кривой.

Кроме использования специальных методов вычислений, о которых будет сказано ниже, большой выигрыш дает аппаратная оптимизация для конкретных типов процессоров.

2.1 Базисные представления элементов поля $GF(2^m)$

На уровне элементов поля на скорость вычислений влияет выбор базиса для представления элементов поля. Элементы расширенного поля Галуа $GF(2^m)$ имеют два основных базисных представления – полиномиальное и нормальное [3, 9].

Каждый элемент поля $GF(2^m)$ в полиномиальном базисе представляется как двоичный полином степени не выше $(m - 1)$:

$$b = a_{m-1}x^{m-1} + a_{m-2}x^{m-2} + \dots + a_1x + a_0.$$

При этом элементы $x^{m-1}, x^{m-2}, \dots, x, 1$, где x – корень неприводимого многочлена $f(t)$ степени m , образуют базис конечного поля $GF(2^m)$ как векторного пространства над полем $GF(2)$.

Элементы $GF(2^m)$ при этом представляются битовыми строками длиной m битов $(a_{m-1}a_{m-2}\dots a_1a_0)$. Единичный элемент по умножению задается битовой строкой $(00\dots001)$. Нулевой элемент (0) задается битовой строкой из всех 0. Сложение элементов битового поля соответствует сложению по модулю 2 битовых строк. Произведение полиномов приводится по двойному модулю $(f(x), 2)$.

Удобно задавать полиномиальный базис примитивным многочленом. При этом вычислительная сложность модульных операций зависит от количества ненулевых коэффициентов в примитивном полиноме. Стандартом [3] рекомендовано использовать трехчлен $f(t) = t^m + t^k + 1, 0 < k < m$ или пятичлен (так как не всегда существует трехчлен заданной степени m) $f(t) = t^m + t^l + t^j + t^k + 1, 0 < k < j < l < m$ указанных степеней. Рекомендованные примитивные многочлены приведены в стандарте [3].

При использовании нормального базиса элементы поля $GF(2^m)$ представляются следующим образом:

$$b = a_{m-1}x^{m-1} + a_{m-2}x^{m-2} + \dots + a_1x + a_0,$$

при этом элементы $\{x^{2^{m-1}}, x^{2^{m-2}}, \dots, x^2, x\}$ должны быть линейно независимыми над $GF(2)$. Нормальный базис существует для каждого конечного поля характеристики 2.

Если число $p = 2m + 1$ – простое и для наименьшего натурального числа k , такого что $2^k \equiv 1 \pmod{p}$ выполняется $p \equiv 1 \pmod{4}$ и $k = 2m$ либо $p \equiv 3 \pmod{4}$ и $k = m$, то нормальный базис называется гауссовым оптимальным нормальным базисом типа 2. Он существует не для всех конечных полей характеристики 2 [3]. Стандарт [3] допускает представление в полиномиальном и гауссовом оптимальном нормальном базисах.

Как показано в [9] для минимизации вычислительной сложности при $m < 129$ более предпочтительным является гауссов оптимальный нормальный базис, а при $m > 129$ – полиномиальный базис. А поскольку в преобразованиях на ЭК используется $m > 160$, то в целом более предпочтительным оказывается использование полиномиального базиса.

2.2 Сложение в различных координатных базисах

Другой способ повышения скорости вычислений – представление точек кривой в различных типах координатных базисов, отличных от аффинных. Основными из них являются проективные координаты, якобиановые координаты, координаты Чудновского и модифицированные якобиановые координаты [10].

Например, в случае представления кривой в проективных координатах значения преобразуются по формулам [11]:

$$x = \frac{X}{Z}, \quad y = \frac{Y}{Z}.$$

Тогда

$$q(x, y) = Q\left(\frac{X}{Z}, \frac{Y}{Z}, 1\right)$$

и обратно,

$$Q(X, Y, Z) = Z^3q\left(\frac{X}{Z}, \frac{Y}{Z}\right).$$

В этом случае $(X, Y, Z) = (\lambda X, \lambda Y, \lambda Z)$.

Кривая над простым полем в проективных координатах имеет вид:

$$Y^2Z = (X^3 + aXZ^2 + bZ^3) \pmod{p}$$

В отличие от вычислений в аффинных координатах во всех остальных случаях наиболее вычислительно трудоемкая операция деления в поле не используется. Вместо этого увеличивается количество умножений.

В [10, 12] показано, что наиболее вычислительно трудоемкими являются вычисления в аффинном базисе из-за присутствия в алгоритмах сложения и удвоения точек операции деления в поле. Повысить производительность криптопреобразований в группах точек ЭК позволяет также использование смешанных координат.

2.3 Методы повышения скорости операции умножения

Существуют различные методы повышения скорости операции умножения. Среди них можно выделить умножение с помощью дискретного преобразования Фурье [11, 13], алгоритм умножения Карацубы-Оффмана [11] и арифметику Монтгомери [11].

Пусть $K[x]$ – кольцо многочленов над полем K и $f(x) = \sum_{i=0}^{n-1} a_i x^i \in K[x]$ – многочлен из этого кольца. Кроме общепринятого задания многочлена n коэффициентами $a_i, i = \overline{0, n-1}$ можно однозначно задать его множеством значений в n точках X_i . При этом для удобства вычислений в качестве x_i можно принять множество корней степени n из 1 в поле K . В этом случае многочлены можно умножать путем "покоординатного" умножения.

Элемент коммутативного кольца, обладающий свойствами

$$\begin{aligned} & 1) \omega \neq 1; \quad 2) \omega^n = 1; \\ & 3) \sum_{j=0}^{n-1} \omega^{jp} = 0 \text{ для } 1 \leq p < n \end{aligned}$$

называется примитивным корнем n -й степени из единицы. Элементы $\omega^0, \omega^1, \dots, \omega^{n-1}$ называются корнями n -й степени из единицы.

Дискретным преобразованием Фурье называется отображение вектора коэффициентов многочлена степени не выше n в вектор значений, определяемых матрицей $(n \times n)$ вида $A[i, j] = \omega^{ij}$ для $0 \leq i, j < n$, где ω – примитивный корень n -й степени из единицы. При этом обратное преобразование осуществляется путем умножения вектора значений на обратную матрицу вида $A^{-1}[i, j] = \left(\frac{1}{n}\right) \omega^{-ij}$ для $0 \leq i, j < n$.

Чтобы перемножить два многочлена, заданных своими значениями в корнях n -й степени из 1, достаточно перемножить пары их значений в соответствующих корнях. Для вычисления коэффициентов многочлена – произведения необходимо выполнить над ним обратное преобразование. То есть $a \otimes b = F^{-1}(F(a)F(b))$.

Если $n = 2^k$, то произведения $y = Ax$ и $x = A^{-1}$ удобно вычисляются. В этом случае преобразование называется быстрым преобразованием Фурье (БПФ). Алгоритм вычисления БПФ приведен в [11, 13].

БПФ также можно использовать для умножения больших чисел [11]. Пусть a, b – n -разрядные двоичные числа от 0 до $2^n - 1$, которые нужно перемножить по модулю $2^{2n} + 1$. Пусть $n = 2^k$, каждое число состоит из d блоков длины $l = n/d$ бит. Для этого для каждого сомножителя выполняется БПФ так, чтобы образы имели двойную длину (защита от переполнения). Затем Фурье-образы покоординатно перемножаются по модулю $2^{2l} + 1$. После этого выполняется обратное БПФ.

Алгоритм умножения Карацубы-Оффмана сводит умножение двух $2n$ -значных чисел к трем умножениям двух n -значных чисел [11].

$2n$ -значное число представляется в виде $A_n + 10^n B_n$. Тогда выполняется тождество:

$$\begin{aligned} & (A_n + 10^n B_n) \cdot (C_n + 10^n D_n) = \\ & = A_n \cdot C_n (10^n + 1) + (D_n - C_n) \cdot (A_n - B_n) \cdot 10^n + \\ & \quad + B_n \cdot D_n (10^{2n} + 10^n). \end{aligned}$$

Для умножения двух многочленов степени не более $2m - 1$ из кольца $F_p[x]$ их необходимо представить в виде:

$$U(x) = U_1(x)x^m + U_0(x); \quad V(x) = V_1(x)x^m + V_0(x),$$

где степень каждого из многочленов $U_1(x), U_0(x), V_1(x), V_0(x)$ не более $m - 1$. Тогда выполняется тождество:

$$UV = U_1 V_1 x^{2m} + ((U_1 + U_0)(V_1 + V_0) - U_1 V_1 - U_0 V_0) x^m + U_0 V_0.$$

Операция умножения многочлена на x^m выполняется сдвигом массива коэффициентов и имеет линейную сложность. Следовательно, вычислительно трудоемкими являются лишь три операции умножения многочленов степени $m - 1$.

В арифметике Монтгомери [11], при вычислении произведения двух длинных чисел A и B по модулю

нечетного числа N , множители представляются своими N -вычетами AR, BR по модулю N . Здесь R – число, равное степени двойки и превышающее N . Предложенный П. Монтгомери метод модульного умножения не требует операции деления. Произведение Монтгомери по модулю N :

$$MP(A, B, N, R) = ABR^{-1} \pmod{N}.$$

В основе вычисления произведения Монтгомери лежит преобразование Монтгомери:

$$TR^{-1} \pmod{N} = \frac{T + MN}{R} \text{ или } TR^{-1} \pmod{N} = \frac{T + MN}{R} - N,$$

где N и R – взаимно простые положительные целые числа, $N' \equiv -N^{-1} \pmod{R}$ и $M \equiv TN' \pmod{R}$. Деление на R выполняется путем сдвига на соответствующее число разрядов.

Для перевода числа в N -вычет необходимо вычислить $MP(A, R^2 \pmod{N}, N, R)$. Обратное преобразование N -вычета в число нужно вычислить $MP(AR \pmod{N}, 1, N, R)$. В [12, 20] приведены алгоритмы вычисления преобразования Монтгомери для многоразрядного числа и модульного умножения больших чисел. Множество N -вычетов с операциями умножения по Монтгомери и сложения является кольцом, изоморфным кольцу классов вычетов по модулю N [12].

2.4 Оптимально расширенные поля

В [14] предложено определенным образом выбирать вид основного поля $GP(p^m)$, где p – простое, m – положительное целое, с целью уменьшения вычислительной сложности операций в нем. Такое оптимально расширенное поле соответствует следующим требованиям:

1) p выбирается возможно большим, но таким, чтобы оно помещалось в разрядную сетку машинного слова. Это позволяет с максимальной эффективностью использовать быстрые процессорные команды;

2) p выбирается таким, чтобы оно было псевдопростым числом Мерсенна (вида $2^n \pm c$ где $\log_2(c) \leq \frac{1}{2}n$). Это позволяет выполнять операцию модульного преобразования над элементами поля с минимальной вычислительной сложностью;

3) m выбирается таким, чтобы в поле существовал несократимый бином вида $x^m - \omega$ для эффективного модульного преобразования в расширенном поле. Степень расширения m может быть тем меньше, чем большим выбрано p .

Элементы такого оптимально расширенного поля в машине представляются с помощью m машинных слов. Благодаря тому, что разрядность элемента $GP(p)$ не превышает разрядности машинного слова, умножение в подполе $GP(p)$ выполняется одной машинной инструкцией, а модульное преобразование по модулю вида 2^n можно выполнить без вычислительно трудоемкой операции деления. Поскольку вычислительная сложность модульного преобразования тем меньше, чем меньше

число членов в полиноме поля, требование 3) обеспечивает уменьшение вычислительной сложности модульного умножения в расширенном поле $GP(p^m)$. Как показано в [14], использование оптимально расширенных полей обеспечивает более высокую производительность по сравнению с блочной арифметикой Монтгомери.

2.5 Комплексное умножение

Эллиптические кривые над алгебраическим замыканием конечного поля имеют тривиальные автоморфизмы, заключающиеся в умножении точки на целое число. В этом случае для отображения $\psi_n: P \rightarrow nP$ справедливо [11]

$$\psi_n(P) + \psi_n(Q) = nP + nQ = n(P + Q) = \psi_n(P + Q).$$

Таким образом, множество автоморфизмов кривой всегда содержит множество целых чисел Z . В случае, когда множество автоморфизмов строго больше Z говорят, что кривая обладает комплексным умножением.

Пусть на кривой $E(GF(q))$ над полем из $q = p^n$ элементов определено отображение

$$\phi: (x, y) \rightarrow (f(x), h(x, y)), \quad (14)$$

где f, h – рациональные функции, такое, что если $(x, y) \in E$, то и $\phi(x, y) \in E$ и $\phi(0) = 0$. По определению это отображение является эндоморфизмом кривой. Пусть группа точек кривой содержит циклическую подгруппу G простого порядка r , причем $\#E(GF(q))$ не делится на r^2 . В случае, когда группа G под действием отображения (14) переходит в себя, то есть когда отображение является автоморфизмом, $\phi(P) = eP$ для некоторого $0 < e < r$. Такое отображение задает комплексное умножение на кривой $E(GF(q))$. Число e легко вычислимо [15].

Важным примером таких отображений является эндоморфизм Фробениуса, определяемый как $\phi: (x, y) \rightarrow (x^q, y^q)$. По определению он переводит точку O в O . Если $\#E(GF(q))$ не делится на r^2 , эндоморфизм Фробениуса имеет собственные значения. Тогда имеет место равенство

$$\phi(P) = eP,$$

где e – собственное значение эндоморфизма Фробениуса [11, 16].

Комплексное умножение для быстро вычисляемых отображений позволяет ускорить процедуру умножения точки на число. Для вычисления произведения $Q = dP$ множитель d представляют в виде

$$d = d_0 + d_1 e,$$

где $|d_0|, |d_1| < \sqrt{r}$.

Доказано [28], что такое представление всегда существует. Затем последовательно вычисляют $d_0 P, d_1 P, e(d_1 P), d_0 P + e(d_1 P)$. Это позволяет ускорить процедуру умножения примерно вдвое [15], при этом произведение точки на e легко вычисляется с помощью комплексного умножения.

Однако комплексное умножение можно также использовать и для снижения стойкости криптоалгоритма, так

как оно задает разбиение конечных точек циклической группы G на классы эквивалентности. Тогда алгоритм Полларда для логарифмирования в произвольной циклической группе вычислимого порядка, основанный на задаче о встрече на графе случайного отображения, можно использовать не для точек, а для классов. В этом случае сложность логарифмирования на кривых с комплексным умножением в \sqrt{m} меньше чем обычно. Однако такое снижение сложности представляется безопасным [16].

ЗАКЛЮЧЕНИЕ

В настоящее время для криптографических целей используются кривые над простым полем или над полем характеристики 2. Арифметические операции в простом поле выполняются достаточно долго. В поле характеристики 2 вычисления проводятся быстрее, особенно при аппаратной реализации, однако неизоморфных кривых над полем характеристики 2 с требуемыми для криптографического приложения свойствами существует довольно немного [1]. Примеры допустимых кривых приведены в стандарте [3]. Показано [16, 17], что операции сложения и умножения в поле характеристики $q = 2^{16} + 1$ микропроцессор выполняет быстро, количество неизоморфных кривых, пригодных для построения криптосистем, исчисляется тысячами для каждой степени расширения. Существуют и другие "удобные" характеристики. Однако в современных стандартах эти результаты еще не нашли отражения.

Основной проблемой эллиптической криптографии, как и вообще всех асимметричных алгоритмов является невысокая скорость криптографических преобразований, хотя на сегодняшний день это наилучший выбор по соотношению "скорость-стойкость". В работе рассмотрены существующие методы повышения скорости вычислений на различных уровнях реализации криптопротоколов на эллиптических кривых.

Одним из резервов повышения быстродействия компьютерных вычислений является параллельная обработка данных на многопроцессорных вычислительных комплексах. Эта техника давно с успехом применяется для криптоанализа. Представляет интерес задача разработки параллельных алгоритмов для прямых преобразований на эллиптических кривых. В этом случае необходимо также уяснить, какое максимальное количество процессоров имеет смысл использовать и какие из приведенных выше методов оптимизации вычислений удастся распараллелить.

ПЕРЕЧЕНЬ ССЫЛОК

1. Ростовцев А.Г. О выборе эллиптической кривой над простым полем для построения криптографических алгоритмов. // Проблемы информационной безопасности. Компьютерные системы. №3, 1999. С. 37–40.
2. ГОСТ Р 34.10–2001. Государственный стандарт Российской Федерации. Информационная технология. Криптографическая защита информации. Процессы формирования и проверки цифровой подписи. М.: Госстандарт России, 2001.

3. ДСТУ 4145–2002. Державний стандарт України. Інформаційні технології. Криптографічний захист інформації. Цифровий підпис, що ґрунтується на еліптичних кривих. Формування та перевірка. Київ:–Держстандарт України, 2003.
4. Don Johnson, Alfred Menezes, Scott Vanstone. The Elliptic Curve Digital Signature Algorithm (ECDSA)// Certicom Research, Canada, 2001.
5. J.H. Silverman "The Arithmetic of Elliptic Curve", GTM 106, Springer-Verlag, New York, 1986
6. Koblitz N. Elliptic Curve Cryptosystems // Mathematics of Computation. 1987. Vol. 48, № 177.Р.203–209.
7. Штанько С.В. Эллиптические кривые в криптографии // Проблемы информационной безопасности. Компьютерные системы. № 2, 2003, С. 65–74.
8. Качко Е.Г., Свиначев А.В., Головашич С.А. Методы и алгоритмы ускорения вычислений в несимметричных преобразованиях на эллиптических кривых. // Радиотехника: Всеукр. межвед. науч.-техн. сб. 2000, Вып. 114. С. 69–74.
9. Горбенко И.Д., Збитнев С.И. Расширенное поле Галуа. Вычислительная сложность простейших операций над расширенным полем. // Радиотехника: Всеукр. межвед. науч.-техн. сб. 2000, Вып. 114. С. 80–89.
10. Збитнев С.И. Проективная геометрия – не все так гладко. / Радиотехника: Всеукр. межвед. науч.-техн. сб. 2003. 9 с.
11. Ростовцев А.Г. Алгебраические основы криптографии. – СПб.: НПО "Мир и семья", ООО "Интерлайн", 2000. – 354 с.
12. Горбенко И.Д., Збитнев С.И., Поляков А.А. Сложность арифметических операций в группах точек эллиптических кривых для криптографических операций. // Радиотехника: Всеукр. межвед. науч.-техн. сб. 2001. Вып. 119. 5 с.
13. Лавриненко Д.И. Применение быстрого преобразования Фурье в криптографических преобразованиях. // Радиотехника: Всеукр. межвед. науч.-техн. сб. 2000, Вып. 114. С. 75–79.
14. Лавриненко Д.И. Оптимально расширенные поля в алгоритмах для эллиптических кривых..// Радиотехника: Всеукр. межвед. науч.-техн. сб. 2001. Вып. 119. С. 56–61.
15. Ростовцев А.Г., Буренкова А.П., Маховенко Е.Б. О комплексном умножении на эллиптических кривых. // Проблемы информационной безопасности. Компьютерные системы. №1, 1999. С. 90–91.
16. Маховенко Е.Б., Ростовцев А.Г. Быстрая арифметика для эллиптических кривых над расширенными полями. // Проблемы информационной безопасности. Компьютерные системы. №2, 1999. С. 50–54.
17. Маховенко Е.Б. Эллиптические кривые над полем характеристики $2^{16} + 1$ с хорошими криптографическими свойствами. // Безопасность информационных технологий. №3, 1997. С. 56–58.

Надійшла 02.04.2004
Після доробки 08.10.2004

В статті проведено огляд методів оптимізації швидкості обчислень при криптографічних перетвореннях в групі крапок еліптичної кривої. Запропоновано використовувати методи паралельного програмування для збільшення швидкості прямих перетворень.

It is conducted the review of methods of optimization of computation rate in cryptographic transformations in the cluster of points of elliptical curve in the article. It was offered to use methods of multiple programming for increasing the rate of direct transformations.

УДК 681.32:007.52

В. И. Дубровин

МЕТОДЫ СОКРАЩЕНИЯ ОБЪЕМА ДАННЫХ

Рассмотрено решение задачи сокращения объема данных при построении математических моделей сложных объектов и процессов путем уменьшения количества признаков за счет исключения малозначимых признаков и уменьшения обучающего множества за счет исключения избыточных примеров. Описана программная реализация и приведены результаты применения предложенных методов при решении практических задач.

ВВЕДЕНИЕ

Управление качеством представляет собой область исследований, находящую широкое применение в самых различных сферах: в проектировании, производстве и эксплуатации изделий, в биологии, медицине и сельском хозяйстве, в экономике и социологии.

При построении автоматических и автоматизированных систем управления качеством, в частности, систем технической и биомедицинской диагностики, возникает задача разработки адекватных математических моделей сложных объектов и процессов.

На практике, как правило, приходится иметь дело с многомерными нелинейными объектами. Сбор, хранение и обработка информации о таких объектах представляет собой достаточно трудоемкую и ресурсоемкую задачу. Поэтому при построении математических моделей возникает необходимость сокращения объема данных, характеризующих моделируемые объекты, таким образом,

чтобы минимизировать количество исходных данных для моделирования и одновременно обеспечить требуемый уровень точности модели [1–5].

Для решения этой задачи в настоящей работе предлагается использовать два подхода:

- уменьшение размерности модели за счет исключения малозначимых признаков путем оценки информативности и отбора признаков;

- уменьшение обучающего множества за счет исключения избыточных экземпляров, что позволяет перейти от массива, содержащего информацию по всем n статистически обследованным объектам, к соответствующей информации по k эталонным образцам ($k \ll n$), где в качестве эталонных берутся специальным образом отобранные наиболее типичные представители классов объектов, полученные в результате разбиения множества объектов на однородные группы.

1 МЕТОДЫ ОЦЕНКИ ИНФОРМАТИВНОСТИ И ОТБОРА ПРИЗНАКОВ

Известно большое количество методов, позволяющих оценивать информативность параметров (признаков). Однако ни один из них по отдельности не способен обеспечить получение адекватного результата для всех практических задач. Поэтому представляется актуальным ис-

пользовать комбинацию различных методов для получения интегрированной оценки информативности параметров (признаков).

Рассмотрим основные методы оценки информативности признаков. Обозначим: S – количество экземпляров в выборке; s – номер текущего экземпляра выборки; N – количество признаков; y^s – целевое значение, сопоставленное s -му экземпляру (номер класса или числовая характеристика).

$$I_i^A = \frac{\sum_{s=1}^S (x_i^s - \bar{x}_i)(y^s - \bar{y})}{\sqrt{\sum_{s=1}^S (x_i^s - \bar{x}_i)^2 \sum_{s=1}^S (y^s - \bar{y})^2}}, \quad i = 1, \dots, N,$$

где x_i^s – значение i -го признака s -го экземпляра выборки, \bar{x}_i , \bar{y} – средние арифметические значений i -го признака и целевого параметра, соответственно.

Весьма перспективной для оценки информативности признаков является методика, основанная на использовании искусственных нейронных сетей (НС) [6, 7], которые позволяют строить точные модели сложных объектов и процессов на основе многомерных наборов точечных данных, характеризующих данные объекты и процессы.

Наиболее простой моделью НС является однослойный перцептрон, который преобразует входной вектор признаков x в выходную скалярную величину y в соответствии с формулой:

$$y = \Psi \left(w_0 + \sum_{i=1}^N w_i x_i \right),$$

где x_i – значение i -го признака экземпляра, подаваемое на i -й вход перцептрона, w_0 – порог нейрона, w_i – вес i -го входа нейрона, Ψ – функция активации, которая является монотонной и ограниченной. В частности, при отрицательных выходах нейрона такой функцией может быть сигмоидная функция $\Psi(x) = 1/(1 + e^{-x})$ [6, 7].

Для того, чтобы осуществить построение модели на основе перцептрона, его необходимо обучить путем предъявления набора примеров. Процесс обучения перцептрона заключается в итерационной настройке его весовых и пороговых коэффициентов. Обучение перцептронов традиционно осуществляют на основе итеративных алгоритмов [6, 7].

После обучения перцептрона его весовые коэффициенты будут определять вклад каждого признака в выходную моделируемую величину. Вычисление относительной *нейросетевой информативности* i -го признака I_i^B предлагается производить по формуле:

$$I_i^B = \frac{|W_i|}{\sum_{j=1}^N |W_j|}, \quad i = 1, \dots, N.$$

Информативность по минимуму энтропии i -го признака, дискретизированного на z интервалов, I_i^F оценивают по формуле [8, 9]:

$$I_i^F = -\sum_z p_z \log_2 p_z,$$

где p_z – вероятность попадания признака x_i в z -й интервал диапазона его изменения.

Рассмотренные методы позволяют оценивать информативность признаков для построения количественных моделей. Однако на практике в большинстве случаев достаточно иметь качественную модель, что, например, имеет место в технической и биомедицинской диагностике. Для этого диапазон значений целевого параметра разбивается (дискретизируется) на заданное количество интервалов, а качественная оценка этого параметра указывает номер интервала, к которому относится экземпляр. В этом случае признаки будут тем информативнее, чем больше они несут информации для разделения экземпляров на классы, соответствующие определенным интервалам целевого параметра. Расчет информативности признаков для дискретизированного целевого параметра предлагается осуществлять на основе следующих методов.

Усредненный модуль коэффициента корреляции i -го признака и целевого параметра, дискретизированного на q интервалов, I_i^C может быть определен следующим образом. Последовательно в цикле для $j = 1, \dots, q$ находят частные коэффициенты корреляции для j -го интервала целевого параметра аналогично I_i^A с единственным отличием: y^s определяет не абсолютное значение целевого параметра, а принадлежность данного экземпляра к j -му интервалу целевого параметра ($y^s = 0$, если значение целевого параметра экземпляра не относится к j -му интервалу целевого параметра, $y^s = 1$ – в противном случае). После чего находят усредненную сумму модулей частных коэффициентов корреляции для всех $j = 1, \dots, q$.

Усредненная нейросетевая информативность i -го признака для целевого параметра, дискретизированного на q интервалов, I_i^E может быть определена следующим образом. Последовательно в цикле для $j = 1, \dots, q$ находят частные нейросетевые информативности для j -го интервала целевого параметра аналогично I_i^B с единственным отличием: y определяет не абсолютное значение целевого параметра, а принадлежность данного экземпляра к j -му интервалу целевого параметра ($y = 0$, если значение целевого параметра экземпляра не относится к j -му интервалу целевого параметра; $y = 1$ – в противном случае). После чего находят усредненную сумму частных нейросетевых информативностей для всех $j = 1, \dots, q$.

Усредненную статистическую информативность i -го признака для целевого параметра, дискретизированного на q интервалов, I_i^D определяют по формуле [6]:

$$I_i^D = \sum_{z=1}^q I_{i,z}^D,$$

где

$$I_{i,z}^D = \frac{(\bar{X}_i^A - \bar{X}_i^B)^2}{\sigma^2}, \quad \text{если } \sigma = \sigma_{X_i^A} = \sigma_{X_i^B};$$

$$I_{i,z}^D = 0,5(\sigma_{x_i^A} - \sigma_{x_i^B}) \left(\frac{1}{(\sigma_{x_i^B})^2} - \frac{1}{(\sigma_{x_i^A})^2} \right) +$$

$$+ 0,5 \left(\frac{1}{(\sigma_{x_i^A})^2} + \frac{1}{(\sigma_{x_i^B})^2} \right) (\bar{x}_i^A - \bar{x}_i^B), \text{ если } \sigma_{x_i^A} \neq \sigma_{x_i^B},$$

где \bar{x}_j – среднее значение i -го признака, σ_{x_j} – среднеквадратическое отклонение i -го признака.

Индексы A и B подразумевают вычисление соответствующей величины только для экземпляров соответствующего класса (т. е. экземпляров, относящихся или не относящихся к z -му интервалу).

Усредненные эвристические оценки информативности признаков для целевого параметра, дискретизированного на q интервалов I_i^G определяются следующим образом. Последовательно в цикле для $j = 1, \dots, q$ находят частные эвристические оценки информативности признаков для j -го интервала целевого параметра. После чего для всех признаков находят суммы частных эвристических оценок информативностей для всех $j = 1, \dots, q$.

Частные эвристические оценки информативности признаков для j -го интервала целевого параметра определяют путем подсчета N_j , $i = 1, \dots, N$ – количества разбиений на интервалы с одинаковым номером класса ($y = 0$, если значение целевого параметра экземпляра не относится к j -му интервалу целевого параметра; $y = 1$ – в противном случае) диапазона изменения каждого i -го признака.

Значение частной эвристической оценки информативности i -го признака для j -го интервала целевого параметра принимают равным $\min(N_j)/N_i$, $j = 1, \dots, N$.

Диагностическая ценность i -го признака, дискретизированного на m интервалов, для целевого параметра, дискретизированного на q интервалов, I_i^H определяется по формуле [3]:

$$I_i^H = 1 + \sum_{j=1}^q \sum_{z=1}^m P(y_j) P(x_{iz}/y_j) \frac{P(x_{iz}/y_j)}{P(x_{iz})},$$

где $P(y_j)$ – вероятность того, что экземпляр относится к j -му интервалу целевого параметра, $P(x_{iz})$ – вероятность попадания i -го признака в z -й интервал, $P(x_{iz}/y_j)$ – вероятность попадания i -го признака в z -й интервал, при условии, что экземпляр попадает в j -й интервал целевого параметра.

Информативность по количеству вносимой информации для i -го признака, дискретизированного на m интервалов, для целевого параметра, дискретизированного на q интервалов, I_i^I определяется по формуле [3]:

$$I_i^I = 1 + \sum_{j=1}^q P(y_j) \sum_{k=1}^q \sum_{z=1}^m P(x_{iz}/y_i) P(y_k/x_{iz}) \log P(y_i/x_{iz}),$$

где $P(y_j)$ – вероятность того, что экземпляр относится к j -му интервалу целевого параметра, $P(x_{iz})$ – вероятность попадания i -го признака в z -й интервал, $P(x_{iz}/y_j)$ – вероятность попадания i -го признака в z -й интервал, при условии, что экземпляр попадает в j -й интервал целевого

параметра, $P(y_j/x_{iz})$ – вероятность того, что экземпляр попадает в j -й интервал целевого параметра, при условии, что i -й признак попадет в z -й интервал.

Для интеграции результатов, полученных после применения рассмотренных методов по отдельности, предлагается использовать относительную нормированную комбинированную оценку значимости признаков I_i^J , для расчета которой предлагается использовать следующую методику.

В начале рассчитываются суммарные оценки информативности признаков по нормированным значениям информативностей для каждого признака:

$$I_i^{J*} = \sum_{k \in \{A, \dots, I\}} \frac{I_i^k - \min(I_i^k)}{\max(I_i^k) - \min(I_i^k)}.$$

Затем производится расчет относительных нормированных комбинированных оценок значимости признаков:

$$I_i^J = \frac{I_i^{J*} - \min(I_i^{J*})}{\max(I_i^{J*}) - \min(I_i^{J*})}.$$

2 ПРОГРАММНАЯ РЕАЛИЗАЦИЯ ПОДСИСТЕМЫ ОТБОРА ПРИЗНАКОВ

Рассмотренные методы оценки информативности и отбора признаков представляют существенный интерес для решения практических задач управления качеством. Однако, к сожалению, в большинстве программных комплексов для управления качеством и математического моделирования (например, в пакетах MATLAB, Mathematica, Statistica и др.) данные методы отсутствуют. Поэтому возникла необходимость разработки программной реализации рассмотренных методов.

В качестве базового языка программирования для реализации процедур отбора признаков был выбран язык программного комплекса MATLAB фирмы MathWorks Inc., поскольку, во-первых, данный комплекс является одним из наиболее распространенных и популярных, во-вторых, он содержит большой арсенал средств для разработки математических приложений, в-третьих, программы, написанные на языке MATLAB, могут быть автоматически переведены в программы на языке Си с помощью компилятора, входящего в комплект поставки программного комплекса MATLAB, в-четвертых, программы на языке MATLAB могут быть легко интегрированы с диагностическим программным комплексом [10, 11].

Для каждого метода отбора признаков была разработана отдельная процедура. При этом все процедуры имели одинаковые типы входных и выходных данных и одинаковый формат вызова. Это позволяет рассматривать совокупность данных процедур как новую подсистему (Toolbox) комплекса MATLAB. Однако непосредственный вызов процедур в среде MATLAB представляется крайне неудобным для практического применения, поскольку требует достаточно высокой программистской квалификации пользователя.

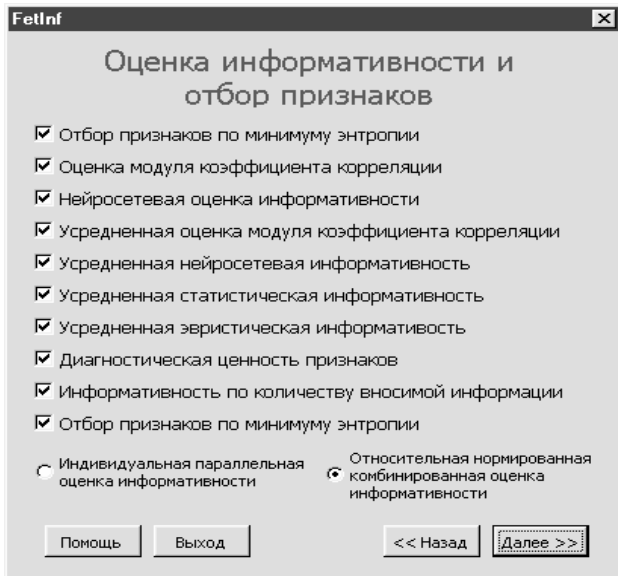


Рисунок 1 – Форма «Меню» интерфейсной программы-оболочки

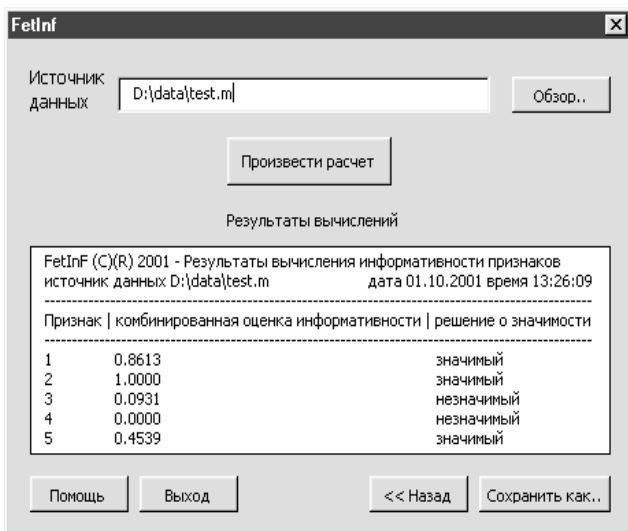


Рисунок 2 – Форма «Данные и результаты» интерфейсной программы-оболочки



Рисунок 3 – Схема взаимодействия пользователя с программными системами

Поэтому гораздо более удобным, чем непосредственная работа с комплексом MATLAB, является использование специально разработанной программной оболочки, которая имеет простой и удобный пользовательский интерфейс и гибко интегрируется как с пакетом MATLAB, так и с диагностическим программным комплексом [10, 11]. Графическое изображение интерфейса программной оболочки представлено на рис. 1 и рис. 2. Схема взаимодействия пользователя, программной оболочки, комплекса MATLAB, диагностического программного комплекса и подсистемы процедур отбора признаков показана на рис. 3.

3 МЕТОД ФОРМИРОВАНИЯ ОБУЧАЮЩЕЙ ВЫБОРКИ

При построении моделей сложных объектов и процессов по точечным данным на основе статистических, вероятностных и топологических методов распознавания образов, а также нейронных сетей возникает задача разбиения исходной выборки данных на обучающую и контрольную.

Традиционно обучающую и контрольную выборки выделяют с помощью случайных чисел или линейным разбиением исходной выборки. Однако при этом в контрольную выборку могут попасть такие экземпляры, которые сильно удалены в метрическом пространстве признаков от экземпляров, попавших в обучающую выборку. В этом случае модель, построенная по обучающей выборке, будет плохо работать даже для контрольной выборки, не говоря уже об адекватности этой модели исследуемому объекту.

Поэтому возникает задача разработки алгоритма, позволяющего разбивать исходную выборку таким образом, чтобы обучающая выборка содержала все экземпляры, находящиеся в узловых точках, представленные в исходной выборке, а контрольная выборка содержала экземпляры близкие (в смысле расстояния) к соответствующим экземплярам обучающей выборки.

Заметим, что в данной работе не ставится задача определения минимально достаточного объема исходной выборки, поскольку в работе предполагается, что исследователь не имеет возможности получить дополнительные экземпляры, а исходная выборка достаточно репрезентативна для получения требуемой модели.

С другой стороны, при обучении нейронных сетей одной из важнейших задач является сокращение времени обучения сети при обеспечении заданного уровня точности прогнозирования, что может быть достигнуто за счет сокращения обучающей выборки путем удаления из нее экземпляров, не находящихся в узловых точках (т. е. избыточных примеров).

Для решения обеих вышеописанных задач предлагается использовать следующий алгоритм.

Шаг 1. Инициализация параметров алгоритма построения обучающей выборки. Задать исходную выборку экземпляров $x_{исх}$ и сопоставленные им номера классов или значения прогнозируемого параметра $t_{исх}$, а также L – количество разбиений исходной выборки.

Занести в переменную N количество признаков, характеризующих экземпляры, а в переменную M – количество экземпляров исходной выборки. Для задач классификации принять ширину допустимого интервала вариации прогнозируемого параметра $dt = 0$, для задач численной оценки прогнозируемого параметра принять $dt = |\max(t_{исх}) - \min(t_{исх})|/L$, где $\min(a)$ и $\max(a)$ – минимальное и максимальное значения вектора a , соответственно. Установить счетчик $newind = 1$.

Шаг 2. Вычислить расстояния между экземплярами исходной выборки.

$$R(p, q) = \begin{cases} \sqrt{\sum_{i=1}^N (x_i^p - x_i^q)^2}, p \neq q; \\ \text{Realmax}, p = q, \end{cases}$$

$$p = 1, \dots, M, \quad q = 1, \dots, M,$$

где Realmax – максимально представимое в ЭВМ число, x_i^q – значение i -го признака p -го экземпляра.

Шаг 3. Найти в матрице расстояний R минимальный элемент $\min x$ и его индексы q и p , а также максимальный элемент $\max x$, при условии, что при нахождении минимума и максимума здесь и далее игнорируются элементы, равные Realmax .

Шаг 4. Принять $a = |\max x - \min x|/(2L)$.

Шаг 5. Если $\min x < \text{Realmax}$, то перейти на шаг 6, в противном случае – перейти на шаг 13.

Шаг 6. Принять: $x_{об}(newind) = x_{исх}(q)$, $t_{об}(newind) = t_{исх}(q)$, где $x_{об}$ и $t_{об}$ – массивы экземпляров обучающей выборки и сопоставленных им прогнозируемых значений, соответственно. Установить: $newind = newind + 1$, значение текущего минимального элемента в строке $teck = R(q, p)$. Найти минимальный элемент $m\min x$ и его индексы $m\min$ и $mp\min$ среди элементов q -й строки матрицы R .

Шаг 7. Если $m\min x < \text{Realmax}$, то перейти на шаг 8, в противном случае – перейти на шаг 11.

Шаг 8. Установить значение указателя удаленного экземпляра из столбца $deleted = 0$ (в матрице R нумерация строк и столбцов должна начинаться с 1).

Шаг 9. Если $|teck - R(q, mp\min)| \leq a$, то перейти на шаг 10, иначе принять: $deleted = mp\min$, $R(q, mp\min) = \text{Realmax}$, $R(mp\min, q) = \text{Realmax}$ и перейти на шаг 11.

Шаг 10. Если $|t_{исх}(mp\min) - t_{исх}(q)| \leq dt$, то принять: $R(v, mp\min) = \text{Realmax}$, $R(mp\min, v) = \text{Realmax}$, $v = 1, \dots, M$, в противном случае – принять: $x_{об}(newind) = x_{исх}(mp\min)$, $t_{об}(newind) = t_{исх}(mp\min)$, $newind = newind + 1$, $R(v, mp\min) = \text{Realmax}$, $R(mp\min, v) = \text{Realmax}$, $v = 1, \dots, M$.

Шаг 11. Найти минимальный элемент $m\min x$ и его индексы $m\min$ и $mp\min$ среди элементов q -й строки матрицы R .

Шаг 12. Перейти на шаг 7.

Шаг 13. Принять: $R(v, q) = \text{Realmax}$, $R(q, v) = \text{Realmax}$, $v = 1, \dots, M$, указатель удаленного экземпляра из строки $dstr = q$. Найти в матрице расстояний R минимальный элемент $\min x$ и его индексы q и p .

Шаг 14. Перейти на шаг 5.

Шаг 15. Если $(deleted \neq dstr)$ и $(deleted > 0)$, тогда принять: $x_{об}(newind) = x_{исх}(deleted)$, $t_{об}(newind) = t_{исх}(deleted)$.

Шаг 16. Останов.

В результате выполнения данного алгоритма для исходной выборки $x_{исх}$ и сопоставленного ей набора значений $t_{исх}$ мы получим обучающую выборку $x_{об}$ и сопоставленный ей набор значений $t_{об}$. Остаток экземпляров из $x_{исх}$ и $t_{исх}$, не вошедших в $x_{об}$ и $t_{об}$ составит контрольную выборку.

Данный метод, также как и методы отбора признаков, был реализован на языке MATLAB в виде отдельной процедуры и служит дополнением диагностического программного комплекса [5, 6].

4 ЭКСПЕРИМЕНТЫ И РЕЗУЛЬТАТЫ

Для оценки применимости рассмотренных методов и апробации программных процедур на их основе осуществлялся отбор признаков для решения различных задач управления качеством. В частности решались: задача диагностики лопаток газотурбинных авиадвигателей по спектрам свободных затухающих колебаний после ударного возбуждения [12–14], задача классификации сельскохозяйственных растений на культурные и сорные [15–18], а также задача моделирования суммарного показателя качества жизни больных бронхитом по данным клинических лабораторно-инструментальных исследований [19–20]. Результаты проведенных экспериментов представлены в таблице 1.

Таблица 1 – Результаты экспериментов по отбору признаков

Задача	Размер исходного набора признаков	Размер сокращенного набора признаков
Диагностика лопаток	10000	1000
Классификация растений	55	21
Моделирование суммарного показателя качества жизни больных бронхитом	47	35

Как видно из таблицы 1, количество признаков после отбора сократилось во всех задачах. Заметим, что для сокращенного набора признаков точность построенных моделей была несколько ниже, чем для исходного набора признаков, что объясняется частичной потерей информации, содержащейся в исключенных признаках. Однако точность моделей, полученных на основе сокращенного набора признаков, удовлетворяла требованиям решаемых задач.

Для исследования алгоритма формирования обучающей выборки на его основе осуществлялось разбиение исходных выборок на обучающие и контрольные для вышеописанных прикладных задач. Результаты экспериментов представлены в таблице 2.

Как видно из табл. 2, применение алгоритма формирования обучающей выборки позволяет существенно

(на 20–60% и более в разных опытах) сократить объем данных, используемых при построении модели. На основе полученных обучающих выборок были построены модели для решения соответствующих задач. Тестирование полученных моделей показало, что они обеспечивают практически такую же точность, что и модели, построенные на всей исходной выборке. Однако модели, полученные на основе сформированной обучающей выборки, были менее сложными, более быстро работающими и менее требовательными к ресурсам ЭВМ.

Таблица 2 – Результаты экспериментов по формированию обучающей выборки

Задача	Размер исходной выборки	Размер построенной обучающей выборки	Размер построенной контрольной выборки
Диагностика лопаток	32	13	19
Классификация растений	248	121	127
Моделирование суммарного показателя качества жизни больных бронхитом	86	49	37

Применение предложенных методик позволяет на практике существенно сокращать описание сложных объектов и процессов, что, в свою очередь, позволяет упростить модели этих объектов и значительно снижает требования к ресурсам ЭВМ.

ПЕРЕЧЕНЬ ССЫЛОК

1. Дубровин В.И. Оценка объема выборки при построении регрессионных моделей. // Системный анализ и принятие решений в задачах автоматизированного обеспечения качества и надежности изделий приборостроения и радиоэлектроники: Тезисы докладов Российской научно-технической конференции. – Махачкала: ПО Азимут, 1991. С. 78.
2. Дубровин В.И. Определение необходимого объема экспериментальных данных для моделирования технологического процесса. // Комплексная механизация и автоматизация производства: Тезисы докладов семинара. – Пенза: ПДЭНТЗ, 1991. С. 16–17.
3. Дубровин В.И. Планирование количества измерений при статистическом описании объектов исследования. // Методы и средства оценки и повышения надежности приборов, устройств, и систем: Тезисы докладов международной научно-технической конференции. – Пенза, ПДЭНТЗ, 1992. С. 79.
4. Дубровин В.И. Минимизация объема экспериментальных данных при управлении качеством. // Моделирование и контроль в задачах обеспечения надежности радиоэлектронных устройств: Материалы международного научно-технического семинара. – Шяуляй, 1992., С. 44.
5. Дубровин В.И. Формирование выборки в задачах моделирования технических систем // Контроль и управление в технических системах: Тезисы докладов научно-технической конференции стран СНГ. – Винница: ВПИ, 1992. С. 61.
6. Дубровин В.И., Субботин С.А. Онлайн-методы управления качеством: гибридная диагностика на основе нейронных сетей // Радиоэлектроника. Информатика. Управление. 2001. № 1. С. 158–163.
7. Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере // Новосибирск: Наука, 1996. – 276 с.
8. Биргер И.А. Техническая диагностика. – М.: Машиностроение, 1978. – 240 с.

9. Мозгалевский А.В., Гаскаров Д.В. Техническая диагностика. – М.: Высшая школа, 1975. – 207 с.
10. Дубровин В.И., Субботин С.А. Программный комплекс нейросетевой диагностики // Программные продукты и системы, 2000. № 3, С. 21–23.
11. Дубровин В.И., Субботин С.А. Нейросетевая подсистема диагностического программного комплекса // Нейрокомпьютеры: разработка и применение, 2001, № 2. С. 55–62.
12. Дубровин В.И., Субботин С.А. Нейросетевая диагностика энергетических установок // Датчики и преобразователи информации систем измерения, контроля и управления / Сборник материалов XII научно-технической конференции с участием зарубежных специалистов. Под ред. проф. В. Н. Азарова. М.: МГИЭМ, 2000. – С. 240–242.
13. Адаменко В.А., Дубровин В.И., Жеманюк П.Д., Субботин С.А. Диагностика лопаток авиадвигателей по спектрам свободных затухающих колебаний после ударного возбуждения // Автоматика – 2000. Міжнародна конференція з автоматичного управління, Львів, 11–15 вересня 2000: Праці у 7 томах. – Т. 5. – Львів: Державний НДІ інформаційної інфраструктури, 2000.– С. 7–13.
14. Адаменко В.А., Басов Ю.Ф., Дубровин В.И., Субботин С.А. Нейросетевая обработка сигналов в задачах диагностики газотурбинных авиадвигателей // Цифровая обработка сигналов и ее применение: 3-я Международная конференция и выставка. – М.: РНТОРЭС им. А.С. Попова, 2000. – С. 40–45.
15. Dubrovin V., Morshchavka S., Piza D., Subbotin S. Plant recognition by genetic algorithm based back-propagation // Proceedings, Remote Sensing 2000: from spectroscopy to remotely sensed spectra. Soil Science Society of America, Bouyocos Conference, Corpus Christi, Texas, October 22–25, 2000. – P. 47–54.
16. Дубровин В.И., Морщавка С.В., Пиза Д.М., Субботин С.А. Нейросетевая идентификация объектов по спектрам // Труды международной конференции «Идентификация систем и задачи управления» SICPRO' 2000. – М.: ИПУ РАН, 2000. – С. 1190–1204 (CD-ROM).
17. Dubrovin V.I., Subbotin S.A., Morshchavka S.V., Piza D.M. The plant recognition on remote sensing results by the feed-forward neural networks // Smart Engineering Systems Design: Neural Networks, Fuzzy Logic, Evolutionary Programming, Data Mining, and Complex Systems, ANNIE 2000: the 10-th Anniversary edition / ed. C. H. Daglietal. – Missouri–Rolla: ASME Press, 2000, vol. 10, P. 697–701.
18. Дубровин В.И., Морщавка С.В., Пиза Д.М., Субботин С.А. Распознавание растений по результатам дистанционного зондирования на основе многослойных нейронных сетей // Математичні машини і системи, 2000. № 2–3. С. 113–119.
19. Качество жизни больных хроническими обструктивными болезнями легких / Е.И. Шмелев, М.В. Беда, Р.В. Jones и др. // Пульмонология. – 1998. № 2. С. 79–81.
20. Перцева Т.О., Ботвінікова Л.А. Медико-соціальні аспекти визначення якості життя у хворих на бронхіальну астму та хронічний обструктивний бронхіт // Український пульмонологічний журнал.–2000, № 1. – С. 19–21.

Надійшла 18.11.2004

Розглянуто розв'язок задачі скорочення обсягу даних при побудові математичних моделей складних об'єктів та процесів шляхом зменшення кількості ознак за рахунок виключення малозначимих ознак та зменшення навчальної множини за рахунок виключення зайвих прикладів. Описана програмна реалізація та наведені результати застосування запропонованих методів при вирішенні практичних задач.

The solution of the data size reduction problem at a construction of mathematical models of complex objects and processes by decreasing of features number at the expense of exception of non-significance features and decreasing of learning set at the expense of exception of redundant examples is considered. The program implementation is described and the results of application of offered methods at solution of the practical tasks are indicated.

И.Ш. Невлюдов, Е.П. Второв, Б.А. Шостак

ОПТИМИЗАЦИЯ ПРОВЕРКИ ЦИФРОВЫХ МОДУЛЕЙ ТЕХНОЛОГИЧЕСКОГО ОБОРУДОВАНИЯ В УСЛОВИЯХ ЧАСТИЧНОЙ НЕОПРЕДЕЛЕННОСТИ

Приведены описание и основные характеристики метода принудительной диагностики цифровых модулей систем управления технологическим оборудованием. Рассмотрены вопросы программной и аппаратной реализации данного метода.

ВВЕДЕНИЕ

При диагностике сложных цифровых модулей систем с числовым программным управлением (ЧПУ) и вычислительных микропроцессорных устройств в условиях частичной неопределенности (отсутствие принципиальной схемы объекта диагностики, отсутствие информационной или функциональной модели некоторых элементов объекта диагностики), а также при отсутствии возможности аппаратного подключения внешней тестовой эмуляции возникают значительные трудности. Эти трудности также обусловлены тем, что при моделировании сложных объектов действует фундаментальное положение кибернетики: при сложности объекта выше некоторого уровня его адекватная (полная модель) не может быть более простой. Поэтому фактически ни одна система ЧПУ, применяемая в технологическом оборудовании, не имеет исчерпывающего математического описания.

Для выхода из создавшейся ситуации при диагностике цифровых систем необходимо использовать три уровня описания:

- алгоритмический;
- функциональный;
- вентильный.

В этом случае полное математическое описание строится по линейному принципу:

$$L = L_{\text{вы}} + L_{\text{фу}} + L_{\text{ал}}, \quad (1)$$

где $L_{\text{вы}}$ – язык вентильного уровня; $L_{\text{фу}}$ – язык функционального уровня; $L_{\text{ал}}$ – язык алгоритмического уровня.

Многоуровневая модель выглядит следующим образом:

$$L = L_{\text{ал}} (L_{\text{фу}} (L_{\text{вы}})).$$

Такая последовательность создания диагностических тестов, как показала практика, является наиболее оптимальной с точки зрения снижения трудоемкости и уменьшения временных характеристик процесса диагностики систем ЧПУ. Это связано с тем, что на каждом этапе создания тестовых наборов, тестовые алгоритмы

осуществляют поиск неисправности с заданным уровнем детализации.

Таким образом, происходит ступенчатое углубление детализации. На рис. 1 изображен процесс поиска кратных неисправностей одного структурного модуля.

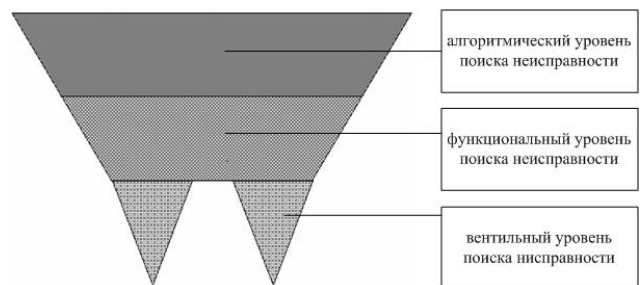


Рисунок 1 – Процесс поиска кратных неисправностей в цифровом модуле

Таким образом, актуальной является задача минимизации времени диагностирования микропроцессорных модулей. Для минимизации можно применить метод экспертных оценок параметров сигналов на выводах цифровых микросхем модуля. При таком подходе возникает проблема создания оптимальной системы принятия решений (о неисправности того или иного дискретного элемента). Под такой системой понимается система, которая минимизирует заданный критерий качества для данного динамического процесса при заданных ограничениях. В случае описанного ниже метода проверки цифровых модулей технологических систем в качестве критерия качества выступает риск принятия неверного решения при заданном входном пространстве состояний, а ограничения заключаются в фиксированной структуре принципиальной схемы и применением логических элементов с открытым коллекторным выходом.

В статье предлагается решить задачу минимизации времени диагностики цифровых модулей посредством применения метода принудительной диагностики. Суть метода принудительной диагностики заключается в подаче внешних тестовых воздействий на определенный исследуемый элемент объекта диагностики посредством специального зонда и анализе полученного с выходов элемента откликов.

Разработанная на основе описанного метода система принудительной диагностики СПД позволяет произво-

дить диагностику микросхем серий ТТЛ и ТТЛШ, не выпаивая их из платы. В настоящий момент база функциональных моделей составляет 246 типов микросхем, как отечественного, так и импортного производства [4].

Общая методика функционирования СПД заключается в следующей последовательности действий:

1. В исследуемом объекте диагностики (ОД) аппаратно блокируются все внутренние генераторы [2];
2. На исследуемый элемент (микросхему) устанавливается зонд;
3. На ОД подается напряжение питания;
4. Активируется программа динамической инициализации, которая производит формирование управляющих воздействий определенной длительности (для микросхем серии 1533 соответственно 250 нс, 500 нс и 750 нс), которые активизирует внутренний динамический функционал исследуемого элемента [3];
5. Проверяется динамический функционал путем сравнения с прототипом из базы знаний [4];
6. Если динамические функционалы совпадают, то элемент считается исправным, и переходят к исследованию следующего элемента, если нет – то производится аналогичный контроль сопряженных с выходными информационными цепями элементов.

Таким образом, проверяются функциональная исправность всех элементов, прототипы (функциональные модели) которых содержатся в базе знаний. Время проверки одного элемента находится в квадратичной зависимости от количества входных выводов исследуемой микросхемы, и в среднем составляет 3 минуты. Подача внешних импульсов на выходные выходы микросхем не приводит к выходу из строя выходных каскадов в связи с малым временем воздействия и ограничениях максимального импульсного тока ($I_{max} = 500 \text{ mA}$) [5]. Ограничение по току необходимо также и для проверки микросхем, входные цепи которых соединены либо с выходными каскадами с открытым коллектором, либо с цепями питания.

Активизировать внутреннюю логическую функцию любой микросхемы из заданной базы можно путем последовательной подачи активных уровней на определенные входы микросхемы. Располагая статическими значениями логических уровней на выводах микросхемы в определенный момент времени, и учитывая изменения сигналов на входах, можно точно определить соответствие выходных сигналов функциональной модели исследуемой микросхемы. Если классифицировать типы входных выводов по функциональной принадлежности, то их можно разделить на три класса: разрешающие, информационные и стробирующие. Соответственно и активизация этих входов происходит в следующей последовательности (рис. 2):

1. Активизация разрешающих входов (момент времени T1);
2. Запись в буферный регистр № 1 состояния логических уровней на выводах микросхемы (момент времени T2);
3. Активизация информационных входов (момент времени T3);

4. Запись в буферный регистр № 2 состояния логических уровней на выводах микросхемы (момент времени T4);

5. Активизация стробирующих входов (момент времени T5);

6. Запись в буферный регистр № 3 состояния логических уровней на выводах микросхемы (момент времени T6);

7. Сброс активных уровней на входных выводах микросхемы (момент времени T7).



Рисунок 2 – Последовательность подачи активных уровней на входы цифровых микросхем

Аппаратная часть СПД подключается к ЭВМ посредством любого 24-разрядного унифицированного интерфейсного модуля сопряжения (рис. 3) [6]. Структурная схема аппаратной части СПД представлена на рис. 4.

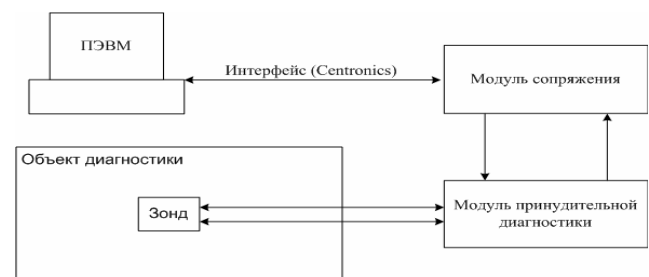


Рисунок 3 – Схема подключения аппаратных узлов

Программное обеспечение СПД состоит из трех частей [7, 8]:

1. Программа, формирующая управляющий и стробирующий код для каждого тестового набора (результатом выполнения программы является файл, содержащий десятичные коды управляющих и стробирующих воздействий);
2. Программа, непосредственно активизирующая работу СПД (результатом выполнения программы является создание временного файла, содержащего значения логических уровней на выводах исследуемой микросхемы при подаче каждого тестового воздействия);
3. Программа, анализирующая полученный на предыдущем этапе файл (результатом выполнения программы является формирование заключения о функциональном

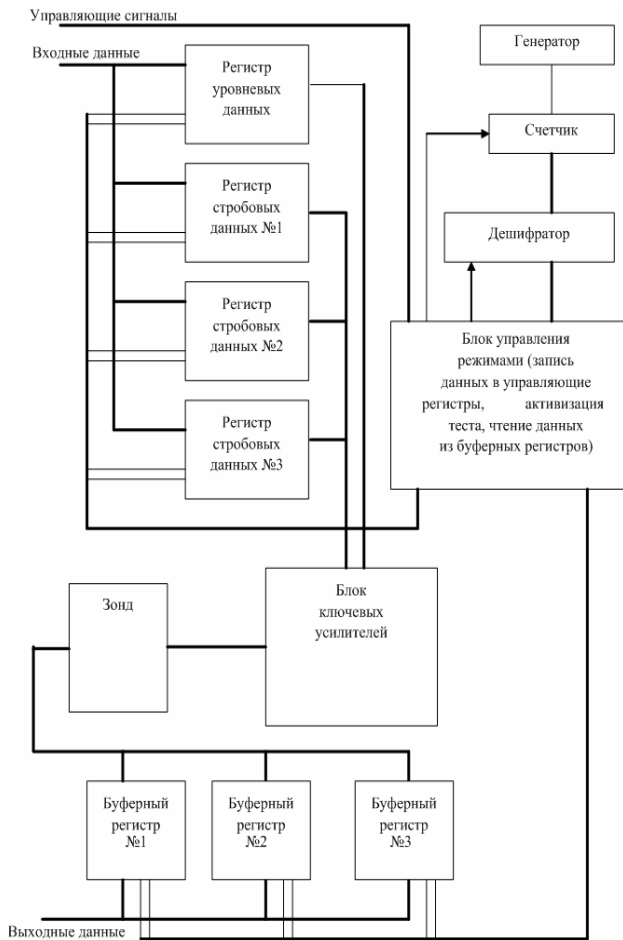


Рисунок 4 – Структурная схема аппаратной части СПД

соответствии исследуемой микросхемы). Функциональные модели микросхем заданы в виде программных модулей производящего типа.

Входными данными при работе ПО, формирующего управляющий и стробирующий код, являются:

- тип микросхемы;
- количество выводов;
- тип выводов (одно- или двунаправленные);
- степень приоритетности (1 – при равнозначных входах, 2 – при наличии разрешающих, сбросовых или установочных входов, 3 – при наличии разрешающих, стробирующих и информационных входов);
- количество независимых элементов в корпусе микросхемы;
- номера выводов подачи питания;
- количество информационных входов;
- количество разрешающих входов;
- количество стробирующих входов.

Алгоритм формирования управляющих и синхронизирующих воздействий зависит от степени приоритетности выводов:

- 1-я степень приоритетности:
 1. Ввод данных;
 2. Формирование синхронизирующего кода;
 3. Ввод данных;
 4. Формирование синхронизирующего кода;

- 5. Формирование управляющего кода;
- 6. Запись кода в базовый файл.
 - 2-я степень приоритетности:
 1. Ввод данных;
 2. Формирование синхронизирующего кода, активирующего разрешающие входы;
 3. Формирование синхронизирующего кода, активирующего информационные входы;
 4. Формирование управляющего кода;
 5. Запись кода в базовый файл.
 - 3-я степень приоритетности:
 1. Ввод данных;
 2. Формирование синхронизирующего кода, активирующего разрешающие входы;
 3. Формирование синхронизирующего кода, активирующего информационные входы;
 4. Формирование синхронизирующего кода, активирующего стробирующие входы;
 5. Формирование управляющего кода;
 6. Запись кода в базовый файл.

Программа анализа данных производит пошаговое извлечение строк состояния сигналов на выводах из файла, сформированного на предыдущем этапе. После извлечения производится обращение к производящей базовой модели исследуемого элемента. Если извлеченные данные совпадают с данными базовой модели, то состояние сигналов на выводах считается корректным и производится извлечение следующей строки состояния. Если извлеченные данные не совпадают с данными базовой модели, то выдается сообщение о неисправности исследуемого элемента.

Функционирование СПД ограничивается следующими факторами:

1. При соединении нескольких выходов логических элементов с открытым коллектором (монтажное ИЛИ) возможно неверное заключение об отказе проверяемого элемента, если его выход заблокирован в нулевом состоянии другим элементом;
2. При замыкании входов логических элементов на шины питания возможна неполная проверка функционирования элемента;
3. При неисправном выходном каскаде предыдущего элемента возможна неполная проверка корректного функционирования элемента;
4. При нанесенном лаковом покрытии возможен не полный контакт зонда с выводами диагностируемого элемента, что приводит к неверной проверке.

ВЫВОДЫ

Описанный выше метод диагностики дискретных элементов цифровых модулей систем ЧПУ обладает несколькими достоинствами. Во-первых, при диагностике отдельных элементов нет необходимости выпаивать их из печатной платы. Во-вторых, нет необходимости формировать тестовые последовательности для проверки этих элементов. В-третьих, появляется возможность производить диагностику элементов цифровых модулей при

отсутствии принципиальных схем, что особенно актуально при диагностике зарубежных систем ЧПУ. В-четвертых, значительно снижается трудоемкость процесса диагностики и требования к квалификации инженера по наладке и испытаниям. В-пятых, метод хорошо автоматизируется при работе на ЭВМ.

ПЕРЕЧЕНЬ ССЫЛОК

1. Нефедов А.В., Савченко А.М., Феоктистов Ю.Ф. Зарубежные интегральные микросхемы для промышленной электронной аппаратуры: Справочник. – М.: Энергоатомиздат, 1989. – 288 с.
2. Фергусон Дж., Макари Л., Уильямс П. Обслуживание микропроцессорных систем. – М.: Мир, 1989. – 336 с.
3. Горбунов В.А., Панфилов Д.И., Преснухин Д.А. Справочное пособие по микропроцессорам и микро-ЭВМ. – М.: Высшая школа, 1988. – 272 с.
4. Самофалов К.Г., Романкевич А.М., Валуйский В.Н. и др. Прикладная теория цифровых автоматов. – Киев.: Вища школа, 1987. – 375 с.

УДК 681.58; 681.32

В. В. Орлов

БЛОК КЛАВИАТУРЫ ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНОЙ СИСТЕМЫ НА БАЗЕ МИКРОКОНТРОЛЛЕРА

Предлагается эффективный алгоритм вычисления весовой функции нажатой клавиши, не требующий арифметических операций, и схема 16-клавишной клавиатуры с минимальным числом дополнительных элементов, использующая 5 линий пространства ввода-вывода микроконтроллера. Клавиатура конструктивно расположена на расстоянии около 300 мм от базового блока микроконтроллерной системы, которая может быть выполнена на микроконтроллерах различного типа.

ВВЕДЕНИЕ

Одной из наиболее динамично развивающихся областей современной вычислительной техники является микроконтроллерная техника. Появление на рынке микроконтроллеров с высокой производительностью, значительным объемом встроенной памяти, мощным вычислительным ядром с системой команд, ориентированной на решение задач управления и контроля, и доступными ценами обусловило их широкое применение в различных областях автоматизации и управления, в том числе и автоматизированных системах научных исследований (АСНИ). Ярким примером современных микроконтроллеров является семейство AVR-микроконтроллеров производства фирмы «ATMEL», завоевавших популярность среди разработчиков информационно-измерительных и микропроцессорных систем управления (МПСУ).

5. Аванесян Г.Р., Левшин В.П. Интегральные микросхемы ТЛ, ТЛШ: Справочник. – М.: Радио и связь, 1991. – 192 с.
6. Новиков Ю.В., Калашников О.А., Гуляев С.Э. Разработка устройств сопряжения для персонального компьютера типа IBM PC. – М.: ЭКОМ, 1997. – 224 с.
7. Кофрон Дж., Лонг В. Расширение микропроцессорных систем / Пер. с англ.: Под ред. П. В. Нестерова. – М.: Машиностроение, 1987. – 320 с.
8. Зельднер Г.А. Программируем на языке Quick Basic 4.5. – М.: АБФ, 1997. – 432 с.

Надійшла 14.06.2004

Приведено опис і основні характеристики методу примусової діагностики цифрових модулів систем керування технологічним устаткуванням. Розглянуто питання програмної і апаратної реалізації даного методу.

The article is developed to the questions of labor input reduction of the process of the technological equipment digital modules' diagnostics and to the increasing of the efficiency of the faults' search in condition of the partial uncertainty. The method of the diagnostics was designed for the solving of the problems, intended for optimization of searching for failure digital element in module of any complication and functional types.

ПОСТАНОВКА ЗАДАЧИ

В интерактивных информационно-измерительных системах и МПСУ в контур управления включен интерфейс «оператор-система», основным элементом которого является клавиатура, предназначенная для ввода информации в систему. Обычно, если число клавиш не превышает 64, применяют так называемую неcodирующую матричную схему, которая представляет собой простую матрицу двоичных переключателей (требуемой размерности), включенных на пересечении строк и колонок матрицы. Идентификация (кодирование) нажатой клавиши в таких клавиатурах выполняется программой. В этом случае наиболее распространенный способ вычисления весовой функции нажатой клавиши – аналитический в соответствии с выражением

$$W = NC + B, \quad (1)$$

где W – весовая функция, N – количество линий возврата, C – номер линии сканирования, B – номер линии возврата.

Примером таких схем могут служить схемы приведенные в [1, 2] и других публикациях по применению микроконтроллеров. Очевидным недостатком такого способа вычисления весовой функции является применение арифметических операций, и в особенности операции умножения, поскольку микроконтроллеры, как правило,

не имеют в своей системе команд операции умножения. Использование подпрограммы умножения приводит к значительному увеличению размера исполняемого кода и снижению быстродействия микроконтроллерной системы в целом.

Так же следует учесть еще одно важное обстоятельство – на практике, по конструктивным соображениям, часто появляется необходимость размещения клавиатуры на некотором расстоянии от базовой платы микроконтроллера.

МЕТОД РЕШЕНИЯ

При решении подобной задачи в ходе разработки блока удаленной клавиатуры информационно-измерительной системы, выполненной на базе AVR микроконтроллера типа AT90S8515, предназначенной для автоматизации тракта регистрации данных магнитного масс-спектрометра МИ-1201Э (НПО «Электрон», г. Сумы, Украина), были изучены особенности функционирования блока и предложено программно-аппаратное решение, позволившее исключить арифметические операции при вычислении весовой функции клавиши.

Первоначально схема клавиатуры представляла собой матрицу 4×4, составленную из кнопок типа ПК8, и по конструктивным требованиям была вынесена от базовой платы микроконтроллера на расстояние 250–300 мм посредством ленточного кабеля. Фрагмент схемы показан на рис. 1.

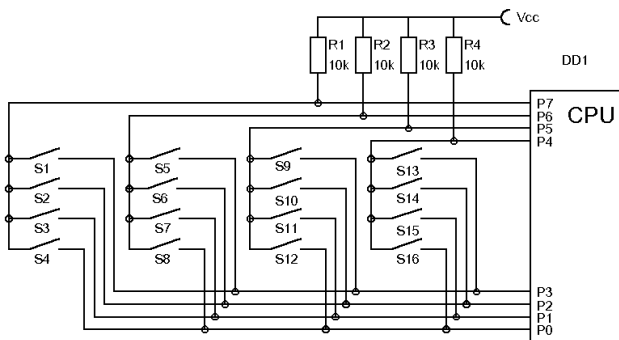


Рисунок 1 – Схема матричной клавиатуры:

S1–S16 – контакты клавиатуры; R1–R4 – подтягивающие резисторы;
DD1 – микроконтроллер

Однако, при практическом использовании данной схемы был обнаружен эффект, приводивший к полной потере работоспособности устройства. Заключалось это в следующем: при обращении управляющей программы к функции чтения клавиатуры, функция, не дожидаясь реального нажатия клавиши, возвращала случайный скэн код клавиши, как правило, первых двух строк матрицы (эффект "самосканирования" матрицы). Подбор номинала подтягивающих резисторов R1–R4 в пределах 1,5...100 кОм приводил к изменению частоты «самосканирования».

Рассматривая электрические характеристики микроконтроллера [3], такие как высокое входное сопротивление линий порта ввода и низкое выходное сопротивление линий порта вывода, можно заметить, что эти параметры обеспечивают хорошие условия для перезарядки распределенной емкости внешней цепи, подключенной к портам ввода-вывода микроконтроллера. На основании этих данных был сделан вывод о емкостной реакции контактов клавиатуры, представляющих собой механические контакты нажимного действия достаточно большой площади.

Дальнейшие эксперименты показали, что работоспособность такой схемы может быть обеспечена только при расположении микропереключателей с уменьшенной площадью контактов в непосредственной близости от выводов микроконтроллера и тщательной трассировке платы. Такой подход к решению проблемы противоречит двум условиям: выбору типа контактов по эргономическим соображениям и удаленному размещению клавиатуры по конструктивным требованиям компоновки устройства.

Для решения вышеизложенной проблемы нами предложено изменение схемотехнического решения блока клавиатуры, основанное на использовании мультиплексора. Фрагмент предлагаемой схемы показан на рис. 2.

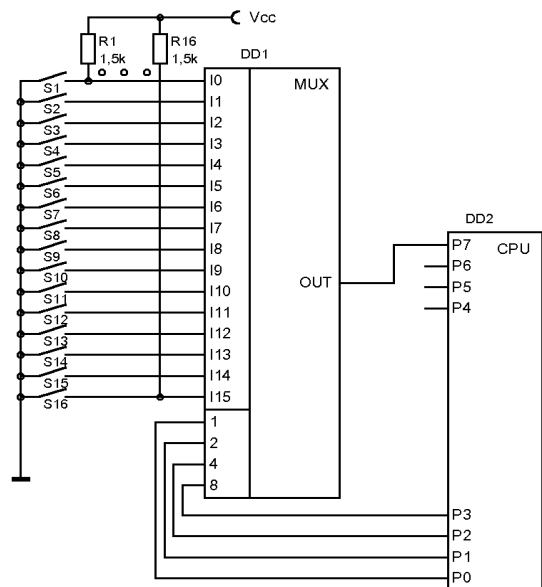


Рисунок 2 – Схема клавиатуры с применением мультиплексора:

S1–S16 – контакты клавиатуры; R1–R16 – подтягивающие резисторы;
DD1 – мультиплексор; DD2 – микроконтроллер

В такой схеме мультиплексор DD1 выполняет функции согласования линии связи контактов клавиатуры с портом ввода-вывода микроконтроллера DD2 и программно-аппаратного шифратора скэн кода нажатой клавиши.

При изготовлении блока возможны два варианта компоновки схемы – размещение мультиплексора на общей плате микроконтроллера и на конструкционной плате клавиатуры. Для уменьшения количества проводников

связи клавиатуры с платой микроконтроллера нами выбран вариант размещения мультиплексора на плате клавиатуры. В этом случае требуется всего 5 проводников.

Рассмотрим принцип действия программно-аппаратного шифратора. Блок-схема предлагаемого алгоритма определения весовой функции представлена на рис. 3.

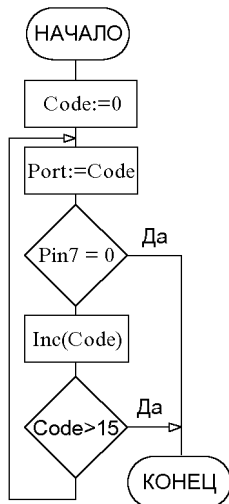


Рисунок 3 – Алгоритм определения весовой функции

Шифратор скэн кода нажатой клавиши работает следующим образом: выход мультиплексора подключен к входу порта ввода микроконтроллера и выполняет функцию бита возврата. При обращении к функции чтения клавиатуры в локальной переменной Code содержится значение от 0 до 15, изменяемое в цикле, которое поступает на адресные входы мультиплексора (переменная Port). Затем анализируется бит возврата (переменная Pin7), имеющий значения 0 – в случае нажатия соответствующей клавиши или 1 – в случае отсутствия события нажатия клавиши. Здесь значения бита возврата указаны для случая прямого выхода мультиплексора. Если мультиплексор имеет инверсный выход, то логику обработки бита возврата необходимо изменить на противоположную. При получении значения 0 бита возврата функция возвращает значение переменной Code. Таким образом осуществляется прямое кодирование номера нажатой клавиши в диапазоне 0...15, ввиду однозначного соответствия номера информационного входа и его адреса, содержащегося в переменной Code.

В целях упрощения блок-схемы в алгоритме не показана обработка ситуации отсутствия события нажатия клавиши. Программное решение в этом случае зависит только от желания программиста. Можно, например, установить старший бит переменной Code или использовать бит T регистра SREG (регистр состояния AVR микроконтроллера).

Подавление дребезга контактов в этой схеме можно производить двумя классическими методами. Либо использовать RS-триггеры в цепи контактов и подключить выход триггера к информационному входу мультиплексора, либо программно, что позволяет существенно уменьшить количество микросхем в устройстве, не накладывает ограничений на тип применяемых контактов (закрывающие или переключающие), достаточно просто реализуется и, наконец, является более современным и изящным методом подавления дребезга контактов.

Особо следует отметить, что опрос клавиатуры можно производить как программно, так и с применением механизма внешнего прерывания, благодаря использованию в схеме одиночной линии возврата.

ВЫВОДЫ

К достоинствам предложенного схемотехнического решения можно отнести простоту, минимальное количество дополнительных элементов и использование меньшего (5 вместо 8) числа разрядов порта ввода-вывода микроконтроллера, что актуально практически в любом случае использования микроконтроллера.

Предлагаемый способ определения весовой функции нажатой клавиши отличаются:

1) отсутствие каких-либо дополнительных арифметических операций вычисления скэн кода (1), что уменьшает размер кода программы и увеличивает быстродействие системы, чем достигается эффективность алгоритма вычисления весовой функции;

2) то обстоятельство, что при наличии клавиш предназначенных для ввода информации в виде цифр в диапазоне 0...9 при соответствующем их подключении к одноименным информационным входам мультиплексора, такой способ декодирования возвращают скэн код, прямо соответствующий необходимой цифре в двоичном формате числа и его значение можно использовать в управляющей программе без дополнительного преобразования.

ЗАКЛЮЧЕНИЕ

Блок клавиатуры, реализованный на основе описанного программно-аппаратного решения, при удалении блока от платы микроконтроллера на расстояние 300 мм посредством неэкранированного ленточного кабеля без проводников защитного заземления (чередование проводов «общий-сигнал-общий») и без использования подтягивающих резисторов в цепи порта ввода-вывода микроконтроллера, показал высокую надежность считывания данных в условиях сложной электромагнитной обстановки работающего масс-спектрометра.

Следует также отметить, что предлагаемый блок клавиатуры можно использовать в МПСУ, выполненной с применением любого типа микроконтроллера.

Данная работа выполнена при частичной поддержке гранта #2155 фонда НТЦУ.

ПЕРЕЧЕНЬ ССЫЛОК

1. Сташин В.В., Урусов А.В., Мологонцева О.Ф. Проектирование цифровых устройств на однокристалльных микроконтроллерах. М.: Энергоатомиздат, 1990. 224 с.
2. Application Note AVR240: 4X4 Keypad-Wake-up on Keypress <http://www.atmel.com>.
3. Евстифеев А.В. Микроконтроллеры AVR семейства Classic фирмы "ATMEL". М.: Издательский дом "Додэка-XXI", 2002. 288 с.

Надійшла 29.12.2003

Запропоновано ефективний алгоритм обчислення вагової функції натиснутої клавіші, що не потребує арифметичних операцій, і схема 16-клавішної клавіатури з мінімальним числом додаткових елементів, яка використовує 5

ліній простору вводу-виводу мікроконтролера. Клавіатуру конструктивно розміщено на відстані близько 300 мм від базового блока мікроконтролерної системи, що може бути виконаною на мікроконтроллерах різних типів.

An efficient algorithm of calculation of the weight function of a pressed key, which does not require arithmetic operations, and a scheme of a 16-key keyboard with minimum of additional elements, which uses 5 lines of the input-output space of a microcontroller, are suggested. The keyboard is located at a distance of about 300mm from the base block of the microcontroller system and can be combined with different types of microcontrollers.

УДК 004.93

М.Ю. Сенашова

ОЦЕНКИ ПОГРЕШНОСТЕЙ ВЫЧИСЛЕНИЯ СЛОЖНОЙ ФУНКЦИИ МНОГИХ ПЕРЕМЕННЫХ И ЕЕ ГРАДИЕНТА

Описаны графы вычисления сложных функций многих переменных, изложен принцип двойственности для вычисления градиента сложной функции. Изложены алгоритмы оценки погрешностей вычислений функции и ее градиента.

ВВЕДЕНИЕ

Если сложная функция многих переменных представима в виде суперпозиции функций меньшего числа переменных, можно построить граф, соответствующий этой функции. Тогда вычисление этой функции может быть описано как движение по графу. Рассматривая затем граф как набор элементов, производящих некоторые вычисления над приходящими к ним данными, можно изучать, как ошибки вычислений элементов влияют на ошибки вычислений всей функции.

Для сложной функции, представленной в виде графа можно решать две задачи, относящиеся к оценке точности ее вычислений: прямую и обратную. Под прямой задачей понимается оценка погрешностей вычислений выходных вершин графа, если заданы погрешности входных вершин графа. Заданные погрешности входных вершин пропускаются по графу в прямом направлении. При этом для каждого нового набора погрешностей входных вершин требуется заново вычислять погрешности выходных вершин. Наиболее близки к этой задаче методы оценки погрешностей счетно-решающих устройств на основе аналоговых элементов [1, 2, 3].

Гораздо больший интерес представляет обратная задача. Под обратной задачей понимается оценка допустимых погрешностей вычислений внутренних и входных вершин графа по заданным погрешностям вычислений его выходных вершин. Для решения обратной задачи задается точность, с которой должно вычисляться значение выходной вершины графа. То есть задается интервал, в котором может изменяться значение выходной вершины графа. За-

тем, проходя по графу в обратном направлении, вычисляются интервалы, в которых могут изменяться значения элементов графа. Эти интервалы вычисляются для всех элементов графа вплоть до входных вершин. Для вычислений интервалов изменения погрешностей строится граф, двойственный исходному. Система связей двойственного графа совпадает с системой связей исходного графа, но все элементы заменяются на двойственные, а сигналы передаются в направлении, противоположном движению сигналов в исходном графе. При этом все значения из интервала изменения значений входных вершин графа, при прямом прохождении по графу попадут в интервал изменения значений выходных вершин графа. При решении обратной задачи мы получаем уже не соотношение «конкретный набор погрешностей для входных вершин – конкретная погрешность для выходной вершины», как это происходит при решении прямой задачи, а соотношение «интервалы изменения значений входных вершин – интервал изменения значения выходной вершины».

ГРАФ ВЫЧИСЛЕНИЯ СЛОЖНОЙ ФУНКЦИИ

Рассматривается сложная функция многих переменных F . Под «сложной» функцией подразумевается, что может быть представлена как суперпозиция некоторого набора «простых» функций f_i от меньшего числа переменных.

Для построения графа, вычисляющего сложную функцию n переменных F , нужно представить эту функцию в виде суперпозиции функций меньшего числа переменных f_i . Функции f_i можно разделить на слои по глубине вхождения в функцию F . В нулевом слое будут находиться константы и переменные, в первом слое – функции, зависящие от констант и переменных нулевого слоя, во втором слое – функции, зависящие от констант, переменных и функций первого слоя и т. д.

По этим слоям удобно строить граф функции F . Вершинами графа будут функции f_i . Если значение вершины f_k является аргументом вершины f_l , то они связываются ребром, направленным от вершины f_k к вершине f_l . Таким образом, получаем ориентированный связный граф, соответствующий функции F . Более подробно построение графа для вычисления сложной функции описано в [4, 5].

С каждой вершиной графа, принадлежащей ненулевому слою, ассоциируется автомат, вычисляющий функцию f_i . Автоматы срабатывают по слоям в дискретные моменты времени (такты). В начальный момент сформированы значения вершин нулевого слоя – известны значения переменных и констант. Они поступают на входы соответствующих автоматов первого слоя. После k -го такта функционирования определены значения вершин, принадлежащих слоям с 0-го по k -й. На $(k+1)$ -м такте автоматы $(k+1)$ -го слоя вычисляют значения вершин $(k+1)$ -го слоя, получая входные сигналы с предыдущих слоев. В результате вычислений получаем значение функции F при конкретных значениях вершин нулевого слоя.

ОЦЕНКА ПОГРЕШНОСТЕЙ ВЫЧИСЛЕНИЯ СЛОЖНОЙ ФУНКЦИИ

Представление сложной функции многих переменных в виде графа удобно еще и для оценки погрешностей вычисления функции, в случае, если аргументы функции имеют погрешности. Для оценки погрешностей вычисления сложной функции нужно построить граф, двойственный исходному графу, вычисляющему значение сложной функции. Этот двойственный граф имеет ту же систему связей, что и исходный, но все элементы исходного графа заменяются на двойственные. Причем сигналы по двойственному графу передаются в направлении, противоположном движению сигналов по исходному графу. Для построения двойственного графа необходимо построить вершины, двойственные вершинам исходного графа. Идея двойственности используется в точности та же, что и в методе обратного распространения ошибки [6]. Отличие состоит в построении двойственных элементов.

Предположим, что заданы значения вершин входного слоя графа v_{in} и для них вычислены значения вершин внутренних слоев и вершин выходного слоя v_{out} . Будем называть эти значения v_{in} , значения вершин внутренних слоев графа, а также значения вершин v_{out} точными. Предположим теперь, что нам требуется вычислить значения сложной функции, то есть значения вершин v_{out} , с некоторой точностью δ . То есть нужно найти такие допустимые погрешности ϵ_i вершин входного слоя v_{in} , что погрешность вершин выходного слоя v_{out} не превысит заданную погрешность δ .

Типичным участком графа является вершина с входящими и выходящими ребрами (рис. 1). Поэтому, чтобы иметь представление о вычислении погрешностей на графе, достаточно выяснить, как вычисляются погрешности для такой вершины.

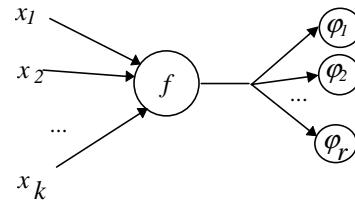


Рисунок 1 – Вершина графа, вычисляющего значение сложной функции

В [7] были получены оценки погрешностей для некоторой вершины f_i графа. Предположим, что вершина f_i имеет k входящих ребер и r выходящих ребер.

Поскольку мы строим вершину двойственного графа для оценки погрешностей, структура связей у нас сохраняется, элементы заменяются на двойственные, а сигналы движутся в направлении, противоположном движению сигналов в исходном графе. Поэтому двойственная вершина будет иметь r входящих и k выходящих ребер. Предположим, что вычисленные при обратном распространении допустимые погрешности ребер, входящих в двойственную вершину f'_i , равны $\epsilon_1, \epsilon_2, \dots, \epsilon_r$. Было показано, что в качестве погрешности вершины f'_i выбирается

$$\epsilon \leq \min\{\epsilon_i\}_{i=1}^r. \quad (1)$$

Для полученной погрешности вершины ϵ было показано, что оценка погрешностей выходящих ребер вершины f'_i вычисляется по формуле:

$$\epsilon_i \leq \epsilon / \sum_{i=1}^k \left| \frac{\partial f_i(x_1, \dots, x_k)}{\partial x_i} \right|. \quad (2)$$

Пользуясь формулами (1) и (2), можно оценить погрешности для любой вершины графа при известных погрешностях выходящих ребер. На рис. 2 показана вершина двойственного графа с оценками погрешностей.

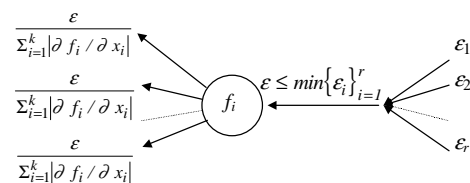


Рисунок 2 – Оценки погрешностей для вершины f_i

При помощи полученных формул можно оценить погрешности только для одного конкретного набора значений вершин v_{in} входного слоя графа. Если задано множество дискретных наборов значений вершин v_{in} , то для оценки погрешностей по всему множеству наборов в целом нужно вычислить погрешности для каждого набора значений v_{in} , а затем для каждой вершины графа выбрать минимум полученных погрешностей.

ГРАФ, ВЫЧИСЛЯЮЩИЙ ГРАДИЕНТ СЛОЖНОЙ ФУНКЦИИ

В [4, 5] было подробно описано построение графа, вычисляющего градиент сложной функции F . Напомним кратко, как строится этот граф. Будем называть граф, вычисляющий значение функции, графом F , а граф, вычисляющий градиент, графом G . Для того, чтобы построить граф G , нужно внести некоторые изменения в граф F . К каждой вершине графа F добавляются дополнительные элементы, вычисляющие частные производные функции f , соответствующей данной вершине, по входам x_i . Граф G имеет ту же структуру связей, что и граф F , но элементы заменяются на двойственные. Точка ветвления графа F заменяется на сумматор в графе G , вершина графа F , вычисляющая значение функции f , заменяется точкой ветвления в графе G , и, наконец, дополнительные элементы, вычисляющие частные производные $\partial f / \partial x_i$, умножаются на сигнал μ , выходящий из точки ветвления графа G . На рис. 3 показана вершина графа F и двойственная ей вершина графа G .

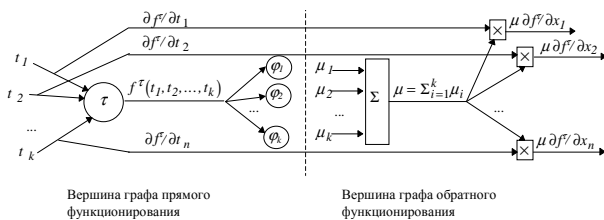


Рисунок 3 – Вершина графа F , вычисляющего значение сложной функции, и вершина графа G , вычисляющего ее градиент

ОЦЕНКИ ПОГРЕШНОСТЕЙ ВЫЧИСЛЕНИЯ ГРАДИЕНТА СЛОЖНОЙ ФУНКЦИИ

Мы получили формулы для оценки погрешностей входных сигналов графа, вычисляющего значение сложной функции многих переменных при заданной погрешности его выходного сигнала.

Рассмотрим теперь такую же задачу для графа, вычисляющего градиент сложной функции. То есть нам требуется оценить, какими могут быть погрешности входных вершин графа прямого функционирования, вычисляющего значение сложной функции F , если заданы допустимые погрешности для выходных вершин графа обратного функционирования, вычисляющего градиент F , т. е. вычислить допустимые погрешности ϵ_i переменных x_1, \dots, x_n , если заданы допустимые погрешности δ_i значений градиента.

Способ построения двойственного графа для оценки погрешностей вычисления сложной функции был описан в п. 3. В п. 4 было описано построение двойственного графа для вычисления градиента сложной функции. Следует различать эти двойственные графы. Как уже говорилось, эти графы имеют одинаковую структуру связей и одинаковое направление прохождения сигналов

по графам. Однако эти графы отличаются элементами и результатами вычисления.

До сих пор мы рассматривали граф, вычисляющий значение функции F , и граф, вычисляющий ее градиент, как два отдельных самостоятельных графа. Поскольку мы хотим оценивать погрешности вычисления градиента сложной функции, то требуется построить «развернутый» граф. «Развернутый» граф, состоит из графа, вычисляющего значение функции F и частные производные вершин по их входам, и двойственного ему графа, вычисляющего градиент этой функции (рис. 4).

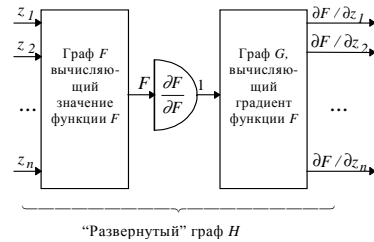


Рисунок 4 – «Развернутый граф», состоящий из графа, вычисляющего значение функции F и графа, вычисляющего градиент этой функции

Как и при вычислении допустимых погрешностей аргументов сложной функции, по заданной погрешности значения самой функции, мы строим граф, двойственный «развернутому» графу. Двойственный граф состоит из двух частей: из графа, двойственного графу F , и графа, двойственного графу G . Этот граф имеет структуру связей, аналогичную структуре связей «развернутого» графа, но элементы «развернутого» графа заменяются на двойственные и сигналы по двойственному графу движутся в направлении, противоположном движению сигналов в «развернутом» графе (рис. 5). Как и выше, сигналами двойственного графа являются погрешности сигналов, проходящих по «развернутому» графу.

Для простоты изложения в дальнейшем будем называть «развернутый граф – графом H , граф, двойственный графу H будем называть графом DH . Граф, двойственный графу G , будем называть графом DG и граф, двойственный графу F графом DF . Все двойственные графы вычисляют допустимые погрешности для соответствующих графов.

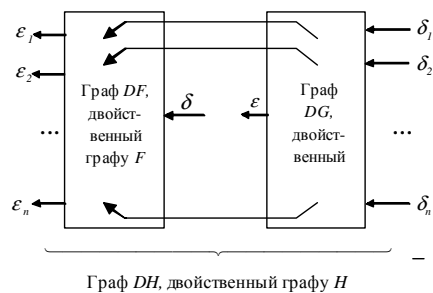


Рисунок 5 – Граф, вычисляющий допустимые погрешности, двойственный «развернутому» графу

Мы задаем значения входных сигналов δ_i графа DG и вычисляем допустимые погрешности вплоть до выходного сигнала этого графа. Затем на вход графа DF подаем погрешность δ и вычисляем допустимые погрешности проходя по элементам этого графа.

Перейдем к более подробному описанию процедуры вычисления допустимых погрешностей для графа H . Для этого рассмотрим изображенные на рис. 3 вершину f графа F и связанную с ней вершину g графа G . Для каждого из элементов этих вершин построим двойственные элементы, вычисляющие допустимые погрешности сигналов, проходящих по этим элементам в прямом направлении.

Входные сигналы графа DH одновременно являются входными сигналами графа DG . Поэтому сначала будем строить вершину \tilde{g} графа DG , двойственную вершине графа G . Вершина g состоит из сумматора, точки ветвления и умножителей, умножающих входящий к ним сигнал на соответствующие частные производные $\partial f / \partial x_i$. Чтобы построить вершину \tilde{g} графа DG , нужно определить двойственные элементы для всех элементов вершины g . Так как выходящие ребра вершины g относятся к умножителям, то входящие ребра вершины \tilde{g} принадлежат элементам, двойственным умножителям. Построим такие элементы. Допустимую погрешность ϵ , входящую к элементу, двойственному умножителю, нужно распределить между допустимой погрешностью сигнала μ (то есть сигнал μ должен изменяться в интервале $[\mu - \epsilon_\mu, \mu + \epsilon_\mu]$) и допустимой погрешностью сигнала $\partial f / \partial x_i$ (то есть сигнал $\partial f / \partial x_i$ должен изменяться в интервале $[\partial f / \partial x_i - \epsilon_\partial, \partial f / \partial x_i + \epsilon_\partial]$). Погрешности ϵ_μ и ϵ_∂ должны быть вычислены таким образом, чтобы произведение сигналов μ и $\partial f / \partial x_i$ с погрешностями отличалось от точного произведения $\mu \partial f / \partial x_i$ не более, чем на ϵ .

Как было показано в [8], для допустимых погрешностей ϵ_μ и ϵ_∂ справедлива следующая оценка: $\epsilon_\mu |\partial f / \partial x_i| + \epsilon_\partial |\mu| + \epsilon_\mu \epsilon_\partial \leq \epsilon$.

Отсюда можем получить формулу для вычисления допустимой погрешности ϵ_∂ :

$$\epsilon_\partial \leq \frac{\epsilon - \epsilon_\mu |\partial f / \partial x_i|}{|\mu| + \epsilon_\mu}.$$

В этой формуле две неизвестных величины: ϵ_∂ и ϵ_μ . Для того, чтобы вычислить ϵ_∂ , необходимо присвоить ϵ_μ какое-либо значение. Представим ϵ_μ в виде $\epsilon_\mu = k\epsilon$, где k – некоторый коэффициент. Тогда предыдущая формула принимает вид:

$$\epsilon_\partial \leq \frac{\epsilon - k\epsilon |\partial f / \partial x_i|}{|\mu| + k\epsilon}.$$

Вынося в числителе множитель ϵ , получаем $\epsilon(1 - k|\partial f / \partial x_i|)$. Для нас имеет смысл только положительное значение ϵ_∂ , поскольку ϵ_∂ характеризует длину интервала, в котором может изменяться величина $\partial f / \partial x_i$. Поэтому нам нужно, чтобы выполнялось неравенство $1 - k|\partial f / \partial x_i| > 0$, то есть $0 < k < \frac{1}{|\partial f / \partial x_i|}$. Выбирая k с учетом

полученного неравенства, находим значение ϵ_μ и затем, используя ϵ_μ , можем вычислить ϵ_∂ .

Следующим элементом, для которого необходимо построить двойственный элемент, является точка ветвления. Точка ветвления передает входящий к ней сигнал μ на умножители. Двойственный точке ветвления элемент, как и в пункте 3, выбирает минимум из допустимых погрешностей, входящих от элементов, двойственных умножителям:

$$\mu = \min\{\mu_i\}.$$

Двойственным элементом к сумматору является точка ветвления. Точка ветвления делит входящую к ней допустимую погрешность ϵ_μ на количество входов сумматора и передает далее полученные величины ϵ_μ / n :

$$\epsilon_i = \epsilon_\mu / n.$$

Мы построили двойственные элементы, составляющие вершину \tilde{g} графа DG . Построим теперь двойственные элементы вершины f графа DF .

Двойственным элементом к точке ветвления, как и выше, является элемент, выбирающий минимум из входящих к нему допустимых погрешностей:

$$\epsilon = \min\{\epsilon_i\}.$$

Двойственным к элементу, вычисляющему значение функции $f(x_1, x_2, \dots, x_n)$, как было показано выше, является элемент, вычисляющий допустимые погрешности сигналов x_1, x_2, \dots, x_n по формуле (2).

Далее находятся элементы, выбирающие минимум из входящих к ним допустимых погрешностей ϵ_∂ частных производных $\partial f / \partial x_i$ и допустимых погрешностей ϵ_i сигналов x_i .

$$\epsilon_i = \min\{\epsilon_{f_i}, \epsilon_{\partial_i}\}.$$

На рис. 6 показаны вершины графов DF и DG , двойственные вершинам графов F и G , вычисляющие допустимые погрешности.

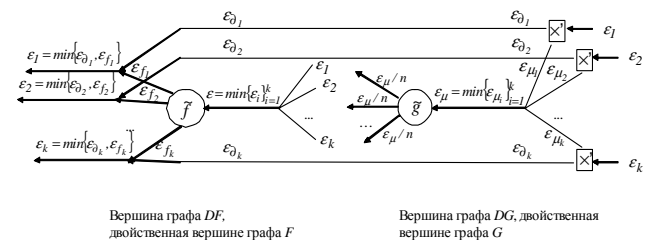


Рисунок 6 – Вершины графа, вычисляющего погрешности градиента

Мы получили формулы, по которым вычисляются допустимые погрешности сигналов на графе DH . Заметим, что для вычисления допустимых погрешностей сложной функции необходимо, чтобы функциональные символы

всех вершин были дифференцируемы. Это условие следует из формулы (2). Для вычисления допустимых погрешностей градиента уже необходимо, чтобы функциональный символ выходной вершины графа F был дважды дифференцируемым. Это опять следует из формулы (2), поскольку эта формула должна применяться к частным производным функционального символа выходной вершины графа F .

Таким образом, мы получили формулы для оценки погрешностей элементов графа, вычисляющего градиент сложной функции многих переменных, если аргументы функции имеют погрешности. Последовательно применяя эти формулы для каждого элемента графа, мы имеем возможность оценить погрешности вычисления градиента сложной функции.

ЗАКЛЮЧЕНИЕ

В работе изложен способ оценки погрешностей вычисления сложной функции многих переменных и ее градиента, если аргументы функции имеют погрешность. Вычисление функции и градиента представлено в виде графа. Это позволяет ускорить процесс вычисления за счет возможности параллельного выполнения вычислений для одного слоя графа. Кроме того, это позволяет легко строить двойственные графы для оценки погрешностей вычисления. При этом на двойственных графах вычисления проводятся так же параллельно.

ПЕРЕЧЕНЬ ССЫЛОК

1. Лебедев А.Н., Смолов В.Б. и др. Проектирование и расчет вычислительных машин непрерывного действия / Под ред. Лебедева А.Н. – М.: Машиностроение, 1966.
2. Лебедев А.Н. Основы теории точности счетно-решающих устройств, ч. 1. – Л.: Изд. ЛЭТИ им. В.И. Ленина, 1964.
3. Маликов И.М., Половко А.М., Романов Н.А., Чукреев П.Л. Основы теории и расчета надежности. – Л.: Судпромгиз, 1960.
4. Горбань А.Н., Сенашова М.Ю. Быстрое дифференцирование, двойственность и обратное распространение

- ошибок // Вычислительные технологии / Издательство СО РАН. – Новосибирск. 1999. Т. 4. С. 55–68.
5. Сенашова М.Ю. Быстрое дифференцирование сложных функций многих переменных // Радиоэлектроника, Информатика, Управление / Запорожский государственный технический университет. – Запорожье. 2000. №1. С. 101–106.
 6. Rummelhart D.E., Hinton G.E., Williams R.J. Learning representations by back-propagating errors. Nature, 1986. V. 323, 533–536.
 7. Сенашова М.Ю. Оценки погрешностей вычисления сложной функции многих переменных // Радиоэлектроника, Информатика, Управление / Запорожский государственный технический университет. – Запорожье. 2001. №2. С. 114–117.
 8. Сенашова М.Ю. Погрешности нейронных сетей. Вычисление погрешностей весов синапсов // Методы нейроинформатики: Сб. научн. трудов / Под ред. А.Н. Горбаня; отв. за вып. М.Г. Доррер. Красноярск; КГТУ. – 1998. – С. 48–64.
 9. Горбань А.Н. Обучение нейронных сетей. СП ПараГраф, М., 1990.
 10. Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере. Новосибирск, Наука, 1996.
 11. Горбань А.Н., Дунин-Барковский В.Л., Кирдин А.Н. и др. Нейроинформатика. Новосибирск. Наука. 1998.
 12. Senashova M. Yu., Gorban A. N., and Wunsch D. Back-Propagation of Accuracy. Proc. IEEE/INNS International Conference of Neural Networks, IEEE, USA, 1997, 1998–2001.
 13. Калмыков С.А., Шокин Ю.И., Юлдашев З.Х. Методы интервального анализа. Сиб. Отд-ние АН СССР, Новосибирск, 1986.
 14. Барцев С.И., Гилев С.Е., Охонин В.А. Принцип двойственности в организации адаптивных систем обработки информации. "Динамика химических и биологических систем", Наука, Новосибирск, 1989, 6–55.
 15. Rummelhart D.E., Hinton G.E., Williams R.J. Learning internal representation by error propagation. "Parallel Distributed Processing: Exploration in the Microstructure of Cognition" MA: MIT Press, Greate Britain, 1986. Vol. 1, pp. 318–362.
 16. Zurada J. M. Introduction to artificial neural systems. PWS Publishing Company, 1992.
 17. Haykin S. Neural networks. A comprehensive foundations. USA, 1994.

Надійшла 10.12.2003
Після доробки 03.06.2004

Описано графи обчислення складних функцій багатьох змінних, викладено принцип двоїстості для обчислення градієнта складної функції. Наведено алгоритми оцінювання похибок обчислень функцій та її градієнта.

Graphs are described for the calculation of composite functions of multiple variables, the duality principle for obtaining the gradient of a composite function is presented. Algorithms for the estimation of the computation error of the composite function and of its gradient are described.

УДК 519.859

А. М. Чугай

МЕТОД ПОИСКА ЛОКАЛЬНЫХ ЭКСТРЕМУМОВ В ЗАДАЧЕ УПАКОВКИ ОДИНАКОВЫХ ЦИЛИНДРОВ В МНОГОУГОЛЬНОЙ ПРИЗМЕ С ЗОНАМИ ЗАПРЕТА

Построена математическая модель и исследованы особенности задачи. Разработан метод поиска локальных экстремумов задачи. Поиск включает следующие этапы: построение крайних точек; поиск лучших крайних точек

модифицированным методом сужающихся окрестностей; локальная оптимизация методом возможных направлений. Для получения крайних точек строится дерево решений. Приведен численный пример.

ВВЕДЕНИЕ

Задача упаковки цилиндров в многоугольной призме с зонами запрета относится к классу задач геометрического проектирования. Задачи геометрического проектирования возникают в различных сферах человеческой деятельности, областях науки и техники и состоят в поиске оптимального размещения множества геометрических объектов относительно области размещения при наличии заданных технологических ограничений и критериев качества размещения. К данному классу задач относятся задачи компоновки оборудования, задачи управления сложными техническими системами, задачи построения генеральных планов промышленных предприятий, оптимального раскроя промышленных материалов, рационального использования отходов, оптимального размещения грузов, задачи размещения модулей на радиоэлектронной плате, задачи календарного планирования и многие другие задачи.

Решению задачи размещения цилиндров разных диаметров в параллелепипеде (задача загрузки трюмов корабля) посвящены работы таких зарубежных авторов как [1, 2]. Данная задача сводится к двумерной задаче размещения кругов разных радиусов в прямоугольнике. Авторы строят математическую модель задачи в терминах целочисленного линейного программирования. Рассматриваются разные эвристические подходы к решению задачи (генетический алгоритм, квази-вероятностный алгоритм и т. п.). В [2] предложены три эвристических алгоритма, которые позволяют решить эту задачу в трехмерной постановке.

Анализ работ по средствам моделирования и решения трехмерных задач размещения позволяет сделать вывод, что подавляющее большинство зарубежных публикаций посвящено размещению прямых параллелепипедов в параллелепипеде или наборе параллелепипедов, которые, как правило, сводятся к двумерным задачам. Предлагаемые методы решения являются преимущественно эвристическими.

Таким образом, необходимо проведение исследований по созданию адекватных математических моделей и разработке эффективных методов решения 3D задач размещения геометрических объектов.

ПОСТАНОВКА ЗАДАЧИ

Пусть имеются одинаковые цилиндры $C_i (i=1, 2, \dots, N)$ радиуса r и задана многоугольная призма, основание которой можно представить в виде многосвязной области P , образованной вычитанием из выпуклого многоугольника областей, которые будем называть зонами запрета на размещение цилиндров:

$$P = \text{cl} \left(P_0 \setminus \left(\bigcup_{j=1}^{q-1} Z_j \right) \right) \subset R^2,$$

где $\text{cl}(A)$ – замыкание A , P_0 – выпуклый многоугольник, Z_j – выпуклые многоугольники или объединение

конечного числа кругов. При этом, цилиндры ориентированы так, что их оси перпендикулярны основанию призмы, которое принадлежит плоскости $ХОУ$. Обозначим координаты центра основания цилиндра C_i через $v_i = (x_i, y_i)$.

Задача. Необходимо разместить в призме максимальное количество одинаковых цилиндров C_i . Так как все цилиндры одинаковые, то каждый цилиндр можно заменить его проекцией на плоскость основания призмы P . Тогда задача из трехмерной постановки может быть сведена к двумерной задаче упаковки максимального количества одинаковых кругов C_i в многосвязной области P .

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ЗАДАЧИ

Решение поставленной задачи можно свести к решению последовательности задач

$$F(u) = \max_{u \in D} F_n(u), \quad (1)$$

где $F_n(u) = \sum_{i=1}^n r_i$, $u = (v_1, v_2, \dots, v_n, r_1, r_2, \dots, r_n)$, $u_i = (v_i, r_i)$, r_i – радиус цилиндра (круга) C_i ; D описывается системой неравенств [3, 4]:

$$\begin{cases} \Phi_i(u_i) \geq 0, i = 1, 2, \dots, n; \\ \Phi_{ij}(u_i, u_j) \geq 0, i < j = 1, 2, \dots, n; \\ r - r_i \geq 0, i = 1, 2, \dots, n; \\ r_i \geq 0, i = 1, 2, \dots, n; \end{cases} \quad (2)$$

$$\Phi_i(u_i) = \min \{ \Omega_i(u_i), \min_{j=1, 2, \dots, m} \{ \theta_{ij}(u_i) \}, \min_{j=m+1, m+2, \dots, q-1} \{ \Psi_{ij}(u_i) \} \};$$

$$\Omega_i(u_i) = \min_{k=1, 2, \dots, n_0} \{ \varphi_{ik}(u_i) \};$$

$$\varphi_{ik}(u_i) = A_{0k}x_i + B_{0k}y_i + \tilde{C}_{i0k},$$

$$A_{0k} = (y_{0k+1} - y_{0k}) / d_{0k},$$

$$B_{0k} = -(x_{0k+1} - x_{0k}) / d_{0k},$$

$$\tilde{C}_{i0k} = C_{0k} - r_i,$$

$$C_{0k} = (x_{0k+1}y_{0k} - x_{0k}y_{0k+1}) / d_{0k},$$

$$d_{0k} = \sqrt{(x_{0k+1} - x_{0k})^2 + (y_{0k+1} - y_{0k})^2},$$

(x_{0k}, y_{0k}) и (x_{0k+1}, y_{0k+1}) – координаты вершин V_{0k} и V_{0k+1} многоугольника P_0 ;

$$\theta_{ij}(u_i) = \max_{k=1, 2, \dots, n_j} \{ \max \{ \chi_{ijk}(u_i), \min \{ \xi_{ijk}(u_i), \omega_{ijk}(u_i) \} \} \};$$

$$\chi_{ijk}(u_i) = A'_{jk}x_i + B'_{jk}y_i + \tilde{C}'_{ijk},$$

$$A'_{jk} = -(y_{jk+1} - y_{jk}) / d_{jk},$$

$$B'_{jk} = (x_{jk+1} - x_{jk}) / d_{jk},$$

$$\tilde{C}'_{ijk} = C'_{jk} - r_i,$$

$$C'_{jk} = -(x_{jk+1}y_{jk} - x_{jk}y_{jk+1}) / d_{jk},$$

$$d_{jk} = \sqrt{(x_{jk+1} - x_{jk})^2 + (y_{jk+1} - y_{jk})^2};$$

(x_{jk}, y_{jk}) и (x_{jk+1}, y_{jk+1}) – координаты вершин V_{jk} и V_{jk+1} многоугольника Z_j , $j = 1, 2, \dots, m$;

$$\xi_{ijk}(u_i) = A^*_{ijk}x_i + B^*_{ijk} + C^*_{ijk},$$

$$A^*_{ijk} = -(y_{ijk}^1 - y_{ijk}^0),$$

$$B^*_{ijk} = x_{ijk}^1 - x_{ijk}^0,$$

$$C^*_{ijk} = -(x_{ijk}^1 y_{ijk}^0 - x_{ijk}^0 y_{ijk}^1),$$

$$(x_{ijk}^0, y_{ijk}^0) = (x_{jk}, y_{jk}) + v_{ijk-1},$$

$$(x_{ijk}^1, y_{ijk}^1) = (x_{ijk}, y_{ijk}) + v_{ijk},$$

$$v_{ijk} = \frac{-r_i}{\sqrt{(A'_{jk})^2 + (B'_{jk})^2}} \begin{pmatrix} A'_{jk} \\ B'_{jk} \end{pmatrix};$$

$$\omega_{ijk}(u_i) = (x_i - x_{jk})^2 + (y_i - y_{jk})^2 - r_i^2;$$

$$\Psi_{ij}(u_i) = \min_{k=1, 2, \dots, n_j} \{\Psi_{ijk}(u_i)\},$$

$$\Psi_{ijk}(u_i) = (x_i - x'_{jk})^2 + (y_i - y'_{jk})^2 - (r_i + r'_{jk})^2,$$

x'_{jk} , y'_{jk} и r'_{jk} – соответственно координаты центра и радиус k -го круга из Z_j -ой области,

$$j = m + 1, m + 2, \dots, q - 1;$$

$$\Phi_{ij}(u_i, u_j) = (x_i - x_j)^2 + (y_i - y_j)^2 - (r_i + r_j)^2.$$

Таким образом, если $F_n(u^*) < nr$, а $F_{n-1}(u^*) = (n-1)r$, то решение задачи (1) достигается в $u^* = (v^*_1, v^*_2, \dots, v^*_{n-1}, r^*_1, r^*_2, \dots, r^*_{n-1})$.

ОСОБЕННОСТИ ЗАДАЧИ

1. Для каждого локального максимума u^* всегда найдется крайняя точка $u' \in D$ [5], определяемая системой из $3n$ уравнений (активных неравенств), входящих в систему (2), такая что $F_n(u^*) = F_n(u')$. (Точка $u' \in D$ называется крайней точкой D , если для любой окрестности $N_{u'} \subset R^{3n}$ точки u' существует гиперплоскость $G \subset R^{3n}$, такая, что $D \cap N_{u'} \cap G = \{u\}$).

2. Локальные максимумы u^* могут быть нестрогими.

3. Задача является многоэкстремальной; число локальных экстремумов u^* значительно больше $n!$, т. е. задача (1) NP – трудная [6];

4. Для того чтобы найти глобальный максимум $F_n(u)$, достаточно перебрать все крайние точки $u' \in D$;

5. В настоящее время не существует метода, позволяющего получить глобальный максимум поставленной задачи, если $n \geq 10$ [5].

МЕТОД РЕШЕНИЯ

Первоначально производится оценка величины $n = \mu$, так что n кругов гарантированно размещаются в P . Для этого можно воспользоваться, например, методом после-

довательно-одиночного размещения. Далее решается задача (1) для $n = \mu + 1$.

Общая стратегия решения задачи (1) может быть представлена в виде следующих этапов:

- 1) построение крайних точек;
- 2) поиск крайних точек с наибольшими значениями целевой функции с помощью модифицированного метода сужающихся окрестностей;
- 3) переход из полученных лучших крайних точек в каждой из сформированных окрестностей к локальным максимумам с помощью метода локальной оптимизации;
- 4) выбор лучшего локального максимума в качестве приближения к оптимальному решению задачи (глобальному максимуму).

ПОСТРОЕНИЕ КРАЙНИХ ТОЧЕК

Процесс получения крайней точки можно представить как процедуру последовательного выделения гиперплоскостей и гиперповерхностей, уравнения (активные неравенства) которых входят в систему (2). Чтобы описать все крайние точки области D , необходимо построить полное дерево решений [7], которое позволяет получить все системы $3n$ уравнений, соответствующие активным неравенствам из системы (2). Так как число крайних точек, значительно больше $n!$, то для решения примеров средней сложности ($n \geq 10$), такой подход не пригоден из-за своей вычислительной сложности, если только не будут найдены очень сильные правила отсекаания. Поэтому для построения крайних точек $u' \in D$ в исследовании используется дерево T возможных размещений кругов в области P , конечные вершины которого определяют системы уравнений, формирующие некоторое подмножество крайних точек из D . Однако весьма значительное множество вершин дерева символизирует несовместные системы.

Дерево решений имеет n уровней (по количеству рассматриваемых кругов). Корню дерева ставится в соответствие пространство R^{3n} . Вершины l -го уровня дерева соответствуют множествам размерности $3(n-l)$, $l = 1, 2, \dots, n$, образованным пересечением множеств размерности $3(n-l+1)$ (полученных на предыдущем уровне) с тремя гиперповерхностями, задаваемыми уравнениями (активными неравенствами) из системы (2), которые содержат переменные x_l, y_l, r_l . При этом, l -й уровень состоит из

$$\prod_{i=1}^l \left(in_0 + (i-1) \sum_{k=1}^i n_k + \frac{1}{2}(i-1)(i-2) \right) \text{ вершин.}$$

Каждой из этих вершин соответствует система уравнений, решение которой (в случае совместности системы) определяет параметры размещения кругов $C_i, i = 1, 2, \dots, l$.

На первом уровне дерева рассматриваются n_0 систем из трех уравнений вида:

$$\Phi_{1k}(u_1) = 0, \Phi_{1k+1}(u_1) = 0, r - r_1 = 0; k = 1, 2, \dots, n_0;$$

которые соответствуют размещению круга C_1 , при котором он касается двух соседних сторон P_0 .

На втором уровне рассматривается $2n_0 + \sum_{i=1}^m n_i$ систем из шести уравнений вида (при этом, первые три уравнения определяются системой, которая соответствует родительской вершине первого уровня):

$$\varphi_{2k}(u_2) = 0, \varphi_{2k+1}(u_1) = 0, r - r_2 = 0; \\ k = 1, 2, \dots, n_0;$$

$$\Phi_{12}(u_1, u_2) = 0, \varphi_{2k}(u_2) = 0, r - r_2 = 0; \\ k = 1, 2, \dots, n_0;$$

$$\Phi_{12}(u_1, u_2) = 0, \varphi_{2jk}(u_2) = 0, r - r_2 = 0; \\ j = 1, 2, \dots, m; k = 1, 2, \dots, n_k;$$

которые соответствуют размещению круга C_2 , при котором он касается либо двух соседних сторон P_0 , либо круга C_1 и одной из сторон P_0 , либо круга C_1 и одной из сторон многоугольной области Z_j .

На третьем уровне рассматривается $3n_0 + 2 \sum_{i=1}^m n_i + 1$ систем из девяти уравнений вида (при этом, первые шесть уравнений определяются системой, которая соответствует родительской вершине второго уровня):

$$\varphi_{3k}(u_3) = 0, \varphi_{3k+1}u_3 = 0, r - r_3 = 0; \\ k = 1, 2, \dots, n_0;$$

$$\Phi_{i3}(u_p, u_3) = 0, \varphi_{3k}(u_3) = 0, r - r_3 = 0; \\ i = 1, 2; k = 1, 2, \dots, n_0;$$

$$\Phi_{i3}(u_p, u_3) = 0, \varphi_{3k}(u_3) = 0, r - r_3 = 0; \\ i = 1, 2; j = 1, 2, \dots, m; k = 1, 2, \dots, n_k;$$

$$\Phi_{13}(u_1, u_3) = 0, \Phi_{23}(u_2, u_3) = 0, r - r_3 = 0;$$

которые соответствуют размещению круга C_3 , при котором он касается либо двух соседних сторон P_0 , либо одного из кругов C_i ($i = 1, 2$) и одной из сторон P_0 , либо одного из кругов C_i ($i = 1, 2$) и одной из сторон многоугольной области Z_j , либо двух кругов C_1 и C_2 .

На четвертом уровне рассматривается $4n_0 + 3 \sum_{i=0}^m n_i + 3$ систем из двенадцати уравнений вида (при этом, первые девять уравнений определяются системой, которая соответствует родительской вершине третьего уровня):

$$\varphi_{4k}(u_4) = 0, \varphi_{4k+1}u_4 = 0, r - r_4 = 0; \\ k = 1, 2, \dots, n_0;$$

$$\Phi_{i4}(u_p, u_4) = 0, \varphi_{4k}(u_4) = 0, r - r_4 = 0; \\ i = 1, 2, 3; k = 1, 2, \dots, n_0;$$

$$\Phi_{i4}(u_p, u_4) = 0, \chi_{4jk}(u_4) = 0, r - r_4 = 0; \\ i = 1, 2, 3; j = 1, 2, \dots, m; k = 1, 2, \dots, n_k;$$

$$\Phi_{14}(u_1, u_4) = 0, \Phi_{24}(u_2, u_4) = 0, r - r_4 = 0;$$

$$\Phi_{14}(u_1, u_4) = 0, \Phi_{34}(u_3, u_4) = 0, r - r_4 = 0;$$

$$\Phi_{24}(u_2, u_4) = 0, \Phi_{34}(u_3, u_4) = 0, r - r_4 = 0;$$

которые соответствуют размещению круга C_4 , при котором он касается либо двух соседних сторон P_0 , либо одного из кругов C_i ($i = 1, 2, 3$) и одной из сторон P_0 , либо одного из кругов C_i ($i = 1, 2, 3$) и одной из сторон многоугольной области Z_j , либо одной из следующих пар кругов: C_1 и C_2 , C_1 и C_3 , C_2 и C_4 . Аналогично четвертому уровню строятся вершины на всех оставшихся уровнях дерева.

Из всех решений описанных выше систем, выбираются только те, которые обеспечивают непересечение C_i и C_j , $j = 1, 2, \dots, i-1$, и принадлежность C_i области P .

В случае, если на l -м уровне в выбранной вершине система является несовместной либо ее решение дает точку вне D при $r_l = r$, то на этом уровне производится поиск ближайшей допустимой вершины, в которой решение соответствующей ей системы дает точку области D при $r_l = r$. Если такую вершину найти не удастся, то на данном l -м уровне выбирается вершина, в которой решение соответствующей ей системы дает точку области D с максимально допустимым значением переменной r_l . Для этого в каждой вершине данного уровня решается вспомогательная оптимизационная задача поиска максимального значения величины r_l , при котором решение соответствующей этой вершине системы уравнений дает точку области D . Для поиска величины r_l , используется метод половинного деления, на каждой итерации которого решается соответствующая рассматриваемой вершине система уравнений при определенном значении r_l . В результате решения этой задачи дополнительно становится активным хотя бы одно из ограничений системы (2), которое не входит в систему, соответствующую выбранной вершине дерева.

Очевидно, что размещение кругов в P по описанному дереву в общем случае не соответствует локальному максимуму целевой функции, но оно является некоторым приближением к ближайшему максимуму. Поэтому для решения задачи предлагается применить способ поиска приближения к глобальному максимуму u^* , состоящий из направленного перебора модифицированным методом сужающихся окрестностей [5, 8] крайних точек области D . Для этого, в соответствии с определенным алгоритмом, случайным образом генерируются последовательности $\pi_i = \{\pi_{i_1}, \pi_{i_2}, \dots, \pi_{i_n}\} \in \Pi$ комбинаторного множества Π , элементы которых определяются номерами вершин на соответствующих уровнях дерева T . В соответствии со сгенерированными последовательностями определяются крайние точки D . Таким образом, каждой последовательности $\pi_i \in \Pi$ ставится во взаимнооднозначное соответствие крайняя точка $u'_i \in D$.

Из полученных в соответствии с модифицированным методом сужающихся окрестностей крайних точек

выбираются точки с наибольшими значениями целевой функции в каждой из окрестностей. Каждая из выбранных таким образом крайних точек u'_t , $t = 1, 2, \dots, \delta$ затем берется в качестве начальной точки для поиска локального максимума u^*_t . В итоге, точка $u^* = \arg \max \{F(u): t = 1, 2, \dots, \delta\}$ выбирается в качестве приближения к глобальному максимуму задачи.

ЛОКАЛЬНАЯ ОПТИМИЗАЦИЯ

Для отыскания локальных максимумов рассматриваемой задачи, которая является задачей нелинейного программирования, применяется метод оптимизации, который основывается на стратегии метода возможных направлений [9, 10]. Идея метода состоит в следующем. Найденная крайняя точка $u'_i \in D$ становится начальной для поиска локального максимума u^*_i . В полученной крайней точке u'_i переопределяется вектор переменных $u = (v_1, v_2, \dots, v_n, r_1, r_2, \dots, r_n)$ таким образом, что из него исключаются радиусы тех кругов, которые приняли значение равное r , тем самым уменьшается общая размерность задачи (1) и вектор переменных принимает вид $u = (v_1, v_2, \dots, v_n, r_{i_1}, r_{i_2}, \dots, r_{i_t})$.

Пусть имеем j -е приближение \bar{u}_j к локальному максимуму (при этом, принимаем $\bar{u}_0 = u'_i$). Для того чтобы улучшить значение целевой функции (1) рассчитывается направление из точки \bar{u}_j , которое будет учитывать локальные свойства множества допустимых решений в окрестности рассматриваемой точки. В точке \bar{u}_j снова переопределяется вектор переменных. При вычислении направления улучшения целевой функции (1) в точке \bar{u}_j строится конус возможных направлений, задаваемый системой линейных неравенств (3):

$$\begin{cases} (\nabla F_n(\bar{u}_j), \lambda^j) \geq d^j \\ (\nabla \bar{\Psi}_1(\bar{u}_j), \lambda^j) \geq d^j \\ (\nabla \bar{\Psi}_2(\bar{u}_j), \lambda^j) \geq d^j \\ \vdots \\ (\nabla \bar{\Psi}_s(\bar{u}_j), \lambda^j) \geq d^j \\ -1 \leq \lambda^j \leq 1 \end{cases} \quad (3)$$

где $F_n(\bar{u}_j)$ – целевая функция задачи (1), $\bar{\Psi}_1(\bar{u}_j), \bar{\Psi}_2(\bar{u}_j), \dots, \bar{\Psi}_s(\bar{u}_j)$ – неравенства из системы (2), являющиеся ϵ -активными в точке \bar{u}_j , $\lambda^j = (\lambda^j_1, \lambda^j_2, \dots, \lambda^j_{2n+\tau})$.

В построенном конусе возможных направлений выбирается вектор λ^j конечной длины, который обеспечивает наибольшее увеличение целевой функции (1). Для этого решается следующая вспомогательная задача: максимизировать величину скалярного произведения α^j при ограничениях, описываемых системой (3). Эта задача представляет собой задачу линейного программирования в $(2n + \tau + 1)$ – пространстве векторов (λ^j, α^j) . Множество допустимых планов этой задачи замкнуто, ограничено и непусто, так как $\lambda^j = 0, \alpha^j = 0$ является

допустимым планом. Значит, вспомогательная задача имеет решение (λ^j, α^j) , причем $\alpha^j \geq 0$. Если $\alpha^j > 0$, то нетрудно показать, что направление λ^j является возможным направлением возрастания целевой функции (1), т. е. точка $\bar{u}_{j+1} = \bar{u}_j + t\lambda^j$ при достаточно малом t принадлежит множеству D и обеспечивает большее значение целевой функции (1), чем \bar{u}_j . Значение шага t , при котором происходит максимальное увеличение целевой функции (1), определяется методом половинного деления. Это позволяет определить расстояние от точки \bar{u}_j до ближайшей в найденном направлении λ^j предельной точки $\bar{u}_{j+1} \in D$.

Если $\alpha^j = 0$, то получается стационарная точка процесса, которая для задачи выпуклого программирования дает решение, а в общем случае требует дополнительных исследований (т. е. полученная точка может быть либо локальным экстремумом, либо нестрогим локальным экстремумом, либо седловой точкой). Предложена специальная методология определения седловых точек, основанная на анализе множителей Лагранжа, полученных при решении модифицированным симплекс-методом [11] вспомогательной задачи поиска вектора (λ^j, α^j) . Выход из седловых точек обеспечивается уменьшением размерности задачи. Для этого в седловой точке фиксируются значения переменных, входящих в ограничения, которым соответствуют положительные множители Лагранжа, определенные на последней итерации решения задачи поиска вектора (λ^j, α^j) в рассматриваемой точке \bar{u}_j . В случае фиксации всех переменных точка \bar{u}_j считается локальным максимумом.

РЕЗУЛЬТАТЫ РЕШЕНИЯ ЗАДАЧИ

Задача. Необходимо упаковать максимальное количество одинаковых цилиндров радиуса $r = 28$ в призму, основанием которой является многосвязная область P , заданная следующими данными: координатами вершин многоугольника $P_0 - V_{01} = (30,400), V_{02} = (190,470), V_{03} = (300,470), V_{04} = (450,400), V_{05} = (520,310), V_{06} = (540,150), V_{07} = (340,30), V_{08} = (120,85), V_{09} = (10,250)$; координатами вершин многоугольника $Z_1: V_{11} = (340,270), V_{12} = (370,340), V_{13} = (420,260), V_{14} = (400,210), V_{15} = (355,220)$; координатами вершин многоугольника $Z_2: V_{21} = (180,380), V_{22} = (280,330), V_{23} = (-250,300), V_{24} = (200,310)$; координатами центров и радиусов кругов области $X_3: (180,180,28), (225,200,32), (220,160,26)$.

Результат решения задачи (1) при $n = 46$ показан на рис. 1. При этом, $r_i = r = 28, i = 1, 2, \dots, n$. Координаты $v_i = (x_i, y_i)$ центров оснований цилиндров представлены в таблице 1.

Результат решения задачи (1) при $n = 47$ показан на рис. 2. При этом, $r_i = r = 28, i = 1, 2, \dots, n - 1$, а $r_n = 20, 7926 < r$.

Таким образом, решение задачи упаковки цилиндров радиуса $r = 28$ в призму с заданным в примере основанием P достигается при $n = 46$.

Таблица 1 – Координаты центров оснований упакованных цилиндров, $n = 46$

i		1	2	3	4	5	6	7
1–7	x_i	316,349	208,873	55,161	287,420	221,457	341,549	47,778
	y_i	249,841	279,203	374,403	297,805	391,045	195,434	318,887
8–14	x_i	152,009	40,035	286,018	159,977	104,003	197,149	450,488
	y_i	287,288	263,408	202,756	399,603	401,814	441,501	215,784
15–21	x_i	334,398	253,151	161,104	96,096	173,665	67,322	98,385
	y_i	328,309	441,646	342,553	290,548	235,642	214,495	167,900
22–28	x_i	105,192	416,612	129,448	319,956	309,113	275,071	387,496
	y_i	345,818	319,139	121,305	114,314	434,844	147,829	398,270
29–35	x_i	335,440	366,596	446,289	498,679	277,415	505,585	117,742
	y_i	60,482	145,334	271,643	251,832	388,660	196,247	238,894
36–42	x_i	382,080	436,669	479,522	180,850	281,092	333,258	236,213
	y_i	91,505	371,436	146,670	99,063	73,993	384,315	107,501
43–46	x_i	472,303	257,087	416,074	431,556	–	–	–
	y_i	325,101	250,713	171,586	117,765	–	–	–

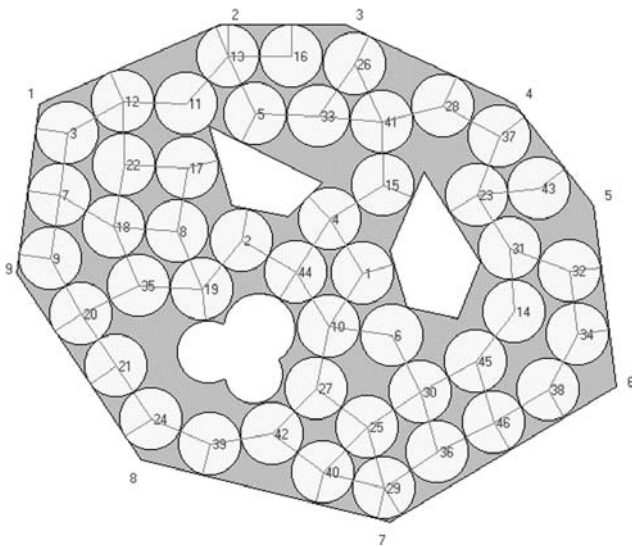


Рисунок 1 – Решение задачи упаковки одинаковых цилиндров, $n = 46$

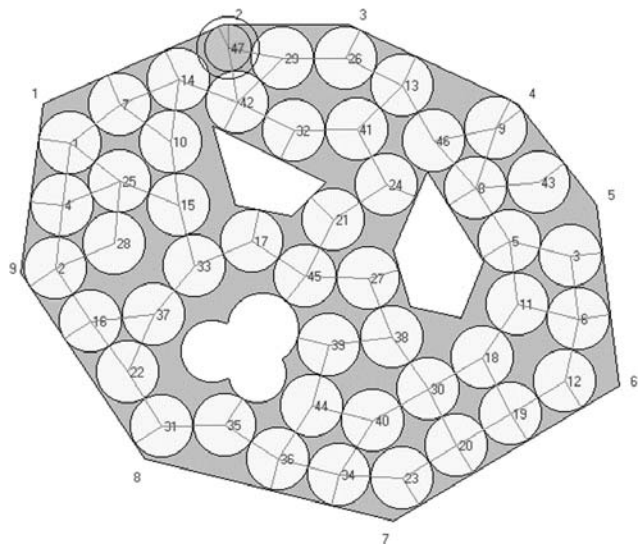


Рисунок 2 – Решение задачи упаковки одинаковых цилиндров, $n = 47$

ВЫВОДЫ

Задача упаковки одинаковых цилиндров в многоугольной призме с зонами запрета сведена к двумерной задаче упаковки одинаковых кругов в выпуклом многоугольнике с зонами запрета. Построена математическая модель задачи и исследованы ее особенности. Разработан метод поиска локальных экстремумов задачи, который основывается на комбинации следующих методов: метода получения крайних точек с помощью дерева решений; модифицированного метода сужающихся окрестностей; метода локальной оптимизации, основанного на стратегии метода возможных направлений. На основа-

нии построенной модели и разработанного метода решения создан программный продукт. Предложенный в статье подход к решению задачи развивает существующие методы решения многоэкстремальных задач оптимизационного геометрического проектирования. Его применение может быть расширено на решение и многих других многоэкстремальных задач, которые требуют организации перебора многочисленных локальных экстремумов.

ПЕРЕЧЕНЬ ССЫЛОК

1. George J.A., George J.M. and Lamar B.W. Packing different-sized circles into a rectangular container // European Journal of Operational Research. – 1995. – N. 84. – P.693–712.

2. George J.A. Multiple container packing: a case study of pipe packing // Journal of the Operational Research Society. – 1996. – N. 47. – P.1098–1109.
3. Stoyan Yu.G. Φ -function and its basic properties // Доп. НАН України. – 2001. – № 8. – С. 112–117.
4. Stoyan Y., Terno J., Scheithauer G., Gil N., Romanova T. Φ -functions for primary 2D-objects. Preprint MATH–NM–15–2001, Technical University of Dresden, 2001.
5. Stoyan Yu., Yaskov G., Scheithauer G. Packing of various radii solid spheres into a parallelepiped. Preprint MATH–NM–17–2001, Technical University of Dresden, 2001.
6. Пападимитриу Х., Стайглиц К. Комбинаторная оптимизация. Алгоритмы и сложность: Пер. с англ. – Москва: Мир, 1985. – 512 с.
7. Стоян Ю.Г., Аристова И.В., Яськов Г.Н. Адаптация метода ветвей и границ для решения задачи размещения прямоугольников с учетом минимально и максимально допустимых расстояний. – Харьков, 1995. – 35 с. – (Препр. НАН Украины. Ин-т пробл. машиностроения; № 384).
8. Стоян Ю.Г., Соколовский В.З. Решение некоторых многоэкстремальных задач методом сужающихся окрестностей. – Киев: Наук. думка, 1980. – 208 с.
9. Polak E. Computation methods in optimization. New York and London: Academic Press, 1971. – 329 p.
10. Зонтендейк Г. Методы возможных направлений. Изд-во иностранной литературы, Москва, 1963. – 176 с.
11. Гил Ф., Мюррей У., Райт М. Практическая оптимизация: Пер. с англ. – Москва: Мир, 1985. – 509 с.

Надійшла 18.05.2004

Після доробки 27.11.2004

Побудована математична модель та досліджені властивості задачі. Розроблено метод пошуку локальних екстремумів задачі. Пошук включає наступні етапи: побудова крайніх точок; пошук кращих крайніх точок модифікованим методом околць, що звужуються; локальна оптимізація методом можливих напрямків. Для отримання крайніх точок будується дерево рішень. Наведено чисельний приклад.

A mathematical model of the problem is constructed and its peculiarities are investigated. A method of searching for local extremums of the problem is developed. Searching includes following stages: construction of extreme points; searching for best extreme points by a modified method of convergent neighbourhood; local optimization by a method of feasible directions. For reception of extreme points a tree of decisions is constructed. A numerical example is given.

НЕЙРОИНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

НЕЙРОИНФОРМАТИКА И ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

NEUROINFORMATICS AND INTELLIGENT SYSTEMS

УДК 681.324: 519.713

В. П. Авраменко, В. В. Калачева, А. В. Цурихин

АДАПТИВНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ СОСТОЯНИЙ ПРОИЗВОДСТВЕННЫХ СИСТЕМ

Исследуются адаптивные методы прогнозирования состояний рабочих систем, процессы в которых описываются динамическими рядами. Состояние системы определяется конечноразностными уравнениями, подобными дифференциальным уравнениям Лагранжа первого рода. Спецификацию прогнозирующей модели состояний выполнено на моделях типа ARMA и ARMAX. На основе стохастического вектора состояний разработана обобщенная векторно-матричная модель и предложен рекуррентный метод прогнозирования калмановского типа. Предложенный метод прогнозирования состояний внедрено в системы поддержки принятия решений для оперативного управления производственными процессами.

ПОСТАНОВКА ПРОБЛЕМЫ ОЦЕНИВАНИЯ СОСТОЯНИЙ ПРОИЗВОДСТВЕННЫХ СИСТЕМ

Важной задачей оперативного управления производством является формирование последовательности управляющих воздействий, которая обеспечит желаемое поведение системы. Применение традиционных методов выработки управляющих воздействий наталкивается на серьезные трудности, связанные с изменением во времени принятых исходных предпосылок и отсутствием достаточного количества информации о свойствах управляемых объектов и условиях их функционирования.

В этой ситуации актуальной становится задача применения адаптивных методов идентификации параметров и состояний управляемого объекта по наблюдениям за входными и выходными сигналами. Математическое описание входных и выходных сигналов управляемого объекта принято представлять стохастическими моделя-

ми гипотетически линейных дискретных систем, на выходе которых формируются процессы с аналогичными корреляционными свойствами из поступающего на вход некоррелированного шума.

Под состоянием производственной системы подразумевается совокупность существенных значений переменных и параметров, характеризующих основные технические, технологические и экономические показатели производственной системы в конкретный момент времени. Если в исследуемой задаче основными показателями выступают производственная мощность предприятия P , численность работающих R и наличие ресурсов S , обеспечивающих выпуск продукции M , то состояние системы $SS(t_1)$ в момент времени t_1 можно описать неизменным параметром P , значениями показателей $R(t_1)$, $S(t_1)$, $M(t_1)$ и вектором состояния (кортежем, если среди показателей системы имеют место качественные величины, не имеющие численных значений)

$$SS(t_1) = \langle P, R(t_1), S(t_1), M(t_1) \rangle,$$

а в момент времени t_2 – показателями P , $R(t_2)$, $S(t_2)$, $M(t_2)$ и вектором состояния

$$SS(t_2) = \langle P, R(t_2), S(t_2), M(t_2) \rangle.$$

Изменение вектора состояния системы $SS(t)$ во времени t можно рассматривать как фазовую траекторию системы в пространстве состояний. Начало траектории может совпадать с состоянием системы в момент принятия плана. На основании начального состояния

можно найти прогнозируемые значения параметров $R(t)$, $S(t)$, $M(t)$ будущих состояний $SS(t)$ и определить последовательность желаемых управляющих воздействий на систему. Конечное состояние системы обычно определяется конечной целью принимаемого решения.

Предположим, что производственная система состоит из объекта управления в виде «черного ящика», устройства управления с прогнозирующей моделью и блока формирования цели управления. Величины, участвующие в исследовании «черного ящика», разобьем на две группы: выходные переменные $Y = [Y^{(1)}, Y^{(2)}, \dots, Y^{(k)}]^T \in R^k$, характеризующие состояние объекта управления, и входные переменные, которые подразделяются на управляющие $X = [X^{(1)}, X^{(2)}, \dots, X^{(s)}]^T \in R^s$ и возмущающие $P = [P^{(1)}, P^{(2)}, \dots, P^{(q)}]^T \in R^q$ воздействия. Внутренние свойства объекта преобразуют управляющие $X(t) \in R^s$ и возмущающие $P(t) \in R^q$ воздействия в выходной сигнал $Y(t) = R^k$, характеризующий состояние объекта в момент времени t .

АНАЛИЗ РАБОТ ПО АДАПТИВНОМУ ПРОГНОЗИРОВАНИЮ СОСТОЯНИЯ ПРОИЗВОДСТВЕННЫХ СИСТЕМ

Первые практические применения методов статистического моделирования относятся ко временам А. М. Лежандра и К. Гаусса, которые дали научное обоснование метода наименьших квадратов и метода максимального правдоподобия. Методы статистического описания нелинейных динамических процессов, представленных временными рядами, получили свое развитие в работах Дж. Бокса и Г. Дженкинса [1].

При моделировании производственных процессов актуальной становится задача оценивания параметров и состояний управляемого объекта по наблюдениям за входными и выходными сигналами. В качестве математического описания входных и выходных сигналов часто используются стохастические модели гипотетических линейных дискретных систем, на выходе которых формируются процессы с аналогичными корреляционными свойствами из поступающего на вход некоррелированного шума [2–3].

Исследование временного ряда традиционно предусматривает построение трендовой, сезонной, циклической и стохастической составляющих композиционной модели производственного процесса. Трендовая составляющая отражает длительные изменения выборочного среднего или дисперсии временного ряда. Во временных рядах могут содержаться сезонные колебания, которые завершаются в течение одного года. Если период колебаний составляет несколько лет, то временной ряд содержит циклические колебания. Составляющая временного ряда, оставшаяся после вычитания из него трендовой, сезонной и циклической компонент, представляет собой стохастическую составляющую, называемую рядом остатков [4–5].

Стохастическая составляющая временного ряда часто описывается моделью типа «черного ящика», на выходе которой создаются стохастические процессы с корреляционными свойствами, аналогичными производственным процессам, из поступающего на вход некоррелированного шума. Различие стохастических моделей ряда остатков определяется свойствами случайных последовательностей и правилами их обработки. Если независимые импульсы $e(t)$ являются случайными нормально распределенными величинами с нулевым средним и дисперсией σ_e^2 , то их последовательность $e(t)$, $e(t-1)$, ... называется «белым шумом». Белый шум $e(t)$ можно трансформировать линейным фильтром в стохастический сигнал $v(t)$ с желаемыми свойствами [6–7].

ИССЛЕДОВАНИЕ МНОЖЕСТВА СТРУКТУР МОДЕЛЕЙ УПРАВЛЯЕМОГО ОБЪЕКТА

Приступая к моделированию производственных процессов, представим модель производственной системы в виде «черного ящика», на вход которого поступает управляющий сигнал $u(t)$, а на выходе которого наблюдается контролируемый сигнал $y(t)$, зашумленный некоторым неконтролируемым сигналом помехи $v(t)$. Требуется найти зависимость выходного сигнала $y(t)$ от управляющего сигнала $u(t)$ в виде отображения $y(t) = \{u(t), v(t)\}$.

При моделировании динамики производства часто используются стационарные линейные модели, соответствующие идеализированному описанию реально протекающих производственных процессов. Если управляющее воздействие $u(t)$ и контролируемый сигнал $y(t)$ относятся к дискретным моментам времени $t_k = kT$ ($k = 1, 2, \dots$), то динамические свойства стационарной линейной незашумленной системы определяются соотношением:

$$y(t) = \sum_{k=1}^{\infty} g(k)u(t-k), \quad t = 1, 2, \dots, \tag{1}$$

где $g(k)$ – весовая функция, формирующая реакцию системы в виде выходного сигнала $y(t)$ на скалярный входной сигнал $u(t)$.

Если ввести оператор сдвига назад вида:

$$q^{-1}u(t) = u(t-1), \quad q^{-2}u(t) = u(t-2), \quad q^{-m}u(t) = u(t-m),$$

то входно-выходное уравнение системы (1) можно записать как

$$y(t) = \sum_{k=1}^{\infty} g(k)u(t-k) = G(q)u(t) \tag{2}$$

или

$$y(t) = G(q)u(t), \tag{3}$$

где $G(q)$ – передаточная функция по управлению, отражающая взаимосвязь между выходной $\{y(k)\}$ и входной $\{u(k)\}$ последовательностями вида

$$G(q) = \sum_{k=1}^{\infty} g(k)q^{-k}. \tag{4}$$

При наличии аддитивной помехи, приложенной к выходу управляемого объекта, динамические свойства линейной стационарной системы описываются соотношением:

$$y(t) = \sum_{k=1}^{\infty} g(k)u(t-k) + v(t) \quad (5)$$

или

$$y(t) = G(q)u(t) + v(t), \quad (6)$$

где $v(t)$ – сигнал помехи в момент времени t , значение которого заранее неизвестно, однако на основании прошлого поведения системы можно сделать обоснованный вывод о его будущих значениях.

Во многих случаях имеются основания считать, что сигнал помехи $v(t)$ представляет реакцию некоторого фильтра

$$v(t) = \sum_{k=0}^{\infty} h(k)e(t-k) = H(q)e(t) \quad (7)$$

или

$$v(t) = H(q)e(t), \quad (8)$$

где $h(k)$ – дискретная весовая функция $h(k)$ на входную последовательность $\{e(t)\}$ взаимно независимых одинаково распределенных случайных величин с некоторой функцией плотности вероятности; $H(q)$ – передаточная функция фильтра помехи, равная

$$H(q) = \sum_{k=0}^{\infty} h(k)q^{-k}. \quad (9)$$

Вместо традиционного фильтра помех с передаточной функцией $H(q)$ и нулевым индексом суммирования ($k=0$) на практике часто используется *монический фильтр помех*, у которого составляющая передаточной функции $h(k)$ при индексе $k=0$ равна единице, т. е. $h|_{k=0} = 1$, а суммирование составляющих передаточной функции начинается с индекса $k=1$

$$H(q) = 1 + \sum_{k=1}^{\infty} h(k)q^{-k}. \quad (10)$$

Уравнение линейной стационарной динамической системы (6) с учетом сигнала аддитивного помехи (10) примет вид [2]:

$$y(t) = G(q)u(t) + H(q)e(t), \quad (11)$$

где $\{e(t)\}$ – последовательность взаимно независимых случайных величин с нулевым средним и конечной дисперсией.

Соотношение (11) принято трактовать как множество структур моделей, из которых для заданной совокупности наблюдений необходимо выбрать наиболее предпочтительное структурное описание. Выбор конкретной структуры модели линейной динамической системы определяется характеристиками передаточных функций

$G(q)$, $H(q)$ и функции плотности распределения вероятности $f_e(\cdot)$ помехи $e(t)$.

При выборе структурного описания множества моделей будем исходить из предположения о наличии желательных свойств в передаточных функциях, основным из которых является рациональность, обеспечивающая представление каждой из передаточных функций правильной дробью. Числитель и знаменатель такой дроби должен описываться конечным набором чисел, т. е. должны отсутствовать нули числителя и полюсы знаменателя передаточных функций.

Предположим, что дробно-рациональные передаточные функции $G(q)$ и $H(q)$ описываются соотношениями:

$$G(q) = \frac{B(q)}{F(q)} \quad (12)$$

и

$$H(q) = \frac{C(q)}{D(q)}, \quad (13)$$

где $B(q)$, $C(q)$ и $F(q)$, $D(q)$ – полиномы числителя и знаменателя, равные

$$B(q) = 1 + b_1q^{-1} + \dots + b_{n_b}q^{-n_b}; \quad (14)$$

$$C(q) = 1 + c_1q^{-1} + \dots + c_{n_c}q^{-n_c}; \quad (15)$$

$$F(q) = 1 + f_1q^{-1} + \dots + a_{n_f}q^{-n_f}; \quad (16)$$

$$D(q) = 1 + d_1q^{-1} + \dots + a_{n_d}q^{-n_d}, \quad (17)$$

тогда множество структур моделей линейной стационарной динамической системы (11) для дробно-рациональных передаточных функций (12) и (13) совпадает с множеством структур моделей Бокса-Дженкинса [1]

$$y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t) \quad (18)$$

или

$$y(t) = \frac{B(q)}{F(q)}u(t) + e(t). \quad (19)$$

Частным случаем множества структур моделей Бокса-Дженкинса (18) для дробно-рациональных передаточных функций вида

$$G(q) = \frac{B(q)}{A(q)} \quad (20)$$

и

$$H(q) = \frac{C(q)}{A(q)}, \quad (21)$$

где $A(q)$ – операторный полином знаменателей $G(q)$ и $H(q)$, равный

$$A(q) = 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a} \quad (22)$$

является множеством структур моделей ошибки уравнения

$$y(t) = \frac{B(q)}{A(q)}u(t) + \frac{C(q)}{A(q)}e(t), \quad (23)$$

которое принято записывать в виде

$$A(q)y(t) = B(q)u(t) + C(q)e(t). \quad (24)$$

В качестве обобщенного множества модельных структур линейной стационарной динамической системы (23) используется уравнение

$$A(q)y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t), \quad (25)$$

которое позволяет реализовать 32 различных варианта множества структурных моделей в зависимости от того, какие из пяти многочленов $A(q)$, $B(q)$, $C(q)$, $D(q)$ и $F(q)$ включены в модель для реализации желаемых динамических свойств модели. Многообразие модельных структур (27) можно расширить введением блоков запаздывания, нелинейности, нестационарности и критериев статистического оценивания и робастности по отношению к исходным предположениям моделирования, внутренним возмущениям и проявлениям внешней среды [2].

СПЕЦИФИКАЦИЯ МОДЕЛЕЙ АДАПТИВНОГО ПРОГНОЗИРОВАНИЯ

Построение математической модели временного ряда начинается с выбора класса структур моделей-претендентов, затем осуществляется выбор предпочтительной структуры из отобранных структур и оценивание параметров модели. Задача спецификации математических моделей состоит в том, чтобы из некоторого множества моделей-претендентов выделить составляющие, допускающие самостоятельное их использование.

Существуют различные способы выбора предпочтительной структуры модели временного ряда из класса отобранных структур. Эти подходы связаны с особенностями реализации выбранного класса моделей и представления случайного процесса. При построении моделей оперативного управления производством наиболее общим классом моделей-претендентов является класс ARIMAX-моделей, компоненты которого могут самостоятельно использоваться как авторегрессионные, стационарные, нестационарные, нелинейные и робастные к различным проявлениям среды модели.

В качестве отправной точкой для спецификации моделей прогнозирования состояний системы примем ARIMAX-модель операторного уравнения $y(t) = G(q)u(t) + H(q)e(t)$ линейной разностной динамической системы с фильтром помех. Если стационарный процесс сигнала помехи $v(t) = H(q)e(t)$, где $\{e(t)\}$ – белый шум с дисперсией y , имеет рациональную передаточную функцию $H(q) = C(q)/A(q)$, где $C(q)$ и $A(q)$ – полиномы числителя и знаменателя передаточной функции $H(q)$, равные

$$C(q) = 1 + c_1q^{-1} + \dots + c_{n_c}q^{-n_c},$$

$$A(q) = 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a},$$

то уравнение помехи представляет собой ARMA(n_a, n_c)-модель вида:

$$\begin{aligned} v(t) + a_1v(t-1) + \dots + a_{n_a}v(t-n_a) = \\ = e(t) + c_1e(t-1) + \dots + c_{n_c}e(t-n_c). \end{aligned}$$

1. Авторегрессионная модель n_a -го порядка AR(n_a)-модель имеет вид:

$$v(t) + a_1v(t-1) + \dots + a_{n_a}v(t-n_a) = \mu + e(t), \quad (26)$$

где $v(t), v(t-1), \dots, v(t-n_a)$ – отсчеты, представляющие собой текущее $v(t)$ и прошлые $v(t-1), \dots, v(t-n_a)$ значения случайного процесса; n_a – количество отсчетов, определяющее число переменных и искомых коэффициентов модели (порядок или структуру модели); μ – константа, отражающая начальные условия; a_1, \dots, a_{n_a} – коэффициенты модели.

Авторегрессионная AR(n_a)-модель представляет собой частный случай ARMA(n_a, n_c)-модели при $n_c = 0$. Если $e(t)$ является белым шумом, то AR(n_a)-модель описывает авторегрессионный процесс n_a -го порядка. Текущее значение авторегрессионного AR(n_a)-процесса выражается как смещенная на μ совокупность предыдущих значений процесса $v(t)$ и импульса $e(t)$. AR(n_a)-модель содержит $n_a + 2$ искомого параметра $\mu, a_1, a_2, \dots, a_{n_a}, \sigma_e^2$, которые вычисляются по результатам наблюдений, где σ_e^2 – дисперсия белого шума $e(t)$.

2. Модель скользящего среднего n_c -го порядка MA(n_c)-модель имеет вид:

$$v(t) = \mu + e(t) + c_1e(t-1) + \dots + c_{n_c}e(t-n_c), \quad (27)$$

из которой следует, что текущее значение процесса $v(t)$ является комбинацией константы μ , текущего $e(t)$ и всех прошлых $e(t-1), \dots, e(t-n_c)$ значений случайной величины $e(t)$, являющейся по предположению белым шумом.

Модель скользящего среднего MA(n_c) представляет собой второй частный случай ARMA(n_a, n_c)-модели при $n_a = 0$. Если $e(t)$ является белым шумом, то MA(n_c)-модель описывает процесс скользящего среднего MA(n_c)-процесс n_c -го порядка. Для построения MA(n_c)-модели необходимо определить $n_c + 2$ неизвестных коэффициента: $\mu, c_1, c_2, \dots, c_{n_c}, \sigma_e^2$. Если в выражении (27) $\mu = 0$, то имеет место чистый MA(n_c)-процесс скользящего среднего n_c -го порядка.

Модели скользящего среднего предпочтительнее использовать в сочетании с авторегрессионными процессами. Это позволяет сосредоточить внимание на самых последних наблюдениях (в отличие от процессов чистой авторегрессии). Прогнозирование следующего наблюдения с помощью скользящего среднего основывается на оценке текущего случайного шума $e(t)$. За пределами следующего наблюдения наилучшим прогнозом является оценка долгосрочного среднего μ , поскольку процесс скользящего среднего забывает все свое прошлое, за исключением последнего наблюдения.

3. Авторегрессионная модель со скользящим средним (n_a, n_c)-го порядка ARMA(n_a, n_c)-модель имеет вид:

$$\begin{aligned} v(t) + a_1 v(t-1) + \dots + a_{n_a} v(t-n_a) = \\ = \mu + e(t) + c_1 e(t-1) + \dots + c_{n_c} e(t-n_c). \end{aligned} \quad (28)$$

которая сочетает в себе процессы авторегрессии и скользящего среднего. Память ARMA(n_a, n_c)-процесса содержит в себе память процесса авторегрессии и память процесса скользящего среднего. В результате получается процесс авторегрессии с улучшенной краткосрочной памятью.

Количество параметров ARMA(n_a, n_c)-модели равно $n_a + n_c + 2$, из них n_a коэффициентов авторегрессии, n_c коэффициентов скользящего среднего и два параметра шумовой последовательности μ и σ_e^2 . Варьированием значений коэффициентов a_1, \dots, a_{n_a} и c_1, \dots, c_{n_c} можно выбрать модель, которая достаточно хорошо опишет любой стационарный производственный процесс набором данных временного ряда.

Многие реальные процессы имеют нестационарный характер, проявляющийся в том, что средние значения и дисперсии стохастических рядов изменяются во времени. Несмотря на то, что средние, относительно которых происходят флуктуации процесса, для определенных интервалов времени могут быть разными, общее поведение рядов с учетом различий относительно средних может оказаться сходным. Для математического описания нестационарных процессов используются авторегрессионные интегрированные модели скользящего среднего ARIMA(n_a, d, n_c), где d – порядок разностного оператора, превращающего нестационарный процесс в стационарный.

4. Авторегрессионная интегрированная модель скользящего среднего (n_a, d, n_c)-го порядка ARIMA (n_a, d, n_c)-модель имеет вид:

$$\begin{aligned} v(t) - v(t-1) = a_1 v(t-1) + \dots + a_{n_a} v(t-n_a) + \\ + \phi[v(t-1) - v(t-2)] + e(t) + c_1 e(t-1) + \dots + c_{n_c} e(t-n_c), \end{aligned} \quad (29)$$

которая получается путем суммирования модели авторегрессии n_a -го порядка, модели скользящего среднего n_c -го порядка и модели принудительного интегрирования временного ряда, обладающего d -м порядком гладкости.

В ARIMA-процессе содержится информация о том, где он находится, как он попал в это состояние, а также о части предыдущего шумового компонента. ARIMA-процесс может служить моделью временного ряда, который является очень гладким и медленно изменяет свое направление. ARIMA-процесс является нестационарным, поскольку его состояния получается суммированием состояний ARMA-процесса. С течением времени такой ряд удаляется все дальше от своего исходного состояния. ARIMA-модели находят свое широкое применение для математического описания поведения нестационарных

временных рядов различных производственных процессов.

5. Авторегрессионная модель выходного сигнала $y(t)$ (n_a, n_b)-го порядка ARX(n_a, n_b)-модель, расширенная управляющим $u(t)$ и возмущающим $e(t)$ воздействиями

$$\begin{aligned} y(t) + a_1 y(t-1) + \dots + a_{n_a} y(t-n_a) = \\ = b_1 u(t-1) + \dots + c_{n_b} u(t-n_b) + e(t), \end{aligned} \quad (30)$$

в которую входит белый шум $e(t)$ как непосредственная ошибка, поэтому ARX(n_a, n_b)-модель часто называют моделью ошибки уравнения. Если $n_a = 0$, то выходной сигнал $y(t)$ описывается моделью с конечной памятью. Такие модели широко применяются при обработке различного рода сигналов. Для построения ARX(n_a, n_b)-модели необходимо определить $n_a + n_b$ настраиваемых параметров.

6. Авторегрессионная ARMAX(n_a, n_b, n_c)-модель выходного сигнала $y(t)$ со скользящим средним помехи $e(t)$ (n_a, n_b, n_c)-го порядка

$$\begin{aligned} y(t) + a_1 y(t-1) + \dots + a_{n_a} y(t-n_a) = \\ = b_1 u(t-1) + \dots + c_{n_b} u(t-n_b) + \\ + b_1 u(t-1) + \dots + c_{n_b} u(t-n_b) + \\ + e(t) + c_1 e(t-1) + \dots + c_{n_c} e(t-n_c) \end{aligned} \quad (31)$$

включает в себя авторегрессионную составляющую по выходному сигналу $y(t)$, инкрементную (экзогенную) составляющую по управляющему воздействию $u(t)$ и шумовую составляющую в виде скользящего среднего помехи $e(t)$. ARMAX-модели стали неутвержденным стандартом построения моделей производственных и экономических процессов. При построении ARMAX-модели необходимо оценить $n_a + n_b + n_c$ настраиваемых параметров модели.

7. Авторегрессионная ARIMAX(n_a, d, n_b, n_c)-модель выходного сигнала $y(t)$ с интегрированным скользящим средним помехи $e(t)$ (n_a, d, n_b, n_c)-го порядка, обладающая d -м порядком гладкости исходных данных.

ПРОГНОЗИРОВАНИЕ СОСТОЯНИЯ ARIMA-ПРОЦЕССОВ

С помощью ARIMA-процессов, наблюдаемых на выходе линейных статистических моделей, можно достаточно точно описывать поведение временных рядов самых различных производственных процессов. ARIMA-процессы образуются из совокупности трех составляющих: процессов авторегрессии, процессов интегрирования и процессов скользящего среднего. Каждая составляющая ARIMA-процесса видоизменяет структуру и параметры ARIMA-модели.

Построение ARIMA-модели предусматривает выбор конкретного типа модели и оценивание параметров модели на основании исходных данных. ARIMA-модель описывает поведение временного ряда минимальным количеством параметров и позволяет определить, в какой

мере каждое наблюдение влияет на будущее и сколько новой полезной информации содержится в каждом наблюдении. Прогнозируемое значение процесса может содержать тренд и сезонный компонент.

Алгоритм адаптивного прогнозирования состояния динамических процессов содержит следующие этапы [8–10]:

1. Предварительный анализ и сглаживание уровней временных рядов производственных процессов.

1.1. Выявление и устранение аномальных значений уровней временного ряда с использованием критерия Ирвина.

1.2. Определение наличия во временном ряду тренда с использованием метода Фостера-Стьюарта и определение степени его гладкости.

1.3. Выявление наличия во временной ряду сезонных колебаний путем проверки на случайность остаточного ряда.

1.4. Фильтрация компонент временного ряда путем выделения из исходного ряда трендовой компоненты, сезонной компоненты, циклической компоненты и случайной компоненты с использованием метода Четверикова.

2. Модели прогнозирования производственных процессов.

2.1. Трендовые модели на основе кривых роста.

2.2. Оценка адекватности и точности трендовых моделей.

2.3. Прогнозирование динамики производства на основе трендовых моделей.

3. Адаптивные модели прогнозирования.

3.1. Модели экспоненциального сглаживания Брауна.

3.2. Модели авторегрессии типа $AR(p)$.

3.3. Модели скользящего среднего типа $MA(q)$.

3.4. Модели авторегрессии и скользящего среднего типа $ARMA(p, q)$.

3.5. Модели авторегрессии и проинтегрированного скользящего среднего $ARIMA(p, q)$.

4. Алгоритм построения $ARMA$ -модели.

4.1. Проверка временного ряда на стационарность.

4.2. Построение автокорреляционных функций временных рядов.

4.3. Выбор нескольких $ARMA$ -спецификаций, исходя из анализа автокорреляционных зависимостей, и определение наилучший из них.

4.4. После выбора наилучшей $ARMA$ -модели применить ее для прогнозирования.

АКТУАЛЬНОСТЬ И НОВИЗНА ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Выполненные исследования показали актуальность и важность проблемы адаптивного прогнозирования в задачах оперативного управления производством. Проанализированы основные типы структур моделей прогнозирования, к которым относятся авторегрессионные ARX -модели с управляющими воздействиями, авторе-

грессионные $ARMAX$ -модели выходного сигнала со скользящим средним помехи, авторегрессионные $ARI-MAX$ -модели выходного сигнала с интегрированным скользящим средним помехи, робастные авторегрессионные $RARIMAX$ -модели, т. е. $ARIMAX$ -модели, устойчивые к отклонениям исходных предположений.

Разработанные адаптивные методы прогнозирования состояний производственных систем весьма удобны для оперативного управления производством в режиме реального времени. По описанному алгоритму разработана программа синтеза модельной зависимости диаметра кристалла от температуры расплава в зоне кристаллизации и скорости вытягивания.

ПЕРЕЧЕНЬ ССЫЛОК

1. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. – М.: Мир, 1974, вып. 1, 406 с., вып. 2, 197 с.
2. Льюнг Л. Идентификация систем. Теория для пользователя. – М.: Наука, 1991. – 432 с.
3. Вероятностно-статистические методы обработки данных в информационных системах / Ю. В. Бородакий и др. – М.: Радио и связь, 2003. – 264 с.
4. Иванов В. В. Анализ временных рядов и прогнозирование экономических показателей. – Харьков: ХНУ им. В. И. Каразина, 1999. – 230 с.
5. Ханк Дж. Э., Уичерн Д. У., Райтс А. Дж. Бизнес-прогнозирование. – М.: Вильямс, 2003. – 656 с.
6. Лук'яненко І. Г., Городніченко Ю. О. Сучасні економетричні методи у фінансах. – К.: Літера, 2002. – 352 с.
7. Тихомиров Н. П., Дорохина Е. Ю. Эконометрика. – М.: Экзамен, 2003. – 512 с.
8. Єріна А. М. Статистичне моделювання та прогнозування. – К.: КНЕУ, 2001. – 170 с.
9. Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов. – М.: Финансы и статистика, 2003. – 416 с.
10. Авраменко В. П., Кротюк И. Г., Петренко В. Н. Спецификация прогнозирующей модели для системы оперативного управления производством / Нови технології. Науковий вісник Інституту економіки і нових технологій ім. Ю. І. Кравченка, вип. 2 (3), 2003. – С. 11–17.

Надійшла 10.05.2004
Після доробки 19.11.2004

Досліджуються адаптивні методи прогнозування станів виробничих систем, процеси в яких описуються динамічними рядами. Стан системи визначається скінченнорізними рівняннями, подібними до диференціальних рівнянь Лагранжа першого роду. Виконано специфікацію прогнозуючої моделі станів на моделях типу $ARMA$ та $ARIMAX$. На основі стохастичного вектора станів розроблена узагальнена векторно-матрична модель і запропоновано рекурентний метод прогнозування калманівського типу. Запропонований метод прогнозування станів впроваджено в системи підтримки прийняття рішень для оперативного управління виробничими процесами.

Adaptive methods for operational systems status forecasting, where processes are described by dynamic series, are investigated in the article. System status is defined by finite-difference equations like Lagrange first-order differential equations. System status predictive model specification is made on $ARMA$ and $ARMAX$ models. Generalized vector-matrix model, which is based on stochastic state vector, is worked out by the authors and Kalman's type recurrent forecasting method is suggested in the article. Proposed status forecasting method is introduced in the expert support systems for manufacturing process on-line control.

Е. В. Бодянский, А. Н. Слипченко

АНСАМБЛЬ НЕЙРОПРЕДИКТОРОВ С ПЕРЕМЕННЫМ ЧИСЛОМ УЗЛОВ

В статье рассматривается задача прогнозирования нелинейных нестационарных временных рядов с помощью ансамбля нейропредикторов. Предлагается алгоритм, позволяющий в реальном времени изменять число нейропредикторов, входящих в ансамбль.

1 ВВЕДЕНИЕ

Задача прогнозирования временных рядов произвольной природы представляет несомненный практический интерес в связи с анализом временных рядов в биологии, технике, экономике и других отраслях. Традиционно для получения прогноза временного ряда используется математическая конструкция, называемая предиктором, которая определенным образом комбинирует имеющуюся информацию о предыдущих значениях исследуемой последовательности и ошибках прогнозирования. В зависимости от вида этой комбинации различают линейные и нелинейные предикторы. Линейные модели предикторов (AR, MA, ARMA) получили широкое распространение [1, 2], однако в последнее время исследователи все чаще сталкиваются с объектами, требующими применения нелинейных подходов [1, 3]. Нелинейные модели предикторов (NAR, NMA, NARMA), являющиеся обобщением соответствующих линейных.

В случае, когда характеристики исследуемого временного ряда изменяются с течением времени, параметры предиктора тоже претерпевают изменения. Для прогнозирования таких последовательностей необходимо производить настройку параметров модели по мере поступления новых данных в реальном времени. Таким образом, предиктор вместе с рекуррентной процедурой оценки параметров является адаптивным и позволяет прогнозировать нестационарные временные ряды.

Одним из важнейших этапов построения предиктора является его структурная идентификация [1, 2]. Для линейных временных рядов существует достаточно строгий подход, основанный на анализе корреляционных функции прогнозируемого ряда. В нелинейном случае таких методов на данный момент не существует. Одним из путей решения этой проблемы является построение нескольких предикторов, различающихся структурой, а затем выбор наилучшего из них на основе некоторого критерия. Это решение, однако, не является в общем случае оптимальным. Например, если структура ряда меняется с течением времени, предиктор, оказавшийся лучшим в один момент времени, может уже не быть таковым в другой момент.

В настоящее время для решения широкого класса задач, связанных с обработкой информации, анализа данных, идентификации, управления и т. п. в условиях априорной и текущей структурной и параметрической

неопределенности широко используются искусственные нейронные сети (ANN) [1, 3]. В [4] было предложено комбинировать выходы нескольких нейронных сетей с целью получения более точного результата, при этом сети имеют одинаковую архитектуру и обучаются по одному и тому же алгоритму, но с разными начальными условиями. Идея комбинирования нескольких прогнозирующих нейронных сетей с целью повышения качества предсказания возникла достаточно давно. В некоторых работах этот подход получил название смесью экспертов [5, 6], в других – оптимальной линейной комбинацией нейронных сетей [4, 7, 8], в третьих – ансамбля нейронных сетей или прогнозирующих моделей [9–12]. Обычно обобщенный прогноз строится при помощи аддитивной свертки, при этом каждому из локальных предикторов присваивается, как правило, фиксированный вес, задающий «вклад» каждой из локальных прогнозирующих моделей в выходной сигнал ансамбля.

Одним из ключевых свойств, обусловивших интерес к нейронным сетям со стороны инженеров, является их возможность изменять свое поведение, адаптируясь к изменяющимся характеристикам моделируемого объекта. При этом, если раньше под адаптивностью сети, как правило, понимали возможность подстройки ее параметров (синаптических весов), то сейчас адаптивность трактуется в более широком смысле – как возможность настройки еще и архитектуры (количества слоев и узлов) [1, 3, 13].

Следует отметить, что в цитированных выше работах ничего не говорится об оптимальности получаемых прогнозов. Кроме того, предложенные алгоритмы не предназначены для работы в реальном времени. В условиях, когда обрабатываемые сигналы изменяют свои свойства, на первый план выходят требования к возможности последовательной обработки поступающей информации и оперативной перестройки весовых коэффициентов и состава ансамбля.

Задачей выбора оптимальной архитектуры сети исследователи интересуются уже довольно давно. На данный момент в этой области имеется устоявшаяся классификация алгоритмов в зависимости от того, на что они ориентированы: на уменьшение или увеличение сложности используемой нейросети [3].

Алгоритмы, которые начинают свою работу с простой архитектуры, а затем усложняют ее путем добавления новых узлов по мере обучения, получили название конструктивных (constructive). В противоположность им деструктивные алгоритмы (pruning, destructive) начинают работу с изначально избыточной сетью, упрощая ее по мере обучения.

Один из первых конструктивных алгоритмов был предложен Платтом для радиально-базисных нейронных

сетей (РБНС) [14] и получил название “resource allocation” [15]. В настоящее время известен также ряд модификаций этой процедуры [16, 17]. Большое распространение получила также каскадно-корреляционная архитектура (CasCor), предложенная С. Фальманом и К. Лебьером [18].

Среди деструктивных алгоритмов наиболее популярными являются “Optimal brain damage” [19] и “Optimal brain surgeon” [20]. В этих методах значимость узла или связи между узлами оценивается по тому, какое изменение в значении целевой функции повлечет за собой его удаление. Для этого анализируется матрица вторых частных производных оптимизируемой функции по настраиваемым параметрам. Обе процедуры обладают довольно значительной вычислительной сложностью. Кроме этого, существенным недостатком является то, что после удаления несущественных узлов необходимо переобучить сеть, что в свою очередь, делает невозможным применение этих алгоритмов для работы в реальном времени. Прочие алгоритмы [21], как правило, являются эвристическими и им недостает универсальности.

В настоящей работе предлагается использовать множество параллельно работающих предикторов, различающихся структурой и объемом используемой предыстории, а затем строить на их основе обобщенный прогноз в виде линейной комбинации локальных прогнозов. Такая совокупность предикторов называется ансамблем. Если производить оценивание параметров предикторов в реальном времени в сочетании с возможностью забывания устаревшей информации и изменения состава ансамбля, то получим конструкцию пригодную для прогнозирования нестационарных временных рядов произвольной природы в условиях априорной и текущей структурной неопределенности. Предлагается архитектура и алгоритм обучения прогнозирующей нейронной метасети, содержащей в своем скрытом слое ансамбль нейропредикторов. Выходной слой метасети представляет собой адаптивный линейный ассоциатор, настройка синаптических весов которого производится в реальном времени. Кроме того, вводится алгоритм структурной адаптации этой метасети, способный работать в реальном времени, т. е. позволяющий производить последовательную обработку информации и не требующий переобучения сети после изменения ее структуры.

2 ПОСТАНОВКА ЗАДАЧИ

Вначале рассмотрим задачу построения обобщенного прогноза на содержательном уровне. Пусть имеется множество нейропредикторов (локальных предикторов). Различие между ними может заключаться в архитектуре, объеме используемой предыстории, алгоритме обучения, начальных условиях и т. п. Все эти предикторы настраиваются с использованием одной и той же обучающей выборки. Необходимо построить обобщенный прогноз в виде взвешенной суммы имеющихся в нашем распоряжении локальных предикторов.

С точки зрения теории нейронных сетей такая конструкция эквивалентна большей метасети, в которой сети, представляющие локальные предикторы работают параллельно. Коэффициенты, с которыми учитываются выходы отдельных ИНС в обобщенном прогнозе могут интерпретироваться как синаптические веса выходного слоя метасети.

Формализация изложенной выше задачи выглядит следующим образом. Пусть имеется различных прогнозирующих моделей (предикторов). Необходимо построить на их основе обобщенный прогноз, который представляет собой линейную комбинацию имеющихся в нашем распоряжении локальных прогнозов:

$$\tilde{y}(k+1) = \sum_{i=1}^J \mu_i(k) \hat{y}_i(k+1) = \mathbf{M}^T(k) \hat{Y}(k+1), \quad (1)$$

где $\mathbf{M}(k) = (\mu_1(k), \dots, \mu_J(k))^T$, $\hat{Y}(k) = (\hat{y}_1(k+1), \dots, \hat{y}_J(k+1))^T$ – $(J \times 1)$ векторы синаптических весов выходного слоя метасети и локальных прогнозов, получаемых с помощью отдельных предикторов соответственно, $k = 0, 1, 2, \dots$ – дискретное время, J – число локальных прогнозирующих моделей.

Логично включить в обобщенный прогноз наряду с локальными прогнозами параметр смещения, что позволит ансамблю работать даже в условиях, когда один или несколько предикторов, входящих в его состав реализуют смещенный прогноз, т. е.

$$\begin{aligned} \tilde{y}(k+1) &= \mu_0(k) + \sum_{i=1}^J \mu_i(k) \hat{y}_i(k+1) = \\ &= \sum_{i=0}^J \mu_i(k) \hat{y}_i(k+1) = \mathbf{M}^T(k) \hat{Y}(k+1), \end{aligned}$$

где $\hat{y}_0(k) \equiv 1$, $\mathbf{M}(k) = (\mu_0(k), \mu_1(k), \dots, \mu_J(k))^T$, $\hat{Y}(k) = (1, \hat{y}_1(k+1), \dots, \hat{y}_J(k+1))^T$.

Таким образом, задача построения обобщенного прогноза сводится к задаче нахождения вектора весов \mathbf{M} , оптимального в некотором смысле. Естественным является требование несмещенности получаемого прогноза, определяемое ограничением

$$\sum_{i=0}^J \mu_i(k) = 1,$$

или в векторной форме

$$\mathbf{M}^T(k) \mathbf{E} = 1,$$

где $\mathbf{E} = (1, 1, \dots, 1)^T$ – $((J+1) \times 1)$ вектор, составленный из единиц.

В нестационарных условиях оценку вектора весов логично строить не на всей выборке данных, а только на некотором скользящем окне. Это позволит локальным ИНС, образующим ансамбль параллельно работающих предикторов, адаптироваться к изменяющимся характеристикам процесса.

3 АЛГОРИТМ ОБУЧЕНИЯ АНСАМБЛЯ НЕЙРОПРЕДИКТОРОВ

Задача построения обобщенного прогноза на основе ансамбля предикторов в момент времени k , сводится к задаче математического программирования с ограничениями в виде равенств:

$$\left\{ \begin{aligned} I(\mathbf{M}(k)) &= \sum_{j=0}^{N-1} (y(k-j) - \tilde{y}(k-j))^2 = \\ &= \sum_{j=0}^{N-1} (y(k-j) - \mathbf{M}^T(k) \hat{\mathbf{Y}}(k-j))^2 \rightarrow \min_{\mathbf{M}(k) \in \Omega}, \quad (2) \\ \Omega: \mathbf{M}^T(k) \mathbf{E} &= 1. \end{aligned} \right.$$

где $\mathbf{M}(k) = (\mu_1(k), \dots, \mu_J(k))^T$, $\hat{\mathbf{Y}}(k) = (\hat{y}_1(k), \dots, \hat{y}_J(k))^T$, $\mathbf{E} = (1, 1, \dots, 1)^T - (J \times 1)$ вектор, составленный из единиц, N – величина скользящего окна, на котором строится оценка. В случае наличия параметра смещения задача построения ансамбля имеет тот же вид, но при этом $\mathbf{M}(k) = (\mu_0(k), \mu_1(k), \dots, \mu_J(k))^T$, $\hat{\mathbf{Y}}(k) = (1, \hat{y}_1(k), \dots, \hat{y}_J(k))^T$, $\mathbf{E} = (1, 1, \dots, 1)^T - ((J+1) \times 1)$ вектор.

Для решения задачи (2) может быть применен метод штрафных функций, что позволит нам перейти к решению задачи безусловной оптимизации:

$$\begin{aligned} I(\mathbf{M}, \rho) &= \sum_{j=0}^{N-1} (y(k-j) - \tilde{y}(k-j))^2 - \rho^2 \left(1 - \sum_{i=0}^J \mu_i(k)\right)^2 = \\ &= \sum_{j=0}^{N-1} (y(k-j) - \mathbf{M}^T \hat{\mathbf{Y}}(k-j))^2 - \rho^2 (1 - \mathbf{M}^T(k) \mathbf{E})^2 \rightarrow \min_{\mathbf{M}(k)}, \quad (3) \end{aligned}$$

где ρ – параметр штрафа. Решением этой задачи является следующее выражение для вектора весовых коэффициентов:

$$\begin{aligned} \mathbf{M}(k, \rho) &= \left(\sum_{j=0}^{N-1} \hat{\mathbf{Y}}(k-j) \hat{\mathbf{Y}}^T(k-j) + \rho^2 \mathbf{E} \mathbf{E}^T \right)^{-1} \times \\ &\times \left(\sum_{j=0}^{N-1} y(k-j) \hat{\mathbf{Y}}(k-j) + \rho^2 \mathbf{E} \right). \quad (4) \end{aligned}$$

Устремим далее ρ к ∞ и окончательно получим:

$$\begin{aligned} \mathbf{M}(k) &= \lim_{\rho \rightarrow \infty} \mathbf{M}(k, \rho) = \lim_{\rho \rightarrow 0} \left(R(k) + \rho^2 \mathbf{E} \mathbf{E}^T \right)^{-1} (F(k) + \rho^2 \mathbf{E}) = \\ &= \mathbf{M}^*(k) + R^{-1}(k) \frac{1 - E^T \mathbf{M}^*(k)}{E^T R^{-1}(k) E} \mathbf{E}, \quad (5) \end{aligned}$$

где $R(k) = \sum_{j=0}^{N-1} \hat{\mathbf{Y}}(k-j) \hat{\mathbf{Y}}^T(k-j)$, $F(k) = \sum_{j=0}^{M-1} y(k) \hat{\mathbf{Y}}(k-j)$,

$\mathbf{M}^*(k) = R^{-1}(k) F(k)$ – оценка, полученная по методу наименьших квадратов на скользящем окне. Для работы в нестационарных условиях целесообразно использовать

текущий рекуррентный метод наименьших квадратов [22], модифицированная версия которого имеет вид:

$$\left\{ \begin{aligned} \tilde{R}^{-1}(k) &= \left(\sum_{j=0}^N \hat{\mathbf{Y}}(k-j) \hat{\mathbf{Y}}^T(k-j) \right)^{-1} = \\ &= R^{-1}(k-1) - \frac{R^{-1}(k-1) \hat{\mathbf{Y}}(k) \hat{\mathbf{Y}}^T(k) R^{-1}(k-1)}{1 + \hat{\mathbf{Y}}^T(k) R^{-1}(k-1) \hat{\mathbf{Y}}(k)}, \\ R^{-1}(k) &= \left(\sum_{j=0}^{N-1} \hat{\mathbf{Y}}(k-j) \hat{\mathbf{Y}}^T(k-j) \right)^{-1} = \\ &= \tilde{R}^{-1}(k) + \frac{\tilde{R}^{-1}(k) \hat{\mathbf{Y}}(k-N) \hat{\mathbf{Y}}^T(k-N) \tilde{R}^{-1}(k)}{1 - \hat{\mathbf{Y}}^T(k-N) \tilde{R}^{-1}(k) \hat{\mathbf{Y}}(k-N)}, \quad (6) \\ \tilde{\mathbf{M}}(k) &= \mathbf{M}^*(k-1) + \frac{R^{-1}(k-1) (y(k) - \hat{\mathbf{Y}}^T(k-1) \hat{\mathbf{Y}}(k))}{1 + \hat{\mathbf{Y}}^T(k) R^{-1}(k-1) \hat{\mathbf{Y}}(k)} \hat{\mathbf{Y}}(k), \\ \mathbf{M}^*(k) &= \tilde{\mathbf{M}}(k) - \frac{\tilde{R}^{-1}(k) (y(k-N) - \tilde{\mathbf{M}}^T(k) \hat{\mathbf{Y}}(k-N))}{1 - \hat{\mathbf{Y}}^T(k-N) \tilde{R}^{-1}(k) \hat{\mathbf{Y}}(k-N)} \times \\ &\times \hat{\mathbf{Y}}(k-N). \end{aligned} \right.$$

Первые два выражения в процедуре (6) реализуют накопление новых и сброс старых данных соответственно (напомним, что алгоритм работает на скользящем окне, состоящем из N последних наблюдений). $\tilde{\mathbf{M}}(k)$ – оценка по методу наименьших квадратов на растущем окне (накопление данных), а $\mathbf{M}^*(k)$ – оценка, построенная на N последних наблюдениях, полученная из $\tilde{\mathbf{M}}(k)$ путем исключения устаревших данных.

Объединяя (6) и (5), приходим к процедуре настройки вектора весовых коэффициентов метасети:

$$\left\{ \begin{aligned} \tilde{R}^{-1}(k) &= \left(\sum_{j=0}^N \hat{\mathbf{Y}}(k-j) \hat{\mathbf{Y}}^T(k-j) \right)^{-1} = \\ &= R^{-1}(k-1) - \frac{R^{-1}(k-1) \hat{\mathbf{Y}}(k) \hat{\mathbf{Y}}^T(k) R^{-1}(k-1)}{1 + \hat{\mathbf{Y}}^T(k) R^{-1}(k-1) \hat{\mathbf{Y}}(k)}, \\ R^{-1}(k) &= \left(\sum_{j=0}^{N-1} \hat{\mathbf{Y}}(k-j) \hat{\mathbf{Y}}^T(k-j) \right)^{-1} = \\ &= \tilde{R}^{-1}(k) + \frac{\tilde{R}^{-1}(k) \hat{\mathbf{Y}}(k-N) \hat{\mathbf{Y}}^T(k-N) \tilde{R}^{-1}(k)}{1 - \hat{\mathbf{Y}}^T(k-N) \tilde{R}^{-1}(k) \hat{\mathbf{Y}}(k-N)}, \quad (7) \\ \tilde{\mathbf{M}}^*(k) &= \mathbf{M}^*(k-1) + \frac{R^{-1}(k-1) (y(k) - \hat{\mathbf{Y}}^T(k) \mathbf{M}^*(k-1))}{1 + \hat{\mathbf{Y}}^T(k) R^{-1}(k-1) \hat{\mathbf{Y}}(k)} \hat{\mathbf{Y}}(k), \\ \mathbf{M}^*(k) &= \tilde{\mathbf{M}}^*(k) - \frac{\tilde{R}^{-1}(k) (y(k-N) - \hat{\mathbf{Y}}^T(k-N) \tilde{\mathbf{M}}^*(k))}{1 - \hat{\mathbf{Y}}^T(k-N) \tilde{R}^{-1}(k) \hat{\mathbf{Y}}(k-N)} \times \\ &\times \hat{\mathbf{Y}}(k-N), \\ \mathbf{M}(k) &= \mathbf{M}^*(k) + R^{-1}(k) \frac{1 - E^T \mathbf{M}^*(k)}{E^T R^{-1}(k) E} \mathbf{E}. \end{aligned} \right.$$

Можно показать, что данная процедура является обобщением оптимального алгоритма адаптивной идентификации, предложенного в [9] и доказатель, что полученный таким образом обобщенный прогноз по точности не уступает ни одному из локальных прогнозов, входящих в его состав.

Использование метода наименьших квадратов (МНК) и его модификаций для оценки синаптических весов выходного слоя метасети позволяет получить строгую и хорошо интерпретируемую оценку значимости каждого локального предиктора, входящего в обобщенный прогноз (1). Очевидно, что если какой-либо элемент вектора весов $M^*(k)$ мал по своей абсолютной величине, то соответствующий предиктор может быть выведен из состава ансамбля без большого ущерба для точности. При этом, однако, оставшиеся синаптические веса не будут нуждаться в перенастройке только в том случае, когда удаляемому узлу (предиктору) соответствует вес в точности равный нулю. Во всех остальных ситуациях потребуется перенастройка параметров сети.

В [23] был предложен алгоритм, основанный на использовании формулы Фробениуса [24] и позволяющий изменять структуру искусственной нейронной сети, обучаемой с использованием стандартного МНК. Одним из основных его преимуществ является возможность обработки данных в реальном масштабе времени. Подобный подход может быть применен и в нашем случае.

Предположим, что в k -й момент времени с помощью (7) была получена оценка синаптических весов выходного слоя метасети:

$$M^*(k) = R_J^{-1}(k)F_J(k),$$

$$M(k) = M^*(k) + R_J^{-1}(k) \frac{1 - E^T M^*(k)}{E^T R_J^{-1}(k) E} E, \quad (8)$$

где индекс J определяет число локальных предикторов в ансамбле.

Как уже отмечалось ранее, абсолютная величина компонентов вектора $M^*(k)$ может быть интерпретирована как значимость соответствующего локального предиктора. Допустим, что параметр $\mu_J(k)$ мал по своей абсолютной величине, и мы хотим вывести соответствующий ему предиктор из ансамбля. Предположение о том, что несущественным является компонент с номером J , не является ограничительным, так как мы всегда можем перенумеровать предикторы таким образом, чтобы наименьшие по абсолютной величине компоненты вектора $M^*(k)$ имели наибольший номер. Такая операция приведет лишь к перестановке строк и столбцов в матрице $R_J(k)$, а также к изменению порядка следования элементов вектора $F_J(k)$. При этом, как известно, перестановка столбцов и/или строк матрицы является преобразованием, не влияющим на последующие матричные операции.

Учитывая тот факт, что матрица $R_J(k)$ – симметрическая, получаем:

$$M_J^*(k) = R_J^{-1}(k)F_J(k) = \begin{pmatrix} R_{J-1}(k) & \beta_{J-1}(k) \\ \beta_{J-1}^T(k) & r_{JJ}(k) \end{pmatrix}^{-1} \begin{pmatrix} F_{J-1}(k) \\ f_J(k) \end{pmatrix}, \quad (9)$$

где $r_{ij}(k)$ – элемент i -й строки j -го столбца матрицы $R_J(k)$, $\beta_{J-1}(k) = (r_{1J}(k), \dots, r_{J-1J}(k))^T = (r_{J1}(k), \dots, r_{JJ-1}(k))^T$.

После несложных преобразований выражения (9) получаем:

$$M_J^*(k) = \begin{pmatrix} M_{J-1}^*(k) - R_{J-1}^{-1}(k)\beta_{J-1}(k)\mu_J(k) \\ \mu_J(k) \end{pmatrix}, \quad (10)$$

что дает нам возможность вывести нейропредиктор с номером J из состава ансамбля (1) и получить уточненные оценки оставшихся параметров сети. При этом для выполнения этой операции используется только информация, накопленную в матрице $R_J(k)$ и векторе $F_J(k)$.

Используя ту же технику, что и выше, можно получить процедуру, позволяющую добавить новый предиктор в ансамбль. Непосредственное применение формулы Фробениуса к выражению типа (9) приводит к следующему алгоритму:

$$M_{J+1}^*(k) = R_{J+1}^{-1}(k)F_{J+1}(k) = \begin{pmatrix} R_J(k) & \beta_J(k) \\ \beta_J^T(k) & r_{J+1J+1}(k) \end{pmatrix}^{-1} \begin{pmatrix} F_J(k) \\ f_{J+1}(k) \end{pmatrix} =$$

$$= \begin{pmatrix} M_J^*(k) + R_J^{-1}(k)\beta_J(k) \frac{\beta_J^T(k)M_J^*(k) - f_{J+1}(k)}{r_{J+1J+1}(k) - \beta_J^T(k)R_J^{-1}(k)\beta_J(k)} \\ \frac{-\beta_J^T(k)M_J^*(k) + f_{J+1}(k)}{r_{J+1J+1}(k) - \beta_J^T(k)R_J^{-1}(k)\beta_J(k)} \end{pmatrix}, \quad (11)$$

где $\beta_J(k) = (r_{1J+1}(k), \dots, r_{JJ+1}(k))^T = (r_{J+11}(k), \dots, r_{J+1J}(k))^T$.

Таким образом, с помощью формулы (11) мы можем добавить в модель (1) новый локальный предиктор, а с помощью формулы (10) – вывести из рассмотрения один из имеющихся. Для того, чтобы эти операции могли проводиться в реальном времени, необходимо накапливать информацию о большем числе предикторов, чем непосредственно используется в данный момент. Так, например, мы можем изначально ввести в рассмотрение избыточное число предикторов H и накапливать информацию в матрице $R_H(k)$ и векторе $F_H(k)$ по мере поступления новых данных. При этом для получения прогноза в текущий момент времени используется только $J < H$ из них. В случае необходимости возможно как уменьшение сложности используемой прогнозирующей модели, так и ее увеличение.

4 РЕЗУЛЬТАТЫ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

Для апробации работы предложенного алгоритма рассмотрим задачу прогнозирования хаотического ряда, составленного из рядов Мандельброта [25] (первые 550 элементов), Мэки-Гласса [26] (следующие 562 элемента) и чисел Вольфа (последние 288 элементов). Задача прогнозирования этой последовательности сильно осложняется тем, что она 2 раза резко меняет свои характеристики.

В качестве локальных предикторов использовались радиально-базисные искусственные нейронные сети [14],

входной слой которых был образован линиями чистой задержки с различной глубиной памяти. Для настройки синаптических весов выходного слоя нейропредикторов использовался модифицированный алгоритм стохастической аппроксимации [27]:

$$\begin{cases} w(k+1) = w(k) + r^{-1}(k)(y(k) - w^T(k)\varphi(x(k)))\varphi(x(k)), \\ r(k) = \alpha r(k-1) + \|\varphi(x(k))\|^2, \end{cases} \quad (12)$$

где α – параметр забывания устаревшей информации. При $\alpha = 0$ алгоритм (12) совпадает с одношаговым алгоритмом Качмажа-Уидроу-Хоффа [28], а при $\alpha = 1$ – с алгоритмом стохастической аппроксимации Гудвина-Рэмеджа-Кейнеса [22].

Всего было синтезировано 14 локальных предикторов, различающихся объемом используемой предыстории и параметром забывания устаревшей информации. В начальный момент времени центры радиально-базисных функций случайно распределялись в пространстве входов, все синаптические веса инициализировались значением 0,5.

Для оценки точности получаемых прогнозов использовалась величина среднеквадратичной ошибки прогнозирования:

$$\text{RMSE}(k, N) = \sqrt{\frac{\sum_{i=1}^N e^2(k+i)}{N}}, \quad (13)$$

где N – объем выборки наблюдения, а $e(k)$ – ошибка прогнозирования тестовой выборки в момент времени k .

В ходе эксперимента было синтезировано два ансамбля нейропредикторов. Один из них имел постоянную структуру (4 предиктора), а другой – адаптируемую. Критерием для управления числом узлов в адаптируемом ансамбле было использование 4-х лучших локальных предикторов из имеющихся в наличии. Такой критерий позволяет наглядно увидеть разницу между обычной и настраиваемой метасетью, так как количество узлов в них одинаково в каждый момент времени. Оценка вектора весов выходного слоя ансамбля строилась на основании последних 100 наблюдений.

Необходимо отметить, что среди локальных предикторов нельзя выделить такой, который был бы лучшим на всем исследуемом временном ряде. На разных участках исследуемого ряда лучшими были различные предикторы.

Результаты работы синтезированных ансамблей приведены в таблице 1. В первом столбце приведено значение $\text{RMSE}(100, 1300)$, которое характеризует прогноз на всем наборе данных, за исключением первых 100 элементов, необходимых для начального обучения. В столбце 2 приведено значения $\text{RMSE}(100, 450)$, в столбце 3 – $\text{RMSE}(551, 562)$, а в столбце 4 – $\text{RMSE}(1113, 288)$, характеризующие точность прогноза рядов Мандельброта, Мэки-Гласса и чисел Вольфа соответственно.

Таблица 1 – Результаты моделирования

	1	2	3	4
Фиксированная архитектура	0.71362	0.64042	0.8507	0.49602
Адаптируемая архитектура	0.57067	0.57211	0.65048	0.36455

Как видно из приведенных данных, ансамбль с адаптируемой архитектурой оказался лучше ансамбля с фиксированной архитектурой не только на исследуемом временном ряде в целом, но и на отдельных его фрагментах. На рассматриваемой выборке преимущество метасети с изменяемой архитектурой относительно фиксированной составило 20%.

5 ЗАКЛЮЧЕНИЕ

Предложен алгоритм изменения количества узлов в ансамбле нейронных сетей. Использование предлагаемого алгоритма позволяет достичь компромисса между сложностью используемой нейромодели и точностью получаемых результатов и проводить обработку данных в реальном масштабе времени.

Результаты имитационного моделирования показывают, что нейронная метасеть с настраиваемым количеством узлов обеспечивает более высокую точность прогнозирования по сравнению с сетью фиксированной структуры.

Научная новизна работы заключается в том, что в ней впервые предложен алгоритм изменения количества узлов в ансамбле нейропредикторов, позволяющий производить обработку данных в реальном времени. Практическая значимость данной работы в возможности получения существенно более точных результатов при тех же вычислительных затратах на настройку модели.

ПЕРЕЧЕНЬ ССЫЛОК

1. Nelles O. Nonlinear System Identification. – Berlin: Springer, 2001. – 431 p.
2. Льюнг Л. Идентификация систем. Теория для пользователя: Пер. с англ./ Под ред. Я.З. Цыпкина. – М.: Наука. Гл. ред. физ.-мат. лит., 1991. – 432 с.
3. Handbook of Neural Computation. IOP Publishing and Oxford University Press, 1997. – 618 p.
4. Hashem S. Optimal linear combinations of neural networks // Neural Networks. – 1997. – 10. – P. 599–614.
5. Jordan M., Lei Xu. Convergence results for the EM approach to mixtures of experts architectures // AIM-1458. – 1993. – P. 33.
6. Tresp V., Taniguchi M. Combining estimators using non-constant weighting function // Advances in Neural Information Processing Systems. – 1995. – 7. – P. 419–426.
7. Hashem S. Effects on collinearity on combining neural networks // Connect. Sci. – 1996. – 8. – P. 315–336.
8. Sharkey A.J.C. On combining artificial neural nets // Connect. Sci. – 1996. – 8. – P. 299–313.
9. Bodyanskiy Ye.V., Vorobyov S.A., Stephan A. Algorithm for adaptive identification of dynamical parametrically nonstationary objects // Journal of Computer and System Sciences International. – 1999. – 38. – №1 – P. 14–18.
10. Naftaly U., Intrator N., Horn D. Optimal ensemble averaging of neural networks // Network: Comput. Neural Syst. – 1997. – 8. – P. 283–296.
11. Optiz D.W., Shavlik J.W. Actively searching for an effective neural network ensemble // Connect. Sci. – 1996. – 8. – P. 337–353.
12. Бодянский Е.В., Слипченко А.Н. Оптимизация ансамбля нейропредикторов // Вестник Национального технического университета "Харьковский политехнический институт". Сборник научных трудов. Тематический выпуск:

- "Информатика и моделирование". – Харьков: НТУ "ХПИ". – 2003. – №26. – С. 35–40.
13. Слипченко А.Н. Управление архитектурой нейронной сети // Тези доповідей учасників VI Міжнародної науково-практичної конференції студентів, аспірантів, молодих вчених 1–3 липня 2004 р., м. Київ. – Київ: Друк. НТУУ "КПІ". – 2004. – С. 81–82.
 14. Poggio T., Girosi F. A Theory of Networks for Approximation and Learning. A.I. Memo No. 1140, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1989.
 15. Platt J. A resource allocating network for function interpolation // Neural Computation. – 1991. – 3. – P. 213–225.
 16. Nag A., Ghosh J. Flexible resource allocating network for noisy data // In SPIE Conf. on Applications and Science of Computational Intelligence, SPIE Proc. Vol. 3390, Orlando, Fl., April 1998. – P.551–559.
 17. Yingwei L., Sundararajan N., Saratchandran P. Performance evaluation of a sequential minimal radial basis function (RBF) neural network learning algorithm // IEEE Trans. on Neural Networks. – 1998. – 9. № 2. P.308–318.
 18. Fahlman S. E., Lebiere C. The cascade-correlation learning architecture. Technical Report CMU-CS-90-100, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1990.
 19. Cun Y. L., Denker J. S., Solla S. A. Optimal Brain Damage // Advances in Neural Information Processing Systems, D.S. Touretzky, Eds. – 1989. – P.598–605.
 20. Hassibi B., Stork D. G. Second-order derivatives for network pruning: Optimal brain surgeon // Advances in Neural Information Processing Systems, Hanson et al., Eds. – 1993. – P. 164–171.
 21. Prechelt L. Connection pruning with static and adaptive pruning schedules // Neurocomputing. – 1997. – 16. P.49–61.
 22. Перельман И.И. Оперативная идентификация объектов управления. – М.: Энергоиздат, 1982. – 272 с.
 23. Бодянский Е.В., Слипченко А.Н. Управление количеством узлов в нейронных сетях, обучаемых в реальном времени // АСУ и приборы автоматики. 2004. № 127. – С. 90–97.
 24. Гантмахер Ф.Р. Теория матриц. – 4-е изд. – М.: Наука. Гл. ред. физ.-мат. лит., 1988. – 552 с.
 25. Mandelbrot B.B. Die fraktale Geometrie der Natur. – Basel: Birkhaeuser Verlag, 1991. – 491 S.
 26. Mackey, M.C., Glass, L.: Oscillation and chaos in physiological control systems. Science 197 (1977) 287–289.
 27. Бодянский Е.В., Михальов О.І., Плісс І.П. Адаптивне виявлення розладнань в об'єктах керування за допомогою нейронних мереж. – Дніпропетровськ: Системні технології, 2000. – 140 с.
 28. Kaczmarz S. Approximate solution of systems of linear equations // Int. J. Control. – 1993 – 53. – P. 1269–1271.

Надійшла 27.09.2004

У роботі розглядається задача прогнозування часових рядів з використанням ансамблю нейропродикторів. Запропоновано алгоритм зміни числа вузлів у ансамблі за умови обробки інформації у реальному часі.

The work is devoted to the task of time-series prediction using ensemble of neuropredictors. An algorithm capable of changing the number of predictors in ensemble in real time is proposed.

УДК 681.3

Л. А. Жуков, Н. В. Решетникова

ИССЛЕДОВАНИЕ РАЗЛИЧНЫХ ВАРИАНТОВ КонтРАСТИРОВАНИЯ И БИНАРИЗАЦИИ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОБРАБОТКИ МЕДИЦИНСКИХ ДАННЫХ

Исследовались различные варианты контрастирования и бинаризации нейронных сетей на примере медицинских данных. Выполнено сравнение качества тестирования для бинаризованных и небинаризованных синапсов.

ВВЕДЕНИЕ

В работе предлагаются отдельные элементы технологии для обработки параметров анализов крови доноров электролизных цехов Красноярского алюминиевого завода, подвергшихся иммунному плазмаферезу. Сбор исходных данных и их первичная обработка проводились на базе Красноярской краевой станции переливания крови за период с 1978 г. по 1992 г. врачом Петровской В.А. [1, 2].

Цель работы: выявить влияние процедур контрастирования и бинаризации на качество тестирования нейросетей и сходство и различие при различных комбинациях этих процедур.

Задачи работы: провести изолированно процедуры бинаризации и контрастирования на примере медицинских данных; провести эксперименты и изучить результаты при различных комбинациях данных подходов,

например, при бинаризации после и до контрастирования. Отметим, что в данной работе не ставилась задача достижения наилучших результатов тестирования.

Исходные данные включают гематологические, биохимические и иммунологические показатели и дополнительно следующие параметры: величина титра антител, число плазмаферезов и некоторые другие характеристики, указывающие на функциональные сдвиги в системе гомеостаза в ответ на введение стафилококкового антигена и под влиянием многолетних плазмаферезов [3, 4].

Для выполнения полномасштабных экспериментов и проверки некоторых частных гипотез был создан общий файл данных, содержащий всю информацию по всем категориям доноров: анодчики – 266 записей, электролизники – 186 записей, крановщики 168, слесари – 198, сварщики – 130, литейщики – 124, студенты – 472. Каждая запись представляет собой данные анализа крови донора, взятого за одно посещение. Следует отметить, что не все изучаемые поля заполнены, что связано со спецификой проведения и обработки анализов крови, т.к. в соответствии с ведомственной медицинской инструкцией некоторые анализы берутся у доноров один раз за несколько сдач крови [5, 6]. Для обработки были использо-

ваны нейронные сети с учителем [7]. Для повышения качества обучения нейронных сетей были реализованы дополнительные операции по предобработке данных. Эти операции позволили включить в число полей дополнительно следующие: идентификация категорий, год, дни недели, сезоны года (зима, осень, весна, лето).

1 ОБЗОР СУЩЕСТВУЮЩИХ МЕТОДОВ ОБРАБОТКИ МЕДИЦИНСКИХ ДАННЫХ С ПОМОЩЬЮ НЕЙРОСЕТЕЙ

Для задач медицинского плана часто характерно приблизительное, реже относительно точное представление о характере взаимосвязи изучаемых параметров и почти никогда не известен точный алгоритм решения задачи. Поэтому для решения таких задач используются методы нейроинформатики [8].

При постановке задачи для нейронной сети не всегда удается точно определить сколько и каких входных данных нужно подавать на вход. Гораздо чаще на вход сети подается избыточный набор входных параметров. Например, при обучении сети постановке диагноза в задачах медицинской диагностики на вход сети подаются все данные, необходимые для постановки диагноза в соответствии с существующими методиками. Как правило, при диагностике заболеваний населения какого-нибудь небольшого региона (например города) можно обойтись меньшим набором исходных данных. Требуется определить, какие данные необходимы для решения конкретной задачи, поставленной для нейронной сети. Кроме того, в ходе решения этой задачи определяются значимости входных сигналов. Следует заметить, что умение определять значимость входных сигналов представляет самостоятельную ценность [9].

Нейронная сеть производит из таблицы данных скрытое знание, т. е. формулирует некоторый навык предсказания или классификации. На практике пытаются привести нейронные сети к наиболее простой логической структуре для чего используется понятие логически прозрачной сети. Логически прозрачной называется максимально упрощенная сеть, по структуре которой легко можно сформировать понятный пользователю алгоритм решения задачи [7, 10].

После получения логически прозрачной нейронной сети обычно наступает этап построения вербального описания. Принцип построения вербального описания достаточно прост. Используемая терминология заимствована из медицины. Входные сигналы называют симптомами, выходные сигналы нейронов первого слоя – синдромами первого уровня и т. п. [7, 9]. Однако, оказывается довольно сложно формализовать и расшифровать данное описание применительно к конкретной задаче, особенно если исследуемые параметры имеют сложные и многоуровневые взаимосвязи. Иногда для ряда задач медицинской тематики характерно неполное упрощение нейросетей, т.е. несколько увеличенный набор оставленных входных параметров и выявление их взаимосвязей,

что обусловлено некоторой консервативностью специалистов-медиков и их представлений о результатах.

Для приведения сетей к логически прозрачному виду используется несколько различных подходов [11], но в данной работе нас интересуют прежде всего бинаризация и контрастирование. Исследователями отмечается, что данные процедуры позволяют существенно упростить структуру нейросетей и выявить набор наиболее значимых входных параметров [7, 9, 11]. Существует два варианта контрастирования: 1) выбираются наиболее значимые параметры и к ним подстраиваются менее значимые – сверху-вниз; 2) удаляются малозначимые параметры, начиная с самого малозначимого – снизу-вверх [7]. В данной работе используется вторая реализация контрастирования [10]. Следует проверять результаты обучения нейросетей после контрастирования и бинаризации путем тестирования, т.к. упрощенная сеть не всегда может решить задачу с требуемой точностью [7]. Считается невозможным добиться при упрощении нейронной сети бинаризации всех связей [2, 9, 12, 13], что зависит от сложности как самой задачи, так и взаимосвязей между параметрами. Нужно иметь в виду, что для данной задачи существует некоторое подмножество различных вариантов сетей после упрощения и соответственно несколько вариантов решения данной задачи [7].

2 ПЕРВИЧНОЕ ОБУЧЕНИЕ И ТЕСТИРОВАНИЕ

Для дальнейшей обработки использовался нейроимитатор NeuroPro версии 0.25 (разработчик В.Г. Царегородцев). Данная программа реализует создание, обучение и тестирование искусственных нейронных сетей с учителем [7, 10]. Нейроимитатор реализует несколько видов упрощения сетей, из которых в данной работе нас интересуют сокращение числа синапсов сети, нейронов, неоднородных (пороговых) входов нейронной сети, бинаризация весов синапсов и неоднородных входов. При упрощении сетей действует жесткое ограничение, согласно которому исходное число слоев остается постоянным (слой не может быть удален), т. е. в слое обязательно присутствует хотя бы один нейрон.

При обучении и тестировании используются 10 нейросетей с 3 слоями нейронов, по 10 нейронов в каждом слое, с точностью 1,5 и характеристикой нейронов 0,1, где точность (ошибка обучения) представляет собой допустимое отклонение выходной величины от ее реального значения. В качестве выходного поля использовалось общее количество лейкоцитов (Х6), количество входных параметров – 48. Объем общей выборки составил 300 записей, объемы обучающей и тестовой выборок – 158 и 142 записи соответственно.

Перекрестным тестированием называют процесс проверки качества обучения нейросетей, при котором используются две выборки: обучающая и тестовая (в данном случае полученная из исходной выборкой четных/нечетных записей), по одной из которых обучают (обучающая), а по другой тестируют (тестовая).

В данном случае процент правильно определенных примеров при перекрестном тестировании составил от 69% до 87% (см. табл. 1).

Таблица 1 – Результаты перекрестного тестирования

	Кол-во верно	% верно	Кол-во неверно	% неверно
Сеть 1	114	80,28	28	19,72
Сеть 2	134	94,37	8	5,63
Сеть 3	124	87,32	18	12,68
Сеть 4	129	90,85	13	9,15
Сеть 5	131	92,25	11	7,75
Сеть 6	106	74,65	36	25,35
Сеть 7	126	88,73	16	11,27
Сеть 8	115	80,99	27	19,01
Сеть 9	124	87,32	18	12,68
Сеть 10	124	87,32	18	12,68
Среднее	122,70	86,41	19,30	13,59

3 КОНТРАСТИРОВАНИЕ

Контрастирование – это процесс сокращения числа параметров, не влияющий на качество обучения, при котором удаляются наиболее незначимые параметры при определении выходного поля [14]. В данном случае используется сокращение числа внутренних синаптических связей, т. е. происходит упрощение внутренней структуры нейронных сетей. Контрастирование показало, что в среднем удалено 94% синапсов. Параллельно сокращению синапсов проходило сокращение числа нейронов и входных параметров. В среднем удалено около 84% входных параметров и 43% нейронов.

При перекрестном тестировании после контрастирования были получены существенно лучшие показатели, чем при полном наборе входных параметров (см. табл. 2). В среднем процент правильно определенных примеров составил 95%. Видимо, это связано с тем, что некоторые входные параметры, слабо влияющие или не влияющие на общее количество лейкоцитов, создают шумовой эффект.

Таблица 2 – Результаты перекрестного тестирования после контрастирования

	Кол-во верно	% верно	Кол-во неверно	% неверно
Сеть 1	153	96,84	5	3,16
Сеть 2	149	94,30	9	5,70
Сеть 3	153	96,84	5	3,16
Сеть 4	143	90,51	15	9,49
Сеть 5	148	93,67	10	6,33
Сеть 6	153	96,84	5	3,16
Сеть 7	149	94,30	9	5,70
Сеть 8	151	95,57	7	4,43
Сеть 9	149	94,30	9	5,70
Сеть 10	155	98,10	3	1,90
Среднее	150,30	95,13	7,70	4,87

Причем, при удалении менее 90% синапсов, менее 70% входных параметров и менее 10% нейронов процент правильно определенных примеров не превышает 91%, но при удалении 95% синапсов – процент правильно определенных примеров составляет от 94 до 97%. Самые лучшие результаты тестирования (97–98%) показывают нейросети с процентом удаленных синапсов – от 94% до 95%, удаленных входов – от 88% до 94%, удаленных нейронов – от 37% до 53%.

Следует отметить, что после контрастирования большая часть синапсов представляла собой неоднородные входы, а именно, на входе/выходе в качестве сигналов передавались константы (некоторые действительные числа). Поэтому дополнительно (после контрастирования) было проведено сокращение числа неоднородных входов, в результате чего сократилось число синапсов и нейронов (графическое представление структуры сетей см. рис. 1, где а и б – два варианта нейросетей).

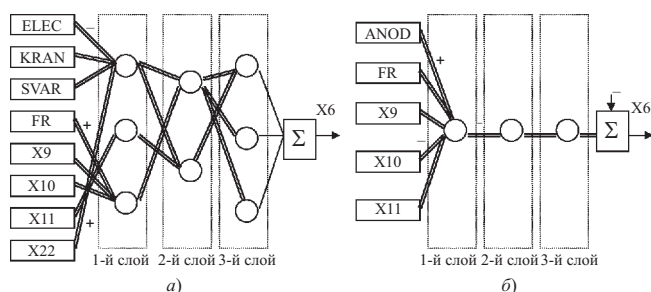


Рисунок 1 – Структура нейронной сети после сокращения неоднородных входов:

○ – нейрон; ↓ – константа (подача постоянного сигнала); → – связь, где вес синапса равен -1; →+ – связь, где вес синапса равен +1; Σ – сумматор; == – связь между элементами сети (небинаризованные синапсы); «-», «+» – знак связи

Было удалено в среднем 65% синапсов (процент рассчитан от оставшегося после контрастирования количества) и 72% нейронов. Причем, чем больше количество удаленных при контрастировании нейронов, тем меньше нейронов удалено при сокращении неоднородных входов. Но для синапсов наблюдается обратная тенденция: чем больше обычных синапсов удалено при контрастировании, тем больше удаляется их при сокращении числа неоднородных входов, что тривиально. Удаляются, в основном, нейроны второго и третьего уровней, минимальная структура состоит из трехслойной сети, каждый слой содержит всего один нейрон (см. рис. 1, б). Что касается перекрестного тестирования, его результаты после данной процедуры не изменились.

4 БИНАРИЗАЦИЯ И КОМБИНИРОВАНИЕ

Процедура бинаризации параметров сети представляет собой приведения весов синапсов к некоторому набору конкретных значений, в данном случае используется их приведение к значениям -1 и 1.

При бинаризации без контрастирования процент бинаризованных синапсов в среднем составляет 84–87%. Причем результаты перекрестного тестирования после бинаризации дают приблизительно такие же результаты, процент верно определенных примеров в среднем составляет от 78% до 83% (см. табл. 3). В среднем лучшие результаты тестирования дают нейросети с большим процентом бинаризованных синапсов.

Таблица 3 – Результаты перекрестного тестирования после бинаризации

	Кол-во бинаризованных синапсов	% бинаризованных синапсов	Кол-во верно	% верно	Кол-во неверно	% неверно
Сеть 1	642	89,04	111	78,17	31	21,83
Сеть 2	603	83,63	134	94,37	8	5,63
Сеть 3	569	78,92	121	85,21	21	14,79
Сеть 4	625	86,69	118	83,10	24	16,90
Сеть 5	643	89,18	120	84,51	22	15,49
Сеть 6	641	88,90	117	82,39	25	17,61
Сеть 7	623	86,41	110	77,46	32	22,54
Сеть 8	640	88,77	116	81,69	26	18,31
Сеть 9	637	88,35	114	80,28	28	19,72
Сеть 10	641	88,90	124	87,32	18	12,68
Среднее	626,40	86,88	118,50	83,45	23,50	16,55

После процедуры контрастирования (контрастирование, затем бинаризация) процент бинаризованных синапсов значительно падает и составляет уже от 29% до 39%. Тенденция получения лучших результатов перекрестного тестирования сохраняется аналогично процедуре контрастирования (см. табл. 4). Результаты перекрестного тестирования практически не отличаются от полученных при контрастировании без бинаризации, процент верно определенных примеров в среднем составляет 95–96%.

Таблица 4 – Результаты тестирования после контрастирования и бинаризации

	Кол-во бинаризованных синапсов	% бинаризованных синапсов	Конечное кол-во синапсов	Кол-во удаленных синапсов	% удаленных синапсов	Кол-во верно	% верно	Кол-во неверно	% неверно
Сеть 1	11	37,93	29	692	95,98	152	96,20	6	3,80
Сеть 2	11	36,67	30	691	95,84	155	98,10	3	1,90
Сеть 3	15	38,46	39	682	94,59	155	98,10	3	1,90
Сеть 4	5	5,95	84	637	88,35	144	91,14	14	8,86
Сеть 5	16	39,02	41	680	94,31	155	98,10	3	1,90
Сеть 6	30	57,69	52	669	92,79	155	98,10	3	1,90
Сеть 7	29	50,00	58	663	91,96	140	88,61	18	11,39
Сеть 8	19	42,22	45	676	93,76	156	98,73	2	1,27
Сеть 9	10	27,78	36	685	95,01	147	93,04	11	6,96
Сеть 10	20	55,56	36	685	95,01	158	100,00	0	0,00
Среднее	16,60	39,13	45	676,00	93,76	151,70	96,01	6,30	3,99

Исходно каждая сеть содержала 721 синапс, 30 нейронов, 48 входных сигналов. При проведении процедуры контрастирования после бинаризации, процент пра-

вильно определенных примеров составляет в среднем 83,5%. Количество удаленных синапсов варьируется от 75 до 692 из 721 возможного, что составляет в среднем 71%, процент удаленных входов в среднем составляет 27%, удаленных нейронов – 33%. Причем, в среднем, чем больше процент бинаризованных синапсов, тем меньше процент удаленных синапсов, входов и нейронов. Удаляется не более 12% синапсов (входные параметры и нейроны не удаляются), если процент бинаризации превышает 87%. Если процент бинаризации не превышает 38%, в среднем удаляется 95% синапсов, 87% входов и 60% нейронов. При значениях процента бинаризации 60–70%, в среднем удаляется 73% синапсов, но не более 11% входных параметров и 27% нейронов. Итак, процент удаления синапсов возрастает пропорционально уменьшению процента бинаризации, что характерно для нейронов и входных параметров.

Аналогично отдельно проведенной процедуре контрастирования, при контрастировании после бинаризации наблюдается относительно большое количество неоднородных входов. Для упрощения структуры дополнительно проводилось сокращение числа неоднородных входов, в ходе которого удалялись синапсы и нейроны. В среднем удалено 27% синапсов (процент от оставшегося после контрастирования количества синапсов) и 45% нейронов. Процент удаленных синапсов после сокращения неоднородных входов увеличивается согласно увеличению удаленных синапсов после контрастирования (см. табл. 5). Т.е. чем больше обычных синапсов удаляется при контрастировании, тем больше их удалено и при сокращении неоднородных входов, что тривиально.

Таблица 5 – Результаты тестирования после сокращения неоднородных входов

	Конечное кол-во синапсов	% удаленных синапсов	Кол-во удаленных синапсов после сокращения неоднородных входов	% удаленных синапсов после сокращения неоднородных входов	Кол-во удаленных нейронов после сокращения неоднородных входов	% удаленных нейронов после сокращения неоднородных входов	Кол-во верно	% верно
Сеть 1	58	91,96	21	36,21	9	69,23	144	91,14
Сеть 2	92	87,24	23	25,00	11	64,71	130	82,28
Сеть 3	29	95,98	21	72,41	9	75,00	152	96,20
Сеть 4	33	95,42	22	66,67	10	76,92	150	94,94
Сеть 5	215	70,18	23	10,70	9	47,37	121	76,58
Сеть 6	638	11,51	38	5,96	1	3,33	110	69,62
Сеть 7	53	92,65	23	43,40	9	64,29	140	88,61
Сеть 8	360	50,07	4	1,11	2	6,90	95	60,13
Сеть 9	215	70,18	11	5,12	5	21,74	110	69,62
Сеть 10	31	95,70	21	67,74	9	75,00	148	93,67
Среднее	172	76,09	20,70	33,43	7,40	50,45	130,00	82,28

При удалении после контрастирования более 95% синапсов, при сокращении неоднородных входов удаляется более 66% синапсов; менее 73% удаленных синапсов – удаляется не более 10% синапсов. Такая же тенденция наблюдается и для нейронов. Минимальная сеть, полу-

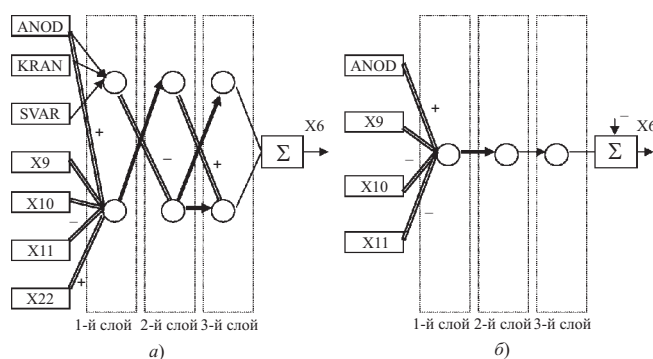


Рисунок 2 – Структура нейронной сети после бинаризации и контрастирования:

○ – нейрон; ↓ – константа (подача постоянного сигнала); → – связь, где вес синапса равен -1; → – связь, где вес синапса равен +1; Σ – сумматор; — – связь между элементами сети (небинаризованные синапсы); «-», «+» – знак связи

ченная в проведенных экспериментах, состоит из трех слоев, по одному нейрону в каждом слое (см. рис. 2, б). Частично бинаризованными являются синапсы нейронов всех слоев. Результаты перекрестного тестирования не отличаются от полученных при контрастировании после бинаризации. Лучшие результаты тестирования дают нейросети с большим процентом удаленных синапсов и нейронов после сокращения неоднородных входов.

5 КОМБИНИРОВАНИЕ БИНАРИЗАЦИИ И КОНТРАСТИРОВАНИЯ

Были проведены эксперименты по последовательному совмещению процедур бинаризации и контрастирования. В первом случае последовательно проводилось контрастирование – бинаризация – контрастирование (КБК), во втором, бинаризация – контрастирование – бинаризация (БКБ).

5.1 Контрастирование-бинаризация-контрастирование

При первой последовательности процедур (КБК) процент удаленных синапсов составил в среднем 94%, дополнительно (после второго контрастирования) удалено в среднем 9% синапсов, 25% входов и 6% нейронов. Заметим, если процент бинаризованных синапсов не превышает 10% (при этом процент удаленных нейронов минимален по сравнению с остальными сетями и не превышает 27%), то повторного контрастирования не происходит.

В среднем процент бинаризованных синапсов не превышает 39%, что связано с высокими процентами удаленных синапсов, входных сигналов и нейронов после первой процедуры контрастирования: в среднем удалено 94% синапсов, 83% входов и 45% нейронов.

Всего обучено 20 сетей, более подробные результаты 10 из них приведены в таблице 6. Процент правильно определенных примеров при перекрестном тестировании составил в среднем 96%. Более низкие результаты при тестировании по сравнению с другими дают нейросети, у которых либо повторного контрастирования не происхо-

дит вообще, либо суммарное количество дополнительно (после второй процедуры контрастирования) удаленных синапсов, нейронов и входов составляет 4–7, но процент бинаризации в этом случае не превышает 28%.

Таблица 6 – Результаты перекрестного тестирования (КБК)

	Кол-во бинаризованных синапсов	% бинаризованных синапсов	Кол-во удаленных синапсов	% удаленных синапсов	Кол-во верно	% верно	Кол-во неверно	% неверно	Дополнительно удалено синапсов
Сеть 1	11	37,93	692	95,98	158	100,00	0	0,00	1
Сеть 2	11	36,67	691	95,84	157	99,37	1	0,63	1
Сеть 3	15	38,46	682	94,59	156	98,73	2	1,27	3
Сеть 4	5	5,95	637	88,35	144	91,14	14	8,86	0
Сеть 5	16	39,02	680	94,31	156	98,73	2	1,27	2
Сеть 6	30	57,69	669	92,79	155	98,10	3	1,90	7
Сеть 7	29	50,00	663	91,96	142	89,87	16	10,13	4
Сеть 8	19	42,22	676	93,76	155	98,10	3	1,90	6
Сеть 9	10	27,78	685	95,01	149	94,30	9	5,70	3
Сеть 10	20	55,56	685	95,01	158	100,00	0	0,00	0
Среднее	16,60	39,13	676	93,76	153,00	96,83	5,00	3,17	2,70

5.2 Бинаризация-контрастирование-бинаризация

При последовательности БКБ процент бинаризованных синапсов составил в среднем 62%, на втором этапе (после второй процедуры бинаризации) бинаризовано в среднем 4% синапсов. Причем, если процент бинаризации (при первой процедуре бинаризации) превышает 88%, то после контрастирования дальнейшей бинаризации не происходит. Процент правильно определенных примеров при перекрестном тестировании в среднем составляет 84% (см. табл. 7). Наилучшие результаты при тестировании дают нейросети, у которых процент дополнительно (после второй бинаризации) бинаризованных синапсов значительно больше остальных (от 6% до 26%).

Таблица 7 – Результаты перекрестного тестирования (БКБ)

	Кол-во бинаризованных синапсов	% бинаризованных синапсов	Кол-во удаленных синапсов	% удаленных синапсов	Кол-во верно	% верно	Кол-во неверно	% неверно	Дополнительно бинаризовано синапсов
Сеть 1	188	80,00	486	67,41	123	86,62	19	13,38	1
Сеть 2	18	38,30	674	93,48	135	95,07	7	4,93	5
Сеть 3	134	68,72	526	72,95	117	82,39	25	17,61	0
Сеть 4	131	70,81	536	74,34	119	83,80	23	16,20	3
Сеть 5	221	80,66	447	62,00	116	81,69	26	18,31	0
Сеть 6	11	34,38	689	95,56	137	96,48	5	3,52	2
Сеть 7	409	82,46	225	31,21	107	75,35	35	24,65	2
Сеть 8	570	88,24	75	10,40	96	67,61	46	32,39	3
Сеть 9	196	81,67	481	66,71	113	79,58	29	20,42	1
Сеть 10	95	69,85	585	81,14	126	88,73	16	11,27	0
Среднее	197,30	69,51	472	65,52	118,9	83,73	23,10	16,27	1,70

Аналогично предыдущим экспериментам по контрастированию и бинаризации здесь выявлено большое количество неоднородных входов. В среднем дополнительно было удалено 56% синапсов (от оставшегося количества) и 64% нейронов. На результатах перекрестного тестирования данная процедура не отразилась.

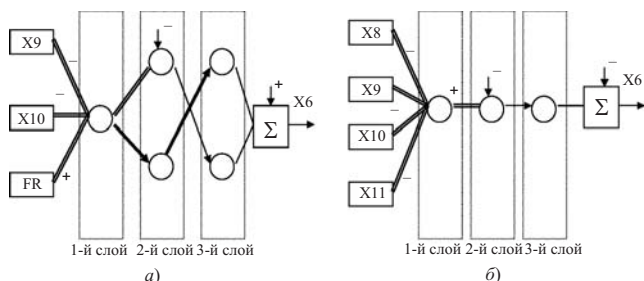


Рисунок 3 – Структура нейронной сети после бинаризации и контрастирования:

○ – нейрон; ↓ – константа (подача постоянного сигнала); → – связь, где вес синапса равен -1; → – связь, где вес синапса равен +1; Σ – сумматор; — — — — — связь между элементами сети (небинаризованные синапсы); «←», «→» – знак связи

Приведем графическое представление вербального описания нескольких нейросетей, полученных после последовательности процедур бинаризации и контрастирования (см. рис. 3). Бинаризованными являются нейроны всех 3 слоев вне зависимости от порядка проведения процедур контрастирования и бинаризации, больше бинаризованных синапсов у нейронов второго слоя. Минимальная структура сети представляет собой 3 слоя, по одному нейрону в каждом слое (см. рис. 3, б).

ЗАКЛЮЧЕНИЕ

В данной работе проведено исследование и сравнение таких методов упрощения нейронных сетей, как бинаризация и контрастирование. Работа проведена на нейронных сетях фиксированного размера и структуры. Отмечены сходства и различия в результатах их применения для задачи предикции на основе медицинских данных иммунного плазмафереза. Были получены результаты перекрестного тестирования для различных вариантов бинаризации, контрастирования внутренних синапсов и неоднородных входов, и комбинации данных процедур. Выявлено снижение процента бинаризованных синапсов после контрастирования и их комбинации, и довольно высокий процент бинаризации до контрастирования. Результаты перекрестного тестирования для данной задачи после отдельных бинаризации и особенно контрастирования по проведенным экспериментам дают результаты в среднем лучше, чем исходные без использования данных процедур. При комбинировании бинаризации и контрастирования результаты тестирования соответствуют первой из проведенных процедур. Следовательно, процедуры контрастирования и бинаризации целесообразно использовать

для повышения качества обучения нейросетей на примере медицинских данных. Возможно использование описанного подхода для обработки различных данных из других предметных областей.

СПИСОК ЛИТЕРАТУРЫ

1. Решетникова Н.В., Петровская В.А., Товбис Е.М., Привалко Е.Ю., Жуков Л.А. Нейросетевая оценка влияния условий труда на параметры крови // Научная сессия МИФИ-2003. V Всероссийская научно-техническая конференция «Нейроинформатика-2003»: Сб. науч. трудов. Ч. 2. М.: МИФИ, 2003. С. 77–85
2. Решетникова Н.В., Петровская В.А., Товбис Е.М., Жуков Л.А. Анализ влияния условий труда на параметры крови с помощью нейросетей // Актуальные вопросы здравоохранения и медицинской науки: Сборник научных трудов / Под ред. д. м. н., проф. О.М. Новикова. Выпуск III. Красноярск, 2003. С. 194–204.
3. Клиническая иммунология: Руководство для практических врачей и врачей-интернов. – Красноярск, 1998. – 87 с.
4. Константинов В.Н. Белковосинтетическая функция печени и факторы неспецифического иммунитета при плазмаферезе у доноров // Автореф. дисс. на соиск. уч. ст. канд. мед. наук. Актюбинск, 1971.
5. Приказ Минздрава РФ от 23 сентября 2002 г. № 295 «Об утверждении «Инструкции по проведению донорского прерывистого плазмафереза»».
6. Демидова Н.В. Функциональные системы организма доноров гипериммунной плазмы. – М.: Медицина, 1987. – 160 с.
7. Горбань А.Н., Россиев Д.А. Нейронные сети на персональных компьютерах. Н.: Наука, 1996.
8. Царегородцев В.Г. Производство полужемпирических знаний из таблиц данных с помощью обучаемых искусственных нейронных сетей // Методы нейроинформатики. – Красноярск: Издательство КГТУ, 1998.
9. Россиев Д.А. Медицинская нейроинформатика / Нейроинформатика Новосибирск: Наука, Сибирская издательская фирма РАН, 1998.
10. Горбань А.Н., Миркес Е.М. Нейронные сети ассоциативной памяти, функционирующие в дискретном времени. Вычислительный центр СО РАН в г. Красноярск. Красноярск, 1997. 23 с., библиогр. 8 назв. (Рукопись деп. в ВИНТИ 17.07.97, № 2436–В97)
11. Гилев С.Е., Коченов Д.А., Миркес Е.М., Россиев Д.А. Контрастирование, оценка значимости параметров, оптимизация их значений и их интерпретация в нейронных сетях. // Нейроинформатика и ее приложения: Материалы III Всероссийского семинара; КГТУ. Красноярск, 1995. С. 66–78.
12. Россиев Д.А., Головенкин С.Е., Назаров Б.В., Шульман В.А., Матушин Г.В. Определение информативности медицинских параметров с помощью нейронной сети // Диагностика, информатика и метрология – 1994. – Тез. научно-технической конференции (г. Санкт-Петербург, 28–30 июня 1994 г.). С.-Петербург, 1994. – С. 348.
13. Оценка влияния плазмафереза на параметры крови с помощью методов статистики и нейросетей / Решетникова Н.В., Жуков Л.А. Красноярск. госуд. техн. ун-т. Красноярск, 2003. 33 с.: ил. 2. – Библиогр.: 15 назв. – Рус. – Рукопись деп. в ВИНТИ 29.09.2003 № 1737–В2003.
14. Гордиенко П.В. Стратегии контрастирования // Нейроинформатика и ее приложения. Тезисы докладов 5 Всероссийского семинара, 3–5 октября 1997 г. / Под ред. А.Н.Горбаня. Красноярск: изд. КГТУ, 1997. С. 69.

Надійшла 09.12.2003
Після доробки 20.05.2004

Досліджені різні варіанти контрастування і бинаризації нейронних мереж на прикладі медичних даних. Виконано порівняння якості тестування для бинаризованих та небинаризованих синапсів.

Various variants of neural networks contrast and binary were investigated by the medical data example. Comparison of testing quality for binary synapse was executed.

М. А. Новотарский

КЛЕТОЧНЫЕ НЕЙРОННЫЕ СЕТИ С ТРАНЗИТНЫМИ ПЕРЕСЫЛКАМИ

В работе дано обоснование актуальности создания более сложных, по сравнению с классическими, структур нейронов, оперирующих с дискретным представлением информации. Рассмотрены принципы функционирования клеточных нейронных сетей в непрерывном и дискретном времени. Описана динамика функционирования дискретной клеточной сети, ориентированной на решение операторных уравнений численными методами. Предложен механизм оптимальных транзитных пересылок для реализации итерационных вычислений на виртуальном шаблоне.

ВВЕДЕНИЕ

Начиная с момента зарождения вычислительной техники и до сегодняшнего времени широко распространено мнение об идеологическом и практическом превосходстве технических конструкций над биологическими. Отсюда вытекает уверенность в том, что человеческая память – это всего лишь несовершенный вариант компьютерной памяти, и для того, чтобы понять секреты работы мозга, следует больше внимания уделять разработке и конструированию компьютерных систем. Последний тезис сыграл отрицательную роль в выборе направлений развития теории искусственных нейронных сетей. Как правило, разработчики интересовались биологическими аналогиями только на самых первых этапах проекта и вскоре теряли к ним интерес. В результате такого подхода искусственные нейронные сети очень быстро превратились в некоторый инструмент, ориентированный на решение ограниченного круга задач. Литература изобилует информацией об использовании искусственных нейронных сетей для задач аппроксимации функций, распознавания образов, классификационных задач, задач адаптивного управления и др. В то же время очень мало сообщений о применении нейронных сетей к вычислительным задачам, хотя наука накопила немало фактов феноменальных вычислительных способностей человека. Возможно, причина такого положения дел кроется в самодостаточности классической модели формального нейрона, предложенной Уорреном С. МакКаллоком и Вальтером Питтсом [1]. Обладая свойствами универсального аппроксиматора, такая структура сама по себе представляет интерес как средство построения моделей. Не следует исключать и субъективный фактор, который заключается в том, что сложилась историческая традиция относить к искусственным нейронным сетям большей частью те архитектуры, которые используют формальный нейрон или незначительные его модификации.

В последнее время наблюдается тенденция отхода от классического понимания теории нейронных сетей. К ярким представителям нового поколения можно отнести

АРТ-сети [2], «нейронный газ» [3], клеточные сети [4] и ряд других. Эти структуры используют более сложные модели нейрона, позволившие расширить круг задач, традиционно решаемых с помощью нейронных сетей. В рамках поддержки данной тенденции в работе рассмотрены подходы к построению дискретных клеточных сетей, использующие в качестве нейрона специализированное вычислительное устройство.

1 КЛЕТОЧНЫЕ НЕЙРОННЫЕ СЕТИ

Среди архитектур нейронных сетей, ориентированных на моделирование вычислительных процессов, особое место занимают клеточные нейронные сети. Подобные структуры широко известны в современной физике как сложные системы, самоорганизующиеся системы, границы хаоса, системы с коллективным поведением и др.

В монографии [5] они названы клеточными нелинейными сетями, а клеточные нейронные сети (КНС) рассматриваются как их разновидность. Согласно [5], КНС – это массово параллельная вычислительная парадигма, определенная на N -мерном дискретном пространстве. Она состоит из N -мерного гомогенного массива элементов (клеток). Структура типовых связей между клетками КНС показана на рис. 1.

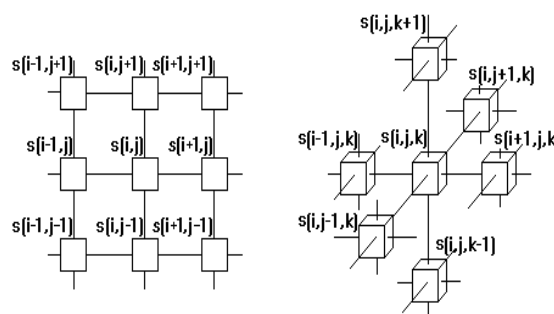


Рисунок 1 – Фрагменты двумерной и трехмерной КНС

Количество связей нейрона с окружающими соседями определяется его индексом окружения α . Структуры КНС, показанные на рис. 1, обладают минимально допустимыми индексами окружения: для двумерной сети – $\alpha = 4$ и для трехмерной сети – $\alpha = 6$. Если сеть обладает индексом окружения, большим за минимальный, то на такой сети может быть задано несколько функций соседства:

$$N_r : H_{\alpha}, N_r(g) = \{z | d(g, z) \leq r\}, \quad (1)$$

где H – множество координат всех клеток, H_α – множество координат соседних клеток, g, z – координатные векторы, r – степень соседства.

Степенью соседства r называют максимальное координатное расстояние между рассматриваемым нейроном и его соседями. Способ определения расстояния задает функция $d(g, z) = |g - z|$.

Каждая клетка представляет собой процессор, который обладает количеством входов, соответствующим индексу окружения, и единственным выходом. Функционирование такого процессора сводится к вычислению параметрических функционалов, заданных в явной или неявной форме с непрерывным или дискретным временем. Общая форма динамического уравнения с явными функционалами для непрерывного времени имеет вид:

$$\begin{aligned} \frac{dx_g(t)}{dt} = & C[x_g(t)] + \sum_{z \in N_r(g)} A_{g,z}(x_g(t-\tau), y_z(t-\tau); p_g) + \\ & + \sum_{z \in N_r(g)} B_{g,z}(x_g(t-\tau), u_z(t-\tau); p_g^B) + I_g(t), \\ y_g(t) = & F(x_g(t-\tau)), \end{aligned} \quad (2)$$

а для дискретного времени:

$$\begin{aligned} x_g(n+1) = & C[x_g(n)] + \sum_{z \in N_r(g)} A_{g,z}(x_g(n-m), y_z(n-m); p_g^A) + \\ & + \sum_{z \in N_r(g)} B_{g,z}(x_g(n-m), u_z(n-m); p_g^B) + I_g(n), \\ y_g(n) = & f(x_g(n-m)), \end{aligned} \quad (3)$$

где x_g – внутреннее состояние клетки $s(g)$, y_z – выходной сигнал клетки $s(z)$, u_z – внешний входной вектор клетки $s(z)$, I_g – смещение, C – локальная мгновенная функция обратной связи, p^A, p^B – массивы параметров, $A_{g,z}, B_{g,z}$ – функционалы, соответствующие связям между клетками $s(g)$ и $s(z)$.

Задача, подлежащая моделированию с помощью КНС, может быть также представлена в виде операционного уравнения:

$$\mathbf{LX} = \mathbf{U}, \quad (4)$$

где \mathbf{L} – дифференциальный оператор, \mathbf{X} – матрица состояния КНС, \mathbf{U} – функция правой части.

Для решения данного уравнения следует воспользоваться одним из известных итерационных методов [6], которые предполагают сведение уравнения (4) к системе разностных уравнений на гомогенной сетке Ω с шагом h :

$$\mathbf{L}_h \mathbf{X}_h = \mathbf{U}_h, \quad (5)$$

где \mathbf{L}_h – разностный оператор, \mathbf{X}_h – разностная матрица состояния, \mathbf{U}_h – сетевая функция.

В простейшем случае устанавливают прямое соответствие нейрона клеточной сети узлу сеточной области. Тогда динамика клетки должна обеспечивать решение одного разностного уравнения с помощью итерационной

формулы, соответствующей выбранному методу. При выборе подходящей итерационной формулы существенную роль играет вид разностного оператора \mathbf{L}_h . Основными критериями выбора разностного оператора являются форма шаблона, на которой он определен, и требования к порядку обновления данных на заданном шаблоне.

Локально-асинхронный метод [7], ориентированный на применение в клеточных нейронных сетях, позволяет сформировать шаблон разностного оператора, ограниченный степенью соседства r , и обеспечить асинхронный режим обмена, снимающий жесткое ограничение на обновления всех данных на шаблоне перед выполнением очередной итерации. Динамика клетки в данном случае определяется итерационной формулой:

$$x_g(n+1) = x_g(n) - \frac{w_g}{h^2} [\mathbf{L}_g x_g(n) - u_g], \quad (6)$$

где \mathbf{L}_g – компонента разностного оператора, $x_g(n)$ – состояние клетки s_g на итерационном шаге n , h – величина шага дискретизации, w_g – внутренний параметр, u_g – значение сеточной функции.

Порядок обновления данных на шаблоне каждой компоненты разностного оператора \mathbf{L}_g определяется хаотической последовательностью $\{\mathbf{J}_n\}_{n=1}^\infty$ непустых множеств $\{1, 2, \dots, \eta\}$. Используя эту последовательность, строят последовательность итераций $\{x_g(n)\}_{n=1}^\infty$ по правилу:

$$x_g(n) = \begin{cases} x_g(n-1), & g \notin \mathbf{J}_n; \\ \mathbf{L}_g[x_1(z_1(n)), x_1(z_2(n)), \dots, x_\eta(z_\eta(n))], & g \in \mathbf{J}_n. \end{cases} \quad (7)$$

Максимальная эффективность асинхронного алгоритма достигается при условии:

$$\mathfrak{R} = \{g \in \{1, 2, \dots, \eta\} \forall n \in \mathbb{N} \exists m \geq n\}: g \in \mathbf{J}_n\}. \quad (8)$$

В этом случае говорят, что хаотическая последовательность $\{\mathbf{J}_n\}_{n=1}^\infty$ имеет максимальный осадок, что обусловлено наличием минимальной последовательности множеств $\{m_g\}_{g=0}^\infty$.

Возрастающую последовательность множеств называют минимальной при условиях:

$$\begin{aligned} 1) & m_0 = 0; \quad 2) \bigcup_{i=m_g+1}^{m_{g+1}} \mathbf{J}_i = \{1, 2, \dots, n\}, \quad g=0, 1, 2, \dots; \\ 3) & \bigcup_{i=m_g+1}^{m_{g+1}-1} \mathbf{J}_i \subset \{1, 2, \dots, n\}, \quad g=0, 1, 2, \dots \end{aligned}$$

Поэтому множество $\{z_g(n)\}_{n=1}^\infty$, $g=0, 1, 2, \dots, \eta$ для асинхронного алгоритма (7) должно удовлетворять условиям: $z_g \leq (n-1)$, $z_i = (n) \rightarrow \infty$.

С целью упрощения алгоритма (7) формирования итерационной последовательности $\{x_g(n)\}_{n=1}^\infty$ вводят ограничение парасжимаемости на дифференциальный оператор \mathbf{L} в уравнении (4). Условием парасжимаемости данного оператора является удовлетворение его неравенству:

$$\|\mathbf{LX} - \Phi\| < \|\mathbf{X} - \Phi\|. \quad (9)$$

В этом случае формирование итерационной последовательности $\{x_g(n)\}_{n=1}^{\infty}$ будет происходить по правилу:

$$x_g(n+1) = L_g(x_g(z_1^g(n)), \dots, x_g(s_{k_g}^g(n))), \quad (10)$$

где $\{g(n) | n = 1, 2, \dots\}$ – последовательность элементов $g(n) \in H$; $Z^g = \{z_{k_g}^g(n)\}_{n=0}^{\infty}$, $g = 1, \dots, H$, $k_g = 1, \dots, k$ – множество целых неотрицательных чисел, удовлетворяющих условиям:

$$\begin{cases} 0 < z_{k_g}^g(n) \leq n, & n > 0, \\ -M \leq z_{k_g}^g(n) \leq 0, & n = 0. \end{cases} \quad (11)$$

Из условия (11) видно, что для стартовой итерации могут быть использованы данные от ранее стартовавших нейронов с некоторым ограничением M . Нейрон, опередивший своих соседей, может использовать устаревшие данные с глубиной запаздывания, максимальное значение которого равняется количеству итераций данного нейрона.

2 АЛГОРИТМ ТРАНЗИТНЫХ ПЕРЕСЫЛОК

В случае выбора шаблона, удовлетворяющего условию соседства (1), нейроны клеточной сети, входящие в один шаблон, имеют прямые связи, которые позволяют получать необходимые данные для вычислений по итерационной формуле (10). Однако существует большая группа задач, не позволяющих получить эффективное решение на локальном шаблоне. Для решения таких задач используют виртуальный шаблон, использующий алгоритм маршрутизации, заданный на циркулянтном графе [8].

Циркулянтным графом $C_N(\rho_1, \rho_2, \dots, \rho_m)$ будем называть регулярный граф с N вершинами и индексом окружения m , в котором каждая вершина n , $0 \leq n < N$ соединена с вершинами $(n \pm \rho_i) \bmod N, 1 \leq i \leq m$ с помощью множества соединений $\{\rho_1, \rho_2, \dots, \rho_m\}$.

Использование циркулянтных графов обусловлено возможностью формирования с их помощью оптимальных структур, характеризующихся минимальным диаметром и минимальным средним расстоянием. Для двумерной структуры с индексом окружения 4 параметры циркулянтного графа $C_N(\rho_1, \rho_2)$ будут иметь вид:

$$\rho_1 = b - 1, \rho_2 = b, b = \left\lceil \sqrt{\frac{N}{2}} \right\rceil. \quad (12)$$

Известны два изоморфных представления циркулянтных графов: хордовое кольцо [9] и мидию сети [10]. Топология мидию сетей подобна топологии клеточных нейронных сетей, поэтому может быть успешно применена для построения алгоритмов транзитных пересылок. На рис. 2 показан пример циркулянтной КНС с параметрами: $b = 4, N = 24$.

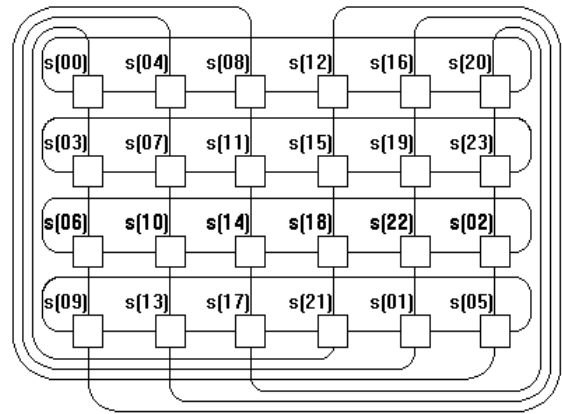


Рисунок 2 – Циркулянтная клеточная нейронная сеть

Алгоритм прокладки трассы в циркулянтной КНС основывается на использовании особенности топологии, базирующейся на том факте, что любые два нейрона обязательно входят во взаимно-пересекающиеся кольцевые структуры [11]. Пусть g – координата нейрона-источника. Тогда номера вертикального V и горизонтального P колец, к которым принадлежит данный нейрон, определяют из выражения:

$$\begin{cases} V = g \bmod (b - 1), \\ P = (b - g \bmod (b - 1)) \bmod b. \end{cases} \quad (13)$$

Кольца нейрона-источника и нейрона-приемника пересекаются в двух местах. Поэтому алгоритм прокладки трассы должен выбирать из четырех возможных вариантов маршрутов оптимальный, характеризующийся минимальным количеством транзитных пересылок. На рис. 3 показан пример оптимального маршрута, связывающего нейроны $s(10)$ и $s(15)$.

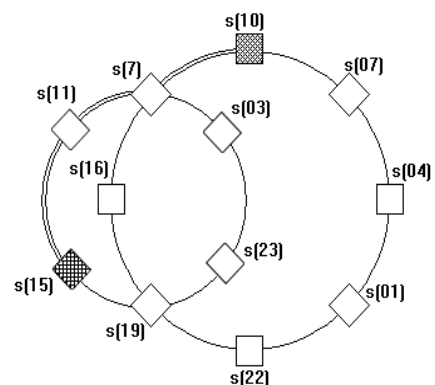


Рисунок 3 – Пример прокладки трассы

Алгоритм прокладки трассы оперирует переменными p и v , которые могут принимать только целые положительные или отрицательные значения. Длина маршрута, пройденного сообщением, определяется из выражения $D = |p| + |v|$ и равна количеству ребер, через которые должно пройти сообщение на пути от источника g

к приемнику z . Минимизацию длины маршрута D проводят с учетом условия:

$$0 \leq m = z - g = v(b-1) + pb < N. \quad (14)$$

Исходя из свойств циркулянтных графов, неравенство (14) можно ограничить сверху $0 \leq m \leq \left\lfloor \frac{N}{2} \right\rfloor$, и тем самым решение поставленной задачи свести к решению уравнения:

$$m = b(v + p) - v, \quad (15)$$

где m, p, v – целые со знаком, $b > 1$.

Пусть (p_0, v_0) – решение уравнения (15), тогда $m = p_0 b + v_0(b-1)$. Введя параметры $\alpha = \left\lfloor \frac{m}{b} \right\rfloor$, $\beta = m \bmod b$, представим решение уравнения (15) в форме:

$$\begin{cases} p_0 = \alpha + b, \\ v_0 = -\beta. \end{cases} \quad (16)$$

Поскольку любое другое решение уравнение (15) может быть представлено в виде

$$m = p_n b + v_n(b-1), \quad p_n = p_0 - n(b-1), \quad v_n = v_0 + nb,$$

система уравнений (16) для любого решения примет вид:

$$\begin{cases} p_n = \alpha + \beta - n(b-1), \\ v_n = -\beta + nb, \end{cases} \quad (17)$$

где n – целое.

Алгоритм, построенный на использовании описанных особенностей циркулянтных графов, позволяет получить кратчайшую (p_s, v_s) и (p_{ns}, v_{ns}) альтернативную трассу взаимодействия между двумя нейронами клеточной сети.

Procedure Transit (g, z)

```
begin
  If  $z \leq g$  then begin  $m := g - s; s := -1$ ; end else
  begin  $m := s - g; s := 1$ ; end;
  If  $m > N \text{ div } 2$  then
  begin  $s := -s; m := N - m$ ; end;
   $v := -(m \bmod b); p := (m \text{ div } b) - v$ ;
   $vv := b + v; pp := p - (b - 1)$ ;
  begin  $v_s := vv; v_{ns} := v; p_s := pp; p_{ns} := p; y_s := y$ ; end;
   $v_s := v_s * s; v_{ns} := v_{ns} * s$ ;
   $p_s := p_s * s; p_{ns} := p_{ns} * s$ ;
end .
```

Использование процедуры **Transit** в каждой клетке нейронной сети дает возможность определить оптимальное направление передачи сообщения, реализуя тем самым локальную компоненту глобального механизма поиска кратчайшего пути. Наличие альтернативного пути передачи является основой для разработки различных механизмов повышения живучести КНС.

ЗАКЛЮЧЕНИЕ

Основные достижения в области построения и исследования искусственных нейронных сетей базируются на использовании классической модели формального нейрона, что существенно ограничивает сферу их применения и затрудняет дальнейшее развитие. В работе сделана попытка расширить вычислительные возможности нейронных клеточных сетей путем применения математического аппарата численных методов. С целью получения возможности использования шаблона, не совпадающего со структурой связей КНС (виртуального шаблона), предложен алгоритм транзитных пересылок, базирующийся на использовании свойств циркулянтных графов.

ПЕРЕЧЕНЬ ССЫЛОК

1. McCulloch W.S., Pitts W. A logical calculus of the ideas immanent in nervous activity // Bulletin of Mathematical Biophysics. – 1943. – № 5. – pp. 115–133.
2. Shih F. Y., Moh J., Chang F. A new ART-based neural architecture for pattern classification and image enhancement without prior knowledge // Pattern Recognition. – 1992. – vol. 25, №. 5. – pp. 533–542.
3. Martinetz T.M., Berkovich S.G., Schulten K.J. "Neural-Gas" Network for Vector Quantization and its Application to the Time-Series Prediction // IEEE Transactions on Neural Networks. – 1993 – vol. 4, №4 – pp. 558–569.
4. Chua L.O., Yang L. Cellular Neural Networks. Theory. // IEEE Transactions on Circuits and Systems. – 1988. – vol. 35, № 10. – pp. 1257–1272.
5. Chua L. O. CNN: A Paradigm for Complexity. – World Scientific, Singapore, 1998. 380 p.
6. Самарский А.А., Николаев Е.С. Методы решения сеточных уравнений. – М.:Наука,1978.–592 с.
7. Новотарский М.А., Нестеренко Б.Б. Штучні нейронні мережі: обчислення.–К.: Ін-т математики НАН України, 2004. – 404 с.
8. Boesh F.T., Wang J. Reliable circulant networks with minimum transmission delay // IEEE Trans. Cir. And Sys.– 1984. – CAS-32. – pp. 1286–1291.
9. Bermond J.C., Comellas F., Hsu D.F. Distributed Loop Computer Network : A Survey // Journal of Parallel and Distributed Computing. – 1995. – vol. 24. – pp. 2–10.
10. Puente V., Izu C., Gregorio G.A., Bevide R., Pallezo J.M., Vallejo F. Rearranging Links to Improve the Performance of Parallel Computers: The Case of Midimew Networks // Proc. Of International Conference on Supercomputing, ICS'2000. – 2000. – pp. 44–53.
11. Нестеренко Б.Б., Новотарский М.А. Мультипроцессорные системы. – Киев: Ин-т математики. – 1995. – 408 с.

Надійшла 30.01.2004
Після доробки 22.06.2004

В роботі дано обґрунтування актуальності створення більш складних, у порівнянні з класичними, структур нейронів, які могли б працювати з дискретним представленням інформації. Розглянуті принципи функціонування кліткових нейронних мереж у неперервному та дискретному часі. Описана динаміка функціонування дискретної кліткової мережі, яка орієнтована на розв'язування операторних рівнянь чисельними методами. Запропоновано механізм оптимальних транзитних пересылок для реалізації ітеративних обчислень на віртуальному шаблоні.

In paper the justification of urgency for making more complicated, in comparison with classical, structures of neurons working with digital representation of the information is given. The principles of operation of cellular neural networks in continuous and discrete time are construed. The dynamic of operation of a discrete cellular network oriented on a solution of the statement equations by the numerical methods is described. The mechanism of optimum transit transmissions for realization of iterative evaluations on a virtual template is offered.

Е. А. Самойлин

КОМПЛЕКСНАЯ ОЦЕНКА АЛГОРИТМОВ ФИЛЬТРАЦИИ-СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ

Для математически строгой оценки эффективности различных методов и алгоритмов фильтрации-сегментации при обработке изображений, предлагается использовать универсальный комплексный целевой функционал, учитывающий различные факторы возникновения шумов и формы описания изображений.

ВВЕДЕНИЕ

Как известно, обработка изображений в различных информационных системах осуществляется в два основных этапа: предварительная обработка и непосредственно распознавание. На предварительном этапе обработки решаются такие задачи как фильтрация, т. е. подавление шумов, помех и сегментация, т. е. выделение контурных линий и информативных признаков.

К настоящему времени разработано достаточно много алгоритмов и методов фильтрации-сегментации изображений [1–11]. Такое их многообразие обусловлено различными факторами возникновения шумов и помех, а также формами представления изображений. Например, известно [2, 3], что для фильтрации белого гауссовского шума наиболее подходит линейная или нелинейная медианная фильтрация, для подавления низкочастотной составляющей шума эффективны спектральные методы и вейвлет-преобразования [2], для фильтрации высокочастотной составляющей – цифровые фильтры [1], фильтрации фона – согласованные и винеровские фильтры [4] и т. д.

Анализ последних работ в области фильтрации и сегментации изображений [1–10] показывает, что модификации различных алгоритмов вызывают появление новых свойств фильтров, которые по-разному влияют на подавление шумов и помех разных факторов. Например, проведенный в работе [5] вычислительный эксперимент показывает, что алгоритм медианной фильтрации, функционирующий в режиме рекурсивного формирования апертуры, обладает существенно более выраженной способностью к сглаживанию низкочастотной составляющей, т. е. приближается к спектральным методам.

С другой стороны, при всем многообразии новых алгоритмов и методов фильтрации-сегментации, возникает проблема выбора решения для конкретной задачи, так как в большинстве случаев алгоритмы моделируются на простейших тестовых изображениях и оценка их проводится визуально, т. е. «на глаз». Например, в [5] представлены результаты модифицированной медианной фильтрации на примере тестовых геометрических фигур, пятен произвольной формы, а также карт облачности, но нет строгой количественной оценки полученных результатов. В работе [6] приведены результаты модели-

рования пространственной фильтрации-сегментации без их математической оценки. В [7] также графически приведены результаты фильтрации-сегментации простейших тестовых изображений однослойной нейронной сетью без численного анализа. Строгая математическая оценка результатов обработки по предлагаемым алгоритмам отсутствует и в [8–10].

Таким образом, можно сделать вывод о том, что в настоящее время, при всём многообразии новых методов и алгоритмов фильтрации-сегментации изображений, отсутствует строгий математический аппарат оценки их эффективности.

Цель работы – обоснование необходимости введения и выбор универсального целевого функционала для строгой математической оценки эффективности алгоритмов фильтрации и сегментации изображений.

1 ПОСТАНОВКА ЗАДАЧИ

Рассматривается задача поиска комплексного целевого функционала для расчета многокритериального показателя, объективно отражающего качество принципиально различных методов и алгоритмов фильтрации-сегментации изображений.

Количество критериев оценки должно соответствовать факторам возникновения шумов и помех, которые перечислены выше, а также основным формам представления и описания изображений. Наиболее общими способами описания являются векторно-матричное представление и представление с использованием детерминированных или стохастических функций, а также на их основе структурное, признаковое и множественное представления [11].

Качество различных методов и алгоритмов целесообразно оценивать мерой близости (схожести) вектора изображения в k -мерном пространстве после фильтрации-сегментации с некоторым идеализированным (эталонным) вектором, который должен быть получен в результате «идеального» алгоритма. Размерность пространства $k = m \times n$, где m, n – соответственно число строк и столбцов изображения. Эталонный вектор изображения должен содержать только компоненты, относящиеся к контурным линиям (идеальная сегментация) и не содержать шумов (идеальная фильтрация).

2 ВЫБОР КОМПЛЕКСНОГО ЦЕЛЕВОГО ФУНКЦИОНАЛА

Как было отмечено, шумы и помехи на изображениях могут быть отнесены к одному из факторов возникно-

вения, кроме того, сами изображения характеризуются многообразием форм их представления и описания. Например, в работе [11] предложено классифицировать методы обработки изображений в соответствии с законом описания их характеристик.

Исходя из этого, искомый комплексный целевой функционал может быть записан в виде матрицы определённой размерности, число элементов вектор-столбцов которой будут соответствовать количеству оцениваемых методов обработки, а число элементов вектор-строк – количеству целевых функций (для расчета отдельных критериев) в зависимости от факторов возникновения шумов и способов описания изображений.

Оценить эффективность фильтрации-сегментации в условиях белого гауссовского шума, для векторно-матричного представления изображений, можно путем определения дисперсии многомерных векторов изображений в k -мерном пространстве [3]. При этом вектор-столбец (транспонированная вектор-строка) дисперсии векторов изображений будет иметь вид

$$F^D_{\langle 0, 1, \dots, N \rangle} = |F^D_0 F^D_1 \dots F^D_N\rangle, \quad (1)$$

$$\text{где } F^D_0 = \frac{\mathbf{M} \left\{ \sum_{i=1}^m \sum_{j=1}^n (I_{m,n} - R_{m,n})^2 \right\}}{\mathbf{M} \left\{ \sum_{i=1}^m \sum_{j=1}^n (I_{m,n} - R_{m,n})^2 \right\}} - \text{нормированная}$$

$$\text{на себя дисперсия исходного изображения с эталонным (равна единице); } F^D_1 = \frac{\mathbf{M} \left\{ \sum_{i=1}^m \sum_{j=1}^n (P^1_{m,n} - R_{m,n})^2 \right\}}{\mathbf{M} \left\{ \sum_{i=1}^m \sum_{j=1}^n (I_{m,n} - R_{m,n})^2 \right\}} -$$

нормированная дисперсия изображения, полученного обработкой исходного по первому из оцениваемых ме-

$$\text{тодов; } F^D_N = \frac{\mathbf{M} \left\{ \sum_{i=1}^m \sum_{j=1}^n (P^N_{m,n} - R_{m,n})^2 \right\}}{\mathbf{M} \left\{ \sum_{i=1}^m \sum_{j=1}^n (I_{m,n} - R_{m,n})^2 \right\}} - \text{нормированная}$$

дисперсия изображения, обработанного последним (N -м) из оцениваемых методов; T – оператор транспонирования; $I_{m,n}$ – исходное изображение (до обработки) размерности $m \times n$; $R_{m,n}$ – эталонное изображение; $P^1_{m,n}$ – изображение после обработки первым оцениваемым методом; $P^N_{m,n}$ – изображение после обработки N -м методом. Эталонный вектор изображения $R_{m,n}$ является требуемым, т. е. который должен быть получен в результате работы «идеального» метода.

Для оценки в условиях фильтрации низкочастотного шума, которая осуществляется спектральными методами, приведением к нулю постоянной составляющей в ограниченном скользящем окне, либо вейвлет-преобразованием [2] и детерминированным представлением изо-

бражений, оптимальным будет определение величины расстояния в k -мерном пространстве непрерывных функций с квадратичной метрикой. При этом второй вектор-столбец (строка) функционала примет вид:

$$F^S_{\langle 0, 1, \dots, N \rangle} = |F^S_0 F^S_1 \dots F^S_N\rangle^T, \quad (2)$$

$$\text{где } F^S_0 = \frac{\left(\int_0^m \int_0^n (I(i, j) - R(i, j))^2 didj \right)^{\frac{1}{2}}}{\left(\int_0^m \int_0^n (I(i, j) - R(i, j))^2 didj \right)^{\frac{1}{2}}} - \text{нормированное}$$

на себя расстояние между исходным и эталонным k -мерными векторами изображений (единица); $F^S_1 =$

$$= \frac{\left(\int_0^m \int_0^n (P^1(i, j) - R(i, j))^2 didj \right)^{\frac{1}{2}}}{\left(\int_0^m \int_0^n (I(i, j) - R(i, j))^2 didj \right)^{\frac{1}{2}}} - \text{нормированное рас-}$$

стояние изображения, полученного обработкой исходного по первому из оцениваемых методов от эталонного;

$$F^S_N = \frac{\left(\int_0^m \int_0^n (P^N(i, j) - R(i, j))^2 didj \right)^{\frac{1}{2}}}{\left(\int_0^m \int_0^n (I(i, j) - R(i, j))^2 didj \right)^{\frac{1}{2}}} - \text{нормированное}$$

расстояние изображения, обработанного последним (N -м) методом от эталонного; $I(i, j)$ – непрерывная функция исходного изображения; $R(i, j)$ – эталонного; $P^1(i, j)$, $P^N(i, j)$ – непрерывные функции после преобразования 1-м и N -м алгоритмом.

Фильтрация-сегментация при высокочастотном шуме, которая осуществляется цифровыми фильтрами [2], как правило, в векторно-матричной форме изображений, оценима нахождением величины евклидова расстояния и угла между интересующими векторами в k -мерном пространстве. Вектор-столбец (строка) евклидовых расстояний будет [12]

$$F^E_{\langle 0, 1, \dots, N \rangle} = |F^E_0 F^E_1 \dots F^E_N\rangle^T, \quad (3)$$

$$\text{где } F^E_0 = \frac{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (I_{m,n} - R_{m,n})^2}}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (I_{m,n} - R_{m,n})^2}} - \text{нормированное на}$$

себя евклидово расстояние между исходным и эталонным k -мерными векторами изображений (единица);

$$F_1^E = \frac{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (P_{m,n}^1 - R_{m,n})^2}}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (I_{m,n} - R_{m,n})^2}} - \text{нормированное евклидово}$$

расстояние изображения, полученного обработкой исходного по первому из оцениваемых алгоритмов

$$F_0^E = \frac{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (P_{m,n}^N - R_{m,n})^2}}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (I_{m,n} - R_{m,n})^2}} - \text{нормированное евклидово}$$

расстояние изображения, обработанного последним (N-м) методом от эталонного.

Вектор-столбец (строка) угла между векторами примет вид [12]:

$$F_{\langle 0,1,\dots,N \rangle}^\theta = \left[F_0^\theta \ F_1^\theta \ \dots \ F_N^\theta \right]^T, \quad (4)$$

$$F_0^\theta = \sqrt{1 - \left(\frac{\sum_{i=1}^m \sum_{j=1}^n (I_{m,n} R_{m,n})}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (I_{m,n})^2} \sqrt{\sum_{i=1}^m \sum_{j=1}^n (R_{m,n})^2}} \right)^2} - \text{нормированный}$$

на себя синус угла между исходным и эталонным k-мерными векторами изображений (единица);

$$F_1^\theta = \sqrt{1 - \left(\frac{\sum_{i=1}^m \sum_{j=1}^n (P_{m,n}^1 R_{m,n})}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (P_{m,n}^1)^2} \sqrt{\sum_{i=1}^m \sum_{j=1}^n (R_{m,n})^2}} \right)^2} - \text{нормированный}$$

синус угла между вектором изображения, полученным обработкой исходного по первому из оцениваемых методов и эталонным; $F_N^\theta =$

$$= \sqrt{1 - \left(\frac{\sum_{i=1}^m \sum_{j=1}^n (P_{m,n}^N R_{m,n})}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (P_{m,n}^N)^2} \sqrt{\sum_{i=1}^m \sum_{j=1}^n (R_{m,n})^2}} \right)^2} - \text{нормированный}$$

синус угла между изображением, обработанным последним (N-м) методом и эталонным.

Оценка алгоритмов при фильтрации текстурных помех и представлении изображений стохастическими функциями, осуществима определением степени тесноты линейной зависимости между функциями, что возможно путем вычисления коэффициента взаимной корреляции случайных векторов. Вектор-столбец (строка) будет определяться в соответствии с выражением:

$$F_{\langle 0,1,\dots,N \rangle}^K = \left[F_0^K \ F_1^K \ \dots \ F_N^K \right]^T, \quad (5)$$

где $F_0^K = 1 - (K[I_{m,n} R_{m,n}])$ – обратная взаимная корреляция исходного с эталонным вектором; $F_1^K = 1 - (K[P_{m,n}^1 R_{m,n}])$ – обратная взаимная корреляция вектора на выходе первого алгоритма с эталонным; $F_N^K = 1 - (K[P_{m,n}^N R_{m,n}])$ – обратная взаимная корреляция изображения, обработанного последним (N-м) методом с эталонным.

Таким образом, полный вид комплексного целевого функционала будет:

$$F_{\langle 0,1,\dots,N \rangle}^{(D,S,E,\theta,K)} = \begin{bmatrix} F_0^D & F_0^S & F_0^E & F_0^\theta & F_0^K \\ F_1^D & F_1^S & F_1^E & F_1^\theta & F_1^K \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ F_N^D & F_N^S & F_N^E & F_N^\theta & F_N^K \end{bmatrix}. \quad (6)$$

3 ОЦЕНКА АЛГОРИТМОВ ФИЛЬТРАЦИИ-СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ

В качестве примера, с использованием выбранного комплексного целевого функционала, оценена эффективность двух различных алгоритмов фильтрации-сегментации полутоновых изображений, т. е. $N = 2$. Первый является классической медианной фильтрацией, которая достаточно эффективна при сглаживании белого гауссовского шума [3]. Второй базируется на применении однослойной нейронной сети, сущность которой изложена в [7].

Указанными алгоритмами производилась обработка исходного тестового изображения размерностью 150×200 элементов, показанного на рис. 1а, с «зашумленной» буквой «А».

Эталонным изображением для тестового (рис. 1а) будет являться показанное на рис. 1б, той же размерности.

Изображение, полученное обработкой по алгоритму медианной фильтрации [3], показано на рис. 2а, а изображение, полученное на выходе однослойной нейронной сети [7] – на рис. 2б.



Рисунок 1 – Исходное (а) и соответствующее ему эталонное (б) изображения

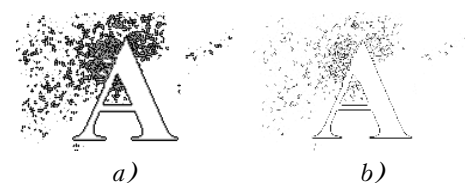


Рисунок 2 – Результат медианной (а) и нейросетевой (б) обработок

Для оценки эффективности алгоритмов, в соответствии с выражениями (1–5), проведен расчет комплексного целевого функционала. Результаты приведены в форме (6) ниже:

$$F_{(0.1.2)}^{(D,S,E,\theta,K)} = \begin{vmatrix} 1 & 1 & 1 & 0.662 & 0.79 \\ 0.914 & 0.85 & 0.933 & 0.868 & 0.71 \\ 3.966 \times 10^{-4} & 0.129 & 0.431 & 0.562 & 0.669 \end{vmatrix}.$$

Первая вектор-строка матрицы представляет собой нормированные на себя критериальные показатели (1–5), т. е. несоответствие исходного изображения (рис. 1а) с эталонным (рис. 1б); вторая – критериальные показатели ошибки медианной фильтрации-сегментации [3] исходного изображения (рис. 2а); третья – показатели ошибки нейросетевого алгоритма [7] фильтрации-сегментации (рис. 2б). Для удобства анализа результатов оценки двух алгоритмов, представим матрицу его значений графически (рис. 3).

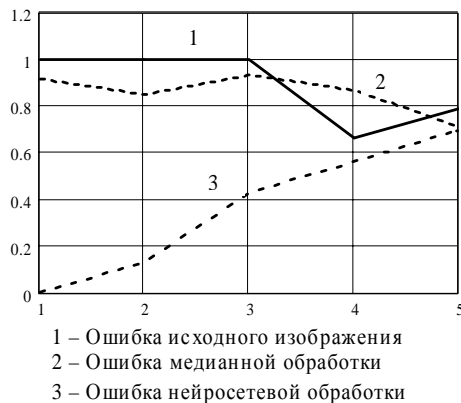


Рисунок 3 – Зависимость ошибки двух алгоритмов от критериев оценки

На рис. 3 по оси аргументов 1 соответствует дисперсии, т. е. выражению (1); 2 – расстоянию в пространстве непрерывных квадратичных функций (2); 3 – евклидову расстоянию (3); 4 – синусу угла между векторами (4); 5 – обратной взаимной корреляции векторов в соответствии с (5). Из рис. 3 видно, что различные критерии оценки двух алгоритмов дают разные их величины. Например, в первом случае, при медианной фильтрации-сегментации, происходит незначительное снижение дисперсии (с 1 до 0.914) а при нейросетевой – спад практически до нуля, т. е. кажущаяся «идеальность» последней. Расстояние между векторами в пространстве квадратичных функций после преобразования медианой, снижается (до 0.85), а после нейронной сети – до значения 0.129 (рис. 3). Расхождение векторов в евклидовом пространстве для медианной обработки практически соответствует дисперсии, а для нейронной сети – уже 0.431. Интересным является зависимость угла между обрабатываемыми векторами в k -мерном пространстве от алгоритмов (4). Если предыдущие критерии обеспечивали некоторое снижение ошибки после медианной

обработки, и более сильное – после нейросетевой, то данный показатель увеличивает ошибку (по сравнению с нормированной) при обработке медианой, и уменьшение при нейросетевой (рис. 3). Согласно расчетам коэффициента взаимной корреляции векторов по выражению (5), эффективность двух алгоритмов обработки изображений является практически одинаковой, при этом ошибка снижается незначительно (рис. 3).

Таким образом, в соответствии с выбранными критериями комплексного целевого функционала, можно сделать вывод о том, что фильтрация белого гауссовского шума и фона в совокупности с сегментацией, успешно реализуется медианным фильтром, и особенно эффективно нейронной сетью. Низкочастотный шум несколько более успешно фильтруется медианой и менее эффективно нейронной сетью по сравнению с белым; а высокочастотный – увеличивается после медианы. Текстурные помехи сравниваемые алгоритмы уменьшают практически одинаково незначительно.

ЗАКЛЮЧЕНИЕ

Предложенный универсальный комплексный целевой функционал позволяет проводить строгую многокритериальную оценку эффективности самых различных алгоритмов и методов фильтрации-сегментации изображений. Согласно комплексной математической оценке, даже единственный алгоритм фильтрации-сегментации характеризуется различной эффективностью по отношению к факторам возникновения шумов и формам представления изображений.

ПЕРЕЧЕНЬ ССЫЛОК

1. Введение в контурный анализ и его приложения к обработке изображений и сигналов. / Под ред. Я.А. Фурмана. – М.: Физматлит, 2002. – 592 с.
2. Панин С.В., Шакиров И.В., Сырямкин В.И., Светлаков А.А. Применение вейвлет-анализа изображений поверхности для изучения процессов пластической деформации и разрушения на мезомасштабном уровне // Автометрия. 2003. Том 39. № 1. С. 37–53.
3. T. S. Huang (ed.), J. – O. Eklundh, B. I. Justusson, H. J. Nussbaumer, S. G. Tyan, S. Zohar: Two-Dimensional Digital Signal Processing II: Transforms and Median Filters (Springer-Verlag, Berlin, Heidelberg, New York 1981).
4. Касперович А.Н. Фильтрация, обнаружение и локализация малоразмерных объектов, полученных при круговом субпиксельном сканировании изображений // Автометрия. 2002. Том 38. № 5. С. 17–29.
5. Воскобойников Ю.Е., Колкер А.Б. Адаптивный алгоритм фильтрации и преобразование изображений в векторный формат // Автометрия. 2002. Том 38. № 4. С. 3–11.
6. Ключко В.К., Ермаков А.А. Алгоритмы фильтрации и сегментации трёхмерных радиолокационных изображений поверхности // Автометрия. 2002. Том 38. № 4. С. 41–48.
7. Серпенинов О.В., Самойлин Е.А. Предварительная нейрообработка изображений в условиях шумов // Радиоэлектроника. Информатика. Управление. 2003. № 1. С. 111–114.
8. Бессарабов И.И. Использование ориентационной сегментации контуров с целью описания объектов на изображении в терминах отрезков // Автометрия. 2002. № 2. С. 67–71.
9. Злобин В.К., Еремеев В.В., Васильев В.М. Стохастическая модель спутниковых изображений и ее исполь-

зование для сегментации природных объектов // Автометрия. 2001. № 2. С. 41–50.

10. Васюков В.Н., Голещихин Д.В. Восстановление полутоновых изображений на основе гиббсовского описания // Автометрия. 2002. № 2. С. 58–66.
11. Соломатин И.И. Обзор методов предварительной обработки тоновых изображений для распознавания искусственных объектов // Изв. Вузов. Приборостроение. 1996. Т. 39. № 7. С. 5–12.
12. Куринной Г.Ч. Математика: Справочник. – Харьков: Фолио; Ростов н/Д: Феникс, 1997. – 463 с.

Надійшла 10.11.2003
Після доробки 21.05.2004

Для математично точної оцінки ефективності різних методів та алгоритмів фільтрації-сегментації при обробці зображень пропонується використовувати універсальний комплексний цільовий функціонал, що враховує різні фактори виникнення шумів і форми опису зображень.

For mathematical of a strict estimation of efficiency of various methods and algorithms of a filtration-segmentation at processing the images, it is offered to use universal complex target function, taking into account the various factors of occurrence of noise and form of the description of the images.

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

ПРОГРЕССИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

PROGRESSIVE INFORMATION TECHNOLOGIES

УДК 681.3.016:681.324

А. А. Завалин, А. А. Зенькович

АВТОМАТИЗИРОВАННОЕ ФОРМИРОВАНИЕ ОБЪЕКТНОГО СЛОВАРЯ РЕЛЯЦИОННОЙ БАЗЫ ДАННЫХ

В статье рассматриваются вопросы практической реализации методики формирования объектного словаря реляционной базы данных. Предложен способ автоматизированного заполнения и поддержки актуальности объектного словаря, подход к именованию объектов и атрибутов, а также приложение для получения справочной информации по объектному представлению базы данных.

ВВЕДЕНИЕ

В настоящее время в большинстве корпоративных информационных систем (ИС) для хранения и обработки больших объемов информации используются реляционные базы данных (РБД). В процессе эксплуатации развивающихся ИС количество таблиц РБД, связей между ними и количество полей в таблицах значительно возрастает. Если в развитии ИС участвует множество разработчиков, то заметно усложняется задача документирования всех изменений в ее структуре. Кроме того, названия хранимых в РБД таблиц и полей часто не соответствуют терминологии предметной области, которую она описывает. Следствием этого является усложнение процесса получения актуальной справочной информации по структуре РБД и извлечения данных из нее.

Для преодоления этих проблем и повышения эффективности работы пользователей и прикладных программистов с РБД предлагается использовать объектное представление, которое отражает структуру РБД, имеет иерархический вид, для объектов и их атрибутов используются общепринятые названия, а несущественные для пользователей данные скрыты [1]. Данные, требующи-

еся для работы с РБД посредством объектного представления и описывающие соответствие между таблицами и объектами, предлагается разместить в объектном словаре данных (ОСД).

Для практической реализации методики формирования объектного представления [2] необходимо рассмотреть процедуры заполнения и обновления ОСД, выработать правила задания названий для объектов и атрибутов, а также рассмотреть некоторые технические аспекты разработки приложения, поддерживающего решение этих задач.

Перед использованием объектного представления ОСД должен быть инициализирован. Эта процедура выполняется однократно и может выполнена в несколько этапов.

ПОЛУЧЕНИЕ МЕТАДАННЫХ РЕЛЯЦИОННОЙ БАЗЫ ДАННЫХ

Исходными данными для процедуры формирования ОСД являются сведения о таблицах, полях и внешних ключах таблиц, т. е. метаданные РБД. Каждое поле – внешний ключ определяет связь между двумя таблицами: первая содержит этот внешний ключ, а вторая содержит первичный ключ, на который ссылается внешний ключ. Метаданные РБД в ОСД могут быть описаны множествами элементов со следующей структурой:

- «Таблица»: название таблицы;
- «Поле таблицы»: название поля; ссылка на таблицу, к которой относится данное поле; тип данных;

– «Связь двух таблиц T_i и T_j »: ссылка на таблицу T_i ; ссылка на поле – внешний ключ в таблице T_i ; ссылка на таблицу T_j ; ссылка на поле – первичный ключ в таблице T_j .

В приложении метаданные РБД можно получить автоматически, используя:

- прямые запросы к системным каталогам системы управления базами данных (СУБД);
- запросы к виртуальным таблицам, представляющим системные каталоги СУБД, в системной схеме INFORMATION_SCHEMA, описанной стандарте SQL92 [3];
- функции JDBC- или ODBC-драйверов [4].

Реализация первого способа приводит к необходимости написания различных функций доступа для различных СУБД, поскольку каждая из них имеет собственные форматы данных и методы доступа к системным каталогам. Два последних способа является предпочтительными, т. к. они обеспечивают унифицированный доступ к метаданным РБД для различных СУБД. Использование специальных функций в JDBC- или ODBC-драйверах позволяет избежать прямого обращения к СУБД благодаря внутренней реализации низкоуровневых функций доступа к системным каталогам и предоставлению единого программного интерфейса для получения метаданных РБД, при этом для использования приложения совместно с другой СУБД достаточно заменить соответствующий драйвер.

Если применяемая СУБД не предоставляет средств для получения метаданных РБД, то их требуется ввести в ОСД вручную.

Результатом выполнения этого этапа являются заполненные множества элементов ОСД, описывающие структуру таблиц РБД и связи между ними.

ФОРМИРОВАНИЕ ОБЪЕКТОВ

Каждый объект соответствует некоторой таблице в РБД. Все объекты подразделяются на два типа [2]: основные – находятся на верхнем уровне иерархии объектов и являются самостоятельными, и вспомогательные – являются составной частью других объектов. Объект включает в себя множество атрибутов, которые подразделяются на две группы: собственные атрибуты, ссылающиеся на поля в таблицах, связанных с объектами, и несобственные атрибуты. Множество несобственных атрибутов, в свою очередь, состоит из атрибутов-подобъектов и атрибутов-указателей на объекты. Атрибут-подобъект связывает объект, которому он принадлежит, со вспомогательным объектом, включая все его атрибуты в состав объекта. Атрибут-указатель связывает объект, которому он принадлежит, с некоторым основным объектом, при этом поля связанного основного объекта не включаются в состав данного объекта.

Для каждого элемента, описывающего таблицу РБД, в ОСД создается элемент, описывающий объект и имеющий то же название, что и у таблицы, и в нем устанавливается ссылка на соответствующую таблицу.

Для каждого элемента ОСД, описывающего поле в таблице, создается элемент, описывающий собственный атрибут объекта, для него устанавливается то же название и тип данных, что и у поля в таблице, и устанавливается ссылка на него.

Для описания атрибутов-подобъектов и атрибутов-указателей на объекты в ОСД вводится множество элементов, каждый из которых определяет соотношение между первым и вторым объектом в каждой паре связанных объектов. Алгоритм, который позволяет автоматически установить типы объектов и связи между ними, основан на анализе связей между таблицами и их последовательной «свертке», и изложен в [2].

Для описания объектов и соотношений между ними в ОСД могут быть использованы множества элементов со следующей структурой:

- «Объект»: название объекта; синонимы названия; ссылка на таблицу, соответствующую объекту; тип объекта (основной, вспомогательный); видимость объекта (видимый, скрытый);
- «Атрибут объекта»: название атрибута; синонимы названия; ссылка на объект, которому принадлежит атрибут; тип данных; видимость атрибута (видимый, скрытый);
- «Связь объектов O_i и O_j »: ссылка на объект O_i ; ссылка на объект O_j ; отношение объекта O_i к объекту O_j (подобъект, указатель на объект).

Результатом выполнения данного этапа являются заполненные множества элементов ОСД, описывающие объекты и связи между ними. При их совместном использовании объектное представление РБД для пользователя можно отобразить в виде дерева объектов.

ИМЕНОВАНИЕ ОБЪЕКТОВ И АТТРИБУТОВ. СИНОНИМЫ НАЗВАНИЙ

Названия таблиц и полей в таблицах определяются разработчиками на этапе проектирования РБД. Как правило, они выбираются достаточно краткими с целью уменьшения размеров SQL-запросов к РБД. Некоторые СУБД не позволяют использовать в названиях таблиц и полей символы различных национальных алфавитов и существенно ограничивают их максимальную длину. Если в создании или модернизации отдельных фрагментов крупных РБД участвуют различные разработчики, то они могут руководствоваться собственным подходом к формированию названий. Для пользователей и прикладных программистов, не имеющих продолжительного опыта работы с РБД, все эти факторы существенно затрудняют понимание сущности данных, хранящихся в таблицах, по их названиям.

С целью облегчения понимания структуры РБД предлагается для всех объектов и их атрибутов в объектном представлении задавать названия, которые являются общепринятыми в той предметной области, которую описывает данная РБД. Для выбора названий используются термины предметной области, описывающие сущности реального мира и хорошо знакомые пользователям. Это в большинстве случаев позволяет

интуитивно использовать информацию в РБД и существенно снизить необходимость детального изучения документации к ней.

Процедура задания имен для объектов и атрибутов в объектном представлении проводится администратором РБД, хорошо знакомым с ее структурой, путем интервьюирования и изучения должностных инструкций пользователей, являющихся специалистами в своей области деятельности и хорошо разбирающихся в соответствующей предметной области. Если РБД содержит фрагменты, описывающие несколько предметных областей, которые связаны с работой пользователей различных профессий, то консультации должны быть проведены со всеми ними с целью формирования согласованного и непротиворечивого пространства имен в объектном представлении.

В задачах пользователей и прикладных программистов, для решения которых требуется работа с РБД, встречаются несколько характерных ситуаций, связанных с необходимостью поиска нужного элемента (объекта или атрибута) в объектном представлении:

– известно название элемента и его расположение в иерархии объектов;

– известно название элемента и факт его присутствия в объектном представлении, но не его точное расположение;

– точное название элемента, факт его присутствия и расположение в иерархии объектов неизвестны.

В первой ситуации нужный элемент может быть обнаружен сразу. Во второй задача может быть решена путем последовательного обхода дерева объектов, начиная с его корня и по направлению к узлам более глубоких уровней, выбирая на каждом шаге наиболее подходящий по названию элемент из нескольких непосредственно связанных с текущим. В третьей ситуации нахождение требуемого элемента приводит к необходимости просмотра всех имеющихся объектов в надежде найти название, похожее на искомое, и для крупных РБД это приводит к существенным временным затратам.

Предлагается облегчить задачу поиска элемента с неизвестным точным названием путем введения множества синонимов названий для каждого объекта и атрибута. Заполнение этих множеств проводится администратором РБД при участии специалистов в необходимых предметных областях. Должно выполняться следующее ограничение: множества синонимов названий различных объектов не должны пересекаться и содержать названия существующих объектов с целью однозначной идентификации объекта по синониму его названия. Для синонимов названий атрибутов вводится требование их уникальности в пределах одного объекта, поскольку в объектном представлении атрибут однозначно идентифицируется названием объекта, к которому он принадлежит, и самим названием атрибута.

Результатом выполнения данного этапа является заполненный ОСД, в котором объекты и атрибуты имеют названия, принятые в предметной области и понятные пользователям.

ВИДИМОСТЬ ОБЪЕКТОВ И АТТРИБУТОВ

Практически в любой РБД в таблицах присутствуют поля, не несущие смысловой нагрузки для пользователей. К ним относятся: автоматически генерируемые уникальные идентификаторы записей в таблицах, используемые в качестве первичных ключей; ссылающиеся на них внешние ключи, обеспечивающие ссылочную целостность данных; поля, применяемые для отслеживания изменений в других полях и т. п. Атрибуты в объектном представлении, соответствующие таким полям, не предоставляют пользователю значимой информации. Для того, чтобы уменьшить общее количество полей в объектах и тем самым сконцентрировать внимание пользователя на существенных полях, в структуру элемента ОСД, описывающего атрибут объекта, вводится признак видимости, который может принимать значения «видимый» или «скрытый» и определяет, будет отображаться данный атрибут или нет.

В РБД могут присутствовать таблицы, используемые в ИС для решения административных или системных задач, но не требующиеся для пользователей, например, журналы запросов, журналы изменения данных в таблицах, таблицы с константами для проверки ограничений целостности при вставке и обновлении записей и другие таблицы, не имеющие прямого отношения к предметной области, которую описывает РБД. Для объектов, связанных с такими таблицами, целесообразно также ввести признак, определяющий их видимость для пользователей.

Поскольку скрытые поля и объекты могут потребоваться при составлении запросов к РБД программистами или администраторами РБД, то в приложение, отображающее структуру объектов, вводится переключатель, определяющий, будут ли принудительно отображаться скрытые элементы или нет.

Установку признака видимости для объектов и атрибутов должен выполнять администратор РБД, руководствуясь сведениями о задачах, решаемых пользователями ИС.

ОБНОВЛЕНИЕ ОБЪЕКТНОГО СЛОВАРЯ ДАННЫХ

В процессе эксплуатации развивающейся ИС структура РБД может изменяться, поэтому в ОСД требуется своевременно вносить изменения для поддержания актуальности объектного представления. Несовпадение ОСД текущей структуре РБД может привести к ошибкам, вызванным попыткой доступа к объектам или атрибутам, связанным с несуществующими или модифицированными таблицами или полями, или к недоступности данных, хранящихся в РБД, но не отраженных в объектном представлении.

Для приведения ОСД к актуальному состоянию возможна его повторная инициализация, однако, она приводит к потере данных, введенных администратором РБД (названия объектов, атрибутов, признаки их видимости, синонимы названий), и существенным затратам

времени на их повторное внесение, поэтому процесс обновления ОСД должен быть неструктурным, инкрементальным и автоматизированным, сводя необходимость вмешательства администратора РБД к минимуму.

Возможны следующие варианты практической реализации системы поддержки актуальности объектного представления РБД:

- периодическое сопоставление текущей структуры РБД с метаданными РБД в ОСД;
- отслеживание действий, приводящих к изменениям в структуре РБД.

Первый подход предполагает периодическое извлечение метаданных РБД во временную память и сравнение их с метаданными РБД в ОСД. Он может быть универсально реализован для различных СУБД, но его существенным недостатком является то, что на период времени между моментом какого-либо изменения в структуре РБД и моментом ближайшего сопоставления данных объектное представление будет рассогласовано с реальной структурой РБД, что может привести к ошибкам при доступе к данным. Для уменьшения вероятности рассогласования следует увеличивать частоту сопоставлений, а т. к. для крупных РБД извлечение метаданных и их сравнение с данными в ОСД является достаточно сложным и ресурсоемким процессом, то это может существенно замедлить работу сервера РБД.

Второй подход основан на том, что реляционные СУБД позволяют изменять структуру РБД только с помощью специальных SQL-запросов, следовательно, при обнаружении таких запросов можно выполнить их анализ и внести соответствующие изменения в ОСД. При этом вероятность возникновения ошибок, являющихся следствием рассогласования объектного представления и текущей структуры РБД, сводится к минимуму, а нагрузка на сервер РБД существенно не возрастает, поэтому данный подход является предпочтительным.

Он может быть реализован путем отслеживания в реальном времени успешно выполненных SQL-запросов к РБД, изменяющих ее структуру. К ним относятся запросы на создание, изменение и удаление таблиц и представлений, и запросы, модифицирующие внешние ключи в таблицах: CREATE TABLE (VIEW), DROP TABLE (VIEW), ALTER TABLE. Определить факт выполнения такого запроса можно путем анализа журнала запросов, имеющегося практически во всех СУБД. В случае, если передача запросов на изменение структуры РБД выполняется только через ODBC-драйвер (например, при использовании универсальных приложений для администрирования РБД), то журналирование может быть включено в настройках источника данных, обеспечивающего подключение к модифицируемой РБД. Если в аналогичном случае используется только JDBC-драйвер, то между приложением и JDBC-драйвером может быть подключен промежуточный драйвер, имеющий тот же программный интерфейс, и выполняющий журналирование и передачу запросов JDBC-драйверу.

Выявив выполнение запроса, изменяющего структуру РБД, производится его разбор с целью выделения изменившихся элементов РБД и автоматического внесения

изменений в элементы ОСД, описывающие метаданные РБД и объектное представление. При добавлении в РБД новых таблиц, полей или представлений, администратор РБД извещается о необходимости задания для них названий и синонимов.

Если используемая СУБД не поддерживает программный интерфейс для извлечения метаданных и возможность журналирования выполняемых запросов, то все изменения в ее структуре требуются вручную отражать в ОСД.

ЗАКЛЮЧЕНИЕ

Предложенный подход позволяет в автоматизированном режиме сформировать объектный словарь РБД, выполнить процедуру именования объектов и атрибутов и обеспечить поддержку его соответствия текущему состоянию РБД. Он реализован в виде приложения, для разработки которого были выбраны следующие средства: операционная система Linux; СУБД PostgreSQL [5]; язык программирования Java; JDBC-драйвер для доступа к СУБД PostgreSQL из приложений на языке Java [6]. Такой выбор обеспечивает работоспособность приложений, использующих объектное представление РБД, на различных программных и аппаратных платформах и возможность использования других СУБД без необходимости внесения изменений в приложение. Кроме того, все использованные средства распространяются свободно. Приложение позволяет решать следующие задачи: инициализация ОСД; установка названий объектов, атрибутов и синонимов с необходимыми проверками на их уникальность; установка признаков видимости для объектов и атрибутов; визуализация иерархического объектного представления РБД с возможностью поиска нужных элементов по названию с указанием произвольного сочетания из следующих областей поиска: названия объектов, названия атрибутов, синонимы названий.

Применение объектного представления РБД позволяет упростить работу пользователей и прикладных программистов благодаря представлению структуры РБД в удобной для восприятия форме, сокрытию неинформативных данных, а также использованию понятных названий объектов и атрибутов.

Дальнейшее развитие предложенного подхода возможно в направлении разработки средств автоматизированного формирования объектного представления для распределенных ИС.

ПЕРЕЧЕНЬ ССЫЛОК

1. Завалин А.А., Кунгурцев А.Б., Блажко А.А. Использование словаря данных в информационных системах с логической формой представления данных // Тр. Одес. политехн. ун-та. – Одесса, 2002. – Вып. 2 (18).
2. Кунгурцев А.Б., Завалин А.А. Методика формирования объектного представления реляционной базы данных // Тр. Одес. политехн. ун-та. – Одесса, 2004. – Вып. 1 (21).
3. Kreigel A., Trukhnov V. SQL Bible. – Chichester, England: Wiley, 2003. – 858 с.
4. Дунаев С. Доступ к базам данных и техника работы в сети. Практические приемы современного программирования. – М.: Диалог-МИФИ, 2000. – 416 с.

5. Гешвинде Э., Шениг Г.-Ю. PostgreSQL. Руководство разработчика и администратора. – СПб.: ООО "ДиаСофт-ЮП", 2002. – 608 с.
6. Хорстманн К., Корнелл Г. Java 2. Библиотека профессионала. Том 2. Тонкости программирования. – М.: Изд. дом "Вильямс", 2002. – 1120 с.

Надійшла 11.1.2003
Після доробки 02.04.2004

В статті розглянуті питання практичної реалізації методики формування об'єктного словника реляційної бази даних. Запропоновано автоматизований спосіб заповнення та підтримки актуальності об'єктного словника, підхід до

іменування об'єктів та атрибутів, а також програма, що дозволяє отримувати довідникову інформацію по об'єкту представленою бази даних.

The article reviews the questions of practical implementation of the object dictionary creation methods for relational database. The automated method of filling and keeping the up-to-date state of object dictionary, the way of naming the objects and attributes and application that provides reference information for database object presentation form are offered.

УДК 681.327.12.001.362

А. А. Каргин, А. В. Григорьев

НЕЧЕТКИЙ МЕТОД К ИДЕНТИФИКАЦИИ ЧЕЛОВЕКА ПО ФОТОПОРТРЕТУ

В публикации рассмотрена проблема поиска неизвестного человека в БД по фотоportрету. Предложены две нечеткие модели представления фото и метод поиска фотографии, основанный на нечетком сравнении с эталоном.

ВВЕДЕНИЕ

Одной из важнейших задач, возникающих в ходе оперативно-розыскной деятельности правоохранительных органов, является быстрая идентификация неизвестных лиц, являющихся организаторами, участниками или свидетелями преступлений; лиц пропавших без вести, а также неопознанных трупов. Из систем криминалистического учета, функционирующих на территории стран СНГ, можно выделить российскую АДИС «Папилон» [1] и отечественную систему информационной поддержки ОВД Украины [2]. В рамках данных систем работают соответственно подсистемы «Словесное описание» («Папилон»), «Пізнання» и «ІБД», позволяющие проводить поиск неизвестного лица в БД по словесному описанию, и поддерживающие регистрацию в БД фотоизображений, однако, ни в какой из приведенных систем не реализован модуль поиска по фотографии. В данной статье предложено два метода нечеткого автоматизированного поиска неопознанного лица в БД по фотоportрету.

ПОСТАНОВКА ЗАДАЧИ

Пусть в БД хранится N изображений человеческих лиц, сделанных по правилам опознавательного фото. По предъявленному тестовому фотоизображению система должна упорядочить изображения из БД по степени соответствия тестовому изображению и выдать в их порядке убывания. Окончательное решение при этом остается за «лицом, принимающим решение». На тестовое изображение также накладывается ограничение – оно

должно быть выполнено по правилам опознавательного фото или близким к нему (допустимы легкие повороты головы, некоторые изменения эмоционального выражения лица, и т. д.). Система должна работать в off-line, в связи с этим на систему не накладывается ограничений по времени – поиск лишь должен быть проведен за разумное время. Предполагается также, что система должна работать в автоматизированном режиме, т. е. в качестве исходных данных наряду с фотоизображениями допускается ввод координат некоторых опорных точек.

МОДЕЛЬ НЕЧЕТКОГО ПРЕДСТАВЛЕНИЯ ФОТОИЗОБРАЖЕНИЯ ЧЕЛОВЕЧЕСКОГО ЛИЦА

Предложено два нечетких метода представления фотоизображения в БД. Оба этих метода базируются на результатах, полученных в области психологии восприятия и психофизики. Известно [3], что в процессе восприятия зрительной информации зрачок совершает движения трех видов: трем (частые легкие колебания), плавное смещение и резкие скачки (саккадические переходы). Установлено, что конфигурация саккадических переходов обусловлена задачей, выполняемой человеком в процессе зрительного восприятия, при этом особый интерес человека привлекает ряд областей изображения. В дальнейшем такие области будем называть «особыми областями».

При идентификации человеческого лица наибольший интерес вызывают области передненосовой, ротовой и зрачковых антропометрических точек, а также область точки, расположенной примерно посередине спинки носа. В пользу выбора данных антропометрических точек говорят также знания в области габитоскопии [4]. Так, в области данных антропометрических точек расположен ряд информативных индивидуализирующих контуров: например, контуры глазной щели, внутреннего и внеш-

него углов глаз, век, бровей; контуры крыльев, спинки и кончика носа; носогубного филтра, носогубных складок; линии верхней и нижней каймы губ и т. д. В связи с этим, выдвинута гипотеза о том, что области указанных точек содержат достаточный для идентификации человеческого лица объем зрительной информации.

Для обеих моделей согласно постановке задачи исходные данные представлены фотоизображениями и координатами некоторых опорных точек. Фотоизображение задано на пиксельной плоскости $P = \{(i, j) | i = \overline{1, N}; j = \overline{1, M}\}$ функцией яркости $f: P \rightarrow [0, 1]$; координаты правой и левой зрачковых точек; передненосовой и ротовой антропометрических точек, а также точки «спинка носа» $p_1, p_2, p_3, p_4, p_5 \in P$ соответственно. Отметим, что координаты точек p_1, p_2, p_3, p_4 вводятся в автоматизированном режиме, а координата точки p_5 находится из соотношения $p_5 = \frac{1}{4}(p_1 + p_2) + \frac{1}{2}p_3$.

Согласно выдвинутому предположению, модель фотоизображения предложено задавать набором моделей особых областей:

$$S = \langle \Omega_1, \Omega_2, \Omega_3, \Omega_4, \Omega_5 \rangle, \quad (1)$$

где $\Omega_1, \Omega_2, \Omega_3, \Omega_4, \Omega_5$ – модели особых областей правой и левой зрачковых точек, областей передненосовой, ротовой точек и области точки «спинка носа» соответственно.

Модель каждой особой области Ω_k предложено задавать набором нечетких множеств, выражающих поведение контура в ряде контрольных точек, принадлежащих k -й особой области. Выбор контрольных точек, в свою очередь, зависит от ряда параметров особой области таких параметров как: центр особой области, задаваемый точкой $p_k \in P$; радиус области R_k и направление области Ψ_k . Последние два параметра предложено сделать зависимыми от взаимного расположения центра описываемой особой области относительно центров других особых областей, иными словами $R_k = H_k(p_1, \dots, p_5)$; $\Psi_k = \Pi_k(p_1, \dots, p_5)$, где $k = \overline{1, 5}$. Радиусы особых областей находятся из выражений: $R_1 = R_2 = \frac{1}{2}\ell_2(p_1, p_2)$, $R_3 = \frac{1}{4}(\ell_2(p_1, p_3) + \ell_2(p_2, p_3))$, $R_4 = \frac{5}{2}\ell_2(p_3, p_2)$, $R_5 = \frac{3}{2}\ell_2(p_4, p_2)$, где $\ell_2(p_i, p_j)$ – евклидово расстояние между точками p_i и p_j . Направления Ψ_k особых областей Ω_k предложено задавать углом между осью ординат и векторами \vec{a}_k , где $\vec{a}_1 = \overrightarrow{p_1 p_2}$, $\vec{a}_2 = \overrightarrow{p_2 p_1}$, $\vec{a}_3 = \overrightarrow{p_1 p_2} + \overrightarrow{p_2 p_1}$, $\vec{a}_4 = \overrightarrow{p_4 p_5}$, $\vec{a}_5 = \overrightarrow{p_5 p_4}$.

Перед описанием модели Ω_k k -й особой области удобнее перейти к полярной системе координат с центром в точке p_k , единичным радиусом R_k и нулевой полярной угловой координатой, соответствующей направлению особой области. Связь между евклидовой (i, j) и полярной (r, φ) системами координат при этом осуществ-

ляется выражениями: $i = i_k + R_k r \sin(\varphi + \Psi_k)$, $j = j_k + R_k r \cos(\varphi + \Psi_k)$. В качестве контрольных точек, поведение контура в которых составляет модель особой области Ω_k , предложено выбрать точки

$$\{(r_i, \varphi_j) | r_i = K^{1-i}; \varphi_j = 2\pi(j-1)/s, i = \overline{1, q}, j = \overline{1, s}\},$$

где q и s задают размерность модели, а $K > 1$ – коэффициент смещения радиальной координаты.

Таким образом, модель особой области представляется набором нечетких множеств, каждое из которых отражает поведение контура в окрестности некоторой контрольной точки:

$$\Omega = \langle C_{11}^{(k)}, \dots, C_{ij}^{(k)}, \dots, C_{qs}^{(k)} \rangle, \quad (2)$$

где $C_{ij}^{(k)}$ – нечеткое множество, задающее поведение контура в контрольной точке (r_i, φ_j) особой области Ω_k .

Сами нечеткие множества $C_{ij}^{(k)}$ строятся на базовом множестве $\Phi = \left\{0, \frac{2\pi}{3}, \frac{4\pi}{3}\right\}$ и имеют вид $C_{ij}^{(k)} = \{\phi_m | \mu(\phi_m)\}$,

где $\phi_m \in \Phi$, и функция принадлежности $\mu(\phi_m)$ выражает уверенность в том, что через контрольную точку (r_i, φ_j) в направлении $\varphi_j + \phi_m$ проходит линия контура.

Для описания процедуры вычисления функции принадлежности нечеткого множества $C_{ij}^{(k)}$ аналогично описанию модели особой области перейдем к более удобной системе координат с началом отсчета в контрольной точке (r_i, φ_j) , нулевым направлением φ_j и единичным отрезком радиальной координаты, соответствующим размеру ε -окрестности контрольной точки. Связь между евклидовой (i, j) и введенной полярной (ρ, ξ) системами координат выражена соотношениями: $i = i^* + \varepsilon \rho \sin(\xi + \Psi_k + \varphi_j)$, $j = j^* + \varepsilon \rho \cos(\xi + \Psi_k + \varphi_j)$, где (i^*, j^*) – координаты контрольной точки в системе координат (i, j) .

Функция принадлежности множества $C_{ij}^{(k)}$ задана усеченной сигмоидальной функцией (4) величиной $DY(\phi)$, называемой наличием контура в направлении ϕ :

$$\mu_C(\phi) = \sigma(DY_\phi); \quad (3)$$

$$\sigma(x) = \frac{2}{1 + e^{-25x}} - 1. \quad (4)$$

Геометрически понятие $DY(\phi)$ наличия контура в направлении ϕ интерпретируется следующим образом: на пиксельной плоскости задано три нечетких области, каждая из которых, с одной стороны, принадлежит области, каждый пиксель которой находится довольно близко к контрольной точке. С другой стороны, первая область, называемая «центральной», также принадлежит области, находящейся на небольшом расстоянии от условной линии, лежащей в направлении ϕ и проходящей через контрольную точку. Две другие области – принад-

лежат области, удаленной от прямой, проходящей через контрольную точку в направлении ϕ , а также одной из двух полуплоскостей, полученных разделением пиксельной плоскости прямой, задающей направление ϕ . Данные области предложено условно называть «положительной периферийной» и «отрицательной периферийной» областями.

Для формализации трех вышеперечисленных областей введены: область $\underline{\Delta} \subset P$, находящаяся довольно близко к контрольной точке; область $\underline{\Delta} \subset P$, близкая к прямой, задающей направление ϕ ; $\underline{\Theta}^+, \underline{\Theta}^- \subset P$ – положительная и отрицательная пиксельные полуплоскости (относительно прямой, задающей направление ϕ) с функциями принадлежности:

$$\mu_{\underline{\Delta}}(p) = \text{MAX}(1 - \rho^2, 0); \quad (5)$$

$$\mu_{\underline{\Delta}}(p) = \text{MAX}[1 - 3 \cdot \rho \sin(\xi - \phi)^2, 0]; \quad (6)$$

$$\mu_{\underline{\Omega}^+}(p) = \begin{cases} 1, & \text{если } \text{sign}[\sin(\xi - \phi)] > 0; \\ 0, & \text{если } \text{sign}[\sin(\xi - \phi)] \leq 0. \end{cases} \quad (7)$$

$$\mu_{\underline{\Omega}^-}(p) = \begin{cases} 1, & \text{если } \text{sign}[\sin(\xi - \phi)] > 0; \\ 0, & \text{если } \text{sign}[\sin(\xi - \phi)] \geq 0. \end{cases} \quad (8)$$

а также следующее определение:

Определение 1. Пусть задано нечеткое множество $\underline{A} \subset U$ и действительная функция $f: U \rightarrow R$. Средним значением функции f на нечетком множестве \underline{A} будем называть выражение:

$$\text{AVG}_f(\underline{A}) = \left(\int_{\underline{A}} f(x) \cdot \mu_{\underline{A}}(x) dx \right) / \left(\int_{\underline{A}} \mu_{\underline{A}}(x) dx \right). \quad (9)$$

Тогда центральная, положительная и отрицательная периферийные области задаются в виде выражений: $\underline{\Delta}^* \underline{\Delta}$, $\underline{\Delta}^* \underline{\Delta}^* \underline{\Theta}^+$ и $\underline{\Delta}^* \underline{\Delta}^* \underline{\Theta}^-$ соответственно, а величина $\text{DY}(\phi)$ представлена выражением:

$$\text{DY}(\phi) = \delta(\text{AVG}_f(\underline{\Delta}^* \underline{\Delta}), \text{AVG}_f(\underline{\Delta}^* \underline{\Delta}^* \underline{\Theta}^+)) \vee \delta(\text{AVG}_f(\underline{\Delta}^* \underline{\Delta}), \text{AVG}_f(\underline{\Delta}^* \underline{\Delta}^* \underline{\Theta}^-)), \quad (10)$$

где $\delta: [0, 1]^2 \rightarrow [0, 1]$ – некоторая метрика.

Предложено две модели, различающиеся выбором метрики δ : первую из которых предложено называть `fuzz_sub`, а вторую – `fuzz_div`. В качестве метрики δ использованы соответственно выражения:

$$\delta(x, y) = |x - y|; \quad (11)$$

$$\delta(x, y) = \begin{cases} 0, & \text{если } x = 0 \text{ и } y = 0; \\ |x - y| / (x \vee y) & \text{в противном случае.} \end{cases} \quad (12)$$

Таким образом, модели `fuzz_sub` и `fuzz_div` представляют собой набор моделей особых областей (1), каждая из которых задана множеством нечетких характеристик контура, измеряемых в ряде контрольных точек, принадлежащих особой области (2). Эти нечеткие харак-

теристики заданы на базовом множестве направлений

$$\Phi = \left\{ 0, \frac{2\pi}{3}, \frac{4\pi}{3} \right\},$$

а в качестве функции принадлежности данных нечетких множеств использовано усиленное сигмоидальной функцией (4) значение (10) наличия контура $\text{DY}(\phi)$ в направлении $\phi \in \Phi$. Различие между двумя моделями заключается в использовании метрики в выражении (10): для модели `fuzz_sub` использована метрика (11), а для `fuzz_div` – (12).

МЕТОД ПОИСКА НЕИЗВЕСТНОГО ЛИЦА В БД ПО ФОТОПОРТРЕТУ

Согласно постановке задачи, задача поиска фотоизображения в БД фактически сводится к процедуре сортировки моделей фото из БД по степени их нечеткого соответствия предъявляемому образу. Последняя процедура, очевидно, сводится к нахождению степени сходства двух нечетких моделей.

Степень сравнения нечетких моделей фотоизображений представлена AVG-критерием:

$$\delta^{(1)}(S, \hat{S}) = \frac{1}{5} \sum_{k=1}^5 \delta^{(2)}(\Omega_k, \hat{\Omega}_k), \quad (13)$$

где $\delta^{(2)}(\Omega_k, \hat{\Omega}_k)$ – степень нечеткого сравнения моделей особых областей, для вычисления которой также использован AVG-критерий:

$$\delta^{(2)}(\Omega_k, \hat{\Omega}_k) = \frac{1}{qs} \sum_{i=1}^q \sum_{j=1}^s \delta^{(3)}(\underline{C}_{ij}^k, \hat{\underline{C}}_{ij}^k), \quad (14)$$

где $\delta^{(3)}(\underline{C}_{ij}^k, \hat{\underline{C}}_{ij}^k)$ – степень сравнения моделей контрольных точек. $\delta^{(3)}(\underline{C}_{ij}^k, \hat{\underline{C}}_{ij}^k)$ предложено выражать одним из выражений:

$$\delta^{(3)}(\underline{C}_{ij}^k, \hat{\underline{C}}_{ij}^k) = 1 - \max_{\phi \in \Phi} |\mu_C(\phi) - \mu_{\hat{C}}(\phi)|; \quad (15)$$

$$\delta^{(3)}(\underline{C}_{ij}^k, \hat{\underline{C}}_{ij}^k) = \begin{cases} 1, & \text{если } \max_{\phi \in \Phi} (\mu_C(\phi) \vee \mu_{\hat{C}}(\phi)) = 0; \\ 1 - \frac{\max_{\phi \in \Phi} |\mu_C(\phi) \vee \mu_{\hat{C}}(\phi)|}{\max_{\phi \in \Phi} (\mu_C(\phi) \vee \mu_{\hat{C}}(\phi))} & \text{в противном случае.} \end{cases} \quad (16)$$

Установлено, что при сравнении нечетких моделей `fuzz_div` более эффективно применять выражение (15), а для `fuzz_sub` – выражение (16). Также установлено, что оптимальные параметры моделей: $q = 22$; $s = 32$; $K = 1,05$; $\varepsilon = 4$ пикселя при размере фотоизображений 256×384 пикселя. Для модели `fuzz_div` также заслуживает внимания модель, сформированная с параметром $\varepsilon = 2$ пикселя. В дальнейшем модели `fuzz_div` с параметром $\varepsilon = 2$ и $\varepsilon = 4$ пикселям будем соответственно обозначать `fuzz_div2` и `fuzz_div4`.

СРАВНЕНИЕ МЕТОДА С ИЗВЕСТНЫМИ ЗАПАДНЫМИ АНАЛОГАМИ

Для тестирования предложенного метода поиска использована методика и база данных FERET [5]. При оценке метода в соответствии с [5] необходимо для каждого предъявляемого тестового образа отсортировать образы из БД в порядке убывания степени соответствия образов из БД тестовому образу. Говорят, что тестовый образ распознан по рангу I , если в отсортированном списке соответствующий тестовой фотографии образ расположен не ниже i -й позиции. Безусловно, зависимость процента распознавания от ранга, получившая название CMS (cumulative matching score), является удачной для оценки метода поиска в БД при данной постановке задачи.

В рамках проекта FERET предложено четыре теста: $fafb$, $fafc$, $dup1$ и $dup2$. БД для всех четырех тестов состоит из 1196 фотоизображений различных людей, сделанных по правилам, допускающим некоторое отклонение от правил опознавательного фото; тестовые же выборки – различны. Тестовый набор для $fafb$ составляют снимки, сделанные через короткий (порядка нескольких минут) промежуток времени после съемки изображения для БД, и тестовые изображения отличаются от изображений из БД, как правило, легким поворотом головы или эмоциональным выражением. Тестовые фотоизображения для $fafc$ отличаются от изображений из БД выбранной фотографической системой и уровнем освещенности; тестовые выборки из $dup1$ и $dup2$ – состоят из фотоизображений, сделанных через определенный период после фото из БД: от 3 до 18 месяцев для $dup1$ и от 18 до 36 месяцев для изображений для $dup2$. Объем тестовых выборок для тестов $fafb$, $fafc$, $dup1$ и $dup2$ составил соответственно 1195, 194 722 и 234 фронтальных фотопортретов.

Говоря о качестве изображений, представленных в БД, следует отметить, что они не всегда в полной мере отвечают требованиям опознавательной съемки: имеются фотографии людей в очках; в парах изображений из БД и тестовой выборки, соответствующих одному лицу, как правило, различаются эмоциональное состояние и положение головы (легкие повороты головы в различные стороны). В связи с этим, можно говорить о том, что условия эксперимента являются более жесткими, нежели те, которые оглашены в постановке задачи.

Для проведения эксперимента разработан программный модуль, формирующий модели фотопортретов и поиск в БД. Эксперимент проведен на ПК класса Pentium-IV 2.8 GHz с ОЗУ объемом 256 Мб, ОС – Windows XP. Средняя скорость сравнения составила 5352,9 тысяч сравнений в секунду при скорости предварительного формирования нечетких моделей более 7 тысяч образов в секунду. В таблице 1 приведены результаты сравнения описанных методов с западными аналогами, такими как: Baseline Correlation (arl_cor) и Baseline Eigenface (arl_ef) от US Army Research Laboratory; системы от NIST на основе eigenface, использующие различные метрики ($ef_hist_dev_ang$, $ef_hist_dev_anm$, и т.д.), ex-

calibur от Excalibur Inc., системы от MIT Media Lab (mit_mar_95 , mit_sep_96); umd (umd_mar_97) от University of Maryland и usc (usc_mar_97) от University of Southern California.

Таблица 1 – Результаты тестирования методов в сравнении с зарубежными аналогами

Название метода / модели	Ранг	Fafb	Fafc	dup1	dup2
Fuz_sub	1	0,9682	0,5	0,49792	0,38034
	10	0,99414	0,76289	0,638	0,62393
	20	0,99498	0,84536	0,69071	0,7094
Fuz_div2	1	0,93054	0,86082	0,41748	0,32479
	10	0,97908	0,94845	0,61997	0,55556
	20	0,98494	0,96392	0,681	0,61966
Fuz_div4	1	0,96736	0,59794	0,50208	0,36325
	10	0,9908	0,8299	0,6699	0,62821
	20	0,99414	0,8866	0,71845	0,73077
Arl_cor	1	0,826778	0,0515464	0,362881	0,17094
	10	0,949791	0,175258	0,540166	0,393162
	20	0,964854	0,201031	0,616343	0,512821
Arl_ef	1	0,796653	0,185567	0,409972	0,222222
	10	0,921339	0,520619	0,596953	0,461538
	20	0,943096	0,587629	0,6759	0,58547
Ef_hist_dev_ang	1	0,701255	0,0721649	0,34072	0,123932
	10	0,882008	0,21134	0,50831	0,307692
	20	0,920502	0,268041	0,566482	0,358974
Ef_hist_dev_anm	1	0,774059	0,237113	0,445983	0,209402
	10	0,931381	0,530928	0,628809	0,363248
	20	0,955649	0,623711	0,689751	0,448718
Ef_hist_dev_l1	1	0,772385	0,257732	0,350415	0,132479
	10	0,922176	0,530928	0,516621	0,320513
	20	0,943096	0,64433	0,565097	0,367521
Ef_hist_dev_l2	1	0,716318	0,0412371	0,331025	0,136752
	10	0,896234	0,170103	0,50831	0,311966
	20	0,932218	0,242268	0,578947	0,401709
Ef_hist_dev_md	1	0,741423	0,231959	0,422438	0,166667
	10	0,907113	0,520619	0,621884	0,337607
	20	0,94477	0,639175	0,671745	0,405983
Ef_hist_dev_ml1	1	0,733054	0,391753	0,304709	0,128205
	10	0,837657	0,649485	0,445983	0,260684
	20	0,866946	0,701031	0,470914	0,299145
Ef_hist_dev_ml2	1	0,771548	0,309278	0,34626	0,128205
	10	0,916318	0,572165	0,522161	0,320513
	20	0,939749	0,659794	0,576177	0,384615
Excalibur	1	0,794142	0,216495	0,414127	0,196581
	10	0,929707	0,541237	0,605263	0,410256
	20	0,948954	0,628866	0,67036	0,5
Mit_mar_95	1	0,83431	0,154639	0,33795	0,17094
	10	0,940586	0,371134	0,540166	0,410256
	20	0,960669	0,505155	0,632964	0,538462
Mit_sep_96	1	0,948117	0,319588	0,576177	0,34188
	10	0,979079	0,670103	0,725762	0,598291
	20	0,98159	0,762887	0,770083	0,683761
Umd_mar_97	1	0,962343	0,587629	0,472299	0,209402
	10	0,990795	0,865979	0,67036	0,534188
	20	0,994142	0,917526	0,722992	0,628205
Usc_mar_97	1	0,949791	0,819588	0,591413	0,521368
	10	0,985774	0,917526	0,722992	0,705128
	20	0,989121	0,948454	0,772853	0,764957

Из таблицы видно, что предложенные нечеткие методы на предложенных тестах показали довольно высокую эффективность. Методы fuz_sub и fuz_div4 показали

среди рассмотренных в таблице методов наилучший результат по самому объемному тесту fafb, и находятся в числе лидеров по остальным трем тестам. Отметим, что fuz_div2, уступая двум названным нечетким методам по тестам fafb, dup1 и dup2, при этом, показал уникальный для рассмотренных в таблице методов результат по тесту fafc – более чем 86-процентное распознавание по рангу 1. По совокупному результату тестирования, т.е. при учете всех четырех тестов, можно сделать вывод, что методы fuz_div4 и fuz_sub наряду с методом usc_mar_97 являются одними из лучших методов, рассмотренных в таблице 1. Сравнивая данные методы, отметим, что некоторое превосходство методов fuz_sub и fuz_div4 на наиболее объемном тесте fafb компенсируется преимуществом метода usc_mar_97 на остальных трех тестах, при этом заметное преимущество метода usc_mar_97 на ранге 1 имеет тенденцию к резкому сокращению на более высоких рангах. Таким образом, по совокупному результату usc_mar_97, fuz_sub и fuz_div4 можно считать методами примерно одного класса.

ВЫВОДЫ

Описанные в данной публикации нечеткие методы и модели показали свою высокую эффективность. Предложенные методы fuz_sub и fuz_div4 показали наилучший результат по тесту fafb, а метод fuz_div2 – по тесту fafc.

УДК 681.518:004.6

А. Б. Кунгурцев

ИССЛЕДОВАНИЕ ИНФОРМАЦИОННЫХ СИСТЕМ ДЛЯ ВНЕДРЕНИЯ МЕХАНИЗМОВ МАТЕРИАЛИЗОВАННЫХ ПРЕДСТАВЛЕНИЙ И ШАБЛОНОВ

Предлагается метод классификации запросов информационной системы (ИС) для формирования материализованных представлений (МП) и шаблонов. Предложена методика статистического исследования ИС для отбора запросов в группы МП и шаблонов, позволяющая обосновать эффективность применения указанных механизмов.

ВВЕДЕНИЕ

Применение механизма материализованных представлений (МП), предусматривающего сохранение результатов запросов, может значительно повысить производительность обработки запросов [1]. В работе [2] проведено исследование параметров запросов и показано, при каких условиях применение МП в конкретной информационной системе (ИС) окажется эффективным. Однако за рамками исследования остался очень важный вопрос – как организовать исследование ИС, каким образом произ-

Также по совокупному результату методы fuz_sub и fuz_div4 являются наряду с методом usc_mar_97 одними из наиболее эффективных методов, и, безусловно, могут быть положены в основу модуля поиска неизвестного лица в БД по фотопортрету.

ПЕРЕЧЕНЬ ССЫЛОК

1. Системы Папилон, АДИС «Папилон» – Режим доступа: <http://www.papillon.ru/adis.html> – название экрана.
2. *Саницкий В.А та ін.* Система інформаційного забезпечення ОВС України: Навчально-практичний посібник / За ред. Л.В. Бородича. – К., Редакційно-видавничий відділ МВС України, ТОВ АНТЕКС, 2000. – 144с., іл., таб.
3. *Шиффман Х. Р.* Ощущение и восприятие. 5-е изд. – СПб.: Питер, 2003. – 928с.: ил.
4. *Снетков В.А., Виниченко И.Ф., Зинин А.М., Освянникова М.Н.* Криминалистическое описание внешности человека: учебное пособие. / Под ред. проф. В. А. Снеткова. – ВНИИ МВД СССР, 1984. – 128 с., 51 рис.
5. *P. Jonathon Phillips & other, The FERET Evaluation Methodology for Face-Recognition Algorithms.* Technical report NISTIR 6264. – 1999.

Після доробки 22.11.2004

В публікації розглянуто проблему пошуку невідомої людини у БД по фотопортрету. Запропоновано дві нечіткі моделі представлення фото і метод пошуку фотографії на основі нечіткого порівняння з еталоном.

This article is dedicated to the problem of finding an unknown person in database by a photo. Two fuzzy models of representing a photo & the method of finding a photo based on fuzzy comparison with the template are offered.

вести классификацию и отбор транзакций для их дальнейшего анализа с целью создания МП.

Эффект от внедрения шаблонов запросов несколько другого рода. Здесь труд прикладного программиста частично перекладывается на плечи конечного пользователя за счет формирования определенных фрагментов запроса к БД непосредственно в момент возникновения потребности в этом запросе. Выгода от внедрения этого механизма заключается в повышении оперативности системы и снижении расходов на программирование. В настоящее время использование шаблонов, как и МП, в основном основано на интуитивной оценке их эффективности.

Целью данной работы является создание методики анализа транзакций реальной ИС, основанной на классификации запросов к системе, позволяющей объективно оценить эффективность внедрения механизма материализованных представлений и шаблонов.

1 КЛАССИФИКАЦИЯ ЗАПРОСОВ

Только запросы типа SELECT [3] являются потенциальными кандидатами на МП.

```
SELECT [ALL | DISTINCT] список_выбора
FROM {имя_таблицы | имя_курсора}[, {имя_таблицы_1
имя_курсора}]
[WHERE условие]
[GROUP BY имя_столбца [, имя_столбца]...]
[HAVING условие]
[ORDER BY {имя_столбца | список_выбора} [ASC] [DESC]
[, {имя_столбца | список_выбора} [ASC | DESC]]...]
```

Пусть Z_1 и Z_2 два запроса к БД. Введем следующие обозначения.

T_1 и T_2 – множество таблиц, используемых в Z_1 и Z_2 соответственно.

P_1 и P_2 – множество полей, используемых в Z_1 и Z_2 соответственно.

U_1 и U_2 – условия выборки для Z_1 и Z_2 соответственно.

S_1 и S_2 – множества полей в списках сортировки для Z_1 и Z_2 соответственно.

С точки зрения МП все запросы типа SELECT можно отнести к трем категориям.

Идентичные запросы ($Z_1 = Z_2$). Кроме запросов полностью совпадающих по тексту к этой категории также следует отнести запросы, отличающиеся формой записи, но не семантикой. Ниже приведены некоторые характерные и наиболее простые случаи.

– В сложном условии не совпадает порядок простых условий:

```
type = 'business' type = 'business'
or price > $10 or advance < &200
or advance < &200 or price > $10
```

– Применены различные способы записи отрицания:

```
where type != 'business' where not type = 'business'
```

– Применены различные формы записи диапазонов:

```
Where ytd_sales >= 4000 where ytd_sales between 4000 and
and ytd_sales <= 12000 12000
```

Один запрос можно пересчитать в другой, используя только результирующую таблицу. Здесь возможен ряд вариантов.

А) $T_1 = T_2$, $P_1 = P_2$, $U_1 = U_2$, отличие заключается в порядке расположения полей в списке выборки. Здесь под равенством P_1 и P_2 понимается равенство множеств элементов списков выбора. Таким образом, порядок элементов не имеет значения. Однако применение к некоторому полю, например, агрегирующей функции, делает его неравным полю, к которому применена другая функция или не применена функция вообще.

Пересчет Z_1 и Z_2 предусматривает перестановку столбцов результирующей таблицы.

В) $T_1 = T_2$, $P_1 \subset P_2$, $U_1 = U_2$. Пересчет Z_1 и Z_2 предусматривает удаление из результирующей таблицы всех $p \in P_2 - P_1$.

С) $T_1 = T_2$, $P_1 = P_2$, $U_1 = U_2$, $S_1 \neq S_2$. Пересчет Z_2 в Z_1 предусматривает выполнение в результирующей таблице новой сортировки в соответствии с S_2 .

Д) $T_1 = T_2$, $P_1 = P_2$, $U_1 = U_2$, $S_1 = S_2$, критерии сортировки в Z_1 и Z_2 не совпадают. Пересчет Z_2 в Z_1 предусматривает новую сортировку в результирующей таблице в соответствии с критерием Z_1 .

Е) $T_1 = T_2$, $P_1 = P_2$, $U_1 \neq U_2$, простое условие u_{2i} из U_2 поглощает простое условие u_{1i} из U_1 (например, условие « $a > 15$ » поглощает условие « $a > 10$ »). Пересчет Z_2 в Z_1 предусматривает применение к результирующей таблице условия u_{1i} .

Ф) $T_1 = T_2$, $P_1 = P_2$, $U_1 \neq U_2$. Отличие в условии заключается в том, что Z_1 использует предложение LIKE. Если подстрока из LIKE в Z_1 применена к столбцу, который в Z_2 выбран целиком, то возможен пересчет.

Г) $T_1 = T_2$, $P_1 \neq P_2$, $U_1 = U_2$. Неравенство списков выбора заключается в использовании в Z_1 агрегирующей функции для некоторого поля p_i . Если это поле в Z_2 выбирается полностью, то возможен пересчет Z_2 по результирующей таблице в Z_1 .

Н) Анализ отличий запросов, для которых $T_1 = T_2$, $P_1 = P_2$, $U_1 \neq U_2$, а также тех, у которых отличие заключается в применении агрегирующих функций позволяет создать «групповой запрос», обслуживающий группу подобных запросов методом пересчета его результирующей таблицы. Например, для двух запросов с условиями « $a > 10$ and $a < 50$ » и « $a > 5$ and $a < 20$ » можно создать групповой запрос с условием « $a > 5$ and $a < 50$ ».

Перед принятием решения о создании группового запроса следует оценить размер его результирующей таблицы.

Один запрос может быть пересчитан в другой с помощью дополнительного запроса к БД. Классификация запросов для этого варианта построения МП еще не завершена.

Шаблоны могут быть созданы для всех видов запросов. Ниже приведена классификация шаблонов для запросов типа UPDATE, INSERT и DELETE. Классификация шаблонов для запросов типа SELECT во многом подобна приведенной выше классификации для МП и поэтому не будет рассматриваться в дальнейшем.

Для запросов типа UPDATE можно применить следующий формат [2].

```
UPDATE <Имя таблицы>
SET <Имя поля> = <Выражение>
...
<Имя поля> = <Выражение>
[WHERE <Условия выбора>];
```

Возможны следующие варианты шаблонов.

А) Имя таблицы произвольное. Отсюда следует, что и все остальные настройки произвольны. Получаем шаблон для фразы UPDATE в самом общем виде. Создание шаблона подобного типа весьма затруднительно, поскольку предусматривает предоставление пользователю полного набора таблиц и их полей, а также проверку корректности формирования условий.

В) Имя таблицы неизменно, поля редактирования и условия разные.

С) Имя таблицы и поля редактирования неизменны, разные значения и условия выбора.

Д) Имя таблицы, поля редактирования и условия выбора неизменны, меняются только значения полей.

Оператор INSERT имеет две основные формы [2]. Первая форма позволяет добавить в таблицу одну строку.

```
INSERT INTO <имя таблицы>
[(<список полей>)]
VALUES (<список значений>);
```

Вторая форма позволяет вставить несколько строк и обычно используется для копирования данных. В упрощенном виде она представлена ниже.

```
INSERT INTO <имя таблицы>
[(<список полей>)]
SELECT <список полей>
FROM <имя таблицы>
WHERE <условие выбора>
```

Рассмотрим варианты такого шаблона для INSERT.

А) Имя таблицы, куда производится вставка произвольно. Тогда и все остальные настройки произвольны. Получаем шаблон для фразы INSERT в самом общем виде. Создание подобного шаблона достаточно трудная задача.

В) Оператор INSERT первой формы. Имя таблицы, куда производится вставка зафиксировано. В этом случае с большой долей вероятности можно утверждать, что и список полей будет фиксированным. В этом случае переменным остается список значений.

С) Оператор INSERT второй формы. Имя таблицы, куда производится вставка зафиксировано. Имя таблицы из фразы FROM, список полей из фразы SELECT и условия выбора произвольны. Работа с таким шаблоном достаточно сложная, однако для большинства ИС список таблиц, из которых можно копировать строки, очевидно, не будет большим. Это позволит в подсказке предоставить не только список таблиц, но и список их полей.

Д) Оператор INSERT второй формы. Имя таблицы, куда производится вставка зафиксировано. Имя таблицы из фразы FROM также зафиксировано. Тогда очень вероятно, что и список полей из фразы SELECT будет фиксированным. В этом случае переменным остается только условие выбора фразы WHERE.

Оператор DELETE имеет следующий формат:

```
DELETE FROM <имя таблицы>
[WHERE <условие выбора>];
```

Для этого оператора возможны два шаблона.

А) Не фиксировано имя таблицы и условие. Получим шаблон общего вида. Трудно представить себе ситуацию, когда конечному пользователю будет предоставлена возможность по своему усмотрению производить удаление из любой таблицы базы данных. Однако, если список таких таблиц будет ограничен функциями рабочего места, то возможно построение шаблона подобного типа.

В) Имя таблицы фиксировано. Переменным является условие.

Вопрос о целесообразности создания МП и шаблонов может быть решен после анализа запросов ИС. Для некоторых запросов решение можно принять на основании опроса конечных пользователей. Однако оценка эффективности такого решения может быть только интуитивной. Обоснованную оценку эффективности применения шаблонов и МП может дать только статистическое исследование выполненных транзакций, которое можно провести по данным журнала транзакций.

2 ПОЛУЧЕНИЕ СТАТИСТИЧЕСКИХ ДАННЫХ

Для статистической обработки предлагается следующая стратегия.

В качестве наибольшей генеральной совокупности принимается совокупность транзакций за один год работы системы. Такой интервал позволяет охватить все периоды колебаний, связанных с сезонностью, отчетностью, анализом деятельности организации и другими факторами, а также учесть тренд, обусловленный, например, развитием ИС. Обработка всех транзакций за год может оказаться практически нереализуемой задачей, поэтому далее обосновывается гипотеза создания репрезентативной выборки.

Если нет истории транзакций, либо она не полная (например, фиксировались только запросы, которые изменяли данные), то предварительное исследование можно выполнить и на меньшем статистическом материале, с последующей коррекцией результатов по мере накопления материала.

Для построения выборки будем исходить из следующих предположений.

Частота появления запросов, которые могут быть «кандидатами» для ПМ и шаблонов соответствует частоте всех запросов. Интенсивность запросов может быть разной в зависимости от дня недели. Поэтому в качестве рабочего дня выбирается каждый n -й, где n не кратно k (k – количество рабочих дней в неделе), что позволяет в большой выборке обеспечить практически равную вероятность включения каждого рабочего дня недели.

Количество элементов в выборке будет определяться наименьшим временным интервалом, который будет учтен при реализации механизмов МП или шаблонов. Например, если в конце каждого месяца предполагается корректировка указанных механизмов, то промежуточный уровень статистических результатов должен относиться к месяцу. В любом случае следует установить больше выборочных значений, чем этого требует принятая доверительная вероятность, например 0.95, поскольку из общей выборки предстоит сформировать вторичные выборки в соответствии с рассмотренными выше типами запросов. Предлагается количество элементов в выборке устанавливать в соответствии с формулой:

$$N = \frac{N_g}{2^k}, \quad (1)$$

где N_g – число элементов в генеральной совокупности; k – целое число.

Из (1) следует, что в выборку включается каждый 2^k элемент генеральной совокупности, например, при $k = 4$ – каждый шестнадцатый. Если выборка сделана и обработана, то при необходимости можно увеличить количество элементов в выборке в два раза без пересчета ранее выбранных запросов путем выбора каждого шестнадцатого запроса, но со смещением на восемь позиций относительно элементов первой выборки.

3 ОБРАБОТКА СТАТИСТИЧЕСКИХ ДАННЫХ

Пусть множество элементов выборки будет

$$Z = \{Z_1, Z_2, \dots, Z_N\},$$

тогда в соответствии с целью дальнейшего анализа нужно определить принадлежность каждого запроса определенной группе запросов в соответствии с ранее приведенной классификацией. Для этого каждый элемент выборки должен быть представлен как m -мерный вектор

$$Z_i = \begin{pmatrix} Z_i^{(1)} \\ Z_i^{(2)} \\ \vdots \\ Z_i^{(m)} \end{pmatrix},$$

где каждый компонент определяет признак запроса.

На основании анализа признаков формируются множества запросов, которые могут быть представлены шаблонами или материализованными представлениями

$$Z^{(j)} = \{Z_1^{(j)}, Z_2^{(j)}, \dots, Z_l^{(j)}\}. \quad (2)$$

Здесь j – признак классификации запросов.

Для шаблонов на этом исследовании ИС заканчивается, поскольку значения l (мощность множества) достаточно для определения эффективности построения шаблона соответствующего типа.

Решить вопрос об эффективности МП можно только с учетом частоты изменения используемых ими данных. Если рассматривать (2), как множество запросов, формирующих некоторое МП, то следует сформировать множество используемых этим МП данных.

$$D^{(j)} = D_1^{(j)} \cup D_2^{(j)} \dots \cup D_l^{(j)}, \quad (3)$$

где $D_i^{(j)} = \{d_{11}, d_{12}, \dots, d_{1q}\}$ – множество данных (полей таблиц), используемых в запросе $Z_i^{(j)}$. Аналогично $T^{(j)}$ будет представлять множество таблиц для этого же МП. Следует отметить, что в (2) могут попасть много одинаковых запросов. Если же будет сформирован групповой запрос, то именно его множество данных будет определено в (3).

Каждое изменение данных из $D^{(j)}$ должно быть зафиксировано счетчиком $k^{(j)}$, для чего удобно использовать операцию присваивания ($:=$). Будем фиксировать

изменения данных в результате запросов UPDATE, INSERT и DELETE, которые попали в выборку Z .

Пусть запрос UPDATE редактирует данные d_1, d_2, \dots, d_n , тогда если $\{d_1, d_2, \dots, d_n\} \cap D^{(j)} \neq \emptyset$, то $k^{(j)} := k^{(j)} + 1$.

Пусть запрос INSERT или DELETE вставляет или удаляет строку в таблице T_i , тогда, если $T_i \notin T^{(j)}$, то $k^{(j)} := k^{(j)} + 1$.

Отношение $r = l/k^{(j)}$ дает предварительную оценку эффективности МП. Очевидно, при $r < 2$ механизмы МП, предусматривающие полный пересчет окажутся неэффективными.

После выбора МП и шаблонов имеет смысл определить доверительную вероятность их появления, учитывая, что на стадии предварительной статистической обработки еще не было известно, какие именно запросы будут представлять интерес, и количество обрабатываемых запросов не могло определяться теми запросами, которые были окончательно выбраны. Если расчеты окажутся неудовлетворительными, то можно увеличить число элементов, как об этом упоминалось выше.

Другим способом улучшения принятых решений является определение тенденции изменения частоты запросов, входящих в множества МП и шаблонов. Если обнаружен тренд для элементов этих множеств, и он не соответствует тренду частоты всех запросов, то можно считать, что происходят качественные изменения в используемой информации в ИС и в скором времени, возможно, придется выполнять новый анализ транзакций и выявлять новые МП и шаблоны.

Следует отметить, что предлагаемые исследования используются в качестве прогноза на последующий период. Если не происходят серьезные изменения в структуре БД, в налоговой, статистической и другой подобной сфере, не наблюдается существенное изменение количества и функций рабочих мест, то есть все основания распространить полученные результаты на текущий период работы системы. Однако даже в этом случае рекомендуется продолжать вести журнал транзакций и периодически, например один раз в месяц, производить из него выборку за новый период.

Результаты обработки такой выборки, особенно при сравнении их с соответствующим периодом в прошлом, позволят своевременно отслеживать изменения в процессах обработки информации и принимать своевременные решения по корректировке множеств МП и шаблонов.

ВЫВОДЫ

Предложенный метод классификации запросов реляционных баз данных и статистической обработки транзакций информационной системы позволяет обосновать выбор запросов для материализованных представлений и шаблонов и значительно повысить эффективность внедрения этих механизмов.

В рамках данного направления исследования представляет интерес разработка метода адаптивной коррекции МП, по результатам текущего исследования ИС.

ПЕРЕЧЕНЬ ССЫЛОК

1. Чаудхари С. Методы оптимизации запросов в реляционных системах // Системы управления базами данных. – 1998. – Вып 3. – С 22. – <http://www/osr.ru/dbms/1998/03/22/htm>.
2. Кунгурцев А. Б., Куок Винь Нгуен Чан. Анализ возможности применения материализованных представлений в информационных системах. Труды ОНПУ, 2003, вып. 2(20). С. 102–106.
3. Джудит С. Боуман, Сандра Л. Эмерсон, Марси Дарновски. Практическое руководство по SQL. К.: Диалектика, 1997. 320 с.

Надійшла 17.11.2003
Після доробки 31.03.2004

Пропонується метод класифікації запитів інформаційної системи (ІС) для формування матеріалізованих уявлень (МУ) і шаблонів. Запропоновано методику статистичного дослідження ІС для добору запитів у групи МУ і шаблонів, що дозволяє обгрунтувати ефективність застосування зазначених механізмів.

The method of classification of inquiries of information system for formation of materialized view and patterns is offered. The technique of statistical research of information systems for selection of inquiries in the groups allowing to prove efficiency application of the specified mechanisms is offered.

УДК 004.272

В. М. Опанасенко, І. Г. Тимошенко

АРХІТЕКТУРНА ОРГАНІЗАЦІЯ РЕКОНФІГУРОВНИХ КОМП'ЮТЕРІВ НА БАЗІ ПЛІС

Виконано аналіз сучасного стану проблеми Reconfigurable computing – механізми розробки комп'ютерних архітектур, які базуються на інтегральних схемах типу FPGA, а також аналіз статичного способу проектування з використанням інструментальних засобів автоматизованого проектування та способу з використанням мови Java. Розглянуті перспективи механізму з використанням Java, які обумовлені перевагами над традиційними методами проектування: перевірка і моделювання проекту на рівні двійкових потоків, підтримка RTR, взаємодія з інструментальним пакетом для проектування систем за технологією IRL – Jbits.

ВСТУП

Розвиток новітніх технологій у світі комп'ютерної техніки й мікроелектроніки та створення програмних логічних інтегральних схем (ПЛІС) високого ступеня інтеграції викликало підвищений інтерес до досліджень в області реконфігуровного комп'ютерингу. Реконфігуровна комп'ютерна система поєднує поняття реконфігуровної структури апаратного пристрою і процесу обробки даних, який виконується цим пристроєм. Відмінність реконфігуровного комп'ютера від традиційного полягає в тому, що структура реконфігуровних пристроїв не є фіксованою і змінюється в залежності від виконуваної задачі чи алгоритму, що дозволяє вносити зміни в готовий і функціонуючий продукт.

Gerald Estrin на початку 1960 років запропонував «комп'ютер з елементами фіксованої та змінної структури» [1]. Ця базова архітектура, яка підтримує програмовані пристрої і програмне забезпечення, лежить в основі реконфігуровних комп'ютерних систем. Дана концепція набагато випередила розвиток технологій у світі мікроелектроніки і, на жаль, він зміг представити тільки наближення своєї розробки. Множина понять, які зараз відкриті організацією з досліджень в області реконфігуровного комп'ютерингу, не були прийняті до уваги в цих дослідженнях.

Втілення основних концепцій реконфігуровної комп'ютерної системи в реальні проекти стало можливим тільки

з появою інтегральних схем типу FPGA (field programmable gate array), які містять масиви конфігуровних логічних блоків і програмованих взаємозв'язків між цими блоками [2]. Логічні блоки можуть бути зконфігуровані на виконання простих чи складних логічних функцій і можуть змінюватися відповідно до заданої вимоги.

МЕХАНІЗМИ РОЗРОБКИ КОМП'ЮТЕРНИХ АРХІТЕКТУР, ЩО БАЗУЮТЬСЯ НА ІНТЕГРАЛЬНИХ СХЕМАХ ТИПУ FPGA

Ємність сучасних FPGA нараховує більше ніж 10 мільйонів логічних вентилів, крім того, FPGA також включають RAM (Read Access Memory), апаратні перемножувачі, які є стандартними компонентами сучасних мікропроцесорів та фіксоване мікропроцесорне ядро. Метод, за яким конфігуруються логічні блоки та взаємозв'язки, специфічний кожному виробнику FPGA. Логічні блоки та взаємозв'язки мають внутрішню структуру, яка зберігає поточну конфігурацію.

Час, потрібний для конфігурації FPGA, називається часом конфігурації. Час конфігурації залежить від серії, ємності FPGA та їх кількості. Для реконфігуровної комп'ютерної системи, яка включає декілька FPGA час конфігурації залежить не тільки від часу конфігурації окремої інтегральної схеми, а й від конфігурації всіх схем. Кристал FPGA може бути сконфігурований послідовно чи паралельно в залежності від розробки системи і, таким чином, час конфігурації може варіювати від сотень наносекунд до кількох секунд. Рисунок 1 ілюструє базові архітектури реконфігуровних систем, з яких походить більшість архітектур.

У Сполучених Штатах Америки ведуться інтенсивні роботи у напрямку досліджень та розробки комп'ютерних архітектур, які базуються на інтегральних схемах типу FPGA з метою прискорення обробки даних. Розробка структури реконфігуровного пристрою і її реалізація в кристалі на вентильному рівні дозволяє

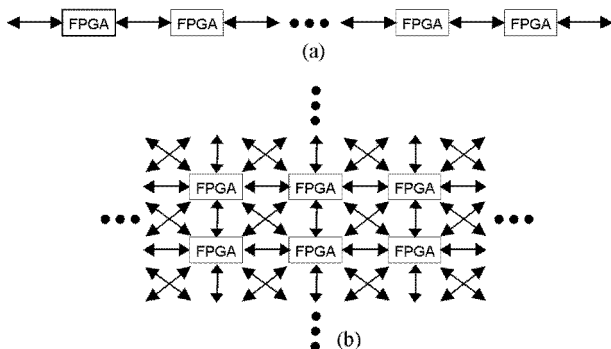


Рисунок 1 – Базові архітектури реконфігурованих систем з множиною FPGA

збільшити швидкість комп'ютера на декілька порядків у порівнянні з традиційними рішеннями.

Не всі алгоритми цифрової обробки даних можуть бути ефективно реалізовані з використанням сучасних інтегральних схем. На Сьомій Міжнародній Конференції «IRTC-2001» обговорювалися такі проблеми як виконання обчислень з плаваючою крапкою та виконання числових операцій з високою точністю [3].

Процес проектування цифрових схем спрощується з появою апаратних мов описів, тому що використання машинних схематичних інструментальних засобів стає складним при наявності великої кількості інтегральних схем [4]. Міжнародним стандартом є мова HDL (Hardware Description Language), спочатку розроблена для цілей проектування, моделювання і документування. Синтез з інструментальними засобами дозволив автоматизувати процес проектування.

До складу САПР ПЛІС вводиться CORE Generator – інструментальний засіб, який надає у розпорядження користувача параметричні логічні Cores, оптимізовані для ПЛІС, які описуються мовою VHDL. Система CORE дозволяє істотно зменшити час розробки нових проектів. У відповідності зі сформульованими технічними вимогами, проектувальник по мережі Internet може одержати оптимізоване для FPGA логічне Core і включити його у свій проект [5].

Істотно скорочується час розробки проекту з використанням System Generator, який є програмним інструментальним засобом для проектування, моделювання та створення систем Digital Signal Processing (DSP) на базі FPGA [6]. Він дозволяє швидко застосовування алгоритмів DSP до пристроїв FPGA. Скорочення процесу розробки проекту, від його концепції до робочих апаратних засобів, досягається за допомогою використання множини компонентів одночасно. Потіки даних, традиційні мови апаратних описів (VHDL і Verilog) та функції, які одержані з мови програмування MATLAB, можуть використовуватися одночасно, моделюватися разом і синтезуватися в робочі апаратні засоби. На рис. 2 зображено процес моделювання системи DSP.

MATLAB інтегрує математичні обчислення, візуалізацію і потужну мову, забезпечуючи гнучке середовище для технічних застосувань. Відкрита архітектура

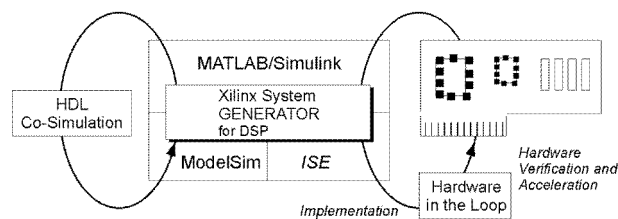


Рисунок 2 – Моделювання системи DSP

робить зручним використання MATLAB для дослідження даних, створення алгоритмів та користувальницьких інструментальних засобів.

Simulink – інтерактивний інструмент для моделювання, імітації та аналізу динамічних систем. Він дозволяє точно описувати, моделювати, оцінювати та уточнювати поведінку системи завдяки стандартним і користувальницьким бібліотекам. Simulink цілісно інтегрується з MATLAB, забезпечуючи швидкий доступ до множини інструментальних засобів проектування та аналізу.

System Generator розширює можливості Simulink моделювання на системному рівні за допомогою bit і cycle-true моделювання схем FPGA [7].

System Generator включає Simulink бібліотеку функціональних блоків для побудови DSP, арифметичних схем і схем цифрової логіки. Розробнику надається можливість комбінування блоків Xilinx з блоками MATLAB і Simulink для створення реалістичних test benches і аналізу даних. Високий рівень абстракції, який надає System Generator, спрощує розробку алгоритмів і їх перевірку.

Крім бібліотеки моделювання на системному рівні System Generator включає генератор коду, який автоматично генерує синтезований VHDL код з Simulink моделі. VHDL код включає IP (intellectual property) блоки, які розроблені для високої продуктивності та інтенсивності в пристроях Xilinx типу FPGA.

Вимоги до пристроїв на базі сучасних FPGA роблять майже неможливим налагодження та перевірку проектів з використанням традиційних логічних методів аналізу [8]. Компанією Xilinx розроблений налагоджувач апаратних засобів ChipScope Pro, який дозволяє перевіряти функціонування схем на базі FPGA. ChipScope Pro включає logic analyzer (ILA) і bus analyzer (IBA). Ці Cores дозволяють переглядати всі внутрішні сигнали, вузли і системні шини в межах FPGA.

Створення ПЛІС високого ступеня інтеграції обумовило появу технології Internet Reconfigurable Logic (IRL). Ця технологія забезпечує дистанційну реконфігурацію структур апаратних пристроїв, включених до мережі і реалізованих на елементній базі ПЛІС, через мережу Internet [9]. Для розробки продуктів на основі технології IRL необхідний розвиток трьох фундаментальних технологій – всеохоплюючого використання мережних технологій, набору програмних засобів на основі мови Java і реконфігурованих кристалів FPGA типу Virtex. Архітектура Virtex на даний час дозволяє розробляти пристрої з кількістю вентилів 10 мільйонів і може

містити в собі апаратну реалізацію інтерпретатора Java програм для Java Virtual Machine (JVM).

Технологія Java дозволяє створювати програми для роботи з файловими системами і з локальними та глобальними комп'ютерними мережами. Java-застосування компілюються в байт-код, який виконується на JVM, яка представляє собою інтерпретатор байт-коду Java, що дозволяє динамічно розширювати систему. При першому запуску Java-програми JVM здійснює збір модулів і встановлює зв'язки між ними, при чому пошук відсутніх модулів здійснюється не тільки в локальній системі, а й у мережі [10].

Перед першим запуском нового застосування віртуальна машина перевіряє його код на належність до байт-коду, на небезпечність команд для системи і локальної мережі, на відповідність з дозволеними операціями, що свідчить про високий ступінь захисту та безпеки мови Java для роботи в мережі. Java забезпечує швидкий цикл компіляції і налагодження програм. Засоби розробки, що містять JVM, забезпечують контроль програмних застосувань на стадії виконання [11]. Істотною рисою архітектури Java є вбудована багатопотоковість, що дає можливість одночасно використовувати кожну функцію даної бібліотеки декількома потоками та виконувати декілька задач одночасно в межах одного застосування, що є актуальним для розподілених систем, коли процеси обміну по мережі можуть протікати одночасно і асинхронно. Об'єктно-орієнтовна парадигма мови Java зручна для організації розподілених обчислень, а також при створенні програмного забезпечення типу клієнт-сервер.

Завдяки наявності JVM мова Java є мобільною [10]. JVM забезпечує абстрагованість скомпільованих Java-програм від апаратної платформи й операційної системи, яка досягається перш за все стандартизацією «бінарного формату коду». Байт-код не залежить від конкретної платформи чи системи, для роботи Java-програми на даній апаратно-програмній платформі достатньо наявності відповідної JVM. Повна специфікація віртуальної Java машини відкрита й загальнодоступна і віртуальна машина може бути реалізована на будь-якій сучасній апаратно-програмній платформі. Система Java в силу своєї високої продуктивності, інтерпретовності і динамічної природи підходить для цілей швидкої розробки надійних програм.

Корпорацією Xilinx був представлений інструментальний пакет JBits (Java Based interface) для проектування систем за технологією IRL, призначений для створення спеціальних програмних застосувань мовою Java, які можуть використовуватися для модифікації апаратних засобів через мережу Internet [12].

JBits бере початок від работ, які пов'язані з кристалами серії XC6200 і базується на Java Environment for Reconfigurable Computing (JERC). JERC надавав у розпорядження Application Programming Interface (API), який дозволяв маніпуляцію з бітами конфігурації низького рівня.

Мотивацією розробки JBits послужила підтримка динамічної реконфігурації для кристалів FPGA серії Vir-

tex. Так як проектування пристрою було обмежене статичними методологіями розробки, інструментальними засобами проектування інтегральних схем – редактором схем або мовою HDL і не було забезпечене програмною підтримкою конфігурації, тому динамічна реконфігурація не була можливою.

На зміну статичному способу проектування прийшло рішення, яке передбачало підтримку програмного забезпечення, яке б дало повний доступ до всіх архітектурних особливостей реконфігуровного пристрою. Цим забезпеченням стала бібліотека скомпільованих Java програм. Так як бібліотека являє собою скомпільовані програми – виходом проектування не є статичний двійковий потік, а є Java код, який виконується і забезпечує керування конфігурацією і даними реконфігуровної логіки.

Перевагою процесу проектування з використанням JBits (рис. 3) є те, що виконувана програма являється тільки довольною частиною скомпільованого Java коду, що робить її переносною на інші системи і дозволяє інтеграцію з іншими частинами системи.

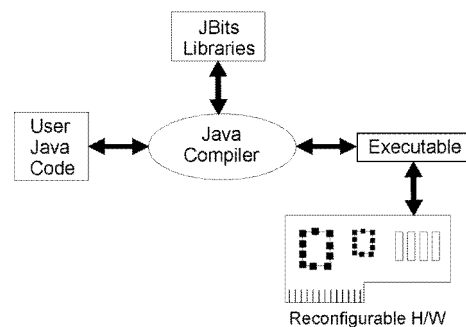


Рисунок 3 – Процес проектування з використанням JBits

JBits – це Java API, який забезпечує доступ до конфігурації пристроїв FPGA серії Virtex [13]. Інтерфейс дозволяє доступ до всіх реконфігуровних ресурсів пристрою для їх швидкої модифікації під програмним керуванням, що, в свою чергу, забезпечує програмну підтримку нових можливостей, які не були реалізовані раніше в пристроях Xilinx [14].

В той час як оточення JBits було представлено вперше, воно вимагало налагодження проекту за допомогою використання фізичних апаратних засобів [13]. Згодом був представлений Device Simulator, який працює на двійкових потоках і емує пристрій FPGA в програмному забезпеченні. Це було важливим кроком у напрямку проектування під програмним керуванням, але не було враховано спосіб моделювання зовнішніх периферійних пристроїв, які взаємодіють з FPGA. Недоліком Virtex Device Simulator було те, що він не мав вбудованої схеми для перенесення даних до / від пристрою.

JBits забезпечує можливість швидкого створення і реконфігурації схем серії Virtex під час виконання програмного застосування за допомогою. Run-Time Reconfiguration (RTR).

Моделювання на рівні пристрою з використанням мови Java має переваги над традиційними методами статичного проектування і налагодження проекту. Перевагами є перевірка проекту на рівні двійкових потоків, підтримка RTR, гнучкість при роботі у середовищі моделювання проекту і середовищі налагодження та контролю.

Використання стандарту HDL для тестування проекту дозволяє моделювати апаратні засоби, включаючи моделювання поведінки. Але цей підхід працює лише на рівні проектування, і не обов'язково забезпечує точну модель поведінки під час запуску проекту на фізичних апаратних засобах. Перевірка правильності проекту з використанням традиційного методу вимагає щоб кожен крок компіляції був без помилок, що є часто неможливим. VHDL проект і середовища налагодження також не підтримують реконфігурацію під час виконання і не забезпечують гнучкість при роботі.

Система JBits. На рис. 4 зображена схема системи JBits. *User Java Application* являє собою написану програму мовою Java, яка використовує JBits інтерфейс для маніпуляції конфігурованими ресурсами FPGA, які включають пошукові таблиці, маршрутизацію та тригери [14]. Кожна функція, яка викликається на рівні JBits інтерфейсу, робить один чи більше запитів до *Bit-level Interface*, який відповідає за розміщення біта у двійковому потоці для будь-якого пристрою серії Virtex. Bit-level Interface взаємодіє з файлом *Bitstream*, який керує двійковим потоком пристрою і підтримує зчитування та запис двійкових потоків з файлу / у файл. Можливість зчитування даних з пристрою і представлення їх у двійковому форматі є необхідністю для динамічної конфігурації.

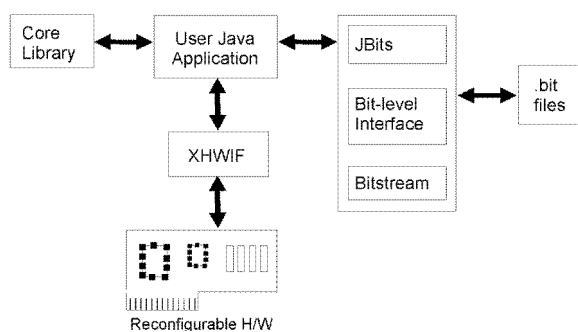


Рисунок 4 – Система JBits

JBits API використовує *Xilinx HardWare InterFace (XHWIF)* – Java інтерфейс, який розроблено для зв'язку зі схемами на базі FPGA і надає у розпорядження методи зчитування/запису цифрових потоків в/із FPGA, методи опису видів і кількості FPGA на платі. *XHWIF* також включає метод інкрименту інтегрованого таймера та метод запису в блоки пам'яті, які розташовані на схемі, якщо вони доступні. Таким чином, інтерфейс описує схему і дозволяє пересилання даних на схему і зі схеми. XHWIF інтерфейс стандартизує спосіб, по якому застосування спілкуються з апаратними засобами таким

чином, що вони, використовуючи той же самий інтерфейс, можуть взаємодіяти з множиною схем. *XHWIF server* дозволяє іншим застосуванням спілкуватися з реконфігурованими схемами, які знаходяться в межах мережі Internet. Ця можливість дозволяє налагоджувати проект без прямого доступу до апаратного забезпечення, крім того, дозволяє доступ до схеми множиною користувачів.

Core Library – це колекція Java програм, які визначають макрокомірки або Cores. Вони за звичай параметризовані і можуть переміщатися в межах пристрою. Прикладами Cores є лічильники, суматори, перемножувачі та інші стандартні логічні функції та функції обчислень.

Налагоджувач BoardScore. Корпорацією Xilinx був розроблений пакет BoardScore, подібний до старого інструменту WebScore для пристроїв серій XC6200, – інтерактивний налагоджувач апаратних засобів, який дозволяє слідкувати за функціонуванням схем на базі FPGA [12]. BoardScore дає можливість налагоджувати схеми FPGA при їх взаємодії з іншими апаратними компонентами конфігурації, відслідковувати внутрішній стан систем, які включають декілька FPGA, контролювати рівні сигналів, а також стан багаторозрядних шин даних [14].

Пакет може використовуватися для віддаленого налагодження пристроїв систем IRL або для спільної розробки автономних систем чи окремих кристалів через мережу Internet різними проектувальними групами.

З використанням XHWIF, двійкові потоки завантажуються для конфігурації FGAs, або зчитуються – для їх аналізу (рис. 5).

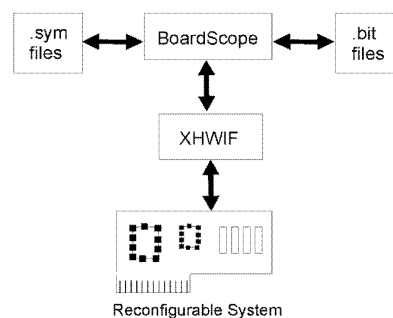


Рисунок 5 – Структура налагоджувача BoardScore

Крім того що BoardScore може взаємодіяти з локальними апаратними засобами та апаратними засобами, які включені до мережі, він може використовувати Device Simulator.

Device Simulator виконує моделювання апаратних засобів, яке засновано на параметрах конфігурації, і може використовуватися коли апаратні засоби недоступні. У такому випадку налагодження з використанням Device Simulator потребує порівняно менше часу, а ніж з прямим використанням апаратних засобів, так як не потребує виконання операції завантаження та зчитуван-

ня двійкових потоків [12]. На даний час BoardScore підтримує пристрої серії Virtex.

Run-Time Reconfiguration. JBits, на відміну від традиційних мов апаратних описів, створює нові можливості в області RTR, забезпечуючи можливість швидкого створення і динамічної модифікації схем серії Virtex під час виконання програмного застосування за допомогою прямого доступу до конфігурації двійкового потоку [13].

RTR визначається як динамічна модифікація апаратних схем FPGA під час виконання. RTR системи розрізняють за визначенням логіки схем і маршрутизації до виконання операції чи під час виконання. RTR системи модифікують схему декілька разів на протязі виконання програмного застосування. Так як не виключається можливість проектування низького рівня, JBits надає у розпорядження необхідні інструментальні засоби для ефективного модифікування чи створення проекту.

RTR досягається за допомогою інтеграції JBits і XHWIF API у JBits технології [14]. Тобто, RTR застосування може робити запити до JBits інтерфейсу для модифікації даних конфігурації у двійковому потоці і явні запити XHWIF для взаємодії з апаратним забезпеченням. Наприклад, RTR застосування виконує запити setConfiguration (пристрій, дані) і getConfiguration () для виконання завантаження і зчитування даних конфігурації.

На даний час JBits 3.0 для Virtex-II містить файли класу для створення Run Time Reconfigurable застосунків і інструментальні засоби, які використовують архітектуру Virtex-II [15]. JBits API дозволяє Java застосуванням динамічно модифікувати Xilinx Virtex-II двійкові потоки.

Обмеження JBits. Поряд з усіма перевагами способу проектування інтегральних схем з використанням JBits існують і деякі обмеження JBits API [14]. Програмне забезпечення вимагає повної специфікації деталей, на відміну від інструментів статичного проектування. JBits інтерфейс підтримує тільки структуровані схеми. Неструктуровані схеми, такі як схеми з випадковою логікою, не задовольняють пряме виконання в JBits застосуваннях.

JBits API вимагає від користувача повного ознайомлення з архітектурою пристроїв. В той час як архітектура пристроїв повністю задокументована фірмою Xilinx, більшість користувачів не мали необхідності ознайомитися з її особливостями. Очікується, що необхідність розуміння архітектури пристроїв буде самим великим бар'єром для схвалення і широкого поширення JBits інтерфейсу чи інших йому подібних інструментів.

В той час як JBits API дозволяє використовувати схеми, які створено стандартними інструментальними засобами розробки, зміною чи реконфігурацією схеми, на рівні JBits виключається можливість використання будь-яких інструментальних засобів аналізу, доступних проектувальникам схем. В JBits відсутня можливість аналізу синхронізації. Навіть невеликі зміни в конфігурації схеми можуть вплинути на її функціональні можливості, такі як синхронізація. Лише одним засобом, який нещодавно був розроблений для налагодження апаратних засобів, є BoardScore.

ВИСНОВКИ

Поява інтегральних схем типу FPGA зробила можливим втілення основних концепцій Reconfigurable computing в реальні проекти.

Синтез мови апаратних засобів HDL з інструментальними засобами дозволив автоматизувати процес проектування.

Процес розробки проекту істотно скорочується з використанням System Generator для DSP, який є програмним інструментальним засобом для проектування, моделювання та створення систем DSP на базі FPGA.

Технології IRL є основою для створення нових продуктів, які можуть динамічно модифікуватися як за допомогою програмного забезпечення, так і за допомогою алгоритму, заданого замовником. Комбінація трьох фундаментальних технологій – всеохоплююче використання мережних технологій, технологія Java та кристали Virtex дозволяє виконувати проектування та розробку зовсім нових IRL-продуктів. Такі продукти можуть поновлюватися та модифікуватися через мережу Internet і розширюватися новими можливостями і особливостями.

Здатність розробляти інтегральні схеми в середовищі програмування, з можливістю швидкого компілювання і налагодження змінює традиційний спосіб проектування FPGA з використанням інструментальних засобів автоматизованого проектування. Технологія Java забезпечує швидкий цикл компіляції, виконання та перевірки програм. Засоби розробки, що містять JVM, забезпечують контроль програмних застосунків на стадії виконання. Завдяки наявності JVM досягається мобільність мови Java, забезпечується абстрагованість компільованих Java-програм від апаратної платформи й операційної системи.

Використання мови Java має багато переваг над традиційними статичними методологіями проектування. Серед них – перевірка проекту на рівні двійкових потоків, гнучкість при роботі у середовищі моделювання і налагодження проекту, взаємодія з JBits API [13].

JBits API, на відміну від традиційних мов апаратних описів, підтримує RTR, забезпечуючи можливість швидкого створення і реконфігурації схем Xilinx типу FPGA серії Virtex під час виконання програмного застосування за допомогою прямого доступу до конфігурації двійкового потоку.

JBits API та програмне забезпечення, яке входить до його складу, BoardScore, XHWIF та Device Simulator надають нову методологію для проектування, розробки і перевірки інтегральних схем [14]. Можливості й особливості XHWIFServer і BoardScore полегшують віддалену розробку і налагодження проекту. Крім того, система JBits може використовуватись як автономний інструмент або як база для створення інших інструментів [15].

Спираючись на виконаний вище аналіз сучасного стану ключових розділів Reconfigurable computing, можна сказати, що реконфігуровна обробка даних, сучасні ПЛІС і розроблені на їх базі засоби обчислювальної техніки представляють собою останні досягнення в області комп'ютерної техніки та мікроелектроніки.

Реконфігуровні пристрої є основою для побудови систем керування та засобів обчислювальної техніки нових поколінь.

ПЕРЕЛІК ПОСИЛАНЬ

1. Seeking Solutions in Configurable Computing / *W. H. Mangione-Smith, B. Hutchings, D. Andrews, A. DeHon* / Available at <http://www.ecs.umass.edu>.
2. *Newport D., Bouldin D.* Using Configurable Computing Systems / Available at <http://www.cs.utk.edu>.
3. *Arnold M.* Reconfiguring an FPGA-based RISC for LNS arithmetic, 7th International Reconfigurable Technology Conference / Available at <http://www.vcc.com/RTC/RTC.html>.
4. FPGA Compiler II / FPGA Express, VHDL Reference Manual, Version 1999.05.- Synopsys Inc., 1999.- 454 p.
5. Xilinx CORE Generator System / Available at <http://www.xilinx.com>.
6. *Shirazi N., Ballagh J.* Put Hardware in the Loop with Xilinx System Generator for DSP, Xcell 05/27/03 / Available at http://www.xilinx.com/publications/xcellonline/xcell_47.
7. *Hwang J.* Xilinx System Generator / Available at http://www.xilinx.com/publications/xcell_43.
8. ChipScope Pro 6.1i / Available at http://www.xilinx.com/ise/verification/chipscope_pro.htm.
9. *Westfeldt W.* Internet Reconfigurable Logic / Available at http://www.xilinx.com/publications/xcell_31.
10. *Вебер Д.* Технология Java в подлиннике. – С-Пб.: БХВ-Петербург, 2001. – 1104 с.
11. *Цишевский В.* Язык и архитектура Java / Available at <http://www.citforum.ru/internet/iinet96/17.shtml>.
12. *Singh S., James-Roxby P.* Lava and JBits: From HDL to Bitstream in Seconds / Available at http://www.xilinx.com/labs/satnam/fccm2001_singh_jamesroxby.pdf.
13. *Ballagh J., Athanas P., Keller E.* Java Debug Hardware Models using JBits / Available at <http://www.ccm.ece.vt.edu>.
14. *Guccione S., Levi D., Sundararajan P.* JBits: Java based interface for reconfigurable computing / Available at <http://klabs.org/richcontent/MAPLDCon99/Abstracts/sundararajan.doc>.
15. JBits 3.0 SDK for Virtex-II / Available at – <http://www.xilinx.com/labs/projects/jbits>.

Надійшла 02.02.2004

Після доробки 17.09.2004

The question of today's Reconfigurable computing – mechanisms of computer hardware design based on FPGA analysis is done also static design method with use of automated design tools and method with use of Java language analysis are done. Perspectives of mechanism with use of Java are considered, which are specified over traditional design methods: bitstream-level debug and simulation, RTR support, interaction with instrumental package for system design on IRL technology – JBits.

Выполнен анализ современного состояния проблемы Reconfigurable computing – механизмы разработки компьютерных архитектур, базирующихся на интегральных схемах типа FPGA, а также анализ статического метода проектирования с использованием инструментальных средств автоматизированного проектирования и метода с использованием языка Java. Рассмотрены перспективы механизма с использованием Java, обусловленные преимуществами над традиционными методами проектирования: проверка и моделирование проекта на уровне двоичных потоков, поддержка RTR, взаимодействие с инструментальным пакетом для проектирования систем по технологии IRL – JBits.

ТЕОРІЯ І МЕТОДИ АВТОМАТИЧНОГО УПРАВЛІННЯ

ТЕОРИЯ И МЕТОДЫ АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ

THEORY AND METHODS OF AUTOMATIC CONTROL

УДК 681.515

И. А. Орловский

НАСТРОЙКА МАСШТАБИРУЮЩИХ КОЭФФИЦИЕНТОВ ФАЗЗИ-РЕГУЛЯТОРА

Для повышения чувствительности фаззи-регулятора (ФР) при нахождении входных переменных в зоне насыщения граничных термов предлагается уточнение выхода регулятора использованием масштабирующих коэффициентов (МК) входных и выходных переменных. Приводится один из вариантов определения МК. Рассмотрена настройка коэффициентов с учетом динамики системы. Дан конкретный пример уточнения выхода регулятора.

Одним из наиболее важных вопросов расчета и настройки ФР является составление необходимого набора правил и определение оптимальных МК для входных и выходной переменных ФР, работающего в соответствии с составленным набором правил. Составление правил опирается на рассуждения (логическое мышление) человека и его поведение при возникновении ошибок в достижении цели. Количество таких правил ограничено, и не должно превышать несколько десятков. При этом необходимый минимальный набор правил может рекурсивно повторяться, позволяя получить оптимальные МК входных и выходной переменных, а тем самым заданную точность управления. Будем исходить из того, что заданный набор правил не является абсурдным и позволяет при правильной настройке МК входных и выходной переменных приводить к снижению ошибки управления.

Задача расчета ФР, а также его адаптивная настройка достаточно часто рассматривается в литературе. В [1] авторы выполняют настройку фаззи ПИД регулятора с помощью генетических алгоритмов, что требует просмотра очень большого количества вариантов. В [2] исследуется скорость настройки (обучения) нечетких систем при различных методах дефаззификации. Делается

вывод, что наилучшим методом дефаззификации при построении прикладных нечетких систем является метод центра тяжести. Настройка диапазонов изменения термов входных и выходных переменных не рассматривается. В [3] авторы исследуют различные методы импликации, агрегации и дефаззификации, а также параметры и количество функций принадлежности входных и выходных переменных, различные фаззи правила. Приводятся некоторые рекомендации, «позволяющие иметь определенную ориентировку выбора параметров фаззи-блока». В [4] рассматривается устойчивость системы с ФР, основанном на объединении множеств моделей системы и принципа нечеткой логики для обеспечения перехода от одной модели к другой. В [5] приводится структура ФР типа Такаги-Сугено-Канга, в которой выход есть линейная комбинация двух модальных регуляторов. Такой подход обеспечивает плавный переход от одной формы настройки к другой. Однако, как отмечает автор, не позволяет учитывать ограничения, накладываемые на промежуточные координаты. Здесь же вводятся в функционал качества специальные функции наказания, учитывающие выход координат системы за допустимые значения. Для повышения качества предлагается изменение границ функций принадлежности ФР.

Для осуществления качественного управления необходима достаточная чувствительность (отношение изменения выходной переменной к вызвавшему его изменению входной переменной за короткий промежуток времени) выхода ФР на всем диапазоне изменения его входных переменных. В ФР достаточно часто присутствуют зоны изменения входной переменной либо на-

боры значений входных переменных, где выходное значение регулятора остается неизменным (чувствительность ФР равно нулю), что снижает качество управления. В этом случае, а также при входных значениях, приближающихся к нулю, необходимо уточнение выхода регулятора. Уточнение выхода регулятора при нахождении значений входных величин вне зоны насыщения граничных термов может быть выполнено с использованием рекурсивного повторения того же набора правил для уменьшенного диапазона входных переменных, в котором они находятся в рассматриваемый момент времени [6]. Нахождение входных переменных в зоне насыщения граничных термов является сигналом о необходимости настройки МК.

Целью статьи является разработка метода настройки МК ФР в процессе управления для заданного набора правил.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Рассмотрим настройку МК входных и выходной переменных для наиболее простого ФР, имеющего один выход и два входа. Структурная схема ФР с МК приведена на рис. 1. Входными переменными ФР являются ошибка регулирования выходной координаты ($x_{1вх}$) и ее производная ($x_{2вх}$). Выход регулятора – $z_{вых}$. На рис. 1 m_1 и m_2 – МК на входе регулятора для входных переменных $x_{1вх}$ и $x_{2вх}$ и m_z – МК на выходе регулятора для выходной переменной $z_{вых}$. При этом

$$x_{1m} = m_1 \cdot x_{1вх}, x_{2m} = m_2 \cdot x_{2вх}, z_{вых} = m_z \cdot z_m \quad (1)$$

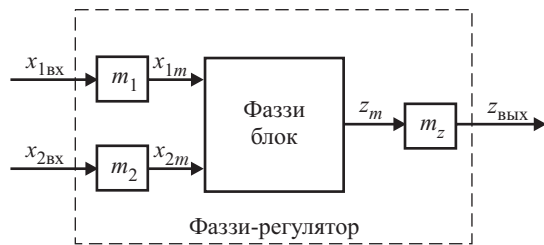


Рисунок 1 – Структурная схема фаззи-регулятора с МК

Термы входных переменных задаются в нормированном диапазоне изменения от -1 до +1. Входные сигналы $x_{1вх}$ и $x_{2вх}$ умножаются на соответствующие коэффициенты m_1, m_2 , что при устойчивой работе системы обеспечивает нахождение входных переменных x_{1m} и x_{2m} в принятом рабочем интервале.

В фаззи-блоке каждой входной и выходной переменной соответствуют пять лингвистических термов:

- PL – Positive Large (положительная большая);
- PS – Positive Small (положительная малая);
- Z – Zero (нуль);
- NS – Negative Small (отрицательная малая);
- NL – Negative Large (отрицательная большая).

Форма граничных функций PL и NL принимается трапецидальной. Форма остальных функций – треугольная. Треугольная и трапецидальная формы функций принадлежности наиболее простые, что уменьшает затраты машинного времени при расчетах. Функции принадлежности размещаются таким образом, что значения входных и выходных переменных принадлежат всегда двум термам внутри рабочего интервала или одному трапецидальному терму вне рабочего интервала. Фаззи правила, аналогичные [7], представлены в таблице 1. Выход ФР вычисляется по алгоритму Мамдани.

Таблица 1 – Фаззи правила

		x_{2m}				
		PL	PS	Z	NS	NL
x_{1m}	PL	PL	PL	PS	PS	Z
	PS	PL	PS	PS	Z	NS
	Z	PL	PS	Z	NS	NL
	NS	PS	Z	NS	NS	NL
	NL	Z	NS	NS	NL	NL

НАСТРОЙКА МК БЕЗ УЧЕТА ДИНАМИКИ

До выполнения настройки МК m_1, m_2, m_z устанавливаются равными 1. Выполняется запуск системы, и если значения входных переменных не находятся в рабочем интервале, то для уточнения выхода выполняется рекурсивное повторение всех правил вывода для части диапазона входных переменных, что достигается пересчетом (настройкой) входных и выходной переменных регулятора [6].

Если значения входных переменных находятся в зоне насыщения граничных термов, то можно выполнить рекурсию в сторону увеличения рабочего интервала входных переменных. В этом случае рабочий интервал входной переменной представляется, как диапазон одного внутреннего нулевого терма для этой переменной, что приводит к новому расположению всех термов этой переменной. Чтобы не изменять диапазоны термов, можно выполнить изменение МК на входе блока для этой переменной. Если значение z_m принадлежит максимальному значению граничного терма, тогда весь диапазон по z можно представить как внутренний терм, и пересчитать положение остальных термов для z .

При не настроенном ФР возникает ситуация, когда две входные переменные x_{1m}, x_{2m} и выходная переменная z_m находятся вне рабочего интервала (рис. 2). Выполняя настройку ФР, будем исходить из того, что используемые правила при оптимальном подборе рабочих интервалов переменных позволяют уменьшить ошибку управления.

Для увеличения рабочих интервалов входных переменных, находящихся в зоне насыщения граничных термов, выполним первую рекурсию формирования выхода фаззи-блока. Для этого рабочий интервал входных переменных представим, как диапазон изменения одного нулевого Z терма. Проведем следующее.

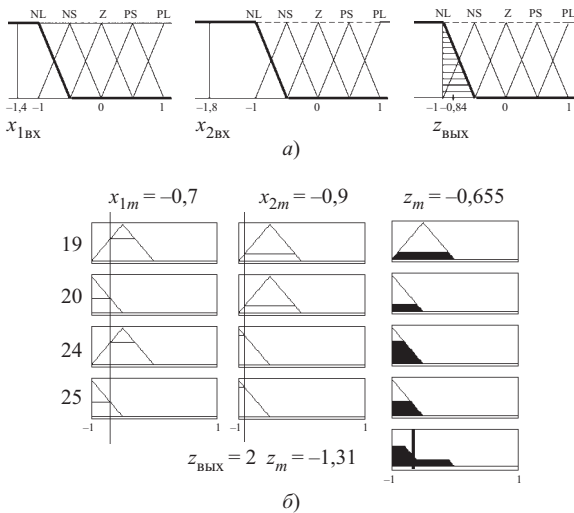


Рисунок 2 – Определение выхода ФР, когда входные переменные имеют одинаковый знак и находятся вне рабочего интервала:

а – без МК; б – уточнение выхода ФР с МК

1. Рассчитаем МК входных и выходной переменных с учетом того, что диапазон изменения нулевого терма составляет k -ю часть рабочего интервала переменной, получим $m_1 = k_1 = 0,5$; $m_2 = k_2 = 0,5$; $m_z = k_z = 0,5$ для термов, представленных на рис. 2.

2. Вычислим значения переменных на входе фаззи-блока x_{1m} , x_{2m} по зависимостям (1)

$$x_{1m} = x_{1вх} \cdot m_1, x_{2m} = x_{2вх} \cdot m_2. \quad (2)$$

3. Определим, с учетом МК, новое значение выхода фаззи-блока z_m , используя те же правила и алгоритм получения результата. Если x_{1m} и x_{2m} вошли в рабочий интервал, то данное правило для термов входных и выходной переменных (рис. 2, а) и системы правил (таблица 1) реализуется с помощью четырех правил (рис. 2, б).

4. Определим МК выходной переменной из уравнения

$$m_z = 1/\sqrt{m_1 \cdot m_2}. \quad (3)$$

5. Рассчитаем значение выхода ФР $z_{вх}$ по зависимости (1)

$$z_{вх} = m_z \cdot z_m. \quad (4)$$

Полученное значение $z_{вх} = -1,31$ после рекурсии, согласно рис. 2, б, превышает значение выхода регулятора до рекурсии $(-0,84)$. Это можно объяснить тем, что значения входных переменных находятся в глубине зоны насыщения на некотором расстоянии от границы рабочего интервала. Если x_{1m} и x_{2m} находятся в зоне насыщения граничных термов, то данный алгоритм повторяется.

Рассмотрим вариант, когда входные переменные принадлежат зоне насыщения граничных термов, а значение выходной переменной принадлежит внутреннему терму

(рис. 3). При выполнении первой рекурсии новые значения переменных на входе фаззи-блока и значение выхода рассчитываются по формулам (2)–(4), при этом $m_1 = 0,5$, $m_2 = 0,5$, $m_z = 0,5$. Из рисунка 3, б видно, что выход $z_{вх} = 0,306$ отличен от нуля и имеет небольшое положительное значение (до рекурсии (рис. 3, а) $z_{вх}$ был равен нулю). Это объясняется тем, что значение входной переменной x_{2m} находилось глубже в зоне насыщения положительного граничного терма, чем значение входной переменной x_{1m} в зоне насыщения отрицательного граничного терма. Если вычисленные значения x_{1m} и x_{2m} по пункту 2 находятся в зоне насыщения граничных термов, то данный алгоритм повторяется.

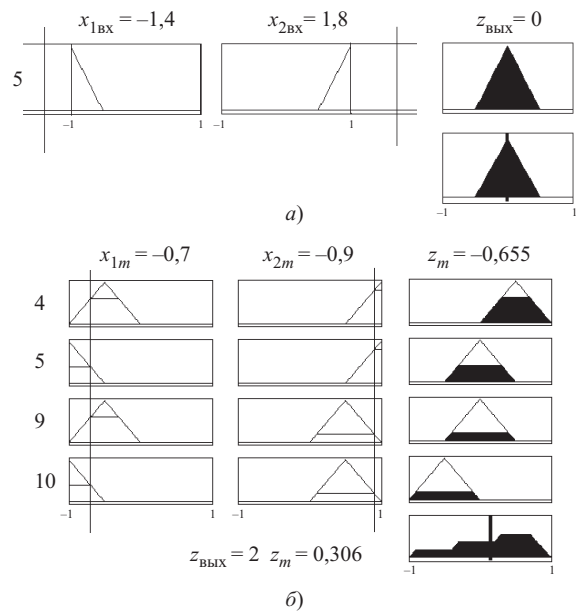


Рисунок 3 – Определение выхода ФР, когда входные переменные разного знака и находятся вне рабочего интервала:

а – без МК; б – уточнение выхода ФР с МК

Рассмотрим вариант, когда значение одной входной переменной $x_{1вх}$ принадлежит зоне насыщения граничного терма, а значение второй входной переменной $x_{2вх}$ и выходной переменной принадлежат внутренним термам (рис. 4). При выполнении первой рекурсии новые значения входных переменных рассчитываются по формулам (2), при этом $m_1 = 0,5$; $m_2 = 1$. Если x_{1m} вышло из зоны насыщения граничного терма, то данное правило реализуется с помощью четырех правил (рис. 4, б), из которых определяется z_m , а далее рассчитывается $z_{вх}$. Значение выхода в этом случае изменяется с $0,379$ на $0,524$, что вызвано учетом глубины насыщения переменной x_{1m} .

Пусть значение входной переменной $x_{1вх}$ находится в рабочем интервале и плавно изменяет свое значение в сторону отрицательной границы рабочего интервала, значение второй входной переменной при этом неизменно. До достижения границы рабочего интервала происходит плавное изменение выходной переменной фаззи-блока. При вхождении в зону насыщения изменяется МК переменной x_1 , что приводит к резкому изменению

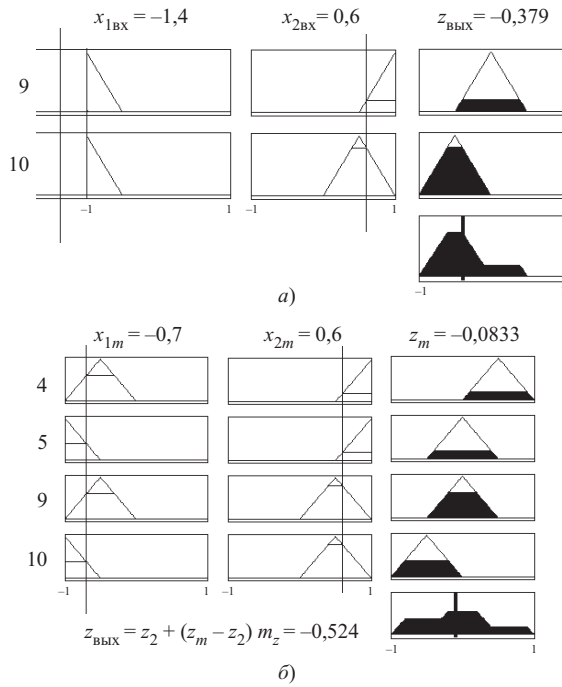


Рисунок 4 – Определение выхода ФР для условия, когда $x_{1вх}$ – вне рабочего интервала, $x_{2вх}$ и $z_{вх}$ – внутри рабочего интервала:

а – без МК; б – уточнение выхода ФР с МК

выходной координаты z_m фаззи-блока (рис. 5). Для качественного управления выход ФР должен изменяться монотонно без резких скачков. Плавный переход на перегибе граничного терма можно осуществлять по специальным уравнениям, либо используя подобную фаззи систему. Пересчет выхода ФР по уравнению

$$z_{вх} = z_2 + (z_m - z_1) \cdot m_z \quad (5)$$

при изменении m_1 обеспечивает нахождение значения выхода на одной из линий, выходящих из точки А (рис. 5). Наклон этих линий зависит от значения МК m_z . МК m_z для более толстой линии (рис. 5) вычисляется по формуле (3). Для рассматриваемого случая $m_z = \sqrt{2}$.

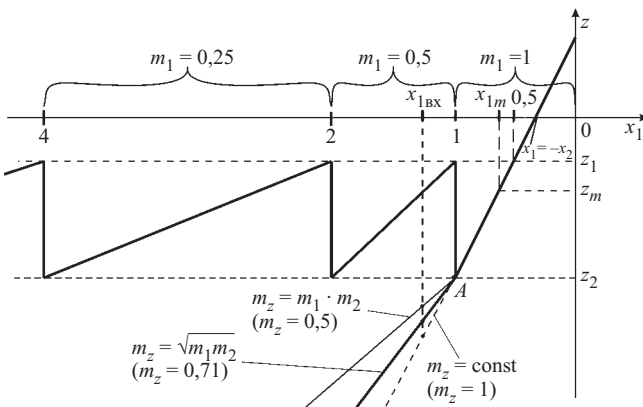


Рисунок 5 – Выход фаззи-блока при изменении переменной x_1 и ее МК m_1 , коэффициент $m_2 = 1$

В уравнении (5) z_1 – выход фаззи-блока, рассчитанный для заданного входного сигнала $x_{2вх}$ и входного сигнала $x_{1вх}$, равного граничному значению рабочего интервала, при этом МК m_1 не изменяется и равен 1; z_2 – выход фаззи-блока рассчитанный для заданного входного сигнала $x_{2вх}$ и входного сигнала $x_{1вх}$, равного граничному значению рабочего интервала, при этом МК m_1 изменяется и для приведенного примера равен 0,5; z_m – выход фаззи-блока для заданных входных переменных $x_{1вх}$, $x_{2вх}$ и измененном МК m_1 ($m_1 = 0,5$).

Рассмотренная выше настройка МК позволяет уточнить выход (повысить чувствительность) ФР, когда одна или обе входных переменных находятся вне рабочего интервала. При работе ФР в замкнутой системе управления необходим такой выбор МК, который обеспечивает устойчивость работы всей системы. Рассмотрим один из вариантов определения МК с учетом динамики системы.

НАСТРОЙКА МК ВХОДНЫХ ПЕРЕМЕННЫХ С УЧЕТОМ ДИНАМИКИ

Изменение ошибки управления и производных ошибки управления представляется в фазовом пространстве. Для двух переменных $x_{1вх}$ и $x_{2вх}$ их взаимное изменение отображается на фазовой плоскости (рис. 6). Каждому терму входных переменных ФР соответствует определенная зона на фазовой плоскости. Пересечение зон определяет соответствующую комбинацию входных переменных. Эта комбинация соответствует правилу, по которому определяется вид выходного терма. На рис. 6 показано движение точки на фазовой плоскости, значение координат которой принадлежат зонам насыщения граничных термов.

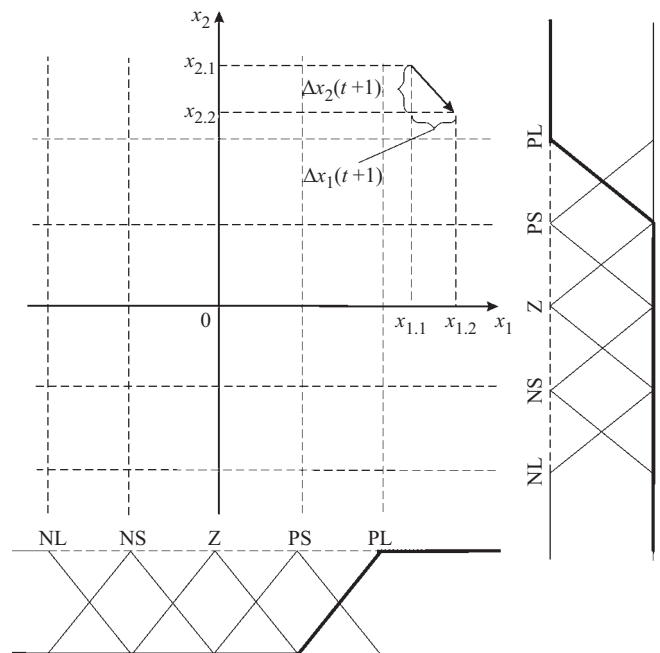


Рисунок 6 – Движение точки на фазовой плоскости вне рабочего интервала входных переменных

Для устойчивой работы системы необходимо, чтобы при движении точки по фазовой плоскости происходило уменьшение радиуса вектора [8], то есть уменьшение функции $L_2 = x_1^2 + x_2^2$. Повлиять на движение точки на фазовой плоскости можно изменением соответствующего влияния на выход ФР входных величин $x_{1\text{ВХ}}$ и $x_{2\text{ВХ}}$. Если фаззи-блок использует в качестве входных переменных не только первую производную ошибки управления, но и производные более высоких порядков, то рассуждая аналогичным образом, можно сформулировать требования к соотношению МК входных переменных ФР.

Переменные фазового пространства приближаются к началу координат, если следующие функции являются убывающими [8]:

$$\begin{aligned} L_2 &= x_1^2 + x_2^2, \\ L_3 &= x_1^2 + x_2^2 + x_3^2, \\ &\dots \\ L_n &= x_1^2 + x_2^2 + x_3^2 + \dots + x_{n-1}^2 + x_n^2, \end{aligned}$$

где x_1 – ошибка управления, x_2, x_3, \dots, x_n – первая, вторая, ..., $(n-1)$ -я производные ошибки управления. Выполнения условия убывания функций $L_i, i = \overline{2, n}$ можно достигнуть соответствующим выбором МК.

Для системы, имеющей только два входных сигнала, соотношение между МК m_1 и m_2 может быть определено по приращениям входных переменных следующим образом. Пусть за короткий интервал времени Δt произошло изменение переменной $x_{1\text{ВХ}}$ на $\Delta x_{1\text{ВХ}}$ с $x_{1.1\text{ВХ}}$ на $x_{1.2\text{ВХ}}$ и переменной $x_{2\text{ВХ}}$ на $\Delta x_{2\text{ВХ}}$ с $x_{2.1\text{ВХ}}$ на $x_{2.2\text{ВХ}}$, как показано на рис. 7. С учетом МК изменения входных переменных произойдут с $x_{1.1m}$ до $x_{2.1m}$ и с $x_{2.1m}$ до $x_{2.2m}$. На рис. 7

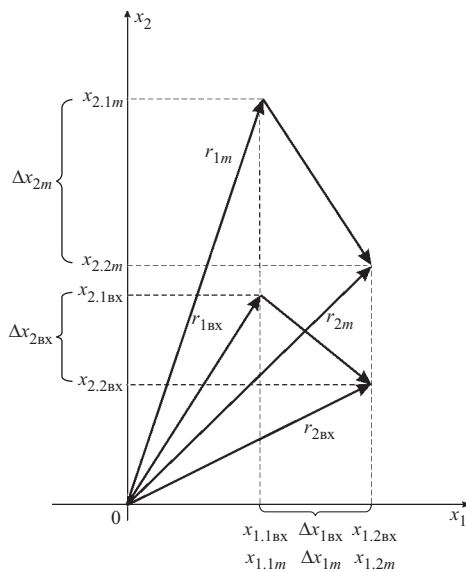


Рисунок 7 – Изменение положения точки на фазовой плоскости за время Δt

$m_1 = 1, m_2 = 2$. Используя выражения (1) и построения на рисунке 7, можно определить разность между радиус-векторами r_{2m} и r_{1m} за время Δt следующим образом:

$$\begin{aligned} (r_{1m})^2 &= (m_1)^2 \cdot (x_{1.1\text{ВХ}})^2 + (m_2)^2 \cdot (x_{2.1\text{ВХ}})^2, \\ (r_{2m})^2 &= (m_1)^2 \cdot (x_{1.2\text{ВХ}})^2 + (m_2)^2 \cdot (x_{2.2\text{ВХ}})^2, \\ \Delta x_{1\text{ВХ}} &= x_{1.2\text{ВХ}} - x_{1.1\text{ВХ}}, \quad \Delta x_{2\text{ВХ}} = x_{2.2\text{ВХ}} - x_{2.1\text{ВХ}}, \\ (r_{2m})^2 - (r_{1m})^2 &= (m_1)^2 \cdot \Delta x_{1\text{ВХ}} \cdot (2x_{1.1\text{ВХ}} + \Delta x_{1\text{ВХ}}) + \\ &\quad + (m_2)^2 \cdot \Delta x_{2\text{ВХ}} \cdot (2x_{2.1\text{ВХ}} + \Delta x_{2\text{ВХ}}), \end{aligned}$$

необходимо выполнение условия $(r_{2m})^2 - (r_{1m})^2 < 0$, отсюда

$$\begin{aligned} (m_1)^2 \cdot \Delta x_{1\text{ВХ}} \cdot (2x_{1.1\text{ВХ}} + \Delta x_{1\text{ВХ}}) + \\ + (m_2)^2 \cdot \Delta x_{2\text{ВХ}} \cdot (2x_{2.1\text{ВХ}} + \Delta x_{2\text{ВХ}}) < 0. \end{aligned}$$

Из неравенства следует:

$$\begin{cases} \text{Если } \Delta x_{2\text{ВХ}} \cdot (x_{2.1\text{ВХ}} + \Delta x_{2\text{ВХ}}) < 0, \\ \text{тогда } (m_2)^2 > (m_1)^2 \cdot \frac{\Delta x_{1\text{ВХ}} \cdot (2x_{1.1\text{ВХ}} + \Delta x_{1\text{ВХ}})}{-\Delta x_{2\text{ВХ}} \cdot (2x_{2.1\text{ВХ}} + \Delta x_{2\text{ВХ}})}, \\ \text{Если } \Delta x_{2\text{ВХ}} \cdot (x_{2.1\text{ВХ}} + \Delta x_{2\text{ВХ}}) > 0, \\ \text{тогда } (m_2)^2 < (m_1)^2 \cdot \frac{\Delta x_{1\text{ВХ}} \cdot (2x_{1.1\text{ВХ}} + \Delta x_{1\text{ВХ}})}{-\Delta x_{2\text{ВХ}} \cdot (2x_{2.1\text{ВХ}} + \Delta x_{2\text{ВХ}})}. \end{cases} \quad (6)$$

Корректировка МК выполняется, если не выполняются условия (6).

Если промежуток времени Δt устремить к нулю, то $\Delta x_{1\text{ВХ}}$ и $\Delta x_{2\text{ВХ}}$ также будут стремиться к нулю. Это позволяет для малых значений Δt упростить неравенства (6) до вида

$$\begin{cases} \text{Если } \Delta x_{2\text{ВХ}} \cdot x_{2.1\text{ВХ}} < 0, \text{ тогда } (m_2)^2 > (m_1)^2 \cdot \Delta_2, \\ \text{Если } \Delta x_{2\text{ВХ}} \cdot x_{2.1\text{ВХ}} > 0, \text{ тогда } (m_2)^2 < (m_1)^2 \cdot \Delta_2, \end{cases}$$

$$\text{где } \Delta_2 = \frac{\Delta x_{1\text{ВХ}} \cdot x_{1.1\text{ВХ}}}{-\Delta x_{2\text{ВХ}} \cdot x_{2.1\text{ВХ}}}.$$

Из полученных неравенств определяется коэффициент соотношения a_2 таким образом, чтобы неравенства выполнялись с необходимым запасом ϑ_2 , например для запаса 20% $\vartheta_2 = 1,2, a_2 = \vartheta_2 \sqrt{|\Delta_2|}$.

В этом случае m_2 определяется из выражения

$$\begin{cases} \text{Если } \Delta x_{2\text{ВХ}} \cdot x_{2.1\text{ВХ}} < 0, \text{ тогда } m_2 = a_2 \cdot m_1, \\ \text{Если } \Delta x_{2\text{ВХ}} \cdot x_{2.1\text{ВХ}} > 0, \text{ тогда } m_2 = (m_1/a_2). \end{cases} \quad (7)$$

Если ввести новую переменную a'_2 , такую, что

$$\begin{cases} \text{Если } \Delta x_{2\text{ВХ}} \cdot x_{2.1\text{ВХ}} < 0, \text{ тогда } a'_2 = a_2, \\ \text{Если } \Delta x_{2\text{ВХ}} \cdot x_{2.1\text{ВХ}} > 0, \text{ тогда } a'_2 = 1/a_2, \end{cases}$$

или, что то же самое, $a'_2 = a_2^{-\text{sign}(\Delta x_{2\text{ВХ}} \cdot x_{2.1\text{ВХ}})}$, тогда

$$m_2 = a'_2 \cdot m_1. \quad (8)$$

Из выражения (8) следует, что МК одной входной переменной может быть вычислен из значения МК второй входной переменной.

Необходимые условия соотношения коэффициентов для трех входных переменных x_1, x_2 и x_3 определяются следующим образом:

$$(m_1)^2 \cdot \Delta x_{1ВХ} \cdot x_{1.1ВХ} + (m_2)^2 \cdot \Delta x_{2ВХ} \cdot x_{2.1ВХ} + (m_3)^2 \cdot \Delta x_{3ВХ} \cdot x_{3.1ВХ} < 0.$$

Используя (7) и (8), получим

$$(m_1)^2 \cdot (\Delta x_{1ВХ} \cdot x_{1.1ВХ} + (a'_2)^2 \cdot \Delta x_{2ВХ} \cdot x_{2.1ВХ}) + (m_3)^2 \cdot \Delta x_{3ВХ} \cdot x_{3.1ВХ} < 0.$$

Введя аналогичные обозначения: ϵ_3 – запас МК m_3 ,

$$\Delta_3 = \frac{\Delta x_{1ВХ} \cdot x_{1.1ВХ} + (a'_2)^2 \cdot \Delta x_{2ВХ} \cdot x_{2.1ВХ}}{-\Delta x_{3ВХ} \cdot x_{3.1ВХ}}, \quad a_3 = \epsilon_3 \sqrt{|\Delta_3|}, \quad \text{а также } a'_3 = a_3^{-\text{sign}(\Delta x_{3ВХ} \cdot x_{3.1ВХ})},$$

получим выражение для МК третьей входной переменной:

$$m_3 = a'_3 \cdot m_1.$$

Аналогичным образом, могут быть вычислены МК для всех входных переменных. МК для x_n определяется следующим образом:

$$\Delta_n = \frac{\Delta x_{1ВХ} \cdot x_{1.1ВХ} + (a'_2)^2 \cdot \Delta x_{2ВХ} \cdot x_{2.1ВХ} + \dots + \Delta x_{n-1ВХ} \cdot x_{n-1.1ВХ}}{-\Delta x_{nВХ} \cdot x_{n.1ВХ}},$$

$$a_n = \epsilon_n \sqrt{|\Delta_n|}, \quad a'_n = a_n^{-\text{sign}(\Delta x_{nВХ} \cdot x_{n.1ВХ})}, \quad m_n = a'_n \cdot m_1,$$

где ϵ_n – запас МК m_n .

Если входные переменные регулятора изменяются таким образом, что соответствующая функция L_i, i – число входных переменных, уменьшается, то изменения МК не выполняются. Если функция L_i возрастает, то определяются желаемые коэффициенты соотношений a'_2, a'_3, \dots, a'_n , из значений которых делается вывод о желаемом управляющем воздействии в данный момент времени. При этом необходимо осуществить такое воздействие, чтобы система направила свое движение в сторону уменьшения функции L_i . Предлагается при выработке управляющего воздействия использовать на входе ФР МК, удовлетворяющие требованиям устойчивости выходной координаты. При этом качество управления во многом будет зависеть от динамики объекта.

НАСТРОЙКА МК ВЫХОДНОЙ ПЕРЕМЕННОЙ С УЧЕТОМ ДИНАМИКИ

Если МК входных переменных вычисляются один через другой, то для оптимальной настройки регулятора необходимо найти оптимальное соотношение между одним из МК входной переменной, например m_1 , и МК

выходной переменной. Если на входы ФР подаются не только ошибка управления и ее первая производная, но 2-я, 3-я, ..., $(n-1)$ -я производные, то МК для этих производных определяются с использованием необходимых значений разностей.

Из уравнения (1) для приращения выходной переменной можно записать $\Delta z_{\text{ВЫХ}} = m_z \cdot \Delta z_m$. При постоянном рабочем интервале выходной переменной изменение выхода выполняется за счет изменения МК Δm_z по формуле

$$\Delta z_{\text{ВЫХ}} = \Delta m_z \cdot z_m, \tag{9}$$

где Δm_z определяется из формулы

$$\Delta m_z = m_z \cdot (\Delta z_m / z_m).$$

Настройку МК выходной переменной можно осуществить с помощью еще одного фаззи-блока. Так как изменение выхода осуществляется изменением МК Δm_z (9), то можно сразу составить правила для определения приращения МК выходной переменной Δm_z . Изменение Δm_z будем определять через изменения входных переменных $\Delta x_{1ВХ}$ и $\Delta x_{2ВХ}$ ФР.

Рассуждения при составлении правил могут быть следующие. Если x_{1m} и x_{2m} находятся в зоне насыщения (вне рабочего интервала) и при этом $\Delta x_{1ВХ}$ и $\Delta x_{2ВХ}$ положительны, то необходимо увеличить воздействие на объект, увеличив выход ФР. Для этого увеличивается МК выходной переменной путем задания приращения Δm_z , например, положительно большим (PL). Если x_{1m} и x_{2m} находятся в зоне насыщения, при этом $\Delta x_{2ВХ}$ близко к нулю, то приращение Δm_z делается положительно малым (PS). Если x_{1m} и x_{2m} в зоне насыщения при этом $\Delta x_{2ВХ}$ отрицательно, то приращение Δm_z необходимо сделать близким к нулю (Z). Если x_{1m} не находится в зоне насыщения, то с учетом пересчета МК x_{2m} также не должно находиться в зоне насыщения. МК m_1, m_2 устанавливаются первоначально равными 1 либо задаются другие значения, исходя из заданной точности управления. Если $\Delta x_{1ВХ}$ и (или) $\Delta x_{2ВХ}$ находится в рабочем интервале, то каждое из утверждений с учетом различных комбинаций термов для $\Delta x_{1ВХ}$ и $\Delta x_{2ВХ}$ дает ряд правил. Правила для нахождения приращения МК Δm_z представлены в таблице 2.

Таблица 2

Δm_z	$\Delta x_{2ВХ}$					
		PL	PS	Z	NS	NL
$\Delta x_{1ВХ}$	PL	PL	PL	PS	PS	Z
	PS	PL	PS	PS	Z	NS
	Z	PL	PS	Z	NS	NL
	NS	PS	Z	NS	NS	NL
	NL	Z	NS	NS	NL	NL

Данные правила являются повторением правил для входных переменных x_{1m} и x_{2m} , где определялось z_m . В случае объекта высокого порядка, если, используя правила для входных переменных и правила первой рекурсии приращений, не удастся получить заданное

качество управления, то аналогичным образом последовательно вводятся правила второй, третьей и так далее рекурсий приращений до достижения заданного качества управления. С помощью использования различной глубины рекурсии приращений МК вычисляются приращения МК различных порядков следующим образом:

$$\begin{aligned} \Delta m_{z_{\text{ВЫХ}}} &= f_1(\Delta x_{1_{\text{ВХ}}}, \Delta x_{2_{\text{ВХ}}}), \\ \Delta^2 m_{z_{\text{ВЫХ}}} &= f_2(\Delta^2 x_{1_{\text{ВХ}}}, \Delta^2 x_{2_{\text{ВХ}}}), \\ &\dots \\ \Delta^n m_{z_{\text{ВЫХ}}} &= f_n(\Delta^n x_{1_{\text{ВХ}}}, \Delta^n x_{2_{\text{ВХ}}}), \end{aligned}$$

где $f_1(), f_2(), \dots, f_n()$ – функции, реализующиеся фаззи-блоками.

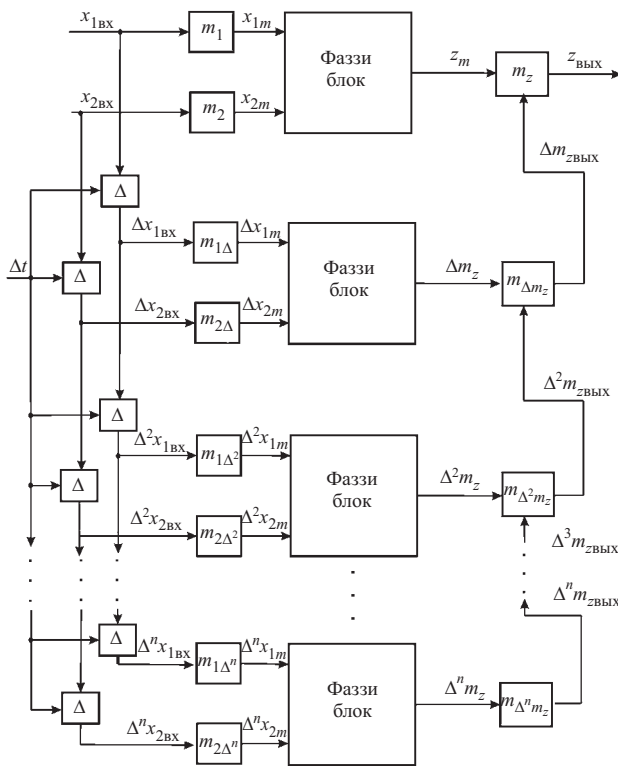


Рисунок 8 – Функциональная схема рекурсивной настройки МК выходной переменной

Функциональная схема ФР с рекурсивной настройкой МК выходной координаты представлена на рисунке 8. Блок со значком Δ вычисляет разность входного сигнала за время Δt .

МОДЕЛИРОВАНИЕ УТОЧНЕНИЯ ВЫХОДА ФР

Уточнение выхода ФР при нахождении одной или двух входных переменных вне рабочего интервала моделировалось в пакете Fuzzy Logic системы Matlab. Для

удобства пронумеруем правила, расположив в ячейках таблицы номера правил (таблица 3).

Таблица 3

		x_{2m}				
		PL	PS	Z	NS	NL
x_{1m}	PL	1	6	11	16	21
	PS	2	7	12	17	22
	Z	3	8	13	18	23
	NS	4	9	14	19	24
	NL	5	10	15	20	25

Результаты моделирования для входных значений $x_{1_{\text{ВХ}}} = -1,4$ и $x_{2_{\text{ВХ}}} = -1,8$ представлены на рассмотренном ранее рисунке 2. При выработке выходного сигнала активными являются правила 19, 20, 24, 25. Входным переменным $x_{1_{\text{ВХ}}} = -1,4$ и $x_{2_{\text{ВХ}}} = 1,8$ соответствует рисунок 3. При выработке выходного сигнала активными являются правила 4, 5, 9, 10. Для значений входных переменных $x_{1_{\text{ВХ}}} = -1,4$ и $x_{2_{\text{ВХ}}} = 0,6$ результаты приведены на рисунке 4. Без изменения МК активными являются правила 5, 10 (рис. 4, а). После изменения МК активны правила 4, 5, 9, 10 (рис. 4, б).

Рассмотренные примеры уточнения выхода ФР с помощью введения МК подтвердили правильность методики настройки МК.

Практическая реализация настройки МК с учетом динамики объекта является задачей дальнейших исследований.

ВЫВОДЫ

Предложенная методика настройки МК входных и выходных переменных при нахождении значений входных переменных вне рабочего интервала позволяет повысить чувствительность ФР и, следовательно, уточнить его выход. Рассмотрен способ уточнения МК с учетом динамики объекта. Выполненное математическое моделирование для конкретных условий подтвердило правильность предложенной настройки МК.

ПЕРЕЧЕНЬ ССЫЛОК

1. K. S. Tang, K. F. Man, G. Chen, S. Kwong An Optimal fuzzy PID Controller // IEEE Transactions on Industrial Electronics. Vol. 48, No. 4, August 2001. Pp. 757–765.
2. Ротштейн А.П., Штовба С.Д. Влияние методов дефаззификации на скорость настройки нечеткой модели // Кибернетика и системный анализ. № 5. 2002. С. 169–176.
3. Клепиков В.Б., Глебов О.Ю., Моисеенко П.Л. Влияние фаззи данных и правил на выходные характеристики фаззи регулятора. // Техн. электродинамика. 2002. С. 43–48.
4. Лозинський А.О. Критерії стійкості систем з нечітким регулятором. Вестник НТУ «ХПИ». Выпуск № 10. Т. 2. Серия «Электротехника, электроника и электропривод». Проблемы автоматизированного электропривода. Теория и практика. Харьков, 2003. С. 510–511.
5. Лозинський А.О. Формування керуючих впливів електромеханічних систем на основі принципів нечіткої логіки // Радіоелектроніка, інформатика, управління. 2000. № 1. С. 156–161.
6. Орловский И.А. Уточнения выхода фаззи-регулятора с помощью рекурсии // Радіоелектроніка, інформатика, управління. 2004. № 1. С. 154–160.

7. Калашников В.І., Палис Ф., Лозинський О.Ю. Основы фазы-логики та фазы-регулювання. Донецьк, Магдебург, Львів, 2000. С. 86.
8. Попов Е.П. Теория нелинейных систем автоматического регулирования и управления. М.: Наука, 1988. С. 255.

Надійшла 30.12.2003
Після доробки 04.06.2004

Для підвищення чутливості фаззі-регулятора (ФР) при знаходженні вхідних змінних у зоні насичення кордонних термів пропонується уточнення виходу регулятора завдяки використанню масштабних коефіцієнтів (МК) вхідних і вихідних змінних. Наводиться один з варіантів визначен-

ня МК. Розглянуто настроювання коефіцієнтів з врахуванням динаміки системи. Наведено конкретний приклад уточнення виходу регулятора.

For increasing of the sensitivity of the fuzzy-regulator while finding input variables in the zone of saturation of bordering terms it's offered a more precise definition of output of the regulator by using zoom coefficients of input and output variables. Here is given one of the variants of finding the zoom coefficients. Here is considered the tuning of coefficients with consideration of dynamics of the system. A real example of the amplification of the regulator output is considered.

УДК 681.511.46

Е. А. Шушляпин, А. Е. Работнов

УПРАВЛЕНИЕ ТЕРМИНАЛЬНЫМИ НЕЛИНЕЙНЫМИ МНОГОИНДЕКСНЫМИ ДИСКРЕТНЫМИ СИСТЕМАМИ МЕТОДОМ КОНЕЧНОГО СОСТОЯНИЯ

Предложена модификация метода конечных состояний для синтеза терминальных управлений многомерными нелинейными системами с математическими моделями в виде совокупностей многоиндексных рекуррентных выражений.

ВВЕДЕНИЕ И ПОСТАНОВКА ЗАДАЧИ

Актуальной задачей теории управления является разработка управляющих методов и алгоритмов для систем с распределенными параметрами, описываемых дифференциальными уравнениями в частных производных (систем трубопроводного транспорта, линий электропередач, управления температурными режимами печей и др.). Несмотря на большое количество задач и математических моделей такого рода, арсенал соответствующих методов управления, особенно для нелинейных систем, невелик. Наиболее часто используется прием приближенной замены уравнений в частных производных обыкновенными дифференциальными уравнениями (например, с помощью метода «прямых» [1]) с существенным расширением размерности фазового пространства, что является недостатком данного подхода. Другой способ, применимый к линейным стационарным системам простого вида, основан на применении преобразования Лапласа [2]. Для некоторых задач оказывается возможным применение вариационного исчисления, принципа максимума, динамического программирования [3]. Предпринимаются попытки получения в аналитическом виде приближенных решений систем уравнений в частных производных [4], на основе которых также можно решать задачу поиска управлений.

В данной работе терминальную задачу управления системой с распределенными параметрами предлагается решать на основе представления ее конечно-разностной

схемой в виде многоиндексной системы рекуррентных соотношений и последующим применением к ней дискретного метода конечного состояния. Метод конечного состояния [5] является представителем методов для решения обратной задачи динамики и основан на использовании переменных и моделей конечного состояния. Данный метод разработан для нормальных дифференциальных многомерных нелинейных систем с гладкими функциями правых частей [6], а также одноиндексных конечно-разностных [7] и смешанных (гибридных) нелинейных систем с аддитивным управлением. Результаты для смешанных систем будут опубликованы в трудах IV Международной конференции «Идентификация систем и задачи управления» SICPRO'05 (Москва, январь 2005 г.). Достоинством перехода от дифференциальной к конечно-разностной схеме применительно к методу конечного состояния является существенное ослабление требований к функциям правых частей соответствующих нормальных форм уравнений. Если в дифференциальном случае требуется непрерывность и дифференцируемость правых частей, то в конечно-разностном случае указанные функции должны быть лишь конечны в каждой точке соответствующей области определения. С точки зрения практических применений конечно-разностная схема удобнее ее дифференциального аналога. Кроме того, решению поставленной задачи применительно к дифференциальной форме должно предшествовать получение аналога формулы Алексева В.М. [8], на которую опирается метод конечного состояния в его непрерывной форме. Такой аналог для уравнений в частных производных нам неизвестен. Для упрощения последующих выкладок рассмотрим вначале двухиндексный случай, соответствующий уравнениям в частных производных с двумя независимыми переменными.

Итак, рассмотрим двухиндексную терминальную задачу с аддитивным управлением в нормальной форме [3]:

$$J = J(x(N_1, N_2)) \rightarrow J^*, \quad (1)$$

$$x(i, j-1) = \Phi_1(i-1, j-1, x(i-1, j-1)) + B_1(i-1, j-1) \cdot u_1(i, j), \quad (2)$$

$$x(i-1, j) = \Phi_2(i-1, j-1, x(i-1, j-1)) + B_2(i-1, j-1) \cdot u_2(i, j), \quad (3)$$

где $i=1, 2, \dots, N_1$; $j=1, 2, \dots, N_2$; J^* – желаемое значение терминального критерия J ; $x(i, 0)$, $x(0, j)$ – заданные «левые» граничные условия; x , Φ_1 , Φ_2 – n -мерные векторы; B_1 – $(n \times n_1)$ -мерная матрица при управлении u_1 , B_2 – $(n \times n_2)$ -мерная матрица при управлении u_2 .

ПОЛУЧЕНИЕ ДВУХИНДЕКСНОЙ МОДЕЛИ КОНЕЧНОГО СОСТОЯНИЯ

Метод конечного состояния основан на использовании переменных конечного состояния (ПКС), зависящих в дискретном двухиндексном случае от двух пар индексов. Уравнения для ПКС как функции первой пары индексов определяются из однородного уравнения, следующего из (1)–(3). В то же время для определения искомого управления ПКС нужны как функции второй пары индексов при конечных значениях первой пары N_1, N_2 соответственно. Вектор ПКС $\bar{x}(N_1, N_2, i-1, j-1, x(i-1, j-1))$ имеет смысл конечного состояния однородной системы, следующей из (1)–(3) и имеющей в дискретные моменты $i-1, j-1$ состояние $x(i-1, j-1)$. Оказывается, что как функция второй пары индексов ПКС зависит от управления (показано ниже), что позволяет при рациональных требованиях к поведению ПКС как функции второй пары индексов его определять. Систему рекуррентных соотношений, определяющей ПКС как функцию второй пары индексов, мы и называем моделью конечного состояния.

Определим переменную конечного состояния (ПКС) как решение однородного уравнения, следующего из (1)–(3), в смысле функции первой пары индексов (первое уравнение – функция k , второе уравнение – функция l).

$$\begin{aligned} \bar{x}(k, l-1, i-1, j-1, x(i-1, j-1)) = \\ = \Phi_1[k-1, l-1, \bar{x}(k-1, l-1, i-1, j-1, x(i-1, j-1))]; \end{aligned} \quad (4)$$

$$\begin{aligned} \bar{x}(k-1, l, i-1, j-1, x(i-1, j-1)) = \\ = \Phi_2[k-1, l-1, \bar{x}(k-1, l-1, i-1, j-1, x(i-1, j-1))], \end{aligned} \quad (5)$$

где $k=i, i+1, \dots, N_1$; $l=j, j+1, \dots, N_2$;

$$\bar{x}(i-1, j-1, i-1, j-1, x(i-1, j-1)) = x(i-1, j-1). \quad (6)$$

В общем случае, по смыслу ПКС, имеет место равенство

$$\bar{x}(k, l, k, l, x(k, l)) = x(k, l), \quad (7)$$

так как первая пара индексов означает дискретный момент вычислений по рекуррентным формулам (4)–(7), а вторая – момент начала при начальных условиях

$x(k, l)$. Следовательно, при совпадающих парах выход будет совпадать с начальным условием.

Далее получим алгоритм для расчета ПКС как функций второй пары индексов. Вначале – для второй пары индексов, отличающихся на единицу от значений первой пары. Это следует из (2) и (6) при $k=i$, $l=j$.

$$\begin{aligned} \bar{x}(i, j-1, i-1, j-1, x(i-1, j-1)) = \\ = \Phi_1[i-1, j-1, \bar{x}(i-1, j-1, i-1, j-1, x(i-1, j-1))] = \\ = \Phi_1[i-1, j-1, x(i-1, j-1)] = \\ = x(i, j-1) - B_1(i-1, j-1) \cdot u_1(i, j); \end{aligned} \quad (8)$$

$$\begin{aligned} \bar{x}(i-1, j, i-1, j-1, x(i-1, j-1)) = \\ = \Phi_2[i-1, j-1, \bar{x}(i-1, j-1, i-1, j-1, x(i-1, j-1))] = \\ = \Phi_2[i-1, j-1, x(i-1, j-1)] = \\ = x(i-1, j) - B_2(i-1, j-1) \cdot u_2(i, j). \end{aligned} \quad (9)$$

В соотношениях (8) и (9) заменяем в соответствии с (7) переменные состояния переменными конечного состояния

$$\begin{aligned} x(i, j-1) = \bar{x}(i, j-1, i, j-1, x(i, j-1)); \\ x(i-1, j) = \bar{x}(i-1, j, i-1, j, x(i-1, j)). \end{aligned}$$

После подстановок и переноса слагаемых в левые части имеем

$$\begin{aligned} \bar{x}(i, j-1, i, j-1, x(i, j-1)) - \bar{x}(i, j-1, i-1, j-1, x(i-1, j-1)) = \\ = B_1(i-1, j-1) \cdot u_1(i, j); \end{aligned} \quad (10)$$

$$\begin{aligned} \bar{x}(i-1, j, i-1, j, x(i-1, j)) - \bar{x}(i-1, j, i-1, j-1, x(i-1, j-1)) = \\ = B_2(i-1, j-1) \cdot u_2(i, j). \end{aligned} \quad (11)$$

Здесь уменьшаемое и вычитаемое в левой части имеют одинаковые первые пары $\{i, j-1\}$ в (10) и $\{i-1, j\}$ в (11), а индексы вторых пар отличаются на единицу: в (10) – первый индекс, в (11) – второй индекс.

Теперь в (10) и (11) будем увеличивать индексы первой пары: в (10) – первый индекс, в (11) – второй индекс. Первое слагаемое (10) запишем с учетом (4) в виде

$$\begin{aligned} \bar{x}(k |_{i, (l-1) |_{j-1}, (i-1) |_{i, j-1}, x(i, j-1)) = \\ = \Phi_1[i-1, j-1, \bar{x}(i-1, j-1, i, j-1, x(i, j-1))]. \end{aligned}$$

Теперь увеличим i первой пары ПКС на единицу. Тогда

$$\begin{aligned} \bar{x}(i+1, j-1, i, j-1, x(i, j-1)) = \\ = \Phi_1[i, j-1, \bar{x}(i, j-1, i, j-1, x(i, j-1))]. \end{aligned} \quad (12)$$

То же сделаем и со вторым слагаемым (10), где первый индекс j второй пары на единицу меньше.

$$\begin{aligned} \bar{x}(i, j-1, i-1, j-1, x(i-1, j-1)) = \\ = \Phi_1[i-1, j-1, \bar{x}(i-1, j-1, i-1, j-1, x(i-1, j-1))]. \end{aligned}$$

Увеличивая первый индекс первой пары, получаем

$$\begin{aligned} \bar{x}(i+1, j-1, i-1, j-1, x(i-1, j-1)) = \\ = \Phi_1[i, j-1, \bar{x}(i, j-1, i-1, j-1, x(i-1, j-1))]. \end{aligned} \quad (13)$$

Таким образом, левая часть (10) с увеличенным на единицу первым индексом первой пары получена. Эта левая часть уже не будет равна $B_1 u_1$. Вычислим ее, учитывая выражение (2). Найдем разность (12) и (13).

$$\begin{aligned} & \bar{x}(i+1, j-1, i-1, j-1, x(i-1, j-1)) - \\ & - \bar{x}(i+1, j-1, i-1, j-1, x(i-1, j-1)) = \\ & = \Phi_1[i, j-1, \bar{x}(i, j-1, i, j-1, x(i, j-1))] - \\ & - \Phi_1[i, j-1, \bar{x}(i, j-1, i-1, j-1, x(i-1, j-1))]. \end{aligned} \quad (14)$$

Используя соотношения (7) и (2), запишем

$$\begin{aligned} & \bar{x}(i, j-1, i, j-1, x(i, j-1)) = x(i, j-1) = \\ & = \Phi_1[i-1, j-1, x(i-1, j-1)] + B_1(i-1, j-1) \cdot u_1(i, j). \end{aligned}$$

Используя соотношения (4) при $k=i$, $l=j$ и (6), запишем

$$\begin{aligned} & \bar{x}(i, j-1, i, j-1, x(i, j-1)) = \\ & = \Phi_1[i-1, j-1, \bar{x}(i-1, j-1, i-1, j-1, x(i-1, j-1))] = \\ & = \Phi_1[i-1, j-1, x(i-1, j-1)]. \end{aligned}$$

Таким образом, разность (14) представлена в виде

$$\begin{aligned} & \bar{x}(i+1, j-1, i, j-1, x(i, j-1)) - \\ & - \bar{x}(i+1, j-1, i-1, j-1, x(i-1, j-1)) = \\ & = \Phi_1[i, j-1, \Phi_1[i-1, j-1, x(i-1, j-1)]] + \\ & + B_1(i-1, j-1) \cdot u_1(i, j) - \\ & - \Phi_1[i, j-1, \Phi_1[i-1, j-1, x(i-1, j-1)]]. \end{aligned} \quad (15)$$

Введем обозначение

$$\Phi_1^{[k]}[i, j, x(i, j)] = \underbrace{\Phi_1[i, j, \Phi_1[i-1, j, \dots, \Phi_1[i-k, j, x(i, j)]]]}_{k \text{ раз}}, \quad (16)$$

$$\Phi_1^{[1]} \equiv \Phi_1. \quad (17)$$

С учетом (16)–(17) разность (15) при $k=l$ равна

$$\begin{aligned} & \bar{x}(i+k, j-1, i, j-1, x(i, j-1)) - \\ & - \bar{x}(i+k, j-1, i-1, j-1, x(i-1, j-1)) = \\ & = \Phi_1^{[k]}[i, j-1, \Phi_1[i-1, j-1, x(i-1, j-1)]] + \\ & + B_1(i-1, j-1) \cdot u_1(i, j) - \\ & - \Phi_1^{[k]}[i, j-1, \Phi_1[i-1, j-1, x(i-1, j-1)]]. \end{aligned} \quad (18)$$

В общем случае при $k > l$ выражение (18) сохраняется. Для $k = N_1 - i$ получаем, перенося второе слагаемое в правую часть,

$$\begin{aligned} & \bar{x}(N_1, j-1, i, j-1, x(i, j-1)) = \\ & = \bar{x}(N_1, j-1, i-1, j-1, x(i-1, j-1)) + \\ & = \Phi_1^{[N_1-i]}[i, j-1, \Phi_1[i-1, j-1, x(i-1, j-1)]] + \\ & + B_1(i-1, j-1) \cdot u_1(i, j) - \\ & - \Phi_1^{[N_1-i]}[i, j-1, \Phi_1[i-1, j-1, x(i-1, j-1)]]. \end{aligned} \quad (19)$$

Проводя аналогичные действия для второго индекса первой пары (11), получаем следующее выражение:

$$\begin{aligned} & \bar{x}(i-1, N_2, i-1, j, x(i-1, j)) = \\ & = \bar{x}(i-1, N_2, i-1, j-1, x(i-1, j-1)) + \\ & = \Phi_2^{[N_2-j]}[i-1, j, \Phi_2[i-1, j-1, x(i-1, j-1)]] + \\ & + B_2(i-1, j-1) \cdot u_2(i, j) - \\ & - \Phi_2^{[N_2-j]}[i-1, j, \Phi_2[i-1, j-1, x(i-1, j-1)]]. \end{aligned} \quad (20)$$

Выражения (19) и (20) представляют собой рекуррентные формулы для вычисления ПКС как функций второй пары аргументов, а именно: (19) – как функцию i при фиксированном j ; (20) – как функцию j при фиксированном i . Такую систему рекуррентных формул, по аналогии с непрерывным случаем [5], назовем «Дискретной многомерной моделью конечного состояния» (ДММКС). Переменные ДММКС при известных граничных условиях (3) и управлениях $u_1(i, j)$, $u_2(i, j)$ могут быть определены как функции аргументов i и j . Однако в методе конечного состояния ДММКС используются косвенно для получения управления.

ПОЛУЧЕНИЕ УПРАВЛЕНИЯ

В соответствии с методикой конструирования управления методом конечного состояния введем так называемую критериальную функцию (КФ) второй пары индексов, заменив аргумент критерия (1) на ПКС

$$\bar{J}(i, j) = J[\bar{x}(N_1, N_2, i, j, x(i, j))] \rightarrow J^*. \quad (21)$$

В силу свойства (7)

$$\bar{J}(N_1, N_2) = J[\bar{x}(N_1, N_2, N_1, N_2, x(N_1, N_2))] = J[x(N_1, N_2)] \equiv J.$$

Таким образом, при выборе управлений u_1 и u_2 таких, что КФ (21) при $i \rightarrow N_1$, $j \rightarrow N_2$ $\bar{J}(N_1, N_2) \rightarrow J^*$, будет решена и эквивалентная задача (1)–(3).

К сожалению, соотношение (21) пока нельзя использовать ввиду его зависимости от неизвестного $\bar{x}(N_1, N_2, i, j, x(i, j))$. По этой причине определим

$$\bar{J}(k, l, i, j) = J[\bar{x}(k, l, i, j, x(i, j))], \quad (22)$$

которое при $k, i \rightarrow N_1$, $l, j \rightarrow N_2$ в силу (7) стремится к J^* . Частными случаями (21) будут

$$\bar{J}(N_1, j-1, i, j-1) = J[\bar{x}(N_1, j-1, i, j-1, x(i, j-1))] \quad (23)$$

и

$$\bar{J}(i-1, N_2, i-1, j) = J[\bar{x}(i-1, N_2, i-1, j, x(i-1, j))], \quad (24)$$

для которых известны \bar{x} (это выражения (18)–(20), зависящие от управляющих воздействий u_1 и u_2).

Для определения u_1 и u_2 зададим желаемое поведение критериальной функции \bar{J} в виде пары рекуррентных формул

$$\bar{J}(N_1, j-1, i, j-1) = f_{J_1}[\bar{J}(N_1, j-1, i-1, j-1)], \quad (25)$$

$$\bar{J}(i-1, N_2, i-1, j) = f_{J_2}[\bar{J}(i-1, N_2, i-1, j-1)] \quad (26)$$

при граничных условиях

$$\bar{J}(N_1, 0, i, 0) = J[\bar{x}(N_1, 0, i, 0, x(i, 0))], \quad (27)$$

$$\bar{J}(0, N_2, 0, j) = J[\bar{x}(0, N_2, 0, j, x(0, j))]. \quad (28)$$

Для получения $\bar{x}(N_1, 0, i, 0, x(i, 0))$ воспользуемся (4) при $k=1, N_1, l=1, j=1$ для каждого $i=1, 2, \dots, N_1$, то есть расчет по (4) производится вначале для $i=1$, и т. д. до $i=N_1$. Аналогично, но по (5) при $k=1, l=1, N_2, i=1$ для каждого $j=1, 2, \dots, N_2$ рассчитывается $\bar{x}(0, N_2, 0, j, x(0, j))$. Следует отметить, что для получения управления выражения (27) и (28) не требуются. Они нужны лишь для построения трехмерных графиков, если есть потребность в визуальном сравнении поверхностей желаемой и фактической критериальных функций. Для получения же управления нужны f_{J_1} и f_{J_2} , которыми необходимо задаться таким образом, чтобы решение (25) и (26) при достижении $i \rightarrow N_1, j \rightarrow N_2$ подходило (точно или приближенно) к J^* .

Определим

$$f_{J_1} = \bar{J}(N_1, j-1, i-1, j-1) + \frac{1}{T_{u_1}}(J^* - \bar{J}(N_1, j-1, i-1, j-1)), \quad (29)$$

$$f_{J_2} = \bar{J}(i-1, N_2, i-1, j-1) + \frac{1}{T_{u_2}}(J^* - \bar{J}(i-1, N_2, i-1, j-1)). \quad (30)$$

Тогда (25) и (26) при $T_{u_1}, T_{u_2} \geq 1$ будут иметь решение, по близкой к экспоненте форме приближающееся к J^* (при $T_{u_1}, T_{u_2} < 1$ разностные уравнения (29) и (30) неустойчивы).

При $T_{u_1} = T_{u_2} = 1$ имеем из (25), (26) с учетом (29)–(30)

$$\bar{J}(N_1, j-1, i, j-1) = J^*, \quad \bar{J}(i-1, N_2, i-1, j) = J^*$$

при любых i и j , больших нуля. То есть задание $T_{u_1} = T_{u_2} = 1$ за один шаг выводит критериальную функцию на уровень J^* , который далее не изменяется. Напомним, что здесь речь идет о теоретическом поведении КФ.

Теперь вернемся к получению управления. Для этого приравняем правые части (22), (23) и (25) и (26) (или в частном случае (29) и (30)) соответственно с подстановкой (19) и (20) вместо \bar{x} в (22) и (23). В итоге получим два скалярных нелинейных алгебраических уравнения относительно векторов $u_1(i, j)$ и $u_2(i, j)$, которые решаются для каждого i и j . Для частного случая f_{J_1}, f_{J_2} имеем

$$\begin{aligned} & J[\bar{x}(N_1, j-1, i-1, j-1, x(i-1, j-1))] + \\ & + \Phi_1^{[N_1-i]}(\arg_{11}(u_1)) - \Phi_1^{[N_1-i]}(\arg_{12}(u_1)) = \\ & = J[\bar{x}(N_1, j-1, i-1, j-1, x(i-1, j-1))] + \\ & + \frac{1}{T_{u_1}} \left[J^* - J[\bar{x}(N_1, j-1, i-1, j-1, x(i-1, j-1))] \right], \quad (31) \end{aligned}$$

$$\begin{aligned} & J[\bar{x}(i-1, N_2, i-1, j-1, x(i-1, j-1))] + \\ & + \Phi_2^{[N_2-j]}(\arg_{21}(u_2)) - \Phi_2^{[N_2-j]}(\arg_{22}(u_2)) = \\ & = J[\bar{x}(i-1, N_2, i-1, j-1, x(i-1, j-1))] + \\ & + \frac{1}{T_{u_2}} \left[J^* - J[\bar{x}(i-1, N_2, i-1, j-1, x(i-1, j-1))] \right], \quad (32) \end{aligned}$$

$$\begin{aligned} \arg_{11} &= N_1, j-1, \Phi_1[i-1, j-1, x(i-1, j-1)] + B_1(i-1, j-1) \cdot u_1(i, j), \\ \arg_{12} &= N_1, j-1, \Phi_1[i-1, j-1, x(i-1, j-1)], \\ \arg_{21} &= i-1, N_2, \Phi_2[i-1, j-1, x(i-1, j-1)] + B_2(i-1, j-1) \cdot u_2(i, j), \\ \arg_{22} &= i-1, N_2, \Phi_2[i-1, j-1, x(i-1, j-1)]. \end{aligned}$$

Присутствующие в (31) и (32) аргументы $\bar{x}(N_1, j-1, i-1, x(i-1, j-1))$, $\bar{x}(i-1, N_2, i-1, j-1, x(i-1, j-1))$ вновь рассчитываются по (4), (5), (6), если $x(i-1, j-1)$ уже известно.

ОБЩИЙ СЛУЧАЙ

Для сокращения записей введем следующие обозначения для множеств индексов:

$$M_p(i) = \{i_1, i_2, \dots, i_p\};$$

$$M_{p \setminus q \setminus +1}(i) = \{i_1, i_2, \dots, i_{q-1}, i_q + 1, i_{q+1}, \dots, i_p\};$$

$$M_{p \setminus q \setminus a}(i) = \{i_1, i_2, \dots, i_{q-1}, a, i_{q+1}, \dots, i_p\}.$$

С учетом принятых обозначений постановка многоиндексной задачи имеет вид

$$J = J(x(M_p(N))) \rightarrow J^*,$$

$$\begin{aligned} x(M_{p \setminus j \setminus +1}(i-1)) &= \Phi_j(M_p(i-1), x(M_p(i-1))) + \\ & + B_j(M_p(i-1)) \cdot u_j(M_p(i)), \end{aligned}$$

где входящие в определение множеств индексов $i_j = 1, 2, \dots, N_j$; $j = 1, 2, \dots, p$; J^* – желаемое значение терминального критерия J ; $x(M_{p \setminus 1 \setminus i_1}(0)), \dots, x(M_{p \setminus p \setminus i_p}(0))$ – заданные «левые» граничные условия; x, Φ_1, \dots, Φ_p – n -мерные векторы; $B_j - (n \times r_j)$ – мерная матрица при управлении u_j .

Следуя схеме вывода (31), (32), получаем для частного случая f_{J_j} следующие p соотношений:

$$\begin{aligned} & J[\bar{x}(M_{p \setminus j \setminus N_j}(i-1), M_p(i-1), x(M_p(i-1))) + \\ & + \Phi_j^{[N_j-i]}(\arg_{j1}(u_j)) - \Phi_j^{[N_j-i]}(\arg_{j2}(u_j))] = f_{J_j}; \\ & f_{J_j} = J \left[\bar{x}(M_{p \setminus j \setminus N_j}(i-1), M_p(i-1), x(M_p(i-1))) \right] + \\ & + \frac{1}{T_{u_j}} \left[J^* - J \left[\bar{x}(M_{p \setminus j \setminus N_j}(i-1), M_p(i-1), x(M_p(i-1))) \right] \right], \quad (33) \\ & j = 1, 2, \dots, p, \end{aligned}$$

где

$$\begin{aligned} \arg_{j1} &= M_{p \setminus j \setminus N_j}(i-1), \Phi_j [M_p(i-1), x(M_p(i-1))] + \\ &+ B_j(M_p(i-1)) \cdot u_j(M_p(i)), \\ \arg_{j2} &= M_{p \setminus j \setminus N_j}(i-1), \Phi_j [M_p(i-1), x(M_p(i-1))]. \end{aligned}$$

ЗАКЛЮЧЕНИЕ

Полученный алгоритм (31), (32) для двухиндексного, (33) для многоиндексного случаев легко реализуем на компьютерах и имеет большие перспективы для расчета управлений системами с распределенными параметрами. При этом возможен расчет не только аддитивных управлений, но и параметрических управлений, а также управляющих воздействий на границах и внутренних слоях пространства переменных. Для этого задача приводится к форме (1)–(3) путем расширения ее размерности и подходящим выбором явных зависимостей от индексов Φ_1, Φ_2, B_1, B_2 .

ПЕРЕЧЕНЬ ССЫЛОК

1. Березин Н.С., Жидков Н.П. Методы вычислений. – Т. 2. – М.: ГИФМЛ, 1962. – 640 с.
2. Деч Г. Руководство к практическому применению преобразования Лапласа и z-преобразования. – М.: Наука, 1971. – 288 с.

3. Лурье К.А. Оптимальное управление в задачах математической физики. – М.: Наука, 1975. – 478 с.
4. Monaco S., Normand-Cirot D. About the solution to nonlinear multi-time-scale differential equations // Proceedings of the III International Conference "System Identification and Control Problems", Moscow, January 28–30, 2004. – М.: ИПУ им. В.А.Трапезникова РАН, 2004. – С. 1212–1223.
5. Шушляпин Е.А. Синтез линейных и нелинейных систем управления конечным положением на основе моделей конечного состояния // Проблемы управления и информатики. – 1997. – №3. С. 10–16.
6. Шушляпин Е.А. Терминальное управление системами с дифференциально-алгебраическими ограничениями методом конечного состояния // Радиоэлектроника. Информатика. Управление. – 2000. – №1. – С. 166–172.
7. Шушляпин Е.А., Подольская О.Г. Управление терминальными нелинейными дискретными системами методом конечного состояния // Радиоэлектроника. Информатика. Управление. – 2003. – №2. – С. 138–142.
8. Алексеев В.М. Об одной оценке возмущений обыкновенных дифференциальных уравнений // Вестн. Москов. ун-та. Сер. 1. Математика, механика. – 1961. – №2. – С. 28–36.

Надійшла 06.10.2004

Запропонована модифікація методу кінцевих станів для синтезу термінальних керувань багатомірними нелінійними системами з математичними моделями у вигляді сукупностей багатойндексних рекурентних виразів.

There was offered a terminal state method modification for terminal control synthesis in multidimensional nonlinear systems with mathematical models in the form of multi-indexed recurrent expression combinations.

КЕРУВАННЯ У ТЕХНІЧНИХ СИСТЕМАХ

УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ СИСТЕМАХ

CONTROL IN TECHNICAL SYSTEMS

УДК 621.365.036

В. В. Осадчий

ОСОБЕННОСТИ АДАПТИВНОГО ПРОГРАММНОГО УПРАВЛЕНИЯ ТЕХНОЛОГИЧЕСКИМИ ПРОЦЕССАМИ ПРОИЗВОДСТВА ОГНЕУПОРНЫХ МАТЕРИАЛОВ

В работе проводится теоретическое обоснование методологии построения взаимосвязанного дуального закона и моделей функционирования иерархических систем адаптивного управления производством огнеупорных материалов на основе закона параметрической и структурной адаптации.

ВВЕДЕНИЕ

Производство огнеупорных материалов для агрегатов металлургии является важнейшим системным обеспечением производства металла и нерудных изделий в сталеплавлении, ферросплавном производстве и других конструкционных материалов для металлургических печей и установок.

Автоматизированное управление технологическими процессами производства никельхромовых и железохромалюминиевых сплавов проводится в воздушной и углеродсодержащей атмосфере, из карбида кремния, молибдена, хромита лантана, диоксида циркония и других компонентов [1]. Технологические процессы производства огнеупорных материалов отличаются значительными системо- и схмотехническими факторами и конструктивным воплощением. В настоящее время созданы и разрабатываются АСУ ТП металлургического назначения: транспортных линий известково-обжигательного цеха, производства кирпича, флюсодоломитного комбината, технологических отделений цеха агломерации и другие [2].

Огнеупорные материалы для футеровок обеспечивают работу десятков и сотен термических установок технологических комплексов только на крупнейших предприятиях Запорожья (Запорожбразив, Мотор Сич, Запорожжкокс, Запорожсталь, Днепропетсталь, Запорожжуглерод, Запорожский алюминиевый комбинат, Запорожферросплав), что обуславливает существенный интерес к вопросам анализа системотехнического обеспечения производства огнеупорных изделий многокомпонентного состава с изменяемой рецептурой и на его основе синтеза схмотехнических решений АСУ ТП.

ОБЩИЕ ПОЛОЖЕНИЯ

В современных условиях важнейшими факторами при создании систем управления технологией являются факторы эффективного использования первичных и вторичных энергоресурсов, экономичного расходования энергетических материалов, высокого качества продукции, структурно-технологической перестройки энергоемких процессов и производств; организация постоянного контроля и разработка мер воздействия на подразделения предприятия; массовое оснащение всех потребителей средствами учета, контроля и регулирования расхода всех видов затрачиваемых энергоресурсов; повышения эффективности управления энергопотреблением и другие [3]. Главным критерием управления производственно-технологическими процессами является стабилизация ре-

жима работы технологической системы в соответствии с заданной программой. Система контролирует работу конвейеров, бункеров-накопителей, питателей, дозаторов и других механизмов, при этом рабочими параметрами являются точность дозирования, производительность агрегатов, разрешающая способность весо-измерительных систем. Операторская станция осуществляет взаимодействие оператора и системы управления в интерактивном режиме через клавиатуру и набор отображаемых на экране монитора унифицированных элементов, составляющих подсистему связи с оператором: мнемосхему, меню, окно ввода данных, информационное окно [4, 5].

Эксплуатация современных АСУ ТП позволит:

- обеспечить координацию управляющих воздействий по всему комплексу параметров процесса, исключить нарушения работы оборудования, предотвратить аварии и увеличить сроки эксплуатации оборудования, снизить расход энергоресурсов за счет оптимизации работы приводов оборудования технологической системы;

- повысить точность измерений, исключив погрешности планиметрической обработки диаграмм при исключении громоздких регистрирующих приборов [5, 6];

- повысить эффективность информационных каналов благодаря представлению и архивированию информации о протекании технологического процесса, ведению протокола технологического режима по сменам с последующим его анализом;

- обеспечить рост культуры производства, освобождение оперативного персонала от непроизводительных потерь рабочего времени за счет возможности анализа качества ведения технологического процесса, получения достоверной информации о состоянии и работе технологического оборудования и регулировать ход технологического процесса в пределах допустимых изменений параметров технологического процесса;

- выдавать сообщения о нарушениях в работе технологического оборудования и отклонениях параметров от заданных значений на рабочую станцию диспетчера-оператора, оперативно вести системный журнал, архивировать системные данные, вести визуальное наблюдение за процессами управления электроприводами и диагностику работы электрооборудования, протоколировать отказы в работе оборудования за требуемый интервал и другие процессы.

Реализация указанных факторов требует использования в АСУ ТП решения новых задач управления [7, 8] на основе перестройки методов планирования и управления. Однако преодоление дуальных трудностей решения многовариантных задач планирования и управления приводят к тому, что разработчики АСУ базируются на простых и примитивных методах планирования процессов управления и регулирования координат, направляя при этом основные усилия на решение задач прямого расчетного характера, используя при этом традиционные принципы планирования от достигнутого и регулирования по отклонению, как правило, связывая с регулированием обычную корректировку.

Экономико-математические методы управления [9] базируются на методах математического программирования

в силу их эффективности и общности моделей, допускающих широкий спектр интерпретаций. В то же время, структура производства вариативна, характеризуется ростом числа рецептов продукции, повышением факторов динамичности и неопределенности. Действие на производственную систему различных помех приводит к тому, что даже хорошо разработанный план на основе применения методов математического программирования может быть неустойчивым. Это вызывает дополнительный расход энергии и материалов, затрат, направленных на коррекцию производственного плана. Данная ситуация снижает доверие к методам математического программирования. Традиционно используемые математические модели [10] задач планирования и регулирования производства только приблизительно описывают реальные экономико-производственные системы с определенной и, как правило, с существенной погрешностью. И даже при этом условии качество принятого решения в большой степени зависит от того, насколько правильно выбран тип модели и достаточно ли информационное обеспечение этой модели. Формирование алгоритмов управления в этих условиях должно осуществляться с помощью решения комплекса экстраполяционных и оптимизационных задач. В связи с этим возникает необходимость как в разработке методов решения отдельных задач, так и в алгоритмическом обеспечении согласования этих решений. Однако, для того чтобы оптимизация стала рабочим инструментом при планировании и регулировании производства, необходимо решить ряд проблем по совершенствованию организации и механизма управления. В данном случае должна быть повышена гибкость структуры производства и алгоритмов управления, что достигается использованием современных подходов к практике организации производства и автоматизированных систем управления на основе ЭВМ.

Целью проводимых нами исследований является решение задачи построения гибкой системы управления, обладающей адаптивностью к изменяющимся условиям производства, на базе реализации возможностей математических методов и средств вычислительной математики. Успешное решение проблемы совершенствования управления производством связано с преодолением следующих трудностей: постановки задачи, адекватной реальному объекту и процессу; оценки и анализа полученных результатов решения задачи с точки зрения их реализации и связи с другими проблемами; внедрения требуемой схемы принятия решений. Для реализации указанных задач на современном этапе развития систем управления производством и технологическими процессами адаптивного характера выделяется параметрическая и структурная адаптация математического обеспечения АСУ ТП. Параметрическая адаптация широко распространена в практике создания библиотеки стандартных программ системы обработки данных [6, 11]. Структурная адаптация обеспечивается с помощью модулей (подпрограмм), обладающих свойствами параметрической адаптации. Количество и разнообразие модулей выбирается достаточным для реализации той или иной функции в пределах области существования системы. В обоих видах присут-

ствуют элементы пассивной и активной адаптации. Элементы, предназначенные для самонастройки математического обеспечения к изменениям параметров, обладают свойствами пассивной адаптации, а элементы, осуществляющие структурную адаптацию информационной базы и математического обеспечения, обладают свойствами активной адаптации. Эволюция и нестационарность производственного объекта приводит к необходимости широкого использования неформальных средств "дооптимизационного" и "послеоптимизационного" анализа, т. е. комплексные методы принятия решения должны включать как процедуры математического программирования, так и процедуры дооптимизационного и послеоптимизационного анализа, а последние должны включать средства имитации и прогнозирования.

Экономикоматематические модели планирования движений системы являются наиболее разработанными в исследованиях проблемы оптимального функционирования экономико-производственных систем, оптимизационные модели имеют не менее ограниченное применение. Это является следствием того, что модели математического моделирования, по которым разрабатывается план управления движением системы, не отражают возможностей и условий его реализации. Реально процедура планирования движения системы должна учитывать возможности реализации плана в процессе его разработки, учитывать прогнозируемые потери, обусловленные необходимостью локализации помех и компенсации отклонений при реализации плана, т. е. удельный вес методов математического программирования в реальной процедуре планирования может не превышать удельный вес формальных и неинформационных средств анализа, а иногда может быть гораздо меньше его. В этом случае применение точных методов решения задач математического программирования, используемых при планировании движений системы, может терять смысл, так как нестационарность среды производства требует применения итерационных методов, обогащенных возможностями учета эволюции данных об объекте, изменения моделей планирования, оценки потерь и анализа реализации плана движений на основе имитационных и прогнозных моделей влияния дестабилизирующих воздействий помех и возмущений. В данном случае требуется адаптация не только методов математического программирования к анализу сложных технологических процессов и производства, но и самого плана движений или управляющего решения к условиям его реализации [10, 11]. К принципиальным особенностям адаптивного планирования управления движением относятся следующие особенности:

– лица, принимающие решения (ЛПР) в области экономико-производственных систем, непосредственно привлекаются к процессу формирования планов и управлению. В данном случае оказывается возможным адаптивный процесс прогнозирования тех или иных решений. Адаптивное планирование предполагает наличие процесса обучения ЛПР. Тем самым обеспечивается сближение процедур планирования и управления с процедурами научного исследования. В современных условиях возникает необходимость предпосылки реализации плана ими-

тационного эксперимента с целью проверки функциональных возможностей системы, который основан на диалоге ЭВМ и человека;

– при адаптивном планировании должны быть разработаны методы эффективного контроля и управления деятельностью системы по реализации плана движений и учета знаний о характеристиках в определенные временные отрезки будущего. Знание может быть определенным, "размытым" и полностью отсутствующим. В первом случае распознавание сравнительно определенных аспектов будущего может оказаться трудной задачей. Их раскрытие требует большого объема исследований. Очевидными они часто становятся только ретроспективно. Во втором случае, если известны статистические характеристики, для оптимизации плановых решений используются различные варианты стохастического программирования. Если же неопределенность не является стохастической, возможно использование элементов теории нечетких множеств и алгоритмов. При сравнительно высокой степени неопределенности и возможности вариантного представления будущего, планирование движений следует осуществлять по вариантам, т. е. для каждого возможного состояния среды в будущем формировать свой план. Используя также игровой подход, гарантированный результат может быть получен на основе экспертных оценок. В случае полного отсутствия знаний или информации о будущем состоянии системы, адаптация системы осуществляется косвенно путем планирования движений по реагированию. Такое планирование движений направлено на создание системы, которая бы могла различать отклонения от ожидаемого и эффективно реагировать на них. В этом случае может быть максимизирована скорость адаптации.

ОСНОВНЫЕ ТЕОРЕТИЧЕСКИЕ РЕЗУЛЬТАТЫ

На основании вышеизложенного можно принять за основной постулат положение, что адаптивная система управления технологическими процессами или производством состоит из двух взаимосвязанных систем: адаптивной системы планирования движений и адаптивной системы регулирования. Основываясь на основных положениях концепции адаптации, следует отметить, что выделенные системы практически идентичны по своей структуре. Функциональная структура каждой из них состоит из следующих взаимосвязанных частей: модели планирования (соответственно регулирования) движений производства; имитационной модели функционирования производственной системы; внутреннего (имитационного) адаптора; внешнего (объектного) адаптора.

Неформальное описание работы адаптивной системы управления будет включать в себя следующие основные элементы. Внешний адаптор на основе анализа характеристик объекта и внешней среды выбирает модель задачи планирования, а также имитационную модель, осуществляя тем самым структурную адаптацию системы управления. Затем по результатам выполнения планов прошлых периодов и прошлых возмущающих воздействий

он подстраивает параметры в модели планирования (регулирующего) и в имитационной модели, включающей модели объекта, среды и системы регулирования. В адаптивной системе планирования, основываясь на полученных параметрах, по модели планирования определяют план управления движением системы и потенциальный эффект. План рассматривается как траектория задания последовательностей плановых состояний, распределенных во времени. По имитационной модели осуществляется имитация реализации плана и оцениваются потери, не позволившие достичь потенциального эффекта. Имитация реализации плана выполняется несколько раз для получения статистически значимых оценок показателей плановых движений. По результатам расчета плана и имитации его выполнения проводятся оценка и анализ приемлемости алгоритмов реализации плана. Если план с учетом его возможной реализации приемлем, то он принимается к реализации. В противном случае внутренний адаптор, основываясь на результатах имитации, подстраивает параметры модели планирования и модели регулирования и работа схемы повторяется, начиная с пересчета плана при новых параметрах. Работа внутреннего адаптора базируется на одном из методов оптимизации в условиях помех.

Фактор "приемлемости" при неформальном описании адаптивной системы управления производством является сложным. Для технических адаптивных систем управления по Д. Дональдсу и Л.А. Задэ система считается приемлемой, если для заданного класса входных воздействий значение ее критерия качества или нескольких критериев попадает в заданную область [12]. Критерий приемлемости в адаптивной системе управления можно представить в виде двух показателей: потенциальный эффект Q , полученный в результате расчета планирования движений и потери Π , образуемые на этапе формирования и обусловленные влиянием помех и необходимостью выбора регулирующих воздействий, направленных на устранение влияния этих помех. Применительно к условиям производства огнеупорных изделий в качестве показателя, который может служить в качестве критерия приемлемости в адаптивной системе планирования производства, является показатель ритмичности работы технологической системы R , определяемый в виде соотношения [10, 11] $R = 1 - \frac{x_{\Pi} - x_{\Phi}^*}{x_{\Pi}}$, x_{Π} , x_{Φ}^* – плановые и фактические значения показателей движения элементов производства соответственно. Фактическое состояние x_{Φ}^* на этапе планирования определяется в результате функционирования произвольной системы, а плановое x_{Π} – в результате решения задачи планирования.

В общем виде за критерий приемлемости может быть принято выражение вида [12] $\Phi = Q - \Pi$, которое подчеркивает отрицательное воздействие потерь. Таким соотношением удобно пользоваться при описании механизма адаптации. Если между показателями, входящими в Q и Π , нет однозначного соответствия, а также не удается привести их к одному показателю, то критерий

приемлемости будет рассматриваться в виде векторного критерия $\vec{\Phi} = \{f_s, s = \overline{1, S}\}$, где f_s – частный критерий оптимальности, S – количество частных критериев. Данным представлением $\vec{\Phi}$ можно пользоваться при описании конкретных алгоритмов. Показатели, определяемые по результатам функционирования системы, имеют вероятностную природу [9], поэтому в качестве управления может выступать либо математическое ожидание [9] $M(\vec{\Phi}) \rightarrow \max$, либо вероятность $P(\Phi \geq \alpha) \rightarrow \max$, ($\alpha - \text{const}$), либо $\alpha \rightarrow \max$, $P(\Phi \geq \alpha) = 1$. Предполагая, что цель движения системы задана как $P(\Phi \geq \alpha) \rightarrow \max$, для дальнейшего анализа вводятся необходимые основные определения. Время принимается дискретным $t = t_0, t_1, t_2, \dots$. Каждый из интервалов времени $[t_k, t_{k+1}]$, в свою очередь, разбивается на N подинтервалов $[\tau_v, \tau_{v+1}]$, $v = \overline{0, N-1}$, $\tau_0 = t_k$; τ_v – моменты времени, принадлежащие интервалам $[t_k, t_{k+1}]$. Величины, связанные с моментом времени t_k , снабжаются индексом k . Величины, связанные с моментом времени τ_v , принадлежащим интервалу $[t_k, t_{k+1}]$, снабжаются индексами k_v (индекс k может быть опущен). Пусть x_k – состояние системы в момент времени t_k . Через x_k обозначается плановая траектория, заданная как $[T+1]$ точками на интервале $[t_k, t_{k+1}]$; $x_k = \{x_v, v = \overline{0, N}\}$. На систему в ходе выполнения траектории x_k действуют помехи: измеряемая r_k и неизмеряемые (не учтенные в моделях) φ_k ; $r_k = \{r_v, v = \overline{0, N}\}$. Несмотря на то, что в каждой конкретной модели смысл φ_k будет различным, используется одинаковый символ φ_k .

Регулирующие воздействия, выдаваемые системой регулирования на интервале $[t_k, t_{k+1}]$ с целью компенсации возмущений, действующих на систему, обозначается как U_k ; $U_k = \{U_v, v = \overline{0, N}\}$. Через w_k обозначаются параметры моделей, входящих в систему планирования и регулирования, $w_k = \{w_v, v = \overline{1, N}\}$. Параметры w_k состоят из двух наборов параметров: параметров, подстраиваемых внутренним адаптором W_k^1 , и параметров, подстраиваемых внешним адаптором W_k^2 , которые, в свою очередь, делятся по принадлежности к различным моделям. В набор параметров W_k^1 входят параметры модели планирования $W_k^{1\Pi}$ и модели регулирования W_k^{1P} , в набор параметров W_k^2 входят параметры модели планирования $W_k^{2\Pi}$, параметры имитационной модели внешней среды W_k^{2B} , имитационной модели производства (объекта регулирования) W_k^{2O} и параметры модели регулирования W_k^{2P} (имитационная модель объекта и модель регулирования входят в качестве составных частей в модель имитации реализации плановых движений системы). Кроме того, если Z_0, Z_1, \dots, Z_t – последовательность элементов множества $\{Z\}$, то упорядоченные наборы (конечные и бесконечные) соответствующих элементов будут обозначаться как $z_s^t = [z_s, \dots, z_t]$, $z_s^\infty = [z_s, z_{s+1}, \dots]$, $0 \leq s \leq t$. Элемент z_s^t при $s > t$ обозначает пустое множество, поэтому $Z_0 = Z_0^0$.

С учетом вышеизложенного вначале рассматривается работа адаптивной системы планирования движений на основе алгоритмов параметрической адаптации. Динамика системы адаптивного планирования определяется контуром внутренней адаптации. Первым по времени работает внешний адаптор и параметры, подстроенные внешним адаптором, используются во всех операторах внутреннего контура адаптации. Системотехническая эволюция объекта описывается уравнением вида $x_{v+1} = X(x_S^V, U_S^V, r_S^V, \varphi_S^V, \xi)$; $X(\cdot)$ – некоторый оператор. Состояние объекта в момент времени τ_{v+1} зависит от предыдущих состояний объекта x_S^V , регулирующих U_S^V и возмущающих воздействий r_S^V , φ_S^V , а также набора параметров объекта ξ . Под параметрами объекта подразумеваются параметры: внешней среды, продукции и производства, включая параметры системы регулирования координат технологии производства. В общем случае оператор $X(\cdot)$ может задаваться алгоритмом различного вида, в том числе стохастическим.

Работа системы планирования движений технологической системы базируется на эволюционном уравнении объекта управления x_{v+1} . Система планирования движений рассчитывает плановую траекторию в дискретные моменты времени t_k на период $[t_k, t_{k+1}]$. Система регулирования осуществляет контроль выполнения объектом плановых заданий и вмешательство в процесс производства с целью стабилизации выходных показателей объекта в дискретные моменты времени τ_v . Расчет плановых траекторий осуществляется по уравнению $x_k = P(Y_k^{\Pi}, Q, W_k^{\Pi})$, где $P(\cdot)$ – оператор (модель) расчета плановой траектории; Y_k^{Π} – учетные данные, полученные подсистемой учета; Q – производительность или набор показателей (потенциальный эффект), характеризующий план χ_k ; W_k^{Π} – модель планирования. С помощью параметров $w_k^{\Pi} = (W_k^{1\Pi}, W_k^{2\Pi})$ учитываются особенности объекта, который характеризуется своим набором параметров ξ .

Для этапа планирования процесс сбора и преобразования информации, осуществляемой подсистемой учета, можно описать сенсорным уравнением [10, 13]: $Y_k^{\Pi} = Y^{\Pi}(x_S^k, \chi_S^{k-1}, U_S^{k-1}, r_S^{k-1}, \varphi_k)$; $S < k$; $Y(\cdot)$ – оператор, отражающий работу задач учета и нормирования, в частности, этим оператором учитываются и потери Π , связанные с неопределенным или неоптимальным функционированием из-за дестабилизирующего воздействия помех r_S^{k-1} , φ_k , U_S^{k-1} – регулирующие воздействия, применяемые системой регулирования для уменьшения влияния помех r_S^{k-1} на плановые траектории; r_S^{k-1} – контролируемые (учитываемые) помехи, действующие в периоды $[t_S, t_{S+1}]$, ..., $[t_{k-1}, t_k]$. Учетные данные y_k^0 задаются в виде функций от времени. Плановая траектория χ_k , полученная по соответствующему выражению для χ_k , поступает в блок, в котором имитируется реализация плана. В результате имитации реализации плана формируются данные, совпадающие по содержанию с y_k^{Π} . Некоторый l -й вариант реализации плановой траектории ге-

нерируется имитационной моделью, работа которой описывается уравнением $\tilde{y}_{k+1}^{le} = U(y_k^{\Pi}, \chi_k, W_k^u, \xi)$; (индекс "u" – имитационный); $U(\cdot)$ – оператор; $w_k = (w_k^{2O}, w_k^{2P}, w_k^{2B})$ – параметры имитационной модели (подстраиваются внешним адаптором); ξ – равномерно распределенная на интервале $[0, 1]$ некоррелированная случайная величина. Параметр $\tilde{y}_{k+1}^{\Pi(l)}$ является многомерным случайным процессом, причем любой наперед заданный тест должен подтверждать гипотезу об эквивалентности выборок $\tilde{y}_{k+1}^{\Pi(l)}$ и Y_{k+1}^{Π} , т. е., имея информацию $\tilde{y}_{k+1}^{\Pi(l)}$, можно подстраивать параметры в операторе планирования $P(\cdot)$. В этом случае подстройка параметров будет осуществляться по соотношению $w_k^{1\Pi(l)} = W1(\bar{Y}_{k+1}^{\Pi(l)}, w_k^{1\Pi(l-1)})$, $l = 1, 2, \dots$; $W_k^{1\Pi(0)} = w_{k-1}^{1\Pi}$; $W1(\cdot)$ – оператор, описывающий работу внутреннего адаптора; $w_k^{1\Pi(l-1)}$ – параметры, полученные внутренним адаптором при формировании плана движения системы в прошлом периоде.

Параметры $w_k^{1\Pi(l)}$ используются в уравнении для χ_k , т. е. параметры w_k^{Π} заменяются на параметры $w_k^{\Pi(l)} = (w_k^{2\Pi}, w_k^{1\Pi(l)})$ и по новому значению проводится имитационное моделирование. Практически вместо оператора χ_k работает оператор $\chi_k^{(l)} = P(y_k^{\Pi}, Q, w_k^{\Pi(l)})$. Данный оператор при $l=0$ совпадает с χ_k . При выборе параметров внутреннего адаптора $W1$ стремится увеличить средний эффект Φ^u , получаемый при имитации функционирования системы: $M(\Phi^u) = M(Q - \Pi^u)$; Π^u – потери, возникающие при функционировании и полученные по имитационной модели $\tilde{y}_{k+1}^{\Pi(l)}$. Значения параметров $w_k^{1\Pi(l)}$, обеспечивающих оптимум функционалу, наиболее полно приспособливают модель планирования к объекту управления, который характеризуется параметрами w^u . Так как все методы оценки параметров являются приближенными, то и план χ_k будет настроен не на реальный объект ξ , а на объект ξ^1 , близкий к ξ . Чтобы подстроить план к объекту ξ , внешний адаптор подстраивает помимо параметров w^u также параметры $w_k^{2\Pi}$. Подстройка этих параметров осуществляется по соотношению $w_{k+1}^2 = W2(y_{k+1}^{\Pi}, w_k^2)$. $W2(\cdot)$ – оператор, описывающий работу внешнего адаптора. При выборе параметров внешний адаптор стремится оптимизировать выражение $M(\Phi) = M(Q - \Pi)$.

Наличие трех групп настраиваемых параметров $(w^1, w^{2\Pi}, w^{2U})$ приводит к необходимости решения в адаптивных системах планирования трех типов задач поиска значений приведенных параметров. Для решения первой задачи вводится некоторая мера $\mu(\{\tilde{y}_k\}, \{y_k\})$ оценки близости законов распределения выборок $\{\tilde{y}_k\}$ и $\{y_k\}$, тогда эта задача, решаемая внешним адаптором, может быть представлена следующим образом: найти параметры настройки w^{2u} , решая задачу вида $w^{2u} = \arg \min \mu(\{\tilde{y}_k\}, \{y_k\})$; $w^{2u} \in W^{2u}$. Вторая задача, решаемая

внешним контуром адаптаций, связана с выбором группы параметров $w^{2\Pi}$ и формализуется следующим образом: $w^{2\Pi} = \arg \max M(Q(\chi(w^{2\Pi})), \Pi)$; $w^{2\Pi} \in W^{2\Pi}$; $M(\cdot)$ – математическое ожидание. Третья задача настройки параметров решается внутренним адаптором. Настраиваются параметры w^1 исходя из решения задачи вида: $w^1 = \arg \max M(\Phi(\chi(w^1, w^{2\Pi}), w^{2\Pi}))$; $w^1 \in W$; $M(\cdot)$ – математическое ожидание величины Φ , полученной по имитационной модели (имитационным реализациям плана).

В целом задачей адаптивной системы планирования движений является получение наиболее приемлемого плана работы системы.

Система приведенных соотношений является адаптивной системой планирования в классе θ по отношению к целевому условию $M(\Phi) = M(Q - \Pi)$, если для любого $\xi \in \theta$ и любого набора начальных значений настраиваемых параметров $w_0^\Pi \in w^\Pi$ существуют такие t^* и l^* , при которых для любых $t \geq t^*$ и $l \geq l^*$ выполняется цель планирования [8].

Формальные соотношения, описывающие работу адаптивной системы регулирования представляются в виде сенсорного уравнения

$$\tilde{y}_v = \tilde{Y}(\tilde{\chi}_s^v, \tilde{u}_s^{v-1}, \chi_k, r_k^v, \varphi_v), \quad s \leq v \leq N; \quad (1)$$

$\tilde{Y}(\cdot)$ – оператор, отражающий работу подсистемы учета на этапе регулирования. Регулирующие воздействия, принимаемые в качестве задающих воздействий на интервале $[\tau_v, \tau_m]$, $\tau_m > \tau_{v+1}$, определяются из следующего соотношения, взаимосвязанного с описываемыми далее процессами имитации и изменения параметров: $\tilde{u}_v^{(n(l+1))} = U(\tilde{y}_v, w^{p(l)})$, где $U(\cdot)$ – оператор, используемый в соотношении $u_v^n = \tilde{U}(\tilde{y}_v, w_v^p, \tilde{Q}_p)$; \tilde{Q}_p – потенциалных эффект, определяемый на этапе регулирования.

Изменяемые параметры w_v^p имеют четыре составляющие. Составляющие w_v^{1p}, w_v^{2p} на этапе регулирования остаются неизменными и численно равны тем значениям, которые были получены на этапе планирования. Параметры, входящие в \tilde{w}_v^{2p} , подстраиваются внешним адаптором (контуром адаптора) этапа регулирования, а \tilde{w}_v^{1p} – внутренним. Алгоритм работы внешнего адаптора этапа регулирования записывается как $\tilde{w}_m^2 = w2(\tilde{y}_v, \tilde{w}_v^2)$, где $\tilde{w}2(\cdot)$ – оператор оценки настраиваемых параметров; $\tilde{w}_m^2 = \{\tilde{w}_{v+1}^2, \dots, \tilde{w}_{v+m}^2\}$; \tilde{w}_v^2 – параметры, изменяемые внешним адаптором этапа регулирования, которые делятся на параметры модели регулирования \tilde{w}_v^{2p} и параметры имитационной модели объекта регулирования w_v^{2u} . Алгоритм работы внутреннего адаптора этапа регулирования записывается в виде рекуррентного уравнения, взаимодействующего с параметром $\tilde{U}_v^{(m(l+1))}$ и имитационными соотношениями:

$$\tilde{w}_{v+m}^{1p(l+1)} = \tilde{w}1(\tilde{y}_v, \tilde{y}_v^{\Pi(l)}, \chi^m, \tilde{w}_{v+m}^{1p(l)}); \quad \tilde{w}_{v+m}^{1p(0)} = \tilde{w}_v^{1p}. \quad (2)$$

Ниже рассматривается алгоритм динамического функционирования внутреннего адаптора системы регулирования параметров производства динасовых, карборундовых и форстеритовых огнеупоров [1], где сенсорное уравнение вида (1) и уравнение внутреннего адаптора (2) представлены нижеследующей системой дифференциальных уравнений:

$$\left. \begin{aligned} \frac{dx_1}{dt} &= -ax_1 + x_1^2 + x_1x_2; & x_1(0) &= d; \\ \frac{dx_2}{dt} &= -bx_2 + cx_1^2 + x_2^2; & x_2(0) &= g. \end{aligned} \right\} \quad (3)$$

Исследование динамики регулирования параметров системы (3) проводится применительно к условиям линии приготовления массы (дозирование, перемешивание, подача массы в бункер прессы НРФ-2500) на основе виброприводов серии JVM. Для указанных механизмов линии коэффициенты системы уравнений (3) равны: $a = -8$; $b = -10$; $c = 0.5$; $d = 0.5$; $g = 0.2$.

Матричное уравнение для системы (3), как результирующее уравнение функций прогноза и коррекции для системы (2) представляется в виде

$$\frac{d[\bar{x}]}{dt} = [A]\bar{x} + [F]\bar{x} = \begin{bmatrix} -8 & 0 \\ 0 & -10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} x_1^2 + x_1x_2 \\ 0.5x_1^2 + x_2^2 \end{bmatrix}, \quad (4)$$

где параметры $x_1 = \omega_1^p$; $x_2 = \omega_2^p$; $x_1(0) = \omega_2^p$; $x_2 = \omega_2^p$.

Алгоритм построения регулирующего сигнала в соответствии с [6, 7] будет иметь вид

$$\bar{x}_{k+1} = [D] \left[\bar{x}_k + \left([E] - \frac{[A]h}{2!} \right) h [\bar{F}](\bar{x}_k, kh) \right], \quad (k = 0, 1, 2, \dots), \quad (5)$$

где [E] – единичная матрица; [A] – матрица спектра вариативности x_i .

Матрица управляющего параметра $[Q(\vec{A}(h))]$ и вектор $\vec{F}(x_0)$ равны

$$D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} -0.8 & 0 \\ 0 & -1 \end{bmatrix} + \begin{bmatrix} 0.32 & 0 \\ 0 & 0.5 \end{bmatrix} = \begin{bmatrix} 0.52 & 0 \\ 0 & 0.5 \end{bmatrix};$$

$$[F(x_0)] = \begin{bmatrix} 0.5^2 + 0.5 * 0.2 \\ 0.5 * 0.5^2 + 0.2^2 \end{bmatrix} = \begin{bmatrix} 0.350 \\ 0.165 \end{bmatrix};$$

$$Q(A(h)) = \left([E] - \frac{[A(h)]}{2!} \right) h = \begin{bmatrix} 0.14 & 0 \\ 0 & 0.15 \end{bmatrix},$$

откуда искомым вектор уравнения динамики определится следующим образом

$$\begin{aligned} \bar{x}_1 &= [D] \left[\bar{x}_0 + \left([E] + \frac{[A][h]}{2} \right) h [F(x_0)] \right] = \\ &= \begin{bmatrix} 0.52 & 0 \\ 0 & 0.5 \end{bmatrix} \left\{ \begin{bmatrix} 0.5 \\ 0.2 \end{bmatrix} + \begin{bmatrix} 0.14 & 0 \\ 0 & 0.15 \end{bmatrix} \begin{bmatrix} 0.350 \\ 0.165 \end{bmatrix} \right\} = \begin{bmatrix} 0.286 \\ 0.112 \end{bmatrix}. \end{aligned} \quad (6)$$

Для каждого шага вектор \bar{x}_2 вычисляется по выражению

$$\bar{x}_2 = [D] \left[\bar{x}_1 + \left([E] + \frac{[A][h]}{2} \right) h [F(\bar{x}_1, kh)] \right] =$$

$$= \begin{bmatrix} 0.52 & 0 \\ 0 & 0.5 \end{bmatrix} \left\{ \begin{bmatrix} 0.286 \\ 0.112 \end{bmatrix} + \begin{bmatrix} 0.14 & 0 \\ 0 & 0.15 \end{bmatrix} \begin{bmatrix} 0.113 \\ 0.005 \end{bmatrix} \right\} = \begin{bmatrix} 0.157 \\ 0.061 \end{bmatrix}$$

и далее управляющие сигналы равны

$$\bar{x}_3 = \begin{bmatrix} 0.084 \\ 0.0321 \end{bmatrix}; \quad \bar{x}_4 = \begin{bmatrix} 0.044 \\ 0.016 \end{bmatrix}; \quad \bar{x}_5 = \begin{bmatrix} 0.023 \\ 0.008 \end{bmatrix}; \quad \bar{x}_6 = \begin{bmatrix} 0.0004 \\ 0.0001 \end{bmatrix} \dots$$

По полученным значениям x_i строится переходная характеристика, представленная на рис. 1.

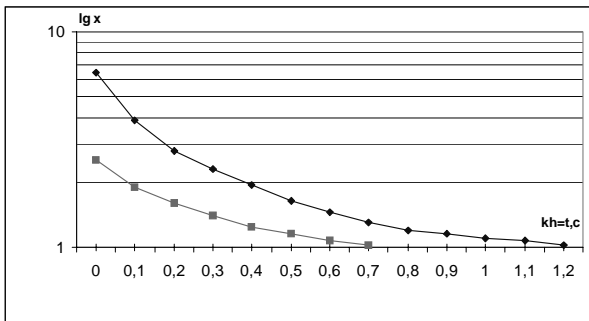


Рисунок 1

После принятия регулирующего воздействия (т. е. процесс адаптации по внутреннему контуру окончен) индекс l опускается. Поскольку на этапе регулирования возможны изменения плановой траектории на интервале $[\tau_v, \tau_m]$, то по аналогии с этапом планирования цель управления представляется в виде потенциального эффекта, получаемого на этапе регулирования \tilde{Q} , и потерь $\tilde{\Pi}$, обусловленных последующей на интервале $[\tau_v, \tau_m]$ работой системы. Потенциальный эффект \tilde{Q} определяется на интервале $[\tau_v, \tau_m]$ относительно траектории, полученной в результате расчета регулирующего воздействия $\tilde{Q} \leq Q$, т. е. целевое условие на этапе регулирования $M(\tilde{\Phi}) = M(\tilde{Q} - \tilde{\Pi}) \rightarrow \max$ может приниматься в качестве цели адаптации во внешнем контуре на этапе регулирования. Аналогично этапу планирования при регулировании также осуществляется имитация реализации регулирующего воздействия, поэтому целью работы внутреннего адаптора этапа регулирования может быть использовано соотношение $M(\tilde{\Phi}^u) = M(\tilde{Q} - \tilde{\Pi}^u) \rightarrow \max$.

Формальные соотношения, обеспечивающие достижения цели $M(\tilde{\Phi}^u)$ и модели системы регулирования, которые используются на этапе регулирования при имитации реализации регулирующих заданий на интервале $[\tau_v, \tau_m]$, аналогичны соотношениям для интервала $[t_k, t_{k+1}]$. Они имеют вид сенсорного уравнения $\bar{y}_\lambda =$

$= \bar{Y}(x_s^\lambda, u_s^{\lambda-1}, \bar{X}_k, \bar{r}_{s,k}^\lambda, \varphi_\lambda)$; $s \leq \tau$; $k\lambda \leq m$. Регулирующие воздействия определяются по соотношению $u_\lambda = U(\tilde{y}_\lambda, w_\lambda^\Pi)$; $\lambda \leq n \leq m$. Работа имитационной модели объекта управления описывается уравнением $\tilde{y}_{\lambda+1} = u(\tilde{y}_\lambda, u_{\lambda-1}, w_\lambda^\Pi)$; $\tilde{y}_\lambda = \bar{y}_\lambda$.

На основании вышеприведенных соотношений и общего алгоритма регулирования формальное определение адаптивных параметров системы регулирования координат сводится к следующему. Система уравнений для $\tilde{y}_v - \bar{y}_{\lambda+1}$ будет адаптивной в классе θ по отношению к целевому условию $M(\tilde{\Phi})$ если для любого $\xi \in \theta$ и любого $w_0^p \in W^p$ существуют такие t^* и l^* , при которых для всех $t \geq t^*$, $l \geq l^*$ выполняются условия цели регулирования. Структурная адаптация выполняется при невозможности учета происходящих изменений средствами параметрической адаптации и заключается в изменении структуры системы управления относительно изменяющихся условий функционирования и классов моделируемых объектов управления. В данном случае могут быть выражены два подхода к решению проблемы структурной адаптации. Первый – выбор модели из фиксированного множества моделей, наиболее адекватно описывающей объект управления, второй – конструирование моделей с требуемыми свойствами, на основании некоторых элементарных составляющих модели. По сравнению с первым подходом второй обеспечивает более гибкую и точную настройку структуры системы управления к изменениям в объекте управления при минимальном участии специалистов в области проектирования систем управления. Количество объектов, к которым может адаптироваться система во втором случае значительно больше, чем в первом.

Структура системы управления будет являться двухконтурной. В первом контуре реализуются функции управления объектом, во втором осуществляется настройка структуры системы управления. Управляющее устройство первого контура включает системы планирования и регулирования, которые будут обладать свойствами параметрической адаптации.

Из вышеизложенного следует, что работа структурного адаптора идентична работе параметрического адаптора и она представима в виде обобщенного соотношения $\text{PR}_{k+1} = WC(\psi_k, y_k, M, M^\Pi, I, \Sigma, D)$, где $WC(\cdot)$ – оператор структурного адаптора системы управления. Математическое обеспечение базируется на основе составляющих PR_{k+1} : ψ – набор признаков; M – модель системы; M^Π – база программных модулей; I – база информационных структур; Σ – граф межмодульных связей; D – база документации. Указанное обеспечение проверяется на адекватность $\mu(\{\tilde{y}_k\}, \{y\}) \leq \bar{\mu}$. В данном случае, если методами структурной адаптации не удастся достичь заданного уровня адекватности, то необходима доработка элементов исходных множеств $\psi, M, M^\Pi, I, \Sigma, D$ неформальными методами.

Сложность процессов производства огнеупорных материалов преодолевается иерархическими подходами к построению энергоэкономической системы управления

технологическими процессами. Одним из таких подходов является реализация высокоэффективной адаптивной системы управления взаимодействующими производственными процессами, основными положениями которой являются следующие.

1. Взаимодействие подсистем в иерархической системе управления базируется на общесистемных принципах управления сложными системами. Сложность объекта приводит к необходимости организации иерархической структуры управления.

2. Временные соотношения в иерархической системе управления производством упорядочены снизу вверх, т. е. управленческие решения вышестоящей подсистемы не могут следовать чаще воздействий подчиненных подсистем.

3. В иерархической системе каждая подсистема наделена функциями стабилизации и координации.

4. Каждая подсистема иерархической системы базируется на алгоритмах параметрической и структурной адаптации.

5. Достижение цели планирования осуществляется процессами координации относительно общей цели системы.

6. Адаптивные системы управления в контуре координации с использованием подстройки по имитационным моделям должны обладать свойством рефлексии по отношению к подсистемам более низкого уровня.

ПЕРЕЧЕНЬ ССЫЛОК

1. Материалы для электротермических установок / Н.В. Большаков, В.И. Бурцев, К.С. Борисанова и др.; Под ред. М.Б. Гутмана. – М.: Энергоатомиздат, 1987. – 296 с.

2. 50 лет деятельности в области информационных технологий и автоматизации: Харьков: ЗАО Тяжпромавтоматика, 2004. – 96 с.
3. Энергобережения в рыночных условиях / Материалы семинара: Київ, 31 березня – 1 квітня 2004р. – Київ: Укр. центр перепідготовки та навчання, 2004. – 223 с.
4. Предко М. Руководство по микроконтроллерам: Т.1. – М.: Постмаркет, 2001. – 416 с.
5. Предко М. Руководство по микроконтроллерам: Т.2. – М.: Постмаркет, 2001. – 488 с.
6. Дорф Р., Бишоп Р. Современные системы управления / Пер. с англ. Б.И. Копылова. – М.: Юнимедиастайл, 2002. – 832с.
7. Моисеев Н.Н. Элементы теории оптимальных систем. – М.: Наука, 1975. – 528 с.
8. Андерсон Т. Статистический анализ временных рядов / Пер. с англ. – М.: Мир, 1976. – 755 с.
9. Боровиков В. СТАТИСТИКА: искусство анализа данных на компьютере. Для профессионалов. – СПб: Питер, 2001. – 656 с.
10. Саридис Дж. Самоорганизующиеся стохастические системы управления / Пер. с англ. – М.: Наука, 1980. – 400 с.
11. Фомин В.Н. Фрадков А.А., Якубович В.А. Адаптивное управление динамическими объектами. – М.: Наука, 1981. – 448 с.
12. Лингер Р., Миляс Х., Уитт Б. Теория и практика структурного программирования / Пер. с англ. – М.: Мир, 1982. – 406 с.
13. Бутенин Н.В., Лунц Я.А., Меркин Д.Р. Курс теоретической механики. – СПб: ЛАНЬ, 1998. – 736 с.

Надійшла 20.04.2004

Після доробки 02.11.2004

В роботі проводиться теоретичне обґрунтування методології побудови взаємозв'язаного дуального закону і моделей функціонування ієрархічних систем адаптивного керування виробництвом вогнетривких матеріалів на основі закону параметричної і структурної адаптації.

In this work the theoretical substantiation of decision methodology of constructing of the interconnected dual law and dynamic functioning models in hierarchical systems of fire-resistant materials production adaptive management is carried out on the basis of the law of parametrical and structural adaptation.

УДК 62-83:621.313.333

Е. М. Потапенко, А. В. Соломаха, Е. Е. Потапенко

КАЛИБРОВКА ДАТЧИКОВ ОДНОФАЗНЫХ ГАРМОНИЧЕСКИХ СИГНАЛОВ С ПОМОЩЬЮ ДИНАМИЧЕСКИХ ФИЛЬТРОВ

Синтезированы два фильтра для оценки неизвестного смещения нуля датчика, измеряющего синусоидальный сигнал. Разработан простой фильтр, позволяющий найти смещения нуля датчика без использования частоты сигнала. Проведено сравнение фильтров по объему вычислений. Работоспособность фильтров подтверждена численным моделированием.

ВВЕДЕНИЕ

Для качественного управления необходимо иметь точные датчики. Одной из самых распространенных ошибок датчиков является смещение нуля. Смещение нуля можно определить последовательной подачей двух

постоянных и равных по модулю сигналов, но с противоположными знаками. Путем алгебраического суммирования установившихся выходных сигналов получается удвоенное смещение нуля. Однако описанный способ калибровки требует введения специального калибровочного режима, и не применим в штатной работе системы, содержащей указанные датчики. В электросетях переменного тока измеряемые сигналы являются синусоидальными. В работе [1] для калибровки датчиков тока и напряжения двухфазных гармонических сигналов в штатном режиме были использованы сами рабочие сигналы и получены хорошие точностные характеристики.

Ниже предлагаются алгоритмы калибровки в штатном режиме работы системы при однофазном синусоидальном полезном сигнале.

ПОСТАНОВКА ЗАДАЧИ

На датчик подается синусоидальный сигнал. На его выходе вследствие наличия смещения нуля будет смещенная синусоида

$$y = x_0 + x_1, \quad (1)$$

где x_0, x_1 – постоянная и гармоническая составляющие выходного сигнала y , т. е. $x_1 = X \cos \omega t$. Здесь ω – частота выходного сигнала. Введем в рассмотрение новую переменную $x_2 = X \sin \omega t$. Тогда правую часть уравнения (1) можно описать следующей системой уравнений:

$$\dot{x}_0 = 0, \quad (2)$$

$$\dot{x}_1 = -\omega x_2, \quad \dot{x}_2 = \omega x_1. \quad (3)$$

Покажем, что система (2), (3) с измерением (1) является полностью наблюдаемой. При $y \equiv 0$ система (1)–(3) принимает вид

$$0 \equiv -\omega x_2, \quad \dot{x}_2 \equiv \omega x_1,$$

откуда следует, что при $|\omega| > 0$ $x_1 \equiv x_2 \equiv 0$, а из (1) тогда следует $x_0 \equiv 0$. Это свидетельствует [2] о полной наблюдаемости системы (1)–(3). Следовательно, для этой системы можно построить наблюдатель для оценки переменных x_0, x_1, x_2 .

СИНТЕЗ НАБЛЮДАТЕЛЯ 1

Наблюдатель для оценки переменных x_0, x_1, x_2 имеет вид

$$\dot{\hat{x}} = A\hat{x} + L(C\hat{x} - y), \quad (4)$$

где

$$\hat{x} = \begin{bmatrix} \hat{x}_0 \\ \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -\omega \\ 0 & \omega & 0 \end{bmatrix}, \quad C = [1 \ 1 \ 0], \quad L = \begin{bmatrix} l_0 \\ l_1 \\ l_2 \end{bmatrix}, \quad (5)$$

l_0, l_1, l_2 – коэффициенты передачи наблюдателя, индексом «^» обозначены оценки вектора $x = [x_0 \ x_1 \ x_2]^T$, который описывается уравнениями

$$\dot{x} = Ax, \quad (6)$$

$$y = Cx. \quad (7)$$

Вычитая из (4) (6) с учетом (7), получим уравнение ошибок наблюдателя $\tilde{x} = \hat{x} - x$

$$\dot{\tilde{x}} = (A + LC)\tilde{x}. \quad (8)$$

Уравнение (8) имеет характеристическое уравнение

$$p^3 - (l_0 + l_1)p^2 + \omega(\omega + l_2)p - \omega^2 l_0 = 0. \quad (9)$$

В качестве стандартного характеристического уравнения примем уравнение

$$p^3 + a_1 \Omega p^2 + a_2 \Omega^2 p + \Omega^3 = 0, \quad (10)$$

где Ω – частота, характеризующая быстродействие наблюдателя.

В случае биномиального распределения корней

$$a_1 = a_2 = 3. \quad (11)$$

В случае распределения корней по Баттерворту

$$a_1 = a_2 = 2. \quad (12)$$

Приравнивая коэффициенты при одинаковых степенях p в уравнениях (9) и (10), найдем

$$l_0 = -\frac{\Omega^3}{\omega^2}, \quad l_1 = \frac{\Omega^3}{\omega^2} - a_1 \Omega, \quad l_2 = \frac{a_2 \Omega^2}{\omega} - \omega. \quad (13)$$

Уравнение наблюдателя (4) в дискретном виде будет выглядеть так:

$$x_0[k] = (x_0[k-1] + l_0 T_0 (x_1[k] - y[k])) / (1 - l_0 T_0),$$

$$x_1[k] = (x_1[k-1] - T_0 \omega [k] x_2[k] + l_1 T_0 (x_0[k] - y[k])) / (1 - l_1 T_0),$$

$$x_2[k] = x_2[k-1] + T_0 \omega [k] x_1[k] + l_2 T_0 (x_0[k] + x_1[k] - y[k]). \quad (14)$$

СИНТЕЗ НАБЛЮДАТЕЛЯ 2

Для уменьшения порядка наблюдателя подставим x_1 из (1) в (3). С учетом (2) получим

$$\dot{y} = -\omega x_2, \quad (15)$$

$$\dot{x}_2 = \omega(y - x_0), \quad (16)$$

$$\dot{x}_0 = 0. \quad (17)$$

Уравнение (15) можно рассматривать как измерение z переменной x_2 для системы (16), (17), т. е.

$$z = -\omega x_2, \quad (18)$$

$$z := \dot{y}. \quad (19)$$

Систему (16)–(18) можно представить в виде

$$\dot{q} = A_q q + B_q y, \quad (20)$$

$$z = C_q q, \quad (21)$$

где

$$q = \begin{bmatrix} x_0 \\ x_2 \end{bmatrix}, \quad A_q = \begin{bmatrix} 0 & 0 \\ -\omega & 0 \end{bmatrix}, \quad B_q = \begin{bmatrix} 0 \\ \omega \end{bmatrix}, \quad C_q = [0 \ -\omega]. \quad (22)$$

Для системы (20), (21) построим наблюдатель в виде

$$\dot{\hat{q}} = A_q \hat{q} + B_q y + L_q (C_q \hat{q} - z), \quad (23)$$

где $L_q = [l_{q1} \ l_{q2}]^T$ – матрица коэффициентов передачи наблюдателя. Подстановка z из (21) в (23) и вычитание из полученного уравнения (20), дают уравнение ошибок на-блюдателя

$$\dot{\tilde{q}} = (A_q + L_q C_q) \tilde{q}. \quad (24)$$

Уравнению (24) соответствует характеристическое уравнение

$$p^2 + \omega l_{q2} p - \omega^2 l_{q1} = 0. \quad (25)$$

В качестве стандартного уравнения примем уравнение

$$p^2 + a_{q1} \Omega_q p + a_{q2} \Omega_q^2 = 0, \quad (26)$$

где Ω_q – частота, характеризующая быстродействие на-блюдателя.

Приравнявая коэффициенты при равных степенях p в уравнениях (25) и (26), найдем

$$l_{q1} = -a_{q2} \frac{\Omega_q^2}{\omega^2}, \quad l_{q2} = a_{q1} \frac{\Omega_q}{\omega}. \quad (27)$$

Для моделирования наблюдателя подставим z из (19) в (23). Тогда

$$\dot{\hat{q}} = A_q \hat{q} + B_q y + L_q (C_q \hat{q} - \dot{y}). \quad (28)$$

С целью устранения необходимости дифференцировать выходной сигнал y , примем обозначение

$$\hat{q} + L_q y = r \Rightarrow \hat{q} = r - L_q y. \quad (29)$$

С учетом обозначения (29) уравнение (28) принимает вид

$$\dot{r} = ((A_q + L_q C_q)r + [B_q - (A_q + L_q C_q)L_q])y. \quad (30)$$

Уравнение (30) в скалярном виде выглядит так:

$$\dot{r}_1 = \omega l_{q1} (-r_2 + l_{q2} y), \quad (31)$$

$$\dot{r}_2 = \omega [-r_1 - l_{q2} r_2 + (1 + l_{q1} + l_{q2}^2) y]. \quad (32)$$

Характеристическое уравнение системы (31), (32) имеет вид (25). При $l_{q2} = 2\text{sign}\omega$, $l_{q1} = -1$ уравнение (25) принимает вид

$$D(p) = (p + |\omega|)^2 = 0. \quad (33)$$

Система (31), (32) в дискретном виде выглядит так:

$$\begin{aligned} r_1[k] &= r_1[k-1] + \omega[k] T_0 l_{q1} (l_{q2} y[k] - r_2[k]), \\ r_2[k] &= \frac{(r_2[k-1] + \omega[k] T_0 ((1 + l_{q1} + l_{q2}^2) y[k] - r_1[k]))}{(1 + \omega[k] T_0 l_{q2})}, \end{aligned} \quad (34)$$

где T_0 – такт счета.

По уравнениям (1), (34) составлена программа моделирования на ПЭВМ. На рис. 1 представлены результаты моделирования с частотой $\omega = 50\text{c}^{-1}$, $X = 1$, $x_0 = 0,5$. В со-

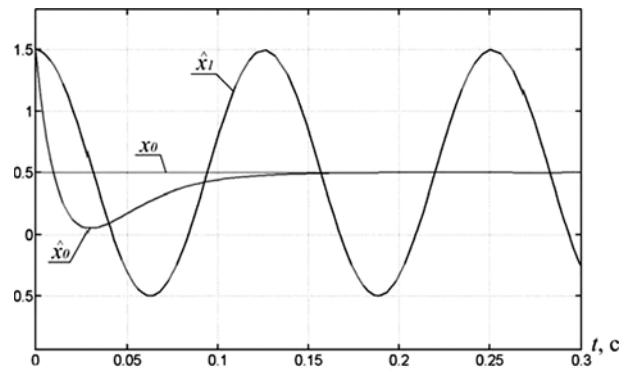


Рисунок 1

ответствии с рис. 1 время калибровки составляет 0,14 с. Как видно из (33), при заданных l_{q1}, l_{q2} время калибровки пропорционально ω .

СИНТЕЗ ФИЛЬТРА

Пропустим измерение (1) через звено с передаточной функцией

$$W_1(p) = \frac{T_1 p}{T_1 p + 1}, \quad (35)$$

в результате чего получим оценку полезного сигнала \hat{x}_1

$$\hat{x}_1 = W_1(p) y. \quad (36)$$

Числитель функции (35) устраняет постоянную составляющую (x_0), а знаменатель позволяет интегрировать сигнал без учета начальных условий. Вычитая (36) из (1) получим оценку смещения нуля датчика

$$\hat{x}_0 = y - \hat{x}_1. \quad (37)$$

Заметим, что чем ниже частота сигнала x_1 , тем больше колебательность оценки (37), поэтому возникает необходимость фильтровать \hat{x}_0 . В результате будем иметь

$$\hat{x}_{0f} = \frac{1}{T_2 p + 1} \hat{x}_0. \quad (38)$$

Для того, чтобы сохранить фазу входного сигнала y неизменной, вычтем (38) из (1), получим оценку полезного сигнала x_1

$$v = y - \hat{x}_{0f}. \quad (39)$$

Алгоритм (35)–(39) в дискретном виде будет выглядеть так:

$$\begin{aligned} \hat{x}_1[k] &= T_1 (\hat{x}_1[k-1] + y[k] - y[k-1]) / (T_1 + T_0), \\ \hat{x}_0[k] &= y[k] - \hat{x}_1[k], \\ \hat{x}_{0f}[k] &= (T_0 \hat{x}_0[k] + T_2 \hat{x}_{0f}[k-1]) / (T_2 + T_0), \\ v[k] &= y[k] - \hat{x}_{0f}[k]. \end{aligned} \quad (40)$$

Было проведено моделирование уравнений (1), (40) с теми же исходными параметрами, что и для наблюдателя 2. Результаты моделирования представлены на рис. 2.

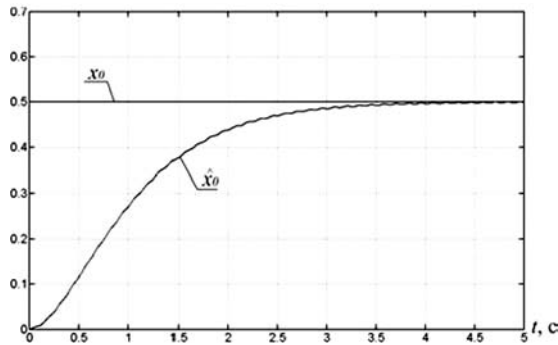


Рисунок 2

Результаты сравнения синтезированных алгоритмов по количеству математических операций сведены в табл. 1.

Таблица 1

Алгоритм	Количество операций сложения (вычитания)	Количество операций умножения (деления)
(14)	9	9
(34)	5	8
(40)	5	6

Как видно из табл. 1, объем вычислений алгоритма наблюдателя (34) по сравнению с алгоритмом (14) меньше примерно в 1,5 раза. Достоинствами разработанного фильтра (40) являются его наименьший объем вычислений и то, что нет необходимости знать частоту входного сигнала, а его недостаток – низкое быстродействие по сравнению с алгоритмами наблюдателей.

ЗАКЛЮЧЕНИЕ

Синтезированы два фильтра (наблюдателя) для оценки неизвестного смещения нуля датчика, измеряющего синусоидальный сигнал. Произведено сравнение наблюдателей по объему вычислений. Объем вычислений во втором наблюдении в 1,5 раза меньше. Разработан простой фильтр, позволяющий найти смещение нуля датчика без использования информации о частоте сигнала. Работоспособность фильтров подтверждена численным моделированием. Сравнение результатов рис. 1 и рис. 2 показывает, что наблюдатель 2 имеет большее быстродействие, чем фильтр. Но поскольку в штатном режиме дрейф смещения нуля имеет большую постоянную времени, то оба результата можно считать удовлетворительными.

ПЕРЕЧЕНЬ ССЫЛОК

1. Потапенко Е.М., Потапенко Е.Е. Оценка двухфазных неопределенных сигналов при наличии погрешностей датчиков // Электротехника (Россия). – 2004. – № 6. – С. 55–57.
2. Квакернаак Х., Сиван Р. Линейные оптимальные системы управления. – М.: Мир, 1977. – 650 с.

Надійшла 26.04.2004

Після доробки 10.11.2004

Синтезовано два фільтри для оцінки невідомого зсуву нуля датчика, що вимірює синусоїдальний сигнал. Розроблено простий фільтр, що дозволяє знайти зсув нуля датчика без використання частоти сигналу. Зроблено порівняння фільтрів за обсягом обчислень. Працездатність фільтрів підтверджена чисельним моделюванням.

Two filters for an estimation of unknown zero creep of the gauge had been synthesized. The simple filter allowing to find a zero creep of the gauge without using the frequency of a signal is developed. The comparison of the observers on volume of calculations is done. Digital modelation approved efficiency of synthesized observers.

АВТОРЫ НОМЕРА

Авраменко Валерий Павлович

доктор технических наук, профессор Харьковского национального университета радиоэлектроники

Алипов Илья Николаевич

кандидат технических наук, доцент Харьковского национального университета радиоэлектроники

Алипов Николай Васильевич

доктор технических наук, профессор Харьковского национального университета радиоэлектроники

Афонин Юрий Сергеевич

студент магистратуры Запорожского национального технического университета

Бодянский Евгений Владимирович

доктор технических наук, профессор Харьковского национального университета радиоэлектроники

Бондарев Виктор Павлович

кандидат физико-математических наук, доцент Запорожского национального технического университета

Вершина Александр Иванович

кандидат технических наук, доцент Запорожского национального технического университета

Воропай Алексей Юрьевич

аспирант Запорожского национального технического университета

Второв Евгений Петрович

кандидат технических наук, доцент Харьковского национального университета радиоэлектроники

Гапоненко Николай Прокопович

кандидат технических наук Запорожского национального технического университета

Гостев Владимир Иванович

доктор технических наук, заведующий кафедрой Государственного университета информационно-коммуникационных технологий, г. Киев

Гребенник Игорь Валериевич

кандидат физико-математических наук, доцент Харьковского национального университета радиоэлектроники

Григорьев Андрей Викторович

аспирант Донецкого национального университета

Долгов Виктор Иванович

доктор технических наук, профессор Харьковского национального университета радиоэлектроники

Дубровин Валерий Иванович

кандидат технических наук, доцент Запорожского национального технического университета

Евсеева Людмила Григорьевна

кандидат физико-математических наук, докторант Института проблем машиностроения им. А.Н. Подгорного НАН Украины

Жуков Леонид Александрович

кандидат технических наук, доцент Сибирского государственного технологического университета

Журавлев Владимир Николаевич,

кандидат технических наук, доцент Запорожского национального технического университета

Завалин Алексей Анатольевич

аспирант Одесского национального политехнического университета

Зенькович Анатолий Александрович

магистр Одесского национального политехнического университета

Зиненко Игорь Иванович

ассистент Запорожского национального технического университета

Кабак Владислав Семенович

кандидат технических наук, декан Запорожского национального технического университета

Калачева Вероника Валериевна

кандидат технических наук, старший научный сотрудник Харьковского университета воздушных сил

Каргин Анатолий Алексеевич

доктор технических наук, заведующий кафедрой Донецкого национального университета

Карпуков Леонид Матвеевич

кандидат технических наук, доцент Запорожского национального технического университета

Касьян Константин Николаевич,

кандидат технических наук, доцент Запорожского национального технического университета

Касьян Николай Николаевич

кандидат технических наук, декан Запорожского национального технического университета

Киричек Галина Григорьевна

ассистент Запорожского национального технического университета

Крищук Владимир Николаевич

кандидат технических наук, заведующий кафедрой Запорожского национального технического университета

Кунгурцев Алексей Борисович

кандидат технических наук, доцент Одесского национального политехнического университета

Лукьянова Валентина Петровна

ведущий инженер Национального Научного Центра "Харьковский Физико-технический институт"

Намлинский Антон Александрович

аспирант Запорожского национального технического университета

Невлюдов Игорь Шакирович

доктор технических наук, заведующий кафедрой Харьковского национального университета радиоэлектроники

Неласая Анна Викторовна

ассистент Запорожского национального технического университета

Новотарский Михаил Анатольевич

кандидат технических наук, старший научный сотрудник Института математики НАН Украины

Онуфриенко Владимир Михайлович

кандидат физико-математических наук, декан Запорожского национального технического университета

Опанасенко Владимир Николаевич

кандидат технических наук, старший научный сотрудник Института кибернетики НАН Украины

Орлов Вадим Владимирович

ведущий инженер Физико-технического института низких температур им. Б.И. Веркина НАН Украины

Орловский Игорь Анатольевич

кандидат технических наук, доцент Запорожского национального технического университета

Осадчий Владимир Владимирович

ассистент Запорожского национального технического университета

Пица Дмитрий Макарович

доктор технических наук, проректор Запорожского национального технического университета

Потапенко Евгений Евгениевич

младший научный сотрудник Запорожского национального технического университета

Потапенко Евгений Михайлович

доктор технических наук, профессор Запорожского национального технического университета

Прохорец Светлана Ивановна

инженер-исследователь Национального Научного Центра "Харьковский Физико-технический институт"

Пулов Роман Дмитриевич

аспирант Запорожского национального технического университета

Пьянков Владимир Павлович

кандидат физико-математических наук, заведующий кафедрой Запорожского национального технического университета

Работнов Александр Евгеньевич

аспирант Севастопольского национального технического университета

Решетникова Наталья Владимировна

аспирант Красноярского государственного технического университета

Романенко Сергей Николаевич

кандидат физико-математических наук, доцент Запорожского национального технического университета

Романова Татьяна Евгеньевна

доктор технических наук, старший научный сотрудник Института проблем машиностроения им. А.Н. Подгорного НАН Украины

Рыбин Вадим Олегович

старший преподаватель Запорожского национального технического университета

Самойлик Сергей Сергеевич

инженер-конструктор ГП ЗМКБ "Прогресс" им. А.Г. Ивченко

Самойлин Евгений Александрович

адъюнкт Ростовского военного института ракетных войск

Самоквит Светлана Алексеевна

аспирант Запорожского национального технического университета

Сенашова Мария Юрьевна

кандидат физико-математических наук, научный сотрудник Института вычислительного моделирования Сибирского отделения Российской академии наук, г. Красноярск

Сидоров Виктор Николаевич

соискатель Северодонецкого технологического института Восточноукраинского национального университета имени В. Даля

Слипченко Александр Николаевич

младший научный сотрудник Харьковского национального университета радиоэлектроники

Соломаха Александр Викторович

аспирант Запорожского национального технического университета

Тимошенко Ирина Григорьевна

аспирант Института кибернетики НАН Украины

Фарафонов Алексей Юрьевич

аспирант Запорожского национального технического университета

Флора Валентин Данилович

кандидат технических наук, доцент Запорожского национального технического университета

Хажмурадов Манан Ахмадович

доктор технических наук, начальник отдела Национального Научного Центра "Харьковский физико-технический институт"

Хиль Михаил Иванович

кандидат технических наук, заведующий кафедрой Северодонецкого технологического института Восточноукраинского национального университета имени В. Даля

Цурихин Алексей Вячеславович

стажер-преподаватель Харьковского государственного экономического университета

Чернобородов Михаил Петрович

аспирант Запорожского национального технического университета

Чернобородова Наталья Петровна

ассистент Запорожского национального технического университета

Чугай Андрей Михайлович

аспирант Института проблем машиностроения НАН Украины им. А.Н. Подгорного

Чумаченко Виталий Павлович

доктор физико-математических наук, профессор Запорожского национального технического университета

Шило Галина Николаевна

кандидат технических наук Запорожского национального технического университета

Шостак Богдан Алексеевич

ведущий инженер ООО "Торговый дом "Агротехника"

Штефан Татьяна Александровна

ассистент Запорожского национального технического университета

Шушляпин Евгений Андреевич

доктор технических наук, профессор Севастопольского национального технического университета

ДЛЯ СВЕДЕНИЯ АВТОРОВ

Журнал "Радіоелектроніка. Інформатика. Управління" предназначен для публикации наиболее значимых научных и практических результатов исследований ученых высших учебных заведений и научных организаций.

Журнал включен в перечень научных изданий Украины, в которых могут публиковаться результаты диссертационных работ на соискание ученых степеней доктора и кандидата технических наук.

Статьи, опубликованные в журнале, реферируются в реферативных журналах и базах данных ВИНТИ (Россия) и "Джерело" (Украина). Рефераты статей размещаются на web-сайте Запорожского национального технического университета (<http://www.zntu.edu.ua/> RIC).

Журнал издается два раза в год и распространяется по подписке в Украине. Подписной индекс 22914.

Последовательность размещения материала статьи: индекс УДК, название статьи, инициалы и фамилия автора(ов), аннотация, текст, перечень ссылок. Рукопись статьи должна быть подписана всеми авторами.

Текст статьи

Принимаются статьи на русском, украинском и английском языках. Размер статьи до 0,5 авторского листа.

Структура статьи

Редакционная коллегия принимает к рассмотрению только научные статьи, которые содержат такие необходимые элементы: постановка проблемы в общем виде и ее связь с важными научными или практическими заданиями; анализ последних исследований и публикаций, в которых начато решение данной проблемы, и на которые опирается автор; выделение нерешенных ранее частей общей проблемы, которым посвящается статья; формулирование целей статьи (постановка задачи); изложение основного материала исследования с полным обоснованием полученных научных результатов, выводы по данному исследованию и перспективы дальнейших исследований в данном направлении.

Требования к оформлению

Статьи принимаются набранные в редакторе Word for Windows (v.6 и выше).

Параметры страницы:

- размер бумаги - А4(210x297);
- ориентация - книжная;
- шрифт - Times New Roman Сур, размер - 12pt;
- междустрочный интервал - полуторный;
- верхнее поле - 20мм, нижнее - 20мм, левое - 25мм, правое - 15мм.

Аннотация

Аннотация является обязательной и подается в начале статьи на русском, украинском (для граждан Украины) и английском языках. Объем аннотации не должен

превышать 40 слов.

Рисунки

Рисунки помещаются в тексте и подаются в отдельных файлах (формат .TIF с разрешением не менее 200 dpi). Рисунки нумеруют и подписывают внизу.

Таблицы

Заголовки таблиц размещаются по центру.

Формулы

Формулы выполняются с помощью встроенного в Word for Windows редактора Microsoft Equation. Формулы нумеруют в круглых скобках справа:

$$Z(\Theta) = 101 \log \left(\frac{y^2}{s} \right) \quad (3)$$

Формулы большого размера записываются в несколько строк.

Перечень ссылок

Перечень ссылок в конце рукописи подается языком оригинала и составляется в порядке упоминания ссылок в тексте и в соответствии со стандартом на библиографическое описание. Ссылки на литературу в тексте обозначаются цифрой в квадратных скобках.

В сведениях об авторах необходимо привести фамилию, имя, отчество(полностью), место работы, должность, ученую степень, рабочий и домашний адрес, телефоны, e-mail.

В редакцию журнала необходимо представить:

- 1) распечатанный текст статьи в 2-х экземплярах;
- 2) экспертное обоснование о возможности опубликования;
- 3) сведения об авторах;
- 4) сопроводительное письмо-ходатайство из организации, где была выполнена работа;
- 5) название статьи, фамилии, имена и отчества авторов на английском языке;
- 6) дискета 3,5" с текстом документа, файлами рисунков, данными о статье на английском языке и сведениями об авторах в формате Word for Windows (v.6 и выше). Файлы можно выслать электронной почтой в виде заархивированных (ZIP-архиватором) файлов.

Статьи, которые не отвечают указанным требованиям, не принимаются к рассмотрению.

Рукописи и дискеты не возвращаются.

Адрес редакции: 69063, г. Запорожье,

ул. Жуковского, 64, ЗНТУ, редакция журнала "PIU"

Тел.: (0612) 63-09-99,

(0612) 69-85-00 - редакционно-издательский отдел

(0612) 69-82-19 - главный редактор

(0612) 63-09-99 - зам. главного редактора

Факс: (0612) 64-22 -74

E-mail: rvv@zntu.edu.ua, TQM@zntu.edu.ua

Запорізький національний технічний університет
Інститут безперервної освіти
Навчальне відділення "Освіта для бізнесу"

Шановні друзі!

**Запрошуємо на програми підготовки та перепідготовки
з нижченаведених напрямків:**

- Система міжнародних розрахунків при виконанні зовнішньо-економічної діяльності.
- Захист інформації в системах телекомунікацій.
- Програмування в середовищі Delphi.
(можлива дистанційна форма навчання - з використанням Internet)
- Системи керування базами даних на прикладі СУБД Informix.
(можлива дистанційна форма навчання - з використанням Internet)
- Фінансовий менеджмент.
- Ціноутворення та аналіз діяльності підприємства.
- Управління персоналом.
- Протокол ділового спілкування.
- Менеджмент організацій.
- Маркетинг: модне слово чи філософія бізнесу?
- Комп'ютерне проектування в системі AutoCAD.
- Бухгалтер, головний бухгалтер фірми.

Тривалість програм – від 1 до 4 міс.

Для навчальних груп підприємств розробляються навчальні програми
з напрямків діяльності кафедр університету.

Адреса: 69063, м. Запоріжжя, вул. Жуковського 46, оф. 616
Тел./факс (0612) 64-29-59, тел. (0612) 69-83-51
e-mail: business-school@zntu.edu.ua
www.zntu.edu.ua

*Що відрізняє сильну фірму від слабкої, так це
насамперед кваліфікація її персоналу, його знання,
мотивація і спрямованість.*

П. Друккер