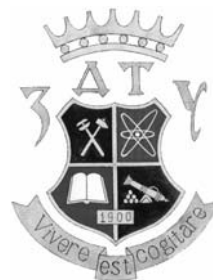


Науковий журнал
Радіоелектроніка
Інформатика
Управління



2001
2(6)

Видається двічі на рік

Зареєстрований **27 травня 1998 року**
Міністерством інформації України,
свідоцтво - серія **КВ № 3264**.

Засновник і видавник -
Запорізький національний технічний університет

Передплатний індекс - 22914

З М І С Т

РАДІОЕЛЕКТРОНІКА	4
АНАЛИТИЧЕСКАЯ ОЦЕНКА И МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ВХОДНОГО ПОТОКА СИГНАЛОВ В АВИАЦИОННЫХ БОРТОВЫХ КОМПЛЕКСАХ ОБОРОНЫ <i>А.М.Бородин, Н.В.Усков</i>	4
АЛГОРИТМ ОПРЕДЕЛЕНИЯ ФУНКЦИЙ ЧУВСТВИТЕЛЬНОСТИ В ДИНАМИКЕ ПО МНОГИМ ВАРЬИРУЕМЫМ ПАРАМЕТРАМ <i>А.И. Вершина, В.С. Кабак, А.Г. Маркин</i>	7
СРЕДНЯЯ ИНТЕНСИВНОСТЬ ПОЛЯ ВБЛИЗИ ФОКУСА ПАРАБОЛИЧЕСКОГО РЕФЛЕКТОРА С КРУПНОМАСШТАБНЫМИ ШЕРОХОВАТОСТЯМИ <i>Р.В.Воробьев</i>	11
АВТОНОМНАЯ СИСТЕМА ФУНКЦИ-УПРАВЛЕНИЯ ТЕПЛОСНАБЖЕНИЕМ <i>В.И.Гостев, В.В.Крайнев, В.В.Гостев</i>	19
МОДЕЛЬ ДЛЯ РАСЧЕТА ТЕНЗОРА ГРИНА МИКРОПОЛОСКОВОЙ СТРУКТУРЫ В ПРОСТРАНСТВЕННОЙ ОБЛАСТИ <i>Л.М.Карпуков</i>	28
ВИЗНАЧЕННЯ НОРМОВАНОГО ПОЛЯ ДОПУСКІВ НА ПАРАМЕТР НАСТРОЙКИ ПРИ АВТОФОКУСУВАННІ ЕЛЕКТРОННОГО МІКРОСКОПА ЗА ЗОБРАЖЕННЯМ <i>А.С.Краснопоясовський, А.М.Скаковська</i>	33
ИНТЕГРОДИФФЕРЕНЦИАЛЬНАЯ МОДЕЛЬ ВЗАИМОДЕЙСТВИЯ МОНОХРОМАТИЧЕСКОЙ ВОЛНЫ С КРУГОВЫМ ЦИЛИНДРОМ (КВАЗИСТАЦИОНАРНЫЙ СЛУЧАЙ) <i>В.М.Онуфриенко, Т.И.Слюсарова</i>	36
ПРОСТРАНСТВЕННО-(ПОЛЯРИЗАЦИОННО-) ВРЕМЕННАЯ ФИЛЬТРАЦИЯ СИГНАЛОВ НА ФОНЕ КОМБИНИРОВАННЫХ ПОМЕХ <i>Д.М.Пица, М.П.Чернобородов, Ю.Л.Мейстер</i>	41

ТРАКТОВЫЙ УСИЛИТЕЛЬ ДЛЯ МАРКЕРНО-СВЯЗНОГО ВОЛНОВОДА В РАДИОВОЛНОВОЙ СИСТЕМЕ УПРАВЛЕНИЯ МАГНИТОЛЕВИТИРУЮЩИМИ ТРАНСПОРТНЫМИ СРЕДСТВАМИ <i>В.Н.Привалов, С.В.Плаксин</i>	ВИЗНАЧЕННЯ ЧАСТОТИ КОМУТАЦІЇ ІМПУЛЬСНОГО ПЕРЕТВОРЮВАЧА <i>В.Д.Флора</i>	44	46
ИНФОРМАТИКА		49	
THE INFLUENCE OF SELF-CONNECTION ON THE PERFORMANCE OF PSEUDO-INVERSE AUTOASSOCIATIVE NETWORKS <i>Dmitry O.Gorodnichy</i>	НЕЧЕТКИЙ ПОДХОД К СИНТЕЗУ КОМПЬЮТЕРНОЙ СЕТИ <i>Л.Г.Комарцова</i>	49	95
PARALLEL IMPLEMENTATION OF NEURAL NETWORK ALGORITHM USING PARALLEL VIRTUAL MACHINE (PVM) SOFTWARE <i>Jamil Ahmad</i>	АНАЛИЗ АЛГОРИТМОВ РЕКОНФИГУРАЦИИ СТРУКТУРЫ ПРОЦЕССОРНОЙ МАТРИЦЫ НА ОСНОВЕ СИГНАЛОВ СОГЛАСИЯ <i>Н.В.Лаходьнова</i>	58	98
REQUIREMENTS PARTICLE NETWORKS: AN APPROACH TO FORMAL SOFTWARE FUNCTIONAL REQUIREMENTS MODELLING <i>Wiwat Vatanaewood, Wanchai Rivepiboon</i>	КОМПЬЮТЕРНО-МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССОВ С ПОСЛЕДЕЙСТВИЕМ В МЕДИЦИНЕ <i>В.П.Марценюк</i>	62	102
АНАЛИЗ ХАРАКТЕРИСТИК МНОГОПОТОКОВЫХ СЕТЕЙ ОБСЛУЖИВАНИЯ <i>А.А.Алиев, Б.Г.Исмаилов</i>	ДИСПЕТЧЕРСКОЕ УПРАВЛЕНИЕ ДВИЖЕНИЕМ ГОРОДСКОГО ПАССАЖИРСКОГО ТРАНСПОРТА С ПОМОЩЬЮ СРЕДСТВ МИКРОПРОЦЕССОРНОЙ ТЕХНИКИ <i>Е.А.Оленев</i>	66	106
МОДЕЛИРОВАНИЕ И ОБРАБОТКА ДАННЫХ СИСТЕМ ТЕСТОВОГО КОНТРОЛЯ ЗАКРЫТОГО ТИПА <i>А.Е.Архипов, С.А.Архипова</i>	ИСПОЛЬЗОВАНИЕ WEB-ТЕХНОЛОГИЙ ДЛЯ АДМИНИСТРИРОВАНИЯ ГЕТЕРОГЕННЫХ ВЫЧИСЛИТЕЛЬНЫХ СЕТЕЙ <i>А.В.Пархоменко, Д.В.Данильченко, А.А.Степаненко</i>	70	110
ОБРАБОТКА СТАТИСТИЧЕСКИХ ДАННЫХ И ОЦЕНКА ЗНАЧИМОСТИ ВЛИЯНИЯ ФАКТОРОВ <i>А.С.Дулесов</i>	ОЦЕНКИ ПОГРЕШНОСТЕЙ ВЫЧИСЛЕНИЯ СЛОЖНОЙ ФУНКЦИИ МНОГИХ ПЕРЕМЕННЫХ <i>М.Ю.Сенашова</i>	76	114
АКТИВНАЯ ЭКСПЕРТНАЯ СИСТЕМА МОНИТОРИНГА И УПРАВЛЕНИЯ РЕМОНТОМ АВИАЦИОННЫХ ГАЗОТУРБИННЫХ ДВИГАТЕЛЕЙ <i>С.В.Жернаков</i>	БИБЛИОТЕКА ФУНКЦИЙ ЧИСЛОВОГО ПОДМНОЖЕСТВА СИСТЕМЫ КОМПЬЮТЕРНОЙ АЛГЕБРЫ АНАЛИТИК-2000 <i>Т.Н.Швалюк</i>	80	117
ВИКОРИСТАННЯ МАТЕМАТИЧНИХ МЕТОДІВ ДЛЯ ОПТИМІЗАЦІЇ ПРОЦЕСІВ ВИРОЩУВАННЯ І ПЕРЕРОБКИ СОЇ <i>А.В.Калініченко</i>	ІНТЕРВАЛИ І ІНТЕРВАЛЬНІ СТРУКТУРИ <i>Г.М.Шило</i>	85	121
ОБ ИНФОРМАЦИОННОЙ ЗНАЧИМОСТИ ЦИФРОВЫХ ИЗОБРАЖЕНИЙ <i>Е.Н.Курсанова, М.Г.Садовский</i>		88	

A FUZZY-TABU SEARCH APPROACH FOR THE PLANT LAYOUT PROBLEM <i>Turkbey Orhan, Alabas Cigdem</i>	126	СИНТЕЗ ОПТИМАЛЬНОГО НЕЧЕТКОГО РЕГУЛЯТОРА ДЛЯ НЕЛИНЕЙНОЙ ДИНАМИЧЕСКОЙ СИСТЕМЫ <i>М.Б.Коломейцева, Д.Л.Хо</i>	153
КОДИРОВАНИЕ СИГНАЛОВ В ЗАДАЧЕ НЕЙРОЭМУЛЯЦИИ <i>Е.В.Бодянский, Н.Е.Кулишова</i>	136	ОБЗОР СОВРЕМЕННЫХ МЕТОДОВ УПРАВЛЕНИЯ СИНХРОННЫМИ ДВИГАТЕЛЯМИ С ПОСТОЯННЫМИ МАГНИТАМИ <i>Д.В.Корельский, Е.М.Потапенко, Е.В.Васильева</i>	155
МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ТЕХНОЛОГИЧЕСКОЙ СХЕМЫ СБОРКИ И ПРОЦЕДУРА ЕЕ ПОСТРОЕНИЯ <i>И.П.Гамаюн</i>	140	КРИТЕРИИ СИНТЕЗА НОМИНАЛЬНЫХ МОДЕЛЕЙ В ЗАДАЧАХ НАБЛЮДЕНИЯ И УПРАВЛЕНИЯ <i>В.И.Кортунов</i>	159
АЛГОРИТМ КЛАССИФИКАЦИИ С ОЦЕНКОЙ ЗНАЧИМОСТИ ПРИЗНАКОВ <i>В.И.Дубровин, С.А.Субботин</i>	145	ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНАЯ СИСТЕМА КОНТРОЛЯ КАЧЕСТВА ФИЛЬТРОВ ГАЗОДИНАМИЧЕСКИМ МЕТОДОМ <i>А.В.Печенов, А.Л.Симаков, Ю.Н.Матросова</i>	163
АВТОМАТИЗАЦИЯ ПРОЦЕССА ИДЕНТИФИКАЦИИ СТАНДАРТНЫХ ДЕТАЛЕЙ, ПОКУПНЫХ КОМПЛЕКТУЮЩИХ ИЗДЕЛИЙ И МАТЕРИАЛОВ <i>Ю.З.Житников, А.А.Фаткин, Т.В.Рулева</i>	150	АВТОРЫ НОМЕРА	168



УКАЗ

ПРЕЗИДЕНТА УКРАЇНИ

Питання надання вищим навчальним закладам статусу національного

Враховуючи загальнодержавне і міжнародне визнання результатів діяльності та вагомий внесок у розвиток національної освіти та науки постановляю:

1. Надати статус національного таким вищим навчальним закладам:

...
Запорізькому державному технічному університету

...
і надалі відповідно іменувати їх

...
Запорізький національний технічний університет

...



Президент України

Л.КУЧМА

м. Київ
7 серпня 2001 року
№ 591/2001

I. РАДИОЭЛЕКТРОНИКА

УДК 621.396.72

АНАЛИТИЧЕСКАЯ ОЦЕНКА И МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ВХОДНОГО ПОТОКА СИГНАЛОВ В АВИАЦИОННЫХ БОРТОВЫХ КОМПЛЕКСАХ ОБОРОНЫ

А.М.Бородин, Н.В.Усков

Об аналитической оценке плотности потока сигналов радиолокационного поля на входе станции РТР при различных граничных условиях. Результаты математического моделирования.

There is analytical estimation of radar signal flow density at RTR station input under different boundary conditions. The results of mathematical modelling are represented.

ВВЕДЕНИЕ

Задача оценки плотности входного потока сигналов является актуальной при проектировании авиационных бортовых комплексов обороны (БКО), устанавливаемых на самолетах, преодолевающих систему противовоздушной обороны (ПВО), с целью увеличения вероятности их сохранения [1].

АНАЛИТИЧЕСКАЯ ОЦЕНКА ВХОДНОГО ПОТОКА СИГНАЛОВ

Одной из важных статистических характеристик радиолокационного поля (РЛП), образуемого РЛС системы ПВО, определяющих входной поток сигналов, является распределение РЛС поля по "энергетическому" параметру Ψ , пропорциональному длине волны (λ) и корню квадратному из потенциала РЛС (РГ). Распределение энергетического параметра в диапазоне средств сухопутного радиолокационного поля определяется выражением [2, стр. 23]:

$$W(\Psi) = \frac{\Psi}{\sigma_{\Psi}^2} e^{-\Psi^2/2\sigma_{\Psi}^2}, \quad (1)$$

где $\Psi = \frac{\lambda}{4\pi} \sqrt{PG}$, σ_{Ψ} - параметр релейского распределения, заключенный в пределах $0, 1 < \sigma_{\Psi} < 0, 5 \text{ км} \cdot \text{Вт}^{1/2}$.

Выбор предела σ_{Ψ} определяется диапазоном работы аппаратуры (типами и количеством РЛС, участвующих в создании радиолокационного поля) и высотой (h) полета самолета (при малых h требуется учитывать маломощные

РЛС, обслуживающие маловысотные ЗРК, что смещает σ_{Ψ} к левому пределу; при этом плотность РЛС на местности увеличивается).

Распределение (1) не учитывает РЛС с квазинепрерывным излучением (КНИ) и позволяет получить среднее число обслуживающих РЛС, а также поток только импульсных сигналов. Действительно, средняя частота (плотность потока) импульсных сигналов ($\bar{\zeta}$) на входе станции радиотехнической разведки (РТР) при круговом приеме, средней плотности работающих на местности РЛС ρ и средней частоте следования зондирующих импульсов \bar{F}_n (без учета квазинепрерывного излучения) равна [2, стр.44,45]:

$$\bar{\zeta} = P(P_{\text{вх}} \geq P_{\text{пор}}) \bar{F}_n(\Psi) \rho (\pi 115^2 h), \quad (2)$$

где $P(P_{\text{вх}} \geq P_{\text{пор}})$ - вероятность превышения входным сигналом порогового уровня $P_{\text{пор}}$ приемника РТР, зависящая от Ψ и h ; $\bar{F}_n(\Psi) = 2250 \cdot \exp(-6\Psi) + 250$ - средняя частота повторения РЛС РЛП (зависимость от Ψ тривиальна: большим Ψ соответствуют меньшие \bar{F}_n).

Выражение $P(P_{\text{вх}} \geq P_{\text{пор}})$ для различных Ψ и h имеет следующий вид [2, стр. 46]:

$$P(P_{\text{вх}} \geq P_{\text{пор}}) \cong \frac{1}{2} \left\{ 1 - \Phi \left[\frac{\ln(\lg P_{\text{пор}} - \lg \frac{\Psi^2}{h} + 8,4) - 0,34}{0,6} \right] \right\}, \quad (3)$$

где $\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ интеграл вероятности при

$P_{\text{пор}} > \frac{\Psi^2}{h} \cdot 10^{-8,4} (\Psi < 10^{4,2} \cdot \sqrt{h \cdot P_{\text{пор}}})$, $P(P_{\text{вх}} \geq P_{\text{пор}}) \cong 1$ в формуле (3).

Для получения зависимости плотности потока импульс-

ных сигналов от всех РЛС с различными частотами повторения и ψ в функции от $P_{\text{пор}}$ необходимо проинтегрировать выражение (2) с учетом распределения $P(P_{\text{вх}} \geq P_{\text{пор}})$ вида (3), т.е. найти среднее значение плотности потока входных сигналов от всего РЛП ($\bar{\eta}$):

$$\bar{\eta} = \int_0^{\infty} P(P_{\text{вх}} \geq P_{\text{пор}}) \bar{F}_n(\psi) W(\psi) \rho \cdot \pi \cdot 115^2 h \cdot d\psi \quad (4)$$

Выражение (4) справедливо при массовом применении авиации, когда в образовании входного потока участвуют практически все РЛС радиолокационного поля.

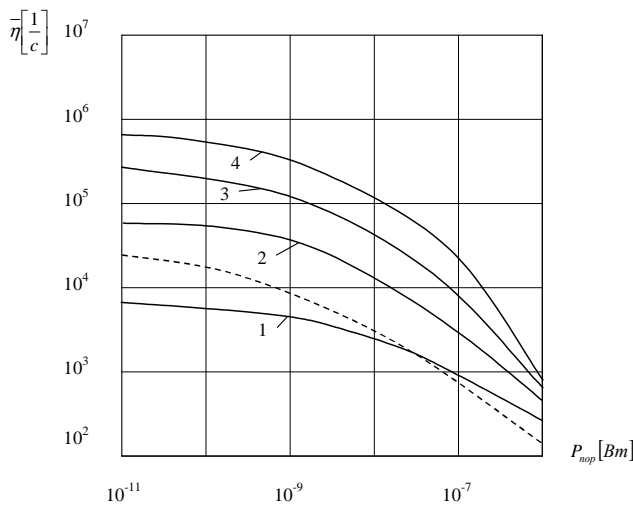


Рисунок 1 - Плотность потока входных сигналов в зависимости от пороговой чувствительности приемника при групповых налетах авиации

На рисунке 1 сплошными линиями проведены результаты интегрирования по формуле (4) в зависимости от пороговой чувствительности приемника $P_{\text{пор}}$ при плотности работающих на местности РЛС $\rho = 10^{-3} 1/\text{км}^2$, $\sigma_{\psi} = 0,3 \text{ км} \cdot \text{Вт}^{1/2}$ для различных высот работы объектов (1 - $h = 0,2 \text{ км}$; 2 - $h = 1 \text{ км}$; 3 - $h = 10 \text{ км}$; 4 - $h = 20 \text{ км}$), и пунктирной линией - зависимость, полученная в тех же условиях для преодоления маловысотного РЛП. ($\sigma_{\psi} = 0,1 \text{ км} \cdot \text{Вт}^{1/2}$, $h = 0,2 \text{ км}$) [1,3].

При локальных действиях авиации необходимо учесть зависимость относительного числа (K_p) работающих РЛС из общего числа РЛС РЛП, от энергетического параметра и высоты полета [2. стр. 48]:

$$K_p = 1 - \exp\left(-17 \frac{\Psi}{h}\right). \quad (5)$$

Последнее выражение позволяет получить зависимость плотности потока импульсных сигналов на входе станции РТР от $P_{\text{пор}}$, когда станция в составе БКО размещается на самолете, действующем одиночно или в составе компактной группы над типовым РЛП на высоте h :

$$\bar{\eta}_{\text{лок}} = \rho \cdot \pi \cdot 115^2 h \times \int_0^{\infty} P(P_{\text{вх}} \geq P_{\text{пор}}) \bar{F}_n(\psi) K_p\left(\frac{\Psi}{h}\right) W(\psi) d\psi, \quad (6)$$

где все функции, входящие в подынтегральное выражение, определяются по приведенным выше соотношениям.

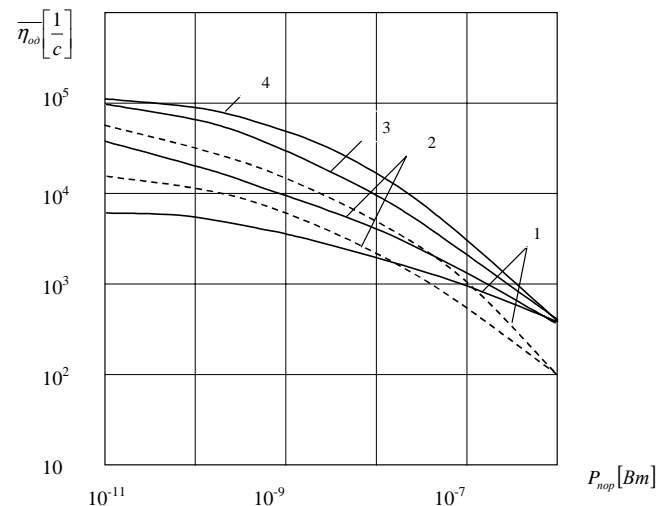


Рисунок 2 - Плотность потока входных сигналов в зависимости от пороговой чувствительности приемника на объекте, действующем одиночно

На рисунке 2 даются результаты интегрирования по формуле (6) для $\rho = 10^{-3} 1/\text{км}^2$, $\sigma_{\psi} = 0,3 \text{ км} \cdot \text{Вт}^{1/2}$ (сплошные линии) и $\sigma_{\psi} = 0,1 \text{ км} \cdot \text{Вт}^{1/2}$ (пунктирные линии) для различных высот работы объектов (1 - $h = 0,2 \text{ км}$; 2 - $h = 1 \text{ км}$; 3 - $h = 10 \text{ км}$; 4 - $h = 20 \text{ км}$).

Из графиков рисунков 1,2 следует, в частности, что плотности потока импульсных сигналов на входе БКО при малых чувствительностях приемников станции РТР порядка $10^{-6} \div 10^{-7} \text{ Вт}$ (приемники прямого усиления) не превышает 1кГц в наиболее насыщенном РЛС сухопутном РЛП. При этом зависимость плотности потока от h и σ_{ψ} сравнительно слабая.

Плотность потока существенно возрастает при больших чувствительностях приемников. Так, при $P_{\text{пор}} = 10^{-10} \text{ Вт}$,

$h = 10 \text{ км}$, $\sigma_{\psi} = 0,3 \text{ км} \cdot \text{Вт}^{1/2}$ плотность потока импульсных сигналов на входе БКО составляет около $8 \cdot 10^4 \text{ 1/с}$ при одиночных или в составе компактной группы полетах и около $2 \cdot 10^5 \text{ 1/с}$ при массовых действиях авиации. Уменьшение рабочего диапазона частот БКО сокращает плотность потока импульсных сигналов минимум в два раза.

РЕЗУЛЬТАТЫ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Необходимо отметить, что выражение (3) для $P(P_{\text{вх}} \geq P_{\text{пор}})$ было найдено из гистограммы $W(\lg P_{\text{вх}})$, которая, в свою очередь, была получена методом статистических испытаний (моделирования) на ЭВМ при числе реализаций $N=10^4$ аппроксимацией нормированного коэффициента усиления антенны РЛС $[G(\theta)/G_0]$ вида:

$$\frac{G(\theta)}{G_0} = e^{-2,78(\theta/\theta_0)^2} + r \frac{0,625}{(\theta/\theta_0)} \cdot [1 - e^{-(\theta/\theta_0)^4}]$$

при $0 \leq \theta \leq \pi$, (7)

где θ_0 - ширина диаграммы направленности РЛС на уровне половинной мощности; r - коэффициент качества позиции, который для "плохой", изобилующей местными предметами позиции, равен 1, а для "хорошей" ("чистой") позиции равен 2.

Гистограмма аппроксимировалась логарифмически - нормальным распределением вида:

$$W(P_{\text{вх}}) = \frac{1}{\sigma(\lg P_{\text{вх}} - \epsilon)\sqrt{2\pi}} \cdot \exp\left\{-\frac{1}{2\sigma^2}[\ln(\lg P_{\text{вх}} - \epsilon) - \delta]^2\right\}, (8)$$

где параметры распределения δ , σ , ϵ находились из гистограмм с использованием методики работы [4].

При моделировании на среднее значение, определяемое формулой (7), накладывалась случайная (фоновая) составляющая, распределенная по нормальному закону с дисперсией $\sigma_{\psi} = 3 - 5 \text{ дБ}$ (использовались $r=1,25$ и $\sigma_{\psi} = 3 \text{ дБ}$) для учета "бокового" излучения РЛС.

Учет РЛС с КНИ может быть произведен либо методом математического моделирования, либо приближенно ана-

литически введением в формулы (4) и (6) дополнительного слагаемого в виде суммы частот повторения сигналов средств перехвата с КНИ при условии, что уровень входных КНИ-сигналов превышает $P_{\text{пор}}$. Сравнение данных математического моделирования, полученных с учетом КНИ (средняя плотность потока порядка $5 \cdot 10^5 \text{ 1/с}$ при $h=5 \text{ км}$ и $P_{\text{пор}} = 10^{-8} \text{ Вт}$), с расчетными данными по формулам (4) и (6) показывает, что они отличаются примерно на (200 - 300) кГц, т.е. на частоту повторения (1-2) РЛС с КНИ-сигналами. Это подтверждает введение дополнительного слагаемого в формулы (4) и (6) для учета КНИ [3].

ЗАКЛЮЧЕНИЕ

Таким образом, при построении БКО для работы на сухопутном РЛП необходимо учитывать фоновый входной поток, содержащий полезные сигналы наземных РЛС, на который накладываются сигналы от 1-2 РЛС КНИ. На морском РЛП, где действуют авиационные ударные соединения и группы, охраняемые и управляемые самолетами ДРЛО "Хокай" и "Авакс", распределение РЛС по площади существенно неравномерное, количество РЛС ограничено и они распределяются в широком диапазоне длин волн. Поэтому оценка входного потока сигналов может быть выполнена обычными расчетами числа облучающих РЛС, вероятностей облучения сигналов с учетом зон обнаружения и вероятностей совпадения импульсных потоков. Расчеты показывают, что "фоновый" поток при этом от корабельных РЛС для аналогичных высот и чувствительностей станции РТР существенно (на один - два порядка) ниже, чем фоновый поток на сухопутном РЛП, а основными атакующими средствами, определяющими входной поток сигналов, являются самолеты с РЛС КНИ.

ПЕРЕЧЕНЬ ССЫЛОК

1. Бородин А.М. Автоматизация выбора оптимальных параметров авиационных бортовых комплексов обороны. Диссертация на соискание ученой степени кандидата технических наук. Томск, 1997 г.
2. Мельников Ю.П. Методы оценки эффективности воздушной радиотехнической разведки. Санкт-Петербург, 1998 г.
3. Усков Н.В. Определение местоположения неразрешаемых по параметрам радиоэлектронных средств. Диссертация на соискание ученой степени кандидата технических наук. Томск, 1975 г.
4. Хан Г., Шапиро С. Статистические модели в инженерных задачах. Москва, "Мир" 1969 г.

АЛГОРИТМ ОПРЕДЕЛЕНИЯ ФУНКЦИЙ ЧУВСТВИТЕЛЬНОСТИ В ДИНАМИКЕ ПО МНОГИМ ВАРЬИРУЕМЫМ ПАРАМЕТРАМ

А.И. Вершина, В.С. Кабак, А.Г. Маркин

Рассматривается построение алгоритма оценки функций чувствительности в динамике, в основе которого лежит предельный случай уменьшения шага.

Розглядається побудова алгоритму визначення функції чутливості у динаміці, в основі якого лежить граничний випадок зменшення кроку.

The construction of algorithm an estimation of functions of sensitivity in dynamics, in which basis on the limiting case of reduction of a step lays.

$$\frac{\partial \dot{x}}{\partial p} = \frac{\partial \dot{x}}{\partial p \partial t} = \frac{\partial \dot{x}}{\partial t \partial p}, \quad (3)$$

отсюда, переходя к конечным приращениям, можно записать

$$\frac{\partial \dot{x}}{\partial p} \approx \frac{\frac{\partial x_{i+1}}{\partial p} - \frac{\partial x_i}{\partial p}}{\Delta t} = \frac{1}{h} \left(\frac{\partial x_{i+1}}{\partial p} - \frac{\partial x_i}{\partial p} \right), \quad (4)$$

где $h = \Delta t$.

После подстановки в выражение (2) и проведения соответствующих преобразований получим

$$\frac{\partial x_{i+1}}{\partial p} = \left(\frac{1}{h_i} \frac{\partial F_i}{\partial \dot{x}} + \frac{\partial F_i}{\partial x} \right)^{-1} \left(\frac{1}{h_i} \frac{\partial F_i}{\partial \dot{x}} \frac{\partial x_i}{\partial p} - \frac{\partial F_i}{\partial p} \right). \quad (5)$$

Данное выражение представляет собой неявный одношаговый метод интегрирования для определения функций чувствительности.

Раскроем скобки в выражении (5):

$$\frac{\partial x_{i+1}}{\partial p} = \left(\frac{1}{h_i} \frac{\partial F_i}{\partial \dot{x}} + \frac{\partial F_i}{\partial x} \right)^{-1} \frac{1}{h_i} \frac{\partial F_i}{\partial \dot{x}} \frac{\partial x_i}{\partial p} - \left(\frac{1}{h_i} \frac{\partial F_i}{\partial \dot{x}} + \frac{\partial F_i}{\partial x} \right)^{-1} \frac{\partial F_i}{\partial p}. \quad (6)$$

Рассмотрим построение алгоритма оценки функций чувствительности по многим варьируемым параметрам, в котором не будет использоваться интегрирование по времени в обратном порядке, как это делается в методе присоединенных схем.

МЕТОДЫ И АЛГОРИТМЫ РЕШЕНИЯ

Представим первую составляющую правой части выражения (6) клеточными матрицами, элементы которых будут соответствовать реактивным и нереактивным составляющим. При этом будем считать, что путем перестановки строк и столбцов реактивные составляющие можно скомпоновать в левом верхнем углу матриц. При этом будет справедливым следующее равенство:

$$\begin{bmatrix} \frac{1}{h} X + G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{h} X & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} X + hG_{11} & hG_{12} \\ hG_{21} & hG_{22} \end{bmatrix} \cdot \begin{bmatrix} X & 0 \\ 0 & 0 \end{bmatrix}, \quad (7)$$

ВВЕДЕНИЕ

Определение функций чувствительности в динамике связано с многократным интегрированием систем нелинейных дифференциальных уравнений высокого порядка. При значительном количестве варьируемых параметров целесообразнее использовать метод присоединенных схем. Однако в динамике такой подход требует проведения интегрирования в обратном порядке времени в каждой временной точке, что фактически делает этот метод неприемлемым для моделирования сложных электронных схем [1-3]. Поиск алгоритмов, позволяющих за один дополнительный просчет определять производные по всем варьируемым параметрам без использования интегрирования в обратном порядке времени [4-5], является актуальной задачей.

ПОСТАНОВКА ЗАДАЧИ

Радиоэлектронная схема описывается системой нелинейных алгебро-дифференциальных уравнений

$$F(\dot{x}, x, p) = 0, \quad (1)$$

где x и \dot{x} - вектор-столбец неизвестных переменных и их производных по времени; p - параметры элементов радиоэлектронной схемы.

Производная по параметрам запишется следующим образом:

$$\frac{\partial F}{\partial \dot{x}} \frac{\partial \dot{x}}{\partial p} + \frac{\partial F}{\partial x} \frac{\partial x}{\partial p} + \frac{\partial F}{\partial p} = 0. \quad (2)$$

Представим производную $\frac{\partial \dot{x}}{\partial p}$ в виде

где X соответствует подматрице $\frac{\partial F}{\partial \dot{x}}$, а $\begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}$ соответствует $\frac{\partial F}{\partial x}$.

Воспользуемся формулой Фробениуса [6], дающей возможность связать блоки прямой и обратной матриц

$$S = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \text{ и } S^{-1} = \begin{bmatrix} K & L \\ M & N \end{bmatrix} \text{ соответственно. Формула}$$

Фробениуса имеет вид:

$$S^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BNCA^{-1} & -A^{-1}BN \\ -NCA^{-1} & N \end{bmatrix}, \quad (8)$$

где $N = (D - CA^{-1}B)^{-1}$.

Применительно к выражению (7) имеем

$$A = X + hG_{11}; \quad B = hG_{12}; \quad C = hG_{21}; \quad D = hG_{22};$$

$$N = [hG_{22} - hG_{21}(R + hG_{11})^{-1}hG_{12}]^{-1}.$$

Отсюда

$$\begin{bmatrix} X + hG_{11} & hG_{12} \\ hG_{21} & hG_{22} \end{bmatrix}^{-1} \cdot \begin{bmatrix} X & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} W_{11} & 0 \\ W_{21} & 0 \end{bmatrix}, \quad (9)$$

где $W_{11} = \{ (X + hG_{11})^{-1} + (X + hG_{11})^{-1}hG_{12} \times [hG_{22} - hG_{21}(X + hG_{11})^{-1}hG_{12}]^{-1} \times hG_{21}(X + hG_{11})^{-1} \} X$,

$$W_{21} = -[hG_{22} - hG_{21}(X + hG_{11})^{-1}hG_{12}]^{-1} \times hG_{21}(X + hG_{11})^{-1}X.$$

Преобразуем выражения для W_{11} и W_{21} вынесением общего множителя h и раскрытием внешних скобок:

$$W_{11} = (X + hG_{11})^{-1}X + h(X + hG_{11})^{-1}G_{12} \times [G_{22} - hG_{21}(X + hG_{11})^{-1}G_{12}]^{-1} \times G_{21}(X + hG_{11})^{-1}X,$$

$$W_{21} = -[G_{22} - hG_{21}(X + hG_{11})^{-1}G_{12}]^{-1}G_{21}(X + hG_{11})^{-1}X.$$

Переходя к пределу при $h \rightarrow 0$, получим

$$\lim_{h \rightarrow 0} W_{11} = \lim_{h \rightarrow 0} ((X + hG_{11})^{-1}X), \quad (10)$$

$$\lim_{h \rightarrow 0} W_{21} = -G_{22}^{-1}G_{21} \lim_{h \rightarrow 0} ((X + hG_{11})^{-1}X). \quad (11)$$

Естественно, все вышесказанное будет справедливым, если существуют соответствующие обратные матрицы.

Рассмотрим случай, когда подматрица X исследуемой схемы имеет обратную матрицу. Тогда выражения (10) и (11) соответственно приобретают вид

$$\lim_{h \rightarrow 0} W_{11} = E, \quad (12)$$

$$\lim_{h \rightarrow 0} W_{21} = -G_{22}^{-1}G_{21}. \quad (13)$$

Отсюда для выражения (9) будем иметь

$$\begin{bmatrix} X + hG_{11} & hG_{12} \\ hG_{21} & hG_{22} \end{bmatrix}^{-1} \cdot \begin{bmatrix} X & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} E & 0 \\ -G_{22}^{-1}G_{21} & 0 \end{bmatrix}. \quad (14)$$

Выражение (12) показывает, что для "реактивной" части схемы при уменьшении шага соответствующие выходные переменные зависят от своих значений на предыдущем шаге и их функции чувствительности могут эффективно вычисляться по всем варьируемым параметрам в процессе интегрирования в прямом порядке времени в процессе получения значений выходных переменных в очередной временной точке. При этом, в соответствии с интересующими нас выходными переменными, определяются нужные строки обратной матрицы $\left(\frac{1}{h_i} \frac{\partial F_i}{\partial \dot{x}} + \frac{\partial F_i}{\partial x} \right)^{-1}$.

Остальные переменные, которым соответствует выражение (13), зависят от переменных, определяемых с помощью выражения (12). Для этих переменных необходимо знать соответствующие коэффициенты вхождения переменных определяемых через выражение (12).

Рассмотрим структуру подматрицы G_{21} , которая отражает связь части схемы без реактивных элементов с реактивной частью. Элементы этой матрицы могут содержать значительное число нулевых столбцов. В результате, при определении функций чувствительности для переменных, соответствующих строкам схемной матрицы без реактивных элементов, необходимо определить только часть коэффициентов, отражающих эту связь.

Умножая правую и левую части выражения (14) на столбец, содержащий единицу в строке i , получим столбец коэффициентов, определяющих влияние производной $\frac{\partial x_i}{\partial p}$:

$$\begin{bmatrix} X+hG_{11} & hG_{12} \\ hG_{21} & hG_{22} \end{bmatrix}^{-1} \cdot \begin{bmatrix} X & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \vdots \\ 1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ 1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \vdots \\ k_j^i \\ \vdots \\ \vdots \\ k_n^i \end{bmatrix}. \quad (15)$$

Для нахождения столбца коэффициентов желательно использовать LU -разложение, полученное при нахождении переменных. С учетом (7), это можно сделать решая следующую систему уравнений, правая часть которой

представляет столбец i матрицы $\begin{bmatrix} \frac{1}{h}X & 0 \\ 0 & 0 \end{bmatrix}$:

$$\begin{bmatrix} hX+G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \vdots \\ \vdots \\ k_j^i \\ \vdots \\ \vdots \\ k_n^i \end{bmatrix} = \begin{bmatrix} hx_{1i} \\ \vdots \\ hx_{mi} \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}. \quad (16)$$

Аналогично определяем коэффициенты для остальных производных, которые необходимы для нужной выходной переменной.

Выражение (6) для определения функций чувствительности приобретает вид:

$$\frac{\partial x_{i+1}}{\partial p} = \begin{bmatrix} \frac{\partial x_i}{\partial p} \\ \sum_{\gamma} \left(k \cdot \frac{\partial x_i}{\partial p} \right)_{\gamma} \end{bmatrix} - \left(\frac{1}{h_i} \frac{\partial F_i}{\partial \dot{x}} + \frac{\partial F_i}{\partial x} \right)^{-1} \frac{\partial F_i}{\partial p}. \quad (17)$$

Для удобства индексы номеров строк (переменных) опущены. Выражение в квадратных скобках представляет матрицу-столбец. Верхняя часть столбца - это подматрица значений производных на предыдущем шаге. Нижняя часть столбца - подматрица вычисляемых значений соответствующих производных предыдущего шага через элементы верхней части столбца.

Таким образом, алгоритм определения функций чувствительности включает в себя следующие шаги.

Шаг 1. Определить переменные, непосредственно связанные с реактивными элементами.

Шаг 2. Выделить переменные, которые соединены с не реактивной частью схемы через не реактивный двухполюсник.

Шаг 3. Выделить переменные, для которых будут определяться функции чувствительности.

Шаг 4. Оценить коэффициенты, которые будут определять значения производных на предыдущем временном шаге, не связанных с реактивной частью.

Шаг 5. Определить строки обратной матрицы, соответствующие выходным переменным, путем решения транспонированной системы уравнений.

Шаг 6. Оценить составляющую функции чувствительности по всем варьируемым параметрам, не связанную с предыдущим временным шагом.

Шаг 7. Определить функцию чувствительности текущего временного шага, используя функции чувствительности предыдущего шага и соответствующие коэффициенты.

Шаг 8. Запомнить значения функций чувствительности текущего шага и перейти к шагу 4 до тех пор, пока не будет просмотрен весь временной интервал.

Предложенный алгоритм излагает основные принципы оценки функций чувствительности в динамике по всем варьируемым параметрам. При этом не используется интегрирование в обратном порядке, как это делается в методе присоединенных схем. Вопрос выбора шага и оценки погрешности в данной работе не рассматривается и, кроме того, необходимо учесть случаи, когда приведенные в выражениях матрицы особенные. В тоже время, существует класс схем, для которых данный алгоритм можно эффективно использовать. Примером может служить оценка влияния конструктивных параметров печатных плат на работу цифровых и аналоговых схем.

РЕЗУЛЬТАТЫ

Показано построение алгоритма определения функций чувствительности по многим варьируемым параметрам в динамике, который позволяет избежать интегрирования в обратном масштабе времени, которое присуще методу присоединенных схем, и в то же время определять за один дополнительный просчет функцию чувствительности одной выходной переменной по всем варьируемым параметрам. Продемонстрируем на простейшем примере полученные результаты. Рассмотрим переходной процесс в RC -цепочке, включенной на источник напряжения E . При составлении уравнений будем считать, что отрицательный потенциал источника напряжения и один из концов емкости заземлены. Исследуем функцию чувствительности напряжения на емкости, которой будем считать выходной переменной.

Напряжение на емкости определяется по формуле

$$U_1(t) = E \left(1 - e^{-\frac{t}{RC}} \right) = E \left(1 - e^{-\frac{tg}{C}} \right), \quad (18)$$

где $g = 1/R$.

Выражения для частных производных по параметрам g и C соответственно равны

$$\frac{\partial U_1}{\partial C} = -\frac{Etg}{C^2} e^{-\frac{tg}{C}}, \quad (19)$$

$$\frac{\partial U_1}{\partial g} = -\frac{Et}{C} e^{-\frac{tg}{C}}. \quad (20)$$

Кроме этого, нам понадобится производная по времени

$$\frac{\partial U_1}{\partial t} = -\frac{Eg}{C} e^{-\frac{tg}{C}}. \quad (21)$$

Система уравнений в расширенном координатном базисе, описывающая переходной процесс, имеет вид:

$$\begin{aligned} F_1(U_1, U_2, i_E) &= C \frac{\partial U_1}{\partial t} - (U_2 - U_1)g = 0, \\ F_2(U_1, U_2, i_E) &= (U_2 - U_1)g + i_E = 0, \\ F_3(U_1, U_2, i_E) &= U_2 - E = 0. \end{aligned} \quad (22)$$

Отсюда можно получить матрицу Якоби:

$$J = \begin{bmatrix} \frac{1}{h}C + g & -g & 0 \\ g & g & 1 \\ 0 & 1 & 0 \end{bmatrix}. \quad (23)$$

Выражение для определения функций чувствительности в соответствии с формулой (5) имеет вид

$$\begin{bmatrix} \frac{\partial}{\partial p} U_1^{n+1} \\ \frac{\partial}{\partial p} U_2^{n+1} \\ \frac{\partial}{\partial p} i_E^{n+1} \end{bmatrix} = J^{-1} \begin{bmatrix} \frac{1}{h}C & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial}{\partial p} U_1^n \\ \frac{\partial}{\partial p} U_2^n \\ \frac{\partial}{\partial p} i_E^n \end{bmatrix} - \begin{bmatrix} \frac{\partial F_1}{\partial p} \\ \frac{\partial F_2}{\partial p} \\ \frac{\partial F_3}{\partial p} \end{bmatrix}. \quad (24)$$

Переходя к предельному случаю с учетом выражения (17), получим

$$\begin{bmatrix} \frac{\partial}{\partial p} U_1^{n+1} \\ \frac{\partial}{\partial p} U_2^{n+1} \\ \frac{\partial}{\partial p} i_E^{n+1} \end{bmatrix} = \begin{bmatrix} E & 0 \\ -G_{22}^{-1}G_{21} & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial}{\partial p} U_1^n \\ \frac{\partial}{\partial p} U_2^n \\ \frac{\partial}{\partial p} i_E^n \end{bmatrix} - J^{-1} \cdot \begin{bmatrix} \frac{\partial F_1}{\partial p} \\ \frac{\partial F_2}{\partial p} \\ \frac{\partial F_3}{\partial p} \end{bmatrix}. \quad (25)$$

Исходя из того, что подматрицы

$$E = [1], \quad G_{22} = \begin{bmatrix} g & 1 \\ 1 & 0 \end{bmatrix}, \quad G_{21} = \begin{bmatrix} -g \\ 0 \end{bmatrix}, \quad (26)$$

получим

$$G_{22}^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & -g \end{bmatrix}, \quad G_{22}^{-1}G_{21} = \begin{bmatrix} 0 \\ -g \end{bmatrix}. \quad (27)$$

Отсюда

$$\begin{aligned} \begin{bmatrix} \frac{\partial}{\partial p} U_1^{n+1} \\ \frac{\partial}{\partial p} U_2^{n+1} \\ \frac{\partial}{\partial p} i_E^{n+1} \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ g & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial}{\partial p} U_1^n \\ \frac{\partial}{\partial p} U_2^n \\ \frac{\partial}{\partial p} i_E^n \end{bmatrix} - J^{-1} \cdot \begin{bmatrix} \frac{\partial F_1}{\partial p} \\ \frac{\partial F_2}{\partial p} \\ \frac{\partial F_3}{\partial p} \end{bmatrix} = \\ &= \begin{bmatrix} \frac{\partial}{\partial p} U_1^n \\ g \cdot \frac{\partial}{\partial p} U_1^n \end{bmatrix} - J^{-1} \cdot \begin{bmatrix} \frac{\partial F_1}{\partial p} \\ \frac{\partial F_2}{\partial p} \\ \frac{\partial F_3}{\partial p} \end{bmatrix}. \end{aligned} \quad (28)$$

Если нас интересуют функции чувствительности напряжения U_1 , то достаточно определить строку матрицы J^{-1} , которая равна

$$\begin{bmatrix} \frac{1}{\frac{1}{h}C + g} & 0 & \frac{g}{\frac{1}{h}C + g} \end{bmatrix}, \quad (29)$$

и значения матриц производных функций по параметрам C и g :

$$\frac{\partial F}{\partial C} = \begin{bmatrix} \frac{\partial U_1}{\partial t} \\ 0 \\ 0 \end{bmatrix}, \quad \frac{\partial F}{\partial g} = \begin{bmatrix} U_1 - E \\ E - U_1 \\ 0 \end{bmatrix}. \quad (30)$$

Выражения для функций чувствительности по параметрам C и g имеют вид:

$$\begin{aligned} \frac{\partial}{\partial C} U_1^{n+1} &= \frac{\partial}{\partial C} U_1^n - \frac{1}{\frac{1}{h}C + g} \cdot \frac{\partial U_1}{\partial t} = \\ &= \frac{\partial}{\partial C} U_1^n - \frac{Eg}{\left(\frac{1}{h}C + g\right)C} \cdot e^{-\frac{tg}{C}}; \end{aligned} \quad (31)$$

$$\frac{\partial}{\partial g} U_1^{n+1} = \frac{\partial}{\partial g} U_1^n + \frac{E}{\frac{1}{h}C + g} \cdot e^{-\frac{tg}{C}}. \quad (32)$$

Точные значения определяются выражениями (19) и (20).

При получении выражений (31) и (32) использованы точные выражения для напряжения (18) и его производной (21), которые в программах моделирования определяются приближенно в процессе интегрирования по времени.

Запишем выражения (31) и (32) в следующем виде:

$$\begin{aligned} \frac{\partial}{\partial C} U_1^{n+1} &= -\frac{Egh}{C^2 + hgC} \cdot \sum_{i=1}^n e^{-\frac{ihg}{C}}, \\ \frac{\partial}{\partial g} U_1^{n+1} &= -\frac{Eh}{C + hg} \cdot \sum_{i=1}^n e^{-\frac{ihg}{C}}. \end{aligned} \quad (33)$$

Суммирование заменим интегралом:

$$\sum_{i=1}^n e^{-\frac{ihg}{C}} \approx \frac{1}{h} \int_{t=0}^t e^{-\frac{gt}{C}} dt = \frac{C}{hg} \left(1 - e^{-\frac{gt}{C}} \right), \quad (34)$$

где $t = ih$ и $dt \approx h$.

После подстановки, получим

$$\begin{aligned} \frac{\partial}{\partial C} U_1^{n+1} &\approx -\frac{Eh}{C + hg} \cdot \left(1 - e^{-\frac{gt}{C}} \right), \\ \frac{\partial}{\partial g} U_1^{n+1} &= -\frac{EC}{(C + hg)g} \cdot \left(1 - e^{-\frac{gt}{C}} \right). \end{aligned} \quad (35)$$

Относительные погрешности ξ_C и ξ_g имеют в данном случае одинаковый вид:

УДК 621.396.677.833.2

СРЕДНЯЯ ИНТЕНСИВНОСТЬ ПОЛЯ ВБЛИЗИ ФОКУСА ПАРАБОЛИЧЕСКОГО РЕФЛЕКТОРА С КРУПНОМАСШТАБНЫМИ ШЕРОХОВАТОСТЯМИ

Р.В.Воробьев

Рассматривается влияние статистических характеристик крупномасштабной шероховатости отражающей поверхности параболического рефлектора на распределение средней интенсивности поля в поперечной фокальной плоскости.

Розглядається вплив статистичних характеристик великомасштабної шорсткості відбиваючої поверхні параболического

$$\xi = 1 - \frac{C^2}{(C + hg)gt} \left(e^{\frac{gt}{C}} - 1 \right). \quad (36)$$

Для $C \ll hg$ и $\frac{gt}{C} \ll 1$, имеем $C + hg \approx C$ и $e^{\frac{gt}{C}} \approx 1 + \frac{gt}{C}$, отсюда видно, что $\xi \approx 0$. При заданном ξ , используя третий член разложения экспоненты, получим $\xi \approx \frac{1}{2} \cdot \frac{gt}{C}$.

Таким образом, получаемая погрешность зависит от значений параметров, интервала времени и выбранного шага. В данной работе общий случай не рассматривался.

ВЫВОДЫ

Поиск путей определения функций чувствительность по многим параметрам в динамических режимах представляет собой актуальную задачу. В некоторых случаях выбор шага интегрирования позволяет за один дополнительный просчет определять производные по всем варьируемым параметрам без использования интегрирования в обратном порядке времени. В работе приведен алгоритм, реализующий такой подход.

ПЕРЕЧЕНЬ ССЫЛОК

1. Петренко А.И. Основы автоматизации проектирования. - К.: Техніка, 1982. - 295 с.
2. Основы построения систем автоматизированного проектирования. Петренко А.И., Семенов О.И. - К.: Вища школа. Головное изд-во, 1984. - 296 с.
3. Табличные методы моделирования электронных схем на ЭЦВМ. Петренко А.И., Власов А.И., Тимченко А.П. - К.: Вища школа, 1977. - 192 с.
4. Вершина А.И., Кузьмина Л.В. Определение функций чувствительности в статике // "Радіоелектроніка. Інформатика. Управління". - Запоріжжя: ЗДТУ. - 2000. - №1. - с.9-12.
5. Вершина А.И., Маркин А.Г. Определение функций чувствительности в динамике // "Радіоелектроніка, інформатика, управління". - Запоріжжя: ЗДТУ. - 2001. - №1. - с.4-8.
6. Сигорский В.П. Математический аппарат инженера. - К.: Техніка, 1977. - 768с.

ного рефлектора на розподіл середньої інтенсивності поля у поперечній фокальній площині.

The influence of the statistical characteristics of a large-scale roughness of the reflecting surface of the parabolic reflector on the distribution of mean intensity of a field in the transversal focal plane is considered.

ВВЕДЕНИЕ

Параболические рефлекторы антенн радиотелескопов, станций спутниковой связи миллиметрового диапазона и гелиоконцентраторов способны фокусировать мощное радиотепловое излучение. При исследовании влияния такой фокусировки на качество работы системы необходимо знать распределение интенсивности поля в поперечной фокальной плоскости. Для идеального параболоида данная задача решена в работе [1].

Однако отражающая поверхность реальных параболических рефлекторов имеет случайные неровности. В общем случае такие неровности представляют собой многомасштабную шероховатость и зависят от производственного и эксплуатационного факторов. В процессе изготовления параболоида на его отражающей поверхности возникает двухмасштабная шероховатость, присущая всем рефлекторам, которая представляет собой наложение мелкомасштабных неровностей на крупномасштабные. Такая шероховатость обусловлена технологией изготовления параболического профиля и видом обработки поверхности металла. Степень шероховатости поверхности зависит от соотношения геометрических характеристик неровности и длины волны, и оценивается параметром Рэлея.

Целью настоящей работы является исследование влияния статистических характеристик крупномасштабной шероховатости отражающей поверхности параболического рефлектора на распределение средней плотности мощности в поперечной фокальной плоскости.

ПОСТАНОВКА ЗАДАЧИ

Рассмотрим нормальное падение плоской линейно-поляризованной волны в раскрыв параболоида. На рисунке 1 представлена схема отражения волны от параболического рефлектора. Уравнение параболоида в сферической системе координат с центром в фокусе запишем как

$$R = f \cdot \sec^2\left(\frac{\theta}{2}\right), \quad (1)$$

где R - расстояние от фокуса до точки на параболоиде; f - фокусное расстояние; θ - угол между фокальной осью и направлением от фокуса до рассматриваемой точки на параболоиде.

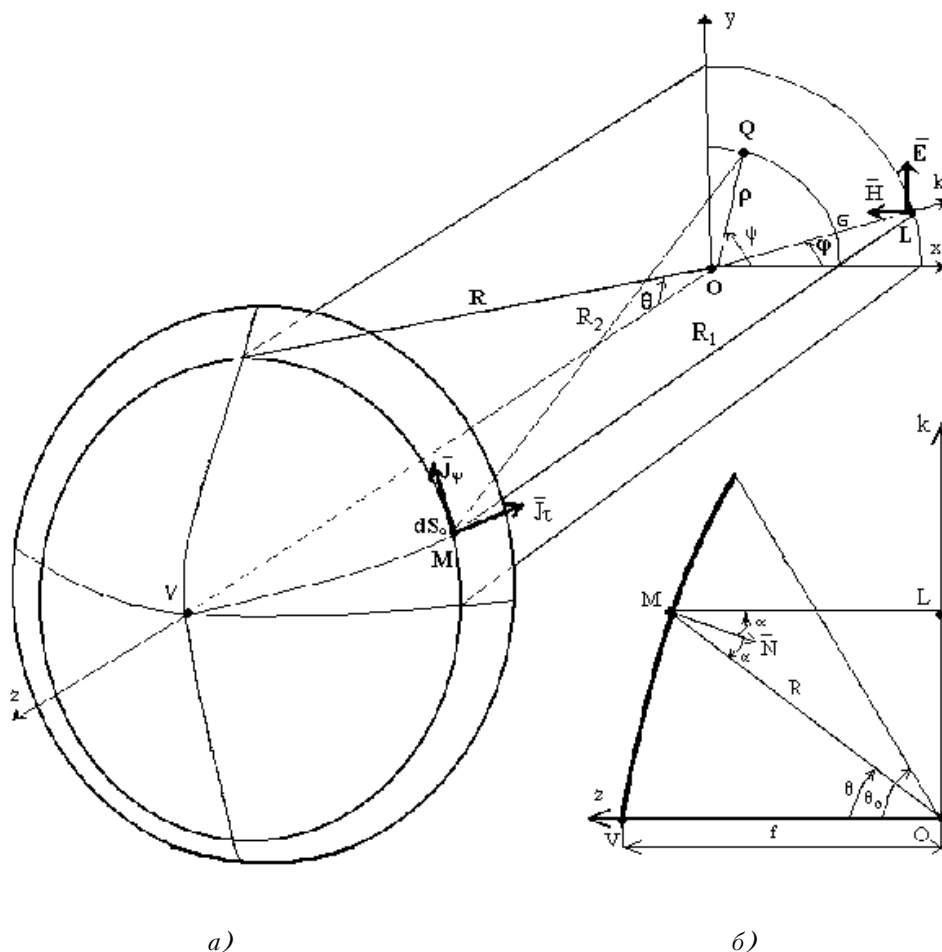


Рисунок 1 - Схема отражения плоской волны от параболоида: а) система координат; б) плоскость падения

Поверхность рефлектора принята идеально проводящей. Величина электрических поверхностных токов, распределенных по теневой стороне рефлектора, считается равной нулю. На отражающей поверхности имеется пологая статистически изотропная и однородная крупномасштабная нормальная шероховатость с нулевым средним значением высоты. Будем считать, что при таких условиях затенения неровностей будут отсутствовать.

Требуется определить среднее значение модуля z -оставляющей вектора Пойтинга в точке $Q(\rho, \gamma, z)$ (рис.1), принадлежащей поперечной фокальной плоскости рефлектора.

ПОЛЕ, СФОКУСИРОВАННОЕ ПАРАБОЛОИДОМ

Поле, рассеянное рефлектором в зоне Фраунгофера описывается уравнениями Кирхгофа-Котлера [2], которые получают путем введения поправки Котлера в уравнении Стреттона [3]:

$$\left. \begin{aligned} \mathbf{E}' &= \frac{-j\omega\mu}{4\pi} \int_S (\mathbf{J}_{\text{Э}} - (\mathbf{J}_{\text{Э}} \nabla r) \nabla r) \Psi dS \\ \mathbf{H}' &= \frac{1}{4\pi} \int_S (\mathbf{J}_{\text{Э}} \times \nabla \Psi) dS \end{aligned} \right\}, \quad (2)$$

где \mathbf{E}' - вектор электрического поля, рассеянного поверхностью S в зоне Фраунгофера; \mathbf{H}' - вектор магнитного поля, рассеянного поверхностью S в зоне Фраунгофера; $\mathbf{J}_{\text{Э}}$ - поверхностный электрический ток; ∇r - орт, направленный в сторону распространения волны; μ - относительная магнитная проницаемость среды распространения; ω - частота, падающего электромагнитного поля; Ψ - функция Грина; ∇ - оператор набла.

Интегрирование производится по отражающей поверхности параболоида. Из уравнений Кирхгофа-Котлера следует, что поле в зоне Фраунгофера рефлектора является поперечным.

Задачи дифракции волн на крупномасштабной шероховатости решаются по методу Кирхгофа. Для применения данного метода перепишем первое уравнение (2) с учетом обозначений, приведенных на рисунке 2:

$$\mathbf{E}' = \frac{-jkZ_0}{2\pi} \int_S \left[((\mathbf{nH}) - ((\mathbf{nH}) \nabla \tilde{R}_2) \nabla \tilde{R}_2) \frac{e^{jk(\tilde{R}_1 + \tilde{R}_2)}}{\tilde{R}_1 \cdot \tilde{R}_2} \right] dS, \quad (3)$$

где Z_0 - характеристическое сопротивление среды; k - волновое число; \mathbf{n} - нормаль к шероховатой отражающей поверхности; \mathbf{H} - вектор падающего магнитного поля; \tilde{R}_1 - расстояние от источника до точки на неровности параболоида; \tilde{R}_2 - расстояние от точки на неровности параболоида до точки наблюдения.

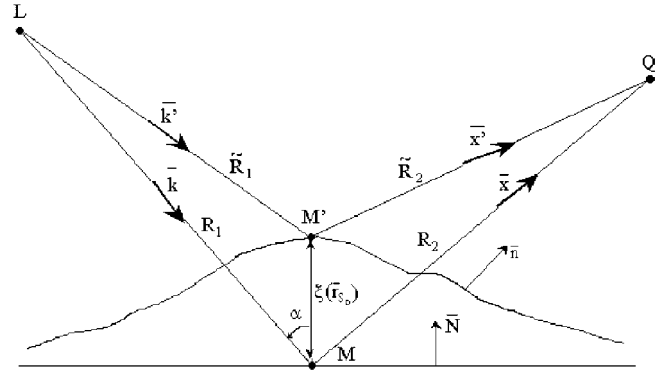


Рисунок 2 - Рассеяние ЭВМ крупномасштабной шероховатостью

Расстояния \tilde{R}_1 и \tilde{R}_2 удобно выразить через расстояния R_1 и R_2 от идеальной поверхности S_0 . Запишем данные преобразования, согласно работе [4], для случая, когда идеальная поверхность является неплоской (параболической) и среднеквадратическая высота неровностей значительно меньше, чем продольный размер зоны Френеля:

$$\tilde{R}_1 + \tilde{R}_2 \approx R_1 + R_2 - \mathbf{N}(\mathbf{x} - \mathbf{k}) \xi(\mathbf{r}_s) = R_{\Sigma}, \quad (4)$$

где R_1 - расстояние от источника до точки на параболоиде; R_2 - расстояние от точки на параболоиде до точки наблюдения; \mathbf{N} - нормаль к поверхности идеального параболоида; \mathbf{x} - орт, направленный в сторону распространения волны, отраженной от идеального параболоида; \mathbf{k} - орт, направленный в сторону распространения волны, падающей на идеальный параболоид; $\xi(\mathbf{r}_s)$ - высота неровности в точке отражения.

Перепишем соотношение (3) с учетом произведенных преобразований:

$$\mathbf{E}' = \frac{-jkZ_0}{2\pi} \int_{S_0} \left[((\mathbf{nH}) - ((\mathbf{nH}) \nabla R_2) \nabla R_2) \frac{e^{jkR_{\Sigma}}}{R_1 R_2 (\mathbf{nN})} \right] dS. \quad (5)$$

При этом интегрирование производится по идеальной поверхности S_0 параболического рефлектора.

Падающее на элемент отражающей поверхности рефлектора dS магнитное поле \mathbf{H} в точке R , θ , ϕ (сферическая система координат с центром в фокусе параболоида), разложим на составляющие в полярной системе координат с центром в фокусе параболоида (рис.1):

$$\left. \begin{aligned} \mathbf{H}_{\sigma} &= -\mathbf{H} \cos(\phi) \\ \mathbf{H}_{\phi} &= \mathbf{H} \sin(\phi) \end{aligned} \right\}. \quad (6)$$

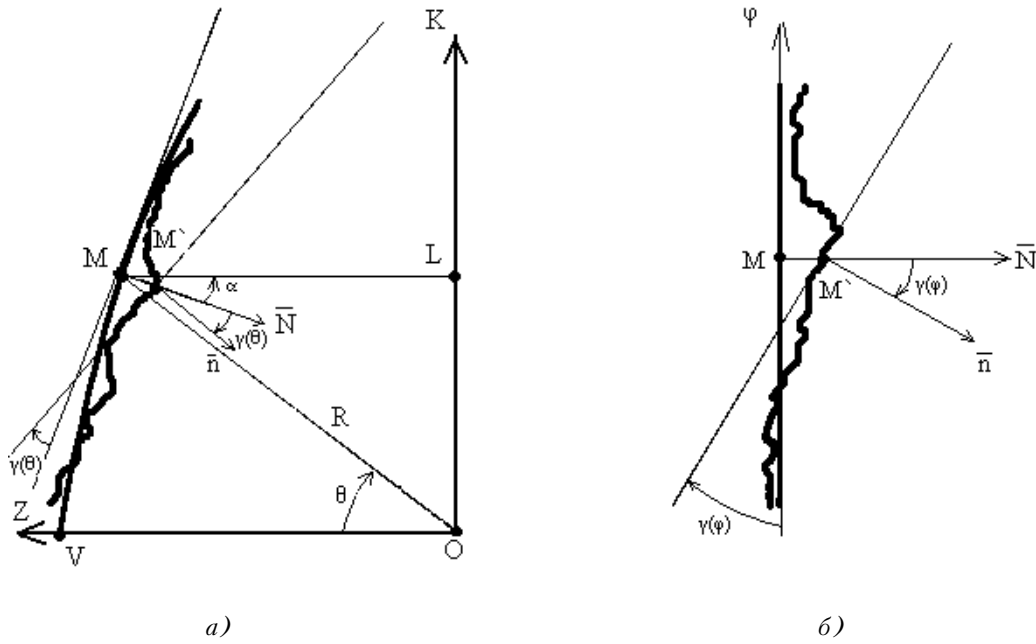


Рисунок 3 - Геометрия крупномасштабной шероховатости: а) вдоль координаты θ ; б) вдоль координаты φ

Если α - угол падения к нормали \mathbf{N} , то для статистически изотропной шероховатой отражающей поверхности параболоида (рис.3) можно записать составляющие поверхностной плотности электрического тока как:

$$\left. \begin{aligned} \mathbf{J}_\tau &= 2\mathbf{H}\sin(\varphi)\cos(\gamma(\varphi, \theta)) \\ \mathbf{J}_\varphi &= 2\mathbf{H}\cos(\varphi)\cos(\alpha + \gamma(\varphi, \theta)) \end{aligned} \right\} \quad (7)$$

Величина $\gamma(\varphi, \theta)$ характеризует угол наклона крупномасштабной шероховатости в точке φ, θ идеального параболоида.

Определим поперечные составляющие поля, рассеянного в направлении \mathbf{x} в сферической системе координат r, β, δ , с центром на элементе отражающей поверхности рефлектора (рис.4).

Поле сферической волны, рассеянной элементом отражающей поверхности dS , определим из соотношения (5). Учтем, что на расстоянии R_1 , соизмеримого с размерами рефлектора, амплитуда падающего поля слабо меняется:

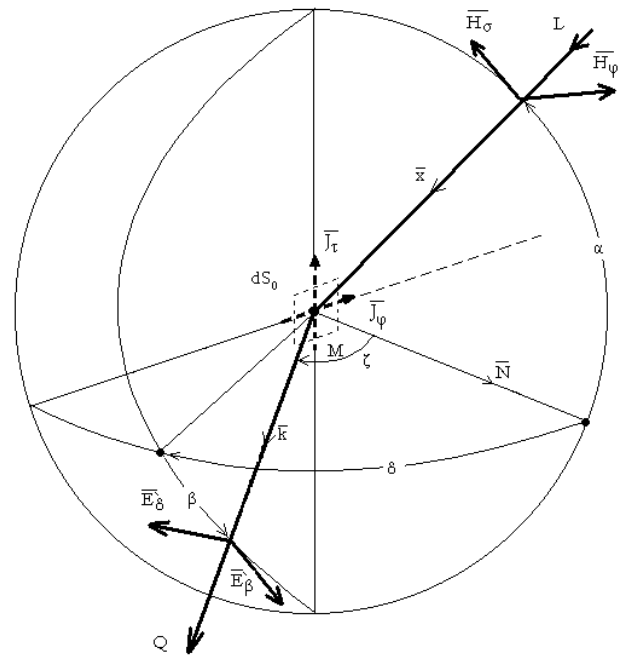


Рисунок 4 - Сферическая система координат на элементе поверхности dS_0

$$\left. \begin{aligned} \mathbf{E}'_\beta &= \frac{jk\mathbf{E}}{2\pi} \left[\sin(\varphi)\cos(\gamma(\varphi, \theta))\cos(\beta) + \right. \\ &\quad \left. + \cos(\varphi)\cos(\alpha + \gamma(\varphi, \theta))\sin(\delta) \right] \frac{e^{jkR_\Sigma}}{R_2(\mathbf{nN})} dS_0 \\ \mathbf{E}'_\delta &= \frac{jk\mathbf{E}}{2\pi} \left[\cos(\varphi)\cos(\alpha + \gamma(\varphi, \theta))\cos(\delta) \right] \frac{e^{jkR_\Sigma}}{R_2(\mathbf{nN})} dS_0 \\ \mathbf{H}'_\beta &= \frac{\mathbf{E}'_\delta}{Z_0} \quad \mathbf{H}'_\delta = -\frac{\mathbf{E}'_\beta}{Z_0} \end{aligned} \right\} \quad (8)$$

где \mathbf{E} - вектор падающего электрического поля.

При определении отраженного поля вблизи оси симметрии рефлектора возможно систему уравнений (8) упростить. Произведем аппроксимацию плоскостью фазового фронта отраженной волны в окрестности оси симметрии рефлектора.

В точке O (рис.1) фаза поля определяется выражением:

$$\Phi_O = k(R\cos(\theta) + R). \quad (9)$$

Согласно работе [1], фаза поля в точке Q можно записать как:

$$\Phi_Q = \Phi_O - k\rho\cos(\varphi - \psi)\sin(\theta). \quad (10)$$

Так как диаграмма направленности элемента отражающей поверхности рефлектора широкая, то можно предположить, что амплитуда поля в окрестности фокуса одинаковая. Согласно работе [1], при $R/\lambda > 50$ амплитудная ошибка не превышает 5% в области аппроксимации плоской волной. Следовательно, в уравнениях (8) для амплитуды можно принять $\delta = 0$, $\beta = \alpha$, $R_2 = R$.

Перепишем уравнения (8) в виде:

$$\left. \begin{aligned} \mathbf{E}'_\beta &= A\sin(\varphi)e^{j\Phi_{Q'}}d\varphi d\theta \\ \mathbf{E}'_\delta &= \frac{A\cos(\varphi)\cos(\alpha + \gamma(\varphi, \theta))}{\cos(\alpha)\cos(\gamma(\varphi, \theta))}e^{j\Phi_{Q'}}d\varphi d\theta \end{aligned} \right\}, \quad (11)$$

где $A = \frac{jk\sigma\mathbf{E}}{2\pi}$; $\Phi_{Q'}$ - набег фазы на расстоянии R_Σ :

$\Phi_{Q'} = \Phi_O - k[(\cos(\zeta) + \cos(\alpha)) \cdot \xi(\varphi, \theta)]$; ζ - угол между нормалью к параболоиду и направлением к точке наблюдения Q .

При условии $\rho/f \ll 1$ $\cos(\zeta)$ можно заменить на $\cos(\alpha)$.

Для определения полного рассеянного рефлектором поля в точке Q разложим поля в цилиндрической системе координат (ρ, ψ, z) (рис.5). В зависимости от составляющей \mathbf{E}'_β , параллельной плоскости падения, поле запишем как:

$$\left. \begin{aligned} \mathbf{E}'_\rho &= -\mathbf{E}'_\beta \cdot \cos(\theta)\cos(\varphi - \psi) \\ \mathbf{E}'_\psi &= -\mathbf{E}'_\beta \cdot \cos(\theta)\sin(\varphi - \psi) \\ \mathbf{E}'_z &= \mathbf{E}'_\beta\sin(\theta) \end{aligned} \right\}. \quad (12)$$

В зависимости от составляющей \mathbf{E}'_δ , перпендикулярной к плоскости падения, поле в точке Q запишем как:

$$\left. \begin{aligned} \mathbf{E}'_\rho &= \mathbf{E}'_\delta \cdot \sin(\varphi - \psi) \\ \mathbf{E}'_\psi &= -\mathbf{E}'_\delta \cdot \cos(\varphi - \psi) \\ \mathbf{E}'_z &= 0 \end{aligned} \right\}. \quad (13)$$

В соотношения (12) подставим \mathbf{E}'_β из (11) и произведем интегрирование по всей поверхности рефлектора:

$$\left. \begin{aligned} \mathbf{E}'_\rho &= -A \int_0^{\theta_0} \left[\cos(\theta) \int_0^{2\pi} \sin(\varphi)\cos(\varphi - \psi) \cdot e^{j\Phi_{Q'}} d\varphi \right] d\theta \\ \mathbf{E}'_\psi &= -A \int_0^{\theta_0} \left[\cos(\theta) \int_0^{2\pi} \sin(\varphi)\sin(\varphi - \psi) \cdot e^{j\Phi_{Q'}} d\varphi \right] d\theta \\ \mathbf{E}'_z &= A \int_0^{\theta_0} \left[\sin(\theta) \int_0^{2\pi} \sin(\varphi) \cdot e^{j\Phi_{Q'}} d\varphi \right] d\theta \\ \mathbf{H}'_\rho &= \frac{\mathbf{E}'_\psi}{Z_\omega \cdot \cos(\theta)} \quad \mathbf{H}'_\psi = \frac{-\mathbf{E}'_\rho}{Z_\omega \cdot \cos(\theta)} \quad \mathbf{H}'_z = 0 \end{aligned} \right\}. \quad (14)$$

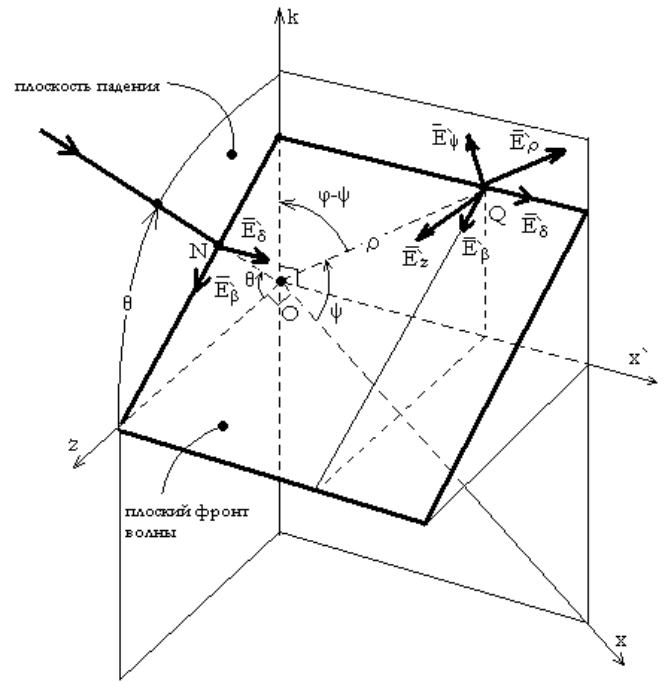


Рисунок 5 - Составляющие отраженного поля в точке Q вблизи оси рефлектора

В соотношения (13) подставим \mathbf{E}'_δ из (11) и произведем интегрирование по всей поверхности рефлектора:

$$\left. \begin{aligned} \mathbf{E}'_\rho &= A \int_0^{\theta_0} \left[B \int_0^{2\pi} \frac{\cos(\varphi)\sin(\varphi - \psi)}{\cos(\gamma(\varphi, \theta))} e^{j\Phi_{Q'}} d\varphi \right] d\theta \\ \mathbf{E}'_\psi &= -A \int_0^{\theta_0} \left[B \int_0^{2\pi} \frac{\cos(\varphi)\cos(\varphi - \psi)}{\cos(\gamma(\varphi, \theta))} e^{j\Phi_{Q'}} d\varphi \right] d\theta \\ \mathbf{H}'_z &= \frac{A}{Z_0} \int_0^{\theta_0} \left[\sin(\theta) B \int_0^{2\pi} \frac{\cos(\varphi)}{\cos(\gamma(\varphi, \theta))} e^{j\Phi_{Q'}} d\varphi \right] d\theta \\ \mathbf{E}'_z &= 0 \quad \mathbf{H}'_\rho = \frac{\mathbf{E}'_\psi \cos(\theta)}{Z_\omega} \quad \mathbf{H}'_\psi = \frac{-\mathbf{E}'_\rho \cos(\theta)}{Z_\omega} \end{aligned} \right\}. \quad (15)$$

где θ_0 - половина угла раскрытия параболоида;

$$B = \frac{\cos(\alpha + \gamma(\varphi, \theta))}{\cos(\alpha)}$$

Уравнения поля (14) и (15) описывают соответственно поперечно-магнитную ТМ- и поперечно-электрическую ТЕ-волны, которые распространяются вдоль оси z со скоростью меньшей, чем скорость света. Эти волны имеют различную радиальную периодичность, следовательно, одновременно граничные условия для обеих волн не выполняются. Полное рассеянное рефлектором поле в точке Q описывается суммой всех составляющих, определенных уравнениями (14) и (15):

$$\left. \begin{aligned} & -B \int_0^{2\pi} \frac{\cos(\varphi)\sin(\varphi - \psi)}{\cos(\gamma(\varphi, \theta))} e^{j\Phi_{Q'}} d\varphi \Big] d\theta \\ \mathbf{E}'_{\psi} = & -A \int_0^{\theta_0} \left[\int_0^{2\pi} \cos(\theta) \sin(\varphi) \sin(\varphi - \psi) \cdot e^{j\Phi_{Q'}} d\varphi + \right. \\ & \left. + B \int_0^{2\pi} \frac{\cos(\varphi)\cos(\varphi - \psi)}{\cos(\gamma(\varphi, \theta))} e^{j\Phi_{Q'}} d\varphi \right] d\theta \\ \mathbf{E}'_z = & A \int_0^{\theta_0} \left[\int_0^{2\pi} \sin(\theta) \sin(\varphi) \cdot e^{j\Phi_{Q'}} d\varphi \right] d\theta \\ \mathbf{H}'_{\rho} = & -\frac{A}{Z_0} \int_0^{\theta_0} \left[\int_0^{2\pi} \sin(\varphi) \sin(\varphi - \psi) \cdot e^{j\Phi_{Q'}} d\varphi + \right. \\ & \left. + B \cos(\theta) \int_0^{2\pi} \frac{\cos(\varphi)\cos(\varphi - \psi)}{\cos(\gamma(\varphi, \theta))} e^{j\Phi_{Q'}} d\varphi \right] d\theta \\ \mathbf{H}'_{\psi} = & \frac{A}{Z_0} \int_0^{\theta_0} \left[\int_0^{2\pi} \sin(\varphi) \cos(\varphi - \psi) \cdot e^{j\Phi_{Q'}} d\varphi - \right. \\ & \left. - B \cos(\theta) \int_0^{2\pi} \frac{\cos(\varphi)\sin(\varphi - \psi)}{\cos(\gamma(\varphi, \theta))} e^{j\Phi_{Q'}} d\varphi \right] d\theta \end{aligned} \right\} \cdot (16)$$

Это приводит к наличию комбинированной волны, распространяющейся вдоль оси рефлектора.

ИНТЕНСИВНОСТИ ПОЛЯ ВБЛИЗИ ФОКУСА

Средняя за период колебания плотность потока мощности сфокусированного поля, проходящего через поперечную фокальную плоскость, определяется реальной частью z -составляющей комплексного вектора Пойтинга:

$$\mathbf{S}'_z = \frac{1}{2} \operatorname{Re}\{(\mathbf{E}'_{\rho} \times \mathbf{H}'^*_{\psi}) - (\mathbf{E}'_{\psi} \times \mathbf{H}'^*_{\rho})\}, \quad (17)$$

где * - знак комплексно-сопряженной величины; Re - реальная часть комплексной величины.

Учитывая ортогональность векторов \mathbf{E}'_{ρ} и \mathbf{H}'_{ψ} , а также векторов \mathbf{H}'_{ρ} и \mathbf{E}'_{ψ} , модуль плотности потока мощности запишем как:

$$|\mathbf{S}'_z| = \frac{1}{2} \operatorname{Re}\{(\mathbf{E}'_{\rho} \cdot \mathbf{H}'^*_{\psi}) - (\mathbf{E}'_{\psi} \cdot \mathbf{H}'^*_{\rho})\}. \quad (18)$$

Учитывая уравнение параболоида (1), запишем ρ - и ψ -составляющие электрического и магнитного поля

$$\left. \begin{aligned} \mathbf{E}'_{\rho} = & -C \int_0^{\theta_0} \int_0^{2\pi} \left[\begin{aligned} & \cos(\theta) \sin(\varphi) \cos(\varphi - \psi) - \\ & -D \cos(\varphi) \sin(\varphi - \psi) \end{aligned} \right] e^{\Phi_{Q'}} d\varphi d\theta \\ \mathbf{H}'^*_{\psi} = & -\frac{C}{Z_0} \int_0^{\theta_0} \int_0^{2\pi} \left[\begin{aligned} & \sin(\varphi) \cos(\varphi - \psi) - \\ & -D \cos(\theta) \cos(\varphi) \sin(\varphi - \psi) \end{aligned} \right] e^{-\Phi_{Q'}} d\varphi d\theta \\ \mathbf{E}'_{\psi} = & -C \int_0^{\theta_0} \int_0^{2\pi} \left[\begin{aligned} & \cos(\theta) \sin(\varphi) \sin(\varphi - \psi) + \\ & + D \cos(\varphi) \cos(\varphi - \psi) \end{aligned} \right] e^{\Phi_{Q'}} d\varphi d\theta \\ \mathbf{H}'^*_{\rho} = & \frac{C}{Z_0} \int_0^{\theta_0} \int_0^{2\pi} \left[\begin{aligned} & \sin(\varphi) \sin(\varphi - \psi) + \\ & + D \cos(\theta) \cos(\varphi) \cos(\varphi - \psi) \end{aligned} \right] e^{-\Phi_{Q'}} d\varphi d\theta \end{aligned} \right\} \cdot (19)$$

где $C = \frac{2j\mathbf{E}f}{\lambda}$;

$$D = 1 - \operatorname{tg}\left(\frac{\theta}{2}\right) \operatorname{tg}(\gamma(\varphi, \theta));$$

$$\Phi_{Q'} = jk(2f - \rho \sin(\theta) \cos(\varphi - \psi) - 2 \cos\left(\frac{\theta}{2}\right) \xi(\varphi, \theta)).$$

Первое слагаемое выражения (18) описывает распределение интенсивности поля в направлении оси y (рис.1), соответственно, второе слагаемое - вдоль оси x . Исходя из симметрии рефлектора, распределение интенсивности поля в поперечной фокальной плоскости одинаково во всех направлениях. Таким образом, достаточно рассмотреть одно из слагаемых выражения (18).

Тогда распределение интенсивности поля в поперечной фокальной плоскости идеального параболоида записывается в виде:

$$\mathbf{S}'_z = \frac{2}{Z_0} (\mathbf{E}fk)^2 \int_0^{\theta_0} \int_0^{2\pi} \left[\begin{aligned} & \operatorname{tg}\left(\frac{\theta}{2}\right) \cdot \left(\cos(\theta) \frac{dJ_1(U)}{dU} + \frac{J_1(U)}{U} \right) \times \\ & \times \operatorname{tg}\left(\frac{\theta}{2}\right) \cdot \left(\frac{dJ_1(U')}{dU'} + \cos(\theta') \frac{J_1(U')}{U'} \right) \end{aligned} \right] d\theta d\varphi, \quad (20)$$

где $U = k\rho \sin(\theta)$ и $U' = k\rho \sin(\theta')$.

На рисунке 7 представлены распределения интенсивности поля в радиальном направлении от фокуса в поперечной фокальной плоскости.

При учете шероховатости отражающей поверхности рефлектора распределение интенсивности поля в поперечной

фокальной плоскости будет также симметрично относительно фокуса. Запишем выражение для распределения интенсивности поля вблизи фокуса параболоида, на отражающей поверхности которого расположена крупномасштабная шероховатость. При этом будем считать $\psi = 0$, т.е. воспользуемся вторым слагаемым выражения (18) и найдем распределение интенсивности поля вдоль оси x .

$$|\bar{S}_z| = \frac{2}{Z_0} \left(\frac{Ef}{\lambda} \right)^2 \int_0^{\theta} \int_0^{\theta} \left[\begin{array}{l} \theta_0 \theta_0 \left[\text{tg}\left(\frac{\theta}{2}\right) \text{tg}\left(\frac{\theta'}{2}\right) \times \right. \\ \left. \int_0^{\varphi} \int_0^{\varphi'} \left[\begin{array}{l} 2\pi 2\pi \left[\cos(\theta) \sin^2(\varphi) + D \cos^2(\varphi) \right] \times \\ \left[\sin^2(\varphi') + D' \cos(\theta') \cos^2(\varphi') \right] \times \\ \left. \times e^{j k \rho (\varphi' - \varphi)} d\varphi d\varphi' \right] \right] \end{array} \right] d\theta d\theta', \quad (21)$$

где $D' = 1 - \text{tg}\left(\frac{\theta}{2}\right) \text{tg}(\gamma(\varphi', \theta'))$;

$$\Phi'_{\varrho'} = jk(2f - \rho \sin(\theta')) \cos(\varphi' - \psi') - 2 \cos\left(\frac{\theta'}{2}\right) \xi(\varphi', \theta').$$

В выражение (21) входят случайные величины $\gamma = \gamma(\varphi, \theta)$, $\gamma' = \gamma(\varphi', \theta')$ и $\xi = \xi(\varphi, \theta)$, $\xi' = \xi(\varphi', \theta')$, характеризующие, соответственно, угол наклона и высоту неровностей отражающей поверхности рефлектора в двух различных точках. Для определения средней интенсивности поля необходимо усреднить соотношение (21). При этом следует учесть, что для однородной шероховатости $\gamma(\varphi, \theta)$ и $\xi(\varphi, \theta)$ являются статистически независимыми величинами [4].

Распределение средней интенсивности поля вдоль оси x можно записать как:

$$|\bar{S}_z| = \frac{2}{Z_0} \left(\frac{Ef}{\lambda} \right)^2 \int_0^{\theta} \int_0^{\theta} \left[\begin{array}{l} \theta_0 \theta_0 \left[\text{tg}\left(\frac{\theta}{2}\right) \text{tg}\left(\frac{\theta'}{2}\right) \times \right. \\ \left. \int_0^{\varphi} \int_0^{\varphi'} \left[\begin{array}{l} S_{\gamma} S_{\xi} e^{-jk\rho \sin(\theta) \cos(\varphi)} \times \\ \left. \times e^{jk\rho \sin(\theta') \cos(\varphi')} d\varphi d\varphi' \right] \right] \end{array} \right] d\theta d\theta', \quad (22)$$

$$\text{где } S_{\xi} = e^{-2jk \left(\cos\left(\frac{\theta}{2}\right) \xi - \cos\left(\frac{\theta'}{2}\right) \xi' \right)}; \quad (23)$$

$$\begin{aligned} S_{\gamma} = & (\cos(\theta) \sin^2(\varphi) + \cos^2(\varphi)) \cdot (\sin^2(\varphi') + \cos(\theta') \cos^2(\varphi')) - \\ & - (\sin^2(\varphi') + \cos(\theta') \cos^2(\varphi')) \cos^2(\varphi) \text{tg}\left(\frac{\theta}{2}\right) \overline{\text{tg}(\gamma)} - \\ & - (\cos(\theta) \sin^2(\varphi) + \cos^2(\varphi)) \cos(\theta') \cos^2(\varphi') \text{tg}\left(\frac{\theta'}{2}\right) \overline{\text{tg}(\gamma')} + \\ & + \text{tg}\left(\frac{\theta}{2}\right) \text{tg}\left(\frac{\theta'}{2}\right) \cos^2(\varphi) \cos^2(\varphi') \cos(\theta') \overline{\text{tg}(\gamma) \text{tg}(\gamma')}. \end{aligned} \quad (24)$$

Чертой обозначено математическое ожидание случайной величины. Рассмотрим отдельно выражение (23) и (24).

Выражение (23) представляет собой двухмерную характеристическую функцию нормальных случайных величин. Согласно работе [5], его можно представить как

$$S_{\xi} = e^{-2k^2 \left(\cos^2\left(\frac{\theta}{2}\right) + \cos^2\left(\frac{\theta'}{2}\right) \right) \cdot (1-w) \cdot \sigma^2}, \quad (25)$$

где w - коэффициент корреляции высот шероховатости между двумя точками с координатами φ, θ и φ', θ' ; σ - среднеквадратическое отклонение высоты.

Из выражения (25) видно, что w зависит от сферических координат. Определим выражение для коэффициента корреляции шероховатости, которая расположена на поверхности параболоида в сферической системе координат с центром в фокусе.

Согласно работе [4], коэффициенты корреляции шероховатостей имеют гауссову форму, которая в общем виде записывается как

$$w(x_1, x_2) = \exp\left\{-\left(\frac{x_1 - x_2}{\rho_k}\right)^2\right\}, \quad (26)$$

где x_1 и x_2 - координаты вдоль поверхности; ρ_k - радиус корреляции.

Чтобы исключить из рассмотрения поверхность второго порядка кривизны (параболоид), целесообразно рассматривать шероховатость на плоской эквивалентной поверхности (круге). Для этого необходимо приравнять площадь отражающей поверхности параболоида и площадь эквивалентной поверхности. При этом радиус эквивалентного круга должен быть равным длине плеча параболы КМ (рис.6).

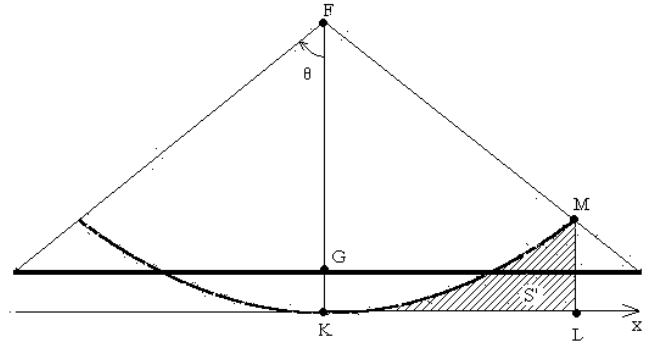


Рисунок 6 - Сечение параболоида и эквивалентная поверхность

Определим площадь, ограниченную плечом параболы и ее проекцией на ось x (на рис.6 заштрихованная площадь S'). Учитывая уравнение параболы в полярной системе координат, запишем:

$$S' = \int_0^{R(\theta_0) \cdot \sin(\theta_0)} \frac{x^2}{4f} dx, \quad (27)$$

где $R(\theta_0)$ - расстояние от фокуса до точки на параболоиде; f - фокусное расстояние; θ_0 - пол угла раскрытия параболоида.

Фигуру KLM можно рассмотреть как треугольник. Зная площадь треугольника S' и две стороны

$$KL = R(\theta_0) \cdot \sin(\theta_0), \text{ и } LM = f - R(\theta_0) \cdot \cos(\theta_0), \quad (28)$$

можно определить сторону KM (плечо параболы) с помощью формулы Герона.

Расстояние от фокуса до эквивалентной поверхности:

$$GF = KM \cdot \text{ctg}(\theta_0). \quad (29)$$

Для изотропной шероховатости в трехмерном случае радиус корреляции является радиусом круга [4]. Учитывая (26), запишем выражение для коэффициента корреляции:

$$w = \exp(-[(GF \cdot \text{tg}(\theta') \cdot \cos(\varphi' - \varphi) - GF \cdot \text{tg}(\theta))^2 + (GF \cdot \text{tg}(\theta') \cdot \sin(\varphi' - \varphi))^2] / \rho_k^2). \quad (30)$$

При рассмотрении выражения (24) следует учесть, что согласно работе [4], для однородной шероховатости в общем виде можно записать:

$$\overline{\text{tg}(\gamma)} = \overline{\text{tg}(\gamma')} = 0$$

$$\overline{\text{tg}(\gamma)\text{tg}(\gamma')} = -\sigma^2 \cdot \frac{d^2 w(x_1 - x_2)}{d(x_1 - x_2)^2}. \quad (31)$$

Для гауссовой корреляционной функции (30) вторая производная равна

$$w'' = ((4 \cdot [GF^2 \cdot (\text{tg}^2(\theta') + \text{tg}^2(\theta)) - (2 \cdot GF \cdot \text{tg}(\theta') \cdot \text{tg}(\theta) \cos(\varphi' - \varphi)] - 2\rho_k^2) / \rho_k^4). \quad (32)$$

Учитывая выражения (22-25, 30-32), запишем окончательное выражение для распределения средней интенсивности поля вдоль оси x :

$$|\bar{S}_z| = \frac{2}{Z_0} \left(\frac{E_f}{\lambda} \right)^2 \int_0^{\theta_0} \int_0^{\theta_0} \int_0^{2\pi} \int_0^{2\pi} \left[\begin{array}{l} \text{tg}(\frac{\theta}{2}) \text{tg}(\frac{\theta'}{2}) \times \\ \left[\begin{array}{l} (\cos(\theta) \sin^2(\varphi) + \cos^2(\varphi)) \times \\ \times (\sin^2(\varphi') + \cos^2(\varphi') \cos^2(\varphi)) - \\ - \text{tg}(\frac{\theta}{2}) \text{tg}(\frac{\theta'}{2}) \cos^2(\varphi) \cos^2(\varphi') \cos^2(\theta) \sigma^2 w'' \\ \times e^{-2k^2 \left(\cos^2(\frac{\theta}{2}) + \cos^2(\frac{\theta'}{2}) \right) (1-w)\sigma^2} \\ \times e^{-jk\rho(\sin(\theta)\cos(\varphi) - \sin(\theta')\cos(\varphi'))} \end{array} \right] \times \end{array} \right] d\varphi d\varphi' d\theta d\theta'. \quad (33)$$

С помощью численного интегрирования выражения (33) было получено распределение средней интенсивности поля в поперечной фокальной плоскости параболоида (рис.7). Параметры параболоида: диаметр раскрыва - 7,75м; фокусное расстояние - 3,25м; половина угла раскрыва $61,6^\circ$; радиус корреляции шероховатости - 2м; среднеквадратическое отклонение высоты шероховато-

сти - 63 мкм. Рассматриваемая длина волны 80 мкм.

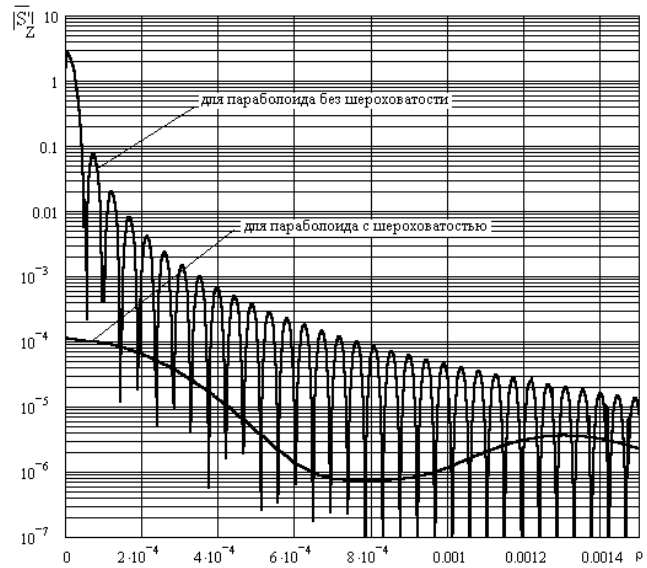


Рисунок 7 - Распределение интенсивности поля в поперечной фокальной плоскости

Таким образом, наличие крупномасштабной шероховатости на отражающей поверхности параболоида приводит к существенному расширению фокального пятна и уменьшению плотности мощности в фокусе.

ЗАКЛЮЧЕНИЕ

В работе получено выражение, описывающее распределение средней интенсивности поля в поперечной фокальной плоскости параболоида при наличии на отражающей поверхности крупномасштабной шероховатости. Крупномасштабные неровности профиля параболоида учтены с помощью метода Кирхгофа. Поле вблизи фокуса определялось в результате решения уравнений Кирхгофа-Котлера.

Результаты исследования могут быть использованы для оценки возможности применения рефлекторов в более коротковолновом диапазоне волн, а также для исследования воздействия радиоизлучения со сплошным спектром на облучатель параболической антенны.

ПЕРЕЧЕНЬ ССЫЛОК

1. Minnett H.C. and Thomas B. MacA. Fields in the image space of symmetrical focusing reflectors./ Proc. IEE, 1968, AP-115, №10, pp.1419-1430.
2. Семенов А.А. Теория электромагнитных волн. -М.: Издательство Московского университета, 1968.-317с.
3. Стрэттон Дж.А. Теория электромагнетизма. -М.: ОГИЗ-Гостехиздат, 1948, - 540с.
4. Басс Ф.Г., Фукс И.М. Рассеяние волн на статистически неровной поверхности. -М.: Наука, 1972, -424с.
5. Ахманов С.А., Дьяков Ю.Е., Чиркин А.С. Введение в статистическую радиофизику и оптику. -М.: Наука, 1981, - 640с.

АВТОНОМНАЯ СИСТЕМА ФУЦЦИ-УПРАВЛЕНИЯ ТЕПЛОСНАБЖЕНИЕМ

В.И.Гостев, В.В.Крайнев, В.В.Гостев

Изложен расчет и представлены результаты математического моделирования автономной системы теплоснабжения с цифровым нечетким регулятором и объектом управления "водогрейный котел + коллектор с трубопроводами + отопляемое здание с помещениями".

Викладено розрахунок і представлені результати математичного моделювання автономної системи теплопостачання з цифровим нечітким регулятором і об'єктом керування "водогрійний котел + колектор із трубопроводами + опалювальна будівля із приміщеннями".

The calculation is explained and the results of mathematical modelling of an autonomous system of a heat-supply with a digital fuzzy controller and plant of control "hot-water boiler + water drum with hot-water manifolds + heated building with rooms" represented.

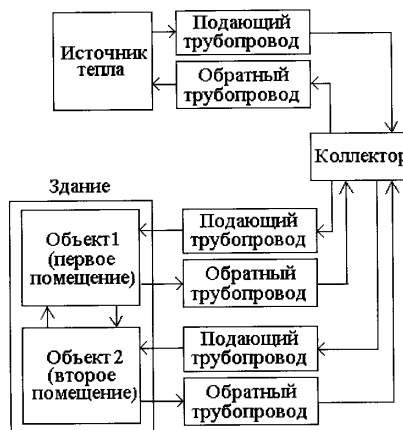


Рисунок 1

ВВЕДЕНИЕ

В работах [1,2] рассмотрены достаточно простые системы автоматического управления объектами типа "водогрейный котел + отопляемое здание (помещение)", работающие на базе нечеткой логики (такие системы носят название систем фуцци-управления [3,4]). Ниже рассмотрена более сложная автономная система фуцци-управления, объект управления в которой состоит из водогрейного котла (источник тепла), который через коллектор обогревает здание, имеющее два отопляемых помещения. Теплоноситель (вода) циркулирует по подающим и обратным трубопроводам. Для управления мощностью водогрейного котла с целью поддержания заданной температуры в помещениях при изменяющейся внешней температуре служит цифровой нечеткий регулятор.

ОПИСАНИЕ ОБЪЕКТА УПРАВЛЕНИЯ И СИНТЕЗ РЕГУЛЯТОРА

Объект управления представлен на рис.1.

Источник тепла - водогрейный котел описывают аperiodическим звеном с передаточной функцией $G_{BK}(s) = K_{BK}/(T_{BK}s + 1)$, в которой K_{BK} - коэффициент передачи (равен максимальной мощности котла P_{max} кВт), а T_{BK} - постоянная времени. Подача топлива G_T регулируется автоматическим клапаном АК. Выходная мощность котла mP_{max} , где m - степень открытия клапана (от 0 до 1). Отметим, что мощность котла в 1кВт соответствует теплу 0,2388 ккал/с или 1000Дж/с.

Подающие и обратные трубопроводы описывают звеньями запаздывания с передаточными функциями вида $k_i e^{-\tau_i s}$, где k_i - коэффициент теплопотерь, а τ_i - транспортная задержка тепла в трубопроводе.

Коллектор перераспределяет общий массовый расход теплоносителя (воды) в системе теплоснабжения в отношении $l_1/l_2 = p_1/p_2$ ($l_1 + l_2 = 1$, $p_1 + p_2 = p$), где $p_1 = l_1 p$ - массовый расход теплоносителя (воды) в первом помещении, $p_2 = l_2 p$ - массовый расход теплоносителя (воды) во втором помещении, p - общий массовый расход теплоносителя (воды) в системе теплоснабжения, кг/с.

Уравнение теплового баланса помещения определяют по формуле:

$$Q_i - Q_{oi} = k_i^* A_i (\theta_{Bi} - \theta_A) + M_i c \frac{d}{dt} (\theta_{Bi} - \theta_A),$$

где Q_i - тепло, поступающее в отопляемое помещение за одну секунду, Дж/с; Q_{oi} - тепло, отводимое из отопляемого помещения за одну секунду, Дж/с; $Q_i - Q_{oi}$ - тепло, передаваемое внутреннему воздуху за одну секунду, Дж/с; k_i^* - общий коэффициент теплопередачи ограждающих конструкций здания, нелинейно зависящий от соотношения температур, Дж/м²/с/°С; A_i - граничная поверхность, нормальная к потоку тепла (площадь наружной поверхности помещения), м²; M_i - масса внутреннего воздуха отопляемого помещения, кг; c - удельная массовая теплоемкость внутреннего воздуха отопляемого помещения, Дж/кг/°С; θ_A - температура окружающей среды, °С; θ_{Bi} - температура внутреннего воздуха отопляемого помещения, °С. Выражение

$k_i^* A_i (\theta_{Bi} - \theta_A)$ описывает тепловые потери помещения в окружающую среду. Выражение $M_i c \frac{d}{dt} (\theta_{Bi} - \theta_A)$ описывает тепло, аккумулирующееся во внутреннем воздухе помещения и обусловленное изменением его температуры. Уравнение теплового баланса справедливо для малых возмущений, когда можно считать, что зависящий от температуры коэффициент k_i^* является постоянной величиной. Кроме того, предполагается, что окружающая среда обладает бесконечно большой массой и что потери тепла помещением не повышают температуру окружающей среды.

За счет тепла, отводимого из отапливаемого помещения, такое помещение, рассматриваемое как объект регулирования, имеет контур внутренней обратной связи с коэффициентом передачи $K_{oci} = c_B p_i$, где c - удельная теплоемкость теплоносителя (воды), $c_B = 4187 \text{ Дж/кг/}^\circ\text{C}$; p_i - массовый расход теплоносителя (воды) в обратном трубопроводе, кг/с. Поэтому структурная схема помещения как объекта регулирования может быть представлена в виде рис.2,а. Преобразованные структурные схемы помещения приведены на рис.2,б,в, где параметры определяются следующим образом:

$$\alpha_i = \frac{1}{M_i c}, \quad b_i = \frac{k_i A_i}{M_i c} = \frac{1}{M_i c R_i}, \quad k_i A_i = 1/R_i.$$

Величина R_i , обратная произведению $k_i A_i$, называется термодинамическим сопротивлением. Начальные условия при интегрировании в схемах на рис.2,а,б различны: $\theta_i(0^+) = \theta_A$, $\theta_i(0^+) = 0$.

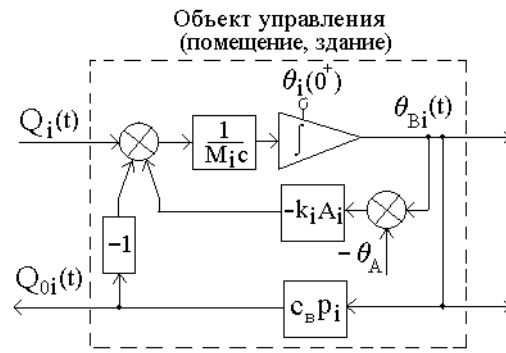
Уравнения тепловых балансов каждого из помещений, имеющих общие разделяющие их конструкции, можно записать в виде:

$$Q_1 - Q_{o1} = k_1^* A_1 (\theta_{B1} - \theta_A) + M_1 c \frac{d}{dt} (\theta_{B1} - \theta_A) - k_o^* A_o (\theta_{B1} - \theta_{B2}),$$

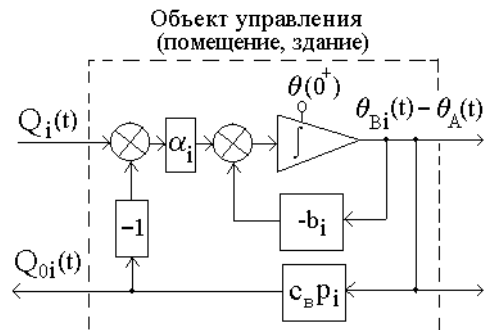
$$Q_2 - Q_{o2} = k_2^* A_2 (\theta_{B2} - \theta_A) + M_2 c \frac{d}{dt} (\theta_{B2} - \theta_A) - k_o^* A_o (\theta_{B2} - \theta_{B1}),$$

где k_o^* - коэффициент теплопередачи общих разделяющих конструкций двух помещений, нелинейно зависящий от соотношения температур в двух помещениях, Дж/м²/с/°С; A_o - граничная поверхность (площадь наружной поверхности, нормальная к потоку тепла) общих разделяющих конструкций двух помещений. Выражения $\pm k_o^* A_o (\theta_{B1} - \theta_{B2})$ описывают переток тепла

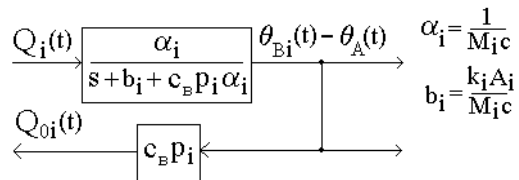
между помещениями.



а)



б)



в)

Рисунок 2

Теперь можно составить структурную схему системы автоматического управления, которая имеет общий объект управления - "водогрейный котел + коллектор с трубопроводами + здание с помещениями" и цифровой нечеткий регулятор. Эта схема приведена на рис.3. Помещения как объекты управления (объект 1 и объект 2) описываются аperiodическими звеньями. Входы объектов 1 и 2 - тепло, передаваемое внутреннему воздуху Q_1 и Q_2 , выходы объектов - разность температур внутреннего воздуха и окружающей среды $\theta_{B1} - \theta_A$ и $\theta_{B2} - \theta_A$ соответственно. Температура окружающей среды θ_A является внешним возмущающим воздействием. Возмущающее воздействие мож-

но представить в виде $\theta_A = \theta_0 + \theta_m \sin \omega t$, где θ_0 - средняя внешняя температура, θ_m - амплитуда суточных изменений внешней температуры, $\omega = 2\pi/(24 \times 3600) 1/c$, $\theta_{уст}$ - требуемая внутренняя температура (уставка). Отметим, что термодинамические свойства реальных зданий нелинейны и изменяются как со временем, так и с изменением погодных условий, поэтому модель (см. рис.3) только приближенно описывает реальный объект.

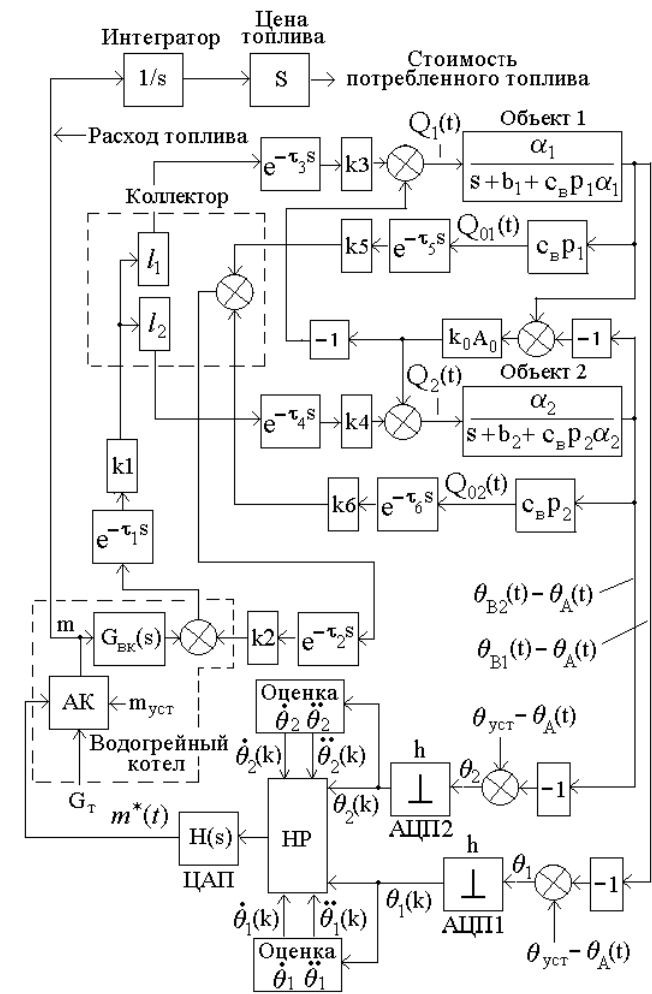


Рисунок 3

Рассмотрим систему водяного отопления одноэтажного здания площадью $300m^2$, работающую от водогрейного котла мощностью 30кВт с модулирующей горелкой. Параметры передаточной функции водогрейного котла: $T_{БК} = 300 c$; $K_{БК} = P_{max} = 30 \text{ кВт} = 30000 \text{ Дж/с}$.

Здание описывается необходимыми теплоизоляционными и геометрическими параметрами: длиной, шириной и высотой здания, размерами и количеством окон, формой крыши, теплопроводностью и толщиной материала стен, окон, крыши.

Расчет проведем для следующих конкретных данных. Здание имеет длину $l = 30 \text{ м}$, ширину $w = 10 \text{ м}$ и высоту

стен $h_{стен} = 4 \text{ м}$. Размеры окон: $h_{окна} = w_{окна} = 1 \text{ м}$. Количество окон $n = 6$. Крышный угол $\theta = 40^\circ$. Коэффициент теплопроводности и толщина материала: $\lambda_{стены} = 0,038 \text{ Дж/м/с/К}$ и $\delta_{стены} = 0,2 \text{ м}$, $\lambda_{окна} = 0,78 \text{ Дж/м/с/К}$ и $\delta_{окна} = 0,01 \text{ м}$, $\lambda_{крыши} = \lambda_{стены}$ и $\delta_{крыши} = \delta_{стены}$. Допустим, что потери тепла через пол пренебрежимо малы, стены и крыша сделаны из одинакового материала, а разделяющей конструкцией является стена, делящая внутренний объем здания в отношении 2/1 (для первого помещения длина $l_{n1} = \frac{2}{3}l_M$ и количество окон $n_1 = 4$, для второго - длина $l_{n2} = \frac{1}{2}l_M$ и количество окон $n_2 = 2$). Коэффициент теплопроводности и толщина материала общих разделяющих конструкций двух помещений $\lambda_o = \lambda_{стены} = 0,038 \text{ Дж/м/с/К}$, $\delta_o = \delta_{стены} = 0,2 \text{ м}$.

Формулы для расчета следующие:

$$M = wl(h_{стены} + \frac{w}{4} \text{tg}\alpha) \rho, M_1 = \frac{2}{3}M, M_2 = \frac{1}{2}M,$$

$$A_1 = (2l_{n1} + w)h_{стены} + \frac{wl_{n1}}{\cos\alpha} + \frac{w^2}{4} \text{tg}\alpha,$$

$$A_{окон1} = n_1 h_{окна} w_{окна},$$

$$A_2 = (2l_{n2} + w)h_{стены} + \frac{wl_{n2}}{\cos\alpha} + \frac{w^2}{4} \text{tg}\alpha,$$

$$A_{окон2} = n_2 h_{окна} w_{окна}, A_{стен1} = A_1 - A_{окон1},$$

$$A_{стен2} = A_2 - A_{окон2}, A_o = h_{стены} w + \frac{w^2}{4} \text{tg}\alpha,$$

$$R_{стены} = \frac{\delta_{стены}}{\lambda_{стены}}, R_{окна} = \frac{\delta_{окна}}{\lambda_{окна}},$$

$$R_1 = \frac{A_{стен1} + A_{окон1}}{\frac{A_{стен1}}{R_{стены}} + \frac{A_{окон1}}{R_{окна}}}, R_2 = \frac{A_{стен2} + A_{окон2}}{\frac{A_{стен2}}{R_{стены}} + \frac{A_{окон2}}{R_{окна}}},$$

$$k_1^* = \frac{1}{R_1}, k_2^* = \frac{1}{R_2}, k_o^* = \frac{1}{R_{стены}},$$

ρ - плотность воздуха на уровне моря ($1,225 \text{ кг/м}^3$).

Удельная массовая теплоемкость внутреннего воздуха отапливаемого помещения $c = 1005,4 \text{ Дж/кг/}^\circ\text{C}$.

Расчетные параметры объектов 1 и 2 следующие (см. рис.3):

$$\alpha_1 = \frac{1}{M_1 c} = 6,658 \cdot 10^{-7} \text{ }^\circ\text{C/Дж},$$

$$\alpha_2 = \frac{1}{M_2 c} = 13,316 \cdot 10^{-7} \text{ }^\circ\text{C/Дж},$$

$$b_1 = \frac{k_1^* A_1}{M_1 c} = 2,682 \cdot 10^{-4} 1/c,$$

$$b_2 = \frac{k_2^* A_2}{M_2 c} = 2,759 \cdot 10^{-4} \text{ 1/с.}$$

Коэффициенты распределения в коллекторе приняты равными: $l_1 = 0,68$, $l_2 = 0,32$. Массовый расход теплоносителя (воды) в системе теплоснабжения первого помещения $p_1 = l_1 p = 0,68 \cdot 0,28 = 0,1904$, кг/с; массовый расход теплоносителя (воды) в системе теплоснабжения второго помещения $p_2 = l_2 p = 0,32 \cdot 0,28 = 0,0896$, кг/с (общий массовый расход теплоносителя в системе теплоснабжения принят равным $p = 0,28$, кг/с). При этом коэффициенты обратной связи для объектов 1 и 2 определяются:

$$K_{oc1} = c_B p_1 = 797,205 \text{ Дж/с/}^\circ\text{C},$$

$$K_{oc2} = c_B p_2 = 375,155 \text{ Дж/с/}^\circ\text{C}.$$

Постоянные времени объектов 1 и 2 (с учетом обратной связи)

$$\frac{1}{b_1 + c_B p_1 \alpha_1} = 1251,56 \text{ с}, \quad \frac{1}{b_2 + c_B p_2 \alpha_2} = 1289,5 \text{ с},$$

коэффициенты передачи объектов 1 и 2 (с учетом обратной связи)

$$\frac{\alpha_1}{b_1 + c_B p_1 \alpha_1} = 8,333 \cdot 10^{-4} \text{ }^\circ\text{C/Дж},$$

$$\frac{\alpha_2}{b_2 + c_B p_2 \alpha_2} = 17,171 \cdot 10^{-4} \text{ }^\circ\text{C/Дж}.$$

Коэффициент взаимосвязи между объектами 1 и 2:

$$k_o^* A_o = 11,585 \text{ Дж/с/}^\circ\text{C}.$$

Коэффициенты теплопотерь и соответствующие транспортные задержки тепла в трубопроводах приняты следующими:

$$k_1=0,95; k_2=0,98; k_3=0,96; k_4=0,965; k_5=0,97;$$

$$k_6=0,973;$$

$$\tau_1=25\text{с}; \tau_2=24\text{с}; \tau_3=24\text{с}; \tau_4=23\text{с}; \tau_5=26\text{с}; \tau_6=25\text{с}.$$

Для простоты решения задачи синтеза нечеткого регулятора будем полагать, что число термов, с помощью которых оцениваются лингвистические переменные (входные и выходные параметры нечеткого регулятора) ошибка первого канала θ_1 и ошибка второго канала θ_2 , скорость изменения (первая производная) ошибки первого канала $\dot{\theta}_1$ и ошибки второго канала $\dot{\theta}_2$, ускорение (вторая производная) ошибки первого канала $\ddot{\theta}_1$ и ошибки второго канала $\ddot{\theta}_2$, управляющее воздействие на объект m , минимально, т.е. равно 2. Отообразим диапазоны

$[\theta_{1\min}, \theta_{1\max}]$, $[\dot{\theta}_{1\min}, \dot{\theta}_{1\max}]$, $[\ddot{\theta}_{1\min}, \ddot{\theta}_{1\max}]$, $[\theta_{2\min}, \theta_{2\max}]$, $[\dot{\theta}_{2\min}, \dot{\theta}_{2\max}]$, $[\ddot{\theta}_{2\min}, \ddot{\theta}_{2\max}]$ и $[m_{\min}, m_{\max}]$ изменения входных и выходного параметров на единое универсальное множество $U_1 = [0, L_i - 1] = [0, 1]$, где $L_i = 2$ - число, соответствующее количеству термов каждой лингвистической переменной x_i , $i = \overline{1, n}$, $n = 4$. При этом пересчет фиксированного значения параметра $x_i^* \in [x_{\text{ни}}, x_{\text{ви}}]$ в соответствующий элемент $u^* \in [0, 1]$ определяется пропорцией $(x_{\text{ви}} - x_{\text{ни}})/(1 - 0) = (x_i^* - x_{\text{ни}})/(u^* - 0)$, на основании которой находим [5]:

$$u_1^* = (\theta_1^* - \theta_{1\min}) / (\theta_{1\max} - \theta_{1\min}), \quad (1)$$

$$u_2^* = (\dot{\theta}_1^* - \dot{\theta}_{1\min}) / (\dot{\theta}_{1\max} - \dot{\theta}_{1\min}), \quad (2)$$

$$u_3^* = (\ddot{\theta}_1^* - \ddot{\theta}_{1\min}) / (\ddot{\theta}_{1\max} - \ddot{\theta}_{1\min}), \quad (3)$$

$$u_4^* = (\theta_2^* - \theta_{2\min}) / (\theta_{2\max} - \theta_{2\min}), \quad (4)$$

$$u_5^* = (\dot{\theta}_2^* - \dot{\theta}_{2\min}) / (\dot{\theta}_{2\max} - \dot{\theta}_{2\min}), \quad (5)$$

$$u_6^* = (\ddot{\theta}_2^* - \ddot{\theta}_{2\min}) / (\ddot{\theta}_{2\max} - \ddot{\theta}_{2\min}), \quad (6)$$

$$u_7^* = (m^* - m_{\min}) / (m_{\max} - m_{\min}). \quad (7)$$

На множестве $U = [0, 1]$ зададим два нечетких подмножества, функции принадлежности (ФП) которых треугольной формы (1 и 2) приведены на рис.4.

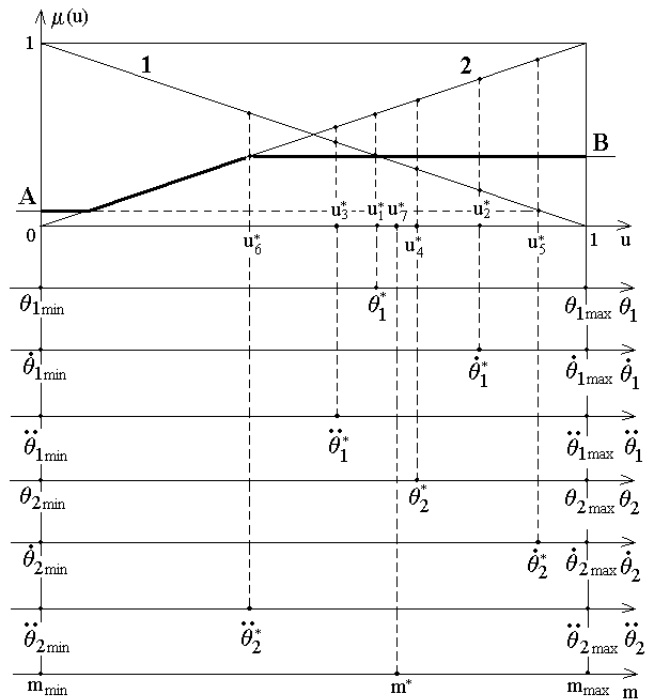


Рисунок 4

Аналитические выражения предложенных ФП для каждой лингвистической величины определяются простыми формулами:

$$\mu^1(u) = 1 - u, u \in [0, 1]; \mu^2(u) = u, u \in [0, 1]. \quad (8)$$

При поступлении на нечеткий регулятор (НР) значений входных переменных θ_1^* , $\dot{\theta}_1^*$, $\ddot{\theta}_1^*$ и θ_2^* , $\dot{\theta}_2^*$, $\ddot{\theta}_2^*$ с шагом квантования h осуществляется расчет величин u_1^* , u_2^* , u_3^* и u_4^* , u_5^* , u_6^* по формулам (1)-(6) и функций принадлежности $\mu^j(u)$, $j = \overline{1, 2}$.

Сформируем лингвистическое правило управления (рабочее правило) нечеткого регулятора в виде:

$$\text{Если } \theta_1^* = a_1^j \text{ и } \dot{\theta}_1^* = a_2^j \text{ и } \ddot{\theta}_1^* = a_3^j \text{ и } \theta_2^* = a_4^j \text{ и } \dot{\theta}_2^* = a_5^j \text{ и } \ddot{\theta}_2^* = a_6^j, \text{ то } m^* = a_7^j, j = \overline{1, 2}.$$

a_1^j , a_2^j и a_3^j - лингвистические оценки ошибки, скорости изменения (первой производной) ошибки и второй производной ошибки первого канала, a_4^j , a_5^j и a_6^j - лингвистические оценки ошибки, скорости изменения (первой производной) ошибки и второй производной ошибки второго канала, рассматриваемые как нечеткие подмножества, определенные на универсальном множестве и выбираемые из терм-множества лингвистических переменных x_i^* , $i = \overline{1, 6}$:

$$a_i^j \in \{\text{отрицательная}(1), \text{положительная}(2)\}.$$

a_7^j - лингвистические оценки управляющего воздействия на объект, выбираемые из терм-множества переменной m^* :

$$a_7^j \in \{\text{уменьшить}(1), \text{увеличить}(2)\}.$$

Пусть $\mu^j(x_i)$ - функция принадлежности параметра $x_i^* \in [x_{\text{ни}}, x_{\text{ви}}]$ нечеткому терму a_i^j , $i = \overline{1, 6}$, $j = \overline{1, 2}$. Тогда μ^{m_j} - зависящая от шести переменных ($x_1 \equiv \theta_1$, $x_2 \equiv \dot{\theta}_1$, $x_3 \equiv \ddot{\theta}_1$, $x_4 \equiv \theta_2$, $x_5 \equiv \dot{\theta}_2$, $x_6 \equiv \ddot{\theta}_2$) функция принадлежности вектора параметров решению (выбранному управляющему воздействию на объект) m_j , $j = \overline{1, 2}$, определяется из системы нечетких логических уравнений:

$$\mu^{m_j} = \mu^j(x_1) \wedge \mu^j(x_2) \wedge \mu^j(x_3) \wedge \mu^j(x_4) \wedge \mu^j(x_5) \wedge \mu^j(x_6). \quad (9)$$

Таким образом, μ^{m_1} - функция принадлежности управляющего воздействия нечеткому множеству "отрицательная", а μ^{m_2} - функция принадлежности управляющего

воздействия нечеткому множеству "положительная". Результирующая функция принадлежности для управляющего воздействия в соответствии с рабочим правилом НР записывается в виде

$$\mu^m = \mu^{m_1} \vee \mu^{m_2}. \quad (10)$$

В выражениях (9) и (10) \wedge - логическое и, \vee - логическое или.

В соответствии с лингвистическими правилами управления, формализованными системой нечетких логических уравнений (9) функция принадлежности управляющего воздействия $\mu^j(u_7)$ нечеткому множеству "уменьшить" ограничена сверху значением:

$$A = \min[\mu^1(u_1^*), \mu^1(u_2^*), \mu^1(u_3^*), \mu^1(u_4^*), \mu^1(u_5^*), \mu^1(u_6^*)], \quad (11)$$

а функция принадлежности управляющего воздействия $\mu^2(u_4)$ нечеткому множеству "увеличить" ограничена сверху значением:

$$B = \min[\mu^2(u_1^*), \mu^2(u_2^*), \mu^2(u_3^*), \mu^2(u_4^*), \mu^2(u_5^*), \mu^2(u_6^*)]. \quad (12)$$

Результирующая функция принадлежности для управляющего воздействия на основании выражения (10) определяется как

$$\mu(u_7) = \mu^1(u_7) \vee \mu^2(u_7), \quad (13)$$

т.е. получается формированием максимума (жирная линия на рис.4)

$$\mu(u_7) = \max[\mu^1(u_7), \mu^2(u_7)]. \quad (14)$$

Для определения конкретного значения управляющего воздействия m^* формируется "результирующая фигура", ограниченная результирующей ФП.

Производится поиск абсциссы "центра тяжести" результирующей фигуры по формуле [3]:

$$u_7^* = \frac{\sum_{k=1}^N (a_{k+1} - a_k)[(2a_{k+1} + a_k)b_{k+1} + (2a_k + a_{k+1})b_k]}{3 \sum_{k=1}^{N-1} (a_{k+1} - a_k)(b_{k+1} + b_k)}, \quad (15)$$

где N - число вершин, a_k , b_k - координаты вершин результирующей фигуры.

Полученное значение u_7^* на основании формулы (7) преобразуется в значение управляющего воздействия (степень открытия автоматического клапана) на общий объект управления

$$m^* = m_{\min} + (m_{\max} - m_{\min})u_7^*. \quad (16)$$

Требуемую уставку в автоматическом клапане определим как

$$m_{\text{уст}}(t) = \frac{\theta_{\text{уст}} - \theta_A(t)}{K_{\text{уст}}} \text{ (отн. ед.)}. \quad (17)$$

При работе системы автоматического управления (см. рис.3) степень открытия автоматического клапана определяется как

$$m(t) = m^*(t) + m_{\text{уст}}(t), \quad (18)$$

где $m^* \equiv m^*(t)$ - управляющее воздействие на общий объект управления, генерируемое нечетким регулятором.

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Выберем шаг квантования (интервал поступления данных в нечеткий регулятор) $h = 300 \text{ с} = 5 \text{ мин}$, а шаг моделирования $h_0 = 10 \text{ с}$.

Расчет и моделирование цифрового нечеткого регулятора производим по формулам (1)-(18). Моделирование общего объекта управления "водогрейный котел + коллектор с трубопроводами + здание с помещениями" выполняем по структурной схеме рис.3.

При оптимизации параметров нечеткого регулятора НР (диапазонов $[\theta_{1\min}, \theta_{1\max}]$, $[\dot{\theta}_{1\min}, \dot{\theta}_{1\max}]$, $[\ddot{\theta}_{1\min}, \ddot{\theta}_{1\max}]$, $[\theta_{2\min}, \theta_{2\max}]$, $[\dot{\theta}_{2\min}, \dot{\theta}_{2\max}]$, $[\ddot{\theta}_{2\min}, \ddot{\theta}_{2\max}]$) и коэффициента $K_{\text{уст}}$ используем квадратичный критерий качества

$$J = \frac{1}{L} \sum_{v=0}^{L-1} \theta_v^2 \Rightarrow \min, \quad (19)$$

где ошибка системы θ_v вычисляется с шагом моделирования h_0 , а число L определяет интервал наблюдения. Оптимальные параметры соответствуют минимальному значению критерия качества, а минимизация критерия качества автоматически приводит к оптимизации переходных процессов в системе управления.

При моделировании требуемая температура внутреннего воздуха здания $\theta_{\text{уст}}$ задавалась равной 20°C , амплитуда суточных изменений внешней температуры θ_m задавалась равной 5°C , а средняя внешняя температура θ_0 принималась соответственно равной $+5^\circ \text{C}$, 0°C , -5°C . Закон суточных изменений внешней температуры принимался синусоидальным (см. рис.5): $\theta_A(t) = \theta_0 + \theta_m \sin[2\pi/(24 \times 3600)]$.

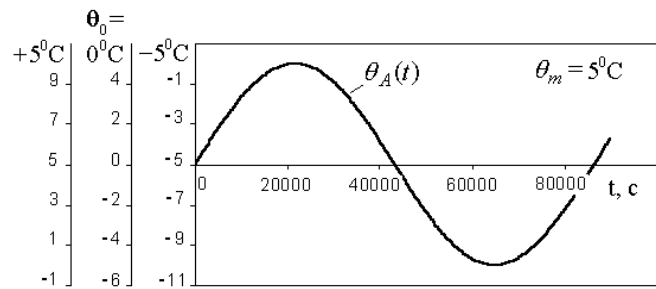


Рисунок 5

Начальные значения температуры в помещениях 1 и 2 принимались равными начальной внешней температуре:

$$\theta_{B1}(0^+) = \theta_{B2}(0^+) = \theta_A(0^+) = \theta_0.$$

При оптимизации получены следующие значения для диапазонов изменения входных и выходного параметров нечеткого регулятора:

$$[\theta_{1\min}, \theta_{1\max}] = [-0, 56, 0, 56],$$

$$[\theta_{2\min}, \theta_{2\max}] = [-0, 55, 0, 55],$$

$$[\dot{\theta}_{1\min}, \dot{\theta}_{1\max}] = [-1, 61 \cdot 10^{-3}, 1, 61 \cdot 10^{-3}],$$

$$[\dot{\theta}_{2\min}, \dot{\theta}_{2\max}] = [-1, 41 \cdot 10^{-3}, 1, 41 \cdot 10^{-3}],$$

$$[\ddot{\theta}_{1\min}, \ddot{\theta}_{1\max}] = [-3, 29 \cdot 10^{-6}, 3, 29 \cdot 10^{-6}],$$

$$[\ddot{\theta}_{2\min}, \ddot{\theta}_{2\max}] = [-5, 32 \cdot 10^{-6}, 5, 32 \cdot 10^{-6}].$$

Диапазон $[m_{\min}, m_{\max}] = [-0, 3, 0, 3]$ выбирается из конструктивных соображений.

Оптимальный коэффициент $K_{\text{уст}}$ в формуле (17) равен 36,08.

Результаты исследования системы автоматического управления (см. рис.3) путем математического моделирования представлены на рис.6-11, где изображены: а - изменение текущей ошибки системы за сутки (время наблюдения 90000с); б - изменение текущей ошибки системы за первых 3 часа (время наблюдения 11000с); в - изменение внутренней температуры помещений за первых 3 часа (время наблюдения 11000с).

На рис.6, 8 и 10 представлены текущие ошибки первого канала управления $\theta_1(t)$ и изменение внутренней температуры первого помещения (объекта 1) $\theta_{B1}(t)$ при средней внешней температуре θ_0 соответственно равной $+5^\circ \text{C}$, 0°C , -5°C . На рис.7, 9 и 11 представлены текущие ошибки второго канала управления $\theta_2(t)$ и изменение внутренней температуры второго помещения (объекта 2) $\theta_{B2}(t)$ при средней внешней температуре θ_0 соответственно равной $+5^\circ \text{C}$, 0°C , -5°C .

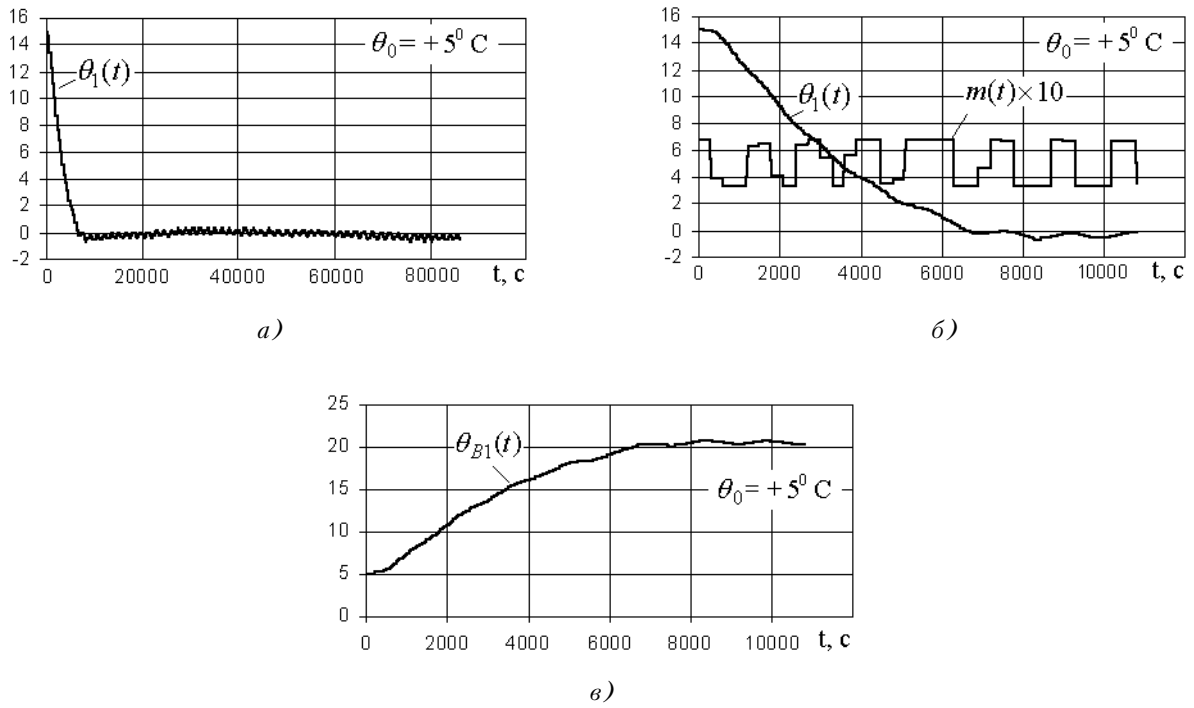


Рисунок 6

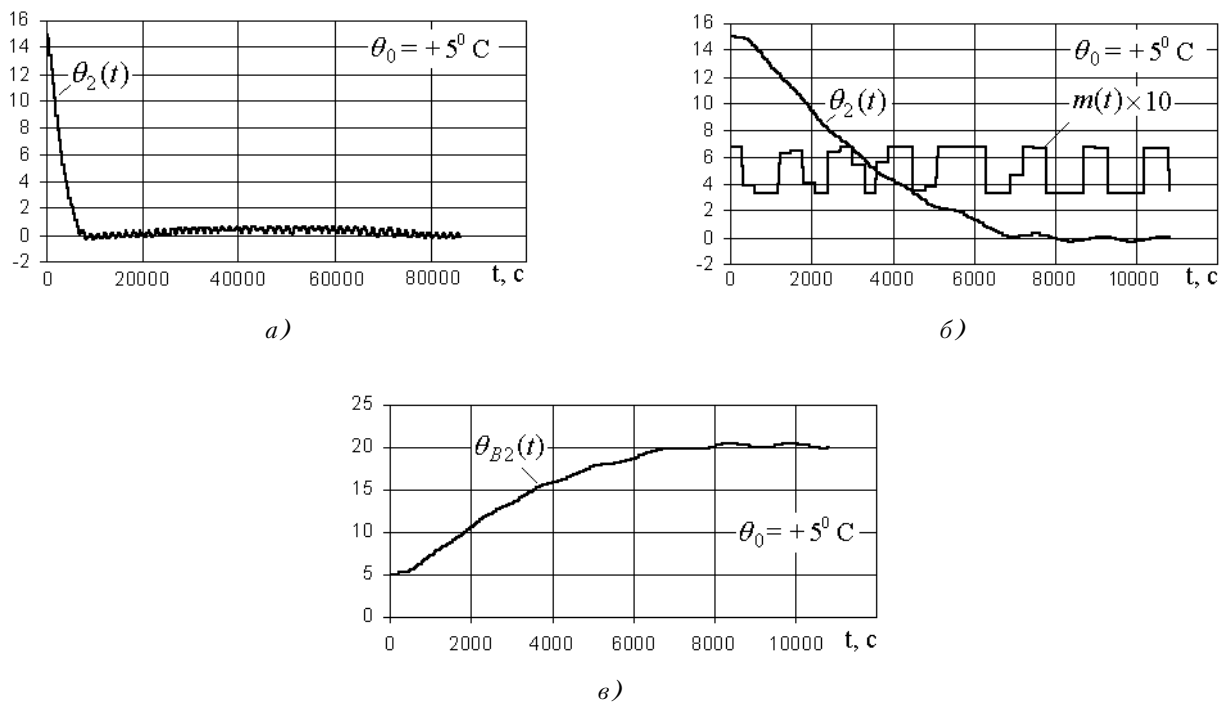
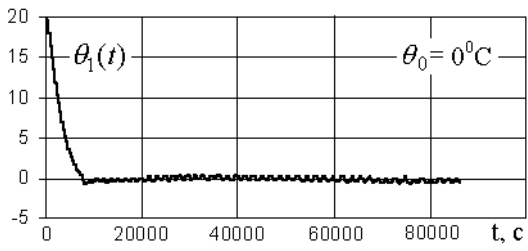
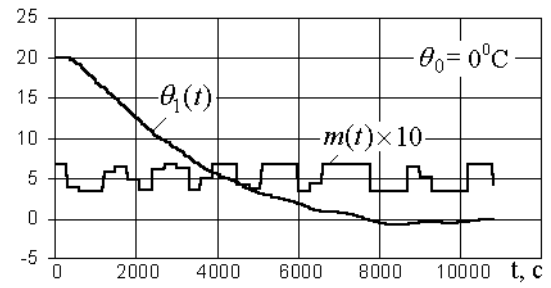


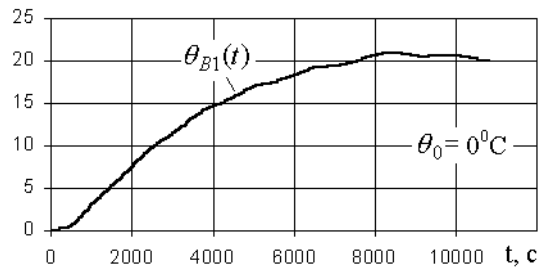
Рисунок 7



a)

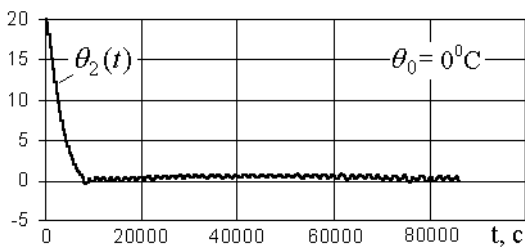


б)

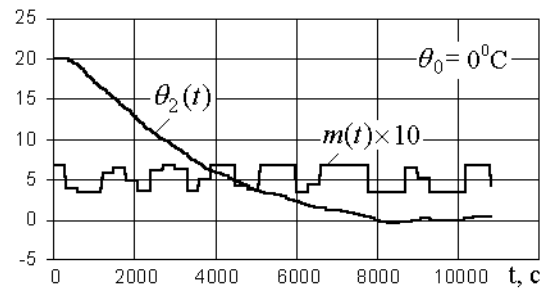


в)

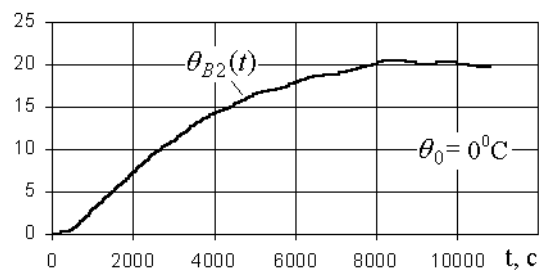
Рисунок 8



a)

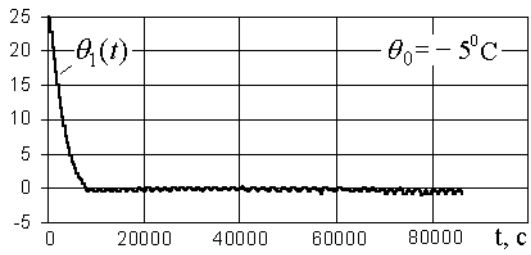


б)

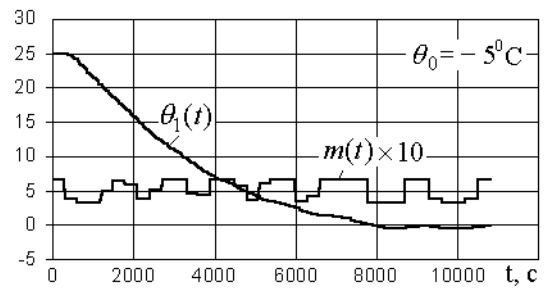


в)

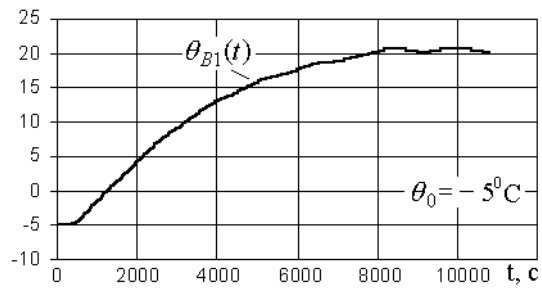
Рисунок 9



a)

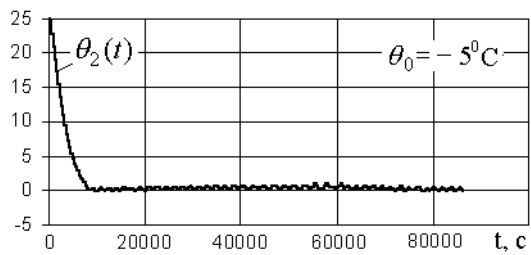


б)

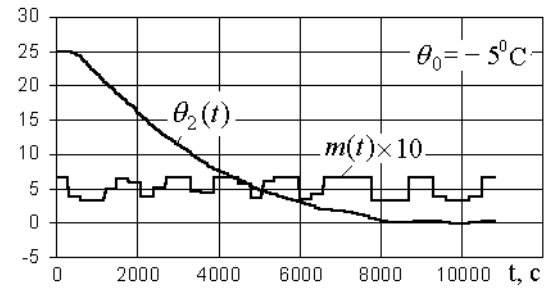


в)

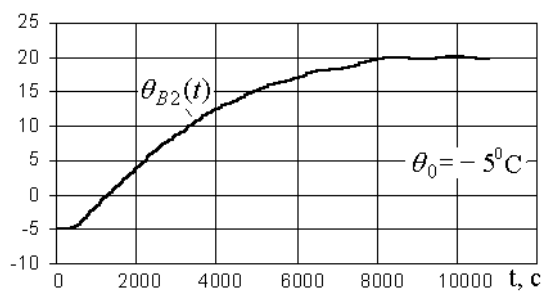
Рисунок 10



a)



б)



в)

Рисунок 11

Для удобства рассмотрения на всех рисунках нулевая ордината для управляющего воздействия $m(t)$ совмещена с ординатой, равной 5, для ошибки.

Текущие ошибки системы $\theta_1(t) = \theta_{уст} - \theta_{B1}(t)$ и $\theta_2(t) = \theta_{уст} - \theta_{B2}(t)$ показывают отличие температуры внутреннего воздуха помещений $\theta_{B1}(t)$ и $\theta_{B2}(t)$ от требуемой $\theta_{уст} = 20^\circ \text{C}$. Во всех случаях текущие ошибки (за исключением начального выброса) не превышают $0,3^\circ \text{C}$.

ВЫВОДЫ

Результаты моделирования (см. рис.6-11) показывают, что нечеткий регулятор обеспечивает качественную работу замкнутой системы автоматического управления при

достаточно сложной модели общего объекта управления (составленной с учетом внутренних контуров обратной связи в объектах 1 и 2, взаимосвязи между объектами, характеристик прямых и обратных трубопроводов).

ПЕРЕЧЕНЬ ССЫЛОК

1. Гостев В.И., Лесовой И.П., Чуприн А.Е. Применение оптимальных по быстродействию цифровых регуляторов для объектов управления с чистым запаздыванием // Радиоэлектроника Информатика Управление.-2000.-N2.-С.6-11.
2. Гостев В.И., Крайнев В.В., Чуприн А.Е. Управление водогрейными котлами на базе нечеткой логики // Автоматизация виробничих процесів.-2001.-N1 (12).-С.108-114.
3. Архангельский В.И., Богаенко И.Н., Грабовский Г.Г., Рюшин Н.А. Досвід розвитку і застосування систем функцій управління // Автоматизация виробничих процесів.-1997.-№2(5).-С.1-10.
4. Архангельский В.И., Богаенко И.Н., Грабовский Г.Г., Рюшин Н.А. Системы функци-управления.-К.: Техника, 1997.- 208 с.
5. Ротштейн А.П. Интеллектуальные технологии идентификации.-Вінниця: "УНІВЕРСУМ - Вінниця", 1999.-320 с.

УДК 621.372.8.01

МОДЕЛЬ ДЛЯ РАСЧЕТА ТЕНЗОРА ГРИНА МИКРОПОЛОСКОВОЙ СТРУКТУРЫ В ПРОСТРАНСТВЕННОЙ ОБЛАСТИ

Л.М.Карпуков

Предложен способ аппроксимации в спектральной области тензора Грина микрополосковой структуры. Составлена ее декомпозиционная модель для расчетов в пространственной области. Получены простые рекуррентные формулы вычисления зависимостей компонент тензора от пространственных координат.

Запропоновано спосіб апроксимації в спектральній області тензора Гріна мікросмужкової структури. Створено її декомпозиційну модель для розрахунків в просторовій області. Одержано прості рекурентні формули обчислювання залежностей компонент тензора від просторових координат.

The method of approximation in spectral domain of the Green's tensor of microstrip structure is proposed. The decomposition model of this structure for calculations in space domain is made. Simple recurrence formulas for calculation of dependencies of tensor components via spatial coordinates are received.

ВВЕДЕНИЕ

Качество проектных работ при конструировании микрополосковых антенн и устройств определяется в значительной степени используемыми методами моделирования и расчета. В электродинамической постановке эффективное решение краевых задач по моделированию микрополосковых структур может быть осуществлено на основе интегрального уравнения, составляемого с помощью тензора Грина для взаимного импеданса \mathbf{Z} , связывающего напряженность электрического поля \mathbf{E} с плотностью

сторонних токов \mathbf{J} . При реализации этого метода моделирования основной объем вычислительных затрат будет определяться выбранным способом нахождения тензора Грина и формой представления результатов расчета его компонент.

В работах по моделированию микрополосковой структуры широко используется спектральное представление компонент тензора Грина [1-3]. Недостатками этого подхода являются значительные вычислительные затраты и ошибки усечений при численном интегрировании несобственных интегралов с бесконечными пределами. Поэтому актуальной является задача непосредственного получения явных зависимостей компонент тензора от пространственных координат. Простое, но недостаточно точное, решение этой задачи получено в [4, 5] путем асимптотической аппроксимации спектрального представления компонент тензора.

Предлагаемый в настоящей работе метод составления явных зависимостей компонент тензора Грина микрополосковой структуры отличается от асимптотической аппроксимации более высокой точностью. Метод основан на разложении спектральных зависимостей в ряды по малому параметру. Для алгоритмизации вычислений вводится декомпозиционная модель микрополосковой структуры. Модель позволяет формировать в свободном пространстве методом многократных переотражений множество вспомогательных токов-изображений, непосредственно по которым составляются искомые формулы для расчета компонент тензора в пространственной области.

ПОСТАНОВКА ЗАДАЧИ

На рисунке 1, а изображено поперечное сечение микрополосковой структуры, состоящей из расположенного на металлическом экране слоя однородного диэлектрика с относительной диэлектрической проницаемостью ϵ_r и толщиной h . Вдоль осей x и y слой имеет бесконечные размеры. На поверхности слоя со стороны свободного пространства в точке с координатами $x_0, y_0, z = h$ размещен сторонний источник тока с плотностью \mathbf{J} .

Выразим тензор взаимного импеданса \mathbf{Z} через компоненты тензора векторного электрического потенциала \mathbf{A} и представим связь составляющих векторов \mathbf{E} и \mathbf{J} следующим образом:

$$\begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = -j\omega \begin{bmatrix} G_{xx} & 0 & 0 \\ 0 & G_{yy} & 0 \\ G_{zx} & G_{zy} & G_{zz} \end{bmatrix} + \frac{1}{k_0^2} \begin{bmatrix} \partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{bmatrix} \cdot \begin{bmatrix} \partial G_{xx}/\partial x + \partial G_{zx}/\partial z \\ \partial G_{yy}/\partial y + \partial G_{zy}/\partial z \\ \partial G_{zz}/\partial z \end{bmatrix}^T \begin{bmatrix} J_x \\ J_y \\ J_z \end{bmatrix}, \quad (1)$$

где $k_0 = \omega\sqrt{\epsilon_0\mu_0}$ - волновое число свободного пространства, T - знак транспонирования.

В спектральной области выражения для компонент $G_{\xi\eta}$ тензора \mathbf{A} могут быть найдены по декомпозиционной модели микрополосковой структуры, ориентированные графы которой изображены на рисунке 1,б [6].

Анализ графов приводит к следующим соотношениям:

$$\left. \begin{aligned} G_{\tau\tau}(k_x, k_y, k_z) &= \frac{\mu_0\delta(k_x, k_y)}{2k_{z0}}(1 + \Gamma_\tau)e^{-k_{z0}(z-h)}, \\ \partial G_{z\tau}(k_x, k_y, k_z)/\partial z &= \\ &= jk_\tau\alpha \frac{\mu_0\delta(k_x, k_y)}{2k_{z0}}(1 + \Gamma_\tau)(1 + \Gamma_z)e^{-k_{z0}(z-h)}, \\ G_{zz}(k_x, k_y, k_z) &= \frac{\mu_0\delta(k_x, k_y)}{2k_{z0}}(1 + \Gamma_z)e^{-k_{z0}(z-h)}, \end{aligned} \right\} \quad (2)$$

$$\text{где } \Gamma_\tau = \frac{\Gamma_k - e^{-2k_{z1}h}}{1 - \Gamma_k e^{-2k_{z1}h}}, \quad \Gamma_z = \frac{\Gamma_\epsilon + \Gamma_n}{1 + \Gamma_\epsilon \Gamma_n}, \quad \Gamma_n = \frac{\Gamma_k + e^{-2k_{z1}h}}{1 + \Gamma_k e^{-2k_{z1}h}},$$

$$\Gamma_\epsilon = \frac{\epsilon_r - 1}{\epsilon_r + 1}, \quad \Gamma_k = \frac{k_{z0} - k_{z1}}{k_{z0} + k_{z1}}, \quad \alpha = \frac{\Gamma_\epsilon}{(1 + \Gamma_\epsilon)},$$

$$\delta(k_x, k_y) = \frac{e^{jk_x x_0 + jk_y y_0}}{(2\pi)^2}, \quad k_{z0}^2 = \gamma^2 - k_0^2, \quad k_{z1}^2 = \gamma^2 - k_0^2 \epsilon_r,$$

$\gamma^2 = k_x^2 + k_y^2$, jk_τ - переменная преобразования Фурье по координате $\tau = x, y$.

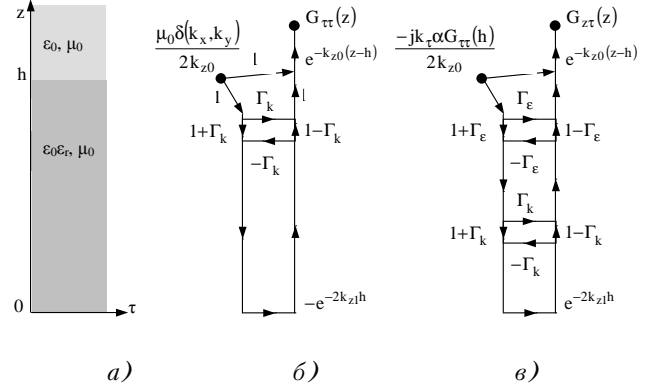


Рисунок 1 - Микрополосковая структура (а) и ориентированные графы ее декомпозиционной модели для горизонтальных (б) и вертикальных (в) компонент тензора

Непосредственное нахождение оригиналов по Фурье-изображениям функций, входящих в (2), возможно только с помощью численных методов интегрирования. Для практических применений наибольший интерес представляют формулы, определяющие в явном виде связь оригиналов с пространственными координатами. Эту связь можно получить в случае аппроксимации Фурье-изображений функций рядами вида:

$$G_{\xi\eta}(k_x, k_y, k_z) \approx \sum_i b_i \frac{e^{-k_{z0}d_i}}{k_{z0}} \delta(k_x, k_y), \quad (3)$$

где b_i, d_i - в общем случае комплексные коэффициенты. Последующее применение соотношения [7]

$$\begin{aligned} \frac{1}{8\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{e^{-k_{z0}d_i - jk_x(x-x_0) - jk_y(y-y_0)}}{k_{z0}} dk_x dk_y &= \\ = \frac{1}{4\pi} \frac{e^{-jk_0\sqrt{(x-x_0)^2 + (y-y_0)^2 + d_i^2}}}{\sqrt{(x-x_0)^2 + (y-y_0)^2 + d_i^2}} \end{aligned} \quad (4)$$

к членам ряда обеспечивает получение оригинала в искомой аналитической форме.

Целью работы является разработка эффективного метода вычисления оригиналов для функций в (2) с использованием приближения Фурье-изображений рядами (3), обладающего более высокой точностью по сравнению с методом асимптотической аппроксимации, который основан на допущении $k_{z1} = k_{z0}$, ограничивающим его применение расчетом полей в ближней зоне на расстояниях r , удовлетворяющих условию $k_0 r \ll 1$ [4, 5].

ТЕОРЕТИЧЕСКИЕ РЕЗУЛЬТАТЫ

Для решения поставленной задачи необходимо аппроксимировать соотношения для коэффициентов отражения Γ_τ и Γ_n из (2) функциями, зависящими от k_{z0} .

Сложность аппроксимации спектрального представления (2) компонент тензора рядами (3) определяется наличием полюсов у функций Γ_τ и Γ_n . Для упрощения вычислений введем условие

$$k_0 h \sqrt{\epsilon_r - 1} < \pi/2, \quad (5)$$

обеспечивающее отсутствие полюсов у Γ_τ и наличие одного полюса у Γ_n [7].

Функция Γ_τ может быть записана в виде

$$\Gamma_\tau = -\frac{\operatorname{ch}(k_{z1}h) - k_{z0}h \frac{\operatorname{sh}(k_{z1}h)}{k_{z1}h}}{\operatorname{ch}(k_{z1}h) + k_{z0}h \frac{\operatorname{sh}(k_{z1}h)}{k_{z1}h}}, \quad (6)$$

позволяющем разложить ее числитель и знаменатель в ряды по степеням $k_{z0}h$ с последующей их аппроксимацией экспоненциальными рядами вида (3). Гладкость функции Γ_τ , имеющая место при выполнении условия (5), позволяет ограничиться при аппроксимации первыми двумя членами экспоненциального ряда и представить ее следующим образом:

$$\Gamma_\tau = \frac{\Gamma_\delta - e^{-k_{z0}h}}{1 - \Gamma_\delta e^{-k_{z0}h}} \cdot e^{-k_{z0}h}, \quad (7)$$

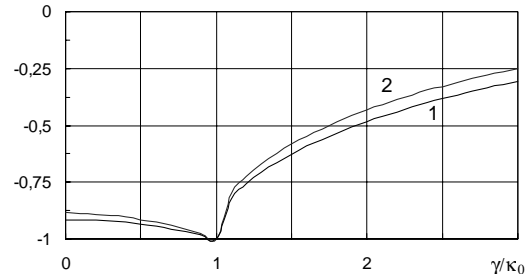
где $\Gamma_\delta = 1 - \delta h \operatorname{ctg}(\delta h)$, $\delta = k_0 \sqrt{\epsilon_r - 1}$.

Для примера на рисунке 2 приведены спектральные зависимости функции $1 + \Gamma_\tau$, рассчитанные по точной формуле (6) и ее аппроксимации (7) на частоте $f = 10$ ГГц для микрополосковой структуры с толщиной подложки $h = 1$ мм и проницаемостью $\epsilon_r = 9,8$. Данные примера соответствуют значению $\delta h = 0,369\pi/2$. Расчеты по (6) и (7), представленные кривой 1 на рисунках, совпали с графической точностью. Кривая 2 на рисунках соответствует асимптотическому приближению к функции $1 + \Gamma_\tau$, вычисляемому по формуле $1 + \Gamma_\tau = 1 - \exp(-2k_{z0}h)$ [5].

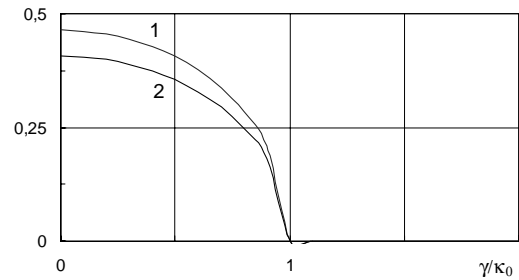
Для перехода из спектральной в пространственную область разложим (7) в геометрическую прогрессию и, применив к членам ряда формулу (4), составим искомое выражение для оригинала $G_{\tau\tau}$ из (2):

$$G_{\tau\tau}(r, h) = \frac{\mu_0}{4\pi} \left[\frac{e^{-jk_0 r}}{r} + \sum_{n=0}^{\infty} \Gamma_\delta^n + 1 \frac{e^{-jk_0 \sqrt{r^2 + (nh+h)^2}}}{\sqrt{r^2 + (nh+h)^2}} - \Gamma_\delta^n \frac{e^{-jk_0 \sqrt{r^2 + (nh+2h)^2}}}{\sqrt{r^2 + (nh+2h)^2}} \right], \quad (8)$$

где $r = \sqrt{(x-x_0)^2 + (y-y_0)^2}$.



а)



б)

Рисунок 2 - Зависимости реальной (а) и мнимой (б) частей функции $1 + \Gamma_\tau$ от отношения γ/k_0 для микрополосковой структуры с параметрами $h = 1$ мм, $\epsilon_r = 9,8$ на частоте $f = 10$ ГГц: кривая 1 - расчет по формулам (6) и (7), кривая 2 - асимптотическое приближение

Процедуру суммирования в этом выражении удобно заменить расчетом по рекуррентной формуле [8, 9]:

$$\left. \begin{aligned} G_{\tau\tau}(r, h) &= \frac{\mu_0}{4\pi} (u_0^+ + u_0^-), \\ u_n^- &= \Gamma_\delta u_{n+1}^+ - u_{n+2}^+ + \Gamma_\delta u_{n+1}^- \end{aligned} \right\} \quad (9)$$

Здесь $u_n^+ = \exp(-jk_0 \sqrt{r^2 + (nh)^2}) / \sqrt{r^2 + (nh)^2}$, начальное условие $u_{N+1}^- = 0$, $N \rightarrow \infty$.

В таблице 1 приведены для сравнения точности аппроксимации значения функции $G_{\tau\tau}(r, h)4\pi/\mu_0$ в зависимости от нормированного расстояния r/λ_0 при $f = 10$ ГГц, $\epsilon_r = 9,8$ и h , равным 0,5 и 1 мм, λ_0 - длина волны в свободном пространстве. В строке 1 размещены результаты численного обратного преобразования Фурье точной формулы (6), в строке 2 - расчеты по формуле (9) и в

строке 3 - результаты асимптотического приближения [5].

Таблица 1

	r/λ_0	1/32	1/16	1/8	1/4	1/2	1	2
h = 1 мм	1	704,8-j7,991	217,9-j7,898	50,54-j7,536	9,146-j6,202	-0,744-j2,421	0,142+j0,615	0,028+j0,154
	2	704,9-j7,992	218,3-j7,901	50,62-j7,537	9,153-j6,204	-0,822-j2,242	0,104+j0,615	0,013+j0,154
	3	641,1-j6,048	186,5-j5,978	40,49-j5,704	7,068-j4,696	-0,609-j1,837	0,077+j0,464	0,009+j0,116
h = 0,5мм	1	359,3-j1,625	71,59-j1,606	12,16-j1,533	1,959-j1,263	-0,081-j0,495	0,058+j0,124	0,017+j0,031
	2	359,5-j1,625	71,63-j1,606	12,16-j1,533	1,960-j1,263	-0,160-j0,495	0,020+j0,124	0,003+j0,031
	3	346,5-j1,522	67,99-j1,504	11,43-j1,436	1,837-j1,182	0,149-j0,464	0,019+j0,116	0,002+j0,029

Основываясь на аппроксимации (7), запишем Γ_n в следующем виде:

$$\Gamma_n = \frac{1 - (Y_m + Y_p)}{1 + (Y_m + Y_p)}, \quad (10)$$

где $Y_m = \frac{1 - \Gamma_m}{1 + \Gamma_m}$, $Y_p = \left(\frac{j\delta}{k_{z0}}\right)^2 \text{sh} \frac{(k_{z0}h)}{[\delta h \text{ctg}(\delta h) + \text{ch}(k_{z0}h)]}$,

$$\Gamma_m = -\Gamma_\tau.$$

Выражение (10) можно интерпретировать как коэффициент отражения от параллельного разветвления линий, к которым подключены нагрузки с коэффициентами отражений Γ_τ и $\Gamma_p = (1 - Y_p)/(1 + Y_p)$. Функция Γ_p , определяя положение полюса в интервале $k_0 < k_{z0} < k_0 \epsilon_r$, формирует спектральную зависимость Γ_n при малых k_{z0} . При больших k_{z0} наибольшее влияние на форму спектральной зависимости Γ_n оказывает функция Γ_m , которая в этом случае может быть аппроксимирована своим асимптотическим представлением: $\Gamma_m = \exp(-2k_{z0}h)$.

В области малых k_{z0} функция Γ_p с учетом условия (5) может быть аппроксимирована выражением

$$\Gamma_p = \frac{k_{z0}/(j\delta h) - j\delta/[1 + \delta h \text{ctg}(\delta h)]}{k_{z0}/(j\delta h) + j\delta/[1 + \delta h \text{ctg}(\delta h)]}.$$

Умножим числитель и знаменатель этого выражения на малое Δ и введем коэффициенты отражений

$$\Gamma_\Delta = \frac{1 - j\delta\Delta}{1 + j\delta\Delta},$$

$$\Gamma_e = \frac{1 - k_{z0}\Delta/(j\delta h_e)}{1 + k_{z0}\Delta/(j\delta h_e)},$$

где $h_e = 1 + \delta h \text{ctg}(\delta h)$.

Аппроксимируем Γ_e экспоненциальной функцией и представим Γ_p в виде:

$$\Gamma_p = \frac{\Gamma_\Delta - e^{-2k_{z0}\Delta/(j\delta h_e)}}{1 - \Gamma_\Delta e^{-2k_{z0}\Delta/(j\delta h_e)}}. \quad (11)$$

Проведенные преобразования выражений для Γ_τ и Γ_n тождественны преобразованию ориентированных графов на рисунках 2,б и 2,в к виду, указанному на рисунках 3,а и 3,б. Шестиполюсник на рисунке 3,б соответствует разветвлению трех одинаковых линий и имеет коэффициенты отражения $\Gamma = -1/3$ и передачи $T = 2/3$.

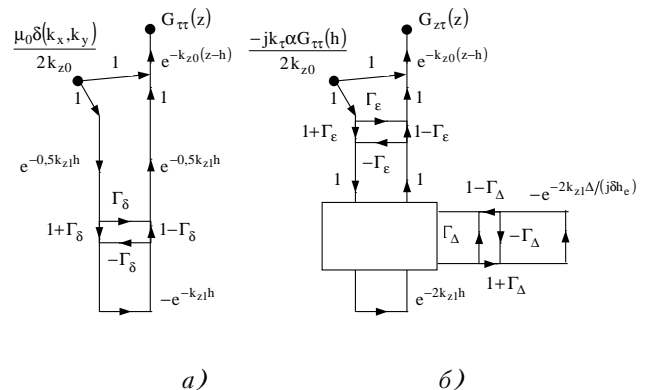


Рисунок 3 - Ориентированные графы модифицированной декомпозиционной модели микрополосковой структуры для горизонтальных (а) и вертикальных (б) компонент тензора

Составим по графу на рис. 3,б систему уравнений модели микрополосковой структуры для вертикальных составляющих:

$$\left. \begin{aligned} \Gamma_z &= S_{11} + S_{12}u_2^- e^{-2k_{z0}h} - S_{13}u_3^- e^{-2k_{z0}\Delta/(j\delta h_e)}, \\ \begin{bmatrix} u_2^- \\ u_3^- \end{bmatrix} &= \begin{bmatrix} S_{22} - S_{23} \\ S_{32} - S_{33} \end{bmatrix} \cdot \begin{bmatrix} u_2^- e^{-2k_{z0}h} \\ u_3^- e^{-2k_{z0}\Delta/(j\delta h_e)} \end{bmatrix} + \begin{bmatrix} S_{21} \\ S_{31} \end{bmatrix}. \end{aligned} \right\} \quad (12)$$

Здесь S_{ik} - элементы матрицы рассеяния разветвления линий с учетом переходов с коэффициентами отражений Γ_ϵ и Γ_Δ :

$$S = \frac{1}{2 + j\delta\Delta(1 - \Gamma_\epsilon)} \begin{bmatrix} 2\Gamma_\epsilon - j\delta\Delta(1 - \Gamma_\epsilon) & 2(1 - \Gamma_\epsilon) & 2j\delta\Delta(1 - \Gamma_\epsilon) \\ 2(1 + \Gamma_\epsilon) & -2\Gamma_\epsilon - j\delta\Delta(1 - \Gamma_\epsilon) & 2j\delta\Delta(1 - \Gamma_\epsilon) \\ 2(1 + \Gamma_\epsilon) & 2(1 - \Gamma_\epsilon) & -2\Gamma_\epsilon + j\delta\Delta(1 - \Gamma_\epsilon) \end{bmatrix} \quad (13)$$

На рисунке 4 приведены спектральные зависимости функции $1 + \Gamma_z$, рассчитанные для данных предыдущего примера по точной формуле и по соотношениям (12). Эти расчеты совпали с графической точностью и представлены кривой 1. Кривая 2 соответствует асимптотическому приближению, вычисляемому по формуле [5]

$$1 + \Gamma_z = (1 + \Gamma_\epsilon) \frac{(1 + e^{-2k_z h})}{(1 + \Gamma_\epsilon e^{-2k_z h})}$$

По уравнениям (12) в соответствии с методикой, изложенной в [8, 9], могут быть составлены рекуррентные формулы для расчета оригиналов Фурье-изображений вертикальных компонент тензора. В частности, для вычисления оригинала функции G_{zz} будет иметь место следующее выражение:

$$G_{zz}(r, h) = \frac{\mu_0}{4\pi} (u_{1,0,0}^+ + S_{11}u_{1,0,0}^+ + S_{12}u_{2,0,0}^- - S_{13}u_{2,0,1}^-), \left. \begin{aligned} \begin{bmatrix} u_{2,n,m}^- \\ u_{3,n,m}^- \end{bmatrix} &= \begin{bmatrix} S_{22} - S_{23} \\ S_{32} - S_{33} \end{bmatrix} \cdot \begin{bmatrix} u_{2,n+1,m}^- \\ u_{3,n,m+1}^- \end{bmatrix} + \begin{bmatrix} S_{21}u_{1,n,m}^+ \\ S_{31}u_{1,n,m}^+ \end{bmatrix} \end{aligned} \right\} \quad (14)$$

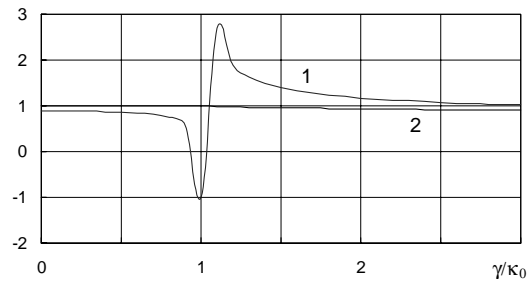
Здесь $u_{1,n,m}^+ = \frac{\exp(-jk_0 \sqrt{r^2 + (n2h + m2\Delta / (j\delta h_\epsilon))^2})}{\sqrt{r^2 + (n2h + m2\Delta / (j\delta h_\epsilon))^2}}$,

начальное условие $u_{2,N+1,M}^- = 0$, $u_{3,N,M+1}^- = 0$, $N \rightarrow \infty$, $M \rightarrow \infty$.

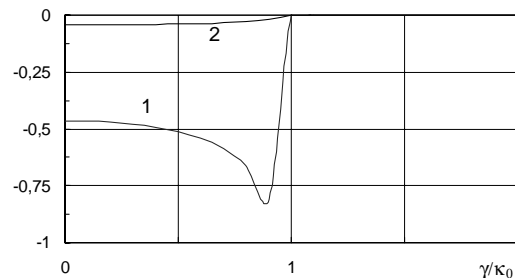
В таблице 2 для тех же данных, которые использовались при составлении таблицы 1, приведены результаты расчета функции $G_{zz}(r, h)4\pi/\mu_0$ путем численного обратного преобразования Фурье точной формулы (строка 1 таблицы), по соотношениям (14) (строка 2) и в асимптотическом приближении [5] (строка 3). Для сходимости вычислений в волновое число $k_0\sqrt{\epsilon_r}$ вносилась отрицательная мнимая часть, составляющая 0,1% его величины.

Таблица 2

	r/λ_0	1/32	1/16	1/8	1/4	1/2	1	2
h = 1мм	1	2090-j570,0	1024-j555,9	364,8-j503,2	-79,55-j323,6	-170,8+j64,94	89,09-j53,73	42,73-j45,31
	2	2106-j568,8	1037-j545,6	362,6-j502,5	-78,66-j324,7	-171,3+j65,31	89,34-j54,01	42,57-j45,47
	3	2002-j416,1	968,7-j408,1	375,9-j377,1	-0,069-j266,6	-133,3+j0,024	66,67-j0,005	33,33-j0,002
h = 0,5мм	1	2119-j483,5	1017+j472,9	378,2-j433,2	-30,91-j295,1	-152,9+j24,03	80,07-j18,67	42,27-j14,69
	2	1025-j488,1	1019-j478,1	380,1-j438,5	-30,82-j296,3	-153,0+j23,92	80,06-j18,65	42,29-j14,63
	3	983,3-j408,2	983,26-j408,1	376,9-j377,2	-0,024-j206,6	-133,3+j0,005	66,67-j0,002	33,33-j0,001



а)



б)

Рисунок 4 - Зависимости реальной (а) и мнимой (б) частей функции $1 + \Gamma_z$ от отношения γ/κ_0 для микрополосковой структуры с параметрами $h = 1$ мм, $\epsilon_r = 9,8$ на частоте $f = 10$ ГГц: кривая 1 - расчет по точной формуле и по соотношениям (12), кривая 2 - асимптотическое приближение

Представленные в таблицах 1 и 2 результаты расчета указывают на высокую точность моделирования компонент тензора Грина микрополосковой структуры предлагаемым методом. В отличие от асимптотического приближения предлагаемый метод расчета воспроизводит с незначительными погрешностями моделируемое поле как в ближней зоне, так и на расстояниях в несколько длин волн от источника. Погрешность моделирования уменьшается с уменьшением толщины h диэлектрического слоя.

ВИВОДИ

Разработанный метод расчета тензора Грина микрополосковой структуры отличается простотой и высокой точностью, обеспечивая моделирование полей на расстояниях порядка нескольких длин волн от источника. Элементы декомпозиционной схемы микрополосковой структуры, адаптированные по предложенной методике к расчету в пространственной области, могут быть непосредственно использованы в качестве базовых при построении декомпозиционных схем многослойных плоско-слоистых структур. Ограничение (5) на толщину диэлектрической пластины микрополосковой структуры не является принципиальным, поскольку пластина может быть представлена декомпозиционной схемой в виде соединения нескольких более тонких слоев диэлектрика.

ПЕРЕЧЕНЬ ССЫЛОК

1. Iton T., Mitra R. Spectral domain approach for calculating the dispersion characteristics of microstrip lines // IEEE Trans. MTT. - 1973. - V. 21. - № 7. - P. 496 - 499.

2. Katehi P., Alexopoulos N. Frequency-dependent characteristics of microstrip discontinuities in millimeter-wave integrated circuits // IEEE Trans. MTT. - 1985. - V. 33. - № 10. - P. 1029 - 1035.
3. Kobayashi M., Ando F. Dispersion characteristics of open microstrip lines // IEEE Trans. MTT. - 1987. - V. 35. - № 2. - P. 784 - 788.
4. Mosig J.R., Sarkar T.K. Comparison of quasi-static and exact electromagnetic fields from a horizontal electric dipole above a lossy dielectric backed by an imperfect ground plane // IEEE Trans. MTT. - 1986. - V. 34. - № 4. - P. 379 - 387.
5. 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. MTT. - 1991. - V. 39. - № 1. - P. 75 - 82.
6. Карпуков Л.М. Алгоритм расчета тензоров Грина для полосково-щелевых структур в слоистой среде // Радиоелектроника. Информатика. Управление. - 1999. - №1. - С. 11 - 15.
7. Марков Г.Т., Чаплин А.Ф. Возбуждение электромагнитных волн. - М.: Энергия, 1967. - 376 с.
8. Карпуков Л.М. Построение и анализ декомпозиционных моделей микрополосковых структур // Радиоелектроника. - 1984. - Т. 27. - № 9. - С. 32 - 36 (Изв. высш. учеб. заведений).
9. Карпуков Л.М. Алгоритм квазистатического моделирования полосковых структур в многослойной анизотропной среде // Радиоелектроника. Информатика. Управление. - 2000. - № 1. - С. 18 - 23.

УДК 681.518:004.93'1

ВИЗНАЧЕННЯ НОРМОВАНОГО ПОЛЯ ДОПУСКІВ НА ПАРАМЕТР НАСТРОЙКИ ПРИ АВТОФОКУСУВАННІ ЕЛЕКТРОННОГО МІКРОСКОПА ЗА ЗОБРАЖЕННЯМ

А.С.Краснопоясовський, А.М.Скаковська

На примере задачи автофокусировки электронного микроскопа, рассматривается определение оптимальной в информационном смысле области значений параметров настройки в рамках метода функционально-статистических испытаний.

На прикладі задачі автофокусування електронного мікроскопа, розглядається визначення оптимальної в інформаційному сенсі області значень параметрів настройки в рамках методу функціонально-статистичних випробувань.

The definition of the optimum field of meaning of the adjustment parameters in the informational sense within the method of the functional-statistical tests is considered on the example of the electronic microscope autofocusing task.

В роботі [1] наведено алгоритм автофокусування растрового електронного мікроскопа (РЕМ) за зображенням в рамках методу функціонально-статистичних випробувань (МФСВ), який ґрунтується на оцінці інформаційної здатності системи класифікаційної настройки (КН). Недоліком цього алгоритму є можливість появи коливального процесу при зміні значення параметра настройки в області його екстремального значення, що призводить до зменшення оперативності процесу автофокусування мікроскопа. З метою усунення цього недоліку розглянемо в рамках МФСВ визначення нормованого поля допусків на параметр настройки, яке забезпечує прийнятну фокусність зображення.

МЕТОДОЛОГІЧНІ ТА ТЕОРЕТИЧНІ ПОЛОЖЕННЯ ПІДХОДУ

Нехай функціональний стан РЕМ характеризується триальтернативною системою оцінок параметра настройки $I_{\text{Л}}$ - струму об'єктивної лінзи: "МЕНШЕ НОРМИ" - клас $X_{\text{М}}^0$, "НОРМА" - клас $X_{\text{Н}}^0$ і "БІЛЬШЕ НОРМИ" - клас $X_{\text{Б}}^0$. На рис.1 наведено розташування цих класів в області значень параметра $I_{\text{Л}}$. Тут A_0 , $A_{\text{Н}}$, $A_{\text{Б}}$ - номінальне значення параметра, нижній і верхній нормовані (експлуатаційні) допуски на параметр настройки відповідно.

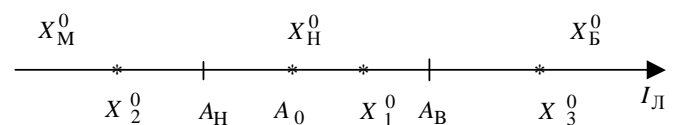


Рисунок 1

Нормовані допуски $A_{\text{Н}}$ і $A_{\text{Б}}$, задають область прийнятної з практичних міркувань фокусності зображення -

область X_H^0 . Допуски A_H і A_B є індивідуальними характеристиками кожного мікроскопа, які залежать від його технічного стану і умов функціонування, тому після кожного входу мікроскопа в робочий режим необхідно здійснювати їх визначення з метою формування вхідних даних для алгоритму автофокусування. Для визначення допусків A_B і A_H застосуємо алгоритм дискримінантного класифікаційного аналізу за МФСВ [2], який дозволяє побудувати оптимальні в інформаційному сенсі роздільні гіперповерхні (РГП) між двома сусідніми класами. За основний клас, який забезпечується поточним значенням струму I_L і відповідає основній гіпотезі γ_1 : функціональний стан мікроскопа "НОРМА", приймемо клас X_1^0 (рис.1). За "сусіда" зліва приймемо клас X_2^0 , який належить області X_M^0 , тобто відповідає гіпотезі γ_2 - "МЕНШЕ НОРМИ", і за "сусіда" справа - клас X_3^0 , який відповідає гіпотезі γ_3 - "БІЛЬШЕ НОРМИ". Оскільки на початку функціонування мікроскопа значення A_0 і допуски A_H і A_B невідомі, то послідовне формування класів X_2^0 і X_3^0 доцільно здійснювати шляхом допускової оцінки зміни однієї з об'єктивних інтегральних характеристик зображення, наприклад, максимальної контрастності зображення, яку визначимо за формулою:

$$\Delta\eta = \max_{\{i,j\}} \eta - \min_{\{i,j\}} \eta, \quad (1)$$

де $\max_{\{i,j\}} \eta$, $\min_{\{i,j\}} \eta$ - максимальна та мінімальна яскравості рецепторного поля відповідно.

За зображення класу X_2^0 вибирається зображення, для якого середня контрастність $\Delta\eta \leq \Delta\eta_{\text{доп}}$ при значеннях $I_L < A_0$ і за зображення класу X_3^0 - зображення, для якого так само $\Delta\eta \leq \Delta\eta_{\text{доп}}$ при $I_L > A_0$. Тут $\Delta\eta_{\text{доп}}$ - гранична середня контрастність зображення, яке відповідає класу X_1^0 .

АЛГОРИТМ ПОБУДОВИ РГП

На підготовчому етапі формуються зображення, які відповідають класам X_1^0 , X_2^0 і X_3^0 . За клас X_1^0 приймається сфокусоване зображення, для якого обчислюється максимальна контрастність $\Delta\eta$ за формулою (1). Для формування класу X_2^0 зменшується струм I_L з кроком ΔI_L , обчислюється поточна максимальна контрастність і перевіряється умова $\eta_{\text{доп}} \leq 0,95\eta_1$. Поточний клас зображення, для якого виконується це відношення, визначається як X_2^0 . Аналогічно формується зображення класу X_3^0 за умови збільшення струму I_L . Визначення нормованого поля допусків здійснюється шляхом послі-

довної побудови РГП для класу X_1^0 між класом X_2^0 (допуск A_H), і між класом X_3^0 (допуск A_B). Розглянемо основні етапи побудови РГП між класами X_1^0 і X_2^0 .

1. Формування масиву реалізацій зображення класу X_1^0 $\{g_{1,i}^{(j)}\}$, $j = \overline{1, n_{\text{min}}}$, де n_{min} - мінімальний обсяг репрезентативної вибірки, який визначається кількістю рядків рецепторного поля.

2. Визначення еталонної реалізації $g_1 \in X_1^0$ шляхом усереднення реалізацій $\{g_{1,i}^{(n)}\}$.

3. Вибір системи контрольних допусків $\{\delta_{k,i}\}$ на значення еталонної реалізації g_1 . При цьому $|\delta_{k,i}| \leq |\delta_{n,i}|$, де $\delta_{n,i}$ - нормоване поле допусків для i -го рецептора.

4. Формування навчальної вибірки - бітового масиву $\{BM1[j, i]\}$, $j, i = \overline{1, 152}$ шляхом порівняння значень кожної реалізації масиву $\{g_{1,i}^{(j)}\}$ з відповідними контрольними допусками $\{\delta_{k,i}\}$:

$$\text{if } g_{1,i}^{(j)} \in \delta_{k,i} \text{ then } BM1[j, i] := 1 \text{ else } BM1[j, i] := 0.$$

5. Формування бітового еталонного вектора x_1 , шляхом статистичного усереднення відповідних стовпчиків матриці $\{BM1[j, i]\}$.

6. Формування навчальної вибірки для класу X_2^0 - масиву $\{BM2[j, i]\}$, яке здійснюється аналогічно пунктам 1, 2, 4 для класу X_1^0 .

7. Формування еталонного вектора x_2 шляхом усереднення відповідних стовпців матриці $\{BM2[j, i]\}$, яке здійснюється аналогічно п.5.

8. Побудова оптимальної РГП для класу X_1^0 , в якості якої розглядається гіперсфера з центром у вершині вектора x_1 , шляхом пошуку максимуму інформаційного критерію функціональної ефективності процесу навчання, в якості якого розглянемо, наприклад, інформаційну міру Шеннона в такій модифікації [2]:

$$E_1 = 1 + 0,5 \left(\frac{D_1}{D_1 + \beta} \log_2 \frac{D_1}{D_1 + \beta} + \frac{D_2}{D_2 + \beta} \log_2 \frac{D_2}{D_2 + \beta} + \frac{\alpha}{\alpha + D_1} \log_2 \frac{\alpha}{\alpha + D_1} + \frac{\beta}{D_1 + \beta} \log_2 \frac{\beta}{D_1 + \beta} \right), \quad (2)$$

де D_1 , D_2 , α , β - точнісні характеристики процесу навчання: перша і друга достовірності, помилки першого та другого роду відповідно.

Пошук оптимального радіуса d_1^* за МФСВ здійснюється за алгоритмом:

- а) $d := 0$;
- б) $d := d + 1$;

- в) обчислення E_1 за формулою (2);
- г) порівняння за умови, що $D_1, D_2 \geq 0,5$:

if $E_1[d] \geq E_1[d-1]$ then п.8б
 else if $((D_1 > 0,5) \text{ and } (D_2 > 0,5))$ then п.8д;

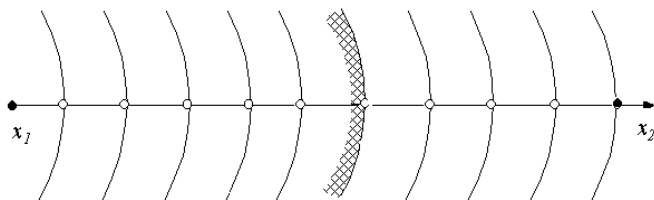
д) визначення екстремального значення радіуса РГП:
 $d_1^* = d_1 - 1$.

9. Визначення нижнього нормованого допуску A_H за формулою $A_H = d_1^* \cdot \chi$, де $\chi = \frac{I_{Л1} - I_{Л2}}{d(x_1 \oplus x_2)}$, де $I_{Л1}$, $I_{Л2}$ значення струму об'єктивної лінзи, які визначають зображення класів X_1^0 і X_2^0 відповідно, $d(x_1 \oplus x_2)$ - міжцентрова кодова відстань.

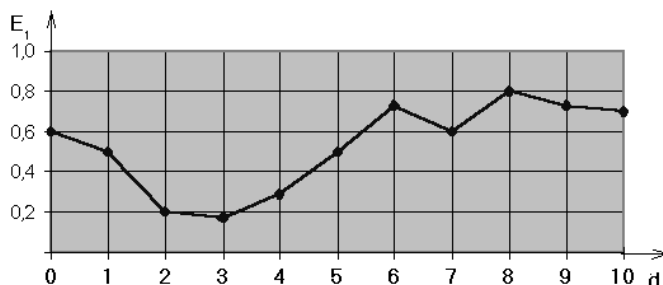
Аналогічно здійснюється визначення верхнього нормованого допуску A_B .

РЕАЛІЗАЦІЯ АЛГОРИТМУ

Алгоритм визначення нормованих допусків для струму об'єктивної лінзи реалізовано для електронного мікроскопа РЕМ-103 виробництва СЕЛМІ (м.Суми, України) після його виходу в робочий режим при прискорюючій напрузі 40 кВ. На рис.2а схематично показано поверхні, які проведено концентровано навколо вершини $x_1 \in X_1^0$ і радіуси яких змінюються на одиницю кодової відстані.



а)



б)

Рисунок 2

При цьому міжцентрова кодова відстань $d(x_1 \oplus x_2) = 10$

при умові, що реалізації класів X_1^0 і X_2^0 склалися з 128 дискрет, тобто крок квантування кривої яскравості для рядка рецепторного поля дорівнював чотирьом пікселям. На рис. 2б наведено графік функції (2) в залежності від радіусу роздільної гіперсфери для класу X_1^0 . В таблиці 1 наведено значення точнісних характеристик і критерію E_1 , обчисленого за формулою (2).

Таблиця 1

k	D_1	α	D_2	β	E_1
1	0,00	1,00	1,00	0,00	0,50
2	0,11	0,89	1,00	0,00	0,40
3	0,31	0,69	0,94	0,06	0,22
4	0,61	0,39	0,86	0,14	0,18
5	0,83	0,17	0,75	0,25	0,31
6	0,97	0,03	0,64	0,36	0,55
7	0,97	0,03	0,54	0,49	0,71
8	0,98	0,02	0,51	0,49	0,61
9	1,00	0,00	0,31	0,69	0,80
10	1,00	0,00	0,17	0,83	0,75
11	1,00	0,00	0,06	0,94	0,70

Аналіз таблиці 1 показує, що робоча область значень критерію (2) знаходиться в межах від $d = 3$ до $d = 7$, в якій екстремальне значення радіуса $d_1^* = 6$.

Оскільки для класу X_1^0 значення струму $I_1 = 1,280$ А, а для класу $X_2^0 - I_2 = 1,268$ А, то різниця $\Delta I_{1,2} = I_1 - I_2 = 12$ мА. Отже A_H визначимо за формулою

$$A_H = I_1 - \frac{\Delta I_{1,2} \cdot d_{1,2}^*}{d(x_1 + x_2)} = 1273 \text{ мА.}$$

На рис.3 наведено залежність критерію (2) від радіуса роздільної гіперсфери класу X_1^0 , оптимальної для розподілу реалізацій класів X_1^0 і X_3^0 .



Рисунок 3

З рис. 3 видно, що для робочої області від $d = 4$ до $d = 10$ оптимальний радіус $d_{1,3}^* = 8$. При $\Delta I_{1,3} = 13$ мА A_B дорівнює 1287 мА. Тоді, розрахункове номінальне значення дорівнює $A_0 = (A_H + A_B)/2 = 1281$ мА.

ВИСНОВКИ

1. Запропонований підхід до визначення нормованих допусків на параметр настройки в рамках алгоритму дискримінантного класифікаційного аналізу за МФСВ дозволяє визначати в просторі параметрів області функціонування складної системи, яка характеризується найвищою функціональною ефективністю системи, що

настроюється.

2. Знання поточних нормованих допусків на параметри настройки спрощує алгоритм настройки і дозволяє підвищити якість і оперативність настройки складної системи.

ПЕРЕЛІК ПОСИЛАНЬ

1. Красноясовський А.С., Скаковська А.М. Автофокусування електронного мікроскопа за зображенням // Обробка сигналів та розпізнавання образів: Праці п'ятої Всеукраїнської міжнародної конференції. - Київ, 2000. - С.183-186.
2. Красноясовський А.С., Черниш А.В. Алгоритм навчання систем розпізнавання за методом функціонально-статистичних випробувань // Вісник Сумського державного університету, 1998. - №2.

УДК 537.874:6

ИНТЕГРОДИФФЕРЕНЦИАЛЬНАЯ МОДЕЛЬ ВЗАИМОДЕЙСТВИЯ МОНОХРОМАТИЧЕСКОЙ ВОЛНЫ С КРУГОВЫМ ЦИЛИНДРОМ (КВАЗИСТАЦИОНАРНЫЙ СЛУЧАЙ)

В.М.Онуфриенко, Т.И.Слюсарова

Представлены результаты дифференциальной моделирования внутреннего поля дифракции для достаточно тонкого цилиндра (квазистационарный случай). Исследование базируется на фрактальном представлении о структуре поверхности цилиндра. Полученная модель позволяет описать с использованием α -характеристик любую заданную степень фрактальности поверхности. Для 0-характеристик результаты совпадают с классическими данными.

Наводяться результати диференціального моделювання внутрішнього поля дифракції для достатньо тонкого циліндра (квазістаціонарний випадок). Дослідження базується на фрактальному уявленні про структуру поверхні циліндра. Отримана модель дозволяє описувати із застосуванням α -характеристик будь-яку задану ступінь фрактальності поверхні. Для 0-характеристик результати співпадають з класичними..

The results of differintegrational modeling of inner field of diffraction for cylinder with fine wall are submitted. The research is based on fractional performance about structure of a surface of a cylinder (quasi-stationary case). The received model allows to describe with use of the α -characteristics any given degree fractality of a surface. For the 0-characteristics the results coincide with the classical data.

ВВЕДЕНИЕ

Изучению явлений взаимодействия электромагнитных волн с фрактальными объектами [1] посвящены работы [2-4]. Так как уравнения Максвелла в интегральной

форме требуют ввода геометрической информации о множествах определения поля и измерения их "величины" (расстояний, окрестностей, площадей, объемов) на контурах, поверхностях и пространственных областях, которые могут быть сильно изрезанными, шероховатыми, пористыми и т. п., то границы объектов в современной электродинамике могут рассматриваться как фрактальные структуры в соответствующих диапазонах масштабов, характеризующиеся одной или несколькими фрактальными размерностями [5-6]. В работе [7] обобщен прием введения покрытия границ и областей определения электромагнитного поля на случай гладкого контура, который на некотором участке имеет фрактальное распределение неоднородностей. Это осуществляется с помощью использования обобщения меры "величины" множества, связанного с выбором некоторой пробной функции $h(\epsilon) = \gamma(\mu) \cdot \epsilon^\mu$ с геометрическим весовым коэффициентом $\gamma(\mu)$, с выбором покрытия рассматриваемого множества точек элементами с характерным размером ϵ (например, ϵ - диаметр покрытия), с образованием μ -меры Хаусдорфа $H_\mu(\epsilon) = \sum h(\epsilon)$, которая служит мерой протяженности и искривленности граничной линии. В общем случае для фрактальных объектов мера $H_\mu(\epsilon)$ равна нулю (для $\mu > 0$) или бесконечности (для $\mu < 0$) при $\epsilon \rightarrow 0$ [8]. Поэтому, вместо сложной процедуры

геометрического построения фрактального множества, нахождения его μ -меры Хаусдорфа с последующими граничными переходами, будем использовать аппарат дробного интегродифференциального исчисления [9]

$${}_a D_x^\alpha(f)(x) = \frac{1}{\Gamma(1-\alpha)} \cdot \frac{d \left(\int_a^x f(t) \cdot \frac{dt}{(x-t)^\alpha} \right)}{dx}, \quad x > a, \quad 0 < \alpha < 1, \quad (1)$$

$${}_a I_x^\alpha(f)(x) = \frac{1}{\Gamma(\alpha)} \cdot \int_a^x f(t) \cdot \frac{dt}{(x-t)^{1-\alpha}}, \quad x > a, \quad \alpha > 0, \quad (2)$$

где $\Gamma(\cdot)$ - гамма-функция Эйлера, $D^\alpha(f)(x)$ и $I^\alpha(f)(x)$ - дробная производная и дробный интеграл порядка α . Примеры применения интегродифференциалов рассмотрены в работах [10-15; 17-18].

ПОСТАНОВКА ЗАДАЧИ

Сформулируем следующую задачу дифракции. Пусть плоская линейно поляризованная электромагнитная моно-

хроматическая волна \vec{E}_0, \vec{H}_0 падает на бесконечный круговой цилиндр радиуса R , поверхность которого обладает фрактальными свойствами. Рассмотрим случай, когда на-

правление вектора \vec{H}_0 совпадает с направлением оси Z , а

направление вектора \vec{E}_0 - с направлением оси Y (перпендикулярная поляризация). Вторичное поле вне цилиндра

обозначим $\vec{E}^{(\alpha)-}, \vec{H}^{(\alpha)-}$, а внутри него - $\vec{E}^{(\alpha)+}, \vec{H}^{(\alpha)+}$.

$\vec{H}_0 = \vec{z}_0 \cdot \dot{H}_0 = \vec{z}_0 \cdot \dot{A} e^{-ik_0 x}$, $\vec{E}_0 = \vec{y}_0 \cdot \dot{E}_0 = \vec{y}_0 \cdot \dot{A} W_0 e^{-ik_0 x}$,

где $k_0 = \omega \cdot \sqrt{\epsilon_0 \mu_0}$, $W_0 = \sqrt{\mu_0 / \epsilon_0}$. Свойства среды внутри и вне цилиндра учитываем с помощью диэлектрической и магнитной проницаемостей: $\dot{\epsilon}, \dot{\mu}$ при $r < R$; ϵ_0, μ_0 при $r > R$.

Особенности поверхности учитываем с помощью α -

характеристик $\vec{E}^{(\alpha)-} = D^\alpha \vec{E}^{(\alpha)-}$, $\vec{H}^{(\alpha)-} = D^\alpha \vec{H}^{(\alpha)-}$ компо-

нент $\vec{E}^{(\alpha)-}, \vec{H}^{(\alpha)-}$ поля. Сами же $D^\alpha \vec{E}^{(\alpha)-+}, D^\alpha \vec{H}^{(\alpha)-+}$

ищем как решения однородного уравнения Гельмгольца $\nabla^2 D^\alpha \vec{E}^{(\alpha)-+} + k_{\pm}^2 \cdot D^\alpha \vec{E}^{(\alpha)-+} = 0$, удовлетворяющие со-

ответствующим граничным условиям, $k_{-}^2 = k_0^2 = \omega^2 \epsilon_0 \mu_0$,

$k_{+}^2 = k^2 = \omega^2 \dot{\epsilon} \dot{\mu}$. Параметры k, W определяются аналогично k_0, W_0 .

РЕШЕНИЕ

Задачу рассмотрим в цилиндрических координатах $x = r \cdot \cos \varphi$, $y = r \cdot \sin \varphi$, $z = z$. Решения построим в виде рядов по функциям Бесселя 1-го рода $J_n(x)$ и 3-го

рода $H_n^{(2)}(x)$ ($J_n(x)$ - для падающей и прошедшей волн, $H_n^{(2)}(x)$ - для отраженной волны, n - порядок функции).

Учитываем также, что внешнее поле дифракции должно

удовлетворять условию излучения. Итак, $\vec{H}_0 = \vec{z}_0 \cdot \dot{A} e^{-ik_0 r \cos \varphi}$. В дальнейшем используем принцип

двойственности: $\dot{\epsilon} \Leftrightarrow -\dot{\mu}$, $\vec{E}_0 \Leftrightarrow \vec{H}_0$, $D^\alpha \vec{E}^{(\alpha)-} \Leftrightarrow D^\alpha \vec{H}^{(\alpha)-}$, $\vec{E}^{(\alpha)-} \Leftrightarrow \vec{H}^{(\alpha)-}$.

Воспользовавшись известным разложением $e^{ix \cdot \cos \beta} = \sum_{n=-\infty}^{+\infty} (-i)^n \cdot J_n(x) \cdot e^{in\beta}$, получим выражения для магнит-

ной составляющей падающей волны

$$\vec{H}_0 = \vec{z}_0 \cdot \dot{H}_0 = \vec{z}_0 \cdot \dot{A} \sum_{n=-\infty}^{+\infty} (-i)^n \cdot J_n(k_0 r) \cdot e^{in\varphi}, \quad r > R \quad (3)$$

и для α -характеристик напряженностей магнитных составляющих дифрагированных внешнего и внутреннего полей соответственно

$$D^\alpha \vec{H}^{(\alpha)-} = \vec{z}_0 \cdot D^\alpha \dot{H}^{(\alpha)-} = \vec{z}_0 \cdot \dot{A} \sum_{n=-\infty}^{+\infty} (-i)^n c_n \cdot (H_n^{(2)}(k_0 r) \cdot e^{in\varphi}), \quad r > R, \quad (4)$$

$$D^\alpha \vec{H}^{(\alpha)+} = \vec{z}_0 \cdot D^\alpha \dot{H}^{(\alpha)+} = \vec{z}_0 \cdot \dot{A} \sum_{n=-\infty}^{+\infty} (-i)^n b_n \cdot (J_n(kr) \cdot e^{in\varphi}), \quad r < R. \quad (5)$$

Граничные условия

$$\begin{aligned} \dot{H}_0 \varphi + D^\alpha \dot{H}_\varphi^{(\alpha)-} &= D^\alpha \dot{H}_\varphi^{(\alpha)+}, \\ \dot{E}_0 + D^\alpha \dot{E}^{(\alpha)-} &= D^\alpha \dot{E}^{(\alpha)+} \end{aligned} \quad (6)$$

записаны, исходя из требования непрерывности \vec{E}, \vec{H}_φ на границе раздела сред.

Для нахождения электрического поля воспользуемся известным соотношением:

$$\vec{E} = \frac{i}{-\omega \dot{\epsilon}} \cdot \text{rot} \vec{H} = \frac{i}{-\omega \dot{\epsilon}} \cdot \left(\vec{r}_0 \cdot \frac{1}{r} \cdot \frac{\partial H}{\partial \varphi} - \varphi_0 \cdot \frac{\partial H}{\partial r} \right). \quad (7)$$

Тогда, на основании формулы (7), для электрических составляющих полей получим:

при $r > R$

$$\vec{E}_0 = \frac{i\dot{A}}{-\omega\epsilon_0} \sum_{n=-\infty}^{+\infty} (-i)^n \left(\vec{r}_0 \frac{in}{r} J_n(k_0 r) - \vec{\Phi}_0 k_0 J_n'(k_0 r) \right) e^{in\varphi}, \quad (8)$$

$$D^\alpha \vec{E}^{\dot{(\alpha)-}} = \frac{i\dot{A}}{-\omega\epsilon_0} \sum_{n=-\infty}^{+\infty} (-i)^n c_n \left(\left(\vec{r}_0 \frac{in}{r} H_n^{(2)}(k_0 r) - \vec{\Phi}_0 k_0 \frac{dH_n^{(2)}(k_0 r)}{dr} \right) e^{in\varphi} \right), \quad (9)$$

при $r < R$

$$D^\alpha \vec{E}^{\dot{(\alpha)+}} = \frac{i\dot{A}}{-\omega\epsilon} \sum_{n=-\infty}^{+\infty} (-i)^n b_n \times \left(\left(\vec{r}_0 \frac{in}{r} J_n(kr) - \vec{\Phi}_0 k J_n'(kr) \right) e^{in\varphi} \right). \quad (10)$$

Неизвестные коэффициенты c_n , b_n подлежат определению. Подставляя в граничные условия

$$-D^\alpha \dot{H}_\varphi^{(\alpha)+} + D^\alpha \dot{H}_\varphi^{(\alpha)-} = -\dot{H}_0 \varphi - D^\alpha \dot{E}^{(\alpha)+} + D^\alpha \dot{E}^{(\alpha)-} = -\dot{E}_0 \quad (11)$$

ряды (3)-(5) и (8)-(10), для определения неизвестных коэффициентов c_n , b_n получим систему:

$$-b_n J_n(k \cdot R) + c_n \cdot H_n^{(2)}(k_0 \cdot R) = -J_n(k_0 \cdot R),$$

$$b_n \cdot \frac{k}{-\epsilon} \cdot J_n'(k \cdot R) - c_n \cdot \frac{k_0}{-\epsilon_0} \cdot H_n^{(2)'}(k_0 \cdot R) = \frac{k_0}{-\epsilon_0} \cdot J_n'(k_0 \cdot R).$$

В итоге, коэффициенты c_n , b_n :

$$b_n = \frac{J_n(k_0 R) H_n^{(2)'}(k_0 R) - J_n'(k_0 R) \cdot H_n^{(2)}(k_0 R)}{J_n(kR) \cdot H_n^{(2)'}(k_0 R) - \frac{W}{W_0} J_n'(kR) \cdot H_n^{(2)}(k_0 R)}, \quad (12)$$

$$c_n = \frac{-J_n(kR) \cdot J_n'(k_0 R) + \frac{W}{W_0} J_n'(kR) \cdot J_n(k_0 R)}{J_n(kR) \cdot H_n^{(2)'}(k_0 R) - \frac{W}{W_0} J_n'(kR) \cdot H_n^{(2)}(k_0 R)}. \quad (13)$$

РЕЗУЛЬТАТЫ И ВЫВОДЫ

Пусть цилиндр (с фрактальными свойствами поверхности) по диаметру $d = 2R$ мал в сравнении с длиной волны λ . Найдем выражения для α -характеристик

$D^\alpha \vec{E}^{\dot{(\alpha)+}}$, $D^\alpha \vec{H}^{\dot{(\alpha)+}}$ компонент и сами компоненты $\vec{E}^{\dot{(\alpha)+}}$, $\vec{H}^{\dot{(\alpha)+}}$ внутреннего дифрагированного поля.

Воспользовавшись формулой $J_n(x) \approx \frac{x^n}{n! \cdot 2^n}$ (при $x \ll 1$)

и ее частными случаями $J_0 x \approx 1$, $J_1(x) \approx \frac{x}{2}$, из выражения (12) находим, что при $\frac{R}{\lambda} \rightarrow 0$ $b_0 = 0$ и

$b_1 = b_{-1} = \frac{2}{k/k_0 + W/W_0}$. При $\frac{R}{\lambda} \rightarrow 0$ в выражении (5) остается только одно слагаемое при $n = 0$, в котором $J_0(kr) \rightarrow 1$. В результате выражение (5) примет вид

$$D^\alpha \vec{H}^{\dot{(\alpha)+}} \Big|_{x=0} = \vec{z}_0 \cdot D^\alpha \dot{H}^{(\alpha)+} \Big|_{x=0} = \vec{z}_0 \dot{H}_0 \Big|_{x=0} = \vec{z}_0 \dot{A}, \quad (14)$$

а после интегрирования выражения (14) по φ получим вектор

$$\vec{H}_\varphi^{\dot{(\alpha)+}} \Big|_{x=0} = \vec{z}_0 \dot{H}_\varphi^{(\alpha)+} \Big|_{x=0},$$

где

$$\begin{aligned} \dot{H}_\varphi^{(\alpha)+} \Big|_{x=0} &= {}_0 I_\varphi^\alpha (D^\alpha \dot{H}^{(\alpha)+} \Big|_{x=0}) = {}_0 I_\varphi^\alpha (\dot{A}) = \\ &= \frac{\dot{A}}{\Gamma(\alpha)} \cdot \int_0^\varphi \frac{dt}{(\varphi-t)^{1-\alpha}} = \frac{\dot{A} \varphi^\alpha}{\Gamma(\alpha+1)}. \end{aligned} \quad (15)$$

Для определения $D^\alpha \vec{E}^{\dot{(\alpha)+}}$ достаточно взять в выражении (5) члены ряда с $n = 1$ и $n = -1$, так как остальные члены (в том числе и нулевой) исчезают при $\frac{R}{\lambda} \rightarrow 0$. Тогда из (10) имеем

$$\begin{aligned} D^\alpha \vec{E}^{\dot{(\alpha)+}} &= \dot{A} \frac{2 \cdot W}{(k/k_0 + W/W_0)} \cdot \left(\left(\vec{r}_0 \frac{i}{2} - \vec{\Phi}_0 \frac{1}{2} \right) e^{i\varphi} - \right. \\ &\left. - \left(\vec{r}_0 \frac{i}{2} + \vec{\Phi}_0 \frac{1}{2} \right) e^{-i\varphi} \right) = -\dot{A} \frac{2 \cdot W}{(k/k_0 + W/W_0)} \times \\ &\times (\vec{r}_0 \sin \varphi + \vec{\Phi}_0 \cos \varphi). \end{aligned} \quad (16)$$

После интегрирования выражения (16) по φ вектор $\vec{E}^{\dot{(\alpha)+}}$ принимает вид

$$\begin{aligned} \vec{E}_\varphi^{\dot{(\alpha)+}} &= \vec{r}_0 \dot{E}_{\varphi r}^{(\alpha)+} + \vec{\Phi}_0 \dot{E}_{\varphi \varphi}^{(\alpha)+} = \\ &= -\dot{A} \frac{1}{\Gamma(\alpha)} \cdot \frac{2 \cdot W}{(k/k_0 + W/W_0)} \int_0^\varphi \frac{(\vec{r}_0 \sin t + \vec{\Phi}_0 \cos t) dt}{(\varphi-t)^{1-\alpha}}. \end{aligned} \quad (17)$$

На рисунках представлены полярные графики модулей векторов $\vec{H}_\varphi^{\dot{(\alpha)+}}$, $\vec{H}^{\dot{(\alpha)+}}$ (рис.1-4), $\vec{E}_{\varphi r}^{\dot{(\alpha)+}} = \vec{r}_0 \dot{E}_{\varphi r}^{(\alpha)+}$, $\vec{E}_{\varphi \varphi}^{\dot{(\alpha)+}} = \vec{\Phi}_0 \dot{E}_{\varphi \varphi}^{(\alpha)+}$, $\vec{E}_r^{\dot{(\alpha)+}} = \vec{r}_0 \dot{E}_r^{(\alpha)+}$, $\vec{E}_\varphi^{\dot{(\alpha)+}} = \vec{\Phi}_0 \dot{E}_\varphi^{(\alpha)+}$

(рис. 5-8) дифрагированного внутреннего поля при различных значениях α для достаточно тонкого цилиндра. Рисунки демонстрируют сравнение фрактального $|\dot{H}_\phi^{(\alpha)+}|$, $|\dot{E}_{\phi r}^{(\alpha)+}|$, $|\dot{E}_{\phi\phi}^{(\alpha)+}|$ (сплошная линия) и классического $|\dot{H}^+|$, $|\dot{E}_r^+|$, $|\dot{E}_\phi^+|$ (пунктирная линия) случаев:

- 1) для магнитного поля: $\dot{A} = 1$, $\varphi = -\pi, -\pi + \frac{\pi}{500}, -\pi + \frac{2\pi}{500}, \dots, \pi$;
- 2) для электрического поля: $\dot{A} = 1$, $\dot{B} = \frac{2 \cdot \dot{W}}{(\dot{k}/k_0 + \dot{W}/W_0)} = 10$, $\varphi = -\pi, -\pi + \frac{\pi}{500}, -\pi + \frac{2\pi}{500}, \dots, \pi$.

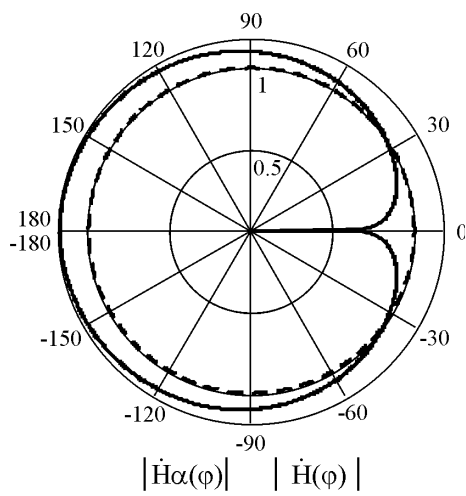


Рисунок 1 - Модуль напряжённости дифрагированного внутреннего магнитного поля при $\alpha = 0, 1$

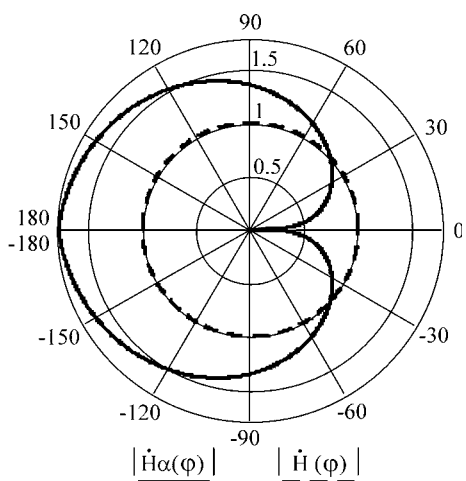


Рисунок 2 - Модуль напряжённости дифрагированного внутреннего магнитного поля при $\alpha = 0, 4$

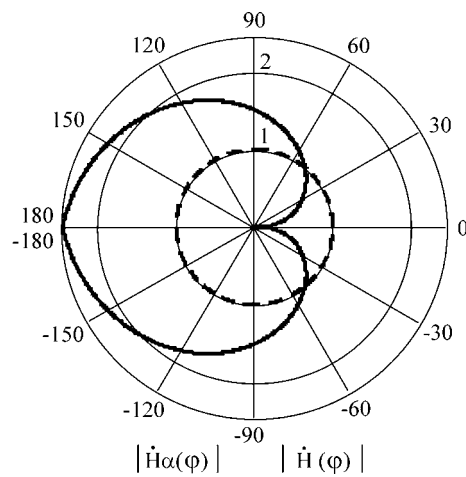


Рисунок 3 - Модуль напряжённости дифрагированного внутреннего магнитного поля при $\alpha = 0, 7$

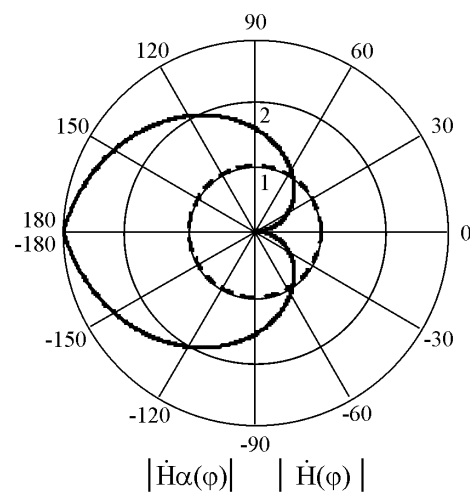


Рисунок 4 - Модуль напряжённости дифрагированного внутреннего магнитного поля при $\alpha = 0, 9$

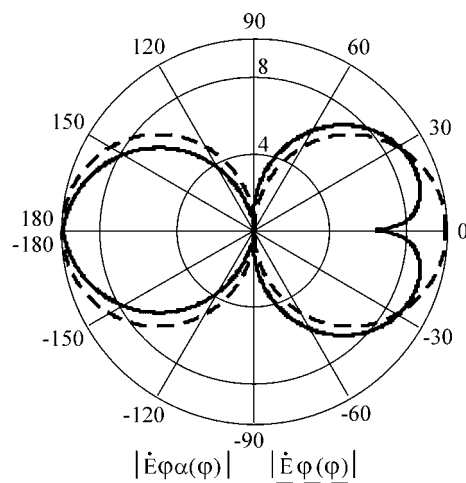


Рисунок 5 - Модуль напряжённости дифрагированного внутреннего электрического поля при $\alpha = 0, 1$
($|\dot{E}_{\phi\phi}^{(\alpha)+}|$ и $|\dot{E}_\phi^+|$)

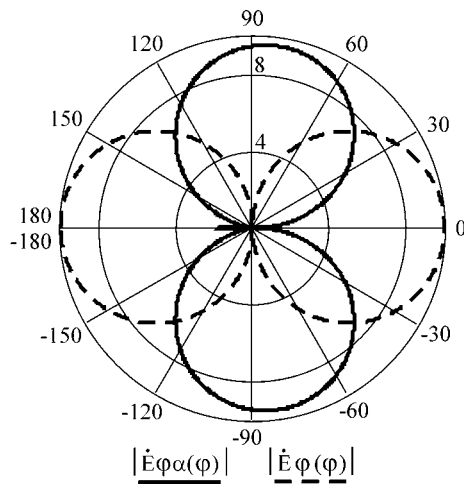


Рисунок 6 - Модуль напряжённости дифрагированного внутреннего электрического поля при $\alpha = 0, 9$

$$(|\dot{E}_{\varphi\alpha}^{(\alpha)+}| \text{ и } |\dot{E}_{\varphi}^{+}|)$$

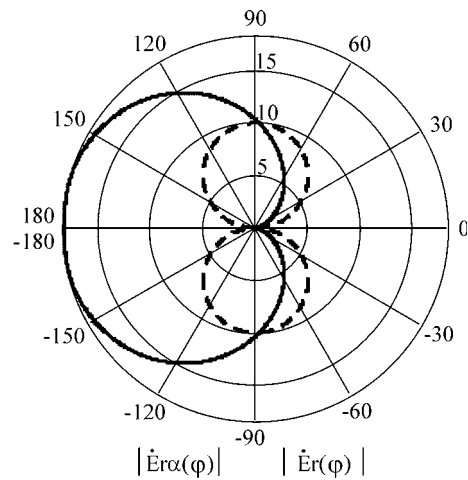


Рисунок 8 - Модуль напряжённости дифрагированного внутреннего электрического поля при $\alpha = 0, 9$

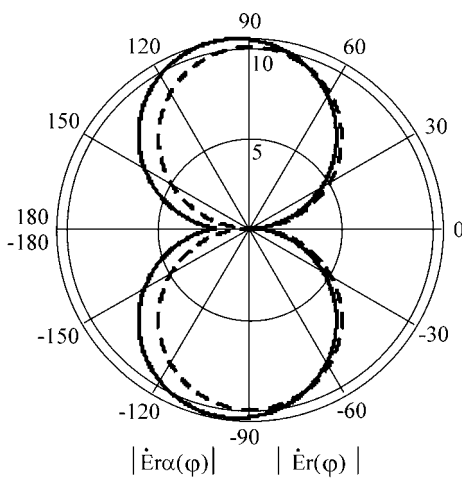


Рисунок 7 - Модуль напряжённости дифрагированного внутреннего электрического поля при $\alpha = 0, 1$

$$(|\dot{E}_{\varphi r}^{(\alpha)+}| \text{ и } |\dot{E}_r^{+}|)$$

При значении $\alpha \rightarrow 0$ результаты проектирования интегродифференциальной модели совпадают с классическими. Анализ выражения (15) показывает, что решение фрактальной задачи по форме совпадает с решением классической, если в качестве волнового импеданса выбрать $\tilde{W}^{(\alpha)} = \tilde{W} \cdot \Gamma(\alpha + 1)\varphi^{-\alpha}$ (зависит от угла φ и порядка α интегродифференциала). Результаты численного расчета подтверждают известные факты смещения диаграммы направленности поля, связанные с шероховатой структурой поверхности. "Фрактальность по углу" (интегрирование по переменной φ) можно интерпретировать как модель описания неоднородного распределения

зарядов и токов на поверхности цилиндра. Результаты исследования позволяют ставить и решать задачи о дифракции на цилиндре волны иной поляризации [17] на фрактальной сфере и др.

ПЕРЕЧЕНЬ ССЫЛОК

1. Федер Е. Фракталы: Пер. с англ. - М.: Мир, 1991. - 254 с.
2. Onufriyenko V. On " α -features" of electrical waves above impedance plane // Proceedings 12th International Conference on Microwaves & Radar. - Krakow (Poland, May 20-22, 1998). - Vol. 1. - P. 212-215.
3. Онуфрієнко В. М. Взаємодія плоскої електромагнітної хвилі з металізованою фрактальною поверхнею // Радиофизика и электроника. - Харьков: ИРЭ им. А.Я.Усикова НАН Украины, 1999. - Т. 4, № 2. - С. 19-20.
4. Онуфрієнко В. М., Лисоконь І. В., Самолчев П. О., Слюсарова Т. І. Електромагнітні хвилі на фрактальній межі розділу двох середовищ // Радіоелектроніка, інформатика, управління. - 1999. - № 1. - С. 20-23.
5. Фракталы в физике // Труды 6-го международного симпозиума по фракталам в физике. - Триест (Италия, 9-12 июля 1985 г.): Пер. с англ./Под ред. Л. Пьетронеро, Э. Тозотти. - М.: Мир, 1988. - 672 с.
6. Працьовитий М. В. Фрактальний підхід у дослідженнях сингулярних розподілів. - Київ: НПУ ім. М. П. Драгоманова, 1998. - 296 с.
7. Онуфрієнко В. М. Фізико-геометрична інтерпретація α -характеристик електромагнітного поля // Радиофизика и электроника. - Харьков: ИРЭ им. А.Я.Усикова НАН Украины, 1999. - Т. 4, № 1. - С. 7-10.
8. Зельдович Я. Б., Соколов Д. Д. Фракталы, подобие, промежуточная асимптотика // УФН, 1985. - Т. 146, № 3. - С. 493-506.
9. Самко С. Г., Килбас А. А., Маричев О. И. Интегралы и производные дробного порядка и некоторые их приложения. - Минск: Наука и техника, 1987. - 688 с.
10. Engheta N. On Fractional Calculus and Fractional Multipoles in Electromagnetism // IEEE Transactions on Antennas and Propagation. - April, 1996. - Vol. 44, NO. 4. - P. 554-566.
11. Engheta N. Electrostatic "Fractional" Image Methods Perfectly Conducting Wedges and Cones // IEEE Transactions on Antennas and Propagation. - December, 1996. - Vol. 44, NO. 12. - P. 1565-1574.
12. Engheta N. On the Role of Fractional Calculus in Electromag-

- netic Theory // IEEE Antennas & Propagation Magazine. - August, 1997. - Vol. 39, NO. 4. - P. 35-46.
13. Vladimir M. Onufriyenko, Petr A. Samolchev, Tatyana I. Slyusarova. Interaction of an electrostatic field with a dielectric body // Conference Proceedings MIKON'2000. - Wroclaw (Poland, May 22-24, 2000). - Vol. 2. - P. 502-505.
 14. Volodymir M. Onufriyenko, Eldar I. Veliev. Electromagnetic theory radiation of electrical and magnetic fractal surface currents // Conference Proceedings ISAP'2000. - Fukuoka (Japan, August 21-25, 2000). - Vol. 3. - P. 1319-1322.
 15. Onufriyenko V. M., Samolchev P. A., Slyusarova T. I. Reflection of a Plane Wave from a Cylinder with Fractal Properties of the Surface (far-field region) // Conference Proceedings MMET'2000. - Kharkov (Ukraine, September 12-15, 2000). - Vol. 2. - P. 420-422.
 16. Никольский В. В. Электродинамика и распространение радиоволн. - М.: Наука, 1973. - 608 с.
 17. Онуфриенко В. М., Слюсарова Т. И. Интегродифференциальная модель взаимодействия монохроматической волны с круговым цилиндром // Радиотехника (Всеукраинский межведомственный научно-технический сборник). - Харьков. - 2001. - № 118. - С. 16-21.
 18. Самолчев П. О., Слюсарова Т. І. Інтегродиференціальна модель взаємодії монохроматичної хвилі з круговим циліндром // Матеріали 5-го Міжнародного молодіжного форуму "РАДІОЕЛЕКТРОНІКА І МОЛОДЕЖЬ В ХХІ ВЕКЕ". - Харків: ХТУРЕ, 24-26 квітня 2001 р. - Ч. 2. - С. 168-169.

УДК 621396.9

ПРОСТРАНСТВЕННО-(ПОЛЯРИЗАЦИОННО-) ВРЕМЕННАЯ ФИЛЬТРАЦИЯ СИГНАЛОВ НА ФОНЕ КОМБИНИРОВАННЫХ ПОМЕХ

Д.М.Пиза, М.П.Чернобородов, Ю.Л.Мейстер

Рассмотрены ограничения, возникающие при обработке сигналов на фоне комбинированных помех. Показано, что одной из причин, ограничивающих возможности пространственно-временной обработки сигналов, является амплитудно-фазовая модуляция пассивной составляющей помехи, принимаемой компенсационными каналами. Оценены модуляционные характеристики различных вариантов построения систем помехозащиты.

Розглянуто обмеження, які виникають за обробки сигналів на фоні комбінованих завад. Показано, що однією з причин обмеження можливостей просторово-часової обробки сигналів є амплітудно-фазова модуляція пасивної складової пасивної завади, яка приймається компенсаційними каналами. Проведена оцінка модуляційних характеристик за різних варіантів побудови систем завадозахисту.

The restrictions arising from signals processed on a background combined noise are considered. This paper shows that one of the causes restricting the possibilities for spatial-temporal signals processing is the gain-phase modulation of a passive component noise received by the compensatory channels. The drive characteristics of various variants for construction of noise-immune systems are estimated.

Предположим, что в дальней зоне на антенную систему импульсно-доплеровской РЛС одновременно воздействуют активная и пассивная компоненты комбинированной помехи. Будем считать, что пространственно-(поляризационно-) временная обработка сигналов выполняется в два этапа. Структурная схема такой обработки приведена на рис. 1.

На первом этапе сигнал подвергается обработке с помощью пространственного (поляризационного) фильтра. В результате такой обработки происходит переход от двух или многих каналов к одному. При этом на первом этапе происходит компенсация активной шумовой помехи. На втором этапе осуществляется временная согласованная обработка. Временная обработка, в соответствии с рис. 1, позволяет также выполнить селекцию полезного сигнала

на фоне пассивной помехи и решает проблему стабилизации уровня ложных тревог (СУЛТ).

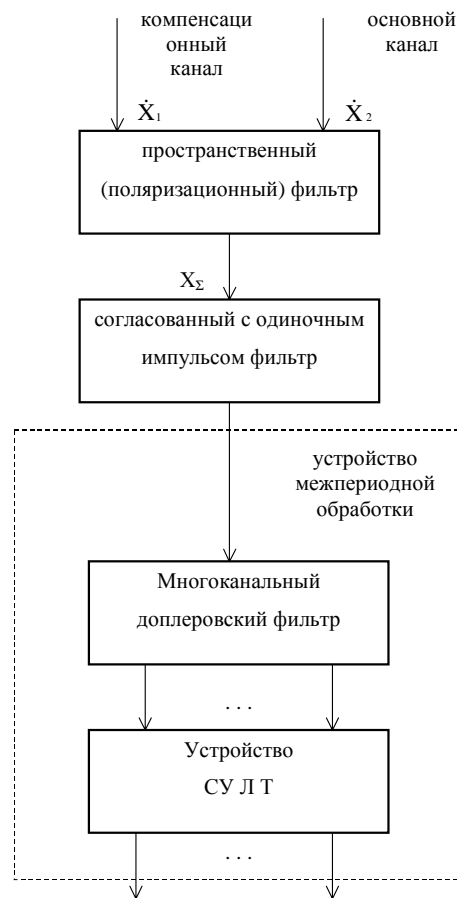


Рисунок 1 - Структурная схема пространственно-(поляризационно-)временного фильтра

Проведем анализ ограничений эффективности такой системы обработки при одновременном воздействии активных и пассивных помех. При этом на входах основного и вспомогательного каналов поляризационного (см. рис. 1) или пространственного фильтров действуют комплексные напряжения

$$\dot{X}_1 = \dot{\xi}_1 + \dot{\eta}_1, \quad (1)$$

$$\dot{X}_2 = \alpha_2 \dot{\xi}_2 + \varepsilon_2 \dot{\eta}_2, \quad (2)$$

где $\dot{\xi}_1, \dot{\xi}_2, \dot{\eta}_1, \dot{\eta}_2$ - комплексные составляющие активной и пассивной компонент комбинированной помехи, действующие в основном (индекс "1") и дополнительном (индекс "2") каналах; α_2 и ε_2 - скаляры, характеризующие различие модулей коэффициентов усиления антенн основного и дополнительного (компенсационного) каналов в направлении на источники активной и пассивной помех.

На выходе пространственного (поляризационного) фильтра комбинированная помеха может быть представлена, как

$$\dot{X}_\Sigma = \dot{X}_1 + \dot{W}\dot{X}_2 = \dot{\xi}_1 + \dot{\eta}_1 + \dot{W}(\alpha_2 \dot{\xi}_2 + \varepsilon_2 \dot{\eta}_2), \quad (3)$$

где \dot{W} - комплексный весовой коэффициент, а ее дисперсия

$$\sigma_\Sigma^2 = (\sigma_\xi^2 + \dot{W}^2 \alpha_2^2 \sigma_\xi^2 + 2\dot{W}\alpha_2 \dot{\rho}_\xi \sigma_\xi^2) + (\sigma_\eta^2 + \dot{W}^2 \varepsilon_2^2 \sigma_\eta^2 + 2\dot{W}\varepsilon_2 \dot{\rho}_\eta \sigma_\eta^2) = \sigma_{\Sigma\xi}^2 + \sigma_{\Sigma\eta}^2. \quad (4)$$

$$\text{Здесь } \dot{\rho}_\xi = \frac{(\overline{\xi_1^* \xi_2})^2}{\overline{\xi_1^2} + \overline{\xi_2^2}} \text{ и } \dot{\rho}_\eta = \frac{(\overline{\eta_1^* \eta_2})^2}{\overline{\eta_1^2} + \overline{\eta_2^2}} \quad (5)$$

комплексные коэффициенты корреляции, соответственно, активной и пассивной компонент комбинированной помехи в каналах приема.

В результате исследования экстремумов частных производных слагаемых уравнения (3), в предположении независимости компонент, получены оптимальные значения весовых коэффициентов

$$\dot{W}_{\xi \text{opt}} = -\frac{\rho_\xi e^{j\vartheta}}{\alpha_2}, \quad (6)$$

$$\dot{W}_{\eta \text{opt}} = -\frac{\rho_\eta e^{j\upsilon}}{\varepsilon_2}. \quad (7)$$

Из выражений (6) и (7) непосредственно следует, что эффективная компенсация помех в адаптивном фильтре возможна в случае, когда весовые коэффициенты сформированы с учетом реальных соотношений уровней помех

и их фазировки в каналах приема. В общем случае $\rho_\xi \neq \rho_\eta, \alpha_2 \neq \varepsilon_2, \vartheta \neq \upsilon$, поэтому пространственный (поляризационный) фильтр не может эффективно минимизировать обе составляющие комбинированной помехи. Адаптивный фильтр, использующий либо пространственные либо поляризационные отличия сигналов и помех, минимизирует компоненту с большей интенсивностью. Если преобладающей является интенсивность активной помехи, то весовые коэффициенты формируются по выражению (6), в противном случае - по выражению (7).

Таким образом, возникновение достаточно мощной пассивной помехи в цепях формирования весовых коэффициентов пространственного (поляризационного) фильтра наряду с имеющейся активной помехой приводит к перестройке весовых коэффициентов и, как следствие, к ухудшению подавления активной маскирующей помехи. При этом, за счет возникающих переходных процессов в компенсационном канале адаптивного фильтра, модулируется пассивная составляющая комбинированной помехи $\dot{\eta}_2$.

Это, в свою очередь, может приводить к существенному снижению эффективности устройств временной межпериодной обработки [1].

Анализ реальной помеховой обстановки дает основание сделать вывод, что даже в те моменты времени, когда одновременно выполняются условия $\alpha_2 = \varepsilon_2$, и $\vartheta = \upsilon$, (т.е. активная и пассивная помехи действуют с одного направления) комбинированная помеха на выходе типовых пространственно- (поляризационно-) временных фильтров не может быть эффективно подавленной. Это обусловлено низким межканальным коэффициентом корреляции разившейся по пространству пассивной помехи, ограничивающей эффективность пространственного (поляризационного) фильтра. Действительно, такая помеха не может быть представлена дельта-коррелированным по угловым координатам процессом [2].

Другим существенным ограничением, обуславливающим низкую эффективность фильтрации пассивной составляющей комбинированной помехи, является флуктуация весовых коэффициентов адаптивных пространственных (поляризационных) фильтров. Флуктуационные ошибки адаптивных фильтров обусловлены ограниченным интервалом усреднения в формирователях весовых коэффициентов. Флуктуация квадратурных составляющих весовых коэффициентов приводит к амплитудно-фазовой модуляции пассивной составляющей помехи, принимаемой компенсационными каналами приема.

Еще одной причиной возникновения модуляции пассивной помехи, принимаемой компенсационными каналами приема, является нестационарность активной составляющей комбинированной помехи особенно при малых значениях частоты повторения импульсов и относительно большом количестве фильтровых каналов устройства межпериодной обработки.

Если представить весовой коэффициент \dot{W} в виде суммы некоторого оптимального значения, определяемого выражением (6), и флуктуационной ошибки

$$\dot{W} = \dot{W}_{\xi_{opt}} + \dot{\epsilon} = \dot{\epsilon} - \rho_{\xi} \frac{e^{j\nu}}{\alpha_2}, \quad (8)$$

то пассивная помеха на выходе адаптивного пространственного (поляризационного) фильтра может быть записана

$$X_{\Sigma\eta} = \dot{\eta}_1 + W_{\xi_{opt}} \epsilon_2 \dot{\eta}_2 + \epsilon_2 \dot{\eta}_2. \quad (9)$$

Анализ показывает, что нестационарная составляющая пассивной помехи на входе устройства межпериодной обработки определяется третьим слагаемым в выражении (9). При этом снижение уровня нестационарной составляющей, ограничивающей эффективность межпериодной обработки, может быть обеспечена уменьшением флуктуационной ошибки $\dot{\epsilon}$.

В этой связи представляло интерес оценить величину флуктуационной ошибки для различных алгоритмов работы вычислителей весовых коэффициентов. На рис. 2 приведены результаты математического моделирования, характеризующие зависимость отклонения весовых коэффициентов σ_w от отношения помеха/собственный шум q_n .

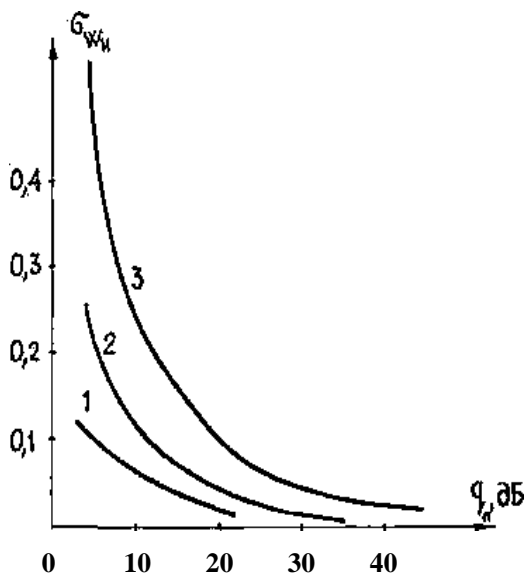


Рисунок 2 - Зависимость среднеквадратических отклонений весовых коэффициентов σ_w от отношения помеха/собственный шум q_n (прямое вычисление весов)

Вычислению среднеквадратических значений предшествовала операция центрирования по реализуемым значениям математического ожидания весовых коэффициентов.

Причем кривыми 1, 2 и 3 приведены соответствующие зависимости для интервалов интегрирования равных 64, 16 и 4 выборки при прямом вычислении весовых коэффициентов [3].

Анализ флуктуационных характеристик проведен также для адаптивных фильтров на базе известных автокомпенсаторов помех с градиентным, а также релейным методами вычисления весовых коэффициентов [3]. На рис. 3 кривыми 1 и 2 приведены зависимости нормированных среднеквадратических отклонений весовых коэффициентов от параметра q_n соответственно, для градиентного и релейного алгоритмов.

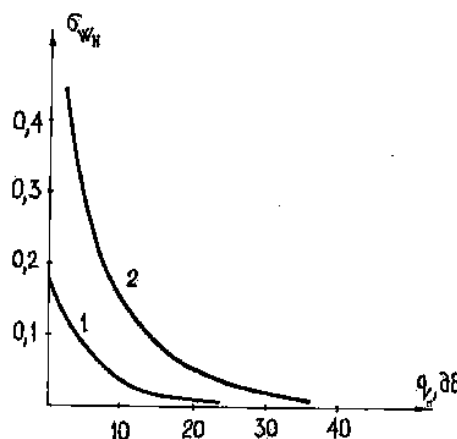


Рисунок 3 - Зависимость среднеквадратических значений σ_w весовых коэффициентов от отношения помеха/собственный шум q_n (1 - градиентный алгоритм, 2 - релейный алгоритм)

Из совместного анализа кривых, приведенных на рис. 2 и 3, следует, что чем меньше интервал интегрирования или чем выше скорость сходимости алгоритма формирования весовых коэффициентов, тем больше флуктуационная ошибка и тем более весомым является ограничение эффективности устройств временной обработки. Флуктуационная ошибка также существенно зависит от действующего отношения помеха/шум q_n . Причем при малых значениях отношения активная помеха/шум флуктуационная ошибка существенно возрастает.

ПЕРЕЧЕНЬ ССЫЛОК

1. Теоретические основы радиолокации. / Под ред. Я.Д.Ширмана - М.: Сов.радио, 1970. - 560 с.
2. Обработка сигналов в многоканальных РЛС /А.П.Лукошкин, С.С.Каринский, А.А.Шаталов и др./ Под ред. А.П.Лукошкина. - М.: Радио и связь, 1983. - 328 с., ил.
3. Пиза Д.М., Татарчук И.К., Томачинский Н.Т. Алгоритмы адаптации цифровых автокомпенсаторов. / Радиопромышленность, № 11, 1993.-с. 34-37.

ТРАКТОВЫЙ УСИЛИТЕЛЬ ДЛЯ МАРКЕРНО-СВЯЗНОГО ВОЛНОВОДА В РАДИОВОЛНОВОЙ СИСТЕМЕ УПРАВЛЕНИЯ МАГНИТОЛЕВИТИРУЮЩИМИ ТРАНСПОРТНЫМИ СРЕДСТВАМИ

В.Н.Привалов, С.В.Плаксин

Представлены аппаратное построение и результаты экспериментального исследования сверхвысокочастотного (СВЧ) усилителя типа модулятор-демодулятор с комбинацией автодинного и параметрического преобразования сигналов, исследуемого в качестве компенсатора потерь магистрального волновода радиоволновой системы управления экипажем с магнитным подвесом.

Представлено апаратну побудову і результати експериментального дослідження надвисокочастотного (НВЧ) підсилювача типу модулятор-демодулятор з комбінацією автодинного і параметричного перетворення сигналів, досліджуваного як компенсатор втрат магистрального хвилеводу радіохвильової системи керування екіпажем з магнітним підвісом.

Presented apparatus-building and results of the experimental study super-high-frequency (SHF) amplifier of type a modulator-demodulator with the combination autodin and parametric transformation of signals, under investigation as a loss compensator main waveguide radiowave managerial system by the crew with the magnetic carriage.

При распространении электромагнитной волны по волноводу любого типа неизбежны потери волны как за счет потерь на затухание в стенках волновода, так и за счет потерь преобразования типов волн и излучения части энергии из продольных щелей в маркерно-связном магистральном волноводе. Естественно, что с увеличением длины транспортной линии эти потери растут и необходима компенсация потерь. Величину потерь нетрудно оценить расчетным путем, а также с большой точностью измерить инструментально. Это позволяет предъявить требования к параметрам трактовых СВЧ-усилителей, определить периодичность их размещения вдоль магистрального маркерно-связного волновода радиоволновой системы управления магнитолевитирующими транспортными средствами.

Естественно, что разработки перспективных видов транспорта должны базироваться на современной радиоэлектронной базе и учитывать тенденции развития соответствующих радиоэлектронных компонентов.

Функциональные устройства на основе объемных полупроводников с отрицательным дифференциальным сопротивлением [1,2], удовлетворяют этому условию и в качестве трактового усилителя для маркерно-связного волновода был выбран сверхвысокочастотный (СВЧ) усилитель типа модулятор-демодулятор на диоде Ганна.

Суть этого технического решения видна из рисунка 1. Полупроводниковая диодная структура 1 в керамическом корпусе размещается в сверхвысокочастотном резонаторе 2, образованном отрезком волновода с короткозамкнутым

поршнем 3. Напряжение питания от источника 4 подается на диод через развязывающий дроссель 5 и фильтр нижних частот 6. Входной СВЧ сигнал частотой ω_c , моделируемый источником 7, через суммирующее устройство 8 и циркулятор 9 поступает на диодную структуру 1, которая генерирует сигнал на частоте резонатора ω_r . За счет нелинейности полупроводниковой диодной структуры на контуре промежуточной частоты 10 развивается сигнал разностной частоты $\omega_{пч} = \omega_c - \omega_r$ (автодинный режим), который поступает на нагрузку 11 и может являться информативным, например, свидетельствовать об исправности участка магистрального волновода и источника сигнала, запрашивающего маркерно-связной волновод. Часть сигнала промежуточной частоты через направленный ответвитель 12 поступает на сигнальный вход параметрического преобразователя 13, ко второму входу которого через фазовращатель 14 подключено второе плечо ферритового циркулятора. Выходное плечо параметрического повышающего преобразователя (рис.2) через систему расфилтровки 15 подсоединено к суммирующему устройству, подключенному к первому плечу циркулятора. Нетрудно показать, что при подаче на параметрический повышающий преобразователь, емкость варикапа в котором под воздействием сигнала ω_r изменяется по закону $C(t) = C_1 \sin \omega_r t$, сигнала промежуточной частоты $U_{пч} = U_1 \cos(\omega_{пч} + \varphi)$, напряжение на комбинационной частоте $\omega_c = \omega_r + \omega_{пч}$,

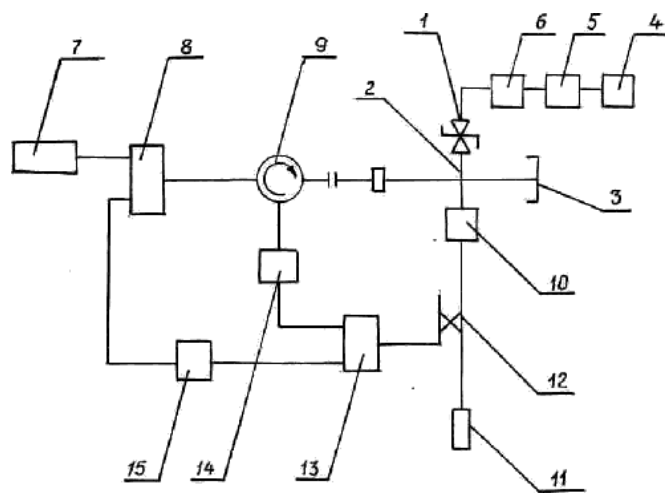


Рисунок 1 - Трактовый усилитель типа модулятор-демодулятор

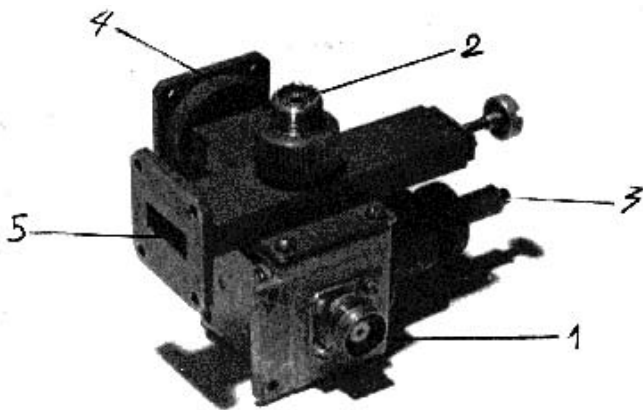


Рисунок 2 - Общий вид повышающего параметрического преобразователя: 1 - сигнальный вход ($\omega_{\text{пч}}$); 2 - ввод напряжения смещения варикапа (1А 401 Б); 3 - орган настройки контура промежуточной частоты ($\omega_{\text{пч}}$); 4 - вход сигнала накачки ($\omega_{\text{г}}$); 5 - вход преобразованного сигнала ($\omega_{\text{с}}$)

$$U_2 = \frac{1}{2} \omega_{\text{с}} U_1 R_{\text{н}} \cos(\omega_{\text{с}} t + \varphi) \quad (1)$$

(здесь $R_{\text{н}}$ - сопротивление нагрузки для частоты $\omega_{\text{с}}$, C_1 - амплитуда изменения емкости) находится в фазе с входным сигналом $\omega_{\text{пч}}$, а добавочное напряжение в сигнальном контуре

$$U_3 = -\frac{1}{4} C_{\text{в1}}^2 U_{\text{пч}} R_{\text{н}} \omega_{\text{с}} \omega_{\text{пч}} R_{\text{пч}} \cos(\omega_{\text{пч}} t + \varphi) \quad (2)$$

находится точно в противофазе с входным сигналом. Регулировкой фазовращателя 15 обеспечивается максимальное значение коэффициента усиления. Соотношения Мэнли - Роу для данного случая могут быть представлены в виде:

$$\frac{P_{\text{г}}}{\omega_{\text{г}}} + \frac{P_{\text{с}}}{\omega_{\text{с}}} = 0, \quad (3)$$

$$\frac{P_{\text{пч}}}{\omega_{\text{пч}}} + \frac{P_{\text{с}}}{\omega_{\text{с}}} = 0, \quad (4)$$

где $P_{\text{пч}}$, $P_{\text{г}}$, $P_{\text{с}}$ - мощности на частоте сигнала $\omega_{\text{пч}}$, частоте накачки $\omega_{\text{г}}$ и преобразованной частоте $\omega_{\text{с}}$, соответственно. При положительном значении $P_{\text{г}}$ уравнение (3) удовлетворяется при отрицательной мощности $P_{\text{с}}$. Соответственно уравнение (4) удовлетворяется при положительном значении мощности на промежуточной частоте. Это означает, что мощность на ПЧ "потребляется" варикапом параметрического преобразователя, модулируемым сигналом накачки, а на суммирующее устройство посту-

пает мощность на частоте $\omega_{\text{с}}$ входного сигнала, усиленного в $\omega_{\text{г}}/\omega_{\text{пч}}$ раз. Максимальное усиление достигается при соответствующих фазовых соотношениях между сигналами на частотах $\omega_{\text{г}}$ и $\omega_{\text{пч}}$, достигаемых регулировкой фазовращателя 15. Выбор частот в соотношении $\omega_{\text{с}} > \omega_{\text{г}}$ обеспечивает:

- а) устойчивость функционирования устройства;
- б) высокую чувствительность, так как вносится минимальный уровень шума со стороны параметрического преобразователя (малое "зашумливание" сигнала);
- в) работоспособность устройства при паразитных изменениях (в пределах полосы пропускания параметрического преобразователя), так как уходы частоты $\omega_{\text{г}}$ повлекут за собой такие же уходы $\omega_{\text{пч}}$, и в результате повышающего преобразования в параметрическом устройстве значение частоты $\omega_{\text{с}}$ останется неизменным. При ином соотношении $\omega_{\text{с}}$ и $\omega_{\text{г}}$ ($\omega_{\text{с}} < \omega_{\text{г}}$) п.п "а" - "в" не выполняются.

Суммирующее устройство может быть реализовано на гибридных кольцах с соответствующей фазовой балансировкой плеч, на квадратурных или щелевых мостах.

Так как предельная чувствительность рассмотренного устройства лишь на несколько децибел ниже чувствительности супергетеродинных преобразователей с внешним гетеродином (это свойство всех генерирующих (автодинных) преобразователей), а мощность, генерируемая полупроводниковой диодной структурой (диодом Ганна), составляет десятки и сотни милливольт, то при взаимодействии в параметрическом преобразователе входного и генерируемого сигналов получают значительные коэффициенты усиления. Следовательно, может существенно повыситься эффективность генерирующих преобразователей, в традиционных построениях которых подавляющая часть генерируемой МЭП-диодами мощности рассеивается в ферритовых вентилях, включаемых в запорном направлении, что создает дополнительные шумы в канале сигнала. Такая схема генерирующего преобразователя выгодно отличается и от режима работы с использованием для последующего усиления многокаскадных усилителей ПЧ.

Разработанный усилитель для компенсации потерь в магистральном маркерно-связном волноводе обладает следующими свойствами:

1. Добавка СВЧ сигнала на входе устройства увеличивает линейный участок коэффициента передачи при сохранении предельной чувствительности (см. рис.3, где приведена передаточная характеристика данного устройства, кривая 1; кривая 2 - для канонического автодинного

режима; $\frac{\omega_{\text{г}}}{2\pi} = 9440$ МГц, $\frac{\omega_{\text{пч}}}{2\pi} = 60$ МГц);

2. При использовании информационного сигнала на промежуточной частоте отпадает необходимость в многокаскадном усилителе ПЧ и сокращаются энергозатраты, связанные с питанием этих усилителей;

3. Уменьшаются тепловые шумы, обусловленные нагревом элементов СВЧ тракта избыточной СВЧ мощностью;

4. Суммарный коэффициент шума также снижается, так как коэффициент шума у параметрических преобразователей существенно меньше, чем у усилителей ПЧ.

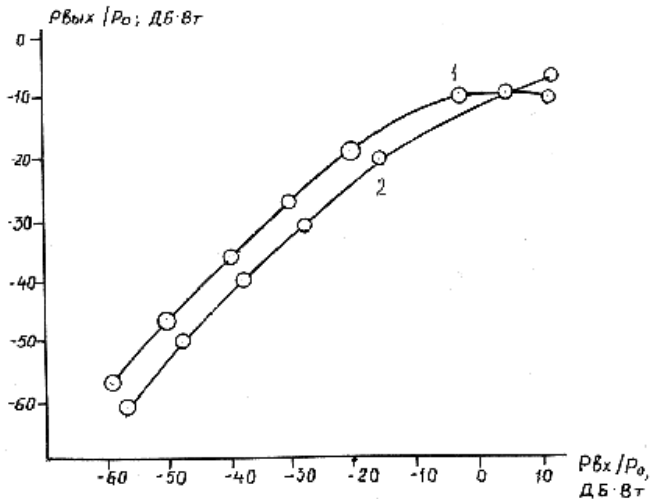


Рисунок 3 - Передаточна характеристика усилителя

Из передаточной характеристики также видно, что в данном усилителе увеличивается на 5-6 дБ коэффициент передачи и повышаются линейность передаточной характеристики на участке малых уровней входных сигналов и предельная чувствительность устройства по сравнению с автодинным преобразователем, построенным по канонической схеме [3] (кривая 2). Столь существенное улучшение показателей разработанного усилителя как компенсатора потерь СВЧ энергии в магистральном маркерно-связном волноводе с учетом высокой энергопрочности позволяет считать, что указанный усилитель может быть рассмотрен в качестве рабочего варианта при выборе окончательного варианта трактового усилителя.

ПЕРЕЧЕНЬ ССЫЛОК

1. Костылев С. А., Гончаров В. В., Соколовский И. И., Челябин А. В. Полупроводники с объемной отрицательной проводимостью в СВЧ-полях. Киев: Наукова думка, 1987.- 141 с.
2. Афромеєв В.И., Привалов В.Н., Яшин А.А. Согласующие устройства гибридных и полупроводниковых интегральных СВЧ-схем. - Киев: Наукова думка, 1989.- 192 с.
3. Lazarus M.I., Bullimore E.D., Novak S. A sensitive millimeter wave self-mixer and a load varian detector // Proc.IEEE.- 1971.- 59, №5-P.812-813

УДК 621.314.6

ВИЗНАЧЕННЯ ЧАСТОТИ КОМУТАЦІЇ ІМПУЛЬСНОГО ПЕРЕТВОРЮВАЧА

В.Д.Флора

Виходячи з виразів для розмахів пульсацій струмів та напруг у різних вузлах найбільш вживаних у практиці схем імпульсних перетворювачів, визначено спосіб розрахунку необхідної частоти комутації.

Исходя из выражений для размахов пульсацій токов и напряжений в различных узлах наиболее употребительных в практике импульсных преобразователей, определен метод расчета необходимой частоты коммутации.

Proceeding from expressions for sweeps of currents and voltage pulsations in various units which are the most common in practice of pulse converters, the method of account of necessary commutation frequency is determined.

Проектуючи імпульсний перетворювач, конструктор вирішує низку питань перед вибором або розрахунком елементів як силової схеми, так й системи керування. Перед розрахунком електромагнітних процесів конструктор повинен вибрати комутатор (тиристорний, або транзисторний, багатофазний чи однофазний, однополярний або двополярний, з широтним, частотним або широтно-частотним регулюванням і т.д.).

Крім того йому необхідно визначити частоту комутації з урахуванням потрібного діапазону регулювання напруги. Визначаючи частоту комутації, можливо виходити з допустимої пульсації певних величин для конкретного

перетворювача.

Для однополярного перетворювача розмах пульсації струмів ΔI_H навантаження [1]:

$$\Delta I_H = \frac{U\alpha(1-\alpha)}{L_H f}, \quad (1)$$

де U - напруга джерела; $\alpha = T_{HK}/T$ - відносна тривалість часу T_{HK} накопичення енергії стосовно періоду комутації T ; $f = 1/T$ - частота комутації; L_H - індуктивність ланцюга навантаження.

Для двополярного імпульсного перетворювача з живленням від одного джерела [2]:

$$\Delta I_H = \frac{2U\alpha_1(1-\alpha_1)}{L_H f}, \quad (2)$$

де $\alpha_1 = T_1/T$ - відносна тривалість T_1 позитивного імпульсу напруги стосовно періоду комутації T .

Для підвищуючого напругу перетворювача розмах пульсації струму накопичувального дроселя [3]:

$$\Delta I_{\text{нк}} = \frac{\left\{ U - I_{\text{нсп}} \left[1 + \frac{\alpha^2}{(1 - \alpha^2)} \right] (R_{\text{дж}} + R_{\text{нк}}) \right\} \alpha}{(L_{\text{дж}} + L_{\text{нк}})f}, \quad (3)$$

де $R_{\text{дж}}$, $L_{\text{дж}}$, $R_{\text{нк}}$, $L_{\text{нк}}$ - відповідно опори та індуктивності джерела та накопичу вального дроселя; $I_{\text{нсп}}$ - середній струм навантаження.

При цьому розмах пульсацій напруги накопичувального конденсатора [3]:

$$\Delta U_C = \frac{I_{\text{нсп}} \left(1 + \frac{I_{\text{нсп}} R_{\text{н}}}{U_{\text{Cср}}} \right)}{Cf}, \quad (4)$$

де $R_{\text{н}}$ - опір навантаження; $U_{\text{Cср}}$ - середня напруга накопичувального конденсатора ємністю C .

Розмах пульсацій струму навантаження [3]:

$$\Delta I_{\text{н}} = \frac{U\alpha}{L_{\text{н}}f(1 - \alpha)}, \quad (5)$$

де $L_{\text{н}}$ - індуктивність ланцюга навантаження.

Для перетворювача з інверсією полярності вихідної напруги розмах пульсації струму накопичувального дроселя [4]:

$$\Delta I_{\text{нк}} = \frac{[U - I_{\text{нсп}}(\alpha^2 R_{\text{дж}} + R_{\text{нк}})]\alpha}{(L_{\text{дж}} + L_{\text{нк}})f}, \quad (6)$$

де $I_{\text{нсп}}$ - середній струм накопичувального дроселя.

Аналогічно можливо прослідкувати й інші одержані вирази для розмахів пульсацій струмів, напруг, швидкостей обертання і т. і. для різних схем імпульсних перетворювачів.

В загальному вигляді всі ці величини можуть бути подані виразом [5]:

$$\Delta X_i = \frac{[X_{\text{јсрнк}} - X_{\text{іср}} R_{\text{н}} \Psi_{\text{нк}}(\alpha, R_m, R_{\text{н}})]\alpha}{fK_{m\text{нк}}}, \quad (7)$$

де ΔX_i - розмах пульсації енергетичного параметра X_i ; $X_{\text{јсрнк}}$, $X_{\text{іср}}$ - середні за період значення енергетичних параметрів; R_m , $R_{\text{н}}$ - дисипативні параметри схеми; $\Psi_{\text{нк}}$ - функція, яка характеризує співвідношення між струмами та опором гілок схеми перетворення під час накопичення енергії; $K_{m\text{нк}}$ - функціональний параметр m -го накопичувального параметра схеми під час накопичення енергії.

Зі співвідношень (1)-(7) можливо записати:

$$\Delta X_i = \frac{X_j \varphi(\alpha)}{K_i f},$$

звідки частота комутації:

$$f = \frac{X_j \varphi(\alpha)}{K_i \Delta X_{i\text{max}}}, \quad (8)$$

де α_e - значення α , при котрому функція $\varphi(\alpha)$ досягає максимуму, або прийнятого значення.

Для перетворювачів однополярних зі зниженням напруги, двополярних, регулювання активного опору R , Γ -подібного фільтру

$$\varphi(\alpha) = \alpha(1 - \alpha). \quad (9)$$

Досліджуючи (9) на максимум, визначаємо $\alpha_e = 0,5$.

В перетворювачі з підвищенням напруги для дроселя-накопичувача, накопичувального конденсатора

$$\varphi(\alpha) = \alpha. \quad (10)$$

А для навантаження

$$\varphi(\alpha) = \frac{\alpha}{1 - \alpha}. \quad (11)$$

Тому для навантаження, якщо $\alpha \rightarrow 1$, то $\varphi(\alpha) \rightarrow \infty$. В такому разі для цього перетворювача доцільно прийняти таке α_e , при якому буде задовільним ККД.

В перетворювачі для імпульсного електричного гальмування з рекуперацією $\varphi(\alpha)$ є складною функцією [6]. Але частота комутації повинна бути $f > f_{\text{гр}}$, де $f_{\text{гр}}$ - гранична частота комутації, при котрій тече безперервний струм у ланцюзі якоря машини. Ця частота визначається параметрами схеми та величиною α [6].

Найчастіше застосовуються однополярні перетворювачі для зниження напруги. В цьому разі, визначаючи частоту комутації, можливо виходити зі співвідношень (1) та (8).

В залежності від різновиду навантаження імпульсного перетворювача (якірна обмотка двигуна, лампа розжарення, нагрівач електричної печі і т.д.) визначається максимально допустимий розмах пульсацій струму навантаження ΔI_{max} . При цьому виходять з умов комутації двигуна постійного струму, допустимих втрат потужності, температури нагріву і т.д. Крім того за умовами регулювання повинні бути визначені або задані мінімальна $U_{\text{срmin}}$ та максимальна $U_{\text{срmax}}$ середні напруги навантаження. Якщо комутатор виконується на транзисторах, то, розраховуючи частоту комутації, необхідно врахувати найбільшу допустиму частоту f_{maxVT} , з якою може тривалий час працювати застосований транзистор.

Враховуючи те, що

$$\alpha_{\min} = \frac{U_{\text{срmin}}}{U} = T_{\text{нк}} f_{\min}; \quad \alpha_{\max} = \frac{U_{\text{срmax}}}{U} = T_{\text{нк}} f_{\max},$$

та співвідношення $T_{\text{нк}} = T_{\text{minVT}} = \frac{1}{f_{\text{maxVT}}}$, одержимо

$$f_{\min} = \alpha_{\min} f_{\text{maxVT}}; \quad f_{\max} = \alpha_{\max} f_{\text{maxVT}}.$$

В такому разі частота комутації може змінюватись в межах

$$f_{\min} < f < f_{\max}. \quad (12)$$

Якщо причиною застосування багатофазного комутатора є необхідність зменшення маси та габаритів вхідного фільтра, то розрахунком низки варіантів при різних частотах комутації визначають прийнятні масогабаритні показники та параметри фільтра. При цьому повинна бути встановлена відповідна частота комутації f . Враховуючи максимальну частоту f_m , з якою може працювати в схемі комутатора напівпровідниковий прилад, підраховують необхідну кількість фаз:

$$m = \frac{f}{f_m}. \quad (13)$$

Якщо ж причиною застосування багатофазного комутатора є необхідність керування великими струмами, які перевищують можливості одного напівпровідникового приладу, то кількість фаз визначається, як

$$m = \frac{I_{\text{нб}}}{I_m}, \quad (14)$$

де $I_{\text{нб}}$ - найбільший можливий струм навантаження; I_m - найбільший струм, яким може керувати один напівпровідниковий прилад. Потім розраховують частоти f та f_m . Частота комутації f розраховується за вимогами оптимізаційних розрахунків вхідного фільтра [7], а f_m - за формулою (13):

$$f_m = \frac{f}{m}.$$

Таким чином, користуючись наведеними співвідношеннями, можливо розрахувати потрібну частоту комутації для найбільш звжваних у практиці схем імпульсних перетворювачів.

ПЕРЕЛІК ПОСИЛАНЬ

1. Флора В.Д. Дослідження режиму зниження напруги однополярного імпульсного регулятора// Радіоелектроніка, інформатика, управління. - 1999. - №1. - С. 40-41.
2. Флора В.Д. Дослідження двополярних імпульсних регуляторів// Радіоелектроніка, інформатика, управління. - 1999. - №2. - С. 48-50.
3. Флора В.Д. Врахування параметрів елементів схеми в режимі підвищення напруги імпульсного регулятора// Радіоелектроніка, інформатика, управління. - 2000. - №1. - С. 38-41.
4. Флора В.Д. Аналіз регульовальних властивостей імпульсного перетворювача з інверсією полярності вихідної напруги.// Збірник наук. праць Львівського державного технічного університету "Львівська політехніка". - 2001. - №2. - С.139-141.
5. Флора В.Д. Метод рівності розмахів пульсацій періодичнозмінних величин в інтервалах часу накопичення та витрати енергії//Електротехніка та електроенергетика. - 2000. - №2. - С. 45-46.
6. Флора В.Д. Режимы імпульсного електричного гальмування електротранспорту// Технічна електродинаміка, тематичний випуск. Силова електроніка та енергоефективність. - 2001. - Ч.2. - С. 29-32.
7. Флора В.Д. Розрахунок Г-подібного фільтра на вході імпульсного регулятора// Електричний журнал. - 1995. - №1. - С. 48-51.

ІНФОРМАТИКА

УДК 681.32:007.58

THE INFLUENCE OF SELF-CONNECTION ON THE PERFORMANCE OF PSEUDO-INVERSE AUTOASSOCIATIVE NETWORKS

Dmitry O.Gorodnichy

The performance of pseudo-inverse autoassociative neural networks proposed by Personnaz et al. [1] and Kanter and Sompolskiy [2] is known to be affected by the value of self-connection of the network. This paper summarizes the results obtained on the phenomenon and describes the desaturation procedure based on partial reduction of self-connection, which is shown to improve the performance of the network. The optimal value of the desaturation coefficient, which is the coefficient self-connection reduction, is obtained.

Как известно, на эффективность псевдо-инверсных автоассоциативных нейронных сетей, предложенных Персоназом, Кантером и Самполским, воздействует значение самосвязанности сети. В настоящей работе обобщаются результаты, описывающие процедуру разнасыщения, основанную на частичном уменьшении самосвязанности, которое улучшает свойства сети. Получено оптимальное значение коэффициента разнасыщения, являющегося коэффициентом уменьшения самосвязанности.

Як відомо, на ефективність псевдо-інверсних автоасоціативних нейронних мереж, запропонованих Персоназом та ін. [1] та Кантером і Самполським [2], впливає значення самосполучення мережі. У роботі наведені результати, що описують процедуру рознасичення, яка ґрунтується на частковому зменшенні самосполучення, яке, як показано, уточнює ефективність мережі. Отримане оптимальне значення коефіцієнта рознасичення, що є коефіцієнтом зменшення самосполучення.

I. INTRODUCTION

An autoassociative memory is a system which, given a noisy version of a prototype (a pattern stored in the memory) as input, retrieves the original prototype as output (Figure 1a).

The idea of using a self-organizing fully connected network of binary neurons for designing an autoassociative memory is mainly attributed to Amari. As early as 1972 he showed [3] that such a network "remembers" some vectors as stable equilibrium states of the network, where the retrieval process consists in converging from an arbitrary state to a stable state. The main questions concerning the network are:

Q1: How to adjust or to calculate the parameters of the network, which are N^2 coefficients of the weight matrix C (N being the number of neurons), so that the desired prototypes $\vec{V}^1, \dots, \vec{V}^M$ become stable states of the network?

Q2: How many prototypes M can be stored by a network of size N and how good is the retrieval capability of the network?

While the answer to the first question gives a learning rule, the answer to second one describes the efficiency of the rule. By 1977 mainly two learning rules had been considered [4], [5]. The first one is the Hebbian, correlation, or outer product rule. The second rule is closely related to Kohonen's generalized inverse approach [6], [7], which has been intensively used for linear associative memories since 1974. It is called the orthogonal, projection, or pseudoinverse learning rule and is the subject of this paper.

This paper is organized with the above two questions in mind. The formal description of autoassociative neural networks and their properties follows in the next section. In Section III we focus our attention on pseudo-inverse neural network (PINNs), which are neural networks designed with the pseudo-inverse learning rule. In Section IV we theoretically study the effect of self-connection reduction on the performance of PINNs, and Section V presents experimental data which show how, by appropriately choosing the coefficient of self-connection reduction, the retrieval capability of PINNs can be considerably enhanced. Conclusions resume the paper.

II. AUTOASSOCIATIVE NEURAL NETWORKS

An autoassociative neural network (AANN) is defined as a self-organizing network of N mutually interconnected two-state neurons, the evolution of which is determined by a synchronous update rule¹:

$$y_i(t+1) = \text{sign}[s_i(t)] = \begin{cases} 1, & \text{if } s_i(t) > 0, \\ -1, & \text{otherwise,} \end{cases} \quad (1)$$

where $s_i(t) = \sum_{j=1}^N C_{ji}y_j(t)$ or in vector form:

$$\vec{Y}(t+1) = \text{sign}[\vec{S}(t)], \quad \vec{S}(t) = C\vec{Y}(t). \quad (2)$$

1. A more general definition includes a threshold and allows asynchronous dynamics. These are not considered in this paper

Column vector $\vec{Y}(t)$ is referred to as the state of the network at time t and C is an $N \times N$ weight matrix (Figure 1).

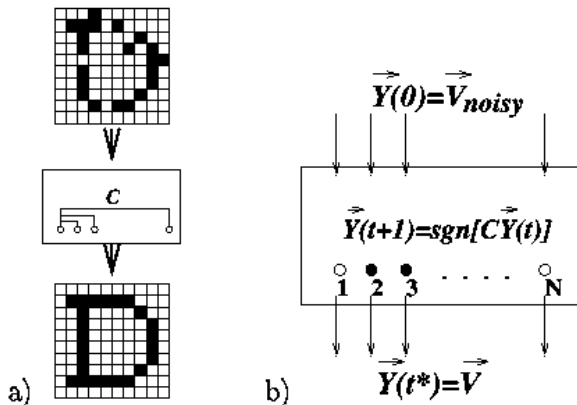


Figure 1 - Retrieval of a pattern \vec{V} by an autoassociative network from an initial noisy pattern \vec{V}_{noisy} . The weights of the network determine the quality of retrieval

The self-organizing feature of the network is attributed to its ability to converge from an arbitrary state to a stable state, which is called an attractor. But how can we be sure that the network converges? What is the nature of these attractors? The following proposition gives the answer.

Proposition 1: As a result of free evolution, an AANN with symmetric weights converges to either a single stable state, called a static attractor, or a cycle consisting of exactly two states, called a dynamic attractor.

Proof: The easiest way to show this [8] is to use the "energy" function proposed by Reznik [9]:

$$E_R(\vec{Y}(t)) = -\frac{1}{2}\vec{Y}^T(t)\vec{S}(t-1) = -\frac{1}{2}\sum_{i=1}^N |s_i(t-1)|. \quad (3)$$

As can be seen, if $C = C^T$

$$\begin{aligned} \vec{Y}^T(t)\vec{S}(t-1) &= \vec{Y}^T(t)C\vec{Y}(t-1) = \\ &= \vec{Y}^T(t-1)C\vec{Y}(t) = \vec{Y}^T(t-1)\vec{S}(t) = \\ &= \vec{Y}^T(t+1)\vec{S}(t) - [\vec{Y}^T(t+1) - \vec{Y}^T(t-1)]\vec{S}(t). \end{aligned} \quad (4)$$

and

$$E_R(\vec{Y}(t)) = E_R(\vec{Y}(t+1)) + \sum_{h=1}^{H(t+1, t-1)} |s_h(t)|, \quad (5)$$

where in the last summation h ranges over the neurons for which $y_h(t-1) \neq y_h(t+1)$, and $H(t+1, t-1)$ is the number of such neurons.

Eq. 5 shows that the "energy" function monotonically decreases with t until the second term of the equation

becomes zero, which happens when $\vec{Y}(t-1) = \vec{Y}(t+1)$. That is, the network is guaranteed to converge to either a stable state: $\vec{Y}(t-1) = \vec{Y}(t) = \vec{Y}(t+1)$, or a cycle: $\vec{Y}(t-1) \neq \vec{Y}(t)$.

This proof can also be used for neural networks with thresholds (see [8]).

Note, that in the above proof we do not use the conventional Hopfield energy [10], which we will use later, defined as

$$E(\vec{Y}(t)) = -\frac{1}{2}\vec{Y}^T(t)\vec{S}(t) = -\frac{1}{2}\vec{Y}^T(t)C\vec{Y}(t), \quad (6)$$

because it has been shown in [11] that in the case of synchronous dynamics this energy is not always monotonically decreasing.

The fact that an AANN may have cycles of length 2 is well known [12]. The above proof however differs from the previously obtained one [13], which is based on the deterministic nature of the evolution process.

A. The update flow technique

Proposition 1 gives us an easy way of identifying dynamic attractors - the indices of the updated neurons should be stored until the next update. If at the next update they are the same, the network is trapped in a dynamic attractor.

In [14] the update flow technique for updating the state of an AANN was introduced. It is based on storing only those neurons which have been updated in the last iteration, instead of storing the whole state vector. This technique is proved to be preferable on account of its high evaluation speed and suitability for parallel implementation. This technique also has another advantage in that it does not require extra processing to check for dynamic attractors.

In Section V we use this technique in simulations.

III. PSEUDO-INVERSE NEURAL NETWORKS

The pseudo-inverse learning rule (PILR) is obtained from the condition that all prototypes are stable states of the network:

$$C\vec{V}^m = \lambda_m \vec{V}^m, \quad (m = 1, \dots, M), \quad \lambda_m > 0, \quad (7)$$

which can be rewritten in matrix form for $\lambda_m = 1$ as

$$CV = V, \quad (8)$$

where V is the matrix made of column prototype vectors.

Resolving this matrix equation for the case of linearly independent prototypes ($M \leq N$) gives us the rule:

$$C = V(V^T V)^{-1} V^T = VV^+, \quad (9)$$

where $V^+ = (V^T V)^{-1} V^T$ is the pseudoinverse of matrix V ,

hence the name of the rule.

The matrix defined by Eq. 9 is the orthogonal projection matrix in the subspace spanned by the prototype vectors $\{\vec{V}^m\}$, which explains the other names of the rule: orthogonal and projection.

Rewriting the PILR (Eq. 9) in scalar form gives

$$C_{ij} = \frac{1}{N} \sum_{m_1, m_2} V_i^{m_1} (C')_{m_1 m_2}^{-1} V_j^{m_2}, \quad (10)$$

$$\text{where } C'_{m_1 m_2} = \frac{1}{N} \sum_{i=1}^N V_i^{m_1} V_i^{m_2}. \quad (11)$$

This formula is not convenient for implementation because of the matrix inversion involved. Therefore, in simulations the iterative formula, which can be obtained from the Greville formula [15], is used¹:

$$C^0 = 0, \\ C^m = C^{m-1} + \frac{(\vec{V}^m - C^{m-1} \vec{V}^m)(\vec{V}^m - C^{m-1} \vec{V}^m)^T}{\|\vec{V}^m - C^{m-1} \vec{V}^m\|^2}, \quad (12) \\ \text{if } \vec{V}^m \neq C^{m-1} \vec{V}^m,$$

or in scalar form

$$C_{ij}^0 = 0, \\ C_{ij}^m = C_{ij}^{m-1} + \frac{(v_i^m - s_i^m)(v_j^m - s_j^m)}{E^2}, \quad (13) \\ \text{where } E^2 = N - \sum_{i=1}^N v_i^m s_i^m, \quad s_k^m = \sum_{i=1}^N C_{ik}^{m-1} v_i^m.$$

If $\vec{V}^m \neq C^{m-1} \vec{V}^m$, which means that the prototype \vec{V}^m is a linear combination of other already stored prototypes, then the weight matrix remains unchanged.

A. Weights of the network

Assume now that the weights have been calculated according to the PILR (Eq. 9). Is it possible to say how many prototypes have been stored by just examining the weights? The following result obtained in [16], [17] gives the answer.

The weights of a FINN satisfy the following conditions:

$$\langle C_{ii} \rangle = \frac{M}{N}, \quad 0 \leq C_{ii} \leq 1, \quad (14)$$

$$\langle C_{ij}^2 \rangle = \frac{M(N-M)}{N^3}, \quad i \neq j, \quad (15)$$

$$\lim_{M \rightarrow N} |C_{ij}| = \langle |C_{ij}| \rangle \quad (16)$$

where $\langle C_{ij} \rangle$ designates the average (arithmetical mean) of values C_{ij} , and $|C_{ij}|$ is the absolute value of C_{ij} . The relationship between the weights and the number of prototypes M as given by the above equations is illustrated in Figure 2,² weight values being approximated as

$$C_{ii} \approx \langle C_{ii} \rangle \text{ and } |C_{ij}| \approx \sqrt{\langle C_{ij}^2 \rangle}. \quad (17)$$

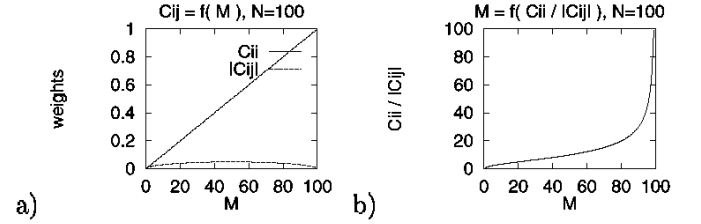


Figure 2 - Relationship between the weights and M .
a) Diagonal C_{ii} and non-diagonal $|C_{ij}|$ weights as functions of M ; b) The ratio $C_{ii}/|C_{ij}|$ indicates the number M of prototypes stored

This result makes PINNs very transparent (as opposed to the "black box" concept). In particular, it can be seen that as M increases, the diagonal weights start dominating the non-diagonal weights. This domination decreases the flexibility of the network, thereby downgrading its associative properties, as will now be shown using the concept of direct attraction radius.

B. Direct attraction radius

It follows from the derivation of the PILR that a PINN can store up to N prototypes as stable states. Yet for an autoassociative memory this is not enough because each prototype should also have a non-zero attraction basin. Therefore, it is important to know the direct attraction radius (DAR) of the network³ which is defined as the largest Hamming distance H_{attr} from within which all vectors are guaranteed to converge to a prototype in a single iteration.

French researchers Personnaz et al. [1] were the first to calculate the DAR of a PINN. They showed that in the case of orthogonal prototypes

$$H_{attr} = \frac{N}{2M}. \quad (18)$$

Another formula for the DAR which relates the retrieval capabilities of the network to the values of its weights has

1. See [6], [9] for the derivation.

2. We use notation $z = f(x, y)$ in the figures to indicate that value z depends on values x and y .

3. The indirect attraction radius will be considered later.

been obtained in [16]:

$$\langle Hattr \rangle = \frac{\frac{1}{2} - \langle C_{ii} \rangle}{\langle |C_{ij}| \rangle}, \quad (19)$$

where $\langle Hattr \rangle$ is the average DAR of a PINN.

Substituting the estimates of weights (Eqs. 14-17) into Eq. 19 gives

$$\langle Hattr \rangle = \frac{\frac{1}{2} - \frac{M}{N}}{\sqrt{\frac{M(N-M)}{N^3}}}. \quad (20)$$

From Eqs. 18 and 20 it follows that a PINN stops exhibiting complete retrieval ($\langle Hattr \rangle < 1$) when $M = N/2$, and Eq. 19 explains this in terms of weights: increasing self-connections reduces the area of attraction. This conclusion will be used in Section IV. Now we outline other properties of PINNs.

C. Energy, cycles, and attractors

It has been shown by Personnaz et al. [1] that the energy of a PINN defined by Eq. 6 is monotonically decreasing during the evolution of the network. This means that, in spite of synchronous dynamics, PINNs do not have dynamic attractors. Rather they evolve until reaching a static attractor corresponding to either a global or a local minimum of the energy in the state space.

Global minima correspond to prototypes and their linear combinations and are given by

$$E(\vec{V}) = -\frac{1}{2} \vec{V}^T C \vec{V} = -\frac{N}{2}. \quad (21)$$

It might be thought that these linear combinations of prototypes introduce a lot of undesirable stable states. However, as shown in [18], [19], this happens very rarely because of the binary nature of neurons. This means that most spurious attractors \vec{Y}^* of the network correspond to local minima of energy, that is:

$$E(\vec{Y}^*) = -\frac{1}{2} \vec{Y}^{*T} C \vec{Y}^* > -\frac{N}{2}. \quad (22)$$

This inequality can be used as a mechanism for detecting spurious attractors. Another mechanism for detecting spurious attractors in PINNs can be found in [20].

D. Implementations of the rule

The bottleneck of the PILR is that it is not local, which is crucial for optical and hardware implementations. Therefore, approximations of the PILR [21], [22], [23], [24] are used, and there exist a number of local iterative methods which converge to the PILR [25], [26], [27], [28], [29]. The most popular of these methods are the Widrow-Hoff rule

[29] and the Gardner formula [28].

Comparisons of the performance of the PILR with that of other learning rules have been done [24], [30], [31], [32], [33] and the PILR remains among the most efficient rules, even if approximations are used.

For optical and hardware implementations of the PILR see [25], [29], [34].

IV. REDUCTION OF SELF-CONNECTION

The influence of self-connection on the performance of AANNs has been studied intensively for the Hebbian rule [35], [36]. In particular, it has been observed that the greater the self-connection, the less sensitive the network and the higher the number of spurious stable states. Kanter and Sompolinsky [2] also indicated that self-connections severely restrict the size of the attraction basins of PINNs. They compared two PINN models: one which contained self-connections and another which did not, and showed that the latter significantly outperforms the former, though the problem with the occurrence of cycles was observed.

The main objective of this and the next sections is to show that rather than being completely removed, selfconnections should only be partially reduced. This increases the attraction radius, improves the retrieval capability of the network, and does not cause too many cycles.

A. Desaturation

As a natural consequence of Eqs. 14, 15 and 19, a modification of the PILR, consisting of partial reduction of self-connections, has been suggested in [16], [17]. More specifically, the modification is described by the equation:

$$C_{ii}^D = D \cdot C_{ii}, \quad (0 < D < 1); \quad C_{ij}^D = C_{ij}, \quad (i \neq j), \quad (23)$$

which is applied after all weights C_{ij} have already been calculated according to the PILR (Eq. 9).

This modification is termed desaturation, and the coefficient of self-connection reduction D is termed the desaturating coefficient. The reasoning behind this is that by applying the modification we restore the balance of weights, thereby allowing a saturated network, i.e. the network which has already lost the retrieval capability as a result of storing too many prototypes, to exhibit retrieval again. In other words, we desaturate the network. The learning rule obtained with the modification (Eq.23) can be rewritten in matrix form as

$$C^D = C - (1 - D)\text{diag}(C_{ii}), \quad 0 < D < 1, \quad (24)$$

where $C = VV^+$ and $\text{diag}(C_{ii}) = \text{diag}(C_{11}, \dots, C_{NN})$ is a diagonal matrix. A network designed with this rule is termed a desaturated network. Before presenting the theory behind desaturation, let us demonstrate how it works on a real problem - the recognition of letters. Fifty English and Ukrainian letters were stored in a PINN, a letter represented as a 10x10 binary pattern. Then desaturation (Eq. 24) was applied to the network (desaturating coefficient D varying

from 0 to 1), and the network was used to retrieve the letters when presented with 9% noise.

Figure 3 shows the retrieval of one letter as observed in the experiments. The overall experimental result was that all letters were retrieved successfully within 3-4 iterations by the network with partially reduced self-connections ($D=0.15$). The performance of the standard network ($D=1$) and of the network with removed self-connections ($D=0$) was poor. In both cases a lot of noise remained. Only one iteration was done in the first case and too many iterations and cycles occurred in the second case. More data from this experiment are given in [16], [17].

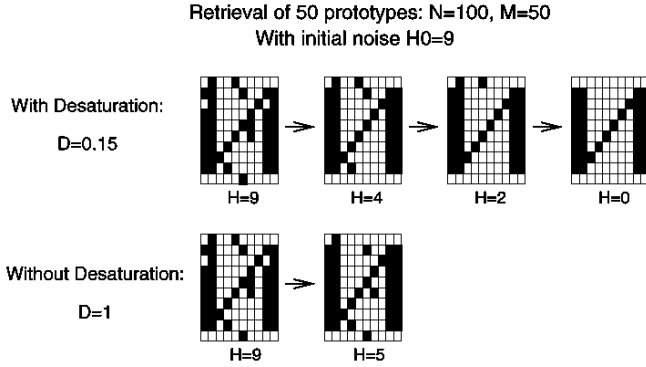


Figure 3 - The retrieval of a letter as observed in the experiments with and without desaturation. Three iterations are executed in the first case, and only one - in the second case

B. Energy considerations

The improvement of the performance can be explained from the "energy" point of view (Section III-C). It can be easily shown using Eq. 14 that changing the diagonal terms does not affect the location of the energy minima in the state space:

$$\begin{aligned}
 E^D(\vec{Y}) &= -\frac{1}{2}\vec{Y}^T C^D \vec{Y} = \\
 &= -\frac{1}{2}\vec{Y}^T C \vec{Y} - \frac{1}{2}\vec{Y}^T (1-D)\text{diag}(C_{ii})\vec{Y} = \\
 &= E(\vec{Y}) - \frac{1}{2}(1-D) \sum_{ii}^N C_{ii} = E(\vec{Y}) - \frac{1}{2}(1-D)M, \quad (25)
 \end{aligned}$$

as $(1-D)M$ is a constant. Yet, changing the weights changes the evolution of the network. In particular, the energy is no longer a monotonically decreasing function of time. This makes the network less prone to getting trapped in local minima and thereby increases the number of iterations. But this may also result in the occurrence of cycles.

Therefore, the changes of the network behaviour incurred by desaturation need to be investigated further.

C. Theory

We want to make sure that the modification does not erase the prototypes from the memory, and does not increase the number of spurious attractors.

Theorem 1: Desaturation preserves prototypes as stable states of the network.

Proof: This follows from Eqs. 24 and 8 and the fact that $D > 0$ and $C_{ii} < 1$ (Eq. 14):

$$\begin{aligned}
 \text{sign}[C^D \vec{V}] &= \text{sign}[C \vec{V} - (1-D)\text{diag}(C_{ii})\vec{V}] = \\
 &= \text{sign}[\vec{V} - (1-D)\text{diag}(C_{ii})\vec{V}] = \vec{V}. \quad (26)
 \end{aligned}$$

Theorem 2: Desaturation decreases the number of spurious static attractors.

Proof: We use the idea of Kanter and Sompolmsky [2].

Consider a state \vec{Y} , which differs from a prototype \vec{V} in only one neuron i ($y_i = -v_i$). Vector \vec{Y} will also be a static attractor if

$$\begin{aligned}
 y_i \left[\sum_{j=1}^N C_{ji}^D y_j \right] &= y_i \left[\sum_{j=1}^N C_{ji} v_j - C_{ii} v_i + C_{ii}^D y_i \right] = \\
 &= v_i [v_i - C_{ii} v_i + D C_{ii} v_i] = v_i [v_i (C_{ii} + D C_{ii} - 1)] > 0, \quad (27)
 \end{aligned}$$

which happens when $C_{ii}(1+D) > 1$.

By decreasing D , we decrease the probability that vector \vec{Y} is a spurious attractor.

These theorems do not yet guarantee that the performance of the network is improved. First we need to know how desaturation affects the DAR of the network (Section III-B), and then we have to find out how many dynamic attractors occur in the desaturated network.

Theorem 3: The average DAR of PINNs increases with the decrease of the desaturating coefficient D as

$$\langle H_{attr} \rangle = \frac{\frac{1}{2}(1-(D+1)\frac{M}{N})}{\sqrt{\frac{M(N-M)}{N^3}}}. \quad (28)$$

Proof: See [37].

The dependency in Eq. 28 is illustrated in Figure 4, which shows the effect of self-connection reduction on the size of the DAR. From the bottom up, the dashed lines correspond to the values $D = 1.0, 0.4$ and 0.2 in Eq. 28. For comparison, we also plot the DAR given by Eq. 18 by a solid line. As can be seen, decreasing self-connections increases the DAR, making it possible for the network to retrieve prototypes even for $M=75$ (as $\langle H_{attr} \rangle > 1$ when $D = 0.2$).

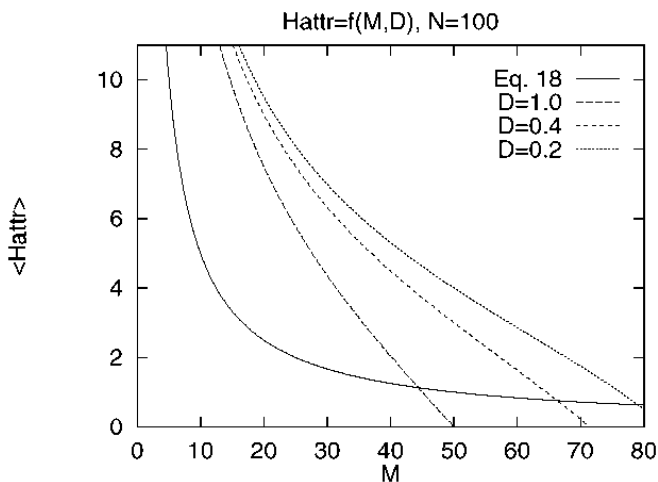


Figure 4 - The direct attraction radius as function of D

This theorem provides the theoretical basis for the observed improvement of the performance of the network caused by the modification, yet it does not explain it completely. This is because the main property of the modification, as mentioned in Section IV-B, is that it allows the network to escape spurious attractors, which increases the indirect attraction radius. However, as theoretical calculation of the indirect attraction radius is known to be hard [38], it can be discovered by simulations only¹. The result of these simulations are presented in the next section, but first we will resolve the "occurrence of cycles".

D. Dynamic attractors

As has already been mentioned, desaturation results in the occurrence of dynamic attractors. These new spurious attractors degenerate the memory. Therefore, it is important to know when they occur and how to avoid them. This is studied in [8], and below we summarize the main results.

Theorem 4. The number of dynamic attractors in a desaturated PINN increases with the number of prototypes M and with the degree of weight reduction $\alpha = 1 - D$.

Proof: The proof from [8] is reproduced in Appendix.

Whether or not the network will be trapped in a dynamic attractor described by Theorem 4, depends on the number of states the network passes through during its evolution, i.e. the number of iterations. The further the initial state of the network is from its final state, the more iterations it takes to reach this state. This explains why the number of cycle occurrences increases not only with M and α , but also with the value of initial noise.

In order to find the area where cycles do not occur, a set of Monte-Carlo simulations has been carried out (see [8] for more details about the setup of the experiment). In Figure 5 the values of D and $H0$ below the curves are those for which cycles are observed with probability no greater than 1%. The

simulations are carried out for a network of size $N = 100$, number of prototypes $M = 20, 40, 60$, and different values of the desaturating coefficient D and initial noise $H0$.

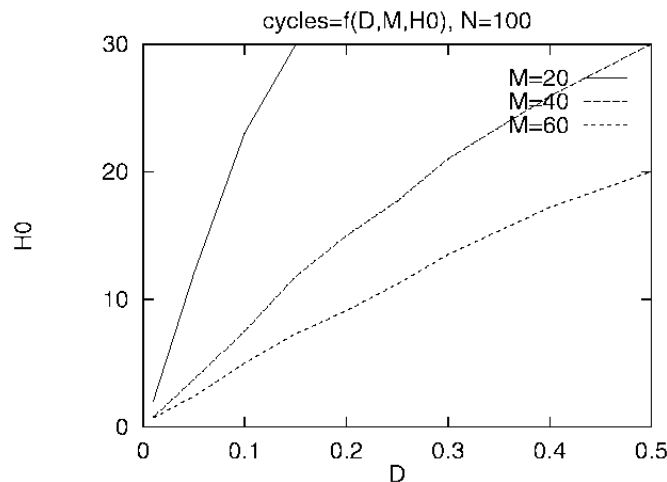


Fig. 5. In the area below the lines the probability of the occurrence of dynamic attractors is not greater than 1%

Simulations also show that when a cycle occurs, there are 2 oscillating neurons in most cases.² This allows us to estimate the value of D that causes the occurrence of such cycles (see Appendix):

$$D < \frac{\langle |C_{ij}| \rangle}{\langle C_{ii} \rangle} \approx \sqrt{\frac{N-M}{NM}}. \tag{29}$$

For example, when $N = 100$ and $M = 0.5N$ a cycle will most likely occur when $D < 0.1$, which is in agreement with the experimental observations. However, as mentioned above, the occurrence of cycles also depends upon the initial noise. For example, from Figure 5 it can be seen that for $M = 0.6N$ and $D = 0.1$ cycles are rarely observed unless the initial noise $H0$ is greater than 4%.

It should also be mentioned that when cycles occur, the performance of the network can be improved by switching from synchronous dynamics to asynchronous dynamics, as suggested in [2].

V. SIMULATIONS

As mentioned before, the increase of the direct attraction radius with desaturation does not fully describe the improvement of the network performance. Knowing how other network parameters are affected by the desaturation is desirable. These parameters are:

1. the error-correction ratio $H/H0$, which is defined as the ratio between the final noise H , measured after a net-

1. Another approach would be to use an approximate treatment for the description of the PINN dynamics as suggested in [39].

2. To be more exact, the number of oscillating neurons varies from 1 to almost N , but the number 2 is the prevailing number especially when $D > 0.05$.

work has converged, and the initial noise H_0 ;

2. the indirect attraction radius (hereafter called simply the attraction radius), which is the largest amount of noise that can be completely eliminated by network iterations;
3. the associative capacity, which is the greatest number of prototypes that can be stored by the network with the attraction radius being greater than zero; and
4. the number of iterations that the network executes before reaching an attractor.

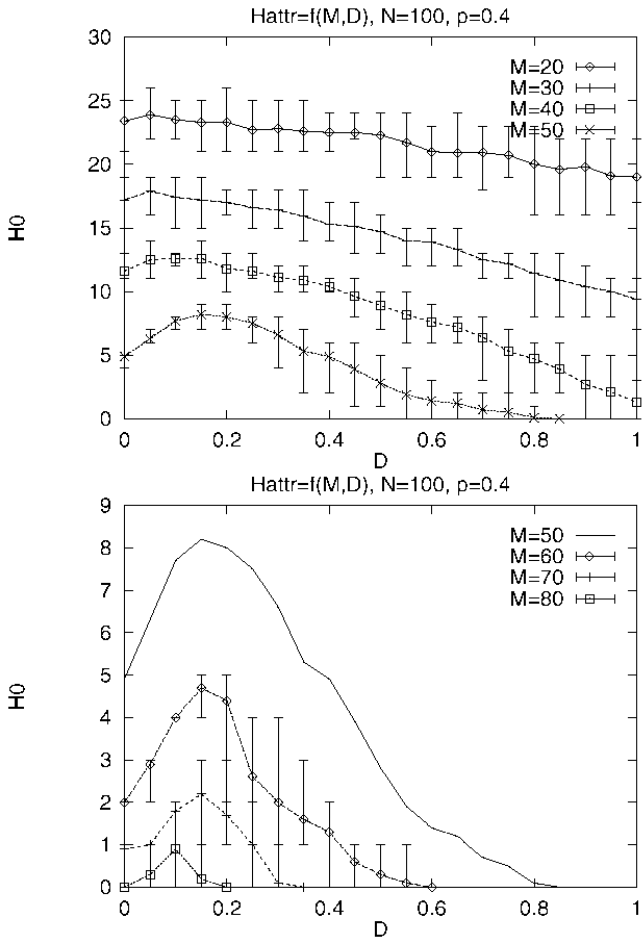


Figure 6 - The increase of the attraction radius as a result of the reduction of self-connection: $C_{ii} = D \cdot C_{ii}$, $p = 0.4$

Whereas the first three parameters describe the improvement of retrieval with desaturation, the last one indicates how much desaturation slows down retrieval. The results concerning the improvement of the error-correction ratio are given in [40], and below we present the results concerning the other parameters.

Extensive Monte-Carlo simulations have been carried out. Prototypes are random vectors, where the probability p of a neuron being in state +1 is the same for all neurons in all

prototypes. The update flow technique described in Section II-A is used¹.

The measurements of the attraction radius are carried out as follows. A network of size $N=100$ is used². We start with $M=20$ prototypes. The coefficient D in Eq. 23 is decreased from 1.0 to 0.0 with an increment 0.05. For each D , the initial noise H_0 is increased until the final noise H is greater than 0.1, as averaged over 10 sets of prototypes and 10 implementations of the initial noise. The initial noise is created by randomly inverting H_0 neurons of a prototype. Then, using Eq. 12, we add 10 new prototypes and repeat the procedure. We record the average value of noise H_0 which can be completely eliminated, its minimal and maximal values, and also the number of occurred iterations.

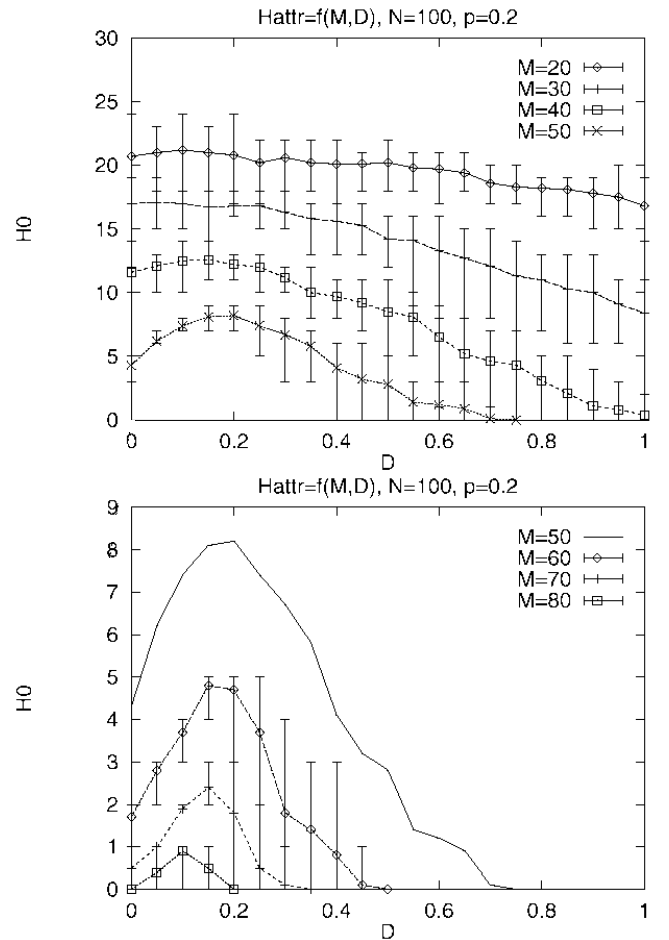


Figure 7 - The increase of the attraction radius as a result of the reduction of self-connection: $C_{ii} = D \cdot C_{ii}$, $p = 0.2$

Figures 6 and 7 plot the results for probabilities $p=0.4$ and $p=0.2$ respectively, where the average values of the attraction radius are drawn as curves and the minimal and

1. The source of the program used in the simulations is available at <ftp://ftp.cs.ualberta.ca/pub/dmitri/PINN>.

2. Data obtained for different N are given in the next subsection.

maximal values of the attraction radius are drawn as vertical bars.

The main results from the simulations in [40] and from the ones described above are as follows.

- While $D \geq 0.15$, a decrease of self-connections always results in a decrease of the error-correction ratio H/H_0 and an increase of the attraction radius of PINNs.

- When the number of stored prototypes is small ($M \leq 0.2N$), the improvement is insignificant. Yet cycles may occur. Therefore, it is not worth using desaturation for small learning ratios M/N .

- The improvement is especially noticeable when $M \geq 0.4N$. More specifically, the desaturation increases the attraction radius of the network from zero (in the case of standard PINNs), to:

- 12.6 (on average) and to 12 (at worst) for $M = 0.4N$,
- 8.2 (on average) and to 7 (at worst) for $M = 0.5N$,
- 4.7 (on average) and to 4 (at worst) for $M = 0.6N$,
- 2.2 (on average) and to 1 (at worst) for $M = 0.7N$,
- 0.9 (on average) for $M = 0.8N$.

- The optimal value for the coefficient of self-connection reduction lies in the interval $0.1 \leq D \leq 0.2$.

- The number of iterations increases insignificantly: from 1 to between 2 and 4.

- Decreasing D below 0.1 is undesirable, as this results in the occurrence of numerous dynamic attractors, thus decreasing the attraction radius. For $D=0$ the performance, although not optimal, is still better than that of the standard model ($D=1$) confirming the result of [2].

- The associative capacity of PINNs is greater than $0.7N$ and is $0.8N$ on average.

A. Scaling a network

In order to use the data presented in the previous subsections, one needs to know how the parameters of the network change with the size of a network, and how they change in the case of non-random prototypes. Here the approach of Kanter and Sompolinsky [2] can be used. In order to compensate for the effect of prototype overlapping, they use another definition for the attraction radius of a network, which can be written for small values of the initial noise as:

$$R = \frac{R_N}{\langle R_{ij} \rangle}, \quad (30)$$

where R_N is the average attraction radius of a network of size N measured by experiments and $\langle R_{ij} \rangle$ is the average distance between the prototypes stored in the network¹.

From this definition it follows that the measured attraction radius R_N increases with the size of the network and decreases when correlated prototypes are stored. This is

exactly what is observed by experiment. Figure 8 shows the observed attraction radius $R_N = Hatr/N$ as a function of the self-connection reduction for different sizes of the network.

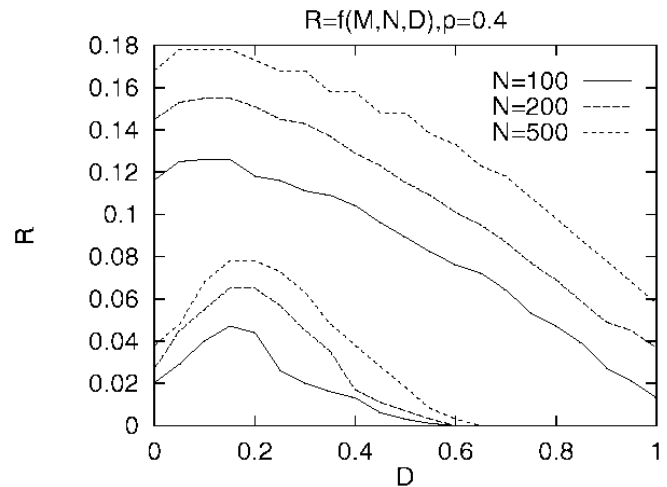


Figure 8 - The increase of the attraction radius $R_N = Hatr/N$ with desaturation for different network sizes N , $M = 0.4N$ (top three curves) and $M = 0.6N$ (bottom three curves)

VI. CONCLUSIONS

We described how reduction of self-connection affects the performance of pseudo-inverse neural networks. We summarized the results obtained on the phenomenon and showed that the best performance of the network is achieved when self-connection is reduced approximately 0.15 times, i.e. when

$$C_{ii} = D \cdot C_{ii}, \text{ where } D = 0.15.$$

In order to efficiently detect cycles, which may occur when self-connection is reduced, the update flow neuro-processing technique described in the paper should be used.

Another contribution of the paper is the desaturation method which, when applied to a pseudo-inverse neural network, restores the balance of network weights, allowing thereby the network to store more prototypes. This method gives rise to developing non-iterative learning techniques for storing continuous information in real-time, which is described in [41].

ACKNOWLEDGEMENTS

The results presented in this paper originally appeared in the author's Ph.D. dissertation [42]. Some of them also appeared in [43].

1. In this section the distance and the attraction radius are scaled according to the size of the network. In particular, $R_N = Hatr/N$.

ADDENDUM

Since this work has been done, other worth mentioning results on pseudo-inverse networks and the influence of self-connection appeared.

Satoshi Matsuda presented results [44] on finding the optimal value of self-connection for Hopfield-like networks, where optimality is considered in terms of finding an optimal solution for an optimization problem the network is applied too.

Giuseppe Grassi et al implemented inverse Greville theorem [45], and also extended the application of pseudoinverse learning rule to heteroassociative and second-order neural networks [46].

Finally, a bulk amount of research on the influence of selfconnection on chaotic neural networks, which is another type of attractor-based networks, and its application to the combinatorial problems has been conducted by Masaya Ohta [47].

APPENDIX

From Proposition 1 it follows that if there exists a dynamic attractor in the network, it consists of exactly two states. Let us calculate the probability that two states of the desaturated network \vec{Y}^{d1} and \vec{Y}^{d2} are a dynamic attractor, i.e. that

$$\begin{cases} \vec{Y}^{d2} = \text{sign}[C^D \vec{Y}^{d1}], \\ \vec{Y}^{d1} = \text{sign}[C^D \vec{Y}^{d2}]. \end{cases} \quad (31)$$

Designating $\Omega = \{i: y_i^{d1} = -y_i^{d2}\}$ as the set of the indices of oscillating neurons in these states, we have

$$\begin{cases} y_i^{d1} \sum_{j=1}^N C_{ij}^D y_j^{d1} < 0, \\ y_i^{d2} \sum_{j=1}^N C_{ij}^D y_j^{d2} < 0, \end{cases} \quad \forall i \in \Omega. \quad (32)$$

Adding one equation to another yields

$$y_i^{d1} \left(\sum_{j=1}^N C_{ij}^D y_j^{d1} - \sum_{j=1}^N C_{ij}^D y_j^{d2} \right) = 2y_i^{d1} \sum_{j \in \{\Omega\}} C_{ij}^D y_j^{d1} < 0. \quad (33)$$

Summing this equation over all oscillating neurons, we have

$$2 \sum_{i \in \Omega} y_i^{d1} \sum_{j \in \Omega} C_{ij}^D y_j^{d1} = 2 \sum_{i, j \in \Omega} C_{ij}^D y_i^{d1} y_j^{d1} < 0 \quad (34)$$

and using Eq. 23, we obtain

$$\sum_{i, j \in \Omega} C_{ij}^D y_i^{d1} y_j^{d1} < (1-D) \sum_{i \in \Omega} C_{ii}. \quad (35)$$

Since for the pseudo-inverse learning rule $C_{ii} \approx \frac{M}{N}$ (Eq. 14),

we obtain that the probability that $\{\vec{Y}^{d1}, \vec{Y}^{d2}\}$ is a dynamic attractor is proportional to M and to $\alpha = 1 - D$, which is the statement of Theorem 4.

Eq. 35 allows us to estimate the value of D , which produces the most number of dynamical attractors. If $\Omega = \{i, j\}$ consists of only two indices, then from Eq. 35 we have

$$2C_{ij} y_i y_j + C_{ii} + C_{jj} < (1-D)(C_{ii} + C_{jj}). \quad (36)$$

From this equation, using the estimates of C_{ij} and C_{ii} (Eqs. 14-17) we obtain

$$D < \frac{\langle |C_{ij}| \rangle}{\langle C_{ii} \rangle} \approx \sqrt{\frac{N-M}{NM}}. \quad (37)$$

REFERENCES

1. L. Personnaz, 1. Glijon and G. Dreyfus. Collective computational properties of neural networks: New learning mechanisms, Phys. Rev. A vol 34, pp. 4217-4228, 1986
2. I. Kanter and H. Sompolinsky. Associative recall of memory with out errors) Phys. Rev. A vol 35, pp. 380-392) 1987.
3. S. Amari. Learning patterns and pattern sequences by self organizing nets of threshold elements, IEEE Transactions on Computers' vol C-219, No.II, pp. 1197-1206, 1971.
4. W.A. Little. The existence of the persistent states in the brain, Mathematical Biosciences, Vol. 19, pp. 101-120, 1974.
5. S. Amari. Neural theory of association and concept formation, Biological Cybernetics, vol 26, pp. 175-185, 1977.
6. T. Kohonen. An adaptive associative memory principle, IEEE Transactions on Computers, C-23, 444-445, 1974.
7. T. Kohonen. Associative memory: A System-Theoretical Approach, Berlin:Springer, 1978.
8. D.O. Gorodnichy and A.M. Reznik. Static and Dynamic Attractors of Auto-associative Neural Networks, Lecture Notes in Computer Science, Vol 1311 (Proc. of 9th Intern. Conf. on Image Analysis and Processing, Florence, Italy, Vol. II), pp. 238-245, Springer, Sept. 1997.
9. A.M. Reznik. Iterative projection learning algorithm for neural networks (in Russian), Kibernetika i Sistemy! Analiz, N 6, pp. 131-141, 1993.
10. J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities, Proc. Nat. Acad. Sci. USA. vol 79, pp. 2554-2558, 1982.
11. K. Cheung, L. Atlas, R. Marks. Synchronous vs asynchronous behavior of Hopfield's CAM neural net, Applied Optics, Vol. 26, No 22, pp. 4808-4813, 1987.
12. P. Peretto. Collective Properties of Neural Networks: A statistical Physics Approach, Biological Cybernetics, 50, 51, 51-62, 1984.
13. A. Frumkin, E. Moses. Physicality of the Little model, Phys. Rev. A vol 34, No 1, pp. 714-716, July 1986.
14. D.O. Gorodnichy, A.M. Reznik. NEUTRAM - A Transputer Based Neural Network Simulator, Proc. of 2nd Intern. Conf. on Software for Multiprocessors and Supercomputers Theory, Practice, Experience (SMS TPE'94), pp. 136-142, Moscow, Russia, 1994.
15. A. Albert. Regression and Moore-Penrose pseudoinverse. Academic New-York, 1972.
16. D.O. Gorodnichy. A way to improve error correction capability of Hopfield associative memory in the case of saturation, HELNET 94-95 International Workshop on Neural Networks Proceedings, vol. 1/11, pp. 198-212, VU University Press, Amsterdam, 1996.
17. D.O. Gorodnichy. Desaturating Coefficient for Projection Learning Rule, Lecture Notes in Computer Science, Vol. 1112 (Intern. Conf. on Artificial Neural Networks Proceedings), pp. 469-476, Bochum, Germany, Springer-Verlag, 1996.
18. A.M. Odiyko. On subspaces spanned by random selection of

- \pm I vectors, J. Combinatorial Theory A, vol. 47, pp. 124-33, 1988.
19. Y. Crama, P.Hansen and B. Jaumard. Detection of Spurious states of Neural networks, IEEE Transactions on NN, Vol. 2, No 1, pp. 165-169, 1991.
 20. J.-D. Gascuel, B.Moobed, M.Weinfeld. An Internal Mechanism for Detecting Parasite Attractors in a Hopfield Network. Neural Computation, 6,5, pp.902-915, 1994.
 21. M. Weinfeld. A fully digital integrated CMOS Hopfield network including the learning algorithm. Proc. of Intern. Workshop on VLSI For Artificial Intelligence , E1-E10, Univ.of Oxford, 1988.
 22. V.S. Dotsenko, N.D. Yarunin and E.A. Dorotheyev. Statistical mechanics of Hopfield-like neural networks with modified interactions, J. Phys. A: Math. Gen 24. pp. 2419-2429, 1991.
 23. A.Engel and M. Weigt. Storage capacity of the truncated projection rule, J. Phys. A: Math. Gen 24, pp. 3707-3709. 1991.
 24. A. Schultz. Five variations of Hopfield Associative memory networks, Journal of Artificial Neural Networks vol. 2, no 3, pp. 285-294, 1995.
 25. B.Telfer and D.Casasent. Updating optical pseudoinverse associative memories. Applied Optics, vol. 28, no. 13. pp. 2518-2528, 1989.
 26. S. Diederich. M. Opper. Learning of correlated patterns in spin-glass networks by local learning rules, Phys. Rev. Lett.-58. N9, 949-952, 1987.
 27. E. Gardner. The space of interactions in neural network models, Phys. Rev. A vol 21, pp. 257-270. 1988.
 28. J. Kratschmar and G. Kohringi Retrieval of neural networks constructed from local and nonlocal learning rules, Journal de Physique. 2, pp. 223-229, Feb. 1990.
 29. A. Johannet, L. Personnaz, G. Dreyfus, J.-D. Gascuel, and M. Weinfeld Specification and Implementation of a Digital Hopfield-Type Associative Memory with On-Chip Training, IEEE Transactions on Neural Networks Vol. 3, N. 3, p. 529, July 1992.
 30. M.H. Hassoun and A.M. Youssef. Autoassociative neural memory capacity and Dynamics. International Joint Conf. on NN (IJCNN'90) Proceedings vol 1, pp. 763-769. San Diego, USA, 1990.
 31. A.N. Michel, J.Si and G. Yen. Analysis and Synthesis of a class of discrete-time neural Networks, IEEE Transactions on NN, Vol. 2, No 1, pp. 29-39, 1991.
 32. S. Coombes and J. Taylor, Using Features for the Storage of Patterns in a Fully Connected Net. Neural Networks 9, pp. 837-844, 1995.
 33. S. Coombes and J. Taylor, The Storage and Stabilisation of Patterns in a Hopfield Net, Neural Network World, 5. (1995) 133-150.
 34. M.A.G. Abyshagur and A.M.Helaly. Neural network training using the bimodal optical computer, Proceedings of SPIE - The International Society for Optical Engineering, vol 1294, pp. 77-83, 1990.
 35. G.R. Gindi, A.F. Gmitro and K. Parthasarathy. Hopfield model associative memory with non-zero diagonal terms in memory matrix. Applied optics 27, pp. 129-134, 1988.
 36. H. Yanai and Sawada Y. Associative memory network composed of neurons with Hysteretic Property, Neural Networks Vol.3, N2, pp. 223-228, 1990.
 37. D.O. Gorodnichy and A.M. Reznik. Increasing Attraction of Pseudo-Inverse Autoassociative Networks, Neural Processing Letters, volume 5, issue 2, pp. 123-127, 1997.
 38. P. Floreen and P. Orponen. Attraction radii in binary Hopfield nets are hard to compute, Neural Computation 5. pp. 812-821, 1993.
 39. R.D. Henkel and M. Opper. Parallel dynamics of the neural network with the pseudoinverse coupling matrix. J, Phys. A: Math, Gen 24, pp. 2201-2218, 1991.
 40. A.M. Reznik, D.O. Gorodnichy. A.S. Sychov. Controlling local self-connection in neural networks designed with the projection learning rule (in Russian). Kibemetika i Sistemnyi Snaliz, N 6, pp 153 - 162. 1996.
 41. A.M. Reznik. Non-Iterative Learning for Neural Networks, CD-ROM Proceedings o IJCNN'99, Washington, July 12-17, 1999.
 42. D.O. Gorodnichy. Investigation and design of high performance fully connected neural networks. PhD dissertation, National Academy of Sciences of Ukraine, Kiev, Ukraine (written in Russian). Available online at ftp://ftp.cs.ualberta.ca/pub/dmitri/PINN 1997.
 43. D.O. Gorodnichy. The Optimal Value of Self-connection ("Best Presentation" award paper). CD-ROM Proceedings of IJCNN'99, Washington, July 12-17, 1999).
 44. S. Matsuda. "Optimal" Hopfield network for Combinatorial Optimization with Linear Cost Function. IEEE Transactions on Neural Network., Vol. 9, No. 6, pp. 1319-1330, 1998.
 45. M. Brucoli, L. Carnimeo, and G. Grassi. Heteroassociative memories via cellular neural networks, Int. Journal of Circuit Theory and Applications, 26, 231-241, 1998.
 46. Ciuseppe Crassi and Giuseppe Acciani Cellular Neural Networks for Information Storage and Retrieval: A New Design Method CD-ROM Proceedings of IJCNN'99, Washington, July 12-17, 1999).
 47. M. Ohta. On the self-feedback controlled chaotic neural network and its application to the N-Queen problem, CD-ROM Proceedings of IJCNN'99, Washington, July 12-17, 1999.

УДК 681.32:007.52

PARALLEL IMPLEMENTATION OF NEURAL NETWORK ALGORITHM USING PARALLEL VIRTUAL MACHINE (PVM) SOFTWARE

Jamil Ahmad

Для повышения скорости работы нейронных сетей предлагается использовать технологию параллельной реализации нейронных сетей. Рассмотрен пример решения задачи распознавания символов нейронной сетью, моделируемой с помощью программного комплекса параллельной виртуальной машины.

Although there are many implementations of artificial neural networks (ANN) are available on sequential machines, but most of these implementations require an immoderate amount of time to train or run ANNs, especially when the ANN models are large. It can be observed that the problem is related to computational power of computer machines. One possible approach for solving this problem could be a parallel implementation of ANNs. Hence, researchers have adopted a number of strategies since the rebirth of ANN in 1986 to implement ANN model in parallel environment. Very few of these strategies are using software platform for such implementations. Therefore, this paper presents a novel technique of implementing ANN for the recognition of characters using Parallel Virtual Machine

(PVM) software package. PVM permits a heterogeneous collection of computers hooked together by a network to be used as a single large parallel computer. Thus large computational problems can be solved more cost effectively by using the aggregate power and memory of many computers. The nodes (neurons) of ANN model are distributed over on the participating computers in the parallel environment to do the needful calculation in parallel. Weights are also adjusted in the same way, if there is any discrepancies between computed and target outputs. Simulation of the study shows that parallel implementation of the ANN produces better results than sequential implementation.

1. INTRODUCTION

The growing importance of artificial neural networks is now widely recognized, it is critical that these models run fast and generate results in real time. As discussed above

that a number of implementations of neural networks are available in sequential form, but most of them are suffered due to its slow speed. One approach for speeding up the implementation of ANN is to implement them on parallel machines.[1] In recent past many attempts have been made to implement ANN models through parallel environment to take full advantage of its structures. There are many ways in which neural networks can be organized to operate in parallel. Some of them are discussed in [2]-[7]. In one particular way, ANN in parallel environment may look like that several different networks, each operating in logical or real parallelism working on the same set of data. Each network will be trained in logical or real parallelism, to make different distinctions. Sometimes, when a set of subset of subtle or complex distinctions are desired, the best solution is to break the problem up into a number of subtasks, each solved by a separate network. A system of the networks can be created. The results of the different networks can be fused or correlated to obtain desired results, i.e., multiple networks are operating in parallel to solve a single problem.[2]

More interesting system design issues arise when we seek to develop systems which can make subtle and complex distinctions from a large body of incoming data.[8] Sometimes, a complex distinction may require that several different types of features need to be extracted from the data, and we may not know in advance what features will be needed. As the network size grows, the representation of feature information can become more distributed, or needed generalizations may fail to be made. Parallel implementation of ANN models provides better solution in such circumstances than sequential implementation. All of attempts, which are made to implement ANN models in parallel environment can be divided into two categories, i.e., software and hardware.[9],[2] However, the ratio of using parallel hardware is very high as compared to software. There is one major problem with the hardware implementation of ANN is the 'portability'. The life of a particular parallel machine is generally only a few years. A user who had developed code for one machine a couple of years ago, often has to rewrite it for another machine if the original machine is no longer available. In the right kind of simulation environment, the users should be able to easily update their implementations for a new machine through the environment.[1]

In this paper we propose a novel method to implement ANN to recognize characters using parallel software platform. PVM software is used to implement basic perceptron model to recognize English characters, which are presented to network in many different fonts. PVM software can be implemented on any kind of machine; the implementation in this environment does not suffer from problems of portability and flexibility (see section 2 for more discussion on PVM).

2. SELECTION OF APPROPRIATE ALGORITHM FOR PVM SOFTWARE

After a through investigations we found that not all of the neural network model can easily be implemented in parallel environment especially those models in which calculations in neurons are interdependent on each other.

Perceptron model is selected to be implemented with PVM software for three reasons; however, other model such as Hopfield and Kohanan etc. can also be implemented with PVM. Interested readers are referred to [1] for further discussion on this issue.

Firstly, Perceptron had perhaps the most far-reaching impact of any of the early neural nets. Secondly, in the Perceptron algorithm the calculations in each neuron of the network don't depend on the calculations in other neurons. Thus, distribution of neurons over different participating computers is easy and possible. Thirdly, under suitable assumptions, it is iterative learning procedure can be proved to converge to the correct weights i.e., the weights that allow the net to produce the correct output value for each of the training input pattern. One of the necessary assumptions is that such weight exists. Details on the perceptron model can be found in [10]-[11]. The algorithm used during simulation is shown in Figure 1.

Step 1:

- 1.1. Initialize all variables
- 1.2. Set weights and bias
- 1.3. Set learning rate α ($0 < \alpha <= 1$)

Step 2:

Determine a stopping condition: if the stopping condition is false, do step 3-7

Step 3: For each training pair $s:t$, do steps 4-6

Step 4: Set activation of input units:

$$x_i = s_i$$

Step 5: Compute response of output unit:

$$y_in = b + \sum_i x_i w_i;$$

$$y = \begin{cases} 1, & \text{if } y_in > \theta \\ 0, & \text{if } -\theta \leq y_in \leq \theta \\ -1, & \text{if } y_in < -\theta \end{cases}$$

Step 6: Update weights and bias if an error occurred for this pattern

if $y \neq t$,

$$w_i(new) = w_i(old) + \alpha t x_i$$

$$b(new) = b(old) + \alpha t$$

else

$$w_i(new) = w_i(old)$$

$$b(new) = b(old)$$

Step 7. Stopping condition (If no weights changed, stop else continue)

Figure 1 - Algorithm used in this research

2.1 What is PVM software?

We have used the PVM software package that permits a heterogeneous collection of computers hooked together by a network to be used as a single large parallel computer. Thus

large computational problems can be solved more cost effectively by using the aggregate power and memory of many computers. The software is very portable and can be compiled on any kind of machine ranging from laptops to super computer CRAYs. [12]

PVM enables users to exploit their existing computer hardware to solve much larger problems at minimal additional cost. Hundreds of sites around the world are using PVM to solve important scientific, industrial, and medical problems. In addition PVM is used as an educational tool to teach parallel programming. With tens of thousands of users, PVM has become the de facto standard for distributed computing worldwide.

3. SIMULATION AND RESULTS

3.1. Input processing and presenting to ANN model

A large number of input patterns were taken to train the network which consists of 7 characters (A, B, C, D, E, J, K). Each pattern is made of 63 pixels (7 column and 9 rows) where pixels are represented with binary values 1s and 0s. However, the input and target outputs are presented into neural network in bipolar form (+1 and -1) since perceptron gives better results in bipolar form.

computers and making the final decision. All of the common and sequential calculations are out by the master routine in the PVM environment. On the other hand slave routine is mainly used for calculations and tasks which can be carried out in parallel. Ideally, there is more than one slave routine to achieve parallelism. The execution starts from master routine where it's first enrolled itself into the Parallel Virtual Memory environment. After successfully enrolling itself to the PVM environment it then distributed data over all slaves. Slaves processes are spawned over different participating host computers available in the PVM environment, the process is shown diagrammatically in Figure 2.

Training is mostly done through slave routines; the input patterns received by master is distributed over slaves for parallel calculations. Nodes (neurons) of the network are assigned and distributed over different slaves, see Figure 3. If there are n slaves, n-1 slaves have equal distribution of the neural network nodes and the last slave handles the remaining of the nodes. Now each slave calculates the weights and biases associated to the nodes of neural network it is handling. In fact, slaves simply run in an infinite loop, it first looks for the special control input, if its an exit message, the slave exit from the loop and exit from the PVM environment. Otherwise, it receives the data, process it, calculates the weights and biases from the data, and gives back the result to master and waiting for next instruction.

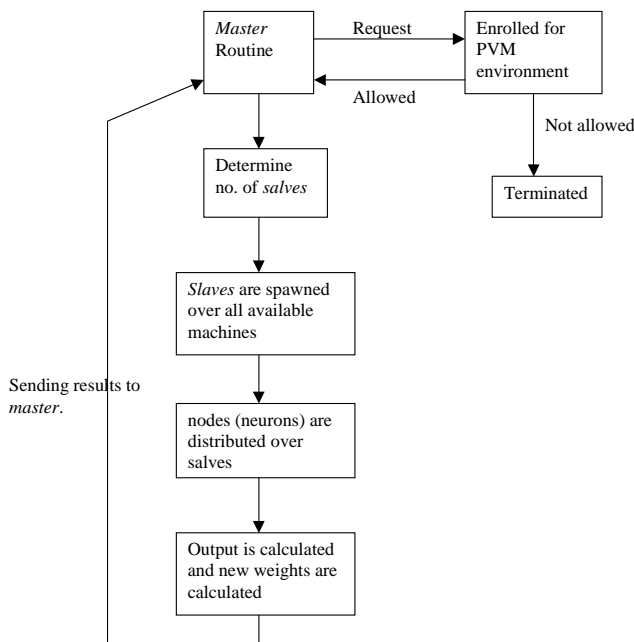


Figure 2 - Operating and dataflow of the system

3.2. Creating Parallel Environment

Usually, PVM implementation consists of two main routines or procedures: they are known as master and slave. The master routine is responsible for initialization of data, receiving input, distributed data over different participating

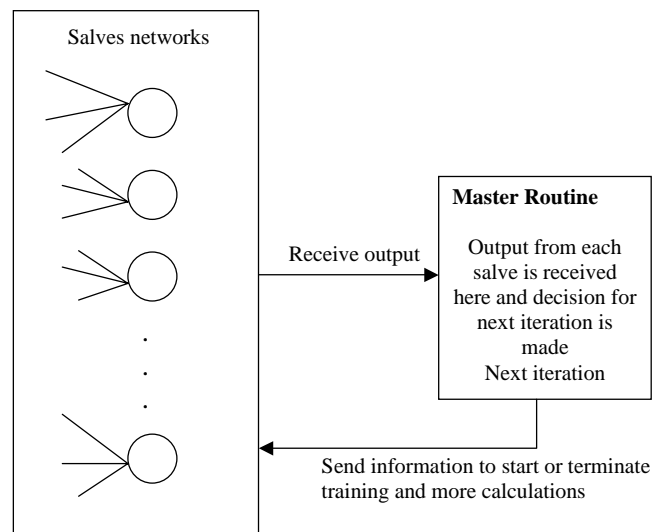


Figure 3 - Structure of Neural network in PVM environment: Master sends and receives data from all slaves, where slaves are responsible for calculation at each neuron in the model

3.3. Analysis of Results

Figure 4 shows the graph of the time taken by the sequential as well as by the parallel implementation of the neural network training for pattern recognition. The same information is also shown in Table 1.



Figure 4 - Graph of the time taken by parallel model with different number of slaves and sequential neural network model

Table 1 - Time taken by parallel model with different number of slaves and sequential implementation of neural network model

Number of Slaves	Time in msec
1	2840
2	1730
3	2025
4	3012
5	3637
6	4143
7	5409
Sequential Time: 2675 msec	

In the graph, the sequential time is represented by line and for the clarity it is drawn against all the slaves. As seen from the graph, if we use one slave, it will take even more time than the sequential, theoretically it should take equal time to the sequential, but due to the network overhead and spawning the slave and initializing the PVM environment also takes time.

The ideal parallel execution is achieved when two slaves were used because the total time taken by the model in this case is recorded to be very low as compare to other sequential and other parallel situations (more than two slaves), see Figure 4. The main reason is that the model needs less time distributing data to and collecting it back from slaves. Results further indicate that a parallel model with three slaves also provides better performance than sequential

implementation. As we increase the number of slave, it needs more time, as the data is such that if we divide it more, its bandwidth is decreased and fewer calculations are done on the slave.

4. CONCLUSION

The simulation results have shown that a PVM based implementation of ANN can be used to solve complex problem in less time as compared to sequential implementation of ANN. PVM based implementation of ANN model provide a good balance of speed and flexibility. The number of slave's plays an important role because of extra time required to manage them in a parallel environment, i.e., network overhead etc. as discussed earlier. As there is no specific formula available to calculate the optimum number of slaves for any given problem, therefore, experimental techniques can be used to determine it. However, slave's population depends on the number of objects to be recognized by the neural network. It could be at-least 25% to 50% of the number of the objects to be recognized. Implementation of back-propagation and Kohanon nets with PVM software are the main considerations for future work.

REFERENCES

- [1] Manavendra M., "Parallel environments for implementing neural networks," Neural Computing Surveys, Vol. 1 pp. 48-60, 1997.
- [2] Skrzypek, J., editor. Neural Network Simulation Environments. Kluwer Academic Publishers, 1993.
- [3] Weigang, L. and Da Silva, N. C., "A study of parallel neural networks," IJCNN'99. International Joint Conference on Neural Networks. Proceedings., Vol. 2, pp. 1113-16, IEEE Service Center, 1999.
- [4] Hammerstrom, D., "A highly parallel digital architecture for neural network emulation," In: Delgado-Frias, J. G. and Moore, W. R. (eds.), VLSI for Artificial Intelligence and Neural Networks, chapter 5.1, pages 357-366. Plenum Press, New York, 1991.
- [5] Gevins, A. S. and Morgan, N. H., "Application of neural network (NN) Signal processing in brain research," IEEE Trans. Acoustics, Speech, and Signal Processing, 36, 1152-1166.
- [6] Hering, D., Khosla, P. and Kumar, B. V. K. V., "The use of modular neural networks in tactile sensing," Proc. Second Int'l. Joint Conference on Neural Networks (Washington, D. C. Jan. 15-19, 1990), II-355-358, 1990.
- [7] Rossen, M. L. and Anderson, J.A., "Representational issues in a neural network model of syllable recognition," Proc. First Int'l. joint conference on Neural Network, (Washington, D. C. June. 18-22, 1989), I-19-26, 1989.
- [8] Casselman, F. and Acres, J. D., "DASA/LARS, a large diagnostic system using neural network," Proc. Second Int'l Joint Conf. On Neural Networks (Washington, D. C. Jan. 15-19, 1990), II-539-542, 1990.
- [9] Nordstrom, T. and B. Svensson, "Using and designing massively parallel computers for artificial neural," Journal of Parallel and Distributed Computing, vol. 14, no. 3, pp. 260-285, 1992.
- [10] Jain, A. K. Mao, J. and Mohiddin, K. M., "Artificial Neural Network: A tutorial," IEEE Computers, pp. 31-44, 1996.
- [11] Rumelhart, D. E., Hinton, G. E. and Williams, R. J., "Learning Internal Representations by error Propagation", in Parallel Distribution Processing: Explorations in the Microstructure of Cognition, vol. 1: Foundations 1986.
- [12] Information on PVM are available at: http://www.epm.ornl.gov/pvm/pvm_home.html

REQUIREMENTS PARTICLE NETWORKS: AN APPROACH TO FORMAL SOFTWARE FUNCTIONAL REQUIREMENTS MODELLING

Wiwat Vatanawood, Wanchai Rivepiboon

Представлен подход к моделированию функциональных требований программного обеспечения. Предложены алгоритмы, реализующие рассматриваемый подход.

In this paper, an approach to software functional requirements modelling using requirements particle networks is presented. In our approach, a set of requirements particles is defined as an essential tool to construct a visual model of software functional requirements specification during the software analysis phase and the relevant formal specification is systematically generated without the experience of writing formal specification. A number of algorithms are presented to perform these formal specification transformations using predefined templates of formal specification schema of the requirements particles. The usability of the requirements particle networks is investigated by conducting a workshop. The result indicates that an analyst with experience in writing data flow diagram is capable to produce a complete and consistent requirements particle networks.

1. INTRODUCTION

There is still a wide gap between the current practice of software requirements engineering and the research on formal specification and software formal development since a formal specification of software system is difficult to write and understand. A number of active researches are conducted in order to extend the capability of software analyst and designer with the specific CASE tools to capture formal specification, including the relevant specification languages to ease the transformation of formal specification [1], [2] and the reverse engineering tasks - to transform program codes to formal specification [3]. It is practical for software analyst and designer to be able to investigate their software system specification in the early stage of the system development.

The Z notation is used in our research to describe formal specification. Since the Z notation is based upon set theory and mathematical logic, the process of constructing proofs can help us to understand the requirements upon a system and can assist us in identifying any hidden assumptions. Proof at the specification stage can make a significant contribution to the quality of software.

Our research is motivated by the work of Jin and Zhu [4] which the formal specification is automatically generated from requirements definition. Our objective is to extend the data flow diagram notations in some manners to explicitly capture precondition predicates happening in the required software system. The technique of decomposition is exploited to obtain a set of requirements primitives so that a well-defined formal specification template can be assigned. Each requirements primitive is expected to guide the identification of requirements patterns in the future work and the reuse of these patterns is feasible.

This paper is organized as follows. Section 2 describes a number of definitions and Algorithms. Section 3 describes the proposed model of software functional requirements specification and section 4 defines formal specification template for requirements particles. The specification procedure is guided in section 5. Section 6 reports the experiment to investigate the usability of the requirements particle networks. Section 7 concludes the paper and the future work is described in section 8.

2. DEFINITIONS AND ALGORITHMS

In this section, a number of definitions and algorithms are introduced and will be used later. The examples are illustrated at the end of this section.

Definition 1. A specification of software functional requirements *SPEC* is considered as a collection of rules that the target software system is obliged to follow as to accomplish user's needs. To provide a formal framework of how to write a unambiguous functional specification for software requirements engineer, the software functional requirements specification is formally defined as $SPEC = (\{RP\}, CPS)$ where $\{RP\}$ is a set of requirements primitives, and *CPS* is a connection description for $\{RP\}$. *RP* and *CPS* are defined in the next paragraph. In our approach, a connection description is defined as a schema in Z.

Definition 2. A requirements primitive *RP* is an essential action expected to perform a particular set of tasks that is required as part of software functional requirements. In our approach, a requirements primitive is formally denoted by $RP = (FS(\Phi), CRP(\Phi))$, where $FS(\Phi)$ represents a set of formal specification defined on each Φ and $CRP(\Phi)$ is a connection description for $FS(\Phi)$.

Definition 3. A requirements particle network is a tuple $\Phi = (V, P, D, ES, EM)$. We define $V = P \cup D$. *P* is a set of vertices called requirements particle vertices. *D* is a set of vertices called data entity vertices. *ES* is a set of edges on $P \times P$, called status edges. *EM* is a set of edges on $V \times V$, called message edges. Let $p, q \in P$ and $d \in D$, edges are written as $p \xrightarrow{Status} q \in ES$, $p \xrightarrow{Message} q \in EM$, $d \xrightarrow{Message} q \in EM$. A requirements particle network is a network of requirements particles and associated data entity. Practically, software functional requirements specification is represented by a set of requirements particle networks.

Definition 4. A requirements particle *P* is an atomic node that performs a specific task which is required as part of a

requirements primitive RP . A requirements particle is formally defined as a collection of ordered attributes $\langle Name, What, Where, Precond, Out, Ack, Nack \rangle$. Let p be a requirements particle, $p.Name$ refers to attribute "Name" while the rest of attributes are referred as $p.What$, $p.Where$, $p.Precond$, $p.Out$, $p.Ack$, $p.Nack$, respectively.

The last six attributes are called communication ports. Each requirements particle communicates to outside world via these communication ports. A number of messages are received via "What" and "Where" port while precondition status is obtained via "Precond" port. "Out" port is used by each particle to forward postconditions to the successive particles. Boolean-value status will be transmitted to successive particles as well via "Ack" and "Nack" port.

Definition 5. A data entity D is the representative of data element or a set of data elements. Data entity D is defined as a ordered pair $\langle Name, Type \rangle$. Let d be a data entity, $d.Name$ refers to attribute "Name" and $d.Type$ refers to attribute "Type".

Definition 6. A set of formal specification $FS(\Phi)$ is formally defined as $\{FST(p, \Phi) | p \in P\}$, where $FST(p, \Phi)$ is a transform function defined in algorithm 1. A well-defined template of formal specification is assigned to match each $p.Name$. A collection of proposed template of formal specification is shown in fig 8 and figure 9.

Algorithm 1. Transform Function $FST(p, \Phi)$

1. Select formal specification template that matches $p.Name$.
2. Substitute all the state variables named "What" with $d.Name$ where $d \xrightarrow{Message} p.What$.
3. Substitute all the state variables named "Where" with $d.Name$ where $d \xrightarrow{Message} p.Where$.
4. Substitute the state variables named "Precond" with the predicates of status edge from $q.Ack$ or $q.Nack$ where $q \in P$ and $q \xrightarrow{Status} p.Precond$.

Algorithm 2. Generating $CRP(\Phi)$

1. Find p where $p, q \in P$ and there exists no such q that $q \xrightarrow{Status} p.Precond$.
2. Define $CRP(\Phi)$ as $FST(p, \Phi) \wedge (CRP(\Phi_{Ack}) \vee CRP(\Phi_{Nack}))$ where Φ_{Ack} is the sub network connecting to $p.Ack$ and Φ_{Nack} is the sub network connecting to $p.Nack$.

Algorithm 3. Generating CPS

Let rp be a requirements primitive, CPS is defined as

$$rp \in RP \vee rp.$$

Example 1. Let s be target $SPEC$ then $s = (\{RP\}, CPS)$ and $rp_1, rp_2, \dots, rp_n \in \{RP\}$. For

requirements primitive rp_1 , see figure 4, such that $rp_1 = (FS(a), CRP(a))$ and let a be a requirements particle network Φ' then $a = (V', P', D', ES', EM')$ and $p_1, p_2, p_3, p_4 \in P'$, $d_1, d_2, d_3, d_4, d_5 \in D'$.

Given that $d_1 \xrightarrow{Message} p_1.What$, $d_2 \xrightarrow{Message} p_1.Where$, $p_1.Out \xrightarrow{Message} p_3.What$, $d_4 \xrightarrow{Message} p_2.What$, $d_3 \xrightarrow{Message} p_2.Where$, $d_2 \xrightarrow{Message} p_3.Where$, $d_5 \xrightarrow{Message} p_4.What$, $d_3 \xrightarrow{Message} p_4.Where \in EM'$, and $p_1.Ack \xrightarrow{Status} p_2.Precond$, $p_1.Nack \xrightarrow{Status} p_3.Precond$, $p_3.Ack \xrightarrow{Status} p_4.Precond \in ES'$.

Then, $FS(a) = \{FST(p_1, a), FST(p_2, a), FST(p_3, a), FST(p_4, a)\}$ and $CRP(a) = FST(p_1, a) \wedge (FST(p_4, a) \vee (FST(p_2, a) \wedge FST(p_3, a)))$. While rp_1 is defined as $rp_1 |$ the value of $CRP(a)$ followed by all formal specification in $FS(a)$. From algorithm 3, CPS is defined as $rp_1 \vee rp_2 \vee \dots \vee rp_n$ and specification s is finally defined as $s |$ the value of CPS followed by all formal specification in $\{RP\}$.

3. A MODEL FOR SOFTWARE FUNCTIONAL REQUIREMENTS SPECIFICATION

In this section, a number of notations are defined to represent requirements particle, data entity and requirements particle networks. Software analyst is provided with a set of requirements particles and data entities to construct software functional requirements specification as desired.

Requirements Particle Notation

Each requirements particle is represented as a circle symbol with 6 communication ports, as shown in figure 1 and figure 2.

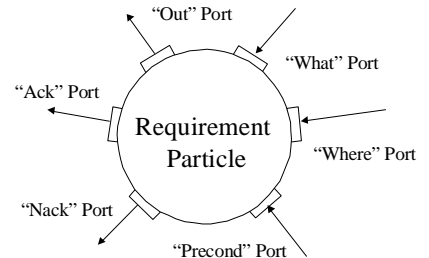


Figure 1 - Requirements Particle Notation

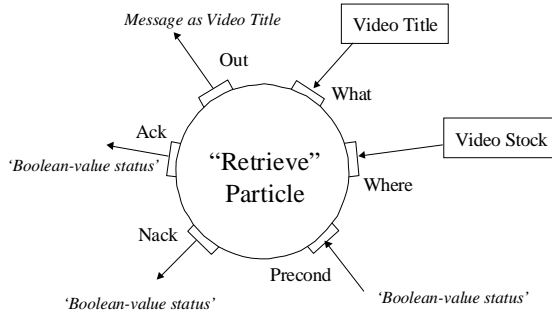


Figure 2 - Sample of Requirements Particle "Retrieve"

Data Entity Notation

Each data entity is represented as a rectangle symbol as shown in figure 3.



Figure 3 - Data Entity Notation and Sample of Its Instance

A Sample of Requirements Particle Networks

To illustrate how to represent a requirements particle network, a sample of video shop's requirements defined in [6] is selected and used in our experiment later. The video shop mentioned has video club members who may hire videos. The functional requirements of video shop system concerns with how to register a new member, how to register new video titles, how to search the video title for hiring and how to manage the hiring record for each member.

A requirements particle network for registering a new video title into video stock data entity is illustrated in figure 4.

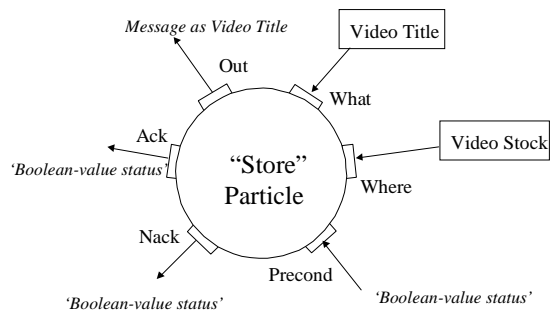


Figure 4 - "Store" Particle Notation

A "Retrieve" particle is required to perform a search of the existing of video title in video stock. If there exists a certain name of video title in video stock, a constant message 'Existing Title' is transmitted to Monitor CRT using "OutDevice" particle. Otherwise, the video title is expected to store into video stock as a new entry using "Store" particle and the constant message 'Register Done' is finally displayed on the Monitor CRT.

4. FORMAL SPECIFICATION TEMPLATE FOR REQUIREMENTS PARTICLE

One of our intentions is to propose a minimum number of simple requirements particles. In this paper, five requirements particles are defined as follows: "Store", "Retrieve", "Remove", "InDevice", "OutDevice" particles. A number of well-defined formal specification items, in Z notations, are defined and assigned to each particle, shown in figure 8 and figure 9. For example, The "Store" particle, shown in figure 5 and defined in figure 6, performs a task of storing messages from "What" port into data entity specified by the message from "Where" port. A number of preconditions will be considered whether the "Store" particle is activated or not. The postconditions will be finally transformed and delivered to "Out" port, "Ack" port and "Nack" port.

Store | [What? : What_Type; Where?, Where! : Where_Type; Ack! : Boolean; Nack! : Boolean; Out! : Out_type; Precond? : Boolean | Precond? \wedge Where! = Where? \cup {What?} \wedge Out! = What? \wedge Ack! = What? \in Where! \wedge Nack! = What? \notin Where!]

Figure 5 - Formal Specification Template for "Store" Particle

Store | [Video Title? : Video Title Type; Video Stock?, Video Stock! : Video Stock Type; Ack! : Boolean; Nack! : Boolean; Out! : Out_type; Precond? : Boolean | Precond? \wedge Video Stock! = Video Stock? \cup {Video Title?} \wedge Out! = Video Title? \wedge Ack! = Video Title? \in Video Stock! \wedge Nack! = Video Title? \notin Video Stock!]

Figure 6 - A Sample of Specification for "Store" Video Title into Video Stock

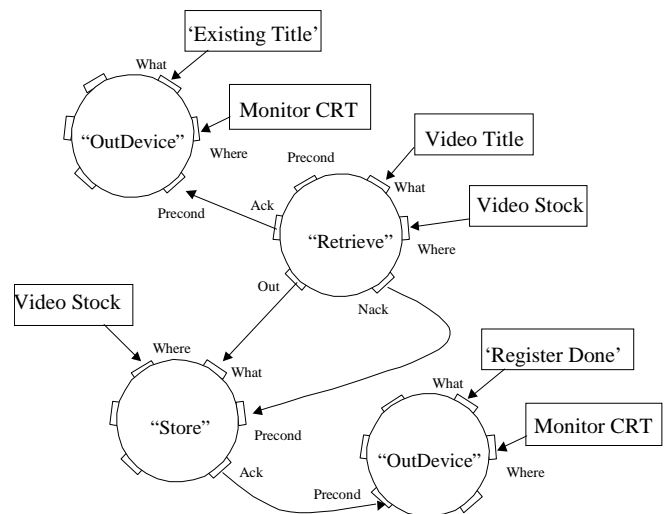


Figure 7 - A Sample of Requirements Particle Network for Registering New Video Title

Store | [What? : What_Type; Where?, Where! : Where_Type; Ack! : Boolean; Nack! : Boolean; Out! : Out_Type; Precond? : Boolean | Precond? \wedge Where! = Where? \cup {What?} \wedge Out! = What? \wedge Ack! = What? \in Where! \wedge Nack! = What? \notin Where!]

Retrieve | [What? : What_Type; Where? : Where_Type; Ack! : Boolean; Nack! : Boolean; Out! : Out_Type; Precond? : Boolean | Precond? \wedge Out! = What? \wedge Ack! = What? \in Where? \wedge Nack! = What? \notin Where?]

Remove | [What? : What_Type; Where?, Where! : Where_Type; Ack! : Boolean; Nack! : Boolean; Out! : Out_Type; Precond? : Boolean | Precond? \wedge Where! = What? \wedge Out! = What? \wedge Ack! = (Where! == What?) \wedge Nack! = !(Where! == What?)]

Where? - {What?} \wedge Out! = What? \wedge Ack! = What? \notin Where! \wedge Nack! = What? \in Where!]

InDevice | [What! : What_Type; Where? : Where_Type; Ack! : Boolean; Nack! : Boolean; Out! : Out_Type; Precond? : Boolean | Precond? \wedge What! = Where? \wedge Out! = What! \wedge Ack! = !({What!} == 0) \wedge Nack! = ({What!} == 0)]

OutDevice | [What? : What_Type; Where! : Where_Type; Ack! : Boolean; Nack! : Boolean; Out! : Out_Type; Precond? : Boolean | Precond? \wedge Where! = What? \wedge Out! = What? \wedge Ack! = (Where! == What?) \wedge Nack! = !(Where! == What?)]

Figure 8 - Formal Specification Templates for "Store", "Retrieve", "Remove", "InDevice", "OutDevice"

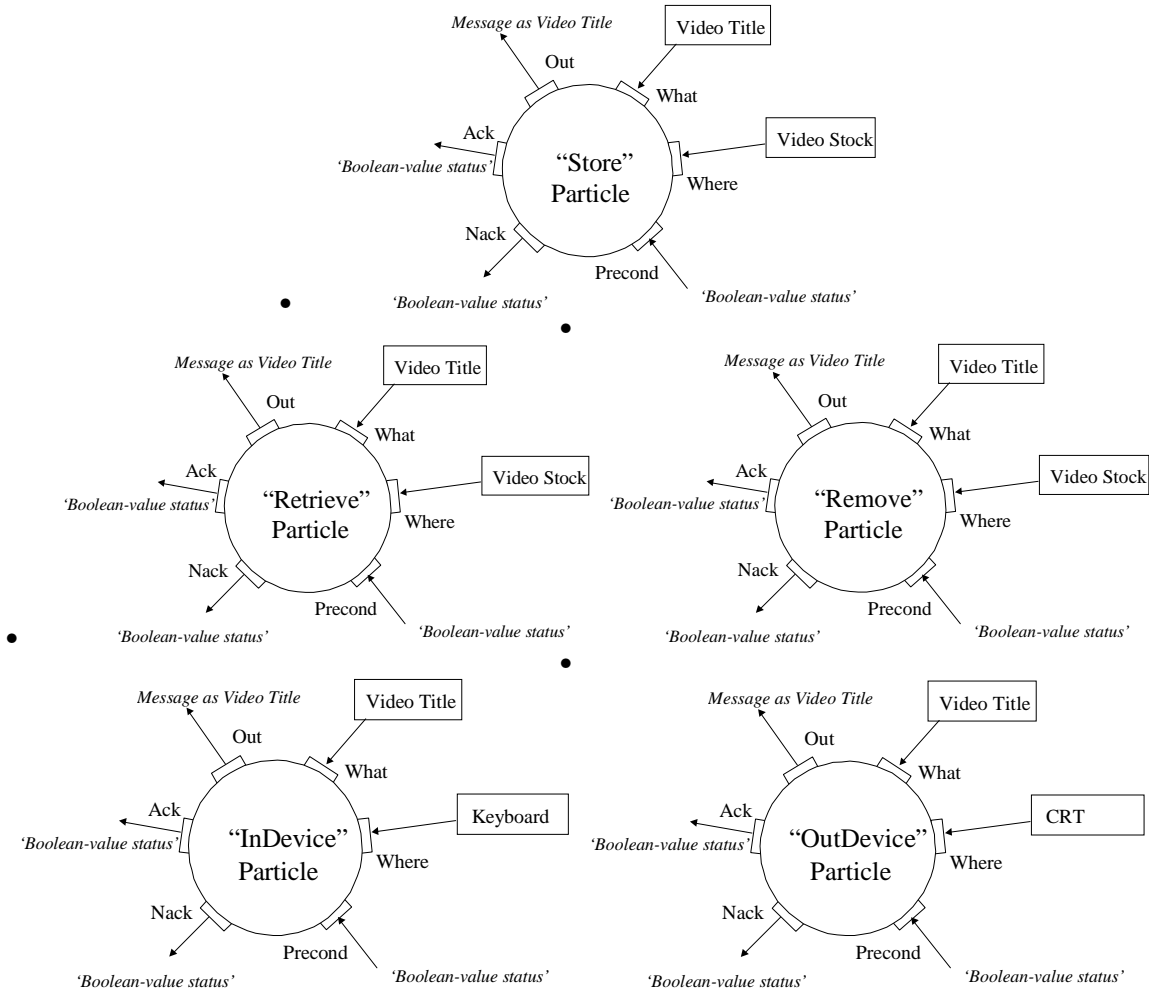


Figure 9 - Particle Notations for "Store", "Retrieve", "Remove", "InDevice", "OutDevice"

5. SPECIFICATION PROCEDURE

In order to develop a software functional requirements specification, the step-by-step procedure is defined as follows:

1. Develop SPEC using decision table approach [5].

2. Decompose SPEC in detail to obtain the final SPEC.
3. Draw requirements particle network for each rule in SPEC.
4. Transform each requirements particle network into formal specification of each requirements primitive with a connection description, using algorithm 1 and 2.

5. Compose the final formal specification, using algorithm 3. A sample of formal specification for "Store" particle shown in figure 5, is specified in figure 7.

6. EXPERIMENT

In order to verify the usability of the requirements particle networks, we conducted a workshop of developing software functional requirements specification of a small software system. A sample of software requirements for "Video Shop" in [6] is selected. More than 80 attendants with experience in writing data flow diagram are gathered. Five requirements primitives are selected as follows: 1) primitive for registering a new video title in the video stock, 2) primitive for searching video stock for the existing of the video title, 3) primitive for registering a new member, 4) primitive for searching the existing of the member name in the member list, and 5) primitive for keeping the video hiring record of each member.

Average time to accomplish the specification procedure is 50 minutes and more than 90% of the attendants produce the complete requirements particle networks. The final formal specifications are consequently mapped from their requirements particle networks without any major complication.

7. CONCLUSION

We have developed an approach to software functional requirements specification using requirements particle networks. Our approach emphasizes that requirements particle is the atomic task and a well-defined formal specification template is relevantly assigned to each requirements parti-

cle. Moreover, we introduce the explicit definition of preconditions in the requirements particle networks so that software analyst is capable to specify when a particular particle is activated in the consequences of preconditions. A number of requirements particles are proposed to deal with store and retrieve functions in software system.

8. FUTURE WORK

We intend to investigate and define more relevant requirements particles for the another part of software system for business information system. A set of requirements particles to perform calculation is required as well. In addition, the reuse features of several common requirements particle networks will be considered.

REFERENCES

- [1] J. J. P. Tsai, T. Weigert, M. Aoyama, A Declarative Approach to Software Requirements Specification Languages, Proceedings of International Computer Languages, 1988, 414-421.
- [2] H. M. Jarvinen, R. K. Suonio, DisCo Specification Language: Marriage of Actions and Objects, Proceedings of Conference on 11th International Distributed Computing Systems, 1991, 142-151.
- [3] B. H. C. Cheng, G. C. Gannod, Abstraction of Formal Specifications from Program Code, Proceedings of IEEE International Conference on Tools for AI, 1991.
- [4] L. Jin, H. Zhu, Automatic Generation of Formal Specification from Requirements Definition, Proceedings of First IEEE Form Engineering Methods, 1997, 243-251.
- [5] W. Vatanawood and P. Chongstitvatana, A Genetic Algorithm Approach to Software Components Search using Software Functional Requirements Checklist, Proceedings of 3rd Annual National Symposium on Computational Science and Engineering (ANSCSE), 1999.
- [6] R. Barden, S. Stepney, D. Cooper, Z in Practice, (Prentice-Hall, 1994).

УДК 681.32

АНАЛИЗ ХАРАКТЕРИСТИК МНОГОПОТОКОВЫХ СЕТЕЙ ОБСЛУЖИВАНИЯ

А.А.Алиев, Б.Г.Исмаилов

Рассмотрены модели многопоточковых компьютерных сетей обслуживания. Поставлены задачи минимизации математического ожидания вероятностной функции потерь информации при минимально необходимой производительности сети. Разработаны процедуры анализа характеристик сети и приведены результаты численных экспериментов.

Itistream computer networks of service are considered. The tasks of minimization of expectation of a probability loss function of the information are posed at minimum of necessary productivity of the network. The procedures of the analysis of the characteristics of the network are developed and the results of numerical experiments are indicated.

В современных сетях компьютеры должны обслуживать большое число источников информации. По этой причине им необходимо работать оперативно и выполнять несколько

операций одновременно. Таким требованиям отвечает мультипроцессорная обработка информации. В таких сетях операционная система позволяет добавить дополнительные процессоры, и она позволяет распределять процессы по нескольким компьютерам, управляя выполняемыми ими задачами. Поэтому с целью ее построения в работе дается анализ характеристик многопоточковых сетей.

В рассмотренной многопоточковой сети имеется m очередей и N мест в очереди. Интенсивность потока заявок, количество заявок в группе, время обслуживания заявок заданы. Характеристики такой сети, в частности, вероятность возникновения заявок на обслуживание, вероятность прихода заявок за время цикла, среднее время цикла и т.д. неизвестны. С целью построения сети необ-

ходимо определить ее характеристики при заданных условиях.

Разработка современных компьютерных систем и сетей сопряжена с применением математических методов моделирования. В условиях случайного характера процессов поступления и обслуживания информации необходимо использование моделей и методов теории массового обслуживания.

В работе рассматривается многопоточковая модель, учитывающая мультипроцессорный принцип построения, позволяющий обеспечить высокие характеристики оперативности и надежности компьютерных сетей; приоритетное обслуживание запросов, ориентированное на повышение общей эффективности компьютерной сети.

На выбор типа используемой модели системы массового обслуживания (СМО) существенно влияет структурная организация анализируемой многопоточковой мультипроцессорной сети.

Теория анализа приоритетных процессов обслуживания достаточно полно разработана в [1].

Однако в отличие от [1] здесь исследуются модели многопоточковых систем с ограничением на время ожидания заявки в очереди.

Рассматриваются многопоточковые компьютерные сети, предназначенные для обслуживания потоков информации поступающих от большего числа источников. Основу системы составляют m периферийных компьютеров ($ПК_i$). Эти $ПК_i$ объединяются в мультипроцессорные сети (МС) с помощью координирующего компьютера (КК). $ПК_i$ организуют сбор и обработку информации. КК соединяет $ПК_i$ с линией связи или терминалом. Конструктивная однородность $ПК_i$ дает возможность наращивания их числа при увеличении количества источников информации, приближения МС к источникам, реализации широкого набора вычислительных процедур.

При организации СМО имеющей много источников информации одним из основных вопросов может являться выбор дисциплины опроса источников.

В обычном случае (без прерывания) КК опрашивает все источники информации по порядку. Если на входе имеется достаточное количество информации, то выполняются операции по ее обслуживанию. В противном случае КК переходит к опросу следующих источников информации.

МС работает в реальном режиме времени, что характерно для управляющих систем и сетей [1].

При моделировании МС основной задачей может являться оптимальное распределение ресурсов сети и программное обеспечение $ПК_i$, которое включает средства программирования и пакет прикладных программ. При этом рассматриваются некоторые вопросы моделирования МС, связанные с оптимальным выбором нагрузочных и структурных параметров. Оптимизация ресурса позволяет получать требуемые значения выбранного критерия функционирования сети в зависимости от параметров обрабатываемой информации.

Одним из показателей качества функционирования МС является минимизация потерь информации, полученной от источников. Следует отметить, что сложность методов

оптимизации ресурсов зависит от вида обработки информации. Если на входе $ПК_i$ дискретная информация, которая поступает в детерминированный момент времени, и известно время обслуживания, то анализ характеристик сети не является трудным. Анализ характеристик сети значительно усложняется при наличии в МС информационных потоков, составляющие которых случайны по моментам поступления.

Если рассматривать сообщения как заявки, а КК, занимающийся сбором сообщений, как обслуживающий прибор, то МС в целом можно рассматривать как однофазную, однолинейную СМО.

Предполагается, что источники заявок выдают стационарные пуассоновские потоки.

Замечание 1. Кроме упрощения анализа, это предположение позволяет получить верхние оценки параметров сети для других законов распределений.

В случае отсутствия потерь интенсивность потока заявок, поступающего на КК равна $\lambda = \sum_{i=1}^m \lambda_i$ и считается,

что в установившемся режиме интенсивности входящих и обслуженных заявок одинаковы.

В сети $ПК_i$ выступают по отношению к КК как устройства ввода-вывода и с помощью критерия эффективности анализируются потери информации как в СМО, имеющей m источников заявок. В каждом источнике имеется конечная очередь заявок; время обслуживания одной заявки и время переключения КК от источника к источнику являются постоянными величинами и заявки могут обслуживаться группами.

В сети имеется m очередей и N мест в очереди. Время обслуживания заявки обозначается $\tau_{обсл}$, Δt - время переключения прибора, λ - интенсивность потока заявок, S - количество заявок в группе.

Запрос на обслуживание возникает в очереди в том случае, когда в ней находятся не менее S заявок, после окончания i -ой очереди прибор переходит к опросу $(i+1)$ -ой, цикл работы СМО завершается опросом m -ой очереди.

Заявки, поступившие в очередь, после того как был произведен ее опрос, рассматриваются в данном цикле.

Оценка характеристик сети осуществляется путем минимизации математического ожидания вероятности потерь информации при заданном количестве источников, т.е.

$$M[P(m, N, S, \lambda, \tau_{обсл})] \rightarrow \min, \lambda \leq \lambda^0, S \leq L_q \leq L^0, (1)$$

где λ^0, L^0 - максимальные допустимые значения λ, L_q ; L_q - количество заявок в очереди.

Однако строгое аналитическое выражение для определения потерь $P(m, N, S, \lambda, \tau_{обсл})$ в таких СМО в настоящее время отсутствует и, следовательно, аналитическое решение постановки (1) представляет большую сложность.

Вместе с тем в таких СМО потери вида (1) могут быть определены с помощью следующих характеристик:

- вероятность возникновения заявок на обслуживание в

n из m очередей как [1]:

$$P_{n+1} = \pi_s(ns\tau_{обсл} + \tau) / (1 - \pi_s((n+1)s\tau_{обсл} + \tau)),$$

где $\tau = m\Delta t$, $\pi_s(ns\tau_{обсл} + \tau)$ - вероятность появления заявок в очереди, с которой начинается цикл;

- заявки на обслуживание заявок появляются при наличии в очереди S и более заявок, при этом для простейшего потока вероятность прихода заявок, за время цикла

$$P_k = \frac{\lambda(ns\tau_{обсл} + m\Delta t)^k}{k!} e^{-\lambda(ns\tau_{обсл} + m\Delta t)};$$

- среднее время цикла

$$\tau_c = \sum_{n=0}^m (ns\tau_{обсл} + m\Delta t)P_n;$$

- вероятность появления заявок в первой очереди

$$\pi = \sum_{n=0}^{\dots} \pi_s(ns\tau_{обсл} + m\Delta t)P_n.$$

Следовательно, вероятность потери заявок

$$P = 1 - \pi_s / \lambda \tau_c. \quad (2)$$

Данное соотношение в нормализованном виде может быть косвенным решением постановки (1). Под нормализацией подразумевается выбор из множества значений P лишь тех, при которых для количества заявок L_q , ожидающих обслуживания, выполняется условие $S \leq L_q \leq L^0$.

Таким образом, при выполнении условия $S \leq L_q \leq L^0$, рассмотренная СМО обеспечивает обслуживание потока заявок в пределах допустимых потерь и, следовательно, система имеет минимально необходимую производительность.

В СМО среди источников заявок в первую очередь выделяют приоритетные. Приоритетность в СМО определяется таким образом, что для группы заявок, ожидающих обслуживания $c_i \mu_i$, является самой большой по значению, т.е. классы приоритетов упорядочиваются согласно соотношению [3]:

$$\mu_1 c_1 \leq \mu_2 c_2 \leq \dots \leq \mu_k c_k, \quad (3)$$

где c_i - стоимость обслуживания за единицу времени.

При решении задачи (1), устанавливая приоритетность по (3), осуществляется выбор из множества значений P лишь тех, при которых выполняется условие $L_q > S$.

Предлагаются прерывания двух типов: прерывания с относительным и абсолютным приоритетами. При этом во

втором случае прерванные заявки вновь поступают в КК и их обслуживание начинается с прерванного места.

Данная задача для небольших значений интенсивностей потоков информации на входе СМО может быть решена при отсутствии приоритетов. Здесь в рассмотренную систему вводятся k различных приоритетных классов ($k \geq 3$), т.е. в системе имеются три крупных приоритетных класса заявок. При функционировании системы внутри потоков может быть дальнейшее разделение на подпотоки. Заявки приоритетного класса K $K = 1, 2, \dots, k$ поступают как пуассоновские потоки с интенсивностью λ_k , каждая заявка из приоритетного класса K имеет время обслуживания, выбранное независимо из экспоненциального распределения со средним значением $1/\mu_k$.

Кроме того, применение приоритетности в данном случае показывает, что значение математического ожидания вероятностной функции потери заявок существенно меньше для прерывания с абсолютным приоритетом, чем для прерывания с относительным приоритетом. Предложенный подход обеспечивает исследование многопоточковых СМО с относительным и абсолютным приоритетами. В его основе лежит решение задачи анализа характеристик компьютерной сети с групповым обслуживанием.

Данная процедура определения характеристик СМО обобщена в виде алгоритма, который имеет следующие шаги:

- Для простейших потоков вводятся значения m, N, S с целью определения основных характеристик СМО.

- Для количества заявок, ожидающих обслуживания проверяется условие $L_q \geq S$. Если данное условие удовлетворяется для относительных и абсолютных приоритетов, то вычисляются и выводятся на печать характеристики СМО.

- Если цикл опроса не завершен, осуществляется переход к первому шагу. В противном случае процесс анализа характеристик СМО завершается.

На основе вышеизложенной процедуры анализа для определения характеристик систем проведены объемные вычислительные эксперименты и получены численные результаты. В этих экспериментах при заданном количестве источников информации определены потери информации в зависимости от интенсивности входного потока.

Потери информации в такой СМО могут возрастать за счет появления потерь в ПК_{*i*}, и МС можно рассматривать как двухфазные СМО [2]. Предполагая что, потоки заявок в обеих фазах пуассоновские, время обслуживания одной заявки, как в первой, так и во второй фазе постоянно. Для оценки параметров СМО можно использовать результаты приведенные здесь для однофазной СМО.

Следует отметить, что при более детальном изучении СМО такого вида целесообразно применять методы имитационного моделирования, так как аналитическое описание двухфазной СМО с не пуассоновскими потоками заявок и ограниченной очередью во второй фазе представляется очень сложным.

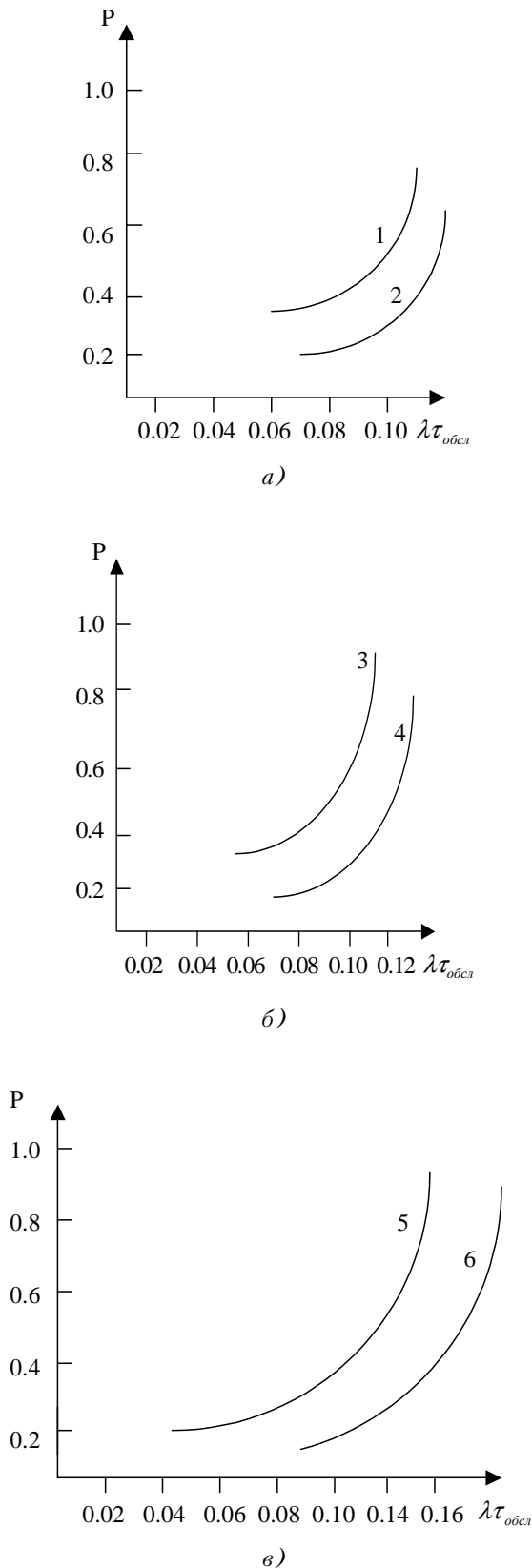


Рисунок 1

В качестве примера, для $\Delta t = 0,1\tau_{обсл}$: а) $m = 4$, $S = N = 1$ (кривая 1); б) $m = 4$, $S = N = 2$ (кривая 2); в) $m = 16$, $S = 1$, $N = 10$ (кривая 3); г) $m = 2,5$, $S = 1$, $N = 10$ (кривая 4); д) $m = 16$, $S = N = 1$ (кривая 5); е) $m = 16$, $S = N = 2$ (кривая 6) исследованы зависимости при различных m , N и S . Из рисунка 1 видно, что при $m = 4$, $S = N = 1$ и $m = 4$, $S = N = 2$ значения кривых 1, 3 вероятностной функции по сравнению с другими кривыми находятся в пределах допустимых потерь и полученные характеристики СМО считаются оптимальными.

ЗАКЛЮЧЕНИЕ

Процедура анализа многопоточковых сетей обслуживания позволяет сделать следующие выводы:

1. Разработанная процедура анализа оптимальных характеристик многопоточковых сетей является удобной для описания сети с групповым обслуживанием. Использование разработанного алгоритма вычисления характеристик сетей может значительно облегчить анализ характеристик подобных сетей.
2. Разработка такой модели позволяет определить не только вероятностно-временные характеристики сети в рамках системы массового обслуживания, но и повысить надежность системы в целом.

ПЕРЕЧЕНЬ ССЫЛОК

1. Меликов А.З., Пономаренко Л.А., Рюмшин Н.А. Математические модели многопоточковых систем обслуживания. - К. 1991, - 264 с.
2. Меликов А.З., Исмаилов Б.Г. О моделировании процесса маршрутизации сообщений в распределенных компьютерных сетях обслуживания. Известия АН Азербайджана. Серия физико-технических и математических наук, XVIII том, №1, 1998 г. - с. 54-56.
3. Исмаилов Б.Г. Анализ характеристик взаимодействия компьютеров распределенных сетей обслуживания. Изв. АН Азерб., Серия физико-технических и математических наук, XIX том, №3-4, 1999г.- с.130-136.

МОДЕЛИРОВАНИЕ И ОБРАБОТКА ДАННЫХ СИСТЕМ ТЕСТОВОГО КОНТРОЛЯ ЗАКРЫТОГО ТИПА

А.Е.Архипов, С.А.Архипова

Предпринята попытка формализации описания процедуры тестового контроля, построение модели, получаемых в ходе тестирования данных, оптимизации их обработки для тестов закрытого типа.

Зроблено спробу формалізації опису процедури тестового контролю, побудови моделі даних, що отримані в процесі тестування, оптимізації їх обробки для тестів закритого типу.

The attempt of formalisation of exposition of a procedure of test monitoring, construction of a model dates, obtained during testing, optimisation of their handling for the tests of the closed type is undertaken.

ВВЕДЕНИЕ

Среди множества информационных технологий, применяемых в социальной сфере, устойчивым интересом и вниманием характеризуются тестовые методы контроля и диагностики, позволяющие реализовать одну из наиболее распространенных форм профессионального отбора, аттестации, контроля и оценивания уровня знаний и умений специалистов в конкретных предметных областях. Очередной всплеск внимания к прикладным аспектам тестирования обусловлен интенсификацией применения тестовых технологий в образовательной сфере, в частности, их очевидная перспективность в системе дистанционного образования, для которого тестовая форма контроля является базовой.

1. ФОРМАЛИЗАЦИЯ ПРОЦЕДУРЫ ТЕСТОВОГО КОНТРОЛЯ

Несмотря на внушительный перечень достоинств и положительных отзывов [1,2], интенсификация и расширение применений тестовых систем сталкивается с определенными ограничениями. Причина этого - большой вес эмпирических приемов в построении и эксплуатации тестовых систем, отсутствие объективных критериев качества и надежности функционирования этих систем, получаемых с их помощью выводов и рекомендаций.

Поэтому актуальна проблема разработки методологии исследования и анализа качества тестовых систем, надежности и объективности результатов тестового контроля, построение методик оптимизации и совершенствования тестов по результатам их практического применения.

Очевидно, исходным пунктом в решении этих задач является формализация процедуры тестового контроля, представление ее некоторой структурой, включающей ряд элементов с вполне определенным назначением и свойствами.

Рассмотрим систему тестового контроля (СТК) уровня знаний, базирующуюся на закрытом тесте, основным элементом которого является множество $\{q_i\}$, $i = \overline{1, N}$ контрольных заданий (вопросов), каждому из которых соответствует по m вариантов возможных ответов $\{v_{11}, \dots, v_{1m}\}, \dots, \{v_{n1}, \dots, v_{nm}\}$, содержащих лишь один правильный, выбрать который - задача тестируемого.

Взаимодействие тестируемого (объект исследования) и СТК можно описать, используя аналогию с процессом измерения некоторой физической величины L , схемой вида:

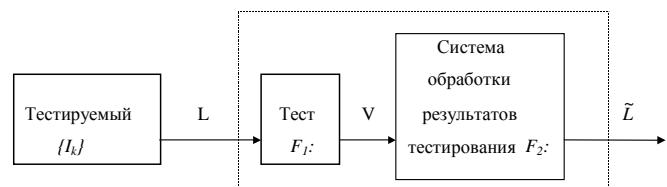


Рисунок 1 - Система тестового контроля (СТК)

Согласно этой схеме тестируемый, обладающий в соответствующей предметной области совокупностью знаний $\{I_k\}$, характеризуемых некоторым интегральным показателем - уровнем знаний L , выполняет (F_1) задание теста, т.е. отвечает на контрольные вопросы, получая в итоге набор $V = v_i$ вариантов ответов. В системе обработки результатов тестирования в ходе анализа этого набора определяется (F_2) оценка \tilde{L} исходного уровня знаний. Таким образом, имеем два последовательных отображения, первое из которых

$$F_1: L \rightarrow V \quad (1)$$

переводит количественное значение измеряемой величины L в значение некоторой промежуточной переменной V , а целью второго

$$F_2: L \rightarrow V \quad (2)$$

является восстановление исходного значения уровня знаний L . Очевидно, в идеальном случае произведению отображений $F_1 \cdot F_2$ должно соответствовать тождество

$F_1 \cdot F_2 \equiv 1$, что обеспечивает эквивалентность переменных $L = \tilde{L}$. К сожалению, сколько-нибудь обоснованное задание алгоритма обратного отображения $F_2 = F_1^{-1}$ оказывается невозможным, так как механизм реализации отображения F_1 , выполняемого непосредственно тестируемым, неизвестен и его формализация обычно не представляется возможной. Поэтому $\tilde{L} = L + \Delta_{LF_2}$ где Δ_{LF_2} - ошибка тестирования, определяющая валидность [1] теста, т.е. принципиальную возможность измерения с помощью СТК именно той характеристики испытуемого, определения которой и является целью проводимого тестового контроля.

По сути используемая в схеме рис.1 переменная L является фиктивной. Ее введение обусловлено исключительно желанием уйти от необходимости оценивания полноты имеющихся у испытуемого знаний в рассматриваемой предметной области, поскольку само понятие знаний (суммы знаний) определяет весьма сложное комплексное явление, многомерное и многосвязное, непосредственное включение которого в схему рис.1 оказывается невозможным. Поэтому появляется переменная L - результат некоторого "предварительно" выполненного отображения суммы знаний тестируемого $\{I_k\}$ в интегральный показатель L уровня знаний.

2. ПОСТАНОВКА ЗАДАЧИ ОБРАБОТКИ ДАННЫХ ТЕСТОВОГО КОНТРОЛЯ

Попытаемся сформулировать набор требований к СТК, выполнение которых позволит обосновать введение вполне определенного алгоритма F_2 обработки результатов тестирования. Пусть полному объему знаний в рассматриваемой предметной области соответствует некоторое количество информации I , допускающее произвольное фрагментирование на частные количества δI , $0 \leq \delta I \leq I$. Предположим, что

а) совокупность контрольных заданий (вопросов) $Q = \{q_i\}$, $i = \overline{1, n}$ составлена из вопросов, ответ на каждый из которых требует одинакового количества информации $I(q_i) = I$, т.е. $I_1 = I_2 = \dots = I_N$;

б) количество взаимной информации $I(q_i \leftrightarrow q_j)$ для произвольной пары контрольных вопросов $q_i, q_j, i \neq j$ равно 0, т.е. для ответа на каждый вопрос теста требуется равный и непересекающийся с другими фрагмент полной информации I , при этом $\sum_{i=1}^N I_i = I$, т.е. частные фрагменты I_i равномерно и полностью покрывают всю информацию I данной предметной области, $I_i = I/N, k = \overline{1, N}$;

в) в ходе тестирования исключена возможность угадывания ответа, т.е. правильный выбор варианта ответа

означает, что тестируемый владеет соответствующей частной информацией.

При выполнении допущений а)-в) уровень знаний L тестируемого можно однозначно оценить в равномерной дискретной шкале $[0, N]$, причем численно $L = n$, $n \leq N$ - количество правильных ответов на контрольные вопросы.

Если полагать, что сложность вопроса определяется количеством информации, необходимым для правильного ответа на него, то тест, удовлетворяющий требованиям а)-в) можно назвать тестом равномерной сложности. Если снять ограничение а), предположив в общем случае $I_1 \neq I_2 \neq \dots \neq I_N$ получим тест неравномерной сложности, для которого уровень знаний L тестируемого определяется формулой

$$L = \sum_{i=1}^N w_i \text{Ind}(q_i), \quad (3)$$

где $\text{Ind}(q_i)$ - индикаторная функция, равная 1 при правильном выборе варианта ответа и 0 - в противном случае, w_i - весовой коэффициент, определяющий информативность i -го вопроса,

$$w_i = NI_i/I. \quad (4)$$

Предложенные модели тестов не единственные, позволяющие обосновать построение алгоритма F_2 обработки результатов тестирования. Можно построить тест, состоящий из ряда упорядоченных последовательностей вопросов нарастающей сложности вида $q_j, q_{j+1}, \dots, q_{j+r}$ причем $I(q_j) \subset I(q_{j+1}) \subset \dots \subset I(q_{j+r})$. Каждая последовательность включает вопросы, относящиеся к одному и тому же тематическому разделу, отличающиеся лишь возрастающим объемом информации, необходимым для правильного ответа на вопрос с более высоким индексом. После тестирования эти последовательности приобретают вид: $\text{Ind}(q_j) = 1, \dots, \text{Ind}(q_{j+r}) = 0, \dots$, т.е. начиная с некоторого индекса $j+r$ сложность вопросов, входящих в последовательность, превышает уровень информированности тестируемого по данной тематике. Естественно, что в подобном тесте ранее приведенные условия а), б) не соблюдаются, однако при выполнении ограничения в) специальный совместный логический анализ ответов тестируемого на вопросы, образующие структуру последовательностей нарастающей сложности, позволяет оценить те объемы информации, которыми владеет тестируемый и вычислить его уровень знаний L .

В реальных тестах можно выделить элементы, соответствующие как моделям тестов равномерной и неравномерной сложности, так и тестам нарастающей сложности, что должно быть учтено при организации обработки результатов тестирования. Для таких тестов наиболее адекватна модель теста неравномерной сложности с наличием ряда вопросов, для которых характерно

частичное перекрытие фрагментов информации. Более подробно эти вопросы рассмотрены в разделе 4.

Кроме того следует учитывать, что процедура тестирования производится в условиях действия различных возмущений случайного характера, в связи с чем на переменную V накладывается случайная составляющая E . При этом даже при выполнении тождества $F_1 \cdot F_2 \equiv 1$ выход СТК будет содержать шумовую составляющую Δ_{LE} :

$$F_2: (V + E) \rightarrow \tilde{L} = (V \rightarrow L) + (E \rightarrow \Delta_{LE}) = L + \Delta_{LE}, \quad (5)$$

определяющую уровень случайного рассеяния значений оценки \tilde{L} , т.е. надежность теста [1]. Основным источником стохастичности в результатах тестирования является угадывание ответа при незнании контрольного вопроса, т.е. невыполнение условия в). Наличие этой особенности реального тестирования требует специального анализа и учета при обработке тестовых данных [3].

3. УЧЕТ ЭФФЕКТА УГАДЫВАНИЯ В РЕЗУЛЬТАТАХ ТЕСТОВОГО КОНТРОЛЯ

Рассмотрим особенности анализа и учета случайной составляющей, обусловленной возможностью угадывания правильного варианта ответа на контрольный вопрос при отсутствии у тестируемого необходимых для ответа знаний.

Анализ имеющей место в этом случае ситуации проводится на модели теста равномерной сложности [2,4], для которого вся информация о предметной области равномерно распределена между N контрольными вопросами, что позволяет в качестве знания L уровня знаний тестируемого в дискретной целочисленной шкале $[0, N]$ использовать количество правильных ответов n , обусловленных действительным знанием контрольных вопросов. Определим общее количество полученных правильных ответов как

$$g = n + e, \quad (6)$$

где e - число случайно угаданных ответов при незнании тестируемым материала контрольного вопроса. В общем случае g - реализация случайной величины

$$G = n + E, \quad (7)$$

вероятностные свойства которой полностью определяются случайной величиной E , представляющей собой вероятностную модель процедуры угадывания правильных вариантов ответа для совокупности из $N - n$ контрольных вопросов. Если m - число равновероятных вариантов ответа на произвольный контрольный вопрос, то случайная величина E может принимать целочисленные значения e в диапазоне $[0, N - n]$, распределенные по биномиальному закону (закон Бернули) [5]:

$$p(e) = c_{N-n}^e \alpha^e \beta^{N-n-e}, \quad (8)$$

где $\alpha = 1/m$, $\beta = 1 - \alpha = (m - 1)/m$, c_{N-n}^e - число сочетаний из $N - n$ по e ,

$$c_{N-n}^e = \frac{(N-n)(N-n-1)\dots(N-n-e+1)}{1 \cdot 2 \cdot 3 \dots e}, \quad (9)$$

которое, учитывая, что $c_{N-n}^0 = 1$ допускает рекуррентный способ вычисления по формуле

$$c_{N-n}^e = c_{N-n}^{e-1} \frac{N-n-e+1}{e}. \quad (10)$$

Математическое ожидание и дисперсия случайной величины E определяются выражениями [5].

$$M\{E\} = (N-n)\alpha = (N-n)/m, \quad (11)$$

$$D\{E\} = (N-n)\alpha\beta = (N-n)(m-1)/m^2. \quad (12)$$

Проанализируем возможность использования общего числа правильных ответов g для оценки уровня знаний тестируемого. Из соотношений

$$M\{G\} = n + M\{E\} = n + (N-n)\alpha, \quad (13)$$

$$D\{G\} = D\{E\} = (N-n)(m-1)/m^2, \quad (14)$$

следует, что лишь при больших значениях n , близких к N , оценка $\tilde{L} = g$ в среднем дает достаточно хорошее приближение к фактическому значению $L = n$, сохраняя при этом относительно низкий уровень рассеяния. С уменьшением n оценка $\tilde{L} = g$ приобретает существенное положительное смещение и увеличившийся разброс своих значений, т.е. не может использоваться в качестве объективного показателя уровня знаний тестируемого.

Скорректировать смещение, значение которого оказывается зависимым от фактического уровня значений L (или, точнее, от неизвестного значения n), позволяет соотношение

$$\tilde{L} = \tilde{n} = g - M\{E\} = g - \alpha N + \alpha \tilde{n}, \quad (15)$$

из которого, после замены неизвестного значения n его оценкой \tilde{n} , получаем выражение для несмещенной оценки \tilde{n} :

$$\tilde{n} = (g - \alpha N) / \beta = (g - \alpha N) / (1 - \alpha), \quad (16)$$

дисперсия которой пропорциональна $D\{G\}$.

Еще один способ улучшения оценки \tilde{L} опирается на использование байесовского подхода [5]. Предположим, по результатам тестирования получено значение g . Соответствующее ему значение n может принимать одно

из целочисленных значений в диапазоне $[0, N]$. Используя формулу (8) можно рассчитать условные вероятности $p(g/n) = p((g-n)/n) = p(e/n)$ для всех n из этого диапазона и найти полную вероятность

$$p(g) = \sum_{n=0}^g p(g/n)p(n), \quad (17)$$

где в качестве априорного значения вероятности $p(n)$ принято значение $p(n) = 1/(N+1)$. Последнее вполне оправдано, ибо до проведения тестирования не было никакой априорной информации относительно уровня знаний тестируемого, который как раз и определяется величиной n , лежащих в целочисленном диапазоне $[0, n]$. Рассчитав $p(g)$, по формуле Байеса, находим апостериорные условные вероятности

$$p(n/g) = p(g/n)p(n)/p(g), \quad 0 \leq n \leq g \quad (18)$$

По полученному условному распределению $p(n/g)$, $0 \leq n \leq g$ можно рассчитать оценку \tilde{n} , приняв в качестве ее условное среднее

$$\tilde{n}/g = \sum_{n=0}^g np(n/g), \quad (19)$$

моду или медиану распределения $p(n/g)$, $0 \leq n \leq g$.

Обычно результаты тестового контроля (ТК) представляются в некоторой шкале оценок (4-балльной, 12-балльной, в общем случае l -балльной). Не вдаваясь в философские аспекты измерения уровня знаний и построения соответствующих шкал, введем формальные соотношения, позволяющие построить подобную шкалу. Полагаем, что $d_1, \dots, d_i, \dots, d_l$ отметки l -балльной шкалы, каждой из которых соответствует некоторый численный отсчет n_i , определенный как количество известных для тестируемого правильных ответов на контрольные вопросы: $n_1, n_2, \dots, n_i, \dots, n_l$, причем $n_1 < n_2 < \dots < n_i < \dots < n_l$, $n_i = i\Delta_n$, $i = \overline{1, l}$, $\Delta_n = N/l$ - целое число, кратное N . Приведенную равномерную шкалу отсчетов $\{n_i\}$ нельзя использовать для объективной оценки уровня знаний тестируемых из-за эффекта угадывания, приводящего к рассмотренному ранее положительному смещению в результатах тестирования. Учесть влияние угадывания можно введением последовательности упорядоченных количеств правильных ответов $\delta_1, \delta_2, \dots, \delta_i, \dots, \delta_l$, включающих как известные тестируемому ответы, так и случайно отгаданные им, "объективно" смещенных по отношению к исходной равномерной шкале: $\delta_1 > n_1$, $\delta_2 > n_2$, ..., $\delta_i > n_i$,

Для расчета значений $\delta_1, \delta_2, \dots, \delta_i, \dots$ используем формулу (13), в соответствии с которой

$$\delta_i = n_i + (N - n_i) \frac{1}{m} = n_i \beta + \frac{N}{m}, \quad i = \overline{1, l}. \quad (20)$$

Для последовательности $\{\delta_i\}$ иногда оказывается удобным введение нулевого отсчета $\delta_0 = N/m$, рассчитываемого по общей формуле (20) для $n_i = 0$. Полагая $n_i = iN/l$, получаем

$$\delta_i = \frac{N}{ml} [i(m-1) + l], \quad i = \overline{1, l}, \quad (21)$$

или

$$\delta_i = \delta_0 + \frac{N}{l} \beta i, \quad i = \overline{1, l}. \quad (22)$$

Так, для 4-балльной шкалы оценок, соответствующим традиционным отметкам "2", "3", "4", "5", имеем (для $N = 96$, $m = 4$, $l = 4$):

Отметки	-	"2"	"3"	"4"	"5"
i	0	1	2	3	4
n_i	-	24	48	72	96
δ_i	24	42	60	78	96

При необходимости в значения $\{n_i\}$ также может быть введен нулевой отсчет n_0 , отражающий субъективные особенности учета знаний в данной предметной области. Это повлечет за собой изменение численных значений последующих отсчетов n_i , δ_i .

Рассмотренные выше методы исключения влияния случайной составляющей на результаты тестирования, в частности коррекция положительных смещений в оценках $\tilde{L} = g$, ориентированы на применение в СТК, для которых справедлива модель теста равномерной сложности. В более общих ситуациях влияние эффекта "угадывания" можно ослабить увеличением числа вариантов m , однако, как отмечается в [1], не рекомендуется применять $m > 7$.

Другим способом устранения влияния случайности в результатах тестирования является введение в тест l вопросов-фильтров, ответы на которые требуют знания одной и той же частной информации I_j . Суть фильтрации в том, что наличие в группе хотя бы одного "нулевого" ответа свидетельствует об угадывании ответов на вопросы, составляющие одну группу. В зачет от группы идет лишь одна оценка, определяемая выражением

$$\text{Ind}(q_i) = \prod_{r=1}^l \text{Ind}(q_{ir}).$$

Некоторые практические методы повышения надежности ТК приведены в [1].

**4. ОБРАБОТКА ДАННЫХ ДЛЯ ТЕСТОВ
НЕРАВНОМЕРНОЙ СЛОЖНОСТИ**

Как уже отмечалось, в реальных СТК последовательность контрольных заданий нельзя описать схемой независимых событий, каждое из которых характеризуется m дискретными равновероятными исходами.

Во-первых, для правильных ответов на контрольные задания требуются различные объемы информации, т.е. задания имеют разную информативность, что следует учитывать при формировании итоговой оценки уровня знаний испытуемого введением специальных весовых коэффициентов для каждого из заданий.

Кроме того, эти объемы информации могут взаимно перекрываться, что выражается в появлении группы заданий с высококоррелированными ответами. Общая информативность подобных групп ниже суммы индивидуальных информативностей контрольных заданий, входящих в группу.

Возможны различные способы задания значений весовых коэффициентов w_i . Один из самых простых - экспертное оценивание информативности I_i отдельных вопросов с последующим определением весов непосредственно по формуле (4).

Попытки более объективного обоснования выбора веса базируются на предположении о существовании взаимосвязи между информативностью I_i вопроса q_i и вероятностью правильного ответа на этот вопрос p_i , и, в частности, сводятся к предположению, что эту взаимосвязь можно описать выражением

$$I_i = c \ln p_i, \tag{23}$$

где c - некоторый коэффициент пропорциональности. Так как фактическое значение p_i не известно, его можно заменить оценкой \tilde{p}_i , в качестве которой предлагается использовать относительную частоту правильных ответов на контрольный вопрос $\tilde{f}_i = k_i/K$ (K - общее количество тестируемых, k_i - число данных ими правильных ответов). Приведенный выше материал позволяет утверждать, что p_i не эквивалентно f_i , так как последнее учитывает эффект угадывания ответа:

$$f_i = p_i + (1 - p_i)/m. \tag{24}$$

Однако при корректном построении теста из оценки \tilde{f}_i можно найти оценку \tilde{p}_i :

$$\tilde{p}_i = \frac{m\tilde{f}_i - 1}{m - 1}. \tag{25}$$

Тогда $I_i \approx -c \ln \tilde{p}_i$, $I \approx -c \sum_{i=1}^K \ln \tilde{p}_i$, $w_i \approx N \ln \tilde{p}_i / \sum_{i=1}^K \ln \tilde{p}_i$.

Две последние формулы справедливы только при отсут-

ствии взаимного пересечения фрагментов информации, соответствующих различным вопросам теста. В противном случае результаты ответов на такие вопросы должны быть взаимосвязанными. Существование и уровень этих взаимосвязей можно оценить, сопоставив и проанализировав последовательности ответов участников теста на упорядоченную совокупность контрольных вопросов q_1, q_2, \dots, q_N представляющие собой состоящие из 0 и 1 бинарные последовательности вида $\{\beta_j\}_i = \{\beta_1, \dots, \beta_K\}_i$, $i = \overline{1, N}$, $\beta_j = \text{Ind}_j(q_i)$, $j = \overline{1, K}$, где K - общее количество тестируемых. Наличие существенных корреляций между ответами, полученными на контрольные вопросы q_r, q_s, \dots, q_t позволяет говорить о взаимосвязи этих ответов, обусловленной в частности тем, что знания тестируемых, требуемые для ответов на эти вопросы, опираются на пересекающиеся объемы информации.

Простейшей моделью, отражающей подобную взаимосвязь, является регрессионное соотношение вида

$$\beta_r = a_s \beta_s + \dots + a_t \beta_t + \xi, \tag{26}$$

коэффициенты которого могут быть рассчитаны по матрице исходных данных $|\beta_{ji}|$ с помощью известных методов оценивания, например, метода наименьших квадратов. Степень тесноты статистической связи между зависимой переменной β_r и совокупностью объясняющих переменных β_s, \dots, β_t измеряется множественным коэффициентом корреляции R , квадрат которого - коэффициент детерминации R^2 показывает, какая доля дисперсии $\sigma^2(\beta_r)$ зависимой переменной определяется совокупным влиянием объясняющих переменных β_s, \dots, β_t :

$$R^2 = 1 - \frac{\sigma_{\xi r}^2}{\sigma^2(\beta_r)}, \tag{27}$$

где $\sigma_{\xi r}^2$ - остаточная дисперсия (средний квадрат) рассеяния значений β_r относительно функции регрессии [5]. Соответственно разность

$$1 - R^2 = \frac{\sigma_{\xi r}^2}{\sigma^2(\beta_r)} \tag{28}$$

характеризует ту долю дисперсии зависимой переменной β_r , которая не может быть объяснена с помощью регрессоров β_s, \dots, β_t , так как опирается на частную (оригинальную) информацию, не требующуюся при формировании ответов на вопросы q_s, \dots, q_t и необходимую только для нахождения правильного ответа на вопрос q_r .

Если β_s, \dots, β_t - система взаимосвязанных случайных величин, то количество информации, содержащейся в

векторе $[\beta_s, \dots, \beta_t]$ о величине β_r определяется выражением [6]:

$$I(\beta_r \leftrightarrow (\beta_s, \dots, \beta_t)) = -\frac{1}{2} \ln(1 - R^2). \quad (29)$$

Эта информация носит взаимный характер, т.е. не является оригинальной, присущей только вопросу q_r , что должно быть учтено при определении веса w_r . Очевидно, чем ближе к 1 коэффициент корреляции R , тем большим оказывается количество взаимной информации $I(\beta_r \leftrightarrow (\beta_s, \dots, \beta_t))$, т.е. тем большая часть знаний тестируемого, необходимых для ответа на вопрос q_r проверяется при проверке правильности ответов на вопросы q_s, \dots, q_t . Чтобы избежать дублирования оценки знаний и, как следствие, завышения уровня знаний, получаемого при проверке сильнокоррелированной системы вопросов, необходимо уменьшить веса соответствующих вопросов. В частности, при практически полной "функциональности" уравнения регрессии $R^2 \approx 1$, взаимная информация $I(\beta_r \leftrightarrow (\beta_s, \dots, \beta_t)) \rightarrow \infty$, т.е. для ответа на вопрос q_r отсутствует принципиальная необходимость в привлечении каких-либо новых знаний относительно тех, что требовались при ответах на вопросы q_s, \dots, q_t , и, очевидно, вес $w_r \rightarrow 0$. Прямо противоположная ситуация характеризуется тем, что $R^2 = 0$, соответственно $I(\beta_r \leftrightarrow (\beta_s, \dots, \beta_t)) = 0$, т.е. вопрос q_r и система вопросов (q_s, \dots, q_t) взаимонезависимы, поэтому вес w_r определяется, как это упоминалось выше, исключительно сложностью контрольного вопроса.

К сожалению, формула (29) справедлива только при нормальности элементов вектора $[\beta_s, \dots, \beta_t]$, в более общем случае, при отличии распределения от нормального или отсутствии сведений о виде распределения, соотношение для расчета взаимной информации $I(\beta_r \leftrightarrow (\beta_s, \dots, \beta_t))$ не известно.

В подобной ситуации для оценивания уровня оригинальной информации, необходимой для ответа на вопрос q_r и, соответственно, нахождения веса w_r этого вопроса можно применить следующий подход. Будем рассматривать данные тестирования группы из K человек, полученные на тесте, содержащем N контрольных вопросов, как результат K -кратных испытаний N -мерной векторной случайной величины $B = [\beta_1, \dots, \beta_N]$. Рассчитав по матрице исходных экспериментально полученных данных $[\beta_{ij}]$ выборочную корреляционную матрицу $[\rho_{ij}]$, выделим совокупность независимых вопросов, статистически не связанных друг с другом и с остальными вопросами теста, а также группы (системы) взаимосвязанных вопросов, характеризующиеся наличием статистически существенных взаимосвязей между элементами внутри группы. Для послед-

них, ставя в соответствие каждому вопросу q_r, q_s, \dots, q_t группы случайную величину $\beta_r, \beta_s, \dots, \beta_t$, строим регрессионные модели вида (26), в которых в качестве зависимой переменной поочередно выступает каждый элемент вектора $[\beta_r, \beta_s, \dots, \beta_t]$. Оценив для каждой модели пары дисперсий $(\sigma_{\xi_r}^2, \sigma^2(\beta_r)), (\sigma_{\xi_s}^2, \sigma^2(\beta_s)), \dots, (\sigma_{\xi_t}^2, \sigma^2(\beta_t))$, определим в соответствии с выражением (28) долю оригинальной информации в каждом контрольном вопросе коэффициентами $c_r = \sigma_{\xi_r}^2 / \sigma^2(\beta_r)$, $c_s = \sigma_{\xi_s}^2 / \sigma^2(\beta_s)$, ..., $c_t = \sigma_{\xi_t}^2 / \sigma^2(\beta_t)$. Используя эти коэффициенты в формуле (23), положив $c = 1$ для совокупности независимых вопросов, определим информативность I_i , $i = \overline{1, N}$ каждого вопроса теста, а затем рассчитаем веса w_i :

$$w_i = I_i / \sum_{i=1}^N I_i. \quad (30)$$

Отметим, что стоящая в знаменателе формулы (30) сумма $\sum_{i=1}^N I_i < I$, так как не учитывает взаимную информацию вида (29) для групп взаимосвязанных вопросов. Изложенная выше методика оценивания весов w_i , $i = \overline{1, N}$, носит эвристический характер, в связи с чем эффективность и целесообразность ее применения устанавливается только в ходе непосредственных апробаций на реальных данных с привлечением существующих методов проверки и анализа качества тестов [1]. Следует отметить, что уровень субъективизма при оценивании значений \tilde{p}_i и R^2 по экспериментальным данным весьма высок и существенно зависит от среднего уровня L подготовки тестируемой группы. Так, низкий общий уровень знаний группы тестируемых ведет к тому, что частота правильных ответов практически на все контрольные вопросы стремится к $\tilde{f} = 1/m$, а степень статистической взаимосвязи этих вопросов - к 0. Для группы с высоким уровнем подготовки \tilde{f} стремится к 1, поэтому понятие статистической взаимосвязи между вопросами теряет смысл.

Наличие у группы тестируемых определенного среднего уровня знаний L приводит к исключению ряда возможных вариантов ответов, повышая вероятность правильного ответа на i -ое контрольное задание: $p_i' = 1/(m - u_i)$, где u_i - число исключенных вариантов.

В итоге приходим к необходимости построения модели теста закрытого типа, связывающей три переменные: средний уровень знаний L , информативность q_i -го контрольного задания I_i , вероятность правильного ответа p_i . Обычно такая модель описывается некоторой сигмоидной

функцией, например, $p_i = 1/(1 + a_1 e^{-(a_{2i} I_i - L)})$, параметры которой оцениваются по результатам обработки экспериментальных данных.

Очевидно, что при $L = 0$ значение $p_i = 1/m$ для любого контрольного задания, что позволяет определить коэффициент a_1 , а при достаточно большом среднем уровне L $p_i \rightarrow 1$. Параметры a_{2i} подбираются таким образом, чтобы совокупность найденных для данного среднего уровня L значений p_i , $i = \overline{1, N}$ после выполнения всех необходимых пересчетов привела бы к тому же значению L среднего по группе тестируемых уровня знаний, т.е. подбор параметров a_{2i} позволяет замкнуть процедуру построения моделей $p_i(a_1, a_{2i}, I_i, L)$, гарантируя тем самым надежность получаемых результатов тестового контроля. Можно более дифференцированно подойти к этой задаче, выделив среди тестируемых несколько групп с различными средними уровнями знаний, что позволит более точно провести параметризацию моделей $p_i(a_1, a_{2i}, I_i, L)$.

Кроме того, с ростом базы данных по результатам тестирования появляются дополнительные возможности по ретроспективному уточнению параметров моделей путем совместной обработки старых и новых данных СТК.

ВЫВОДЫ

Применение закрытых тестов позволяет построить быстрые процедуры оценки уровня знаний с полностью автоматизированным циклом обработки результатов тестового контроля. Однако обосновать надежность получаемых при этом количественных оценок уровня знаний сложно из-за эвристического характера применяемых методик обработки тестовых данных. Введение процедуры подбора весовых коэффициентов для индивидуального учета сложности контрольных вопросов с последующей перепроверкой и уточнением этих коэффициентов в ходе ретроспективной обработки данных тестового контроля позволяет избежать субъективизма в получаемых результатах и снять связанные с этим ограничения в применении СТК.

ПЕРЕЧЕНЬ ССЫЛОК

1. Аванесов В.С. Тесты в социологическом исследовании. -М.: Наука, 1982.-200с.
2. Дружинин В.Н. Экспериментальная психология. -М.: ИНФРА-М, 1997.-256с.
3. Архипов А.Е. Формализация и моделирование тестовых систем закрытого типа. // Системні технології. Системи керування, контролю та технічної діагностики: Збірн. наук. праць. - Вип. 5. - Дніпропетровськ, 1999. -С. 29-33.
4. Архипов А.Е., Архипова С.А. Исключение влияния случайных факторов на результаты тестового контроля. // Системні технології. Системи керування, контролю та технічної діагностики: Збірн. наук. праць. - Вип. 5. - Дніпропетровськ, 1999. -С. 34-37.
5. Колемаев В.А., Калинина В.Н. Теория вероятностей и математическая статистика. -М.: ИНФРА-М, 1997.-302с.
6. Антомонов Ю.Г., Красникова Л.И., Чораян О.Г. Методы математической бтологии.-К.: Вища шк., 1982.- 240с.

УДК 681.32

ОБРАБОТКА СТАТИСТИЧЕСКИХ ДАННЫХ И ОЦЕНКА ЗНАЧИМОСТИ ВЛИЯНИЯ ФАКТОРОВ

А.С.Дулесов

Рассматриваются возможности формирования математических моделей при помощи нейронных сетей и оценка значимости влияния факторов на исследуемые параметры системы

The possibility of forming mathematical models on the basis of neural networks and valuation of significance of influencing factors to the searching parameters of system are considered

Современный этап научно-технического прогресса характеризуется функционированием многих технических, экономических и других систем при комплексном воздействии различных факторов, оказывающих влияние на параметры этих систем. Количественная оценка такого влияния может быть выражена посредством математических моделей

$$Y_j = f(X_1, \dots, X_k, \dots, X_m), j = 1, 2, \dots, n, \quad (1)$$

где Y_j - j -й выходной параметр (ВП) системы; X_k - k -й воздействующий фактор (ВФ); n , m - число учитываемых ВП и ВФ соответственно.

Выражение (1) отражает количественную меру взаимосвязи системы и внешней среды. Часто, ввиду большой сложности протекающих процессов (1) получают в виде уравнения квазилинейной регрессии

$$Y_j = \sum_{l=1}^L a_l z_l, j = 1, 2, \dots, n, \quad (2)$$

где a_l - l -й коэффициент регрессии, z_l - l -й условный фактор, представляющий собой как собственно сам ВФ (например, $z_l = X_1$), так и его функциональное преобра-

зование (например, $z_l = \lg X_1$) и/или сочетание ВФ (например, $z_l = X_1 X_2$), причем $z_l = 1$.

Говоря о необходимости построения (2) следует отметить, что ВФ, случайные величины, подчиняющиеся нормальному закону распределения (для линейной регрессии - дополнительно должны быть независимыми). Формирование (2) непосредственно связано с применением моделей регрессии и планированием эксперимента. Однако, все более жесткие требования предъявляются со стороны практики анализа и обработки информации. Поэтому работы по совершенствованию аналитических моделей продолжаются и в их число вошли нейросетевые.

Рассмотрим некоторые положительные качества, присущие НС (перцептрон) при формировании (1). Как известно из теории математической статистики, "камнем преткновения" является наличие среди всей общности ВФ случайных и фиктивных факторов.

Случайные воздействующие факторы ϵ колеблются по величине, могут быть положительными и отрицательными, нормально и ненормально распределенными, зависимыми, либо независимыми и на первый взгляд не видны "невооруженным глазом". Пренебрежение ϵ может привести к потере адекватности модели отражающей поведение системы. Их "малозначительность" воздействия на систему на одних участках времени может обернуться "значимостью" на других.

При формировании математической модели вида (2) приходится полагаться на условие: $\epsilon \sim N(0, \sigma^2)$, т.е. ϵ - нормально распределенная случайная величина со средним 0 и дисперсией σ^2 . Фактически считается, что исследователь имеет дело с так называемым "шумом", который является, как бы помехой и это не всегда оправдано.

Использование НС в практике прогнозирования, анализа или управления не означает принятия предыдущего условия. Дело в том, что эффективность обучения НС зависит не только от алгоритма обучения, но и от обучающей выборки [1]. Необходимо сформировать выборку такого количественного и качественного состава из числа всевозможных факторов, которые позволили бы получить после обучения НС высокий уровень адекватности нейросетевой модели по отношению к существующей природе воздействия факторов на исследуемые параметры.

Одним из распространенных подходов в решении данной задачи является следующий. По ходу обучения в выборку добавляются или из неё извлекаются факторы, в зависимости от реакции сети при её обучении. Эффективность формирования выборки будет напрямую зависеть от применяемых в данном случае методов. Одним из них предлагается считать метод экспертных оценок. Эксперт для каждого обучающего примера сопоставляет величину получившейся погрешности (как разность между выходным значением НС и заданным в примере) с желаемой. Если соотношение не в пользу эксперта, то он принимает решение о расширении выборки. От знаний эксперта будет зависеть процесс обучения, т.е. форми-

рование выборки позволит алгоритму обучения эффективно "выйти" из локальных минимумов функции ошибки.

Замечание. Может возникнуть соблазн включить в модель как можно больше объясняющих факторов с целью получить лучшую подгонку. Однако следует иметь в виду, что точность оценок при этом уменьшается (погрешность возрастает), поскольку начинает проявляться сильная положительная корреляция между факторами. При повышении положительной корреляции ухудшается способность НС к обучению.

Фиктивные факторы могут принимать всего два значения или, в более общей ситуации, дискретное множество значений. Необходимость в рассмотрении таких переменных возникает довольно часто в тех случаях, когда требуется принимать во внимание какой-либо качественный признак. Фиктивные факторы, несмотря на свою внешнюю простоту, являются весьма гибким инструментом при исследовании влияния качественных признаков. Их внедрение в практику моделирования позволяет строить и оценивать так называемые кусочно-линейные модели, которые можно применять для исследования структурных изменений исследуемой системы. Если при формировании регрессионной зависимости (2) для учета в ней фиктивных факторов требуются дополнительные условия [2], то в них нет необходимости при моделировании (1) с помощью НС.

Таким образом, использование НС при формировании математической модели (1) не столь чувствительно к понятиям "случайность" и "фиктивность" переменных, что нельзя сказать о формировании (2) посредством регрессионных моделей.

Обозначим еще одно из преимуществ нейросетевого моделирования. В частности речь пойдет об анализе возможного поведения системы через предварительно формируемые связанные и одновременные уравнения.

При моделировании достаточно сложных (например, экономических) систем часто приходится вводить не одно, а несколько связанных между собой уравнений, т.е. описывать модель системой одновременных уравнений

$$\begin{aligned} Y &= f(a, X); \\ X &= g(b, Y), \end{aligned} \quad (3)$$

где a, b - коэффициенты. Системы такого вида могут использоваться при решении задач производства и отражают взаимосвязь производств через производственные функции предприятий.

Для реализации задачи моделирования одновременных уравнений посредством НС, (3) перепишем в виде:

$$\begin{aligned} Y &= f(X, Z_i), \quad i = 1, 2, \dots, M, \\ X &= g(Y, S_j), \quad j = 1, 2, \dots, T, \end{aligned} \quad (4)$$

где Z_i, S_j - факторы воздействия.

Решение (4) видится в совместном использовании двух НС (перцептронов) по схеме на рис. 1. обучение каждой НС осуществляется отдельно, при условии, что все

параметры известны. Результат оценки X и Y посредством тестирования НС будет зависеть от факторов воздействия Z и S , которые в соответствии с моделированием внешних условий будут подаваться на входы нейронных сетей.

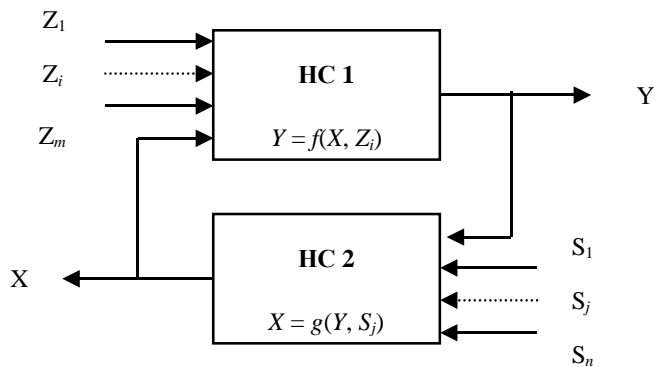


Рисунок 1 - Структура нейросетевой модели решения связанных уравнений

Что касается моделирования на основе построения и последующего решения одновременных уравнений, то в простейшей форме их можно представить в виде системы:

$$\begin{aligned} X &= f(Y, Z, \epsilon); \\ X &= g(Y, S, u), \end{aligned} \quad (5)$$

где ϵ , u - случайные факторы.

Система уравнений данного вида нашла свое применение при оценке взаимодействия спроса и предложения на рынке товаров и услуг.

В модели (5) Y и X определяются одновременно и поэтому обе эти переменные считаются эндогенными. В отличие от них, Z и S являются экзогенными переменными. Подчеркнем, что разделение переменных на эндогенные и экзогенные определяется содержательной стороной модели. Предполагается, что в каждом уравнении экзогенные переменные некоррелированы с ошибкой. Если данное предположение подтверждается, то очевидно применение МНК-оценки определения регрессоров.

Наличие множества случайных факторов ϵ , u и их коррелированность с экзогенными переменными лишают исследователя возможности формирования (5) в явном виде. Применение же НС позволяет формировать модель при условии отсутствия положительной некоррелированности ϵ и Z , u и S . Для более успешного использования (5) необходимо совместное применение двух НС (по каждому из уравнений). После их обучения баланс между моделями в (5) достигается посредством управления параметром X каждой НС по схеме рис. 2.

Построение математических моделей вида (1) и (2) возможно, как правило, по результатам испытаний или слежения систем. Поэтому в рамках математической статистики сформировались такие направления, как регрессионный анализ и теория планирования эксперимента, а так-

же нейросетевое моделирование. При их использовании выявляется ряд вопросов, среди которых, в частности, следует выделить:

1. Как повысить объективность оценки степени однородности результатов, отражающих поведение системы при различных объемах их выборок?
2. Какова количественная оценка влияния того или иного ВФ на ВП системы?

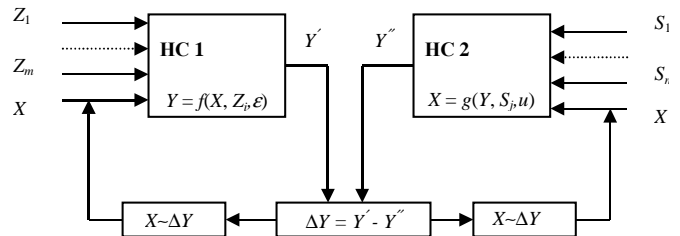


Рисунок 2 - Структура нейросетевой модели решения одновременных уравнений

Решение первого вопроса связано с предварительным анализом полученных результатов. В частности они оценивались по критерию Кохрена [3]

$$G = \max_{i=1}^N \left(\frac{D_i}{\sum_{i=1}^N D_i} \right), \quad (6)$$

где D_i - дисперсия j -го ВП системы при i -м сочетании ВФ; N - число сочетаний (примеров для обучения НС). Величина D_i в свою очередь определяется по формуле

$$D_i = \sum_{k=1}^m (y_{ki} - y_i)^2 / (n - 1), \quad i = 1, 2, \dots, N, \quad (7)$$

где $y_i = (y_{1i} + y_{2i} + \dots + y_{ni}) / n$ - среднее значение j -го ВП системы при i -м сочетании ВФ; n - объем выборки.

Выражения (6) и (7) позволяют судить о вероятности отсутствия случайных ошибок в результате сочетания ВФ. Полученная по (6) величина G сопоставляется с табличным значением [3] доверительной вероятности α однородности ВП системы, которая определяется исходя из наличия степеней свободы $\nu_1 = n - 1$ и $\nu_2 = N$.

Если же объем выборок n_i систем в каждом i -м испытании непостоянен (для НС имеются пробелы в примерах для обучения), то применение формулы (7) становится невозможным, поскольку она справедлива только при $n = n_i = \text{const}$.

Оценка значимости (решение второго вопроса) влияния ВФ на ВП исследуемых систем при наличии математических моделей (1) и (2) осуществляется с использованием критерия Стьюдента [4]. Если же (1) имеет общий нелинейный вид или в (2) условные факторы z_j не

совпадают с ВФ, то оценка значимости влияния собственно того или иного ВФ на ВП системы резко осложняется.

Признавая возможность применения критерия Кохрена для оценки однородности результатов наблюдений с использованием формулы (6), выражение (7) для расчета дисперсий D_i несколько модернизируется [5].

$$D_i = \sum_{k=1}^{n_i} (y_{ki} - y_i)^2 / (n_i - 1), \quad n_i > 1, \quad i = 1, 2, \dots, N. \quad (8)$$

Среднее значение j -го ВП системы при i -м сочетании ВФ вычисляется по формуле

$$y_i = \sum_{k=1}^{n_i} y_{ki} / n_i, \quad i = 1, 2, \dots, N. \quad (9)$$

величина степеней свободы

$$v_1 = \sum_{i=1}^N n_i / (N - 1), \quad v_2 = N.$$

Для оценки однородности результатов наблюдений полученных при использовании нейросетевых моделей, кроме (8) и (9) можно воспользоваться и (6), (7), но при условии заполнения пробелов по методу [6].

Второй вопрос, как отмечалось ранее, связан с задачей определения значимости влияния ВФ на ВП системы и предполагает наличие математической модели (1) или (2). Для оценки значимости факторов в (1) используется метод полунормированных градиентов, а в (2) - метод повторяемости факторов (см. например, [5]). Что касается использования этих методов в случае применения нейросетевого моделирования, то они не приемлемы, так как с помощью НС невозможно получить математическую модель в явном виде. С целью решения данной проблемы представим далее один из алгоритмов оценки влияния ВФ на ПФ в случае нейросетевого моделирования.

Пусть имеется обученная НС (персептрон) на входе которой ВФ ($x_k, k = 1, 2, \dots, m$), на выходе ПФ ($y_j, j = 1, 2, \dots, n$).

Обучающая выборка - таблица данных, в которой строки соответствуют объектам (примерам), а столбцы - факторам x_k и параметрам y_j .

Алгоритм оценки влияния факторов:

1. По каждому из столбцов определяется размах вариации для $x_k, R_k = x_{k\max} - x_{k\min}$ и $y_j, R_j = y_{j\max} - y_{j\min}$.

2. На вход НС подаются центральные значения $x_{k0} = 0,5(x_{k\max} + x_{k\min})$.

3. Варьируя первым фактором x_1 (при $x_2, \dots, x_m = \text{const}$) в пределах размаха, на входе НС фиксируются

значения $\Delta y_{j1} = y_{j1\max} - y_{j1\min}$, т.е. $\Delta y_{11}, \Delta y_{21}, \dots, \Delta y_{n1}$.

4. Определяется вектор значений $[\eta_{11}, \eta_{21}, \dots, \eta_{j1}, \dots, \eta_{n1}]$, позволяющий оценить уровень влияния фактора x_1 на каждый из выходных параметров, $\eta_{j1} = \Delta y_{j1} / R_j$.

5. Для оценки уровня влияния следующего $k+1$ фактора процесс начинается с п.3. Если оценка влияния m -го фактора завершена, переходим к следующему пункту.

6. Определяются по каждому из факторов $\eta_{ji} = \sum_{j=1}^n \eta_{ji}$, ($j = 1, 2, \dots, m$), которые сопоставляются между собой. Тем самым оценивается влияние каждого из ВФ на весь спектр ВП.

В заключение следует отметить, что построение математических моделей отражающих количественную меру взаимосвязи исследуемой системы и внешней среды требует разработки и применения все более совершенных моделей, одной из которых является нейросетевая. Обладая рядом положительных качеств (например, способностью аппроксимировать с высоким уровнем значимости функции системы), она позволяет исследователю при оценке поведения исследуемой системы учитывать качественное и количественное влияние факторов на исследуемые параметры системы, формировать систему связанных между собой уравнений и осуществлять поиск точки равновесия в одновременных уравнениях.

Не смотря на кажущееся отсутствие внутреннего аппарата оценки и обоснования достоверности полученных результатов, как это делается при использовании методов прикладной статистики, нейросетевое моделирование вполне конкурентоспособно регрессионным моделям. Предложенные методы и алгоритм дают возможность улучшать степень однородности результатов поведения исследуемой системы и оценивать уровень влияния каждого из факторов на выходные параметры системы.

ПЕРЕЧЕНЬ ССЫЛОК

1. Беркульцев М.В., Дьячук А.К., Оркин С.Д. Применение генетического алгоритма к построению минимально допустимой обучающей выборки для нейросетевой системы принятия решений // Изв. РАН. Теория и системы управления. 1999, №5. С. 172-176.
2. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Уч. пособие. - М.: Дело, 1998. - 248 с.
3. Статистические методы в инженерных исследованиях. Лабораторный практикум / Под ред. Г.К. Круга. - М.: Высш. шк., 1983. - 208 с.
4. Афифи А., Эйзен С. Статистический анализ: Подход с использованием ЭВМ / Пер. с англ. - М.: Мир, 1982. - 488 с.
5. Критенко М.И., Таранцев А.А., Щербаков Ю.Г. О проблемах предварительного анализа результатов многофакторных испытаний и оценки значимости влияния факторов // Изв. РАН. Теория и системы управления. - 1996. - №1. - С.174-176.
6. Горбань А.Н., Россиев А.А. Нейросетевой итерационный метод главных кривых для данных с пробелами // Изв. РАН. Теория и системы управления. 1999, №5. С. 62-167.

АКТИВНАЯ ЭКСПЕРТНАЯ СИСТЕМА МОНИТОРИНГА И УПРАВЛЕНИЯ РЕМОНТОМ АВИАЦИОННЫХ ГАЗОТУРБИННЫХ ДВИГАТЕЛЕЙ

С.В.Жернаков

В статье рассматривается применение активных экспертных систем с нейросетевыми базами знаний для процесса информационного мониторинга параметров авиационного газотурбинного двигателя. Предлагается оригинальный подход к определению тренда параметров газотурбинного двигателя на базе нейросетей, реализующих динамическую базу знаний активной экспертной системы в процессе эксплуатации двигателя. Предложена методика по разработке подобных нейросетевых баз знаний.

The article considers active expert systems that include neuronet knowledge bases. The systems are used in information monitoring of aviation gas turbine engines parameters. The described method is new in a field of determination of the aviation gas turbine engines parameters trend. It includes the use of dynamic knowledge bases of active expert systems during engines exploitation. The method of neuronet knowledge base developing is proposed.

ВВЕДЕНИЕ

Современные экспертные системы (ЭС) постоянно динамично развиваются. Сегодня трудно представить себе какую либо отрасль промышленного производства без участия ЭС. С одной стороны это экономически выгодно, а с другой существенно повышается качество управления.

В сложившихся условиях наряду с традиционной классификацией ЭС: статические и динамические, в последнее время в отечественной и зарубежной периодике выделяют класс - активных ЭС [1-5]. От динамических ЭС (ЭС реального времени) (ЭС РВ), активные ЭС отличаются участием человеческого фактора в контуре управления. Так, если в ЭС РВ на долю человека-оператора в процессе принятия решений может приходиться 30 ÷ 50%, то в активных ЭС этот процент участия сводится к минимуму 5 ÷ 10% или совсем исключается [2,3].

Совершенно очевидно, что при таком подходе к организации активных ЭС и внедрения ее в процесс управления необходимо учесть ряд факторов, способствующих качественному улучшению ее функционирования [4-7]:

- наличие тесного информационного взаимодействия управляющей системы с окружающей внешней средой с применением специально организованных информационных каналов связи;
- принципиальная открытость системы с целью повышения ее интеллектуальности и совершенствования собственного поведения;
- наличие механизмов прогноза изменения окружающей среды и поведения системы;
- построение управляющей системы на основе многоуровневой иерархической структуры, удовлетворяющей следующему правилу: по мере повышения ранга иерархии про-

исходит повышение интеллектуальности системы и снижение требований к ее точности и наоборот;

- сохраняемость функционирования при частичном разрыве связей или потере управляющих воздействий от высших уровней иерархии управляющей системы.

Иными словами, активная ЭС должна быть легко перестраиваемой (адаптируемой) к внешним изменениям, для чего в ее составе необходимо присутствие следующих соподчиненных уровней [5-7]:

- обучения;
- самоорганизации (перестройки);
- прогноза (регноза) событий (ситуаций);
- работа с базами событий (базами данных) (БД) и знаний (БЗ);
- формирования решений;
- планирования операций по реализации сформированного решения;
- адаптации;
- исполнительный.

При этом, первые пять перечисленных образуют стратегический уровень активной ЭС, остальные выполняют ее тактические функции. Функциональная схема активной ЭС представлена на рисунке 1.

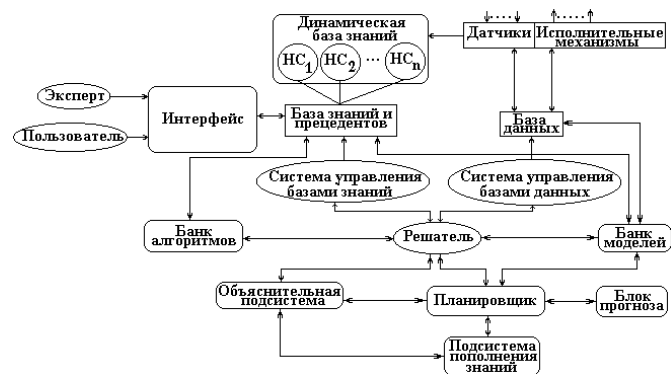


Рисунок 1 - Процесс взаимодействия подсистем активной ЭС

Решатель (машина логического вывода) активной ЭС является комплексным, так как наряду с известными методами и знаниями (логика предикатов, семантические сети, фреймы, продукционный вывод) в активной ЭС могут применяться методы, основанные на мягких вычислениях (нечеткая логика (НЛ), генетические алгоритмы (ГА), нейронные сети (НС), когнитивные сети (КС), вероят-

ностный вывод (ВВ) (эвристики). Комбинация методов и расширение их элементами мягких вычислений с подключением к планировщику банков алгоритмов и моделей увеличивает мобильность вычислительного процесса решателем активной ЭС и как следствие качество принимаемых решений. Имея мощный решатель, активная ЭС сравнительно легко адаптируется к внешней динамической модели, позволяя ставить и решать прямые, обратные и смешанные задачи.

В процессе функционирования сложного динамического объекта (СДО), например, авиационного газотурбинного двигателя (ГТД), подключенная к нему активной ЭС позволяет в темпе реального времени осуществлять моделирование, прогнозирование и оценку эффективности работы силовой установки.

Базы знаний активной ЭС хранят декларативные и процедурные знания. К процедурным относятся базы концептуальных знаний (БКЗ): понятия в виде формул, зависимости, таблицы, процедуры и т.д. К декларативным относятся базы экспертных знаний (БЭЗ), носящие описательный (качественный) характер. При этом БКЗ и БЭЗ тесно взаимодействуют между собой, постоянно осуществляя проверку на непротиворечивость (избыточность) знаний. В процессе взаимодействия с объектом и собственной гетерогенной БЗ активная ЭС осуществляет обучение и самообучение. В режиме сканирования в темпе реального времени проверяются факты и знания. Новая ситуация "образует" прецедент и запоминается в БЗ. Элементы традиционных средств моделирования в активной ЭС осуществляют:

- математическое (имитационное) моделирование ГТД;
- хранение априорных и апостериорных данных в БД активной ЭС (исходная информация и результаты испытаний).

Дополнительная "гибкость" и мобильность БЗ в активной ЭС достигается путем сопряжения моделей искусственного интеллекта и математической модели (ММ) исследуемого ГТД. А для этого у ММ авиационного двигателя необходимо учесть:

- требования к ММ;
- сочетание детерминированной и стохастической моделей;
- механизмы работы с ММ;
- обучение и формализация в среде активной ЭС ММ ГТД с проверкой ее адекватности.

Перечисленное выше, позволяет увеличить точность, достоверность и корректность функционирования активной ЭС.

В процессе мониторинга и управления эксплуатацией ГТД активная ЭС способна полностью контролировать параметры, анализировать (моделировать) текущую ситуацию с прогнозом ее развития в ГТД (информация с датчиков).

Одной из классических задач мониторинга параметров ГТД является разладка (определение тренда контролируемых данных). В общем случае тренд-анализ позволяет контролировать временной ряд, образованный последовательностью значений контролируемых показателей, и определяют наличие тренда: изменений (разладки) в этом ряду. Ценность тренд-анализа в современных активных

ЭС очень велика, так как позволяет выявить дефекты на ранней стадии их развития (даже если значения контролируемых параметров находятся в допустимых пределах).

1. ПОСТАНОВКА ЗАДАЧИ

Обозначим через $x(t)$, $t = 1, 2, \dots, N$ последовательность дискретных наблюдений

$$x(t) = f(t) + \xi(t) \quad (1)$$

на фоне помехи $\xi(t)$ с нулевым средним и дисперсией σ^2 .

В качестве моделей трендов будем использовать множество полиномов

$$f_j(t) = \sum_{s=0}^{j-1} c_{sj} t^s, \quad (j = 1, 2, \dots, n) \quad (2)$$

с неизвестными коэффициентами c_{sj} , где j - индекс вида модели.

При текущем оценивании модель (2) удобно представить в виде

$$f_j(t + \Delta t) = \sum_{s=0}^{j-1} f_j^{(s)}(t) \cdot \Delta t^s / s!, \quad (j = 1, 2, \dots, n), \quad (3)$$

где Δt - время, отсчитываемое от текущего момента времени t ; $f_j^{(s)}(t)$ - s -я производная функции $f_j(t)$. Будем определять значения функции $f_j(t)$ по скользящей выборке наблюдений $x(t - N + 1), x(t - N + 2), \dots, x(t)$ постоянного объема N , что позволяет отслеживать изменение коэффициентов c_{sj} модели (2). Регулярные данные соответствуют наличию определенной закономерности. Нарушение этой закономерности происходит при изменении коэффициентов c_{sj} в (2). Задача состоит в построении нейросетевого обнаружителя (динамическая база знаний активной ЭС), позволяющего в результате обработки наблюдений $x(t)$ установить факты нарушения закономерности и моменты времени появления этих нарушений (трендов).

2. РЕШЕНИЕ ЗАДАЧИ В СРЕДЕ АКТИВНОЙ ЭС

БЗ и прецедентов активной ЭС может хранить следующую информацию:

- оценку случайности расхождения между заданными математическим ожиданием и выборочным средним (параметрические методы, которые требуют знаний априорной информации об объекте (ГТД), обычно среднеквадратичного отклонения исследуемого параметра);
- оценку принадлежности двух выборок одной генераль-

ной совокупности (непараметрические методы, не требующие априорной информации классические критерии: Хальда-Аббе и его модификации [8]);

- анализ тренда контролируемых параметров на базе рекуррентных нейросетей.

Описание классических критериев выявления тренда: Хальда-Аббе; Неймана-Пирсона; модифицированного r -критерия; интегрального S -критерия можно найти, например, в [8].

Автором проведена сравнительная оценка эффективности тренд-анализа нейросетевого и классических критериев. Сравнительное исследование критериев проводилось на основе имитационного моделирования, что позволило осуществить проверку в широком диапазоне изменения, погрешностей измерений и интенсивности проявления тренда. Значение контролируемого параметра равно сумме детерминированной основы и случайной нормально распределенной помехи с дисперсией ξ . Детерминированная составляющая постоянна на интервале $[0, t_0]$, а затем изменяется линейно с темпом $a = \text{tg}(\alpha)$ ($1/c$), (где a - интенсивность тренда). В ходе моделирования значение a варьировалось в диапазоне $[0, 01; 1]$; а значение ξ в диапазоне $[0, 001; 1]$. При моделировании для настройки ММ ГТД использовалась выборочная дисперсия, рассчитанная на интервале стационарности $[0, t_0]$.

Начиная с момента t_0 , вычислялись значения критериев и проверялось наличие тренда. Эффективность критериев оценивалась временем срабатывания критериев от начала тренда τ_0 до момента времени, соответствующего выявлению тренда $\tau_{\text{зап}}$.

3. ВЫБОР НС

Для решения данной задачи необходимо реализовать на базе рекуррентных НС два последовательно соединенных фильтра - низкой частоты (НЧ) и высокой частоты (ВЧ). При этом НЧ-фильтр "пропускает" постоянную составляющую $f_j(t)$ и "срезает" помеху $\xi(t)$, а ВЧ-фильтр пропускает $f_j^{(s)}(t)$ и "срезает" $f_j(t)$ и помеху $\xi(t)$. Реализация НЧ и ВЧ-фильтров на базе рекуррентных НС показана на рисунке 2.

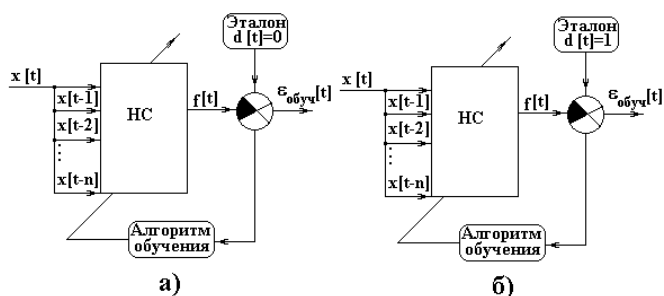


Рисунок 2 - НЧ (а) и ВЧ (б) - фильтры на базе рекуррентных НС

Из теории НС [9-11] известно, что статические архитектуры НС способны аппроксимировать многомерные, нелинейные статические функции. Идентификация динамических систем, наоборот требует модели с соответствующими элементами запоминания. Поэтому статические полноразмерные НС должны расширяться динамическими структурами. Одной из возможностей динамического расширения является добавление внешних фильтров, которые реализуют динамическую модель вне сети. К таким НС с внешней динамикой относятся [9-11]:

- нелинейные модели с обратной связью с выхода;
- нелинейные модели с конечным импульсным ответом;
- нелинейные ортогональные модели базисных функций.

Данные варианты различаются тем, что реализуются соответствующими внешними фильтрами. Структура внешнего фильтра показана на рисунке 3.

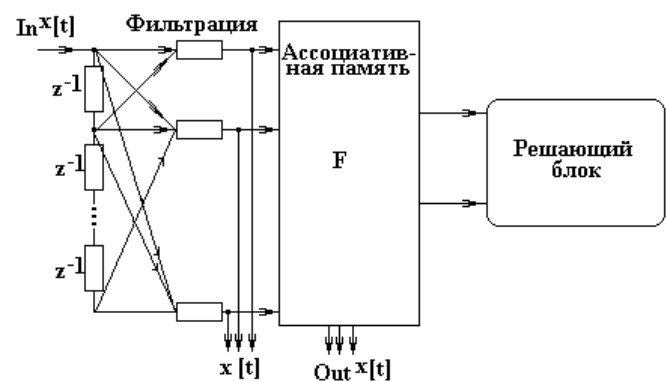


Рисунок 3 - Схема внешнего фильтра

При предварительной обработке считается, что функции $f(t)$ и $\xi(t)$ не коррелированы. Требуется, чтобы вектор $Out^x(t)$ выходных величин фильтра $Out_l^x(t)$, $l = 1, N$, представляющий собой реакцию на внешнее воздействие, приближался к желаемой функции от полезного сигнала:

$$Out^x(t) \approx Ff(t), \tag{4}$$

где $F = (F_l)$ - некоторый векторный оператор, описывающий отображение множества полезных сигналов в выходные сигналы фильтра [12].

В качестве меры приближения $Out^x(t)$ к $Ff(t)$ в общем случае можно выбрать функционал:

$$J = J\{\varphi[Ff(t) - Out^x(t)]\}, \tag{5}$$

где $\varphi[\bullet]$ - некоторая мера векторной функции.

В простейшем случае [12] (рис.3) входной сигнал подается на множество последовательно соединенных функциональных элементов, обладающих задержкой z^{-1} (в синапсах). Их входные величины, представляются в

виде сигналов $In^x(t - kz^{-1})$, $k=1, N$ с весами w_{jk} , формируя вектор оценок полезных сигналов $(x_j(t))$, на основе которого с помощью сети, реализующей матрицу операторов (F_{lj}) , формируется вектор выходных сигналов $(Out_l^x(t))$:

$$Out_l^x(t) = F_{lj} \left[\sum_k W_{jk} In^x(t - kz^{-1}) \right], \quad (6)$$

где $l=1, N$.

Задачей фильтрации является воспроизведение полезного сигнала на фоне шумов и выполнение требуемого преобразования.

Для решения данной задачи необходимо минимизировать среднеквадратичное отклонение оценки полезного сигнала $x_j(t)$ от ожидаемого j - полезного сигнала $f_j(t)$, характеризующего соответствующий полезный результат НС-фильтра, т.е. найти:

$$\min_{W_{jk}} M \left\{ \sum_j \left[f_j(t) - \sum_k W_{jk} In^x(t - kz^{-1}) \right]^2 \right\}, \quad (7)$$

где M - математическое ожидание.

Согласно данному критерию могут быть реализованы классические алгоритмы адаптации фильтра с использованием априорной информации о полезном сигнале и шуме [9-14].

Исходя из выше изложенного, автором, для решения поставленной задачи в качестве статической НС, реализующей НЧ-фильтр, был выбран персептрон; для ВЧ-фильтра - НС РБФ (Радиально-базисная функция).

Сигнал, имеющий N отсчетов $x = [x_1, \dots, x_N]$ может быть аппроксимирован НС с G нейронами в скрытом слое следующими уравнениями:

$$\text{для персептрона } f(t) = \sum_{i=0}^G W_i^{(0)} q(\bar{W}_i^{(h)T} \bar{t}), \quad (8)$$

$$\text{для РБФ } f(t) = \sum_{i=0}^G W_i^{(0)} R_i(\bar{r}, \bar{W}_i^{(h)}), \quad (9)$$

где $q[\bullet]$ - различные типы базисных функций многослойного персептрона, обладающих скалярным аргументом (исходная N -мерная задача аппроксимации путем весовой суперпозиции разлагается на простые скалярные базисные функции; сжатие N -размерного входного пространства до 1-мерного входа $f(\bullet)$ осуществляется путем скалярного произведения $\bar{W}_i^{(h)T} \bar{t}$); $R(\bullet)$ - взвешенные базисные функции РБФ (каждая базисная функция

реализуется отдельным нейроном).

Решающее правило для ансамбля НС, реализующего НЧ и ВЧ фильтры имеет следующий вид:

$$\alpha = \frac{\sum_{i=1}^{\dots} [f_{j+1}(t) - f_j(t)]^2}{t} \geq C, \quad (10)$$

где числитель выражения (10) означает накопление суммы отклонений контролируемых параметров (C - порог срабатывания (чувствительность) НС; при $C = 0$ (нормальный режим работы), при $\alpha \geq C$ (тренд)).

4. ЭКСПЕРИМЕНТЫ И РЕЗУЛЬТАТЫ

В процессе математического моделирования на ансамблевой НС (персептрон - РБФ), реализующих рекуррентные фильтры (динамическая база знаний активной ЭС) в сравнении с классическими критериями обнаружения тренда параметров ГТД были получены следующие результаты (рис.4), которые показывают изменение величины $3\sigma/a$ для 5%-ного уровня значимости, соответствующего вероятности принятия правильного решения $P=0,95$.

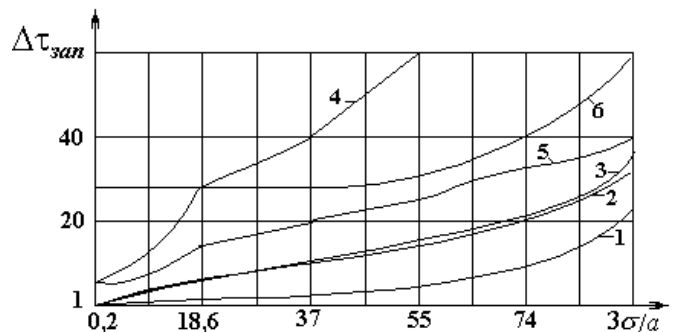


Рисунок 4 - Характеристики критериев тренда для 5% уровня значимости в безразмерных координатах

На рисунке 4 цифрами обозначены: 1 - нейросетевой критерий; 2 - s -критерий; 3 - S'' -критерий; 4 - r' -критерий Хальда-Аббе; 5 - модифицированный r' -критерий; 6 - u -критерий.

Изменяя в обучающей выборке (50 данных): σ_y - среднеквадратичное отклонение ошибки измерения; α - угол наклона тренда; τ_0 - момент появления тренда; было установлено, что нейросетевые критерии эффективнее классических (имеют лучшую чувствительность) распознавания (появления) момента разладки (тренда) контролируемых параметров ГТД, даже в условиях сильных помех (табл.1).

Таблица 1 Сравнительная оценка определения тренда

Критерии	Измеряемая выборка	Время появления тренда (измерения)	Качество распознавания (%)	Качество распознавания тренда при изменении		
				σ_y (%)	α (%)	τ_0 (чувствительность)
Классические	50	7-8	95	70-95	60-95	10-25 (измерений)
Нейросетевые	50	4-5	100	95-100	95-100	3-5 (измерений)

В настоящее время при решении сложных комплексных задач информационного мониторинга и эксплуатации авиационных двигателей могут успешно применяться сложные ансамблевые НС (НА) [15-17], которые по сравнению с обычными полносвязными НС позволяют получить на практике дополнительные преимущества: декомпозиция сложного динамического объекта (СДО) (его систем) на ряд простых объектов (подсистем); НА проще перестраиваются под изменяющиеся внешние условия (в классе адаптивных, самонастраивающихся систем); структура НА может быть оптимизирована под конкретную задачу; быстрдействие и точность НА значительно выше, чем классических полносвязных НС; НА обеспечивают лучшую аппроксимацию кусочно-непрерывных функций.

Перечисленные выше преимущества НА над обычными полносвязными НС, дают возможность их дальнейшего применения при решении задач информационного мониторинга и управления эксплуатацией ГТД.

Методика комплексного мониторинга параметров ГТД на базе рекуррентного НА:

1. Получение обучающей выборки на N режимах нормально работающего двигателя в темпе реального времени;
2. Получение обучающей выборки на N режимах двигателя с трендом параметров в темпе реального времени;
3. Выбор архитектуры рекуррентных НС;
4. Выбор алгоритмов обучения;
5. Обучение, тестирование и распознавание тренда параметров ГТД в режиме реального времени;
6. Мониторинг параметров ГТД рекуррентным НА.
7. Адаптация рекуррентного НА в среде активной ЭС.

Результаты численного моделирования свидетельствуют о возможности решения задач комплексного мониторинга и управления эксплуатацией сложных технических объектов, например, ГТД на базе активных ЭС, позволяющих наряду с классическими критериями выявления тренда параметров, применять качественно новые нейросетевые критерии, расширяющие и дополняющие классические критерии, повышающие достоверность информации при диагностике и контроле параметров и на этапах принятия решений.

Разработка авиационных ГТД V - VI поколений, усложнение их технических систем и подсистем, а также возросшие требования к безопасности полетов привели к необходимости создания интеллектуальных систем способных выполнять некоторые функции человека-эксперта, способствовать в поиске оптимального решения, выдавать

советы и рекомендации в темпе реального времени в процессе комплексного мониторинга и управления эксплуатацией ГТД.

Подход, предложенный автором, апробирован им в среде активной ЭС на базе RTWorks 5.1 совместно с СУБД INFORMIX и нейроимитатором NeuroSolutions 3.0 показавшими высокую эффективность при решении задач комплексного мониторинга и управления эксплуатацией (контроля, диагностики, отладки и прогнозирования параметров) ГТД.

ПЕРЕЧЕНЬ ССЫЛОК

1. Захаров В.Н. Интеллектуальные системы управления: основные понятия и определения. //Теория и системы управления. Известия РАН. №3 1997.
2. Бурков В.Н., Новиков Д.А. Теория активных систем: состояние и перспективы. - М.:СИНТЕГ, 1999. - 127с.
3. Новиков Д.А., Петраков С.Н. Курс теории активных систем.- М.: СИНТЕГ, 1999. - 105с.
4. Попов Э.В., Фомин И.Б., Кисель Е.Б., Шапот М.Д. Статические и динамические экспертные системы. - М.: Финансы и статистика, 1996. - 315с.
5. Васильев В.И. Искусственный интеллект в системах управления и обработки информации. //Вестник УГАТУ. №1, 2000. - С.133-140.
6. Кусимов С.Т., Ильясов Б.Г., Васильев В.И. и др. Управление динамическими системами в условиях неопределенности. - М.:Наука, 1998. - 625с.
7. Кусимов С.Т., Ильясов Б.Г., Васильев В.И. и др. Проблемы проектирования и развития систем автоматического управления и контроля ГТД. - М.:Машиностроение, 1999. - 685с.
8. Епифанов С.В., Кузнецов Б.И., Богаенко И.Н. и др. Синтез систем управления и диагностирования газотурбинных двигателей. - Киев: Техника,1998. - 315с.
9. Глушкин А.И. Теория нейронных сетей.- М.:ИПРЖР, 2000.- 416с.
10. Горбань А.Н. Обучение нейронных сетей. - М.: СП Параграф, 1990. - 125с.
11. Архангельский В.И., Богаенко И.Н., Грабовский Г.Г. и др. Нейронные сети в системах автоматизации. - Киев: Техника, 1999. - 362с.
12. Зозуля Ю.И. Интеллектуальные системы обработки информации на основе нейросетевых технологий. - Уфа: УГАТУ, 2000. - 138с.
13. Бывайков М.Е. Алгоритм обнаружения изменения вида модели при текущем оценивании. //Автоматика и телемеханика, №5,1993. - С.82-93
14. Вишняков А.Н., Цыпкин Я.З. Обнаружение нарушений закономерностей по наблюдаемым данным при наличии помех.//Автоматика и телемеханика, 12, 1991. - С.128-137
15. Жернаков С.В. Об одной методике построения гибридных экспертных систем диагностики и контроля ГТД в условиях неопределенности. // Авиационно-космическая техника и технологии: Сб.науч.тр.- Харьков: ХАИ, 1998. Вып.5. - С.242-249
16. Жернаков С.В. Диагностика и контроль параметров ГТД гибридными экспертными системами. // Авиационно-космическая техника и технологии: Сб.науч.тр.- Харьков: ХАИ, 1999. Вып.9. - С.325-333
17. Жернаков С.В. Контроль и диагностика ГТД гибридными экспертными системами. - Жуковский: ЦАГИ, 1999, вып.2640. - С.215-220

ВИКОРИСТАННЯ МАТЕМАТИЧНИХ МЕТОДІВ ДЛЯ ОПТИМІЗАЦІЇ ПРОЦЕСІВ ВИРОЩУВАННЯ І ПЕРЕРОБКИ СОЇ

А.В.Калініченко

В работе представлена информационная модель биологического процесса симбиотической азотфиксации для сои. Модель позволяет получить высококачественную, экологически чистую продукцию, оптимизировать процесс подбора макро- и микросимбионтов, а также определить значения внешних факторов, при которых растения могут приносить полноценный урожай.

У роботі представлена інформаційна модель біологічного процесу симбіотичної азотфіксації для сої. Модель дозволяє одержати високоякісну, екологічно чисту продукцію, оптимізувати процес підбору макро- і микросимбіонтів, а також визначити значення зовнішніх факторів, при яких рослини можуть приносити повноцінний врожай.

The goal of the work was to investigate the influence of the biological nitrogen-fixation process on receiving high-quality soy-bean products by means of creating of informational model of this process, which in its turn had the aim to find out resulting factors.

У ХХ століття практично в усіх науках, особливо природничо-математичного циклу, ефективно використовується методика моделювання, створення якої підготовлене розвитком людського пізнання і тісно пов'язане з перетворенням математики в універсальну мову і проникненням її в усі галузі суспільного знання. Математичне моделювання в біології на сучасному етапі є одним із найефективніших прийомів дослідження біологічних систем і знаходження найбільш оптимальних форм розв'язання проблем, що дозволяє проаналізувати кількісні характеристики предметів як у статичному, так і динамічному режимах, широко використовуючи при цьому електронно-обчислювальну техніку (ЕОМ) [1,2,3].

Серед пріоритетних практичних завдань моделювання природних систем є використання моделей для підвищення врожайності сільськогосподарських культур, оптимізації процесу їх вирощування, переробки тощо. Виняткову увагу вчених привертає, зокрема, проблема підвищення врожайності бобових культур, високоефективне використання продуктів їх переробки як особливо цінних кормових культур, багатих на білок [4,5,6].

Глобальні екологічні проблеми, енергетична криза, повсюдне забруднення ґрунту, водоймищ, атмосфери вимагають застосування екологічно чистих шляхів постачання рослин зв'язаним азотом. Оптимальною для цих умов є мікробіологічна фіксація атмосферного азоту, яка, крім того, є енергоекономічною, бо здійснюється завдяки енергії Сонця [7, 8, 9].

Важливим прийомом підвищення ефективності симбіотичної азотфіксації є внесення в ґрунт препаратів бульбочкових бактерій. Оскільки ключовою культурою у вирішенні проблеми дефіциту повноцінного білка на півдні України є соя, об'єктом спостереження ми вибрали саме

цю культуру. Проведені дослідження показали, що в районах традиційного вирощування бобових культур приріст врожаю сої від інокуляції складає 2-4 ц/га, а в нових районах вирощування сої, де відсутні спонтанні бульбочкові бактерії, - близько 9 ц/га [4, 10].

Наявність великої кількості сортів сої і штамів азотфіксуючих бактерій призводить до необхідності вирішення оптимізаційних завдань, причому в короткі терміни. Такі проблеми доцільніше розв'язувати шляхом математичного моделювання за допомогою ЕОМ, тобто шляхом проведення математичного експерименту [11, 12, 13, 14].

Метою нашого дослідження була побудова математичної моделі біологічного процесу симбіотичної азотфіксації, яка надасть можливість зробити вибір раціональних стратегій вирощування сої і цілеспрямовано спланувати її сільськогосподарське виробництво, створити відповідні прогнози з селекції високоефективних штамів азотфіксуючих бактерій і рослин в залежності від різних умов і визначити ті умови зовнішнього середовища, за яких рослини сої можуть приносити повноцінний, екологічно чистий, високоякісний врожай.

Модель побудована на основі теорії інформаційних процесів з урахуванням наявності випадковості у природних явищах. Зроблено кореляційний і ентропійний аналіз біологічного процесу.

Біологічна система вивчалася протягом деякого часу як дискретний нестационарний випадковий процес з дискретним часом, що має у загальному випадку α діючих факторів (параметрів) та p кінцевих факторів (параметрів). Як діючі фактори можна розглядати різновиди штамів азотфіксуючих бактерій, сортів сої, ґрунтово-кліматичні умови, рівень азотного та фосфорного живлення ґрунту, вид посіву бобових, методи захисту рослин, тощо. Кінцевими факторами можуть бути приріст урожаю, показники його якості, кількість і маса бульбочок і т.п.

Виходячи з поставлених завдань, у дослідженні було розглянуто вплив одного вхідного фактору (штам бактерій для фіксованого сорту або сорт сої для фіксованого штаму азотфіксуючих бактерій) на вісім вихідних. Усі інші вхідні фактори були взяті як прийнятні для нормальної життєдіяльності, розвитку рослини сої та одержання повноцінного врожаю зерна і високоякісних продуктів його первинної переробки.

У представленій роботі розглянуто наявність одного діючого параметра B (на вибір береться різновид штаму азотфіксуючих бактерій або сорту сої) та восьми кінцевих параметрів $A_k(t, B)$, $k = \overline{1, 8}$; $t \in \{0, t_1, t_2, \dots, t_{N-1}, T\}$ (де A_1 - кількість бульбочок, A_2 - маса бульбочок, A_3 - їх нітрогеназна активність, A_4 - урожай зерна, A_5 - кіль-

кість накопиченого у рослині азоту, A_6 - кількість накопиченого білка у рослині, A_7 - кількість накопиченого азоту у ґрунті, A_8 - показник підвищення родючості ґрунту або родючості наступних культур). Фактори A_k , $k = \overline{1, 8}$ взаємозалежні. Діючий параметр B може знаходитися у m станах із відповідними значеннями b_j , $j = \overline{1, m}$, а кінцеві параметри A_k у n_k станах із значеннями $a_{k i_k}$, $i_k = \overline{1, n_k}$, $k = \overline{1, 8}$, причому всі стани, як параметру B , так і параметрів $A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8$, несумісні між собою.

Зважаючи на вищезазначене, було розглянуто два дискретних простори, які позначимо B та A , елементами яких є точки b_j , $j = \overline{1, m}$ та $\left\{ a_{1 i_1}, a_{2 i_2}, a_{3 i_3}, a_{4 i_4}, a_{5 i_5}, a_{6 i_6}, a_{7 i_7}, a_{8 i_8} \right\}$, $i_k = \overline{1, n_k}$, $k = \overline{1, 8}$, а простір $B \times A$ можемо вважати простором стану випадкового процесу.

Визначено, в якій мірі вхідний параметр визначає вихідні для фіксованих сортів сої та різних штабів азотфіксуючих бактерій (можна розглядати й навпаки: фіксовані штаби бульбочкових бактерій для різних сортів сої). Для цього оцінено інформаційний зміст сигналів кожного часового перетину процесу t_p , $p = \overline{0, N}$ на основі одержаних розподілів ймовірностей дослідних даних на множині можливих значень, яка є обмеженою. З нестационарності багатовимірною динамічного випадкового процесу впливає неможливість знаходження традиційними шляхами інтервалу кореляції процесу, тобто такого $\tau = t_p - t_l$, $p, l = \overline{1, N}$, $t_p < t_l$ для досить малого $\varepsilon > 0$, для якого $\rho_k(\tau/b_j) = \rho(t_p, t_p/b_j) \leq \varepsilon$ при $\Delta t > \tau$, де ρ - нормована кореляційна функція. Тому обчислено приблизні інтервали кореляції перетинів випадкового процесу, що дає можливість кількісно оцінювати інформаційний зміст сигналів за будь-яким з перетинів, що належать до інтервалу, не враховуючи значення станів параметрів на попередніх перетинах, крім початкового перетину інтервалу. Тобто, якщо є необхідність, обчислення проводяться за кожним з перетинів інтервалу або тільки для кінцевого перетину (в межах інтервалу кореляції перетину).

У моменти t_1, t_2, \dots , які входять до інтервалу кореляції $[0, \tau]$ початкового перетину, після одержання значень станів інформативних параметрів, джерело може характеризуватися розподілом ймовірностей:

$$p(b_j; t_p/a_{1 i_1}, a_{2 i_2}, a_{3 i_3}), \quad i_k = \overline{1, n_k}, \quad k = 1, 2, 3.$$

Визначено міру невизначеності вхідної властивості B , тобто ентропію джерела повідомлень $H(B)$, а також ентропії параметрів A_1, A_2, A_3 в моменти t_p , $p = 1, 2, \dots$, які

входять до інтервалу кореляції перетину t_0 , $H(A_1, A_2, A_3)_p$, $H(B, A_1, A_2, A_3)_p$ та умовні ентропії $H(A_1, A_2, A_3/B)_p$, $H(A_1, A_2, A_3/b_j)_p$.

Знайдено середню кількість інформації, що загублена внаслідок шуму (ненадійність) $H(B/A_1, A_2, A_3)_p$ та середню кількість інформації, необхідної для визначення перешкоди (ентропія шуму) $H(A_1, A_2, A_3/B)_p$. Якби шуми були відсутні, то кожному значенню інформативних параметрів відповідало б деяке конкретне значення параметра B і $H(B/A_1, A_2, A_3)_p = 0$, тобто ентропія $H(A_1, A_2, A_3)_p$ давала б повну інформацію про значення параметра B (штам азотфіксуючих бактерій).

Визначено кількість інформації, яка міститься у вихідних параметрах відносно вхідного в моменти t_p , $p = 1, 2, \dots$ $I(B, A_1, A_2, A_3)_p$, що дасть можливість з'ясувати наскільки інформація про значення вхідного фактора B зменшує невизначеність стану системи на даному перетині. Знаючи кількість інформації, яка передається від різних вхідних факторів до вихідних, можна з'ясувати міру залежності між ними.

Якщо момент t_l виходить за межі інтервалу кореляції перетину t_0 , $t_l - t_0 > \tau$, після одержання значень станів інформаційних параметрів A_1, A_2, A_3 джерело повідомлень охарактеризується розподілом ймовірностей $p(b_j/a_{1 i_1}, a_{2 i_2}, a_{3 i_3}; t_p/a_{1 i_1}, a_{2 i_2}, a_{3 i_3}; t_l)$, $i_k = \overline{1, n_k}$, $k = \overline{1, 3}$, тобто ймовірність того, що обраний сорт був оброблений штамом j за умови, що перетини процесу в моменти t_p і t_l входять до інтервалу кореляції τ_p ($t_p - t_l < \tau_p$).

При $t_l = t_N$ розглянуто параметри A_4, A_5, A_6, A_7, A_8 і відповідно розподіл ймовірностей

$$p(a_{1 i_1}, a_{2 i_2}, a_{3 i_3}, a_{4 i_4}, a_{5 i_5}, a_{6 i_6}, a_{7 i_7}, a_{8 i_8}; t_N),$$

$$p(b_j/a_{1 i_1}, a_{2 i_2}, a_{3 i_3}, a_{4 i_4}, a_{5 i_5}, a_{6 i_6}, a_{7 i_7}, a_{8 i_8}; t_p/a_{1 i_1}, a_{2 i_2}, a_{3 i_3}, a_{4 i_4}, a_{5 i_5}, a_{6 i_6}, a_{7 i_7}, a_{8 i_8}; t_N),$$

$$i_k = \overline{1, n_k}, \quad k = \overline{1, 8}.$$

Обчислено ентропії параметрів B, A_1, A_2, A_3 в моменти t_l : $H(A_1, A_2, A_3)_l$, $H(A_1, A_2, A_3; A_1, A_2, A_3)_p, l$, $H(A_1, A_2, A_3/b_j; A_1, A_2, A_3)$, $H(A_1, A_2, A_3/A_1, A_2, A_3; l)_p$, $H(A_1, A_2, A_3/b_j)/A_1, A_2, A_3; l)_p$, $H(A_1, A_2, A_3/A_1, A_2, A_3; p)_l$, $H(A_1, A_2, A_3/(A_1, A_2, A_3; p/b_j))_l$. $H(A_1, A_2, A_3/A_1, A_2, A_3; l)_p$ - середня кількість

інформації, загубленої внаслідок шуму, а $H(A_1, A_2, A_3/A_1, A_2, A_3; p)_I$ - ентропія шуму на проміжку $\Delta t = t_l - t_p < \tau$.

Визначено також кількість інформації, яка міститься в інформаційних параметрах у момент t_l відносно моменту t_p (для визначення (контролю) міри залежності між перетинами): $I(A_1, A_2, A_3; t_l; A_1, A_2, A_3; t_p)$ та $I(A_1, A_2, A_3; t_l; (A_1, A_2, A_3; t_p/b_j))$.

При $t_l = t_N$ розраховано відповідні значення ентропій $H(A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8)_N$ та ін.

Ентропійний аналіз надав можливість з'ясувати, який з часових інтервалів вміщує найбільше корисної інформації і з'ясувати міру залежності між вхідними і вихідними факторами.

Побудована інформаційна модель дає змогу вирішити такі завдання:

- одержати високоякісну соєву сировину для переробки без нітратів, сторонніх хімічних сполук і збагачену додатковою кількістю білка, жиру, незамінних амінокислот;
- дослідити розвиток процесу в динаміці та його вплив на якість переробленої продукції за визначеними часовими перетинами або за періодами розвитку рослин;
- одержати прогноз очікуваних значень урожаю зерна для первинної переробки і якості одержаної продукції, кількості накопиченого азоту в ґрунті тощо.

Таким чином, у результаті проведеного дослідження:

- 1) вивчено та систематизовано дані польових дослідів за кілька років, досліджено симбіотичну систему і виділено вхідні та вихідні, внутрішні і зовнішні фактори. Серед внутрішніх факторів особливе значення має маса бульбочок з високим рівнем нітрогеназної активності, яка в свою чергу впливає на підвищення врожайності і якості рослин;
- 2) створено структурно-логічну схему взаємодії елементів симбіотичної азотфіксуючої системи, яка дає можливість об'єктивно досліджувати потоки інформації в ній;
- 3) на основі теорії інформаційних процесів та представленої структурно-логічної схеми розроблено математичну модель процесу біологічної фіксації азоту соєю, яка враховує наявність випадковості у природних явищах;
- 4) на основі інформаційної моделі отримані ймовірні значення врожаю зерна сої і його біохімічних властивостей, які використано для розробки удосконаленої технології вирощування і первинної переробки;
- 5) розроблені й пройшли апробацію методи підбору комплементарного матеріалу серед мікро- і макросимбіонтів, на основі яких запропонована удосконалена технологія вирощування і переробки сої, яка обмежує використання мінеральних добрив і зменшує вміст нітратів у зерні сої;
- 6) рекомендовано для підвищення показників якості зерна сої застосовувати можливості екологічно чистої симбіотичної системи. Встановлено, що найкращими партнерами по симбіозу є штам бульбочкових бактерій 6346 і

сорти сої Кріпиш, Одеська 124, Аркадія одеська, Юг 40, Харосой;

7) встановлено, що соєвий шрот дослідного варіанта вміщував протеїну 40,45%, жиру 3,01%, в контролі відповідно 39,61% і 2,80%; в екструдаті дослідного варіанта сої протеїну містилося 39,98%, жиру 21,85%, контрольного відповідно - 38,89%, жиру 19,97%. Сума амінокислот дослідного варіанта була вищою по відношенню до контрольного, а щодо вмісту глютамінової кислоти, то він був вищий у 2-3 рази.

Запропонована екологічно чиста технологія застосування можливостей симбіотичної системи є природоохоронним, енергозберігаючим, прибутковим заходом: коефіцієнт енергетичної ефективності виробництва соєвих шротів становив 2,45; екструдатів - 2,65, в контролі він відповідно був 1,33 і 1,51.

Проведене дослідження пропонується використовувати:

- при виборі раціональної стратегії вирощування сої;
- при цілеспрямованому плануванні її сільськогосподарського виробництва;
- при селекції високоефективних штамів азотфіксуючих бактерій або рослин сої в залежності від різних зовнішніх умов;
- при визначенні таких умов зовнішнього середовища, при яких рослини сої зможуть приносити повноцінний урожай;
- при підборі комплементарного матеріалу серед макро- і мікро симбіонтів.

ПЕРЕЛІК ПОСИЛАНЬ

1. Ивахненко А.Г., Юрачковский Ю.П. Моделирование сложных систем по экспериментальным данным. - М.: Радио и связь, 1987. - 120с.
2. Кремянский В.И. Информация и системный подход в биологии. - М.: Знание, 1980. - 64с.
3. Математические методы в биологии. /Под ред. Ю.А.Митропольского. - Киев: Наукова думка, 1977. - 192с.
4. Бабич О.А. Сучасне виробництво та використання сої. - К.: Урожай, 1993. - 429 с.
5. Duffy Z., Chung C., Boast C., Franklin M. A simulation model of biophysico-chemical transformation of nitrogen in tile-drained corn belt soil. // J.Env.Qual. - 1975. - V.4. - №4. - P.477-486.
6. Baldwin J.P., Nyc P.H., Tinker P.B. Uptake of solutes by multiple root system from soil. III. A model for calculating the solute uptake by a randomly dispersed root system developing in a finite volume of soil. // Plant.a.Soil. - 1973. - V.38. - №3. - P.621-635.
7. Гачок В.П., Жохин А.С. Кинетическая модель биохимической азотфиксации. - К.: Би.и. - 1985. - 31с.
8. Гельцер Ф.Ю. Симбиоз с микроорганизмами - основа жизни растений. - М.:Изд-во МСХА, 1990. - 134с.
9. Grant R.F. Simulation of ecological controls on nitrification // Soil Biol. and Biochem. - 1994. - 26, №3. - P.305-315.
10. Мишустин Е.Н., Шильникова В.К. Биологическая фиксация атмосферного азота. - М.: Наука, 1968. - 531с.
11. Торний Дж.Г.М. Математические модели в физиологии растений.: Пер. с. англ. Д.М. Гроздинского. - К.:Наукова думка, 1982. - 310с.
12. Малащенко Ю.Р., Мучник Ф.В., Романовская В.А., Садовников Ю.С. Математические модели и ЭВМ в микробиологической практике. - К.:Наукова думка, 1980. - 195с.
13. Глушков В.М., Иванов В.В., Яненко В.М. Моделирование развивающихся систем. - М.: Наука, 1983. - 350с.
14. Дромашко С.Е., Френкель Г.И., Дубовской Б.О. О возможности исследования генетических систем с помощью информационно-логического подхода.// Генетика. - 1995. - Т.31. - №1. - С.139-143.

ОБ ИНФОРМАЦИОННОЙ ЗНАЧИМОСТИ ЦИФРОВЫХ ИЗОБРАЖЕНИЙ

Е.Н.Кирсанова, М.Г.Садовский

Предложена характеристика цифровых изображений (значимость) и метод ее определения через условную энтропию распределения малых фрагментов изображения относительно ожидаемого. Степень структурированности - мера отличия реального распределения малых фрагментов изображения от "равновесного". Показана величина информационной значимости изображений различной структуры. Обсуждаются приложения этой величины.

New method to estimate images is proposed. Information capacity is the specific entropy of small fragments distribution versus the expected one. It discretizes the random and periodic images from the complex ones. The data obtained for different patterns are discussed.

ВВЕДЕНИЕ

Цифровые изображения (ЦИ) являются широко распространенным объектом при работе с задачами распознавания образов, сцен и т.п. ЦИ являются основой компьютерного представления графики в случае растровых изображений, широко встречающихся как в приложениях, так и в фундаментальных задачах информатики. Изучение тех свойств ЦИ, которые обусловлены его дискретной природой, является важнейшей теоретической задачей, а также представляет прикладной интерес для ряда специальных задач. Так, проблема поиска, выделения и описания тех или иных структур в ЦИ представляет собой актуальную задачу некоторых разделов теории распознавания образов, теории адаптивных процессов, теории алгоритмов [1-3]. При этом под структурой ЦИ могут пониматься различные свойства - от "семантических" (контуры, образы, сцены и т.п.) до тех, которые имеют лишь косвенное отношение к этим первым, и определяются исключительно взаимным расположением тех дискретных единиц, которые составляют изображение, например, пикселей [4, 5]. Очевиден тот факт, что структуры второго типа существенно зависят от способа их определения. При этом не приходится утверждать, что какие-то из выделяемых структур "правильные", а другие - "неправильные", скорее здесь можно вести речь о том, насколько вновь выделяемые структуры коррелируют с представляющими прикладной интерес для исследователей. В этой ситуации возникает задача создания своего рода библиотеки различных структур (и методов из выделения и описания) и каталога взаимосвязей между ними.

Настоящая работа посвящена описанию метода выделения тех структур в ЦИ, которые обусловлены исключительно взаимным расположением пикселей (либо иных элементарных единиц) в них. Отметим сразу, что проблема исследования тех свойств ЦИ, которые обусловлены взаимным расположением пикселей (либо иных элементарных его единиц) оказывается существенно связанной с тем форматом представления ЦИ, который использован в

данном конкретном приложении. Несмотря на то, что сам метод не зависит от конкретного вида представления ЦИ, его программная, а зачастую и алгоритмическая реализация оказываются существенно зависящими от него. Мы будем излагать развиваемый здесь метод на примере простейших ЦИ лишь для наглядности, помня о том, что это никак не ограничивает его общности. Предложенный метод выделения и описания структур в ЦИ

1) опирается только на информацию о частотах сравнительно малых фрагментов фиксированного размера (пиксели либо их сочетания), встречающихся в изучаемом ЦИ,

2) позволяет выделять те структуры, которые определяются информационными и статистическими характеристиками ЦИ,

3) не требует привлечения дополнительной (по отношению к знанию частот отдельных фрагментов) информации,

4) позволяет, как выделять информационные структуры в обширных ЦИ (например, в аэрокосмических снимках), так и сравнивать между собой ЦИ.

Метод основан на вычислении условной энтропии реального распределения всех фрагментов фиксированного размера относительно их ожидаемого распределения; впредь будем называть такую величину (информационной) ценностью. Под ожидаемым распределением фрагментов (пикселей либо их комбинаций) понимается такое распределение фрагментов заданного размера, которое можно получить, опираясь только на распределения аналогичных фрагментов меньшего размера. Подчеркнем также, что развитый здесь метод не может непосредственно выделять такие структуры ЦИ, как контур, либо иные структуры "семантического" ряда. Сказанное не означает, однако, что развитый здесь метод не имеет никакого отношения к поиску и выделению "традиционных" структур в ЦИ. Условия применения развитого здесь метода к данному классу задач и тип тех ЦИ (например, текстуры), для которых информационные и статистические характеристики будут мощным дополнительным методом выявления этих "традиционных" структур в ЦИ требует специального анализа. Перейдем к строгим утверждениям и точным формулировкам.

ОПИСАНИЕ МЕТОДА

Рассмотрим ЦИ размером $N \times M$ элементов. Точнее, пусть имеется решетка, в узлах которой могут стоять символы, являющиеся элементами какого-либо алфавита. Выделим в ней область размером $N \times M$. Будем рассматривать только такие области, для которых в каждом узле решетки находится какой-нибудь элемент (связные области). Любой связный квадрат размером $q \times q$ узлов, $1 \leq q \leq \min\{N, M\}$ встречающийся в ней будем называть

смальтой (размера q). Будем называть верхний левый (для определенности) узел смальты ключом; положение смальты будет определяться положением (координатами) ключа внутри рассматриваемой области. Понятно, что ключ может пробежать $(N - q + 1) \times (M - q + 1)$ положение - часть из них окажутся недоступными из-за конечности размера мозаики. Совокупность всех смальт данного размера, встречающихся в рассматриваемом ЦИ, будем называть q -носителем. Сопоставив каждой смальте из q -носителя число ее копий в изучаемом ЦИ, получим конечную мозаику M_q (размера q). Данная конструкция является конечным объектом и требует для ее исследования соответствующих подходов и техники.

Переход к частотам в мозаике означает переход от конечного ЦИ к ансамблю всех изображений, порождающих данную мозаику. Не обсуждая здесь далее различия между конечной и частотной мозаиками, отметим только, что построение частотной мозаики требует использования одного технического преобразования исходного ЦИ: для построения частотной мозаики его необходимо замкнуть в тор. Тогда ключ смальты будет пробегать $N \times M$ положений, а частота смальты будет определяться как отношение числа копий данной смальты, подсчитанного на торе, к общему числу смальт ($= N \times M$).

Очевидна нормировка для частотной мозаики:

$$\sum_s f_s^{(q)} = 1. \quad (1)$$

Здесь индекс s перечисляет смальты, а $f_s^{(q)}$ обозначает частоту этой смальты; верхний индекс указывает на ее размер.

Предложенный в настоящей статье метод выделения и описания структурированности ЦИ опирается на идею сравнения заданного распределения (мозаики) с равновесным. Если в нашем распоряжении имеется равновесное распределение, то относительную энтропию некоторого заданного распределения относительно этого равновесного можно вычислить всегда. Действительно [6, 7, 8], такая энтропия одного распределения Φ относительно другого ("равновесного") Φ^* равна:

$$S = \int \Phi \cdot \ln\left(\frac{\Phi}{\Phi^*}\right) d\mu. \quad (2)$$

Не обсуждая пока вопроса о том, какую именно мозаику следует считать равновесной, заметим, что сравнение ЦИ с помощью вычисления энтропии одной мозаики относительно другой требует, чтобы носитель одной из них ("равновесной") полностью содержал в себе носитель другой. Такое вложение нельзя гарантировать заранее. В нашем случае эта проблема легко преодолевается, что связано с тем, что в качестве "равновесного" распределения будет использоваться мозаика, для которой реализуется гипотеза о наиболее вероятном продолжении смальт.

Определение "равновесной" мозаики тесно связано с за-

дачей перехода от мозаик одного размера к мозаикам другого. Переход от мозаики размера q к мозаике размера t , $1 \leq t < q$ всегда однозначен; для этого частоты мозаики M_q необходимо просуммировать по любой угловой паре строки и столбца; угловой парой будем называть такую, которая пересекается по элементу, расположенному в одной из вершин смальты¹. Обратный переход далеко не всегда однозначен. Условием однозначности перехода от мозаики размера q к мозаике размера q_1 , $q_1 > q$ является единственность существования любой пары смальт размера q , пересекающихся по подсмальте размера $q - 1$. При этом для всякой смальты размера q должно найтись три других смальты того же размера, единственным образом пересекающихся по подсмальте размера $q - 1$; очевидно, что для любой смальты размера $q - 1$ должно найтись шесть уникальных пар смальт размера q для того, чтобы переход от мозаики размера q к мозаике размера q_1 был единственным.

Если переход от мозаики M_q размера q к мозаике M_{q+s} размера $q + s$ не может быть совершен однозначно, тогда вместо одной мозаики (размера $q + s$) возникнет их семейство. В работах [9-12] предложен общий принцип выбора той мозаики большего размера, которую следует считать порожденной (восстановленной) по заданной: среди всех возможных мозаик размера $q + s$ следует выбрать ту, у которой энтропия максимальна. Данный принцип означает, что в этой мозаике содержатся все те смальты (размера $q + s$), которые могут рассматриваться как продолжения смальт исходной мозаики, а их частоты являются наиболее ожидаемыми. Такая восстановленная мозаика не несет в себе никакой дополнительной, априорной информации. Понятно, что в общем случае возможно восстановление мозаик сколь угодно большого размера по заданной. Использование сформулированного выше экстремального принципа позволяет вычислить частоты смальт в восстановленной мозаике явным образом. Рисунок 1 иллюстрирует процедуру восстановления. Мы остановимся здесь на случае восстановления мозаики размера $q+1$:

$$\tilde{f}_s^{(q+1)} = \frac{f_{\langle 1 \rangle}^{(q)} \cdot f_{\langle 2 \rangle}^{(q)} \cdot f_{\langle 3 \rangle}^{(q)} \cdot f_{\langle 4 \rangle}^{(q)}}{(f_{\langle \text{kern} \rangle}^{(q-1)})^3}. \quad (3)$$

Здесь $\tilde{f}_s^{(q+1)}$ - частота восстанавливаемой смальты s' (размера $q+1$), $f_{\langle i \rangle}^{(q)}$ - частота соответствующей смальты, занимающей i -ю (из четырех возможных - см. рис. 1) позицию, $f_{\langle \text{kern} \rangle}^{(q-1)}$ - частота центральной смальты. Для случая $q = 1$, имеем

$$\tilde{f}_s^{(2)} = f_{\langle 1 \rangle}^{(1)} \cdot f_{\langle 2 \rangle}^{(1)} \cdot f_{\langle 3 \rangle}^{(1)} \cdot f_{\langle 4 \rangle}^{(1)}. \quad (4)$$

1.Замыкание ЦИ в тор было необходимо для того, чтобы все эти четыре суммы совпадали между собой.

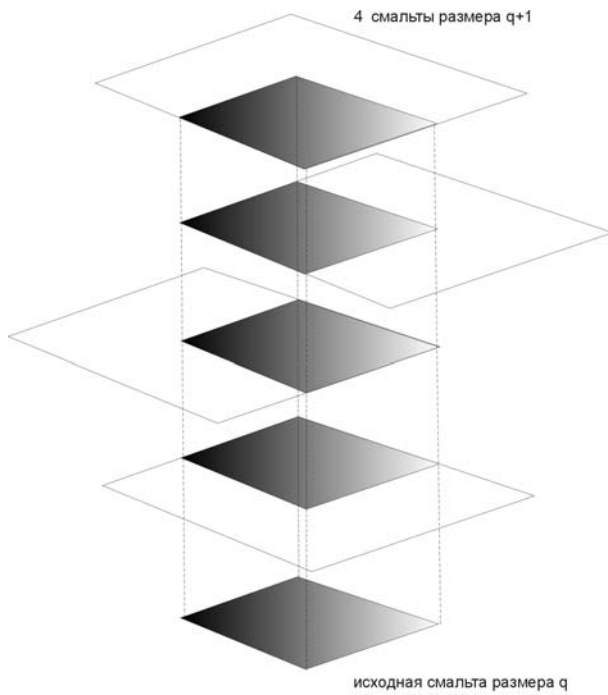


Рисунок 1 - Процедура восстановления смальты

Описав процедуру восстановления мозаики большего размера по заданной, перейдем к описанию метода информационного анализа ЦИ. Итак, в распоряжении исследователя имеется мозаики размера $q = 1, 2, 3, \dots$ и задача исследования состоит в том, чтобы сравнить реальные частоты смальт размера q и те их частоты этого размера, которые можно ожидать, оперируя лишь знанием частот смальт меньшего размера. Для этого необходимо рассмотреть реальную мозаику размера q и восстановленную до того же размера мозаику по мозаикам меньшего размера $t, 1 \leq t < q$. Подчеркнем, что восстановленная согласно (1, 2) мозаика будет содержать все смальты, встречающиеся в реальной мозаике и, возможно, еще некоторые. С точки зрения описываемого метода несущественно, по каким именно мозаикам меньшего размера t проводится восстановление; всюду далее в данной работе, однако, будут рассматриваться мозаики $\tilde{M}_q(q-1)$, восстановленные по реальным M_{q-1} размера $q-1$.

Итак, имея восстановленную мозаику размера q , можно всегда вычислить ее условную энтропию относительно восстановленной:

$$S(\tilde{M}_q(q-1)|M_q) = \sum_{s'} \left(f_s^{(q)} \cdot \ln \frac{f_s^{(q)}}{f_s^{(q)}} \right). \quad (5)$$

Подставляя (1) либо (2) в (3), имеем

$$S(\tilde{M}_q(q-1)|M_q) = \sum_{s'} f_s^{(q)} \cdot \ln f_s^{(q)} - \sum_{j'}^4 \left(\sum_{s'} f_s^{(q)} \cdot \ln f_{(j)}^{(q-1)} \right) + 3 \sum_{s'} f_s^{(q)} \cdot \ln f_{(kern)}^{(q-2)}. \quad (6)$$

Суммируя по "лишним" индексам, соответствующим тем строкам и столбцам смальты размера q , которые не входят в подсмальту размера $q-1$ либо $q-2$, получаем окончательно

$$S(\tilde{M}_q(q-1)|M_q) = 4S_{q-1} - S_q - 3S_{q-2}, \quad (7)$$

где S_i - абсолютная энтропия мозаики размера i . Для случая $q = 2(5)$ имеет вид:

$$S(\tilde{M}_2(1)|M_2) = 4S_1 - S_2. \quad (8)$$

Подчеркнем еще раз, что для вычисления условной энтропии реальной мозаики размера q относительно восстановленной до того же размера не требуется проводить реальное ее восстановление (по мозаикам меньшего размера).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Обсудим подробнее проблему выбора тестовых изображений. Несмотря на то, что метод, развитый выше, применим формально к любым мозаикам, содержательные результаты могут быть получены лишь для таких мозаик, которые достаточно далеки от "равновесных". Не обсуждая в деталях, какие именно мозаики следует считать равновесными и, что именно следует в мозаиках понимать под равновесием, отметим только, что для получения содержательных (нетривиальных) результатов необходимо, чтобы в мозаике не содержалось слишком много уникальных (т.е. имеющих частоту $(N \times M)^{-1}$) смальт. Именно здесь проявляется специфика машинного представления растровых ЦИ: если к построению мозаики подойти формально, то для 256-цветного изображения зачастую уже мозаика размера (2×2) оказывается полностью уникальной. Понятно, что здесь многое зависит от исходного изображения, однако проблему это не снимает.

Другая проблема заключается в том, чтобы выбрать такие ЦИ, которые можно было бы считать эталонными, и для которых, соответственно, получаемые результаты имели достаточно простую интерпретацию. Естественным объектом такого рода является ЦИ, представляющее собой случайное поле с черными и белыми пикселями, распределенными случайно и нескоррелированно, с фиксированной вероятностью (например, с равными вероятностями появления пикселей каждого цвета). Возможны и иные монохромные ЦИ, используемые как эталонные: ЦИ, имеющие простую (например, периодическую) структуру, для них результаты определения информационной значимости могут легко быть получены аналитически и, тем самым, поддаются простой интерпретации. Если такого рода периодические ЦИ подвергнуть случайным мутациям, то можно исследовать влияние случайных изменений на уровень информационной значимости.

Здесь следует сделать одно важное замечание. Говоря об информационной ценности того или иного ЦИ, мы всегда имеем в виду лишь те свойства, которые порождены закономерностями взаимного расположения пикселей в

нем. Информационная ценность не имеет никакого отношения к семантике ЦИ - его смысловому содержанию, значению тех образов, которые на нем изображены и т.п. Изображение, воспринимаемое наблюдателем как имеющее существенное значение и порождающее у него различные реакции в отношении изображенного объекта, может иметь крайне низкую информационную ценность, понимаемую в смысле условной энтропии реальной мозаики относительно восстановленной. Видимо, наиболее ярким примером изображения такого рода следует считать знаменитый "Черный квадрат" К. Малевича.

Еще одним типом эталонных ЦИ могут служить такие,

которые с одной стороны достаточно близки по своим внешним признакам к случайным изображениям, а с другой - имеют достаточно простое и ясное порождающее правило. Одномерным аналогом таких изображений могут служить некоторые разложения действительных чисел в бесконечную непериодическую k -ичную дробь, либо иные последовательности, порождаемые неким преобразованием и обладающие статистическими свойствами, полностью аналогичными свойствам истинно случайных последовательностей. Среди изображений также встречаются аналоги подобных символьных последовательностей - это так называемые фракталы.

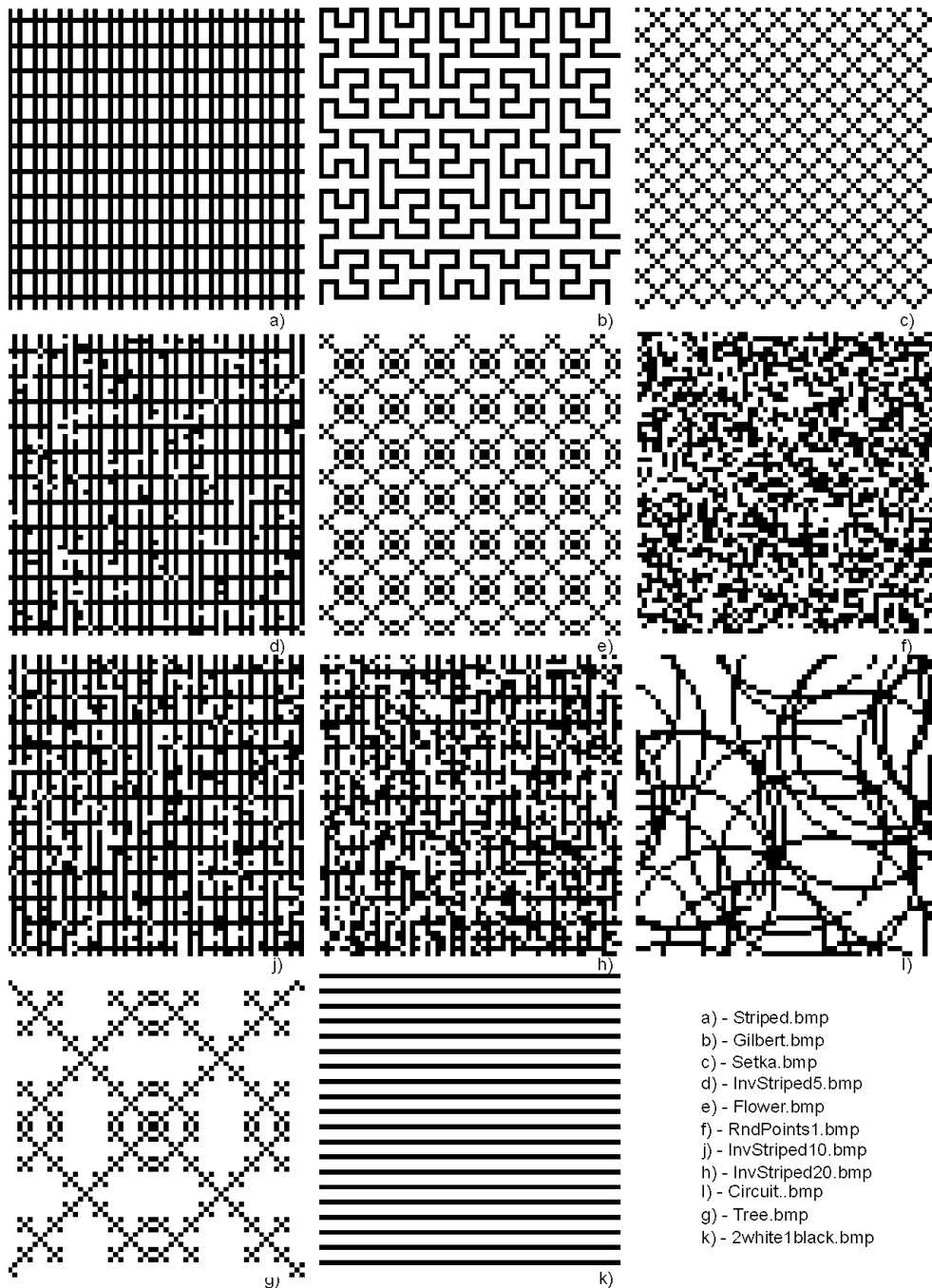


Рисунок 2 - Исследуемые изображения

Анализ реальных ЦИ сталкивается еще с одной проблемой - это проблема конечности анализируемого исходного ЦИ. Заранее невозможно сказать, что именно - структура исходного ЦИ или конечность рассматриваемого изображения оказали влияние на тот или иной эффект при определении информационной значимости ЦИ (точнее - его мозаики). Одним из способов получения ответа на этот вопрос является анализ таких ЦИ, для которых можно предполагать, что их внутренняя структура весьма однородна по пространству, а кроме того, возможно получение тех или иных теоретических оценок. Тест-объектом для такого рода верификации метода являются случайные ЦИ: изображения, состоящие из черных и белых пикселей, разбросанных случайно и нескоррелировано по фрагменту решетки, причем число пикселей каждого цвета близко к его вероятности в бесконечном ЦИ.

На рисунке 2 показаны фрагменты исследуемых ЦИ размером 60×60 , для которых оценивалась информационная ценность. Исходный размер всех изображений равен 310×310 пикселей. Все расчеты приведены для размеров мозаик от 2 до 10 пикселей. Такой выбор размера не случаен, поскольку для большинства представленных ЦИ на размере 10 все смальты мозаик почти уникальны, и значения информационной ценности либо равны нулю, либо слабо изменяются.

Хорошим тест-объектом, позволяющим получить содержательные и простые теоретические оценки влияния эффектов конечности и эффектов структуры исходного изображения, являются различные ковры - периодические замощения плоскости. Несмотря на то, что характерный размер уникальности q^* для таких изображений, как правило, весьма велик (он имеет порядок $N - q^*$, где N - раз-

мер исходного ЦИ), тем не менее, единственность продолжения наступает также очень быстро - для смальт размера порядка $q^* + 1$. Тем самым, можно с уверенностью говорить, что изменения значений условной энтропии реальной мозаики размера $l \approx (N - q^*)/2$ обусловлены исключительно краевыми эффектами (эффектами конечности исходного изображения). Поскольку теоретически оценить значения периодических ковров весьма просто, постольку представляется интересным изучение влияния случайного шума - случайных искажений, вносимых исследователем в исходно правильную периодическую структуру. Случайный шум заключался в изменении цвета пиксела (черный заменялся на белый и наоборот) с различными интенсивностями. В таком эксперименте использовался *Striped.bmp* (Рис. 2а) и его искажения с уровнем шумов 5, 10 и 20 % (Рис. 2d, j, h соответственно). Кривые изменения значений условной энтропии реальной мозаики относительно восстановленной для этих ЦИ по мере увеличения размера q смальт в мозаике показаны на рисунке 3. Для *Striped.bmp* пик величины приходится на размер 3, а ее минимум - на 6 (все смальты уникальны), что соответствует периодам, наблюдаемым в ЦИ. Максимум кривой для *InvStriped20.bmp* на размере 5 объясняется преобладанием уровня шума над периодической структурой ЦИ. И соответственно, малый уровень шума не способен замаскировать структуру ЦИ, однако, заметно искажает ЦИ, что создает обманчиво гладкую кривую. Иначе говоря, малый уровень шумов порождает интерференцию двух эффектов - периодической структуры исходного ЦИ и случайного изображения, что может создавать дополнительные трудности при анализе информационных характеристики тех или иных изображений.

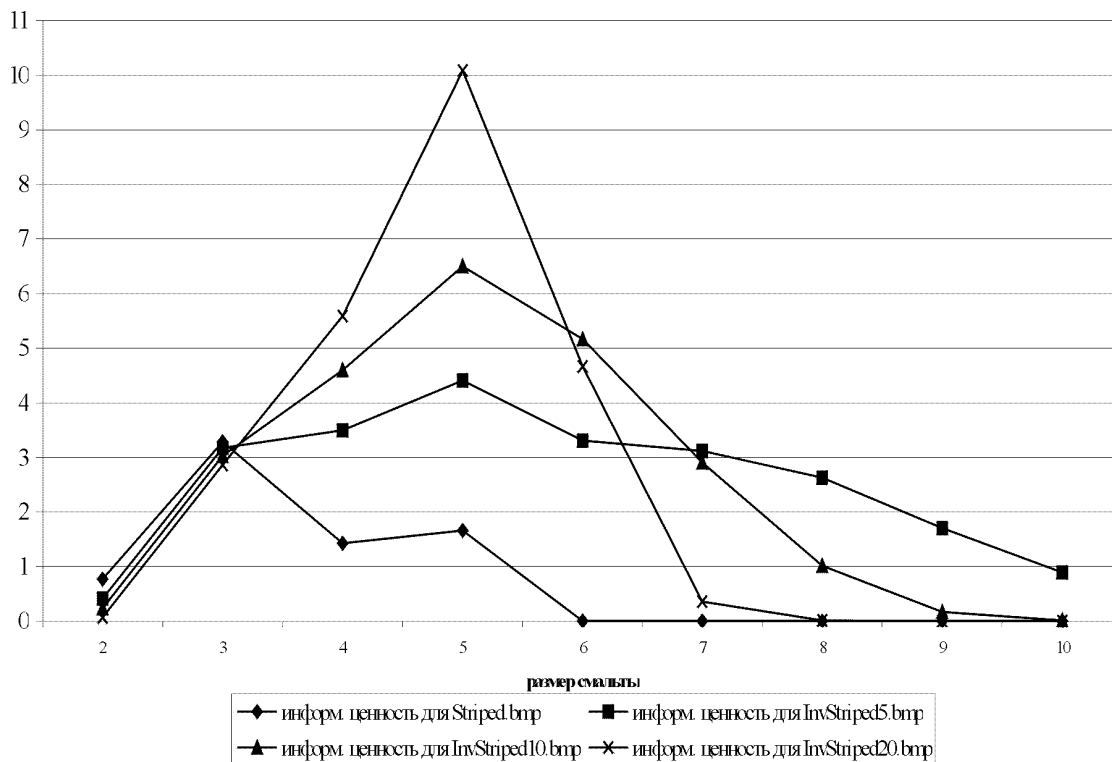


Рисунок 3 - Семейство *Striped.bmp*

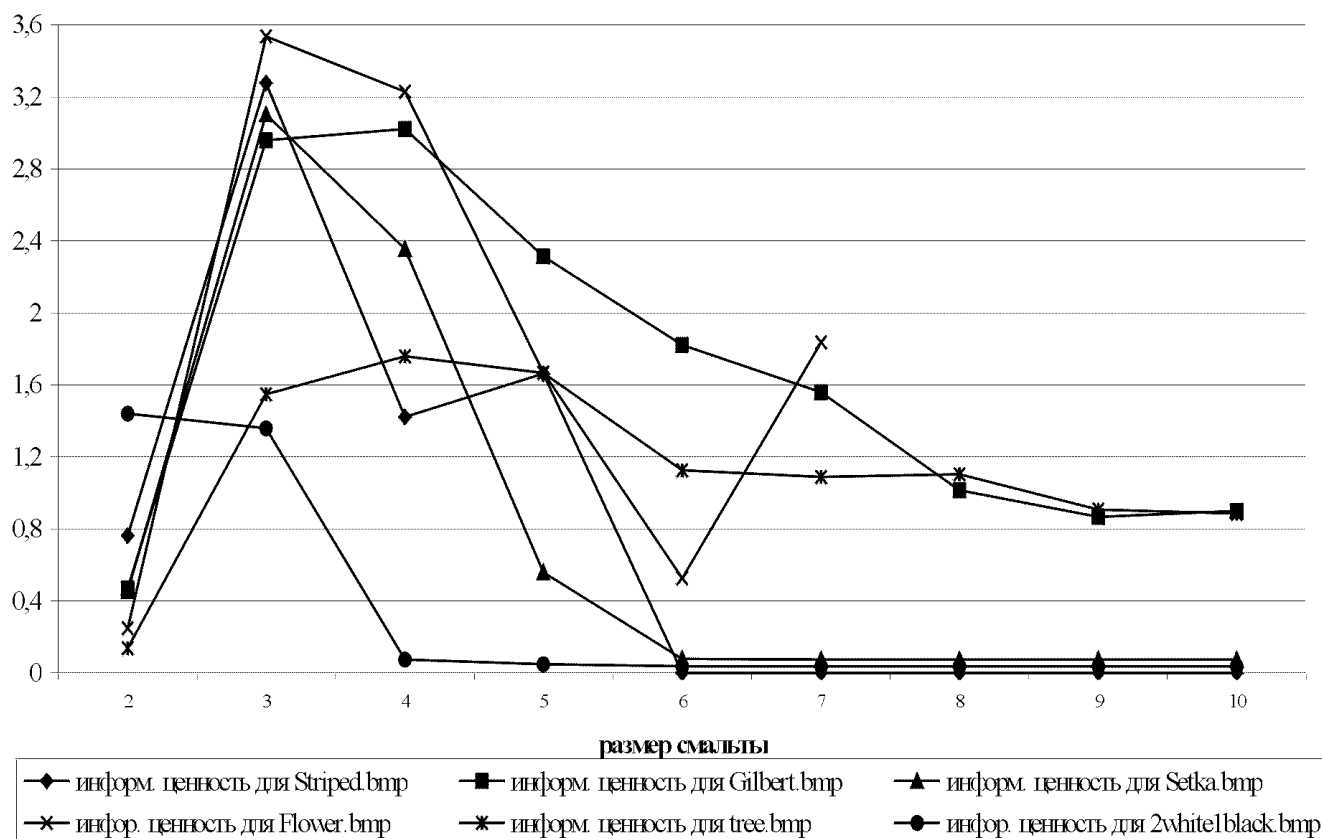


Рисунок 4 - Информационная ценность для ЦИ с периодической структурой

Другим важным тест - объектом, имеющим достаточно ясную и просто интерпретируемую структуру, является развертка Гильберта - пример фрактала. В отличие от Striped.bmp, развертка Гильберта не имеет самопересечений и по этому топологическому показателю существенно отличается от периодической структуры. Однако, с точки зрения набора малых фрагментов, развертка Гильберта мало чем отличается от Striped.bmp.

На рисунке 4 показаны характеристики всех ЦИ, имеющих периодическую структуру.

В отличие от случайных точек на плоскости (RndPoints1.bmp) и периодических структур, для развертки Гильберта возникновение смальт критического размера q^* всегда обусловлено конечностью исходного ЦИ; в случае бесконечной развертки, занимающей всю плоскость, для смальты любого размера всегда найдется еще одна ее копия в этом изображении и, следовательно, они будут иметь различные (неединственные) продолжения. С другой стороны, развертка Гильберта является весьма специальным объектом с точки зрения нашего анализа: она имеет слишком смещенное распределение черных и белых пикселей.

Рассмотрим результаты определения информационной ценности изображения, состоящего из случайно расположенных черных и белых пикселей. Теоретически, следует ожидать, что для случайных изображений не будет наблюдаться никакого заметного роста информационной ценности мозаики того или иного размера. Действительно, по-

скольку вероятность появления любой смальты (заданного размера q) есть произведение вероятностей появления смальты размера 1 (пиксела соответствующего цвета), постольку для идеального случая - бесконечного ЦИ - не будет наблюдаться никакого прироста (и, соответственно, убывания) значений условной энтропии. С другой стороны, эффект конечности изображения проявляется в том, что начиная с определенного размера q^* , все смальты становятся во-первых, уникальными, а во-вторых, имеют единственное продолжение. В этом случае также никакого изменения значений условной энтропии не наблюдается: восстановление может быть продолжено до мозаик сколь угодно большого размера, вплоть до всего исходного изображения. Наблюдаемый на размере 5 пик значения условной энтропии объясняется двумя факторами: во-первых, конечностью изображения, а во-вторых, неточностями в работе генератора случайных чисел, с помощью которого порождалось данное случайное ЦИ. Конечность исходного ЦИ означает, что в нем представлена смещенная выборка из всех мыслимых смальт размера 5×5 - общее число смальт такого размера столь велико, что распределение, полученное на исходном ЦИ весьма неточно представляет таковое, получаемое для бесконечного изображения. Другая возможная причина - наличие корреляций в появлении пикселей разных цветов. Результаты вычисления информационной ценности всех ЦИ показаны на рисунке 5.

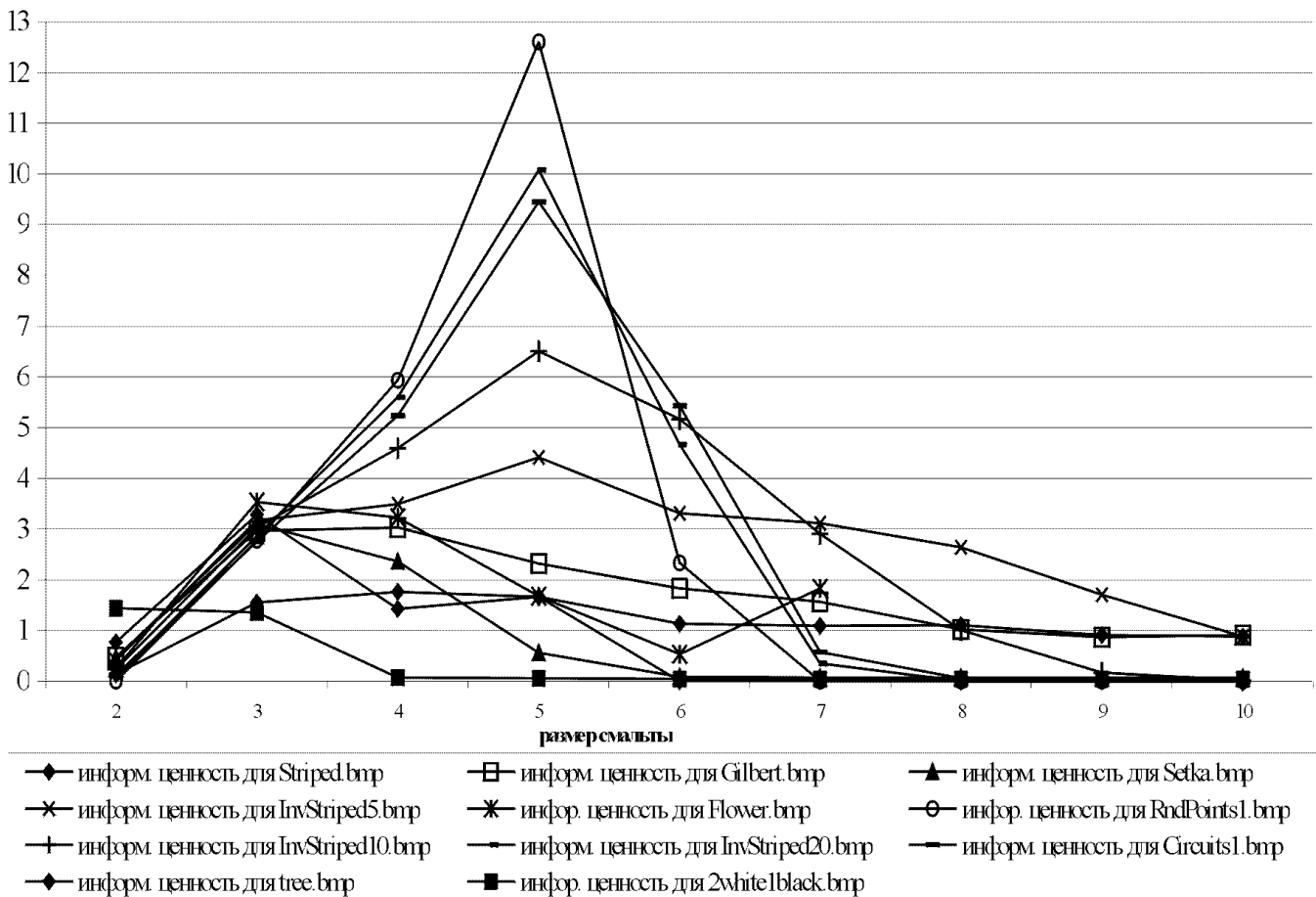


Рисунок 5 - Условная энтропия реальной мозаики относительно восстановленной

Сравнение реальных частот малых фрагментов ЦИ с наиболее ожидаемыми позволяет оценить информационную ценность мозаики. Информационная ценность - величина, зависящая от размера q смальты в мозаике. Тем самым, само по себе ЦИ характеризуется не одним параметром, а набором значений условной энтропии для размеров смальт от 1 до q^* . Два различных ЦИ, имеющие одинаковую информационную ценность на одном размере смальт, могут существенно различаться на другом. Еще одно важное свойство этого показателя заключается в том, что он принципиально не связан с "семантическими" характеристиками изображения. Это не означает, что такие нелокальные характеристики ЦИ, как связность контура, число его самопересечений и т.п. являются "плохими"; информационная ценность изображения и его "семантические" характеристики независимы и взаимодополняют друг друга. Такой показатель, как информационная ценность, может быть полезным при анализе текстур и подобных им изображений, имеющих сложную, нерегулярную, мелкозернистую структуру, плохо поддающуюся описанию и анализу в стандартных терминах теории распознавания образов. Вообще, введение в набор средств анализа информационной ценности ЦИ является шагом в развитии методологии исследования дискретных объектов с помощью энтропийных и статистических методов; одно из ближайших продолжений в этом направлении - поиск и выделение информационно значимых смальт в анализируемом ЦИ.

ЗАКЛЮЧЕНИЕ

В статье введена новая информационная характеристика растрового изображения - информационная ценность его мозаики (набора малых фрагментов), и представлены результаты определения этой характеристики для различных растровых изображений. Показано, что ЦИ различной структуры (случайные точки, фракталы, множества Мандельброта, развертка Гильберта, периодические и почти периодические изображения) различаются по такому показателю как информационная ценность мозаики того или иного размера. Указанные различия выявляют некоторые внутренние характеристики структуры изучаемого ЦИ.

Подчеркнем еще раз, что развитый в настоящей статье метод информационного анализа ЦИ заметно отличается по смыслу и содержанию получаемых результатов от тех методов, которые используются при обработке изображений, распознавании образов и т. п. Основным отличием является то, что результаты, получаемые энтропийными методами, носят "нелокальный" характер: они характеризуют изображение в целом, не выделяя и не описывая в нем локальные характеристики - контуры, точки самопересечений и т. п. Сказанное означает, что информационная ценность изображения есть показатель его отличия от случайного ЦИ, либо от ЦИ с весьма упорядоченной структурированностью - периодического и т.п. Прикладные задачи, однако, зачастую требуют

извлечения иного знания из ЦИ; развитый в настоящей статье метод несомненно может быть использован для этого, однако поиск и установление связи между энтропийными характеристиками растрового изображения и методами распознавания образом, анализа сцен и т.д. требует отдельных исследований.

ПЕРЕЧЕНЬ ССЫЛОК

1. Witkin A., Kass M. Reaction-diffusion textures // Computer Graphics, 1991, vol.25, pp.299-308.
2. Luciano Da Fontoura Costa, Roberto Marcondes Cesar Jr. Shape Analysis And Classification: Theory And Practice
3. Song Chun Zhu, Ying Nian Wu, David Mamford Minimax Entropy Principle and Its Application to Texture Modeling // Neural Computation, vol.9, (1997), pp.1627-1660.
4. Bergen J.R., Adelson E.H. Theories of visual texture perception. / Spatial Vision (D.Regan, ed.), Boca Raton, FL: CRC Press, 1991, 462 p.
5. Winkler G. Image analysis, random fields and dynamic Monte Carlo methods. Berlin: Springer-Verlag, (1995), 547 p.
6. Балеску Р. Равновесная и неравновесная статистическая механика. М.: Мир, 1976. Т.1.
7. Бугаенко Н.Н., Горбань А.Н., Карлин И.В. Универсальное разложение трехчастичной функции распределения // Теор. и матем. физика (1990), т. 28, с.430-441.
8. Горбань А.Н. Обход равновесия. Новосибирск: Наука, Сибирское Отд., 1984. 386 с.
9. Bugaenko N.N., Gorban A.N., Sadovsky M.G. Maximum entropy method in analysis of genetic text and measurement of its information content // Open System & Information Dynamics, 1998. v.5, № 3, pp. 265-278.
10. Gorban' A.N., Popova T.G., Sadovsky M.G. Classification of symbol sequences over thier frequency dictionaries: towards the connection between structure and natural taxonomy // Open Systems & Information Dynamics, 2000, v.7, № 1, pp.1-17.
11. Kirsanova E.N., Sadovsky M.G. Entropy approach to a comparison of images // Open System & Information Dynamics, 2001, (в печати).
12. Кирсанова Е.Н., Садовский М.Г. Метод статистического сравнения объектов// Радиоэлектроника. Информатика. Управление.- 2000.- №2.- С.71-82.

УДК 681.32

НЕЧЕТКИЙ ПОДХОД К СИНТЕЗУ КОМПЬЮТЕРНОЙ СЕТИ

Л.Г.Комарцова

Предложена методология нечетного вывода архитектуры компьютерной сети на основе проблемно-настраиваемой экспертной систем.

The methodology of fuzzy constraction of computer network architecture on the basis of problem ajusted expert system.

ВВЕДЕНИЕ

Неточности и неопределенности в задании исходных данных, наличие нечеткой постановки задачи по выбору характеристик компьютерной сети на начальном этапе ее проектирования приводят к необходимости использования методов нечеткой математики. Аппарат нечеткой математики позволяет формализовать понятия, которыми оперирует эксперт или проектировщик при описании своих представлений о реальной системе. Наиболее существенные особенности нечетких моделей реальных систем состоят в следующем:

1) большая гибкость по сравнению с традиционными четкими, т.к. они позволяют описывать знания и опыт эксперта в привычной для него форме;

2) большая адекватность реальному миру, поскольку позволяют получить решение, по точности соотносимое с исходными данными;

3) возможность в ряде случаев более быстрого получения окончательного результата, чем на "точных" моделях, в силу специфического построения и простоты используемых нечетких операций.

Нечеткие методы, используемые при выводе в базе знаний экспертной системы (ЭС), в которой хранятся формализованные знания эксперта о данной предметной области, можно охарактеризовать тремя отличительными чертами:

1) использование так называемых лингвистических переменных вместо числовых переменных или в дополнение к ним;

2) простые отношения между переменными описываются с помощью нечетких высказываний;

3) сложные отношения описываются нечеткими алгоритмами.

В статье предлагается методология нечеткого вывода архитектуры компьютерной сети на основе проблемно настраиваемой экспертной системы.

СТРУКТУРА РАЗРАБАТЫВАЕМОЙ ЭКСПЕРТНОЙ СИСТЕМЫ (ЭС)

ЭС включает следующие компоненты: интерфейс с проектировщиком (в виде диалогового монитора); рабочую память; базу фактов, хранящую список рабочих нечетких переменных и их значений; базу правил, содержащую знания экспертов об архитектурах вычислительных сетей в формате вида: ЕСЛИ условие ТО заключение (условие - последовательность логических операций сравнения, соединенных логическими связками И, ИЛИ; заключение - присвоение значений нечетким переменным); машину вывода - ядро системы, осуществляющую вывод путем выполнения правил в зависимости от состояния переменных, находящихся в рабочей памяти. Рабочая память экспертной системы - это множество нечетких переменных, содержащих входные данные (задание на проектирование), результаты работы машины вывода и некоторые промежуточные результаты. Методология построения проблемно настраиваемой ЭС включает несколько стадий. На первой стадии осуществляется описание переменных, характеризующих проектируемую сеть, с помощью нечетких переменных.

ОПРЕДЕЛЕНИЕ ЛИНГВИСТИЧЕСКИХ ПЕРЕМЕННЫХ И ИХ ИСПОЛЬЗОВАНИЕ

Лингвистической переменной называется переменная, значением которой являются нечеткие подмножества, выраженные в форме слов или предложений на естественном или искусственном языке. Формально лингвистическая переменная задается набором [1]:

$$\{X, T(X), U, G, M\},$$

где X - название этой переменной; $T(X)$ - терм-множество переменной X , т.е. множество лингвистических значений; U - универсум или область определения этой переменной; G - синтаксическое правило, порождающее названия значений переменной X ; M - семантическое правило, которое ставит в соответствие каждому значению лингвистической переменной ее смысл.

В отличие от классической теории множеств, оперирующей понятием принадлежности и не принадлежности элемента множеству, теория нечетких множеств допускает различную степень принадлежности к ним, определяемую функцией принадлежности элемента, значения которой изменяются в интервале $[0,1]$. При этом границы интервала характеризуют соответственно полную не принадлежность (0) или полную принадлежность (1) элемента нечеткому множеству. Могут использоваться и другие интервалы $[1,100]$, $[1,5]$ и т. д. Таким образом, смысл лингвистического значения X характеризуется некоторой функцией принадлежности: $\mu:U \rightarrow [0,1]$, которая каждому элементу $u \in U$ ставит в соответствие число из интервала $[0,1]$.

Пусть X - это *Пользователи* для одноранговой локальной сети (количество пользователей в таких сетях обычно не превышает 10). Тогда терм-множество лингвистической переменной *Пользователи* может иметь следующие значения:

$$T(\text{Пользователи}) = \text{мало} + \text{не очень мало} + \text{не мало} + \text{не много} + \text{не очень много} + \text{много}.$$

Здесь "+" означает объединение. Для лингвистической переменной *Пользователи* числовая переменная "пользователи", принимающая конкретные числовые значения 1, 2, ..., является базовой переменной. Нечеткое ограничение на значение базовой переменной характеризуется функцией принадлежности μ , которая каждому значению базовой переменной (x_i) ставит в соответствие число из интервала $[0,1]$.

На рисунке 1 представлены в виде графиков функции принадлежности для значений лингвистической переменной: "мало", "средне", "много". Такое представление показывает, что использование лингвистических значений может быть рассмотрено как сжатие данных, когда одним значением охватывается целый интервал допустимых значений базовой переменной. Это позволяет рассматривать каждое значение лингвистической переменной в виде нечеткого подмножества, причем эти подмножества могут пересекаться. Кроме того, использование лингвистических переменных для представления параметров вычислитель-

ных сетей по сравнению с обычным числовым кодированием имеет ряд других преимуществ: 1) переход от одного значения лингвистической переменной к другому является постепенным, а не скачкообразным, что дает возможности использовать более простые методы оперирования данными в процедуре проектирования сетей; 2) универсализация представления количественных (например, число пользователей) и качественных (например, тип операционной системы) параметров.

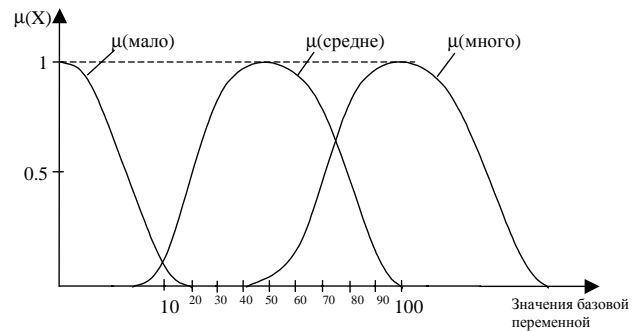


Рисунок 1 - Функции принадлежности лингвистической переменной Пользователи

Конкретные функции принадлежности создаются путем фиксации вопросов и ответов эксперту либо из экспериментов. Функция принадлежности первичных термов, входящих в терм-множество, вычисляется по формуле:

$$\mu(X) = \sum_i \mu_i(x_i)/x_i, (X \in [0,1]; \mu_i(x_i) \in [0,1]).$$

Для примера с пользователями одноранговой сети (количество пользователей меньше 10) имеем:

$$\mu(\text{"мало"}) = 1/1 + 0,98/2 + 0,9/3 + 0,8/4 + 0,5/5 + 0,3/6 + 0,15/7 + 0,1/8 + 0,08/9 + 0/10. \tag{1}$$

Здесь, например, 0,9/3 означает $\mu(3) = 0,9$.

Эта функция принадлежности запоминается в базе знаний экспертной системы и может быть затем использована для оперирования понятием "мало". Формулу (1) можно запомнить как информацию в одномерном массиве, индексы в котором соответствуют элементам (1..10). Аналогично можно определить функции принадлежности для других значений лингвистической переменной, например, "средне", "много" (рис. 1).

В случае, если совокупность элементов терма-множества не исчерпывает возможных значений лингвистических переменных (ЛП), вместо его расширения можно ввести в рассмотрение модифицированную функцию принадлежности, задаваемую, например, таблицей (таблица 1). Допустим, что необходимо ввести в рассмотрение новое значение лингвистической переменной "очень мало", тогда функция принадлежности представляется в виде следующей формулы:

$$\mu(\text{"очень мало"}) = 1/1 + (0,98)^2/2 + (0,9)^2/3 + (0,8)^2/4 + (0,5)^2/5 + (0,3)^2/6 + (0,15)^2/7 + (0,1)^2/8 + (0,08)^2/9 + (0)^2/10. \tag{2}$$

Таблица 1

Модификатор значений ЛП	Параметр модификации α	Значение функции принадлежности μ_k^α
Не	-	$1 - \mu_k$
<почти>	0,75	$\mu_k^{0,75}$
<достаточно>	1,25	$\mu_k^{1,25}$
<очень>	2,0	μ_k^2

Числовая оценка i -го значения лингвистической переменной может быть определена как лингвистическое среднее $C(X)$ модифицированной функции принадлежности:

$$C_k(X) = \frac{\sum \mu_k^\alpha(x_i)x_i}{\sum \mu_k^\alpha(x_i)} \quad (3)$$

ПОСТРОЕНИЕ ДЕРЕВА РЕШЕНИЙ

Следующая стадия построения проблемно настраиваемой экспертной системы - выявление множества правил, аккумулирующих знания экспертов. Эта стадия требует классификации переменных, характеризующих проектируемую сеть, и определения комбинаций этих переменных, которые ведут к возможным решениям. Дерево решений строится в режиме диалога с экспертом. На основе полученного иерархического, в общем случае, дерева решений строится множество правил в заданном формате базы правил. Для этого просматривается каждая ветвь дерева от корня к каждому листу и записываются все промежуточные лингвистические переменные, соединенные связками И, ИЛИ, в условия правил и в заключение. Структура решающего дерева может уточняться в процессе функционирования ЭС путем добавления новых ветвей или удалением старых. Примеры записи правил представлены ниже.

```

if {тип=одноранговая} {цена=низкая} {сегменты=10-100} then
{топология=кольцевая}
if {тип=одноранговая} {цена=средняя} {сегменты=10-100} then
{топология=звездообразная}
if {тип=одноранговая} {цена=низкая} then {носитель=витая пара}
if {тип=с выд.сервером} {цена=высокая} then {носитель=оптоволокно}
if {сохранность=средняя} {носитель=витая пара} {сегменты=10-100} then
{соединение=мосты}
if {сохранность=высокая} {носитель=витая пара} {сегменты=10-100} then
{соединение=маршрутизаторы}

```

ФУНКЦИОНИРОВАНИЕ ПРОБЛЕМНО НАСТРАИВАЕМОЙ ЭКСПЕРТНОЙ СИСТЕМЫ

Для реализации разработанной методологии определения характеристик архитектуры вычислительной сети разработана визуальная инструментальная среда в виде

оболочку ЭС, в которой, заполняя базу фактов и базу правил, можно настраивать ЭС на решение задач определенного класса. Интерфейс с разработчиком представлен на рисунке 2. Особенностью системы является комплексный подход к решению проблемы вывода: на основе единой базы знаний (включающей базу фактов и базу правил) реализованы несколько различных машин вывода: 1) на основе схемы Шортлифа [2]; 2) на основе нечеткой логики Заде [1]; 3) на основе композиционного правила вывода [2]. Такой подход позволяет выбирать стратегию вывода в зависимости от сложности решаемых задач.

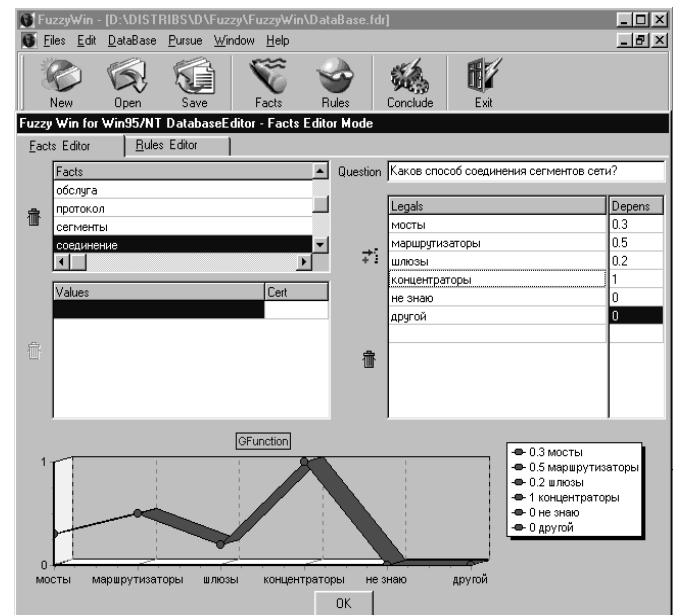


Рисунок 2 - Интерфейсное окно синтеза вычислительной сети

ВЫВОД

Результаты испытаний разработанной ЭС нечеткого вывода при решении ряда гипотетических и реальных примеров проектирования вычислительной сети показали хорошее совпадение с экспертными оценками (от 80 до 90%), что определялось при дальнейшем функционировании сети с выбранными характеристиками. Это позволяет говорить о правомочности использования данного подхода к решению задачи выбора характеристик проектируемой сети на основе ЭС с нечетким выводом.

ПЕРЕЧЕНЬ ССЫЛОК

1. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. - М.: Мир, 1976. - 186 с.
2. Экспертные системы. Принципы работы и примеры. // Под ред. Р.Форсайта. - М.: Радио и связь, 1987. - 220 с.

АНАЛИЗ АЛГОРИТМОВ РЕКОНФИГУРАЦИИ СТРУКТУРЫ ПРОЦЕССОРНОЙ МАТРИЦЫ НА ОСНОВЕ СИГНАЛОВ СОГЛАСИЯ

Н.В.Лаходынова

Проводится моделирование алгоритмов реконфигурации структуры процессорной матрицы с отказами, основанных на сигналах согласия между процессорными элементами. Цель моделирования - исследование работоспособности и эффективности алгоритмов реконфигурации, оценка требований, предъявляемых к надежности отдельных процессорных элементов и связей между ними.

Modeling of reconfiguration of processor matrix algorithm with refusal based on signals of agreement. The aim of modeling is researching of efficiency of algorithm's reconfiguration, evaluation of the requirements for safety of separate processor and their connections.

ПОСТАНОВКА ЗАДАЧИ

Алгоритмы реконфигурации [1-3] процессорной матрицы (ПМ) на СБИС обеспечивают отказоустойчивость, то есть сохранение работоспособности системы при множественных отказах. Матрица, выполненная на неразрезной СБИС, представляет собой решетку, в узлах которой находятся процессорные элементы (ПЭ). Избыточная система связей между ПЭ позволяет вкладывать в физическую структуру матрицы решетки заданного типа. Количество физических соседей одного ПЭ может достигать 14-16. Количество логических соседей определяется типом решетки. Так в квадратной решетке ПЭ(i, j), i, j - физические координаты, имеет связи с четырьмя соседями ($i-1, j$), ($i, j-1$), ($i+1, j$), ($i, j+1$). В решетке, сопряженной с квадратной, кроме вертикальных и горизонтальных соседей ПЭ(i, j) имеет связи с ближайшими соседями, расположенными по диагонали: ($i-1, j-1$), ($i-1, j+1$), ($i+1, j-1$), ($i+1, j+1$). Наибольший интерес представляют квадратная решетка (К), а также треугольная (Т), шестиугольная (Ш), сопряженная с квадратной (K^+) и некоторые другие типы решеток. При кратных отказах ПЭ логическая структура исходной решетки нарушается, матрица теряет работоспособность. При изготовлении СБИС по неразрезной технологии замена ПЭ исключена. Применение алгоритмов реконфигурации структуры ПМ позволяет максимально использовать исправную часть матрицы для получения исходной решетки. По мере появления новых отказов ПЭ в процессе функционирования реконфигурация структуры дает возможность построения ПМ меньших размеров, что обеспечивает живучесть системы в целом. Таким образом, задача реконфигурации структуры ПМ сводится к отображению логической структуры исходной решетки в избыточную физическую структуру. Избыточность структуры ПМ предполагает наличие резервных ПЭ и линий связи между ними.

Ранее мы развивали два направления в разработке алгоритмов реконфигурации. Первое направление основано на использовании сигналов неисправности ПЭ. Множество таких сигналов образует синдром неисправностей, на основе которого происходит перестройка структуры ПМ. Резервные элементы располагаются по краям матрицы [1], или назначаются программно [2]. Реконфигурация структуры производится схемами коммутационного окружения аппаратно или программно, при этом предполагается их абсолютная надежность.

Второе направление представляется более перспективным и заключается в следующем. Каждый ПЭ решает тестовую задачу и сравнивает результаты тестирования с результатами соседей. По результатам сравнения создается список согласных с ним соседей. На основе логических номеров согласных соседей ПЭ(i, j) вычисляет свой логический номер в требуемой решетке, где все соседи согласны между собой и образуют консенсус [3,4]. При несогласии между ПЭ структура решетки создается путем использования резервных связей. Резервные элементы специально не выделяются. Алгоритм реконфигурации не зависит от физической структуры связей. Все связи функционально эквивалентны. Не требуется аппаратуры самодиагностики. Этот подход уместно назвать методом консенсуса. Независимо от подхода и способа реализации результатом работы алгоритма реконфигурации должна быть решетка необходимого или максимально возможного размера, либо сообщение об отказе (невозможности создать такую решетку). Учитывая, что ПМ может содержать десятки тысяч ПЭ, а также то, что каждый ПЭ может иметь большое количество соседей, актуальным становится вопрос проверки работоспособности указанных алгоритмов, исследование их эффективности. Особый интерес представляет оценка минимальных требований к надежности ПЭ на неразрезных СБИС.

1. АРХИТЕКТУРА СИСТЕМЫ С ОТКАЗОУСТОЙЧИВОЙ ПРОЦЕССОРНОЙ МАТРИЦЕЙ

Система, использующая отказоустойчивую ПМ, согласно [3,4], состоит из трех подсистем: мониторинговой, решающего поля и массовой памяти. Решающее поле - это множество ПЭ на одной или нескольких неразрезных пластинах. ПЭ соединены двумя каналами: регулярным и магистральным. Регулярный канал обеспечивает локальные информационные связи между соседними (достаточно близко расположенными друг к другу) ПЭ, а также связывает решающее поле с массовой памятью с параллельным доступом. Связи, обеспечиваемые регулярным

каналом, функционально эквивалентны, информация передается по связи в обоих направлениях. Магистральный канал связывает все ПЭ с мониторной подсистемой, включающей в себя управляющую ЭВМ, и служит для передачи общей для всех ПЭ информации. Каждый ПЭ имеет триггер активности, который блокирует прием информации по связям с соседями, с которыми он не согласен, а также запоминание вновь вычисленных значений в памяти ПЭ.

2. ВОЛНОВОЙ АЛГОРИТМ ПОИСКА КОНСЕНСУСА

В соответствии с [3, 4], реконфигурация структуры ПМ на основе сигналов согласия проводится в четыре этапа: прямой ход, обратный ход, фиксация и настройка. На этапе прямого хода каждый ПЭ составляет список своих возможных логических номеров, основываясь на логических номерах согласных с ним соседей. Начиная с верхнего левого угла ПЭ передают свои возможные логические номера согласным соседям с большими физическими номерами. Если физические номера несравнимы, то передача делается только в одном направлении, например от $(i + 1, j)$ к $(i, j + 1)$. ПЭ получает логический номер (i, j) , если возможные логические номера согласных с ним соседей удовлетворяют определенным условиям.

Для квадратной решетки ПЭ получает логический номер (i, j) , если среди соседей с которыми он согласен имеется пара с логическими номерами (i_1, j_1) , (i_2, j_2) , таких, что выполняются условия

$$[|i_1 - i_2| = 1] \ \& \ [|j_1 - j_2| = 1], \quad (1)$$

$$[(i_1 < i_2) \rightarrow (j_1 > j_2)] \vee [(i_1 > i_2) \rightarrow (j_1 < j_2)]. \quad (2)$$

Для треугольной решетки требуется тройка соседей с логическими номерами (i_0, j_0) , (i_1, j_1) , (i_2, j_2) , удовлетворяющих условиям (1), (2) и еще одному дополнительному условию

$$(i_0, j_0) = (\min(i_1, i_2), \min(j_1, j_2)). \quad (3)$$

Для решетки K^+ требуется четверка соседей (i_0, j_0) , (i_1, j_1) , (i_2, j_2) , (i_3, j_3) и выполнение условий (1), (2), (3) и дополнительное условие

$$(i_3, j_3) = (\max(i_1, i_2) + 1, \min(j_1, j_2)). \quad (4)$$

Для шестиугольной решетки требуются условия (1) и (2), если четности номеров строк и столбцов совпадают. В противном случае достаточно одного передающего и согласного соседа слева.

Логический номер (i, j) во всех случаях вычисляется как

$$(i, j) = (\max(i_1, i_2), \max(j_1, j_2)).$$

Фиктивные элементы левого и верхнего краев не могут

принимать логических номеров с ненулевым номером столбца и строки соответственно. Такая договоренность обеспечивает связь других ПЭ с краями пластины. В результате прямого хода каждый ПЭ имеет множество возможных логических номеров. Количество логических номеров ПЭ можно ограничить, меняя направления передачи номеров соседям. Если из списка логических номеров ПЭ выбирать только его верхнюю границу, то будет задана текущая решетка максимального размера. Фиктивные элементы (входы и выходы пластины) нумеруются как продолжение ПМ и снабжаются признаком края пластины. Если в результате прямого хода правый и нижний края ПМ оказались недостижимы, вырабатывается глобальный сигнал фатального отказа. Реконфигурация невозможна. В противном случае начинается обратный ход. При обратном ходе направления передачи логических номеров меняются на противоположные. Возможные логические номера вычисляются аналогично вычислениям прямого хода, а логический номер (i, j) вычисляется как $(i, j) = (\min(i_1, i_2), \min(j_1, j_2))$, где (i_1, i_2) , (j_1, j_2) - логические номера соседей, удовлетворяющих условиям для решетки заданного типа. Логический номер (i, j) считается допустимым для ПЭ, если он уже был получен в результате прямого хода. После обратного хода в списке допустимых логических номеров ПЭ остается пересечение множеств логических номеров прямого и обратного хода. Каждый ПЭ, имеющий несколько возможных логических номеров вырабатывает сигнал неоднозначности, который по магистральному каналу поступает в управляющую ЭВМ. На этапе фиксации выбирается единственный вариант решетки, после чего выполняется настройка физической структуры связей на полученную логическую нумерацию.

3. МОДЕЛИРОВАНИЕ ВОЛНОВОГО АЛГОРИТМА

Волновой алгоритм реконфигурации структуры ПМ является параллельным. Прием, передача и вычисление логических номеров делаются множеством ПЭ одновременно. Для последовательной реализации алгоритма достаточно выполнять прямой ход алгоритма, двигаясь по столбцам, начиная с левого верхнего угла матрицы. Для проверки работоспособности алгоритма и оценки его эффективности в среде fortran90 разработана программа, моделирующая волновой алгоритм. Процессорный элемент представлен структурой данных производного типа, которая включает в себя массив сигналов согласия с соседями, признаки фиктивных ПЭ (входов и выходов пластины), массив возможных логических номеров ПЭ. Процессорная матрица представляет собой матрицу (двумерный массив) с элементами типа "процессорный элемент". Таким образом, физические номера ПЭ - это номера строк и столбцов ПМ. Для накопления логических номеров согласных соседей и вычисления возможных логических номеров с целью экономии памяти, используется массив, не входящий в структуру ПЭ. Физическая структура связей задается неявно в массивах, содержащих сигналы согласия ПЭ с соседями. Процедуры вычисления

логических номеров ПЭ для разных типов решеток представлены подпрограммами. Программная реализация волнового алгоритма позволяет убедиться путем перебора различных конфигураций согласных и несогласных соседей, что результатом работы алгоритма является либо решетка заданного типа, либо фатальный отказ - сообщение о невозможности реконфигурации.

4. ОЦЕНКА ЭФФЕКТИВНОСТИ ВОЛНОВОГО АЛГОРИТМА

Эффективность алгоритмов реконфигурации обычно характеризуется вероятностью сохранения работоспособности ПМ. Под вероятностью сохранения работоспособности понимается отношение числа конфигураций, не дающих фатального отказа, к общему числу испытаний при заданной доле исправных ПЭ. Для алгоритмов реконфигурации, основанных на синдроме неисправности, соответствующие исследования проводились в работах [1, 5]. Вероятность сохранения работоспособности $P(n, p)$ рассматривалась в зависимости от вероятности исправности ПЭ, где n - количество ПЭ в матрице, p - вероятность исправности отдельного ПЭ. При этом считалось, что коммутационное окружение и связи между ПЭ абсолютно надежны. Следующие предположения позволяют сравнить эффективность волнового алгоритма и алгоритмов рассмотренных в [1, 5].

Будем считать, что два ПЭ согласны друг с другом, если оба они исправны и связь между ними исправна. Тогда вероятность того, что ПЭ имеет хотя бы одного согласного соседа, равна

$$p_c = p(1 - (1 - pr)^k), \tag{5}$$

где p - вероятность исправности ПЭ, r - вероятность исправности связи, k - количество физических соседей ПЭ. Для сравнения эффективности алгоритмов рассматривалась структура, в которой каждый ПЭ имеет 12 физических соседей, $k = 12$; при $r = 1$ (связи абсолютно надежны), $p_c = p(1 - (1 - p)^{12})$. Для этого случая на рисунке 1 представлена зависимость $P(n, p)$ для волнового алгоритма (А) и алгоритма свободного захвата с программируемым резервом (В) - самого эффективного алгоритма из [1,2], основанного на синдроме неисправности. Результаты получены для матрицы размером 20×20 . Учитывая результаты моделирования и уменьшение аппаратных издержек, можно сделать вывод о большей эффективности волнового алгоритма.

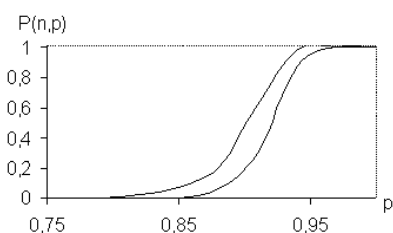


Рисунок 1

5. ВОЛНОВОЙ АЛГОРИТМ И ЗАДАЧА ПРОСАЧИВАНИЯ

Методы теории просачивания применялись для оценки минимальных требований к надежности отдельного ПЭ, обеспечивающей живучесть ПМ [5]. Под живучестью вычислительной системы обычно понимается вектор, элементами которого являются вероятности существования связанной подсистемы (кластера) заданного размера при вероятности исправности узла p и вероятности исправности связи r . В [5] рассматривалась классическая задача просачивания - задача узлов. Процессорные элементы (узлы решетки) исправны с вероятностью p и неисправны с вероятностью $(1 - p)$, узлы выходят из строя (становятся непросачиваемыми) независимо друг от друга. Связи считаются абсолютно надежными. В решетке есть просачивание, если существует связанное подмножество просачиваемых узлов, соединяющее левый и правый края решетки. Задача просачивания по связям, при условии, что узлы абсолютно надежны, сводится к задаче узлов. Бесконечные решетки, в которых исследуется существование бесконечных связанных подграфов (кластеров), дают асимптотические оценки для реальных систем. В реальных системах ПЭ (узлы) и связи ненадежны, причем около 40% неисправностей приходится на долю соединений. Таким образом, более адекватной представляется смешанная задача просачивания, когда и узлы и связи могут быть неисправными (непросачиваемыми). Следует заметить, что классическим задачам просачивания посвящено большое количество публикаций. Смешанная задача мало изучена, практические оценки значений вероятностей просачивания автору неизвестны.

Пусть $P(n, p, r, t)$ - вероятность просачивания в решетке, содержащей n узлов, p - вероятность просачивания узла, r - вероятность просачивания связи, t - тип решетки. Исходным утверждением теории просачивания для задачи узлов ($r = 1$) является утверждение о существовании критической вероятности p_H (порога просачивания), такой, что бесконечный кластер в бесконечной решетке существует с вероятностью 1 при $p \geq p_H$, и с вероятностью 0 при $p < p_H$. Аналогичное утверждение имеет место и для задачи связей. Для конечных решеток функция $P(n, p, r, t)$ имеет точку перегиба и при увеличении количества узлов приближается к ступенчатой. Точка перегиба дает приближенное значение p_H [5]. Значение критической вероятности зависит от типа решетки. Критические вероятности p_H позволяют оценить влияние надежности ПЭ и структуры связей на отказоустойчивость и живучесть процессорной матрицы в целом. Моделирование волнового алгоритма позволяет исследовать просачивание информации по пластине от входов к выходам по согласным ПЭ. При этом в качестве физической структуры берется исследуемая решетка (К, Т и т.д.). Логические номера ПЭ не вычисляются. Каждый ПЭ передает согласным с ним соседям признак связности с левым краем решетки. Таким образом, если ПЭ, находящийся на правом краю пластины имеет признак связности, то он

принадлежит кластеру, соединяющему левый и правый края пластины. При моделировании просачивания по согласным ПЭ вероятность просачивания $P(n, p_c, t)$, так же как и в классической задаче узлов, приближается к ступенчатой при увеличении n . Так же существует критическое значение доли ПЭ, имеющих хотя бы одного согласного соседа, p_{cH} . Если $p_c < p_{cH}$, то функция $P(n, p_c, t)$ близка к нулю, в противном случае $P(n, p_c, t)$ приближается к 1. Таким образом, доля ПЭ, имеющих хотя бы одного согласного соседа, меньше порогового значения, то ПМ рассыпается на несвязные куски, если доля таких ПЭ больше порогового значения, то при достаточно больших n , (уже при $n > 9$) вероятность существования кластера, соединяющего входы и выходы, близка к 1. На рисунке 2 приводятся результаты моделирования просачивания по согласным ПЭ для некоторых типов решеток при $n = 20$. Точки перегиба соответствующих кривых дают приближенные значения p_{cH} .

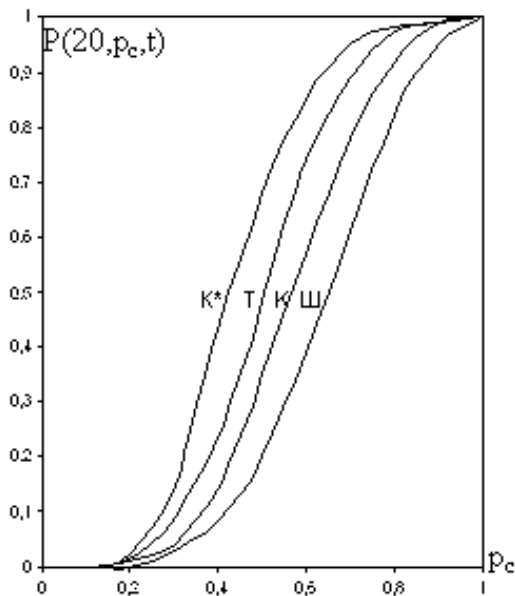


Рисунок 2

Таблица 1

Пороги просачивания				
Решетка	k	p_H	r_H	p_{cH}
К	4	0,59	0,5	0,53
Т	6	0,5	0,35	0,46
Ш	3	0,7	0,65	0,64
К*	8	0,41	0,25	0,39

Порог просачивания по согласным ПЭ характеризует живучесть системы с учетом ненадежности узлов и связей. В исследуемом алгоритме неисправности процессорных элементов и связей не рассматриваются. Неисправности представлены только сигналами согласия между соседними ПЭ. Тем не менее интересно попутно оценить вероятность просачивания по ненадежным связям и ненадежным узлам в зависимости от p и r . Сделать это можно не прибегая к непосредственному моделированию смешанной задачи.

Вероятность того, что ПЭ имеет хотя бы одного согласного соседа, p_c является функцией от p и r , вероятностей исправности узла и связи, соответственно. Используя формулу (5), можно построить кривые при $p_c = c = \text{const}$, вдоль которых вероятность просачивания через процессорную матрицу $P(n, p, r, t)$ постоянна и равна $P(n, p_c, t)$. Таким образом, аппроксимировав значения $P(n, p_c, t)$ (например, методом наименьших квадратов), можно получить значения $P(n, p, r, t)$ для любых p и r . Ниже в таблице 2 приводятся значения $P(20, p, r, K)$, на рисунке 3 изображена соответствующая поверхность. Три знака после запятой в значениях $P(20, p, r, K)$ получены в результате аппроксимации и используются для построения поверхности. Для оценки значений вероятности просачивания нужны только два знака после запятой. Следует заметить, что данные приведенные в таблице, получены попутно при моделировании волнового алгоритма и при определенных предположениях. Значения вероятности просачивания при $p = 1$ выглядят несколько завышенными и могут быть использованы в качестве верхних грубых оценки; при $r = 1$ результаты совпадают с ранее полученными результатами для задачи узлов. Более точные результаты может дать непосредственное моделирование смешанной задачи просачивания.

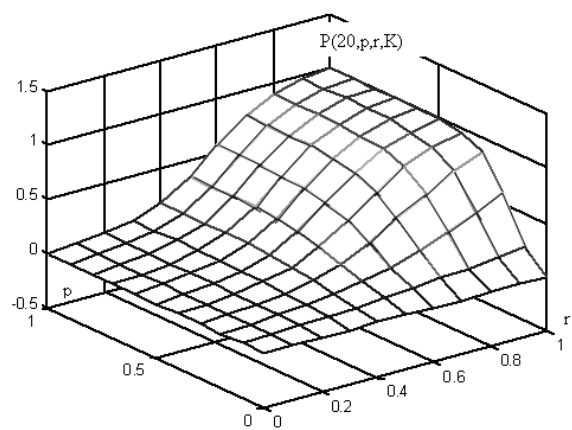


Рисунок 3

Таблиця 2

$p \backslash r$	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
1	0	0,034	0,118	0,228	0,348	0,469	0,585	0,694	0,799	0,900	1
0,9	0	0,031	0,109	0,215	0,333	0,454	0,573	0,687	0,795	0,899	0,999
0,8	0	0,028	0,100	0,200	0,315	0,435	0,556	0,674	0,787	0,895	0,998
0,7	0	0,025	0,091	0,183	0,293	0,411	0,532	0,653	0,770	0,883	0,992
0,6	0	0,022	0,080	0,164	0,267	0,380	0,499	0,621	0,742	0,860	0,974
0,5	0	0,019	0,069	0,143	0,236	0,342	0,456	0,575	0,696	0,818	0,938
0,4	0	0,015	0,057	0,120	0,201	0,295	0,400	0,512	0,629	0,749	0,870
0,3	0	0,012	0,044	0,094	0,160	0,239	0,329	0,427	0,533	0,644	0,759
0,2	0	0,008	0,003	0,067	0,113	0,172	0,240	0,317	0,402	0,493	0,590
0,1	0	0,004	0,015	0,034	0,060	0,093	0,132	0,176	0,227	0,283	0,344
0	0	0	0	0	0	0	0	0	0	0	0

ЗАКЛЮЧЕНИЕ

Моделирование волнового алгоритма реконфигурации структуры процессорной матрицы на основе сигналов согласия позволило убедиться в его работоспособности и преимуществе по сравнению с другими алгоритмами реконфигурации, оценить, какая часть процессорных элементов и связей может отказать без полного разрушения системы на отдельные несвязные куски.

ПЕРЕЧЕНЬ ССЫЛОК

1. Сами М., Стефанелли Р. Перестраиваемые архитектуры матричных процессорных СБИС. -ТИИЭР, Т.74, № 5. 1986. - С. 107-118.
2. Воробьев В.А., Лаходынова Н.В. Процессорная матрица с перестраиваемой структурой и перестраиваемым резервом // Автометрия, № 5. 1994. - С. 90-98.
3. Воробьев В.А., Лаходынова Н.В. Реконфигурация отказоустойчивой процессорной матрицы на основе сигналов согласия // Автометрия, № 6.-1997
4. Воробьев В.А., Лаходынова Н.В. Вложение решеток в процессорную матрицу с отказами на основе сигналов согласия // Новые информационные технологии в исследовании дискретных структур: Доклады всероссийской конференции. - Екатеринбург: УрО РАН, 1998.- С. 153-156.
5. Воробьев В.А., Лаходынова Н.В. Пределы надежности однородных вычислительных структур// Известия АН СС СР. Техническая кибернетика. 1989

УДК 517.91 + 681.3

КОМПЬЮТЕРНО-МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССОВ С ПОСЛЕДЕЙСТВИЕМ В МЕДИЦИНЕ

В.П.Марценюк

В работе предложен метод численного интегрирования систем с последствием, применимый для компьютерной реализации. Компьютерное описание проведено с использованием объектно-ориентированного подхода. Показано применение метода на задачах медицинской информатики.

У роботі запропоновано метод чисельного інтегрування систем з післядією, який припускає комп'ютерну реалізацію. Комп'ютерний опис здійснений із застосуванням об'єктно-орієнтованого підходу. Показано використання методу на задачах медичної інформатики.

The method of numerical integration of hereditary systems is proposed. In this work it allows computer implementation. Computer description is performed using object-oriented approach. Applications of this method for medical informatics problems are shown.

ВВЕДЕНИЕ

Динамические системы с последствием играют важную роль в экономике, биологии, медицине, инженерии. Многочисленные приложения эредитарных систем показаны в работах [1-3]. В то же время широкое практическое использование указанного класса систем сопряжено с определенными трудностями, возникающими при численном интегрировании (решении) соответствующих дифференциальных уравнений. Одной из причин является, по-видимому, бесконечность задач. Показано [4,5], что системы с последствием можно рассматривать как динамические системы в банаховом пространстве. В результате отсутствия признанных методов численного решения эредитарных систем популярны математические

пакеты (например, MATHCAD, MATHLAB, MAPLE) не содержат программных средств решения уравнений с последствием.

ПОСТАНОВКА ЗАДАЧИ

Рассматривается класс динамических систем, описываемых дифференциальными уравнениями с запаздыванием:

$$y'(x) = F(x, y(x), y(x - \tau_1(x)), y(x - \tau_2(x)), \dots, y(x - \tau_m(x))), x > 0. \quad (1)$$

Здесь $x \in R$ - независимая переменная, $y(x) \in R^n$ - фазовая траектория системы, $\tau_1(x), \tau_2(x), \dots, \tau_m(x)$ - переменные запаздывания в системе, $F(\bullet) \in R^n$ - функционал, который может быть даже кусочно-непрерывным. Предполагается, что функции $\tau_i(x)$, $i = \overline{1, m}$ принимают на положительной полуоси только неотрицательные значения и ограничены снизу и сверху. Поведение системы на начальном промежутке времени считается заданным:

$$y(x) = \phi(x), x \in [\max_{x > 0, i = \overline{1, m}} \tau_i(x), 0]. \quad (2)$$

Здесь $\phi(x)$ - функция, допускающая решение начальной задачи (1), (2).

В данной работе представлен метод, пригодный для компьютерного решения начальных задач общего вида (1), (2).

МЕТОДЫ И АЛГОРИТМЫ РЕШЕНИЯ

При численном интегрировании обыкновенных дифференциальных уравнений наибольшее развитие получил метод Рунге-Кутты и многочисленные его модификации. Применительно к задачам (1), (2) метод Рунге-Кутты может быть применен лишь в случае одного постоянного запаздывания τ . Причем это может быть лишь метод с постоянной длиной шага, приводящий к значительным погрешностям и отличающийся плохой устойчивостью. При рассмотрении более сложных систем и использовании более совершенных методов с переменной длиной шага возникают следующие трудности:

- при рассмотрении нескольких переменных запаздываний не всегда известно насколько большой диапазон предыдущих состояний системы требуется сохранять для вычисления последующих;
- не всегда сохраняются именно те значения, которые будут востребованы на будущих шагах;
- неизвестно, какая должна быть плотность сохраняемых значений.

Все вышеуказанные трудности позволяет избежать глобальная аппроксимация решения с помощью многочлена четвертой степени. Подходящими методами для плотной выдачи решений являются методы типа Адамса, непрерывные методы Рунге-Кутты. Ввиду легкости программирования и превосходства над другими интерполяционными

методами [3] в данной работе выбран метод Дормана и Принса порядка 5(4).

Программное построение решения задачи (1), (2) строится на использовании объектно-ориентированного подхода, который на сегодня получил наиболее полное развитие в системе программирования Microsoft Visual Java++ [6], применяемой в данной работе.

На рисунке 1 приведена схема классов, включенных в проект (Workspace) по решению задач (1), (2). Здесь DelaySystemSolution - класс, в котором описывается начальное поведение и правые части системы (1) а также производится численное интегрирование методом Дормана-Принса; GraphConstruction - класс-апплет, осуществляющий визуальное отображение графиков решений, их более тщательное отображение и изучение; FunctionList - класс - "связанный список", предназначенный для решения задачи одновременного вывода графиков нескольких компонент решения в одной плоскости; ColorList - класс, предназначенный для "разноцветного" отображения графиков разных компонент решения; LinkedList - класс - "связанный список", предназначенный для отображения решения задачи (1), (2) в разных пределах изменения x ; BoundsLocation - класс для хранения пределов изменения x . Стрелками обозначены направления взаимодействия классов.

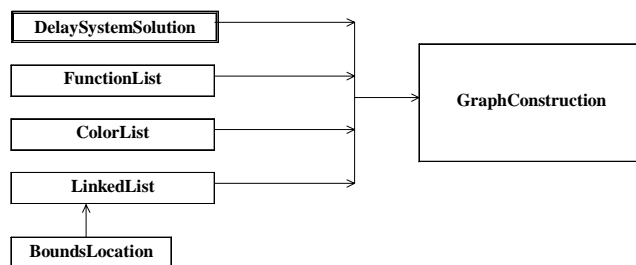


Рисунок 1 - Схема классов проекта решения задач (1), (2)

Класс GraphConstruction является типичной библиотекой машинной графики имеющей богатые визуальные возможности по изучению поведения системы на самых мелких интервалах изменения x , одновременному выводу нескольких графиков в одной плоскости. Класс использует многопоточность, а с целью избежания мерцания используется метод двойной буферизации.

Ниже указаны главные переменные и методы класса GraphConstruction.

```

public class GraphConstruction extends java.applet.Applet implements Runnable
{
    DelaySystemSolution SampleSystemSolution;
    Thread m_Graphics = null; // поток класса
    private double m_x0; // левый предел x
    private double m_x1; // правый предел x
    // второй буфер изображения
    private Image m_image; // внеэкранный изображение
    private Graphics m_g; // графический объект внеэкранный изображения
    double f(int i, double x) // метод класса, указывающий на функцию,
    // график которой следует отобразить
    {
        return SampleSystemSolution.ylag(i, x);
    }
}
    
```

Объект класса DelaySystemSolution, ассоциированный с исследуемой системой с запаздыванием создается в классе GraphConstruction и является его переменной.

Далее остановимся на строении класса DelaySystemSolution. Ниже указаны его главные переменные.

```
public class DelaySystemSolution
{
    double[] y; // переменная-массив для хранения начального значения и
    // искомого решения в точке xend
    double x0; // начальная точка для x
    private double xstor[]; // массив точек, в которых производилось интегрирование
    private double ystor[][]; // массив найденных решений
    private double c1[][]; // массивы коэффициентов аппроксимирующих многочленов,
    private double c2[][]; // в которых хранятся коэффициенты при слагаемых
    private double c3[][]; // соответственно первой, второй, третьей, четвертой
    private double c4[][]; // степеней
    int n; // размерность системы; может быть <= 51
}
```

Метод fcn, объявленный как `public double[] fcn(double x, double y[])` возвращает массив значений правых частей системы (1). Ниже будет показана его реализация на примере решения задачи из медицины.

Метод phi, объявленный как `public double phi(int i, double x)` возвращает значение начальной функции $\phi(x)$, вычисленное для i -й компоненты решения. Нахождение приближенного решения задачи (1), (2) осуществляется в классе DelaySystemSolution следующим образом. В конструктор класса DelaySystemSolution включен вызов метода retard, объявленного как:

```
public void retard(int n, double x, double y[], double xend, double eps,
double hmax, double h).
```

Далее, после успешного шага интегрирования метод retard вызывает метод store, который записывает в переменные xstor, c1, c2, c3, c4 коэффициенты многочленов вместе с соответствующими значениями x . После этого при вызове метода `ylag(i, x)` производится поиск нужного интервала для x и вычисляется соответствующий многочлен для i -го решения. Таким образом с помощью метода `ylag` можно получить i -е решение в произвольной точке x . Это свойство используется для задания запаздывания в методе fcn.

РЕЗУЛЬТАТЫ

В качестве примера рассмотрим задачу о реакции клетки на рентгеновое облучение. Предположим, что облучение продолжается в течение времени T (и начинается при $t = 0$), а затем прекращается. Предполагается, что клетки обладают способностью восполнять недостаток или устранять излишек этого вещества, но их реакция происходит с запаздыванием, равным τ . Пусть $x(t)$ - концентрация вещества в клетке, подвергнутой облучению, a - постоянная облучения, зависящая от степени облучения, b - постоянная, показывающая реакцию клетки на отклонение от равновесной концентрации x_0 . Очевидно, что $x(t) \equiv x_0, t \leq 0$. Исходя из сделанных предположений, концентрация описывается уравнениями:

$$\frac{dx(t)}{dt} = ax(t) + b(x(t - \tau) - x_0), 0 \leq t < T,$$

$$\frac{dx(t)}{dt} = b(x(t - \tau) - x_0), t \geq T. \tag{3}$$

Решения уравнений могут быть найдены методом, предложенным в данной работе. Для этого следует создать следующий метод fcn:

```
public double[] fcn(double x, double y[])
{
    //this method describes "right side" of system
    double T = 20.;
    double tau = 1.;
    double[] dRight_side = new double[Math.max(n, nn) + 1];
    if((x>0) & (x<T))
    {
        dRight_side[1] = m_a * y[1] + m_b * (ylag(1, x - tau) - phi(1,0));
    }
    else
    {
        dRight_side[1] = m_b * (ylag(1, x - tau) - phi(1,0));
    }
    return dRight_side;
}
```

Предполагается, что m_a и m_b являются переменными класса DelaySystemSolution, предназначенными для хранения значений a, b . На рисунке 2 представлено решение указанной задачи для случая $a = -1, b = -1, T = 20, \tau = 1, x_0 = 1$.

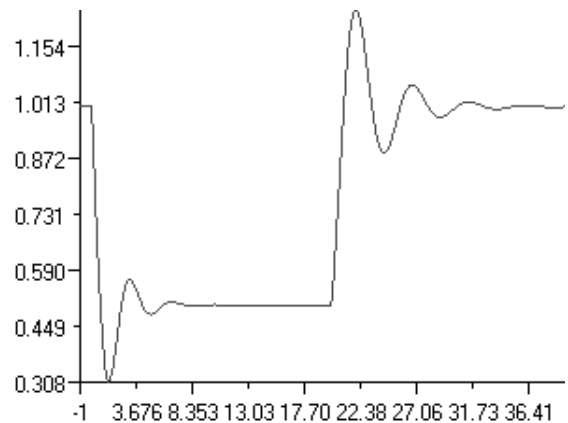


Рисунок 2 - Изменение концентрации вещества в результате рентгеновского облучения

В качестве следующего примера рассматривается модель воспалительного процесса инфекционной природы. В общем виде математическая модель иммунитета описана в работе [2]. Она универсальна и справедлива не только для воспалительного процесса, но и для инфекционного заражения организма. В модели учитываются следующие определяющие течение процесса факторы:

- 1) популяция антигенов V , размножающихся в организме;
- 2) популяция антителообразующих клеток (плазматических клеток) C ;
- 3) количество антител (иммуноглобулинов) F в организме;

4) степень поражения органа m .

Уравнения, определяющие динамику перечисленных факторов, имеют вид

$$\begin{aligned} \frac{dV}{dt} &= (\beta - \gamma F)V, \\ \frac{dC}{dt} &= \xi(m)\alpha V(t - \tau)F(t - \tau) - \mu_c(C - C_0), \\ \frac{dF}{dt} &= \rho C - (\mu_f + \eta\gamma V)F, \quad \frac{dm}{dt} = \sigma V - \mu_m m \end{aligned}$$

с начальными условиями при $t \in [-\tau, 0]$;

$$V(t) = V_0, \quad F(t) = F_0, \quad C(t) = C_0, \quad m(t) = 0.$$

Здесь β - коэффициент размножения антигена; γ - коэффициент, определяющий вероятность нейтрализации антигена антителом; α - коэффициент, обуславливающий вероятность встречи антиген-антитело; μ_c - коэффициент, обратный времени жизни плазмноклеток; ρ - скорость производства антител одной плазмноклеткой; μ_f - коэффициент, обратнопропорциональный времени распада антител; η - число антител, требующихся на нейтрализацию одного антигена; σ - коэффициент, определяющий скорость гибели клеток засчет повреждающего действия антигена; μ_m - коэффициент, учитывающий скорость восстановления пораженного органа; τ - фаза запаздывания (время, в течение которого осуществляется формирование каскада плазмноклеток); $\xi(m)$ - непрерывная невозрастающая функция ($0 \leq \xi(m) \leq 1$), характеризующая нарушение нормального функционирования иммунной системы из-за значительного поражения органа-мишени. Перечисленные параметры положительны и являются специфическими как для вида антигена, так и для органа и конкретного организма.

С помощью разработанной компьютерной модели проведено качественное исследование воспалительного процесса в случае, когда $\beta = 2$, $\gamma = 0,8$, $\alpha = 10^4$, $\mu_c = 0,5$, $\rho = 0,17$, $\mu_f = 0,17$, $\eta = 10$, $\mu_m = 0,12$, $\tau = 0,5$,

$$\xi(m) = \begin{cases} 1, & m \leq 0,1 \\ (1 - m)/(10/9), & 0,1 \leq m \leq 1. \end{cases}$$

При $t \in [-\tau, 0]$ справедливы следующие начальные условия: $V(t) = \max(0, x + 10^{-6})$, $C(t) = 1$, $F(t) = 1$, $m(t) = 0$.

Проведенное моделирование показывает, что время повторного проявления воспалительного процесса и степень его активности зависят от коэффициента σ , что согласуется с экспериментальными данными (см. рисунки 3, 4).

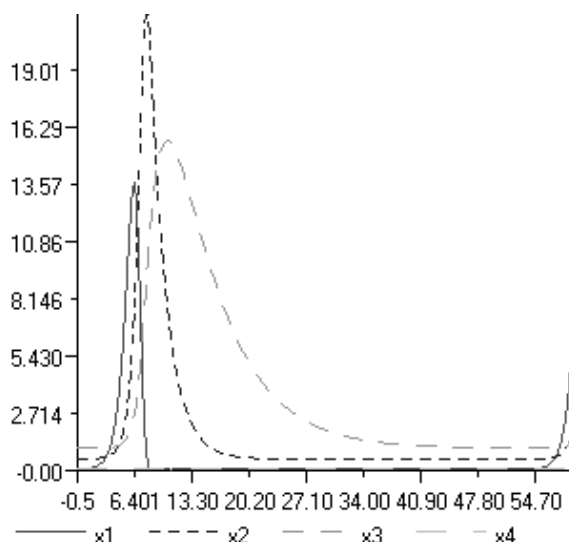


Рисунок 3 - Модель воспалительного процесса инфекционной природы при $\sigma = 2$. Здесь $x1 = 10^4 V$, $x2 = 0,5C$, $x3 = F$, $x4 = 10m$

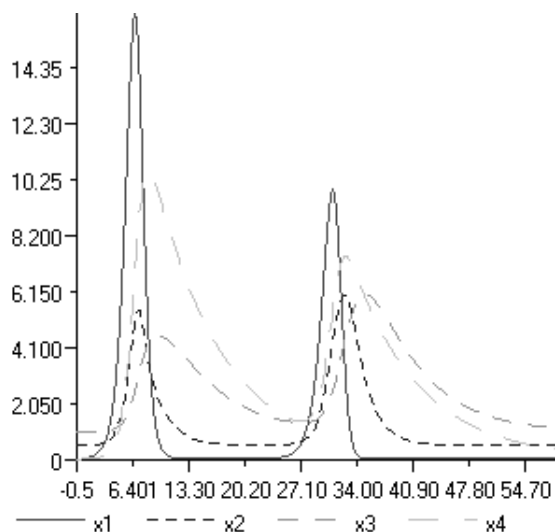


Рисунок 4 - Модель воспалительного процесса инфекционной природы при $\sigma = 300$. Здесь $x1 = 10^4 V$, $x2 = 0,5C$, $x3 = F$, $x4 = 10m$

ВЫВОДЫ

В работе представлена модель, предназначенная для исследования динамических систем с последствием. Она опирается на математический аппарат численных методов Дормана-Принса и на объектно-ориентированный подход современных инструментальных систем. Модель доведена до компьютерной программы, способной выполняться с Web-узла. Модель протестирована на задачах медицинской информатики: 1) реакция клетки на рентгеновское облучение; 2) воспалительный процесс инфекционного характера.

ПЕРЕЧЕНЬ ССЫЛОК

1. Минцер О.А., Молотков В.Н., Угаров Б.Н. и др. Биологическая и медицинская кибернетика. Справочник. - К.: Наукова думка, 1989. - 375 с.
2. Марчук Г.И. Математические модели в иммунологии. - М.: Наука, 1980. - 264с.
3. Хайрер Д., Нерсетт В., Ваннер Т. Решение обыкновенных дифференциальных уравнений. Нежесткие задачи. - М.: Мир, 1990. - 412с.
4. Хейл Дж. Теория функционально-дифференциальных уравнений. - М.: Мир, 1984. - 421с.
5. Красовский Н.Н. Некоторые задачи теории устойчивости движения. - М.: Физматгиз, 1959. - 212с.
6. Дэвис Стефен Р. Программирование на Microsoft Visual Java++/Пер.с англ. - М.: Издательский отдел "Русская редакция" TOO "Channel Trading Ltd.", 1997. - 376с.: ил.

УДК 681.3:519.21+656.072

ДИСПЕТЧЕРСКОЕ УПРАВЛЕНИЕ ДВИЖЕНИЕМ ГОРОДСКОГО ПАССАЖИРСКОГО ТРАНСПОРТА С ПОМОЩЬЮ СРЕДСТВ МИКРОПРОЦЕССОРНОЙ ТЕХНИКИ

Е.А.Оленев

В статье рассматривается модель многосвязного управления перевозочным процессом на городском пассажирском транспорте.

The model of the multilinked management of transportation process on city passenger transport is considered.

Управление транспортной системой города представляет собой очень сложный процесс, поэтому моделирование его с помощью точных математических выражений существенно затруднено. Если при этом учесть, что использование сложных формул предполагает определение большого числа параметров, измерение которых связано со значительными аппаратными затратами, а порой и просто невозможно, то становится очевидной практическая невозможность создания полностью адекватной модели управления. Описание любого реального физического процесса на транспортной сети города связано, как правило, с объективно присутствующими при этом неопределенностями, обусловленными погрешностями прогнозирования развития перевозочного процесса и неопределенностью возмущений внешней среды. Поэтому целесообразно применение двух подходов для учета указанных неопределенностей - стохастический, использующий в своей основе вероятностное пространство случайных событий, и нечеткий, базирующийся на теории нечетких множеств. Построение стохастических моделей, описывающих динамику перевозочного процесса, достаточно полно описано в работах [1, 2, 3], поэтому рассмотрим нечеткое управление, которое является одной из попыток автоматизировать операции анализа информации и выработки управляющих воздействий, которые до сих пор выполняются диспетчерским аппаратом. Идея использования нечеткой логики для диспетчерского управления, дает возможность создать хорошую, качественную модель системы, которая объединяет многосвязное управление по прогнозируемым моделям в единый технологический процесс.

При управлении ходом перевозочного процесса следует учитывать то обстоятельство, что отрицательные последствия, связанные с отклонением от нормального протекания этого процесса, нельзя, как правило, быстро устрани-

нить, поэтому для снижения возможных потерь, которые могут иметь место, необходимо постоянное моделирование картины пассажирских перевозок, обеспечивающее прогнозирование результатов управления. Возникновение конкретных ситуаций, ухудшающих пассажирские перевозки, сопровождается выдачей прогноза развития последующих событий, благодаря чему система организует движение транспортных единиц таким образом, чтобы нормализовать процесс перевоза.

На рис. 1 представлена структурная схема системы многосвязного управления ходом перевозочного процесса по прогнозируемым моделям, которая работает следующим образом.

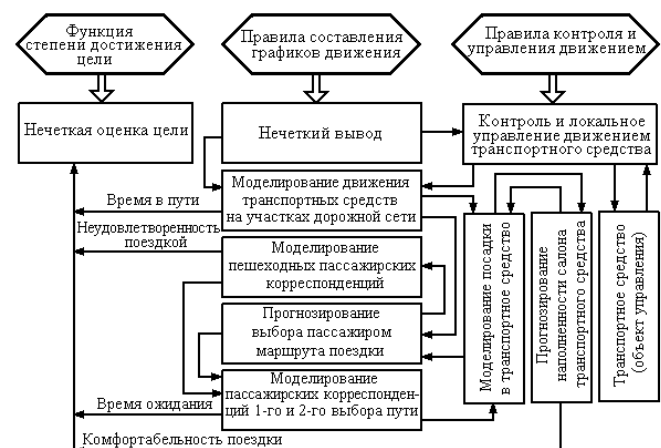


Рисунок 1 - Структурная схема системы многосвязного управления

Системой на основе информации о контроле выполнения графиков движения и в соответствии с математической моделью, параметры которой определяются статистическими исследованиями, моделируется движение транспортной единицы на участках дорожной сети города.

Получаемые при этом данные о предполагаемых моментах прибытия транспортной единицы на остановочные пункты маршрута дают возможность, используя сведения о времени, отводимом на поездку, осуществлять прогнозирование выбора пассажиром маршрута поездки, то есть в зависимости от складывающихся обстоятельств прогнозировать выбор им соответствующего множества маршрутов для поездки или пешеходного пути. Кроме того, с доверительной вероятностью 0,9 определяется также время прибытия транспортной единицы на конечный остановочный пункт маршрута.

Прогнозирование пассажирских корреспонденций 1-го и 2-го выбора пути подвергается системой дальнейшему моделированию с целью более конкретного определения размера этого множества корреспонденций, приходящегося на каждый маршрут. Кроме того, с помощью модели роста эмоционального напряжения для каждого класса пассажирских корреспонденций рассчитывается (в условных единицах) значение энергии, которую готов затратить пассажир при совершении посадки.

Исходя из расчетного уровня указанной энергии, производится моделирование посадки в транспортное средство, причем эта операция выполняется последовательно для всех остановочных пунктов маршрута. Суть данного моделирования сводится к отбору пассажирских корреспонденций, чей уровень энергии превосходит требуемое его значение для посадки в настоящий момент. Данный уровень определяется наполненностью салона транспортной единицы - чем больше наполненность, тем выше уровень и тем больше энергии затрачивает пассажир на посадку в транспортное средство и проезд в нем. Таким образом, в процессе моделирования посадки на каждом остановочном пункте маршрута пассажирские корреспонденции разделяются на совершивших и не совершивших посадку в транспортное средство. По предполагаемому количеству пассажирских корреспонденций, сделавших посадку, прогнозируется величина наполненности салона транспортной единицы на данном остановочном пункте, которая на последующем остановочном пункте маршрута снова используется для моделирования посадки и оценки комфортабельности поездки.

Множество оставшихся на остановочном пункте пассажирских корреспонденций подвергаются системой дальнейшему прогнозированию с целью определения маршрутов выбора поездки и последующего моделирование нагрузки на каждый конкретный маршрут. При необходимости система корректирует план движения транспортных единиц. Если в процессе прогнозирования маршрута поездки предпочтение будет отдано 3-му выбору, то система начнет моделирование пешеходных корреспонденций. Это производится путем подсчета моментов времени достижения ими остановочных пунктов транспортной сети, пересекаемых графом пешеходного пути. В расчетный момент времени прихода на остановочный пункт система проверяет условия хода перевозочного процесса, побудившие сделать 3-й выбор. Если они становятся благоприятными для совершения поездки, то пассажирская корреспонденция включается в процесс моделирования нагрузки на маршрут и посадки в транспортное средство вместе с другими корреспонденциями этого остановочного пункта.

При наличии неблагоприятных условий для посадки моделируется переход до следующего остановочного пункта. Наличие пешеходных корреспонденций свидетельствует о плохом качестве предоставления транспортных услуг, которое в большинстве случаев является следствием перехода процесса движения транспортных средств в нештатное состояние. Подлежат также оцениванию значения времени в пути и времени ожидания, которые получаются соответственно при моделировании движения транспортной единицы на участках дорожной сети и формирования пассажирских корреспонденций 1-го и 2-го выборов пути.

Оцениваемые величины посредством функции степени достижения цели приобретают значение нечеткой оценки цели управления. На основании оценки с учетом правил составления графиков движения делается нечеткий вывод о закреплении за транспортной единицей спланированного графика движения и о последующем контроле его выполнения или о повторном планировании этого графика с целью изменения в лучшую сторону значений оцениваемых величин.

Контроль за ходом перевозочного процесса позволяет не только отслеживать его динамику, получать параметры, необходимые для составления графиков движения, и оценки труда задействованных в технологическом процессе перевоза лиц, но и посредством логических выводов, выполняемых при получении информации с дорожной сети, выявлять неблагоприятные ситуации и принимать меры для нормализации перевозок.

С помощью локального управления, проводимого на остановочных пунктах, управляющие параметры представляются водителям транспортных единиц и пассажирам в удобной для восприятия форме. Причем состав информационных данных и их значения постоянно изменяются периферийным процессором в соответствии с протекающим временем, а также корректируются при необходимости в связи с меняющейся картиной перевозочного процесса.

Естественно, что система должна вырабатывать выходные параметры в реальном масштабе времени, определяемом в общем случае темпом движения транспортных единиц, их количеством, а также числом остановочных пунктов на дорожной сети города. Однако поскольку количество транспортных единиц и режим их движения по маршрутам определяется спланированным множеством графиков движения, то поток заявок, поступающий в систему на обслуживание, может быть определен из графиков движения, по которым в настоящее время множеством K транспортных единиц выполняются рейсы. Так как скорость движения транспортной единицы зависит от времени суток, состояния дорожного покрытия и ряда других факторов, то интенсивность потока заявок, поступающего в систему, будет являться переменной величиной. Если принять за $\lambda_k(t)$ интенсивность потока заявок, создаваемого k -й транспортной единицей, то интенсивность $\lambda(t)$ суммарного потока заявок на обслуживание, поступающих в систему с дорожной сети, будет иметь вид

$$\lambda(t) = \sum_{k=1}^K \lambda_k(t).$$

При этом поток будет близкий к нестационарному пуассоновскому.

Система хранит в памяти предполагаемых моментов прихода транспортных единиц на остановочные пункты, которая, как легко заметить, соответствует предполагаемым моментам поступления заявок на обслуживание. Это позволяет заранее вычислять параметры суммарного потока заявок и управлять вычислительным процессом в системе. Представляя модель в виде системы массового обслуживания с ограниченным временем нахождения заявки в очереди, можно получить все основные характеристики этой системы [4]. Используя полученные характеристики, можно изменять распределение заявок, добиваясь при этом более эффективного их обслуживания [5]. Так, например, при увеличении расчетной длины очереди на обслуживание следует наделять приоритетом заявки, поступающие с участков транспортной сети с большой интенсивностью движения.

Следует заметить, что для сокращения времени обслуживания заявок применено раздельное управление на двух уровнях. Управление на верхнем уровне осуществляется центральной частью системы, анализирующей состояние объектов управления и среды, в которой они функционируют. В основном управление на этом уровне заключается в автоматическом планировании графиков движения транспортных единиц. Управление на низком уровне ведется периферийной частью системы (периферийном процессором), которая преобразует в зависимости от ситуации, складывающейся на остановочных пунктах, данные, получаемые из ее центральной части.

Система диспетчерского управления характеризуется составом основных целей управления, достижение которых должно оцениваться при нечетком управлении с помощью функции степени достижения цели. Существуют разные критерии оценки транспортных услуг [6, 7], но все они, по мнению автора, могут быть сведены к двум основным - времени, затрачиваемом пассажиром на ожидание транспортного средства и проезд в нем, и комфортабельности поездки. Поэтому автор считает возможным оценивать степень достижения цели управления с помощью обобщенной функции, представляющей собой композицию двух частных функций: зависимости эффективности (Э) поездки от коэффициента снижения комфортабельности (Ξ) и зависимости эффективности поездки от времени (T), которое затрачивает пассажир на ее совершение (включая время ожидания посадки).

Первая частная функция, определяющая интересы транспортного предприятия и пассажиров на основании показателя наполненности салона транспортного средства, может иметь вид, показанный на рис. 2. Можно утверждать, что существует некоторое оптимальное число (Z₀) пассажиров (Z₁ ≤ Z₀ ≤ Z₂), совершающих поездку в данной транспортной единице, которое окупает расходы на эксплуатацию (вместе с прибылью) и оправдывает

затраты (включая дотацию), понесенные пассажиром за оплату проезда. При уменьшении количества перевозимых пассажиров (Z₀ < Z₁) снижается рентабельность перевоз-

ки, а при увеличении этого числа (Z₀ < Z₂) пассажиры испытывают дискомфорт и повышенную транспортную усталость. Все это делает поездку малоэффективной. Очевидно, что при возрастании платы за проезд график функции будет смещаться влево. Аналогичным образом будет выглядеть и вторая частная функция, отражающая главным образом соотношение между интервалом движения транспортных средств и временем, отводимым пассажиром на поездку. Следует заметить, что сокращение времени ожидания требует увеличения парка подвижного состава, что связано с большими материальными затратами. Поэтому величина времени ожидания должна быть разумной.

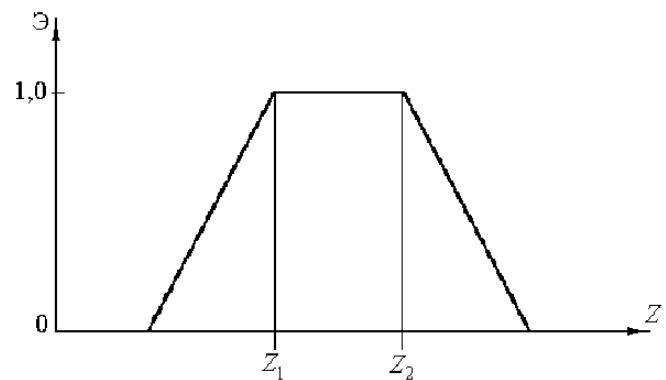


Рисунок 2 - Частная функция, определяющая критерий эффективности поездки

Значительное уменьшение этого времени при имеющемся на маршруте количестве транспортных единиц может привести к нежелательному увеличению последующего интервала движения, а следовательно, к возможной перегрузке транспортного средства.

Продолжительное ожидание общественного транспорта провоцирует пассажира ко второму и к третьему выбору пути, из-за чего возрастает вероятность прибытия в пункт назначения с опозданием. К такому же последствию ведет увеличение времени в пути. Исходя из сказанного, можно сделать вывод, что для каждого класса поездок существует определенный временной интервал, ограничивающий продолжительность поездки, соблюдение которого наиболее благоприятно отражается на эффективности поездки.

Физический смысл критерия эффективности поездки заключается в достигаемом уровне паритета затрат, понесенных организатором и потребителем на выполнение поездки, при данной ее продолжительности и комфортабельности.

Очевидно, что наибольший уровень паритета (равный 1) достигается при равенстве указанных затрат.

Более точно частные функции можно определить, изучив затратный механизм транспортного предприятия,

дотационный уровень финансирования этого предприятия местными администрациями, а также пассажиропотоки и соотношение количества льготных пассажиров и пассажиров, в той или иной степени оплачивающих свой проезд. Действительное соотношение указанных величин в каждом городе зависит от множества факторов, выявление которых производится при комплексном обследовании пассажирских перевозок.

Следует заметить, что в работе [8] дана стоимостная оценка затрат (потерь), которые несет пассажир при длительном ожидании на остановочном пункте транспортного средства. Однако в этом случае можно исходить и из предложенного автором критерия допустимого отклонения времени поездки пассажира, приняв за нулевой уровень эффективности поездки значение времени, случайное превышение которого приводило бы (по теории вероятности) пассажира к опозданию не более 1 - 2 раз в год при совершении им регулярных ежедневных поездок на работу.

Обобщенная функция степени достижения цели управления изображена на рис. 3. Поскольку параметры этой функции не могут быть взаимоднозначно определены для всех классов пассажирских корреспонденций, различных состояний дорожного покрытия и метеоусловий, то необходимо подготовить несколько обобщенных функций, с тем, чтобы их можно было автоматически выбирать в зависимости от условий работы городского пассажирского транспорта.

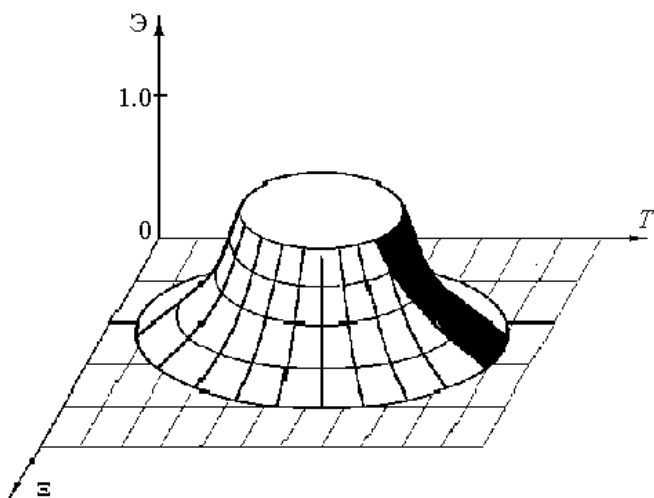


Рисунок 3 - Обобщенная функция степени достижения цели управления

Следует заметить, что главная цель управления - безопасность также достигается, хотя параметры влияющих на нее факторов не входят в явном виде в аргументы обобщенной функции. Рассмотрим три, затрагиваемые при управлении, составляющие безопасности движения:

- скорость движения транспортного средства;
- квалификация водителя;
- продолжительность предоставляемого водителю в процессе работы отдыха.

Для каждого участка дорожной сети необходимо установить аргументированные значения предельной скорости движения транспортных единиц, зависящие от состояния среды, в которой эти единицы функционируют. Планирование графиков движения производится с учетом максимально допустимой скорости, что преднамеренно исключает ее превышение в процессе движения, а постоянный контроль скоростного режима дисциплинирует работу водителей. Именно непостоянство скорости движения в основном и предопределяет разновидность вариантов трехмерной обобщенной функции степени достижения цели управления.

Квалификация водителей, достаточность их подготовки (владение всесторонней оценкой дорожной обстановки), требуемые для вождения транспортных средств по маршрутам той или иной категории сложности, могут быть выявлены как заранее (в процессе статистических исследований), так и непосредственно во время повседневной работы. Поскольку при управлении используются данные о номере транспортной единицы, которые при желании всегда можно дополнить информацией о квалификации водителя, то появляется возможность подбирать для каждой категории маршрута водителей с соответствующей подготовкой, а также предотвратить переключение (в процессе перевозок) на сложный маршрут транспортной единицы, управляемой неопытным водителем.

Во время работы водителям необходимо предоставлять минимально необходимый отдых. Короткий отдых может ухудшить работу водителя, т.е. повредить безопасности движения. Слишком большой - невыгоден. Причем, для отдыха важно выбирать не только его продолжительность, но и время его предоставления в течение смены. При этом следует предусмотреть обеденный перерыв. Поэтому все данные о режиме работы водителей заносятся также в память системы и используются при составлении сведений о моментах готовности водителей отправить транспортное средство в очередной рейс.

Постоянный учет числа транспортных единиц, прекращающих участие в перевозочном процессе из-за технической неисправности, повышает ответственность за выпуск на линию исправного подвижного состава, что способствует безаварийной работе общественного транспорта.

Таким образом, вероятностные модели параметры которых определяются статистическими исследованиями транспортной сети, принадлежат к первому пути учета неопределенностей, а нечеткое управление с прогнозированием - ко второму. Такое объединение дает возможность посредством многосвязного управления осуществлять автоматизацию операций анализа поступающей с транспортной сети информации и выработки управляющих воздействий, используя для этого средства микропроцессорной техники. Такая система нового поколения способна собрать и обработать большое количество информации, которую можно использовать для подъема пассажирских перевозок на более высокий качественный уровень. Посредством такой диспетчерской системы пассажиры получают как бы эффективный защитно-правовой механизм, с помощью которого оплата, произведенная ими за проезд в городском транспорте, сопоставляется с качеством предоставляемых транспортными предприятиями услуг.

ПЕРЕЧЕНЬ ССЫЛОК

1. Оленев Е.А. Прогнозирование выбора пассажиром маршрута поездки//Автоматика и телемеханика. -2000. - № 5.
2. Оленев Е.А. Пассажиры городского общественного транспорта - элементы активной системы перевозочного процесса//Труды Института проблем управления им.Трапезникова РАН. Москва, Институт проблем управления им. Трапезникова РАН, 26 - 28 сент., 2000г.
3. Оленев Е.А. Моделирование динамики перевозочного процесса на общественном транспорте// Труды Института проблем управления им.Трапезникова РАН. Москва, Институт проблем управления им. Трапезникова РАН, 2-4 окт., 2001г.
4. Новиков О. А., Петухов С. И. Прикладные вопросы теории массового обслуживания. - М.: Сов. радио, 1969.
5. Башарин Г. П., Бочаров П. П., Коган Я. А. Анализ очередей в вычислительных сетях. - М.: Наука, 1989.
6. Рудакова Т. В., Новгородцева Т. Ю. Классификация автобусных маршрутов города с использованием комплексного анализа/ Иркут. гос. экон. акад. - Иркутск, 1997. - С.126 - 135. - Деп. в ВИНТИ 11.02.97, № 417 - В97.
7. Chlastaez M. Le palmars 95 des villes. Vie rail. - 1995. - №2521. - P.8.
8. Павленко Г. П. и др. Автоматизированные системы диспетчерского управления движением пассажирского городского транспорта. - М.: Транспорт, 1979.

УДК 681.32

ИСПОЛЬЗОВАНИЕ WEB-ТЕХНОЛОГИЙ ДЛЯ АДМИНИСТРИРОВАНИЯ ГЕТЕРОГЕННЫХ ВЫЧИСЛИТЕЛЬНЫХ СЕТЕЙ

А.В.Пархоменко, Д.В.Данильченко, А.А.Степаненко

В статье рассматриваются современные подходы к администрированию корпоративных сетей. Предлагается подход, базирующийся на использовании web-технологий для администрирования гетерогенных корпоративных сетей.

В статті розглядаються сучасні підходи до адміністрування корпоративних мереж. Пропонується підхід, який базується на використанні web-технологій в системах керування гетерогенними корпоративними мережами.

In this article were considered modern approaches for administration of corporative networks. Propose approach, which based on using of web-technologies for administration heterogenous corporative networks.

Средства сетевого и системного администрирования никогда не занимали доминирующих позиций в корпоративных информационных системах (ИС). Традиционно отводившаяся им вспомогательная роль привела к тому, что структура и функции ПО данного класса оказались в прямой зависимости от архитектуры вычислительных

систем и эволюционировали вместе с ними.

Основная проблема, с которой сталкиваются администраторы информационных систем, заключается в том, что вместо однородной среды администраторам (и имеющимся в их распоряжении инструментальным средствам) приходится иметь дело с многообразием ресурсов, таких как различные компьютерные платформы, активное сетевое оборудование или программное обеспечение [1]. Эта гетерогенность требует решения нескольких больших классов задач администрирования (рис.1)

Типичным применением различных ОС в корпоративных информационных системах является следующее. Рабочие станции пользователей работают под управлением ОС MS Windows 9x/NT, на файловых серверах используется ОС Novell NetWare, на сервере приложений применяется MS Windows NT/2000 Server, функции шлюза в Internet и сервера удаленного доступа выполняет компьютер, работающий под управлением Linux/UNIX.

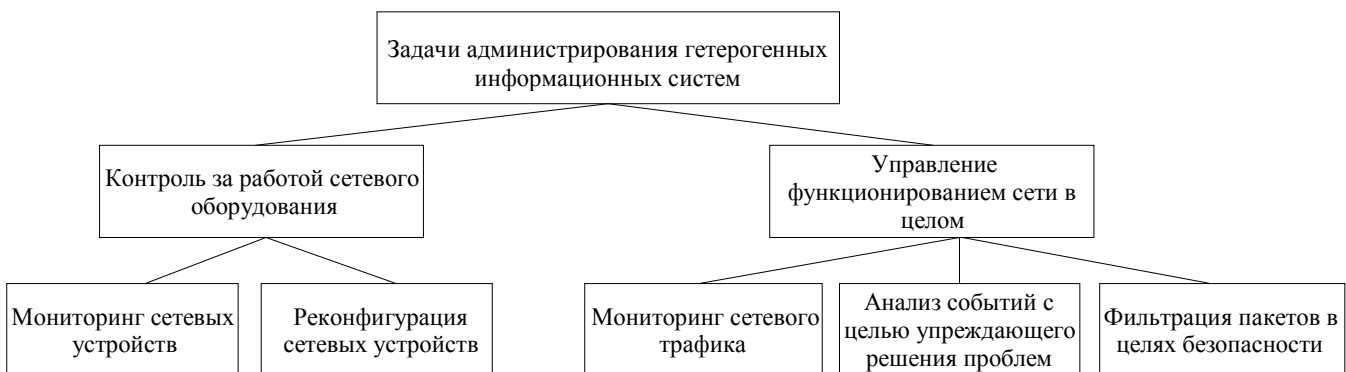


Рисунок 1 - Классификация задач администрирования гетерогенных информационных систем

Использование различных программно-аппаратных платформ обуславливается тем, что каждая операционная система имеет свои особенности. Системы, работающие под управлением UNIX, имеют большую степень защищенности и более высокую производительность. UNIX используется для управления электронной почтой, новостями или для получения доступа к Internet, а также при проектировании в приложениях, интенсивно использующих центральные процессоры компьютеров, таких как CAD/CAM. UNIX представляет надежную, устойчивую, многозадачную, многопроцессорную, многопользовательскую, хорошо масштабируемую среду. Программное обеспечение, работающее под управлением ОС семейства Windows, имеет удобный пользовательский интерфейс, что нашло отражение в его высокой популярности среди пользователей ПК. Серверы и рабочие станции, работающие под управлением ОС семейства Windows, дешевле, чем работающие под управлением UNIX, их легче администрировать, для них существует больше программного обеспечения сторонних разработчиков. Файловые серверы, работающие под управлением ОС Novell NetWare, показывают более высокую производительность, чем файл-серверы, работающие под управлением ОС семейства Windows или UNIX. Они более надежны, чем Windows и менее дорогие, чем UNIX.

Применение различных операционных систем также обусловлено удовлетворением требований пользователей к производительности корпоративной системы в целом или инструментальным средствам программного обеспечения [2].

Если для отдельной платформы задачи администрирования решаются применением стандартных средств для

этой платформы, то в гетерогенных средах интеграция и взаимодействие программных продуктов, предназначенных для администрирования, является достаточно сложной задачей (рис.2).

Сложность администрирования гетерогенной среды заключается в следующем. Различные операционные системы, активное сетевое оборудование настолько не похожи архитектурно, что только высококвалифицированный специалист с узкой специализацией может эффективно обслуживать только "свой" участок работы. Тем не менее, большинство задач администрирования очень похожи между собой, отличие заключается лишь в методе их реализации. Для UNIX-подобных систем команды и утилиты, применяемые для администрирования, это консольные приложения со множеством параметров и модификаторов-ключей. ОС семейства Windows имеют более наглядный, но не более удобный графический интерфейс средств администрирования. Для Windows написание сценариев, при помощи которых, подобно UNIX, можно облегчить процесс администрирования, практически невозможно. Подобное положение дел приводит к значительным затратам на содержание персонала, и значительным непроизводственным потерям. Windows, UNIX, NetWare представляют собой сложные операционные системы и освоение каждой из них занимает много времени. Чем больше степень разнородности, тем большее количество высококвалифицированных специалистов требуется привлечь для администрирования системы, тем выше затраты на обучение и содержание обслуживающего персонала.

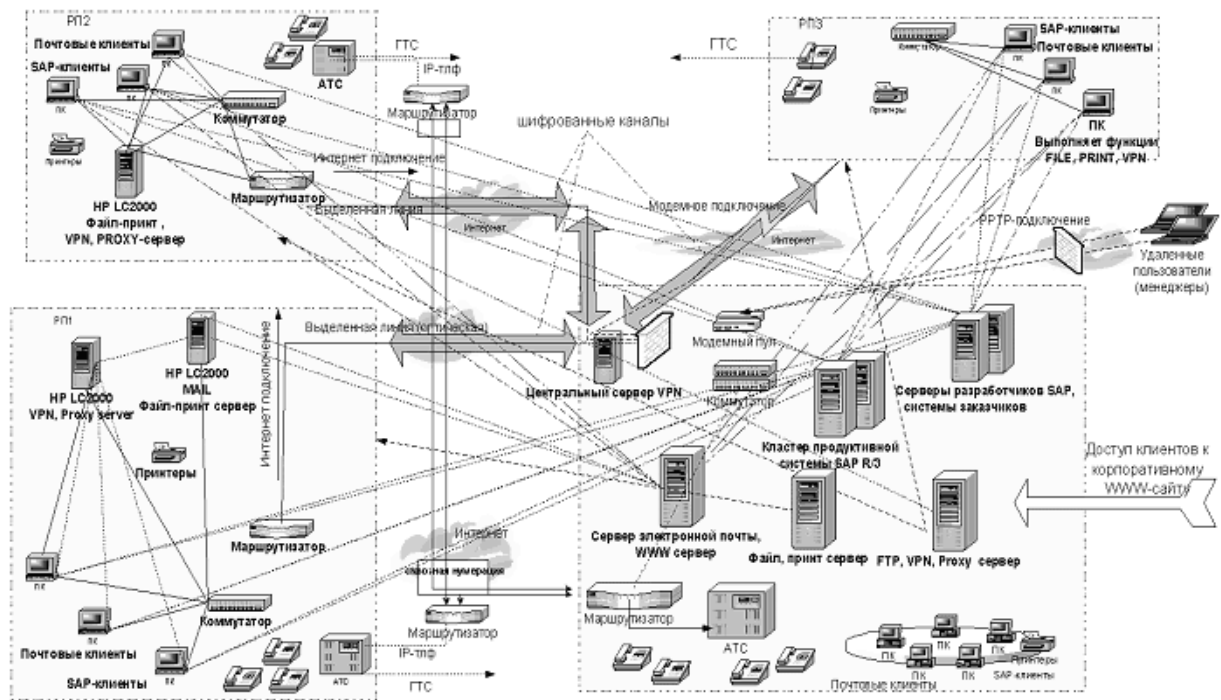


Рисунок 2 - Структура современной вычислительной сети

Наиболее распространенным подходом к контролю состояния информационной системы является централизованный подход, когда администратор контролирует состояние и управляет ИС с одной рабочей станции. Наряду с централизованным подходом существует т.н. локальный подход, когда администратор управляет некоторым сетевым устройством, непосредственно находясь "на расстоянии вытянутой руки" от него [3].

Потребность в контроле над сетью в целом с одной управляющей станции вызвала к жизни различные архитектуры платформ и приложений администрирования. Наибольшее распространение среди них получила двухуровневая распределенная архитектура "менеджер - агенты". Программа-менеджер функционирует на управляющей консоли, постоянно взаимодействуя с модулями-агентами, которые запускаются в отдельных устройствах сети. На агенты в такой схеме возлагаются функции сбора локальных данных о параметрах работы контролируемого ресурса, внесение изменений в его конфигурацию по запросу от менеджера, предоставление последнему административной информации.

Несмотря на очевидные удобства двухуровневой схемы, ее применение в реальной сетевой среде приводит к возрастанию объемов служебного трафика и, как следствие, снижению эффективной пропускной способности, доступной приложениям [1]. Этот эффект оказывается особенно заметен в сложных сегментированных сетях, содержащих большое количество активных устройств. В качестве частичного решения проблемы исчерпания пропускной способности была предложена трехуровневая схема, в которой часть управляющих функций делегировалась важнейшим сетевым узлам. Инсталлированные в этих узлах программы-менеджеры через собственную сеть агентов управляют работой "подотчетных" им устройств и в то же время сами выступают в роли агентов по отношению к основной программе-менеджеру (менеджеру менеджеров), запущенной на управляющей станции. В результате основная часть служебного трафика оказывается локализованной в отдельных сетевых сегментах, поскольку "общение" локальных менеджеров с административной консолью осуществляется только тогда, когда в этом действительно возникает необходимость.

Необходимость контролировать работу разнообразного оборудования в гетерогенной среде требует унификации основных управляющих процедур. Упомянутая схема "менеджер - агенты" нашла выражение в протоколе Simple Network Management Protocol (SNMP), который быстро стал базовым протоколом сетевого администрирования, и в стандарте дистанционного мониторинга Remote Monitoring(RMON). Управление настольными системами обычно осуществляется на базе стандарта Desktop Management Interface (DMI), разработанного организацией Desktop Management Task Force (DMTF).

Таким образом, в существующем программном обеспечении, предназначенном для удаленного администрирования, можно выделить несколько наиболее распространенных подходов.

TELNET-ДОСТУП

При помощи Telnet администратор получает доступ к консоли компьютера. Однако, в Windows не все базовые действия можно осуществлять при помощи доступа к компьютеру таким способом. Например, управление сетевыми интерфейсами, параметрами запуска/остановка служб, редактирование реестра, настройка параметров рабочего стола невозможно. В UNIX практически все ПО можно администрировать при помощи telnet-доступа, однако, различие синтаксиса команд и способов их выполнения сильно отличается от windows-систем в силу их различной архитектуры.

ИСПОЛЬЗОВАНИЕ ГРАФИЧЕСКИХ "ТОНКИХ" КЛИЕНТОВ

Вычислительные системы с применением "тонких" клиентов, вообще говоря, напоминают возврат к эпохе мэйнфреймов, однако пользователей больше не удовлетворяют терминалы для ввода и вывода текста. Все уже привыкли к более сложным графическим интерфейсам пользователя. Для администрирования операционных систем с графическим интерфейсом пользователя систем необходим доступ к рабочему столу, что дает большее количество возможностей управления ОС. Типичными представителями систем, обеспечивающих доступ к рабочему столу ПК, являются VNC компании AT&T и Remote Desktop компании Computer Associates. Логически VNC (и Remote Desktop) разделен на две части: клиентскую и серверную. Серверная часть запускается на сервере, к графическому терминалу которого нужно получить доступ. Клиентская часть запускается на том компьютере, с которого необходимо получить доступ к графическому терминалу. Релизы клиентской и серверной части VNC существуют для Windows и UNIX, что позволяет из UNIX получить доступ к графической "консоли" (рабочему столу) Windows и наоборот (Remote Desktop - только для Windows). Однако, при низкоскоростных соединениях использование VNC становится затруднительным, т.к. необходимо передавать большие объемы данных для нормального отображения содержимого экрана компьютера-сервера на мониторе компьютера-клиента.

СИСТЕМЫ НА БАЗЕ SNMP/DMI

К ним относятся системы типа OpenView компании Hewlett-Packard, Spectrum компании Cabletron, Sun Net-Manager/Solstice компании Sun Microsystems и линии SystemView компании IBM. Эти системы предоставляют рабочий материал - API разработчика, стандартный пользовательский интерфейс и некоторые базовые данные о сети (например, карту сети и базу данных атрибутов узлов) - необходимый минимум, позволяющий производителям оборудования разрабатывать приложения для настройки и мониторинга своих продуктов. Мощные многофункциональные системы управления HP OpenView или Sun NetManager, существуют и совершенствуются уже

много лет, однако для управления неоднородными сетями они пока оснащены явно недостаточно. Система OpenView компании Hewlett-Packard хорошо справляется с управлением сетями, построенными в основном на оборудовании компании Hewlett-Packard, система Spectrum компании Cabletron демонстрирует впечатляющие функциональные возможности при управлении сетями, построенными на оборудовании именно этой компании и т.д. [5]

В качестве недостатков описанных подходов можно выделить:

- различие интерфейсов пользователя ПО, предназначенного для администрирования;
- увеличение служебного трафика и уменьшение производительности серверного ПО за счет преобразования и передачи большого количества служебной информации между менеджером и агентом;
- затруднено использование систем управления в сетях, состоящих из оборудования разных производителей (системы управления начинают ошибаться и принимать одни устройства за другие, а при управлении этими устройствами поддерживают только их основные функции);
- зависимость ПО, предназначенного для реактивного администрирования (как менеджеров, так и агентов), от платформы функционирования.

Поэтому, необходимы механизмы и программные продукты, позволяющие снизить затраты на администрирование информационных систем масштаба предприятия, унифицировать интерфейсы ПО, предназначенного для администрирования.

Для решения задач администрирования гетерогенных информационных сред может быть использован подход, базирующийся на использовании web-технологий. Сильные стороны Web - в первую очередь, это простота применения, а также независимость клиентской и серверной частей от операционной системы.

В настоящее время некоторые софтверные компании работают над разработкой стандартов для систем управления на базе web, например WBEM [4], что подтверждает актуальность подобных систем.

В данной работе предлагается новый подход к администрированию гетерогенных ИС. В основе подхода лежит классическая двухуровневая модель "менеджер-агент".

В качестве транспорта данных предлагается использование протокола HTTP. Такой выбор, в первую очередь, обусловлен тем, что HTTP является универсальным транспортным протоколом для передачи информации на базе Web. HTTP распространен практически повсеместно, он реализован во всех браузерах Web на каждой платформе. Кроме того, в наши дни практически любая ОС - как серверная, так и настольная - имеет сервер Web. Более того, практически всякое межсетевое устройство, сетевой принтер или другой управляемый объект, появившийся за

последние два года, содержит встроенный сервер Web, благодаря которому конфигурация и мониторинг могут производиться непосредственно из браузера. По информации Rapid Logic, компании-разработчика инструментария для быстрой разработки встроенных управляющих приложений, необходимая для реализации демона HTTP, т. е. сервера Web, память в операционной системе реального времени не превышает 8 Кбайт [5].

Для представления служебной информации было принято решение использовать кросс-платформенное, унифицированное средство Extensible Markup Language (XML). XML представляет собой общепризнанный, стандартизированный межплатформенный механизм управления, хранения и передачи информации. Рекомендация XML, выданная консорциумом World Wide Web Consortium, является международным эталоном, который был одобрен и принят такими крупными производителями программного обеспечения как Microsoft, Netscape, Sun, IBM и многими другими.

Совместное использование XML и HTTP в рамках единой системы позволит решить некоторые проблемы управления гетерогенными ИС, а именно: накладывает не такие высокие требования на служебный трафик и платформу функционирования по сравнению с SNMP/DMI-системами и создает гибкую, высоконадежную, не зависящую от платформы среду обмена информацией. Программы разбора XML существуют для многих платформ и на многих языках программирования. Имеющийся для многих сред разработки программного обеспечения инструментарий разработки позволяет обращаться к структурам данных, содержащимся в представленных с помощью XML данных.

Вышеизложенный подход к администрированию гетерогенных сред является принципиально новым и на сегодняшний день не нашел реализации в продуктах, предназначенных для удаленного администрирования гетерогенных информационных систем, поэтому разработка программного продукта на основе предложенного подхода является актуальной научной задачей.

ПЕРЕЧЕНЬ ССЫЛОК

1. П. Иванов. Управление информационными системами: базовые концепции и тенденции развития. ОТКРЫТЫЕ СИСТЕМЫ, №4, 1999, С. 12-26.
2. David Gunter. "Windows NT and UNIX Integration Guide", Osborne, McGraw Hill, 1997.
3. Стив Штайнке. Состояние платформ. LAN/ЖУРНАЛ СЕТЕВЫХ РЕШЕНИЙ, №3, 1998, С. 97-106.
4. Стив Штайнке. Управление сетями и системами с помощью XML. LAN/ЖУРНАЛ СЕТЕВЫХ РЕШЕНИЙ, №11, 1999, С. 25-47.
5. С. Джей Янг, Джейсон Най. Разные "тонкие" клиенты. PC Magazine/RE, №10, 2000, С. 11-34.

ОЦЕНКИ ПОГРЕШНОСТЕЙ ВЫЧИСЛЕНИЯ СЛОЖНОЙ ФУНКЦИИ МНОГИХ ПЕРЕМЕННЫХ

М.Ю.Сенашова

Описан способ представления сложной функции многих переменных в виде графа. Получены формулы для оценки погрешностей вычисления сложной функции. Оценки погрешностей вычисляются на основе графа, двойственного графу, вычисляющему значение функции. Особенно просто формулы оценки погрешностей выглядят для нейронных сетей.

The method of construction of a graf of a compound function of many variables is described. Formulas for error estimations of computation of the compound function are received. The error estimations are computed on a base of dual graf. The formulas for error estimations are especially simple for neural networks.

ВВЕДЕНИЕ

Задача оценки точности вычислений актуальна с тех пор, как люди начали вычислять. Погрешности вычислений возникают из-за неточных входных данных, из-за погрешностей, вносимых на отдельных этапах вычисления, из-за погрешностей самих методов вычисления.

Каждый раз, когда возникает новая область практики вычислений, вместе с ней появляется и необходимость оценивать погрешности этих вычислений.

Интерес к нейронным сетям переживает в настоящее время очередной всплеск. Это вызвано тем, что они являются удобным и достаточно простым инструментом для создания различных экспертных систем, решения задач классификации и извлечения знаний из данных и, кроме того, дают возможность высокопараллельных реализаций.

Рассматривая нейронную сеть как набор элементов, производящих некоторые вычисления над приходящими к ним данными, можно оценивать точность вычислений как для сети в целом, так и для каждого отдельного элемента. Для нейронной сети можно решать две задачи, относящиеся к оценке точности ее работы: прямую и обратную.

Под прямой задачей понимается оценка погрешностей выходных сигналов сети, если заданы погрешности входных сигналов и погрешности элементов сети. Наиболее близки к этой задаче методы оценки погрешностей счетно-решающих устройств на основе аналоговых элементов.

Гораздо больший интерес представляет обратная задача. Под обратной задачей понимается оценка погрешностей внутренних сигналов сети, ее входных сигналов и весов синапсов по заданным погрешностям выходных сигналов сети.

Нейронную сеть можно рассматривать как сложную функцию многих переменных, состоящую из суперпозиции функций, вычисляемых нейронами. Однако, в нейронах, в основном, используются функции одного переменного. В данной работе рассматривается обобщение нейронных сетей - сложные функции, представимые как суперпозиция функций нескольких переменных.

1. ГРАФ ВЫЧИСЛЕНИЯ СЛОЖНОЙ ФУНКЦИИ

Рассматривается сложная функция многих переменных F . Под "сложной" функцией подразумевается, что F 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 [1, 2].

Возьмем, например, выражение для сложной функции:

$$F(x_1, x_2, x_3) = f_5(f_3(x_1, f_1(x_1, x_2)), f_4(f_1(x_1, x_2), f_2(x_2, x_3)))$$

На рисунке 1 изображен граф, соответствующий этой функции. Окружностями изображены вершины графа, S_i - разбиение функций f_i на слои по глубине вхождения в функцию F .

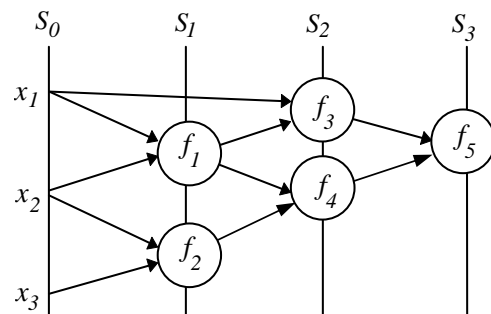


Рисунок 1 - Граф сложной функции

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 -ый. 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. Более подробно построение графа для сложной функции описано в [1].

Типичным участком графа является его вершина с входными и выходными связями. Для того чтобы иметь представление о функционировании графа в целом, достаточно выяснить, как функционирует вершина графа. Поэтому в дальнейшем, для простоты рассуждений, мы будем рассматривать не весь граф, а некоторую вершину графа.

2. 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. Как было указано выше, возможны два способа оценки погрешностей, возникающих при вычислении значения сложной функции многих переменных: прямой и обратный. Под прямым способом понимается оценка погрешности значения сложной функции, если заданы погрешности аргументов, которые одновременно являются входными вершинами графа. В этом случае по графу сложной функции пропускается конкретный набор реальных погрешностей и для них вычисляется конкретное значение погрешности функции, которая является выходной вершиной графа. Однако при этом трудно судить, как изменится погрешность значения функции, если немного изменить погрешности аргументов. Для каждого нового набора значений погрешностей аргументов требуется заново вычислять погрешности значения функции.

Гораздо удобнее использовать обратный способ. Под обратным способом понимается оценка погрешностей переменных, при условии, что задана погрешность значения сложной функции. При этом для всех ребер и вершин графа, переходя от слоя к слою в обратном порядке, вычисляются интервалы, в которых могут изменяться значения ребер и вершин. В этом случае для любого значения из интервалов, в которых могут изменяться погрешности вершин входного слоя, погрешность вершин выходного слоя не превысит заданную. Рассмотрим обратный способ оценки погрешностей.

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 движе-

нию 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 [2]. Отличие состоит в построении двойственных элементов.

Для общности рассуждений предполагаем, что граф имеет несколько выходных вершин v_{out} .

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. Поэтому, чтобы иметь представление о вычислении погрешностей на графе, достаточно выяснить, как вычисляются погрешности для такой вершины.

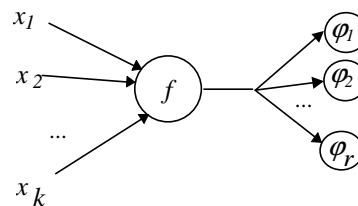


Рисунок 2 - Вершина графа, вычисляющего значение сложной функции

Рассмотрим вычисление 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 $\epsilon_1, \epsilon_2, \dots, \epsilon_r$. Чтобы попасть в интервалы погрешностей выходных вершин, нам требуется налагать наиболее жесткие требования на вычисляемую оценку погрешности вершины f_i' . Поэтому в качестве погрешности вершины f_i' выбирается

$$\epsilon \leq \min \{ \epsilon_i \}_{i=1}^r. \quad (1)$$

Пусть функция f_i является непрерывно дифференцируемой. Предположим, что погрешность двойственной вершины f_i' , вычисленная при обратном распространении, равна ϵ . Требуется вычислить погрешности ϵ_i выходящих ребер вершины f_i' такие, чтобы погрешность значения f_i не превышала ϵ .

Рассмотрим линейное приближение функции f_i в окрестности точки (x_1, \dots, x_k) :

$$f_i(x_1 \pm \epsilon_1, \dots, x_k \pm \epsilon_k) \approx f_i(x_1, \dots, x_k) + \sum_{i=1}^k \frac{\partial f_i(x_1, \dots, x_k)}{\partial x_i} \cdot (\pm \epsilon_i).$$

Для достижения необходимой точности вычислений точное значение функции $f(x_1, \dots, x_k)$ не должно отличаться от значения функции с погрешностями входящих ребер $f(x_1 \pm \epsilon_1, \dots, x_k \pm \epsilon_k)$ более, чем на ϵ :

$$|f_i(x_1 \pm \epsilon_1, \dots, x_k \pm \epsilon_k) - f_i(x_1, \dots, x_k)| \leq \epsilon. \quad (2)$$

Рассмотрим модуль этой разности, заменяя значение функции с погрешностями входящих ребер на линейное приближение в окрестности точки (x_1, \dots, x_k) :

$$|f_i(x_1 \pm \epsilon_1, \dots, x_k \pm \epsilon_k) - f_i(x_1, \dots, x_k)| \approx \left| f_i(x_1, \dots, x_k) + \sum_{i=1}^k \frac{\partial f_i(x_1, \dots, x_k)}{\partial x_i} \cdot (\pm \epsilon_i) - f_i(x_1, \dots, x_k) \right| = \left| \sum_{i=1}^k \frac{\partial f_i(x_1, \dots, x_k)}{\partial x_i} \cdot (\pm \epsilon_i) \right|.$$

Таким образом, для выполнения неравенства (2) требуется, чтобы выполнялось условие:

$$\max_{v_i} \left| \sum_{i=1}^k \frac{\partial f_i(x_1, \dots, x_k)}{\partial x_i} \cdot \epsilon_i \cdot v_i \right| \leq \epsilon,$$

где $v_i \in \{-1, 1\}$.

В предположении, что все ϵ_i равны между собой, предыдущее неравенство принимает вид:

$$\begin{aligned} \max_{v_i} \left| \sum_{i=1}^k \frac{\partial f_i(x_1, \dots, x_k)}{\partial x_i} \cdot \epsilon_i \cdot v_i \right| &= \\ = \epsilon_i \cdot \sum_{i=1}^k \left| \frac{\partial f_i(x_1, \dots, x_k)}{\partial x_i} \right| &\leq \epsilon. \end{aligned}$$

Отсюда получаем оценку для погрешностей выходящих ребер вершины f_i' :

$$\epsilon_i \leq \epsilon / \sum_{i=1}^k \left| \frac{\partial f_i(x_1, \dots, x_k)}{\partial x_i} \right|. \quad (3)$$

Пользуясь этой формулой и формулой (1) для ребер, входящих в вершину f_i' , можно оценить погрешности для любой вершины графа при известных погрешностях выходящих ребер. На рисунке 3 показана вершина двойственного графа с оценками погрешностей.

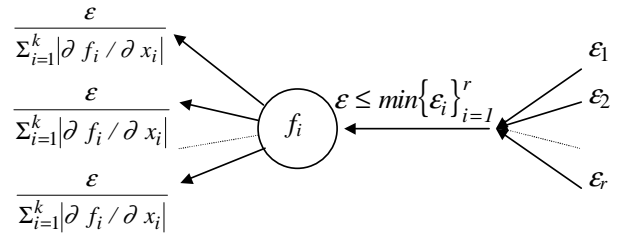


Рисунок 3 - Оценки погрешностей для вершины f_i .

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.

Если вершины входного слоя v_{in} принимают значения из некоторой области $A_i \leq x_i \leq B_i$, $i = \overline{1, n}$, где n - число вершин входного слоя, A_i, B_i - действительные числа, то можно оценить погрешности вычислений на графе не для одного конкретного набора входных значений, а для всей области входных значений в целом.

Для этого нужно интервалы, в которых изменяются значения вершин входного слоя, пропустить по сети в прямом направлении. Это делается следующим образом. Значения x_i вершин входного слоя заменяются интервалами $[A_i, B_i]$, в которых эти значения могут изменяться. Эти интервалы будут значениями ребер, входящих в вершины первого слоя графа. Затем нужно вычислить интервал изменения значений для функций $f_i(x_1, \dots, x_k)$ вершин первого слоя, где $x_i \in [A_i, B_i]$ на области D , представляющей собой k -мерный прямоугольник, ограниченный соответствующими интервалами $[A_i, B_i]$. В качестве левого конца интервала выбирается минимум функции на области D , в качестве правого конца интервала - максимум. Полученный интервал передается на вершины графа следующего слоя по ребрам, выходящим из вершины f_i . Для всех последующих слоев вычисления проводятся аналогично. После того, как для всех вершин и ребер графа вычислены интервалы изменения их значений, можно вычислять соответствующие погрешности, двигаясь в обратном направлении. При этом формула (1) остается без изменений, а формула (3) принимает вид:

$$\varepsilon_i \leq \varepsilon / \sum_{i=1}^n \max_D \left| \frac{\partial f^T}{\partial z_i} \right|. \quad (4)$$

Применяя формулы (1) и (4), можно вычислить погрешности для всей области значений входных вершин.

Особенно просто формулы оценок погрешностей выглядят для нейронных сетей [3,4]. Функциями вершин графа нейронной сети будут либо адаптивные сумматоры

$\sum_{i=1}^n \alpha_i \cdot x_i$, либо функции одной переменной $f(x)$. Для

адаптивного сумматора погрешности входных ребер оцениваются как $\varepsilon_i \leq \varepsilon / \sum_{i=1}^n |\alpha_i|$, где ε_i - погрешности

входных ребер, ε - погрешность вершины, α_i - веса си-напсов сумматора. Для функции с одной входной связью оценки погрешностей выглядят следующим образом: $\varepsilon_1 \leq \varepsilon / |f'(Z)|$, где ε_1 - погрешность входного ребра, ε - погрешность выходного ребра, Z - точное значение входного ребра вершины f .

УДК 681.32

БИБЛИОТЕКА ФУНКЦИЙ ЧИСЛОВОГО ПОДМНОЖЕСТВА СИСТЕМЫ КОМПЬЮТЕРНОЙ АЛГЕБРЫ АНАЛИТИК-2000

Т.Н.Швалюк

Настоящая работа содержит краткое описание библиотеки операций и функций арифметического подмножества системы компьютерной алгебры АНАЛИТИК-2000. Функции и операции библиотеки определены над целыми, рациональными и десятичными числами произвольной длины. Библиотеку можно использовать автономно для расширения области применения численных методов при необходимости вычислений с высокой точностью.

Present work contains the brief description of library of operations and functions of a arithmetic subset of a system of computer algebra ANALYTIC-2000. The functions and operation of library are determined on whole, rational and decimal numbers of any length. The library can be used independently for expansion of the area of application of numerical methods if necessary calculations with high accuracy.

Алгебры чисел с мантиссами произвольной длины и алгебра рациональных чисел (АРЧ) являются принципиально необходимой составной частью ориентированных на реализацию аналитических и численно-аналитических методов систем программирования (систем компьютерной алгебры).

Такие алгебры входили в состав всех языков семейства АНАЛИТИК, которые под руководством академика В.М.Глушкова, начиная с середины 60-х годов, разрабатывались в институте Математических Машин и Систем НАН Украины. Эти алгебры аппаратно реализовывались

ЗАКЛЮЧЕНИЕ

Полученные в работе формулы позволяют оценивать погрешности вычислений в вершинах графов сложных функций многих переменных при заданных погрешностях выходных вершин этих графов. Особенно просто формулы оценки погрешностей вычислений выглядят для элементов нейронных сетей.

ПЕРЕЧЕНЬ ССЫЛОК

1. 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.
2. Горбань А.Н. Обучение нейронных сетей. М.: изд. СССР-США СП "ПараГраф", 1990. 160 с.
3. Senashova Masha Yu., Gorban Alexander N., and Wunsch Donald. Back-Propagation of Accuracy // Proc. IEEE/INNS International Conference of Neural Networks, Houston, IEEE. - 1997. - P.1998-2001.
4. Сенашова М.Ю. Оценки погрешностей сигналов в нейронных сетях // Вычислительные технологии / Издательство СО РАН. - Новосибирск. 2000. Т. 5, №3. С. 83 - 109.

на ЭВМ серии МИР [1-3], а с конца 80-х годов программно реализуются на ПЭВМ [4-8]. Языком реализации становится Си [9]. Реализация таких алгебр в системах аналитических преобразований оказалась естественной, так как эти системы изначально ориентированы на объекты, большого объема со сложной структурой [10-15].

Арифметическое подмножество новой версии семейства (АНАЛИТИК-2000) образует достаточно полная библиоте-ка элементарных функций. Она расширена так же за счет некоторых специальных функций языка АНАЛИТИК-2000.

Библиотека может использоваться автономно в режиме диалога или в составе других вычислительных систем.

В зависимости от заданного в программе режима вычислений функций выполняются в различных модификациях. Так в режиме вычислений в алгебре рациональных чисел трансцендентные функции от целых, рациональных или символьных аргументов не выполняются. Например, $\cos(1/3)$ в этом режиме не вычисляется. Библиотека так же частично реализует вычисления в алгебре комплексных чисел.

Пример 1.

$$\sqrt{-4} = i*2.$$

Пример 2.

$$\ln(-2.1) = 0.7419 + i*\pi.$$

Такие соотношения намного облегчают использование символьных преобразований при программной реализации выкладок в области комплексных чисел.

Включение рациональных чисел в системы компьютерной алгебры в первую очередь связано с проблемой приводимости или точного нуля, так как при приведении подобных и других сопоставлениях символьные выражения не позволяют определять масштаб коэффициентов и проводить обычные для численных методов округления.

Пример 3.

В выражении $a + 10^{-1000} * b$ нельзя автоматически пренебречь вторым членом в процессе численно-аналитических преобразований, так как при нечисловых значениях a и b не видны масштабы этих величин.

Для реализации численно-аналитических методов АРЧ также необходима при определении свойств функций в

0.119898	0.856415	0.856415	0.171283	0.1027698
0.1641975	0.1172839	0.117284	0.234568	0.1407408
0.2869378	0.2049556	0.2049556	0.409911	0.2459466
0.943139	0.673671	0.673671	0.1347341	0.808404
0.745249	0.532321	0.532321	0.1064641	0.638785
0.2577296	0.1840926	0.1840926	0.3681851	0.2209111
0.1623903	0.1159930	0.1159931	0.2319861	0.1391917
0.2043867	0.1459905	0.1459905	0.2919824	0.1751893
0.2591862	0.1851330	0.1851330	0.3702666	0.2221599
0.3012919	0.2152085	0.2152085	0.4304172	0.2582503

методом Гаусса при получении 6 верных знаков для подтверждения результата приходится выполнять вычисления трижды: с разрядностью 6 знаков (результат $-3.02601E-52$), затем для проверки с удвоенной разрядностью 12 знаков (результат $-3.02605920028E-52$) и затем 24 знаками (результат $-3.02605920012582099257137E-52$). Процесс заканчивается, когда первых шесть знаков при втором и третьем вычислениях совпадают. В алгебре рациональных чисел точный результат: $30260592 * 10^{-60}$ получается втрое быстрее¹.

Общим видом объекта АРЧ в составе АНАЛИТИК-2000 является пара чисел разделенных знаком деления: p/q (подмножество для $q=1$ образует кольцо положительных и отрицательных целых чисел произвольной длины).

Сигнатура алгебры состоит из операций сложения, вычитания, умножения, деления и возведения в целую степень. Реализация операций над рациональными числами использует специально разработанные эффективные алгоритмы работы с числами произвольной длины (для описания которых планируется отдельная подобная публикация).

Особое значение для реализации АРЧ имеет алгоритм автоматического сокращения числителя и знаменателя в рациональном числе (деление целых - если знаменатель результата равен 1). Реализация АРЧ в системе АНАЛИТИК-2000 отличается тем, что при таком сокращении удалось избежать отдельного деления числи-

тельных полюсов, точек ветвления, при интегрировании быстроосциллирующих функций и исследовании других особенностей.

Пример 4.

У функции $y = x^{p/q}$ характер ветвления зависит от точного значения (четности или нечетности) p и q . Что тоже не допускает никаких округлений.

Эффективность реализации АРЧ в составе языков семейства АНАЛИТИК часто делает выгодным использование точных числовых преобразований при применении некоторых численных методов. Так при решении систем плохо обусловленных линейных уравнений часто вместо работы с десятичными числами оказывается более выгодной работа в алгебре рациональных чисел.

Пример 5.

Для вычисления определителя:

0.856415	0.685132	0.1198981	0.342566	0.685132
0.117284	0.938272	0.1641976	0.469136	0.938272
0.2049555	0.1639644	0.2869377	0.819822	0.1639644
0.673670	0.538936	0.943138	0.269468	0.538936
0.532320	0.425856	0.745248	0.212928	0.425856
0.1840926	0.147274	0.2577295	0.73637	0.147274
0.1159931	0.927945	0.1623903	0.463972	0.927944
0.1459912	0.1167931	0.2043881	0.583962	0.1167931
0.1851333	0.1481067	0.2591868	0.740535	0.1481067
0.2152086	0.1721669	0.3012921	0.860835	0.1721670

теля и знаменателя на общий наибольший делитель. Вместо этого параллельно с рекуррентной процедурой алгоритма Эвклида числитель и знаменатель отдельно вычисляются по рекуррентным формулам соответственно:

$$u_{i+1} = u_i + \beta_{(ui-1)},$$

$$v_{i+1} = v_i + \beta_{(vi-1)},$$

где β - число нулевых разрядов в конце результата $i+1$ шага алгоритма Эвклида.

Всего библиотека содержит 38 операций и функций в область определения, которых входят объекты числового подмножества языка АНАЛИТИК-2000. Эти функции условно делятся на следующие группы:

арифметические операции: +, -, *, /, **;

тригонометрические функции: sin, cos, tg, ctg;

обратные тригонометрические функции: arcsin, arccos, arctg, arcctg;

гиперболические тригонометрические функции: sh, ch, th, cth;

логарифмические функции: lg, ln;

exp(x) - e^x ;

sign(x) - $\text{sign}(x) = 1$, если $x > 0$; $\text{sign}(x) = 0$, если $x = 0$; $\text{sign}(x) = -1$, если $x < 0$;

abs(x) - абсолютное значение аргумента;

ent(x) - целая часть x , т.е. наибольшее целое, не превосходящее x ;

1. Точное время зависит от конкретной ЭВМ. Так на Intel Celeron 500Mhz / RAM 64Mb / HDD 10Gb общее время трех итераций 5 мс, время точного счета 1,5 мс.

int(x) - целая часть x (отбрасывание дробной части);
 frac(x) - дробная часть x , определяемая как $x - \text{int}(x)$;
 sqrt(x) - \sqrt{x} ;
 gas(x) - преобразование десятичного x в рациональное;
 min(V) - минимальное значение на множестве всех компонент вектора V ;
 max(V) - максимальное значение на множестве всех компонент вектора V ;
 gcd(x,y) - наибольший общий делитель двух целых чисел;
 ПОРЯДОК(x) - двоичный порядок x , представленного во внутреннем формате языка АНАЛИТИК-2000;
 МАНТИССА(x) - мантисса x , представленного во внутреннем формате языка АНАЛИТИК-2000;
 NUM(x) - числитель;
 DEN(x) - знаменатель;
 ing - определенный интеграл (по формуле Симпсона);
 sum - сумма;
 prod - произведение;
 agr - агрегация. Функция является свернутым представлением вектора. Результатом выполнения является вектор, аналогично тому, как sum и prod являются свернутыми суммами и произведениями последовательных слагаемых и сомножителей.

Во внутреннем представлении аргументы и значения перечисленных функций являются специально организованными массивами, над которыми определены соответствующие алгоритмы.

К числу основных алгоритмов относятся алгоритм вычисления косинуса, через который вычисляются все остальные тригонометрические функции. Он состоит из двух этапов, на первом этапе вычисляется $\cos(x/2^n)$, где n зависит от заданной точности (разрядности). На втором этапе с помощью рекуррентной формулы

$$\cos(x/2^i) = 2 * \cos(x/2^{(i-1)})^{**}2 - 1$$

вычисляется $\cos(x)$.

Базовой функцией для вычисления логарифмов и гиперболических функций является экспонента, которая аналогично вычисляется через $\exp(x/2^{**}n)$. И затем с помощью рекуррентной формулы

$$\exp(x/2^i) = \exp(x/2^{(i-1)})^{**}2$$

вычисляется $\exp(x)$.

При вычислении \ln , \sqrt{x} , arccos и деления чисел с плавающей точкой строится итерационный процесс.

Например, результат деления a/b получается через вычисление $1/b$, для которого используется итерационная формула

$$y_{i+1} = y_i + y_i * (1 - b * y_i),$$

где y_{i+1} - текущее приближение и $y \rightarrow 1/b$.

Вычисление таких функций как sign, abs, ent и других получается в результате достаточно простых преобразований соответствующих массивов. Эффективность их вычисления достигается за счет некоторых приемов программирования, которые также будут приведены в планируемой отдельной публикации.

Так как методы и алгоритмы западных систем обычно не публикуются, единственной возможностью сравнения эффективности является сравнение кривых зависимости времени от числа десятичных знаков. Из приведенных ниже графиков видно, что эффективность основных функций библиотеки мало отличается от эффективности функций таких популярных в настоящее время систем как MATHEMATICA и REDUCE.

На рисунках 1-3 изображено типичное соотношение эффективности для АНАЛИТИК-2000, MATHEMATICA и REDUCE на примерах функций \ln , \sin , \arcsin . Графики даны в логарифмической шкале, время в относительных единицах, которые зависят от частоты процессора¹.

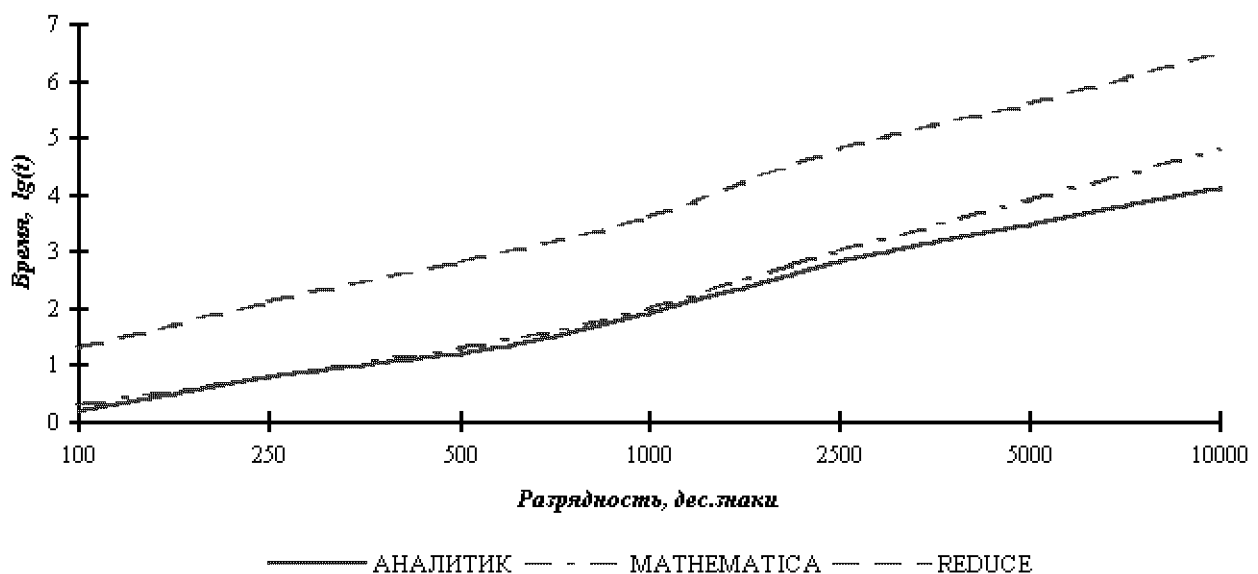


Рисунок 1 - Время выполнения функции \ln в зависимости от числа десятичных знаков мантиссы

1. Для эксперимента использовался Intel Celeron 500Mhz / RAM 64Mb / HDD 10Gb.

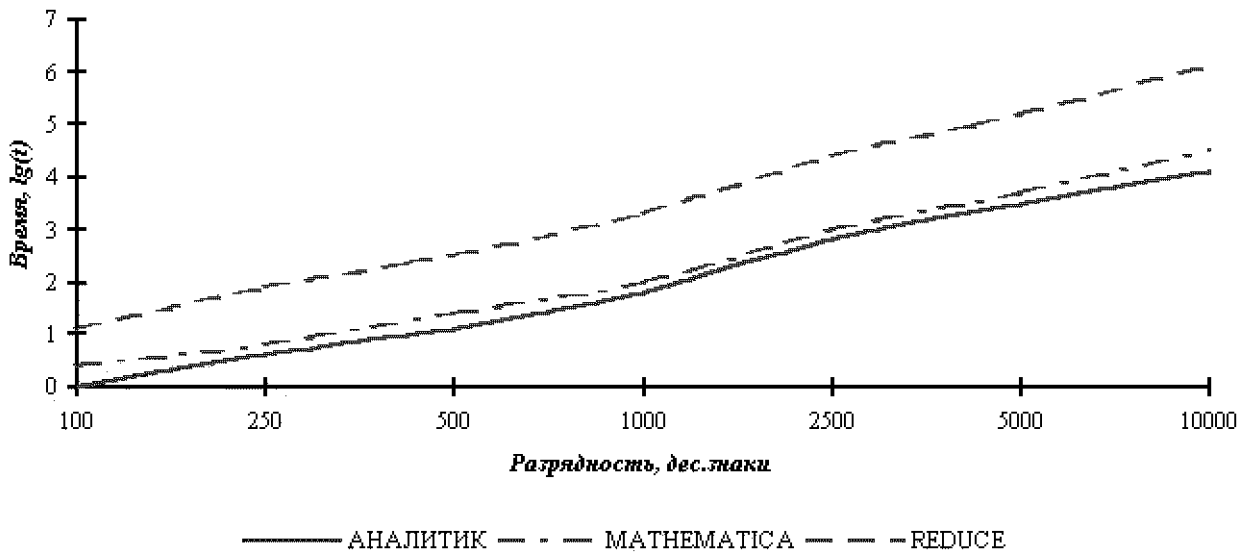


Рисунок 2 - Время выполнения функции \sin в зависимости от числа десятичных знаков мантиссы

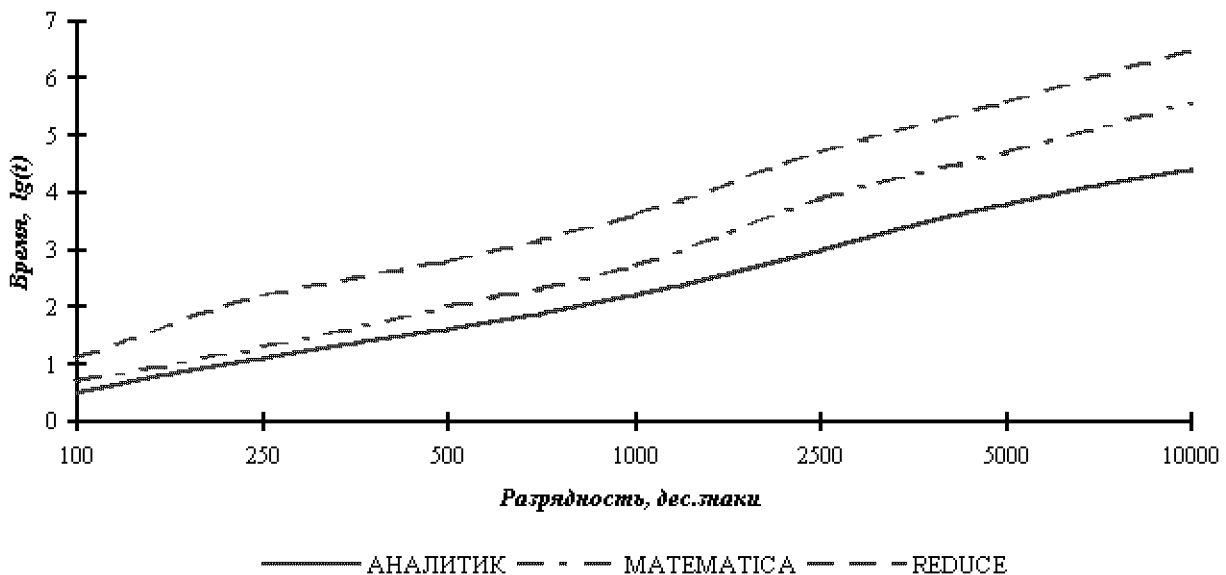


Рисунок 3 - Время выполнения функции \arcsin в зависимости от числа десятичных знаков мантиссы

Необходимо отметить, что при оперировании с длинными десятичными числами также существуют большие теоретические и практические возможности увеличения эффективности за счет распараллеливания вычислительного процесса при использовании современной вычислительной техники.

Автор глубоко благодарен Алексеенко Т.Н., Кондрашову С.В., Теслеру Г.С., Фишману Ю.С. за неоценимую помощь в работе.

ПЕРЕЧЕНЬ ССЫЛОК

1. АНАЛИТИК (Алгоритмический язык для описания вычислительных процессов с использованием аналитических преобразований)/ Глушков В.М., Боднарчук В.Г., Гринченко Т.А.,

Дородницына А.А., Клименко В.П., Летичевский А.А., Погребинский С.Б., Стогний А.А., Фишман Ю.С. // Кибернетика - №3. - Киев, 1971. - С.102 - 134.
 2. АНАЛИТИК-74 / Глушков В.М., Гринченко Т.А., Дородницына А.А., Драг А.М., Капитонова Ю.В., Клименко В.П., Крес Л.Н., Летичевский А.А., Погребинский С.Б., Савчак О.Н., Стогний А.А., Фишман Ю.С. // Кибернетика - №5. - Киев, 1978. - С.114-147.
 3. АНАЛИТИК-79 / Глушков В.М., Гринченко Т.А., Дородницына А.А., Драг А.М., Клименко В.П., Погребинский С.Б., Савчак О.Н., Фишман Ю.С., Царюк Н.П. // Киев, 1983. - 72 с. - (Препр. /АН УССР Ин-т кибернетики; 83-12).
 4. АНАЛИТИК-89 / Бублик Б.А., Фишман Ю.С. и др. // Харьков, 1989. - 78 с. - Деп. в ВИНТИ 29.06.89, N 4305-689
 5. The Programming System ANALYTIC-93 / A.A.Morosov, V.P.Klimenko, J.C.Fishman // Computer algebra applications. -St.Petersburg, 1993. - P.34-36.
 6. АНАЛИТИК-93/Морозов А.А., Клименко В.П., Фишман Ю.С., Бублик Б.А. и др.//Кибернетика и системный анализ - №5. - Киев, 1995. - С.127-156.
 7. АНАЛИТИК-2000 - ЯЗЫК КОМПЬЮТЕРНОЙ АЛГЕБРЫ. (Ориентированный на задачи, требующие высокого уровня ис-

куственного интеллекта.) / Клименко В.П., Фишман Ю.С., Ляхов А.Л., Кондрашов С.В., Швалюк Т.Н. // Четвертый сибирский конгресс по прикладной и индустриальной математике (ИНПРИМ-2000). - Тезисы докладов. - Часть IV - Новосибирск, 2000. - С.107-108.

8. АНАЛИТИК-2000 / Морозов А.А., Клименко В.П., Фишман Ю.С., Ляхов А.Л., Кондрашов С.В., Швалюк Т.Н. // Математические машины и системы. - 2001. - №1. - С.65-101.

9. Kernighan B.W. and Ritchie D.M. (1978). The C Programming Language, Prentice-Hall, Englewood Cliffs, New Jersey.

10. Вычисление элементарных функций на ЭВМ / Благовещенский Ю.В., Теслер Г.С.// К.: Техника, 1977. - 208 с.

11. Вычисление функций на ЭВМ / Попов Б.А., Теслер Г.С. // К.: Наукова думка, 1984. - 600с.

12. Арифметические операции над числами произвольной длины и произвольного диапазона / Дородницына А.А. // Кибернетика. - 1967. - № 4. - С.39-42.

13. Компьютерная алгебра / Абрамов С.А., Зима Е.В., Ростовцев В.А.// Программирование. №5.-Москва, 1992.-С.4-25.

14. Системы символьной математики Mathematica 2 и Mathematica 3 / Дьяконов В.П.// М.: СК ПРЕСС, 1998. - 320 с.

15. REDUCE. User's Manual. Version 3.4. / Hearn A.C. // The RAND Corporation, Santa Monica, 1991. - 200 p.

УДК 519.863

ІНТЕРВАЛИ І ІНТЕРВАЛЬНІ СТРУКТУРИ

Г.М.Шило

Рассмотрены виды интервальных входных данных и интервальных структур. Исследованы их основные свойства. Показано использование интервальных структур для оценки функций и расчета допусков с учетом внешних воздействий.

Розглянуто види інтервальних вхідних даних та інтервальних структур. Досліджені їх основні властивості. Показано використання інтервальних структур для оцінювання функцій і обчислення допусків з урахуванням дії зовнішніх впливів.

Kinds of interval input data and interval structures are considered. Their main properties are investigated. The using of interval structures to estimate functions and calculate tolerances taking into account external perturbations is shown.

ВСТУП

Існують різні способи опису невизначеностей даних: імовірнісні моделі, теорія невизначених множин, інтервальна математика [1, 2]. Перевагою інтервальної математики є можливість обчислення гарантованих розв'язків. При цьому необхідно мати інформацію лише про межі параметрів моделі. Правила виконання арифметичних операцій залежать від виду вхідних даних.

Арифметичні операції над інтервальними параметрами, які змінюються взаємно незалежно виконуються за правилами класичної інтервальної арифметики [3] або її розширень [4-6]. Однак досить часто параметри математичних моделей мають взаємно однозначну відповідність. В цьому випадку існуючі інтервальні арифметики не можуть застосовуватись в повному обсязі.

Ідея оцінки дійсного числа інтервалом, яка покладена в основу інтервальної математики, для оцінки множин набуває вигляду двохступеневих інтервальних структур. Такі задачі виникають, наприклад при оцінюванні функцій [7] і врахуванні дії навколишнього середовища [8]. Можлива різна комбінація межових точок, які характеризують множину (рис.1). При цьому утворюються інтервальні структури з плаваючими межами (рис.1, а), з плаваючими інтервалами (рис.1, б) та твіні (рис.1, в). З цих структур найбільш досліджені властивості твінів [7], які застосовуються для оцінювання функцій.

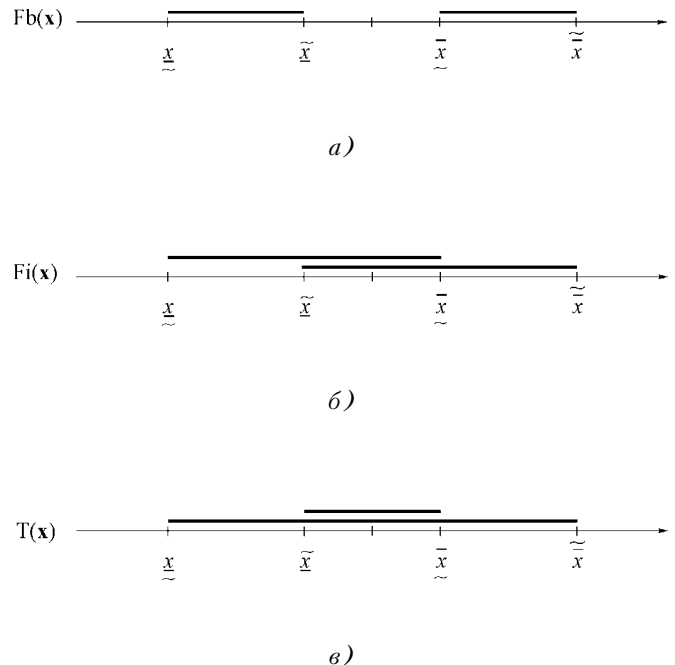


Рисунок 1 - Інтервальні структури: а) з плаваючими межами; б) з плаваючими інтервалами; в) твіні

Метою роботи є дослідження властивостей інших інтервальних структур при різних видах інтервальних даних.

1. ОСНОВНІ ОЗНАЧЕННЯ

ОЗНАЧЕННЯ 1. Інтервальні дані **a** та **b** називаються взаємно однозначними, якщо між кожним елементом інтервалу **a** існує взаємно однозначна відповідність з кожним елементом інтервалу **b**.

ОЗНАЧЕННЯ 2. Комутаційним інтервалом називається інтервал, який визначається співвідношеннями:

$$\underline{\underline{\mathbf{a}}} = [\underline{\underline{a}}; \overline{\overline{a}}] = \begin{cases} \underline{\underline{a}}, & (x < x_0) \\ \overline{\overline{a}}, & (x > x_0) \end{cases}, \quad (1)$$

де $\underline{\underline{a}}$ і $\overline{\overline{a}}$ - нижня і верхня межі комутаційного інтервалу; x_0 - значення параметру x , яке розмежовує використання меж $\underline{\underline{a}}$ і $\overline{\overline{a}}$.

ОЗНАЧЕННЯ 3. Інтервальною структурою з плаваючими межами називається структура виду:

$$\text{Fb}(\mathbf{x}) = (\underline{\mathbf{x}}; \overline{\mathbf{x}}) = ([\underline{x}_m; \underline{x}_p]; [\overline{x}_m; \overline{x}_p]), \quad (2)$$

де $\underline{\mathbf{x}}$ і $\overline{\mathbf{x}}$ - нижня і верхня інтервальні межі параметра x ; x_m і x_p - значення меж інтервалу відповідно при нижньому і верхньому значенню зовнішнього впливу.

ОЗНАЧЕННЯ 4. Інтервальною структурою з плаваючими інтервалами називається структура:

$$\text{Fi}(\mathbf{x}) = (\mathbf{x}_m; \mathbf{x}_p) = ([\underline{x}_m; \overline{x}_m]; [\underline{x}_p; \overline{x}_p]), \quad (3)$$

де \mathbf{x}_m і \mathbf{x}_p - інтервальний параметр при нижньому і верхньому значенні зовнішнього впливу.

Твін це пара інтервалів, які визначають внутрішню і зовнішню оцінку [7]. З використанням прийнятих означень твін записується у вигляді:

$$\text{T}(\mathbf{x}) = (\mathbf{x}_l; \mathbf{x}_g) = ([\underline{x}_p; \overline{x}_m]; [\underline{x}_m; \overline{x}_p]), \quad (4)$$

де \mathbf{x}_l та \mathbf{x}_g - внутрішня і зовнішня оцінки інтервалу \mathbf{x} .

2. АРИФМЕТИЧНІ ОПЕРАЦІЇ НАД ІНТЕРВАЛЬНИМИ СТРУКТУРАМИ

ОЗНАЧЕННЯ 5. Арифметичні операції між взаємно однозначними інтервалами визначаються співвідношеннями

$$\mathbf{a} * \mathbf{b} = [\underline{a}; \overline{a}] * [\underline{b}; \overline{b}] = [\underline{a} * \underline{b}; \overline{a} * \overline{b}], \quad (5)$$

$$\left(\frac{d}{dx} (\mathbf{a}(x) * \mathbf{b}(x)) \neq 0, \quad x \in \mathbf{x} \right),$$

де $* \in \{+, -, \cdot, /\}$.

В співвідношеннях (5) використовуються інтервальні розширення функцій $\mathbf{a}(\mathbf{x})$ та $\mathbf{b}(\mathbf{x})$, де значення інтервалу \mathbf{x} береться однаковим.

ОЗНАЧЕННЯ 6. Арифметичні операції з комутаційним інтервалом виконуються за співвідношеннями:

$$\underline{\underline{\mathbf{a}}} * \mathbf{b} = [\underline{\underline{a}}; \overline{\overline{a}}] * [\underline{b}; \overline{b}] = \begin{cases} [\underline{\underline{a}} * \underline{b}; \overline{\overline{a}} * \overline{b}], & (x_0 \in \mathbf{x}) \\ [\underline{\underline{a}} * \underline{b}; \underline{\underline{a}} * \overline{b}], & (x < x_0, \forall x \in \mathbf{x}) \\ [\overline{\overline{a}} * \underline{b}; \overline{\overline{a}} * \overline{b}], & (x > x_0, \forall x \in \mathbf{x}) \end{cases} \quad (6)$$

де $* \in \{+, -, \cdot, /\}$.

НАСЛІДОК 1. Якщо у комутаційному інтервалі $\underline{\underline{a}} = \overline{\overline{a}} = a$, то при виконанні арифметичних операцій використовуються співвідношення:

$$\underline{\underline{\mathbf{a}}} * \mathbf{b} = a * [\underline{b}; \overline{b}] = [a * \underline{b}; a * \overline{b}], \quad (7)$$

$$\underline{\underline{-1}} * \mathbf{b} = \underline{\underline{-1}} \cdot [\underline{b}; \overline{b}] = [-\underline{b}; -\overline{b}]. \quad (8)$$

Останній вираз означає, що множення на $\underline{\underline{-1}}$ змінює орієнтацію інтервалу на протилежну.

Арифметичні операції над кожним із видів інтервальних структур повинні виконуватись за своїми правилами. Такі правила вже розроблені для твінів - твінна арифметика [7]. Ці правила можна було б використовувати для інших інтервальних структур з допомогою наочних перетворень, які витікають із виразів (2)-(4). Однак, найбільш просто, з використанням звичайної інтервальної арифметики, виконуються операції над структурами з плаваючими інтервалами.

ТЕОРЕМА 1 (плаваючих інтервалів). В результаті арифметичної операції над інтервальними структурами з плаваючими інтервалами утворюється множина у вигляді інтервальної структури з плаваючими інтервалами:

$$\text{Fi}(\mathbf{a}) * \text{Fi}(\mathbf{x}) = (\mathbf{a}_m; \mathbf{a}_p) * (\mathbf{x}_m; \mathbf{x}_p) = (\mathbf{a}_m * \mathbf{x}_m; \mathbf{a}_p * \mathbf{x}_p) = \text{Fi}(\mathbf{a} * \mathbf{x}), \quad (9)$$

елементи якої визначаються за правилами інтервальних операцій відповідних до структури вихідних даних, де $* \in \{+, -, \cdot, /\}$.

Доведення. Будемо вважати, що перехід від стану a_m , x_m до стану a_p , x_p відбувається миттєво, а інтервальні структури $\text{Fi}(\mathbf{a})$ та $\text{Fi}(\mathbf{x})$ не мають пам'яті. Це означає, що інформація про попередній стан інтервальної структури не доходить до наступного стану, тобто в системі $\text{Fi}(\mathbf{a}) * \text{Fi}(\mathbf{x})$ не виникає взаємодії між інтервалами, які віднесено до різних станів. Звідси витікає, що в інтервальній системі $\text{Fi}(\mathbf{a}) * \text{Fi}(\mathbf{x})$ можуть взаємодіяти тільки інтервали \mathbf{a}_m та \mathbf{x}_m і \mathbf{a}_p та \mathbf{x}_p , а для інших комбінацій

$$\mathbf{a}_m * \mathbf{x}_p = \{\emptyset\}, \quad \mathbf{a}_p * \mathbf{x}_m = \{\emptyset\}.$$

В результаті утворюється інтервальна структура з плаваючими інтервалами:

$$\text{Fi}(\mathbf{a} * \mathbf{x}) = (\mathbf{a}_m * \mathbf{x}_m; \mathbf{a}_p * \mathbf{x}_p).$$

Операції над інтервалами \mathbf{a}_m та \mathbf{x}_m проводяться за правилами, які визначаються видом вихідних даних.

НАСЛІДОК 2. Арифметичні операції над інтервальними структурами з плаваючими межами виконуються за

алгоритмом:

1. Інтервальні структури $Fb(\mathbf{a})$ та $Fb(\mathbf{x})$ з допомогою співвідношень (2) і (3) відображаються в структури з плаваючими інтервалами:

$$Fb(\mathbf{a}) \mapsto Fi(\mathbf{a}); Fb(\mathbf{x}) \mapsto Fi(\mathbf{x}).$$

2. З допомогою співвідношень (9) виконуються арифметичні операції над структурами з плаваючими інтервалами:

$$Fi(\mathbf{a}) * Fi(\mathbf{x}) = Fi(\mathbf{a} * \mathbf{x}).$$

3. Утворена інтервальна структура відображається в структуру з плаваючими межами:

$$Fi(\mathbf{a} * \mathbf{x}) \mapsto Fb(\mathbf{a} * \mathbf{x}).$$

НАСЛІДОК 3. Арифметичні операції над твінами виконуються за алгоритмом:

1. Твіни $T(\mathbf{a})$ та $T(\mathbf{x})$ з допомогою співвідношень (3) та (4) відображаються у інтервальні структури з плаваючими інтервалами:

$$T(\mathbf{a}) \mapsto Fi(\mathbf{a}); T(\mathbf{x}) \mapsto Fi(\mathbf{x}).$$

2. З допомогою співвідношень (9) виконуються арифметичні операції над структурами з плаваючими інтервалами

$$Fi(\mathbf{a}) * Fi(\mathbf{x}) = Fi(\mathbf{a} * \mathbf{x}).$$

3. Утворена інтервальна структура відображається у твінну структуру

$$Fi(\mathbf{a} * \mathbf{x}) \mapsto T(\mathbf{a} * \mathbf{x}).$$

ПРИКЛАД 1. За допомогою інтервальних структур з плаваючими інтервалами виконати внутрішню і зовнішню оцінку функції

$$f(x) = (10e^x - x)(\text{tg}x - x^2)$$

з трьома значущими цифрами після десяткової точки при $x \in [1,4; 1,5]$.

Межі інтервалів при внутрішніх оцінках округлюються у внутрішню сторону, при зовнішніх - у зовнішню. Обчислення проводяться за правилами (5) арифметики взаємно однозначних інтервалів і зводиться до утворення інтервальної структури з плаваючими інтервалами:

$$Fi(f(\mathbf{x})) = ([150,219; 513,267], [150,296; 513,429]).$$

Перетворення до структури з плаваючими інтервалами у твін

$$Fi(f(\mathbf{x})) \mapsto T(f(\mathbf{x})) = ([150,296; 513,267], [150,219; 513,429])$$

дозволяє провести сукупне оцінювання функції.

Перший твінний інтервал дає внутрішню оцінку функції $f(x)$, другий - зовнішню. Оцінки, отримані за допомогою інтервальних структур з плаваючими інтервалами, збігаються з оцінками тієї ж функції наведені в [7], де використовувалась твінна арифметика. Перевагою наведеного підходу є можливість використання більш простої інтервальної арифметики.

Кожна з інтервальних структур має свою область застосування. Твіни традиційно використовуються для внутрішньої і зовнішньої оцінок складних функцій при округлені. Інтервали з плаваючими межами зручно використовувати для оцінки стабільності пристроїв.

3. ВИКОРИСТАННЯ ІНТЕРВАЛЬНИХ СТРУКТУР ДЛЯ РОЗРАХУНКУ ДОПУСКІВ

Під дією зовнішніх впливів (зміна температури, вологості, опромінювання, старіння та інші чинники) значення параметрів радіоелектронних пристроїв змінюється (рис.2). При взаємозалежній дії впливів значення параметра набуває вигляду:

$$x = x_n + \Delta x_T + \Delta x_t + \Delta x_\beta + \Delta x_v + \dots, \quad (10)$$

де x_n - номінальне значення параметра при нормальних умовах навколишнього середовища; $\Delta x_T, \Delta x_t, \Delta x_\beta, \Delta x_v$ - приріст значення параметра, обумовлений зміною температури, старінням, зміною вологості та інтенсивності опромінювання.

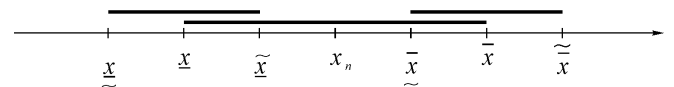


Рисунок 2 - Зміна меж параметрів при зовнішніх впливах

В загальному випадку кожен із цих приростів має нелінійну залежність від величини зовнішніх впливів

$$\Delta x_d = x_n \cdot f(d),$$

де $f(d)$ - функція зовнішнього впливу.

Але у багатьох випадках їх графіки апроксимуються лінійними або кусково-неперервними функціями. Наприклад, приріст параметра, обумовлений дією температури, апроксимується двома прямими [9], одна із яких використовується при температурах більших нормальної, а друга - при менших. Для кожної із ділянок приріст параметра записується у вигляді:

$$\Delta \underline{x} = x_n \alpha_T (T_0 - \underline{T}); \Delta \bar{x} = x_n \bar{\alpha}_T (\bar{T} - T_0) \quad (11)$$

де $\underline{\alpha}_T$ і $\bar{\alpha}_T$ - температурний коефіцієнт параметра відносно при збільшенні і зменшенні температури; T_0 - температура при нормальних умовах навколишнього середовища; T і \bar{T} - нижнє і верхнє значення температури.

З використанням (11) можна співвідношення (10) перетворити до вигляду:

$$x = x_n + x_n \Delta T \left(\alpha_T + \frac{\Delta x_t}{x_n \cdot \Delta T} + \frac{\Delta x_\beta}{x_n \cdot \Delta T} + \frac{\Delta x_v}{x_n \cdot \Delta T} + \dots \right) = x_n (1 + \alpha \cdot d), \quad (12)$$

де $\Delta T = T - T_0$ - приріст температури; α - зведений коефіцієнт зовнішнього впливу; $d = \Delta T$ - величина зовнішнього впливу (приріст температури).

Зняття дії зовнішніх чинників повертає більшість параметрів до їх початкового стану, за виключенням прояви незворотних змін, обумовлених старінням, деякими видами опроміювання і таке інше. Отже виникає явище, коли під час експлуатації РЕА її параметри плавають в деяких межах при відповідній зміні зовнішніх впливів. При цьому утворюються інтервали з плаваючими межами.

З використанням співвідношень (12) та (6) межі інтервальної структури $Fb(\mathbf{x})$ набувають вигляду:

$$\underline{x} = [\underline{x}_m; \underline{x}_p] = [\underline{x}(1 + \underline{\alpha}_- d_m); \underline{x}(1 + \bar{\alpha}_- d_p)]; \quad (13)$$

$$\bar{x} = [\bar{x}_m; \bar{x}_p] = [\bar{x}(1 + \underline{\alpha}_- d_m); \bar{x}(1 + \bar{\alpha}_- d_p)]. \quad (14)$$

З допомогою інтервалів з плаваючими межами можна отримати умови компенсації зовнішніх впливів.

ВЛАСТИВІСТЬ 1 (компенсації). При виконанні арифметичних операцій над інтервальними структурами з плаваючими межами

$$Fb(\mathbf{a}) * Fb(\mathbf{x}) = Fb(\mathbf{a} * \mathbf{x})$$

положення меж результату не залежить від величини зовнішніх впливів, якщо виконуються умови:

для операції складання

$$\frac{a}{x} = \frac{\bar{a}}{\bar{x}} = -\frac{\underline{\alpha}_- x}{\underline{\alpha}_- a} = -\frac{\bar{\alpha}_- x}{\bar{\alpha}_- a}, \quad (15)$$

для операції віднімання

$$\frac{a}{x} = \frac{\bar{a}}{\bar{x}} = -\frac{\underline{\alpha}_- x}{\underline{\alpha}_- a} = \frac{\bar{\alpha}_- x}{\bar{\alpha}_- a}, \quad (16)$$

де $* \in \{+, -, \cdot, /\}$; α_a і α_x - коефіцієнти зовнішньої дії на параметри a та x .

Доведення. Перетворимо інтервальні структури з пла-

ваючими межами в структури з плаваючими інтервалами і виконаємо арифметичну операцію додавання над цими структурами:

$$Fi(\mathbf{a}) * Fi(\mathbf{x}) = (a_m + x_m; a_p + x_p) = ([\underline{a}_m + \underline{x}_m; \bar{a}_m + \bar{x}_m]; [\underline{a}_p + \underline{x}_p; \bar{a}_p + \bar{x}_p]) = Fi(\mathbf{a} * \mathbf{x}). \quad (17)$$

Інтервальна структура (17) відображається в структуру з плаваючими межами

$$Fb(\mathbf{a} * \mathbf{x}) = ([\underline{a}_m + \underline{x}_m; \underline{a}_p + \underline{x}_p]; [\bar{a}_m + \bar{x}_m; \bar{a}_p + \bar{x}_p]).$$

Положення меж цієї структури не змінюється при будь-яких значеннях зовнішніх впливів, якщо виконуються умови:

$$\underline{a}_m + \underline{x}_m = \underline{a}_p + \underline{x}_p; \bar{a}_m + \bar{x}_m = \bar{a}_p + \bar{x}_p. \quad (18)$$

При підстановці співвідношень (13) та (14) і аналогічних виразів для параметра a умови (18) розпадаються і утворюються чотири умови:

$$\frac{a}{x} = -\frac{\underline{\alpha}_- x}{\underline{\alpha}_- a}; \frac{a}{x} = -\frac{\bar{\alpha}_- x}{\bar{\alpha}_- a}; \frac{\bar{a}}{\bar{x}} = -\frac{\underline{\alpha}_- x}{\underline{\alpha}_- a}; \frac{\bar{a}}{\bar{x}} = -\frac{\bar{\alpha}_- x}{\bar{\alpha}_- a}, \quad (19)$$

де α_a та α_x - коефіцієнти зовнішніх дій параметрів a та x .

Умови (19) легко об'єднуються в одне співвідношення (15). Для операції віднімання з урахуванням взаємної залежності інтервалів \mathbf{a} та \mathbf{x} умови компенсації набувають вигляду:

$$\underline{a}_m - \bar{x}_m = \underline{a}_p - \bar{x}_p; \bar{a}_m - \underline{x}_m = \bar{a}_p - \underline{x}_p,$$

а підстановка меж із (13) та (14) перетворює ці умови до співвідношення (16).

У загальному випадку при операціях множення і ділення підтримувати незмінність меж інтервалів неможливо оскільки умови рівності меж інтервалів залежать від величини зовнішньої дії, наприклад, для операції множення:

$$\underline{\alpha}_- a + \underline{\alpha}_- x + \underline{\alpha}_- a \underline{\alpha}_- x d_b = 0.$$

Однак залежність положення меж інтервалів значно зменшується якщо для операції множення виконати умови:

$$\left. \begin{aligned} \underline{\alpha}_- a + \underline{\alpha}_- x = 0, & \left(\underline{\alpha}_- a \gg \underline{\alpha}_- a \underline{\alpha}_- x d_m, \underline{\alpha}_- x \gg \underline{\alpha}_- a \underline{\alpha}_- x d_m \right) \\ \bar{\alpha}_- a + \bar{\alpha}_- x = 0, & \left(\bar{\alpha}_- a \gg \bar{\alpha}_- a \bar{\alpha}_- x d_m, \bar{\alpha}_- x \gg \bar{\alpha}_- a \bar{\alpha}_- x d_m \right) \end{aligned} \right\} \quad (20)$$

Аналогічній поведінці меж при операціях ділення відповідає зміна знаку перед елементами α_a або α_x в співвідношеннях (20).

4. ВИСНОВКИ

Обчислення математичних моделей, які у своєму складі мають інтервальні структури з плаваючими межами, відбувається шляхом відображення їх у структури з плаваючими інтервалами, правила виконання арифметичних операцій з якими залежать від виду інтервальних даних і розроблені в даній роботі. Запропоновані інтервальні структури відображаються також у твіни, які використовуються для оцінювання функцій. Застосування відображення інтервальних структур значно полегшує виконання арифметичних операцій з твінами. Вірогідність результатів підтверджується збігом оцінок, отриманих запропонованим методом і відомих із публікацій.

Інтервальні структури з плаваючими межами застосовуються для урахування дії зовнішніх впливів. В цьому випадку формування елементів структур відбувається з допомогою арифметичних операцій над взаємоднозначними інтервалами.

ПЕРЕЛІК ПОСИЛАНЬ

1. А.П. Вошинин, Г.Р. Сотиров. Оптимизация в условиях неопределенности. -М.: МЭИ, 1989.- 224с.
2. А.И. Орлов. Современная прикладная статистика //Заводская лаборатория.- Т.64, №3.- С.52-60.
3. Г. Алефельд, Ю. Херцбергер. Введение в интервальные вычисления:Пер. с англ.- М.:Мир,1987.- 360с.
4. E. Kaucher. Interval Analysis in the Extended Interval Space IR//Computing Suppl.-1980.- №2.- p.65-79.
5. В.Н. Кришук, Г.Н. Шило. Анализ интервальных моделей при расчете допусков //Радиоэлектроника, Информатика, Управление.- 1999.- №1.- С.66-70.
6. S.M. Markov. Extended interval arithmetic.-C.R. Acad. Bulgare Sci., v.30, 1977.- p.1239-1242.
7. V.M. Nesterov. Interval and Twin Arithmetics// Reliable Computing, №3, 1997.- p.369-380.
8. Кришук В.М., Шило Г.М., Гапоненко М.П. Інтервальний розрахунок допусків при зовнішніх впливах //Вісник державного університету "Львівська політехніка" "Радиоелектроніка та телекомунікації".- Львів.-2000.-№387.- С.191-196
9. Цветков А.Ф. Методы расчета допусков в радиоэлектронной аппаратуре. -Рязань: РРТИ, 1970.- 131с.

III. УПРАВЛІННЯ

УДК 681.32

A FUZZY-TABU SEARCH APPROACH FOR THE PLANT LAYOUT PROBLEM

Turkbey Orhan, Alabas Cigdem

Расходы на систему погрузочно-разгрузочных работ составляют почти 20-50% общей себестоимости производственной системы. Вследствие этого, проблема размещения предприятий - одна из важнейших задач. Но, так как проблема размещения предприятий - это NP-полная проблема, имеющая комбинаторную структуру, размещение предприятий является плохо исследованной областью. В настоящей работе для проблемы размещения предприятий, которая формализуется как квадратичная проблема назначения, используется запрещающий алгоритм поиска, который является одним из интеллектуальных эвристических методов. Отличный от других известных подходов запрещающего поиска, разработанный алгоритм использует теорию нечетких множеств. Разработанный алгоритм запрещающего поиска сравнивается с классическим алгоритмом запрещающего поиска на тестовых задачах различной размерности и методом случайного поиска согласно определенным критериям эффективности функционирования типа "качество решения и номер найденной точки в пространстве решений". Согласно полученным результатам показано, что в отношении задачи размещения предприятий нечеткий запрещающий алгоритм поиска превосходит другие алгоритмы.

Materials handling system cost expenditures make up nearly 20-50% of the total production cost of a manufacturing system. Owing to this, the plant layout problem is one of the subjects which are discussed mostly. But, because the plant layout problem is a NP-complete problem and has a combinatorial structure, the plant layout is still an investigation area. In the content of this study, for the plant layout problem, which is formalized as Quadratic Assignment Problem, Tabu Search Algorithm, which is one of the Smart Heuristic Techniques, is used. Different from other tabu search/search approaches in the literature, the developed algorithm uses Fuzzy Set Theory. The developed Tabu Search Algorithm in this study is compared with Classical Tabu Search Algorithm on test problems in various sizes and Random Search Method according to some performance criteria such as "solution quality and the searched point number in the solution space". According to received results, it's seen that in the plant layout problem Fuzzy-Tabu Search Algorithm is superior to other algorithms.

1. INTRODUCTION

Plant Layout Problem (PLP) has a strategical location in the success of manufacturing systems. The main reason for this fact is that, 20-50% of the total manufacturing cost originates from material management expenses. A good solution for the PLP will assist the general productivity of the system. On the contrary, a bad layout will lead to accumulation of mid stores, collidation of material transport systems, increment of preparation times and finally will lead to forma-

tion of long tails in the system (Jajodia et al. 1992). PLP may require high cost investments in long term. Moreover, relocation or improvement of a present plant/facility is not only expensive and time consuming but also this prevents activities of the workers and material flow (Sule, 1988). But the strategical importance of plant layout design should not be neglected.

The most popular formulation used in solution of PLPs is Quadratic Assignment Problem (QAP). The other formulations used for this purpose are quadratic set covering, linear integer programming, mixed integer programming, graph theoretic models (Alvarenga et al. 2000).

Koopmans and Beckman (1957) formulized PLP as QAP for the first time. The reason for naming this problem as "quadratic assignment" is that, the variables of the object function are second degree polinomial functions and the limits also resemble linear assignment problem's limits. The aims of QAP is to assign n facilities (machines, departments or work stations) to n places optimally by minimizing the material transport cost, which is defined as multiplication of work flow and the distance travelled, as much as possible.

QAP formulation is given in equation 1.

$$\begin{aligned} \text{Minimize} \quad & \sum_{i=1}^n \sum_{p=1}^n \sum_{j=1}^n \sum_{q=1}^n f_{ip} d_{jq} x_{ij} x_{pq} \\ \text{Subject to:} \quad & \sum_{j=1}^n x_{ij} = 1 \quad \forall i \text{ and } j \\ & \sum_{i=1}^n x_{ij} = 1 \quad \forall j \text{ and } j \\ & x_{ij} = 0 \text{ or } 1. \end{aligned} \quad (1)$$

In the model given above x_{ij} decision variable will be 1 if facility i is assigned to area j , otherwise it will be 0. The work flow between facilities i and p is expressed as f_{ip} , the distance between areas j and q is expressed as d_{jq} . The object function in this problem is quadratic and it has a non-convex structure. There are more than one local optimum solutions for QAP and there are $n!$ solution points in the solution space.

2. LITERATURE SURVEY FOR QAP

QAP is a NP-complete problem. Gilmore (1963), Lawler (1963), Burkard and Stratmann (1978), Bazaraa and Elshafei (1979), Bazaraa and Sherali (1980) developed procedures that give the optimum results for this problem. But the calculation ability of the optimum procedures provides the solution of the layout problems which have at most 20 facilities (Resende et al. 1995).

Because of the combinatorial structure of the PLP and the inadequency of the optimal procedures in solution of the large sized problems, researchers directed themselves to development of the heuristic algorithms on this subject. These heuristics are classified in two, as construction and improvement methods. Construction methods start the operation with an empty area and locate the facilities by choosing them consequently till the layout is completed. On the other hand in the improvement methods, the operation starts on a present layout. This initial layout can be a result of structuring algorithm or be selected randomly. Improvement methods try to reach better layout designs by changing the places of the facility units reciprocally on the basis of the present layout. Foulds (1983), Kusiak (1990) and Burkard (1990) presented detailed surveys on heuristic algorithms.

Researchers are using smart heuristic approaches for QAP recently because smart heuristic techniques such as Genetic Algorithms (GAs), Simulated Annealing (SA), Tabu Search (TS) and Artificial Neural Networks have been applied to many problems till that time since they were developed and they gave good solutions for these problems.

SA, which is presented as independent from each other by Kirkpatrick et al. (1983) and Cerny (1985), is a probabilistic method imitating the physical annealing procedure and many researchers used this method in order to solve various PLPs. Kim and Kim (1998), Chwif et al. (1998), Wang et al. (1998), Bazarganlari and Kaebnick (1997), Meller and Bozer (1996), Yip and Pao (1994) are the researchers who have applied SA to PLPs most recently.

GAs, which is another smart heuristic algorithm, imitates the "survival of the fittest" principle of the nature. Holland (1975), DeJong (1975) and Goldberg (1989) are avant-gardes of GAs. GAs have been used in order to solve various PLPs till today. Zhang et al. (2000), Al-Hakim (2000), Gau and Meller (1999), Kochhar and Heragu (1999), Hamamoto et al. (1999), Tam and Chan (1998), Mak et al. (1998), Suresh et al. (1995) applied GAs on PLPs successfully recently.

TS, which is designed modernly by Glover (1989,1990), is a smart heuristic technique assisting the local search procedures in order to be saved from optimal traps. Abninnour-Helm and Hadley (2000), Alvarenga et al. (2000), Chittratanawat and Noble (1999), Chiang and Chiang (1998), Chiang and Kouvelis (1996) are the most recent studies on TS for PLPs.

Artificial Neural Networks, which take origin from studies of McCulloch and Pitts (1943) and Rosenblatt (1962), imi-

tates the working principle of the human brain. Rossin et al. (1999), Chung (1999), Kazurhiro et al. (1996), Liao (1994) applied Artificial Neural Networks on PLP. On the other hand, Badiru and Arif (1996) and Raoot and Rakshit (1994) applied Fuzzy Set (FS) (Zadeh, 1965) based approaches on PLP.

3. TABU SEARCH AND FUZZY SET THEORY

The philosophy of TS method is claimed by Glover and Greenberg (1989) and Hansen (1986) for the first time independently from each other. But, this name is given to the method by Glover (1989,1990) who brought the method to its present modern appearance. TS shows the route to neighborhood search methods in order to be saved from local optimal traps.

TS technique begins the search operation by any beginning solution. Adjacent solutions received from the small changes done on this solution makes up the localization of the present solution and conversion of a solution into another solution is named as moving. Classical neighborhood search methods moves to another solution with better aim values from the present neighborhood solution, they choose this solution as the new present solution. This operation is repeated till all of the adjacent solutions in the neighborhood have a worsen aim value and subsequently the solution found in the end becomes a local optimal point.

TS technique uses various methods in order to save the local search procedures from local traps. Primarily, the best solution in the present neighborhood of the solution is chosen as the new solution though it's worsen than the present solution. But the neighborhood solutions received by the prohibited movements are not included in this selection operation. TS uses tabu list(s) in order to make the opposite of a movement done before. For doing this, while moving to a solution from the present solution another movement changing this movement into its previous solution is listed in the tabu list and prohibited, in a way a tabu is made. The movements listed in the tabu list stay in this list during a certain iteration time and later on they are let to be done by ending their tabu situation. This mechanism forming the short term memory of TS prevents formation of cycles during the search and provides saving from local optimal traps by changing the route of the search. Search operation continues with the selection of the adjacent solution which has the best aim value and is not a tabu as the new present solution from the neighborhood of the present solution. These operations are repeated repeatedly till a stopping condition such as maximum iteration number is provided. Moreover, under some special circumstances TS lets some movements to be made although they are tabu. Among these special circumstances, tabu demolition principle, which is used most widely, is let to be done when tabu has a better aim value than the one of the best solution that is received by a movement throughout the search.

On the other hand some properties of the movements

made are kept in the long term memory of the tabu. In the long term memory usually the permanent and temporariness frequencies of the movements-meaning the iteration number selected for forming the present solution and iteration number of taking out of the present solution are kept. This frequency information is used for strengthening the search by focusing it on certain areas and/or for varifying the search by directing it to different areas in the solution space. In an efficient TS algorithm strengthening and varifying strategies carry a big importance. TS have been applied on many problems since it's claimed till now and quite a lot good results are obtained. Glover and Laguna (1993) give a related literature survey on the application areas of this method.

Fuzzy Set Theory (FST) was spelled in mid-60s for the first time. Zadeh (1965) set up bases of this theory. FST is developed in order to manage with the uncertainties which are not statistical in structure. The difference between fuzzy and classic sets is their membership degrees. A member in the classic set is a member of this set or not, thus membership function is a (0,1) binary function. On the contrary, in the fuzzy set membership,function can take any value between 0 and 1. In this case, a fractional membership degree defines the partial membership of this member for that set. This property is very useful when there are not clear borders defining the main aspects of the set. One of the main advantages of the fuzzy set function is that, it provides deegred passage from the situation of being a member of the set to the situation of not being a member (Hassanein and Cherlopalle, 1999).

In this study the reason for the necessity of assuming TS and FST together is the systematic path showing in TS. The strengthening and diversity strategies used in TS depend on intensifying on good solution areas and reaching different solution areas. But the balance between these two strategies is important for the performance of the algorithm. Thus, there're no definite answers for these questions such as "how much does the research intensify on the good solution areas ? and when one should direct towards the different solution areas ?". Essentially, these questions are internally related and their answers vary depending on the course of the search. That's why in this study, a new fuzzy-tabu search algorithm is developed by using a pre-defined fuzzy membership func-

tion in order to find out the answer to these questions during the search.

4. FUZZY-TABU SEARCH ALGORITHM FOR QAP

In this study for the PLP which is formulized as QAP, an heuristic algorithm depending on tabu search is developed. This algorithm is completely different from the other TS applications in the literature because it contains FS properties. The area structure used by the developed FTS, management of *tabu_list* and the determination of fuzzy *tabu tenures* are presented in the following parts.

4.1 Region structure

In the FTS algorithm, any solution vector in the solution space shows which machine (or facility, unit) is appointed to which area. The sequence of machine numbers and the areas where these machines are appointed to are explained according to the coding structure used in FTS algorithm. As an example, for a 5 size PLP, $x = \{2, 1, 3, 5, 4\}$ is the solution vector, the placement of machine 1 to area 2, machine 2 to area 1, machine 3 to area 3, machine 4 to area 5 and machine 5 to area 4 are shown. Here it's assumed that all of the sizes of all machines are suitable for placing to every area.

According to the region structure the adjacents of a present solution are received by the inter-change movements which are possible to do on this solution, that means the places of the chosen machines are changed reciprocally. Adjacent solutions received by the inter-change movements of x -vector are given in figure 1 as a sample.

According to inter-change movement mechanism, if n shows the size of the problem (area/number of machines), $n(n-1)/2$ adjacent solutions of any solutions may be received. FTS algorithm, which is a repeatedly method, investigates the region of the present solution that's made up of $n(n-1)/2$ adjacent solution in each iteration. Also, for big sized problems ($n > 20$), it's possible to limit adjacent solutions of the region width in one of its subsets.

Table 1 - Adjacent solutions received from inter-change movement of $x = \{2, 1, 3, 5, 4\}$

Inter-Changes		1	2	3	4
1	Adjacents received from inter-change of machine 2	$x_1' = \{1, 2, 3, 5, 4\}$	$x_2' = \{3, 1, 2, 5, 4\}$	$x_3' = \{5, 1, 3, 2, 4\}$	$x_4' = \{4, 1, 3, 5, 2\}$
2	Adjacents received from inter-change of machine 1	$x_5' = \{2, 3, 1, 5, 4\}$	$x_6' = \{2, 5, 3, 1, 4\}$	$x_7' = \{2, 4, 3, 5, 1\}$	Same as (1,1)
3	Adjacents received from inter-change of machine 3	$x_8' = \{2, 1, 5, 3, 4\}$	$x_9' = \{2, 1, 4, 5, 3\}$	Same as (1,2)	Same as (2,1)
4	Adjacents received from inter-change of machine 5	$x_{10}' = \{2, 1, 3, 4, 5\}$	Same as (1,4)	Same as (2,3)	Same as (3,1)

4.2 Tabu list management

In the FTS the movements which ruin a preformed inter-change movement that converts the present solution into its previous situation, are prohibited along a certain repetition number which is shown by *tabu-tenure* (they're taken into tabu class), thus this provides reaching different solution regions by changing the direction of the search in the solution space. This principle lets the FTS algorithms to be saved from the local best traps. For this purpose, a tabu list named as *tabu_list[i, j]* ($i = 1, 2, \dots, n, j = 1, 2, \dots, n$) is kept.

Tabu_list[i, j] method can be explained shortly as; the present solution $x[k]$ ($k = 1, 2, \dots, n$) shows any sequence of machines (style of location to the areas) and the adjacent solution obtained by the inter-change between machine a in area p and machine b in area q is accepted as the new present solution. In this situation, *tabu_list[i, j]* is updated as follows in order to show the *iter* repetition number for the purpose of prohibiting the departures of machines a and b from the newly assigned q and p areas sequentially,

$$(tabu_list[a, q] = iter) \text{ and } (tabu_list[b, p] = iter) .$$

Tabu_list[i, j] takes the value of the first iteration since an inter-change movement has become a tabu and during the following repetition number as *tabu_tenure* the related movement becomes the tabu. Thus, in the following reviews of the algorithm, the inter-change of machine c in area r and machine d in area s becomes tabu if one of the unequalnesses is provided.

$$(iter \leq tabu_list[c, r] + tabu_tenure) \text{ or } (iter \leq tabu_list[d, s] + tabu_tenure) .$$

Lets explain the *tabu_list[i, j]* list management with an example; the present solution shows sequence $x[k]$, $x[1] = 2, x[2] = 1, x[3] = 3, x[4] = 5, x[5] = 4$ (that means $x = \{2, 1, 3, 5, 4\}$) and the adjacent solution received by inter-change of machines 2nd and 5th in the first iteration is accepted as the new present solution. In this situation these assignments below occur:

$$\begin{aligned} Tabu_list[2, 5] &= 1, \\ Tabu_list[4, 1] &= 1, \\ x[1] &= 4, x[5] = 2. \end{aligned}$$

In the second repetition, if the adjacent solution obtained from the inter-change of machines 2 and 5 in the present solution provides one of the two inequations " $(2 \leq tabu_list[2, 5] + tabu_tenure)$ or " $(2 \leq tabu_list[5, 4] + tabu_tenure)$ ", it's a tabu. Here, because $tabu_list[2, 5] = 1$ this adjacent solution will be a tabu when the *tabu_tenure* is 1 or a value more than 1. In the FTS algorithm which value will be taken by the *tabu_tenure* is determined according to a fuzzy membership function. This subject will be emphasized in the next part.

4.3 Fuzzy-Tabu tenure

The studies made on TSA up to today show that the *tabu_tenure* has a rather efficacy on the performance of the algorithm. Choosing the *tabu_tenure* short leads to focussing of the rearch to certain areas in the solution space (that means strengthening it locally) while choosing it longer leads to direction of the search towards different areas in the solution space (that means diversity in general). But, if the *tabu_tenure* is chosen shorter than it has to be, the same solutions will be reached at the end of a certain number of repetitions, thus cycles will be formed, on the contrary if it's chosen longer than normal, the quality of the solution will decrease because we will not be able to quit from bad areas in the solution space. In both conditions the solutions will be far away from the global optimum. Due to this fact, in the literature, we encounter the studies in which dynamic tabu tenures are used instead of constant tabu tenures. Beside this, strengthening and diversity strategies are considered together in the tabu search algorithm. Using both of these two strategies and the balance between them is really important in terms of algorithm's success.

In this study, a different approach is developed in order to determine the *tabu_tenure*. The basic logic of this approach is both to diverse and strengthen the direction of the search by choosing the *tabu_tenure* shorter for the opposite of the movements which are repeated seldomly during the search and by choosing the *tabu_tenure* longer for the opposite of the movements which are repeated frequently. But at this point the relative answers to these question "How often? or How much frequent?" show the necessity of using the fuzzy set theory. At the end of the trials, it's understood that the fuzzy membership function which forms a same-directioned linear relation between the assignment frequency of the machines to areas and *tabu_tenure*, is suitable for the related plant layout problem. *Tabu_tenure_lower* shows the lower limit of *tabu_tenure*, *tabu_tenure_upper* shows the upper limit of *tabu_tenure* and μ_{ij} shows the frequency of the assignment of machine i and j which is normalized between 0 and 1, the choosen fuzzy membership function is shown in equation 2. Graph of the fuzzy membership function is given in figure 1.

$$tabu_tenure_{ij} = tabu_tenure_lower + (\mu_{ij} - 1)(tabu_tenure_lower - tabu_tenure_upper) . \tag{2}$$

The *tabu_tenure* of an inter-change movement which attains machine i to area j is calculated by equation 2. In order to determine μ_{ij} membership value, a *frequency[i, j]* ($i = 1, \dots, n, j = 1, \dots, n$) list of the accomplished attainments that are done by the inter-change movement of each iteration is formed. This list, which works as a counter, is updated as below after a movement, making the new present solution in which inter-change between machine a on area p and machine b on area q occurs.

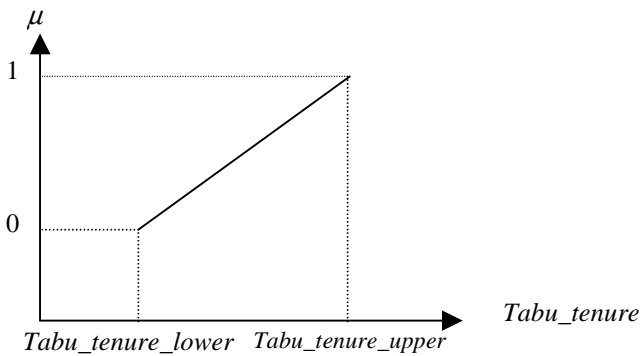


Figure 1 - Fuzzy membership function for determination of tabu tenures

$$\begin{aligned} \text{frequency}[a, q] &= \text{frequency}[a, q] + 1, \\ \text{frequency}[b, p] &= \text{frequency}[b, p] + 1. \end{aligned}$$

$\text{Frequency}[i, j]$ list is converted into μ_{ij} membership function values after normalizing it between 0 and 1. The formula given in equation 3 is used for this operation.

$$\mu_{ij} = \frac{\text{frequency}[i, j] - \min(\text{frequency}[i, j])}{\max(\text{frequency}[i, j]) - \min(\text{frequency}[i, j])}. \quad (3)$$

4.4 Steps of FTS algorithm

FTS algorithm, which is an iterational algorithm, starts (runs) the search by considering the randomly chosen beginning solution as the present solution and determines its adjacent solutions in the region of the present solution according to the inter-change movement mechanism explained in part 4.1. Among these adjacent the one which has the best aim function value, that means the lowest material handling cost and which is not a tabu regarding the fuzzy tabu_tenure in part 4.3 according to the given tabu list management in part 4.2 is chosen as the new present solution. But, when tabu destroying criterion is taken into account, if the value of the aim function of the adjacent solution, which is obtained by an in-change movement classified as tabu, is better than the value of the best aim function found during the search, this adjacent solution is accepted as the new present solution even if it is a tabu. The $\text{tabu_list}[i, j]$ and $\text{frequency}[i, j]$ ($i = 1, \dots, n, j = 1, \dots, n$) lists are updated after choosing the new present solutions. All of these procedures are repeated till the maximum iteration number, which is determined before, is reached. The steps of FTS algorithm showing the f aim function is given below.

Step 1. Choose an x initial solution randomly and assign it as the present solution:

$$x_{pres} = x.$$

The best solution $x_{the\ best} = x_{pres}$.

Give the initial value to the iteration counter: $iter = 1$.

Start the search with empty tabu and frequency lists.

Step 2. Find the adjacent of the x_{pres} solution according to inter-change movement mechanism.

$$x_{kom}^{(k)}, k = 1, \dots, n(n-1)/2.$$

Step 3. For $k = 1, \dots, n(n-1)/2$, calculate $f(x_{adj}^{(k)})$ values. Assume the adjacent solution which has the best $f(x_{adj}^{(k)})$ value and is not a tabu or provides $f(x_{adj}^{(k)}) < f(x_{the\ best})$ criterion is $x_{adj}^{(s)}$.

Step 4. Update the present solution: $x_{pres} = x_{adj}^{(s)}$.

If $f(x_{pres}) < f(x_{the\ best})$, $x_{the\ best} = x_{pres}$.

Step 5. Update the tabu and frequency lists.

Convert the frequency list into membership function by normalizing it.

Step 6. $iter = iter + 1$.

If $iter > max_iter$, stop, if not go back to step 2.

5. TRIAL STUDY

In order to show the efficacy of the developed FTS algorithm, this algorithm is compared with Classical Tabu Search (CTS) and Random Search (RS) algorithms, which use constant tabu_tenure . The only difference between CTS and FTS algorithms is that CTS algorithm considers a constant tabu_tenure after pre-trials. CTS algorithm considers the same principles completely with algorithm. The only difference between them is that tabu tenures are determined according to fuzzy membership function in FTS algorithm while a constant tabu_tenure is used is CTS algorithm. The basic aim in comparison of FTS and CTS algorithms is to determine whether using the fuzzy tabu_tenure increases the efficacy of fuzzy search method on QAP or not. On the other hand in RS algorithm, which is the simplest heuristic search method, a certain sampling is done from the solution space randomly and the solution with the lowest aim function value among these solutions is accepted as the result of the related trial. The sampling size, which is the only parameter of RS algorithm, is accepted equal to the number of solutions searched by FTS and CTS algorithms in solution space in order to provide the comparison on the same base.

In order to compare the performances of the algorithm, their small and large sized QAPs we produced randomly. For small problems with 5, 6, 7, 8 and 9 sizes, the optimum solutions are found by the enumeration method. Inter-plant distance and flow matrices for some of these small-scaled problems are given in appendix 1. Optimum solutions for 10, 20 and 30 sized problems are not known. Also, in each size there are 5 different QAP (different distance and work flow matrices). According to this, FTS, CTS, RS algorithms are compared on $8 \times 5 = 40$ QAPs. But, before the comparisons for the purpose of determining the parameter values of CTS

and FTS algorithms, the situation, in which the algorithms have the highest success for the chosen different parameter values, is searched. The pre-trials made show that, CTS and FTS algorithms have a higher success when their parameters depend on problem size (n). In CTS algorithm $tabu_tenure$ is used as the only parameter. In FTS algorithm the parameters are the upper and lower limits of the $tabu_tenure$ function which is given in equation (2) ($tabu_tenure_lower$, $tabu_tenure_upper$). The average successes of CTS and FTS algorithms in different parameter levels determined according to the problem size are given in table 2 and table 3 sequentially. For the mentioned success the obtained average aim value is considered as criterion.

Table 2 - Average success of CTS algorithm in different parameter levels

N	Tabu_tenure			
	n/3	n/2	n	2n
5	15602,6	15602,6	15557,0	15557,0
6	16149,8	16149,8	15978,0	15978,0
7	25070,8	25021,0	24671,2	24613,0
8	29222,0	29158,4	28518,0	28518,0
9	43107,8	42969,0	42833,8	42695,0
10	43719,2	43719,2	43719,2	43470,0
20	702533,0	698096,8	692972,0	694188,2
30	480507,8	480785,2	479428,6	480155,6

In table 2, it is seen that the best result is obtained for 5-10 sized problems when the $tabu_tenure$ used in CTS algorithm is taken as two times of problem size. The success of CTS algorithm is higher when $tabu_tenure$ is considered as equal to problem size for 20 and 30 sized problems.

Table 3 - Average success of FTS algorithm in different parameter levels

N	tabu_tenure_lower - tabu_tenure_upper				
	n/3-n/2	n/3-n	n/2-n	n/2-n2	n-n2
5	15602,6	15557,0	15557,0	15557,0	15557,0
6	16107,2	15978,0	15978,0	15978,0	15978,0
7	24826,4	24807,6	24807,6	24613,0	24613,0
8	29158,4	28654,8	28881,2	28518,0	28518,0
9	42969,0	42833,8	42833,8	42695,0	42695,0
10	43719,2	43533,8	43702,4	43470,0	43470,0
20	697941,2	693958,2	694002,8	692881,6	694495,8
30	479572,2	479690,6	479557,2	480604,0	481149,6

It's understood that in 5-10 sized problems $tabu_tenure_lower$ and $tabu_tenure_upper$ parameters can be chosen as $n/2$ and $2n$ or n and $2n$ from table 3 which is formed in order to determine the parameters of FTS algorithm. In our study $n/2$ and $2n$ values are chosen. On the other hand, for 20 sized problems the best success is obtained when $tabu_tenure_lower$ and $tabu_tenure_upper$ parameters are chosen as $n/2$ and $2n$, for 30 sized problems the best success is obtained when they're chosen as $n/2$ and n .

After determining the parameter values, algorithms are tried on 40 different QAP for 5 times for each of them. Though FTS and CTS algorithms have deterministic structures, trials are made by choosing the starting solutions randomly. For problems with 5, 6, 7, 8 and 9 facilities the stopping condition of FTS and CTS algorithms are considered as finding the best solution or the highest repetition number that is 1.000. The highest repetition number is chosen 1.500 for size 10, 2.500 for 20 and 3.500 for 30. The stopping condition for RS algorithm is considered as the condition when the sample size is equal to the number of solutions searched by FTS and CTS in the related problems.

Average results of the problems with the best results obtained by the algorithms are given in table 4 while the data in table 4 are more generalized according to problem sizes by taking their average in table 5. Again in table 6 the best layout plans for small size problems are given. The average results of large size problems with an unknown best solution are seen in table 7. Here are the abbreviations used in these tables: SSS is the size of solution space, OMHC is the optimum material handling cost, OPDC is the optimum percentage of deviation from cost, SPSS is the searched percentage of solution space and FMHC is the found material handling cost.

It's seen in table 4 that FTS algorithm can find the best solution in all trials done for small sized problems by searching only a small part of the solution space. CTS algorithm again reaches the best solution in all trials done except the problems 7/5 and 8/4. But the solution space search percentage of CTS algorithm reaches to higher values than that of FTS algorithm as the size of the problem gets larger. On the other hand RS algorithm shows a worse performance than FTS and CTS algorithms do in terms of not only the deviation from the best solution but also the solution space search percentage. Although RS algorithm found the best solution in all trials for problems with 5 size, it searched nearly the whole solution space. Table 5, which is formed according to the averages of the data on basic of size, shows the higher success of FTS algorithms than that of CTS and RS algorithms in all small problem sizes.

Table 4 - Average success of the algorithms in all small sized test problems

QAP			FTS		CTS		RS	
Size/No	SSS	OMHC	OPDC%	SPSS%	OPDC%	SPSS%	OPDC%	SPSS%
5/1	120	317	0	11,83	0	9,5	0	100
5/2	120	5708	0	2,33	0	2,33	0	95,50
5/3	120	15557	0	3,83	0	3,83	0	100
5/4	120	9828	0	5,67	0	5,67	0	100
5/5	120	12198	0	4,67	0	4,67	0	89,01
6/1	720	18444	0	2,06	0	1,42	0,62	100
6/2	720	15978	0	1,61	0	1,50	0,27	88,64
6/3	720	19292	0	1,56	0	1,56	0,03	70,40
6/4	720	7953	0	3,64	0	6,19	1,36	94,58
6/5	720	17930	0	2,36	0	3,14	0,16	100
7/1	5040	14597	0	0,33	0	0,87	1,03	15,17
7/2	5040	26875	0	0,42	0	0,79	0,34	18,12
7/3	5040	24613	0	0,52	0	0,44	1,37	16,62
7/4	5040	24896	0	0,47	0	0,54	1,20	15,30
7/5	5040	21652	0	0,54	0,31	4,59	2,25	18,21
8/1	40320	34684	0	0,09	0	0,19	2,39	2,48
8/2	40320	32000	0	0,19	0	0,42	3,59	2,48
8/3	40320	28518	0	0,40	0	0,14	4,93	2,48
8/4	40320	31870	0	0,09	0,09	0,59	3,44	2,48
8/5	40320	25698	0	0,07	0	0,16	8,06	2,48
9/1	362880	30076	0	0,04	0	0,05	6,08	0,28
9/2	362880	42787	0	0,02	0	0,02	4,22	0,28
9/3	362880	34875	0	0,005	0	0,009	6,19	0,28
9/4	362880	36620	0	0,007	0	0,010	4,27	0,28
9/5	362880	42695	0	0,008	0	0,010	3,65	0,28

Table 5 - Average success of the algorithms according to small problem sizes

Size	FTS		CTS		RS	
	Ave.OPDC	Ave.SPSS	Ave.OPDC	Ave.SPSS	Ave.OPDC	Ave.SPSS
5	0	5,667	0	5,200	0	96,902
6	0	2,246	0	2,762	0,488	90,724
7	0	0,456	0,062	1,446	1,238	16,684
8	0	0,168	0,018	0,300	4,482	2,480
9	0	0,016	0	0,020	4,882	0,280
Average	0	1,711	0,016	1,946	2,218	41,414

Table 6 - The best layout plans for small sized problems

Size/No	The best layout plan	Size/No	The best layout plan
5/1	5 1 3 2 4	7/4	2 1 5 7 6 3 4
5/2	1 5 2 4 3	7/5	7 6 2 4 1 3 5
5/3	4 3 2 5 1	8/1	2 5 7 4 3 1 8 6
5/4	3 5 1 4 2	8/2	7 8 3 6 5 1 4 2
5/5	3 1 4 5 2	8/3	5 1 8 6 4 3 7 2
6/1	3 4 1 2 5 6	8/4	7 6 5 1 2 4 3 8
6/2	2 1 3 4 6 5	8/5	6 5 2 1 4 7 3 8
6/3	4 1 3 2 5 6	9/1	1 3 6 9 4 8 5 7 2
6/4	3 4 2 1 5 6	9/2	5 3 9 8 7 1 4 6 2
6/5	6 1 2 3 4 5	9/3	9 1 4 7 5 3 6 2 8
7/1	7 6 5 4 2 1 3	9/4	8 2 1 6 4 7 3 9 5
7/2	1 7 4 5 3 6 2	9/5	7 4 2 5 3 8 6 1 9
7/3	5 3 1 4 2 6 7		

Таблица 7 - Average material handling costs found by the algorithms in big sized test problems

Size/No	FTS	CTS	RS
10/1	45157,0	45157,0	47621,8
10/2	43470,0	43470,0	46094,4
10/3	44282,0	44282,0	47735,8
10/4	42252,0	42252,0	45416,0
10/5	54287,0	54347,8	57124,0
20/1	692881,6	692972,0	1050269,6
20/2	210781,0	211294,8	226100,4
20/3	211765,6	212208,4	228372,0
20/4	202466,2	202912,0	219157,8
20/5	204522,0	205455,0	221068,2
30/1	470554,4	471890,0	507328,4
30/2	482664,6	482770,6	519662,4
30/3	442195,6	443678,8	481238,0
30/4	356066,2	356505,6	390911,2
30/5	479557,2	479428,6	519897,0

In table 7 It's seen that RS algorithm has a rather worse success rate than FTS and CTS algorithms in all problem sizes in the comparisons made on big sized problems. Success rates of FTS and CTS algorithms are same in problems with size 10 except the problem 5. FTS algorithm is more efficient than CTS algorithm for problem 5. In the table 7, it's again seen that the usage of the fuzzy tabu tenures gives more successful results for problems with size 20. Similarly, FTS algorithm finds better results than CTS algorithm does in problems with size 30 except problem 5. The average cost of the solutions found by CTS algorithm in problem 5 with size 30 is much more lower. For big sized problems, RS algorithm

has rather worse success but FTS algorithm can reach much more better results than CTS algorithm can do in general.

In figure 2 the convergence velocities of FTS, CTS, RS algorithms on a layout problem with 20 facilities are given. As it is seen RS algorithm converges to a much more worse solution than FTS and CTS algorithms do. FTS algorithm reaches much more better solutions than CTS algorithm does even if it finds more costing solution than CTS does at the beginning of the search.

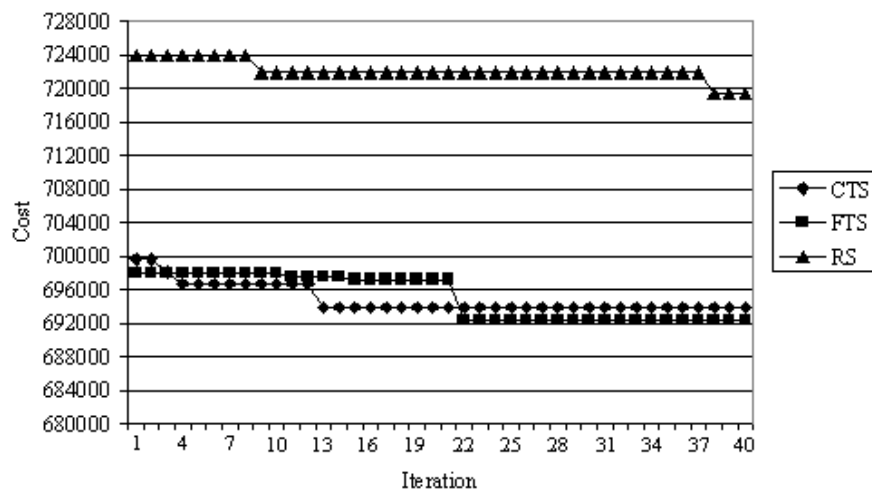


Figure 2 - Convergences of FTS, CTS and RS algorithms for a problem with 20 facilities

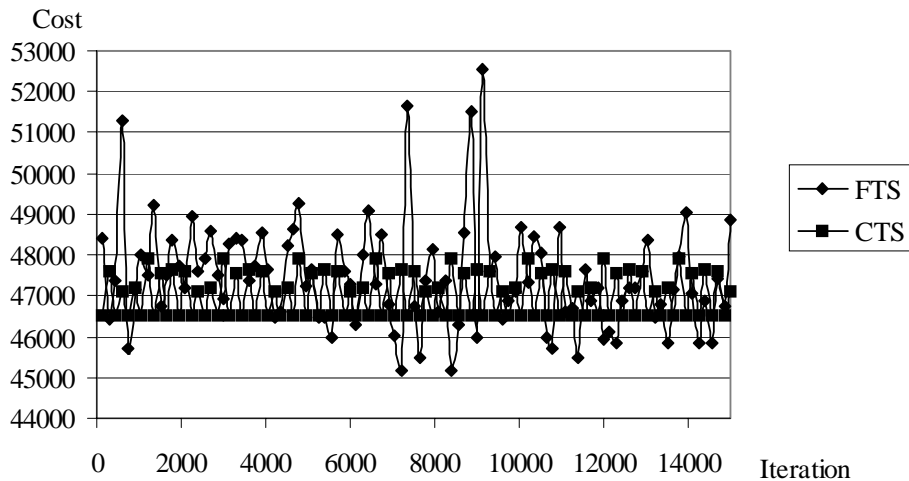


Figure 3 - Search behaviours of FTS and CTS algorithms

In figure 3 search behaviors of FTS and CTS algorithms on a layout problem with 10 facilities are seen. As understood from the figure the search in CTS algorithm intensifies at certain regions while it is directed towards different regions in FTS algorithm and eventually it reaches to better solutions than CTS algorithm does. Thus the efficacy of the fuzzy tabu tenures are shown as the only difference between these two algorithms is the usage of the fuzzy *tabu_tenure*.

6. SUMMARY

Because facility layout problem has a combinatorial structure and it is NP-complete, researchers are directed towards using heuristic approaches for facility layout. The first approaches applied to facility layout problem, which is modelled as QAP generally, have former or improver structures. These heuristics or their combinations can find a best but local solution for the problem. Due to this fact, intelligent heuristic techniques such as genetic algorithms, simulated annealing, tabu search and artificial neural networks are applied to QAPs and really good results are obtained.

In this study a tabu search algorithm which uses fuzzy tabu tenures is developed and it's tried on various sized QAPs. The developed algorithm is different from the tabu search approaches in the literature in terms of the usage of fuzzy tabu tenures. The developed tabu search algorithm is compared with classical tabu search algorithm, which uses static *tabu_tenure* and random search algorithm. Thus it is aimed to show whether using the fuzzy *tabu_tenure* will give better results or not. The obtained results show that fuzzy tabu search algorithms is more dominant than other algorithms in terms of the quality of the solutions found and the number of the points searched in solution space.

In the continuation of this study we aim to investigate whether tabu search algorithm using fuzzy tabu tenures is also effective on some other NP-hard problems or not. Also,

whether different membership functions for tabu tenures will increase the success of the algorithm or not will be investigated.

APPENDIX 1

Material handling cost matrices for some small sized problems.

Problem 5/1: Distance and flow matrix.

	1	2	3	4	5		1	2	3	4	5
1	-	2	6	10	7	1	-	0	3	5	10
2		-	8	10	9	2		-	11	0	6
3			-	3	4	3			-	8	9
4				-	12	4				-	7
5					-	5					-

Problem 6/2: Distance and flow matrix.

	1	2	3	4	5	6		1	2	3	4	5	6
1	-	97	22	92	11	57	1	-	16	43	42	9	37
2		-	56	12	70	60	2		-	33	11	25	45
3			-	52	13	15	3			-	16	37	35
4				-	23	24	4				-	35	36
5					-	76	5					-	16
6						-	6						-

Problem 7/3: Distance and flow matrix.

	1	2	3	4	5	6	7		1	2	3	4	5	6	7
1	-	70	85	74	22	15	18	1	-	45	18	16	16	19	2
2		-	51	55	76	86	36	2		-	32	24	36	2	33
3			-	70	39	40	84	3			-	32	12	16	36
4				-	45	77	11	4				-	17	22	44
5					-	84	54	5					-	31	32
6						-	57	6						-	45
7							-	7							-

Problem 8/4: Distance and flow matrix.

	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8	
1	-	83	13	66	71	48	28	46		1	-	9	6	44	34	21	18	40
2		-	91	72	94	62	58	81		2		-	12	30	23	25	2	10
3			-	31	44	53	58	64		3			-	10	17	16	32	34
4				-	93	17	64	44		4				-	28	13	18	43
5					-	28	67	97		5					-	12	40	28
6						-	92	53		6						-	17	29
7							-	26		7							-	23
8								-		8								-

Problem 9/5: Distance and flow matrix.

	1	2	3	4	5	6	7	8	9
1	-	87	22	24	89	25	61	17	81
2		-	21	91	21	62	75	69	71
3			-	73	73	59	83	61	52
4				-	69	53	22	21	69
5					-	69	87	68	98
6						-	59	21	29
7							-	25	23
8								-	90
9									-

	1	2	3	4	5	6	7	8	9
1	-	16	25	17	32	31	37	29	2
2		-	3	43	33	23	36	20	31
3			-	44	23	30	12	29	11
4				-	6	7	9	22	18
5					-	20	40	30	24
6						-	59	25	17
7							-	45	31
8								-	44
9									-

REFERENCES

[1] Abninnour-Helm, S. & Hadley, S.W. (2000). Tabu search based heuristics for multi-floor facility layout. *International Journal of Production Research*, 38(2), 365-383.

[2] Al-Hakim, L. (2000). On solving facility layout problems using genetic algorithms. *International Journal of Production Research*, 38(11), 2573-2582.

[3] Alvarenga, A.G., Negreiros-Gomes, F. & Mestria, M. (2000). Metaheuristic methods for a class of the facility layout problems. *Journal of Intelligent Manufacturing*, 11, 421-430.

[4] Badiru, A.B. & Arif, A. (1996). FLEXPART: Facility layout expert system using fuzzy linguistic relationship codes. *IIE Transactions*, 28(4), 295-308.

[5] Bazaraa, M.S. & Elshafei, A.N. (1979). An exact branch and bound procedure for quadratic assignment problems. *Naval Research Logistic Quarterly*, 26, 109-121.

[6] Bazaraa, M.S. & Sherali, A.N. (1980). Bender's partitioning scheme applied to a new formulation of the quadratic assignment problem. *Naval Research Logistic Quarterly*, 27(1), 29-41.

[7] BazarganLari, M. & Kaebnick, H. (1997). An approach to the machine layout problem in a cellular manufacturing environment. *Production Planning & Control*, 8(1), 41-55.

[8] Burkard, R.E. & Stratmann, K.H. (1978). Numerical investigation on quadratic assignment problems. *Naval Research Logistic Quarterly*, 25, 129-148.

[9] Burkard, R.E. (1990). Locations with spatial interactions: the quadratic assignment problem. *Discrete Location Theory*, (Editors: Mirchandani, P.B., Francis, R.L.), Wiley, Berlin.

[10] Cerny, V., 1985, Thermodynamical approach to traveling salesman problem: an efficient simulation algorithm, *Journal of Optimization Theory and Applications*, 45(1), 41-51.

[11] Chiang, W.C. & Chiang, C. (1998). Intelligent local search strategies for solving facility layout problems. *European Journal of Operational Research*, 106(2-3), 457-488.

[12] Chiang, W.C. & Kouvelis, P. (1996). An improved tabu search heuristic for solving facility layout design problems. *International Journal of Production Research*, 34(9), 2565-2585.

[13] Chittaranawat, S., & Noble, J.S. (1999). An integrated approach for facility layout, P/D location and material handling system design. *International Journal of Production Research*, 37(3), 683-706.

[14] Chung, Y.K. (1999). A neuro-based expert system for facility layout construction. *Journal of Intelligent Manufacturing*, 10(5), 359-385.

[15] Chwif, L., Barretto, M.R.P. & Moscato, L.A. (1998). A solution to the facility layout problem using simulated annealing. *Computers in Industry*, 36(1-2), 125-132.

[16] De Jong, K.A. (1975). An analysis of the behaviour of a class of genetic adaptive systems. PhD Thesis, University of Michigan.

[17] Foulds, L.R. (1983). Techniques for facilities layout: deciding which pairs of activities should be adjacent. *Management Science*, 29(12), 1414-1426.

[18] Gau, K.Y. & Meller, R.D. (1999). An iterative facility layout algorithm. *International Journal of Production Research*, 37(16), 3739-3758.

[19] Gilmore, P.C. (1963). Optimal and sub-optimal algorithms for the quadratic assignment problem. *SIAM Journal*, 10(2), 205-313.

[20] Glover, F., 1989, Tabu search - part I, *ORSA Journal on Computing*, 1(3), 190-206.

[21] Glover, F., 1990, Tabu search - part II, *ORSA Journal on Computing*, 2(1), 4-32.

[22] Glover, F., Greenberg, H.J., 1989, New approaches for heuristic search: a bilateral lineage with artificial intelligence, *European Journal of Operational Research*, 39, 119-130.

[23] Glover, F., Laguna, M., 1993, *Modern Heuristic Techniques For Combinatorial Problems*, Blackwell Syntefic Publications.

[24] Glover, F., Laguna, M., 1997, *Tabu Search*, Kluwer Academic Publishers, Boston.

[25] Goldberg, D.E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley.

[26] Hamamoto, S., Yih, Y. & Salvendy, G. (1999). Development and validation of genetic algorithm-based facility layout-a case study in the pharmaceutical industry. *International Journal of Production Research*, 37(4), 749-768.

[27] Hansen, P., 1986, The steepest ascent mildest descent heuristic for combinatorial programming, *Congress on Numerical Methods in Combinatorial Optimization*, Capri, Italy.

[28] Hassanein, A.A. & Cherlopalle, V. (1999). Fuzzy sets theory and range estimating. *1999 AACE International Transactions*, 4, 1-9.

[29] Holland, J.H. (1975). *Adaptation in natural and artificial systems*. University of Michigan Press.

[30] Jajodia, S., Minis, I., Harhalakis, G. & Proth, J.M. (1992). CLASS: Computerized layout solutions using simulated annealing. *International Journal of Production Research*, 30(1), 95-108.

[31] Kazurhiro, T., Sunil, B. & Yoshiyasu, T. (1996). A neural network approach to facility layout problems. *European Journal of Operational Research*, 83(3), 556-570.

[32] Kim, J.G. & Kim Y.D. (1998). A space partitioning method for facility layout problems with shape constraints. *IIE Transactions*, 30(10), 947-957.

[33] Kirkpatrick, S., Gelatt Jr., C.D., Vecchi, M.P., 1983, Optimization by simulated annealing, *Science*, 220(4598), 671-680.

[34] Kochhar, J.S. & Heragu, S.S. (1999). Facility layout design in a changing environment. *International Journal of Production Research*, 37(11), 2429-2446.

[35] Koopmans, T.C. & Beckman, M.J. (1957). Assignment problems and the location of economic activities. *Econometrica*, 25(1), 53-76.

[36] Kusiak, A. (1990). *Intelligent manufacturing systems*. Prentice-Hall, Englewood Cliffs, NJ.

[37] Lawler, E.L. (1963). The quadratic assignment problem. *Management Science*, 9(4), 586-599.

[38] Liao, T.W. (1994). Design of line-type cellular manufacturing systems for minimum operating and material-handling costs. *International Journal of Production Research*, 32(2), 387-397.

[39] Mak, K.L., Wong, Y.S. & Chan, F.T.S. (1998). A genetic algorithm for facility layout problems. *Computer Integrated Manufacturing Systems*, 11(1-2), 113-127.

[40] McCulloch, W.S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity.

- [41] Meller, R.D. & Bozer, Y.A. (1996). A new simulated annealing algorithm for the facility layout problem. *International Journal of Production Research*, 34(6), 1675-1692.
- [42] Raoot, A.D. & Rakshit, A. (1994). A fuzzy heuristic for the quadratic assignment formulation to facility layout problem. *International Journal of Production Research*, 32(3), 563-581.
- [43] Resende, M.G.C., Ramakrishnan, K.G. & Drezner, Z. (1995). Computing lower bound for the quadratic assignment problem with an interior point algorithm for linear programming. *Operations Research*, 43(5), 781-791.
- [44] Rosenblatt, F. (1962). *Principles of neurodynamics*. Washington, DC: Spartan Books.
- [45] Rossin, D.F., Springer, M.C. & Klein, B.D. (1999). New complexity measures for the facility layout problem: an empirical study using traditional and neural network analysis. *Computers & Industrial Engineering*, 36(3), 585-602.
- [46] Sule, D.R. (1988). *Manufacturing facilities*. PWS-KENT, MA.
- [47] Suresh, G., Vinod, V.V. & Sahu, S. (1995). A genetic algorithm for facility layout. *International Journal of Production Research*, 33(12), 3411-3423.
- [48] Tam, K.Y. & Chan, S.K. (1998). Solving facility layout problems with geometric constraints using parallel genetic algorithms: experimentation and findings. *International Journal of Production Research*, 36(12), 3253-3272.
- [49] Wang, T.Y., Lin, H.C. & Wu, K.B. (1998). An improved simulated annealing for facility layout problems in cellular manufacturing systems. *Computers & Industrial Engineering*, 34(2), 309-319.
- [50] Yip, P.P.C. & Pao, Y.H. (1994). A guided evolutionary simulated annealing approach to the quadratic assignment problem. *IEEE Transactions on System Man and Cybernetics*, 24(9), 1383-1387.
- [51] Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353.
- [52] Zhang, G.Q., Xue, J. & Lai, K.K. (2000). A genetic algorithm based heuristic for adjacent paper-real layout problem. *International Journal of Production Research*, 38(14), 3343-3356.

УДК 681.513:519.7

КОДИРОВАНИЕ СИГНАЛОВ В ЗАДАЧЕ НЕЙРОЭМУЛЯЦИИ

Е.В.Бодянский, Н.Е.Кулишова

В статье рассматриваются вопросы создания адаптивных искусственных нейронных сетей, работающих в реальном времени. Для обработки параметров, значения которых лежат в различных диапазонах, предложены кодирующие и декодирующие нейроны. Приведена схема нейроэмуляции объекта с применением сети, включающей нейроны-кодеры и декодеры.

В статті розглядаються питання створення адаптивних нейронних мереж, які працюють в реальному часі. Для обробки параметрів, значення яких лежать в різних діапазонах, пропонуються кодуючі та декодуючі нейрони. Наведено схему нейроемуляції об'єкту із вживанням мережі, що містить нейрони-кодерів та декодерів.

Questions of real-time adaptive neuron networks creation are researching in this paper. Parameters of real objects lies in different intervals. To process such values koding and decoding neurons are proposed. Sceme of object's neuroemulation using network with koding and decoding neurons is given.

ВВЕДЕНИЕ

Задача управления динамическими стохастическими объектами в условиях неопределенности сформировала одно из ключевых направлений в современной теории автоматического управления (ТАУ) - аппарат адаптивных систем, среди которых широкое распространение получили не прямые системы, или адаптивные системы управления с идентификатором в цепи обратной связи (АСИ) [1]. В этих системах параллельно объекту управления подключается настраиваемая модель, параметры которой уточняются в реальном времени с помощью адаптивного идентификатора, реализующего ту или иную рекуррентную процедуру оценивания. Полученные оценки параметров объекта используются далее для расчета управляющих воздействий. В качестве основного недостатка таких систем можно отметить то, что процедура их синтеза в зна-

чительной мере опирается на линеаризацию объекта управления, а в качестве настраиваемых моделей используются линейные по параметрам структуры.

Необходимость управления существенно нелинейными объектами в условиях не только параметрической, но и структурной неопределенности привела к созданию принципиально новых алгоритмов и методов управления, связанных с нейросетевыми технологиями, и сформировала новое направление в ТАУ, известное как нейроуправление [2-9]. В рамках этого подхода широкое распространение получила схема управления с нейроконтроллером и нейроэмулятором, который фактически выполняет функцию адаптивного идентификатора [7, 10-12]. В процессе обучения нейроэмулятора, в нем формируется информация, используемая далее нейроконтроллером для формирования управляющих сигналов. Естественно, что такие системы, в значительной мере обладающие интеллектуальными свойствами, характеризуются большими функциональными возможностями, чем ставшие уже традиционными АСИ.

1. ЦЕНТРИРОВАНИЕ И НОРМИРОВАНИЕ ПЕРЕМЕННЫХ В РЕАЛЬНОМ ВРЕМЕНИ

Некоторые проблемы, связанные с синтезом нейроэмуляторов, определяются необходимостью предварительной обработки поступающих на них сигналов, несущих информацию о входных и выходных переменных объекта. Такая обработка предполагается и в некоторых схемах АСИ с целью повышения скорости сходимости процесса идентификации [1] и предусматривает центрирование и нормирование наблюдаемых переменных объекта.

При этом для объекта управления с m входами $x_i(k)$, $i = 1, 2, \dots, m$ и с p выходами $y_l(k)$, $l = 1, 2, \dots, p$

(здесь $k = 0, 1, 2, \dots$ - текущее дискретное время) формируется p линейных моделей вида

$$\hat{y}_k(k) = \hat{a}_{l0} + \sum_{i=1}^m \hat{a}_{li} x_i(k), \quad (1)$$

где \hat{a}_{li} - настраиваемый параметр модели, описывающий связь между l -ым выходом и i -ым входом объекта. Переход к кодированным переменным

$$\tilde{y}(k) = \frac{y(k) - \bar{y}(k)}{\sigma_y(k)}; \tilde{x}_i(k) = \frac{x_i(k) - \bar{x}_i(k)}{\sigma_{x_i}(k)} \quad (2)$$

(здесь средние и стандарты вычисляются согласно известным выражениям

$$\bar{z}(k) = \frac{1}{k} \sum_{j=1}^k z(j), \sigma_z(k) = \sqrt{\frac{1}{k} \sum_{j=1}^k (z(j) - \bar{z}(j))^2}$$

приводит к моделям в форме

$$\hat{y}_l(k) = \sum_{i=1}^m \tilde{a}_{li} \tilde{x}_i(k), \quad (3)$$

у которых

$$\bar{y}_l(k) = \bar{x}_i(k) = 0, \sigma_{\tilde{y}_l(k)} = \sigma_{\tilde{x}_i(k)} = 1,$$

а параметры \tilde{a}_{li} - суть частные коэффициенты корреляции

$$-1 \leq \tilde{a}_{li} = r_{y_l x_i} \leq 1.$$

Между параметрами моделей (1) и (3) существует однозначная связь вида

$$\tilde{a}_{l0} = 0, \tilde{a}_{li} = \frac{\sigma_{x_i} \hat{a}_{li}}{\sigma_{y_l}}, \hat{a}_{l0} = \left(\bar{y}_l - \sum_{i=1}^m \tilde{a}_{li} \frac{\sigma_{y_l} \bar{x}_i}{\sigma_{x_i}} \right),$$

однако структура (3) может быть настроена гораздо быстрее, хотя в общем случае без такой предварительной обработки сигналов можно было бы и обойтись.

В задачах нейроэмуляции возникает более сложная ситуация. Дело в том, что модели формальных нейронов используют функции активации, области значений которых лежат в достаточно узких пределах: обычно это интервалы $[0;1]$ для сигмоидальной функции и $[-1;1]$ для гиперболического тангенса. В реальности области определения переменных реального объекта управления гораздо шире этих интервалов и подача фактических значений сигналов на эмулятор приводит к параличу нейросети. В связи с этим используются различные процедуры предварительной обработки сигналов (центрирование, нормирование, масштабирование, шкалирование)

[13-16], ставящие своей целью избежать этого крайне нежелательного явления. Любая из перечисленных процедур в отдельности не в состоянии предусмотреть все ситуации, возникающие при работе с реальным объектом, а потому представляется целесообразным ввести унифицированный алгоритм обработки и кодирования сигналов объекта перед их подачей на нейроэмулятор.

При работе в реальном времени оценки среднего и стандарта целесообразно получать с помощью рекуррентных процедур

$$\begin{aligned} \bar{z}(k) &= \bar{z}(k-1) + \frac{1}{k}(z(k) - \bar{z}(k-1)), \\ \sigma_z^2(k) &= \sigma_z^2(k-1) + \frac{1}{k}((z(k) - \bar{z}(k))^2 - \sigma_z^2(k)), \\ \sigma_z(k) &= \sqrt{\sigma_z^2(k)}, \end{aligned}$$

реализовать которые можно с помощью элементарной схемы, понятной без дополнительных пояснений. Заметим лишь, что символом z^{-1} обозначен элемент чистого запаздывания, т.е. $z^{-1}z(k) = z(k-1)$.

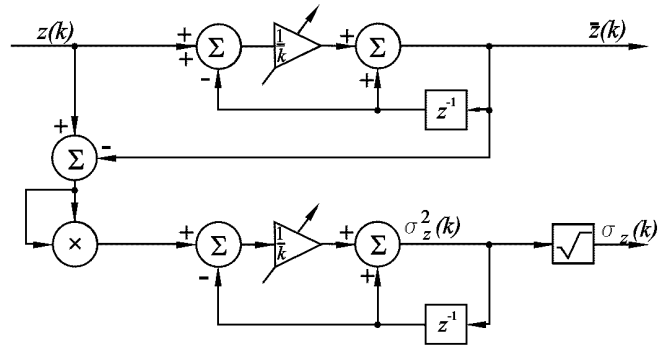


Рисунок 1 - Схема рекуррентного оценивания среднего и стандарта стохастической последовательности

Данная схема позволяет осуществлять центрирование и нормирование переменных в темпе с процессом адаптивной идентификации.

2. ПРЕОБРАЗОВАНИЕ СИГНАЛА С УЧЕТОМ ОБЛАСТИ ОПРЕДЕЛЕНИЯ ФУНКЦИЙ АКТИВАЦИИ

Для нейроэмуляции реальных объектов центрирования и нормирования сигнала по схеме 1 оказывается недостаточно. Если считать сигналы объекта стохастическими последовательностями, подчиняющимися, например, нормальному закону распределения, то область изменения центрированного и нормированного сигнала $\tilde{z}(k)$ будет лежать, как минимум, в интервале $[-3;3]$, что шире области определения стандартных функций активации. В связи с

этим необходимо провести масштабирование сигнала $\tilde{z}(k)$ с помощью линейного преобразования

$$\check{z} = b_0 + b\tilde{z}, \quad (4)$$

где параметры b_0 и b задают требуемые интервалы.

Вводя интервал возможного изменения сигнала $\tilde{z}(k)$

$$\Delta_{\check{z}} = \check{z}_{\max} - \check{z}_{\min}$$

и интервал варьирования масштабированной переменной $\check{z}(k)$;

$$\Delta_z = \check{z}_{\max} - \check{z}_{\min}$$

(обычно полагается $0, 1 \leq \check{z}(k) \leq 0, 9$ для сигмоидальной функции и $-0, 9 \leq \check{z}(k) \leq 0, 9$ для гиперболического тангенса [15]), а также масштабировующий параметр

$$\Delta_z = \frac{\Delta_{\check{z}}}{\check{z}_{\max} - \check{z}_{\min}} = \frac{\check{z}_{\max} - \check{z}_{\min}}{\check{z}_{\max} - \check{z}_{\min}},$$

несложно записать преобразование (4) сигнала $\tilde{z}(k)$ в виде

$$\check{z}(k) = \Delta_{z(k)}\tilde{z}(k) + (\check{z}_{\min} - \Delta_{z(k)}\tilde{z}_{\min})$$

или с учетом (2)

$$\begin{aligned} \check{z}(k) &= \Delta_z \left(\frac{z(k) - \bar{z}(k)}{\sigma_z(k)} - \check{z}_{\min} \right) + \check{z}_{\min} = \\ &= \frac{\Delta_z}{\sigma_z(k)} z(k) + \left(\check{z}_{\min} - \Delta_z \left(\frac{\bar{z}(k)}{\sigma_z(k)} + \check{z}_{\min} \right) \right). \end{aligned} \quad (5)$$

Для реализации выражения (5) целесообразно ввести в рассмотрение нейрон-кодер входных сигналов нейроэмулятора $x_i(k)$, схема которого приведена на рисунке 2, а условное обозначение - на рисунке 3.

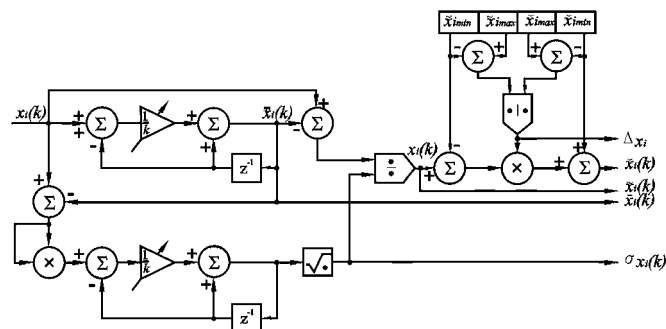


Рисунок 2 - Схема нейрона-кодера входного сигнала

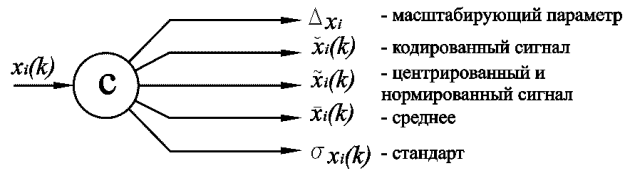


Рисунок 3 - Условное обозначение нейрона-кодера

Не менее важную проблему, чем кодирование входных сигналов, представляет обратное преобразование - декодирование к реальным значениям переменных объекта управления. Для решения этой задачи можно воспользоваться соотношениями

$$\begin{cases} \tilde{z}(k) = \check{z}(k)\Delta_z^{-1} + (\check{z}_{\min} - \Delta_z^{-1}\check{z}_{\min}), \\ z(k) = \check{z}(k)\sigma_z + \bar{z}(k), \end{cases}$$

или в объединенной форме

$$z(k) = \frac{\sigma_z(k)}{\Delta_z} \check{z}(k) + (\bar{z}(k) + \sigma_z(k)(\check{z}_{\min} - \check{z}_{\min}\Delta_z^{-1})),$$

которая может быть реализована в виде нейрона-декодера выходного сигнала $\hat{y}_i(k)$, приведенного на рис. 4, 5.

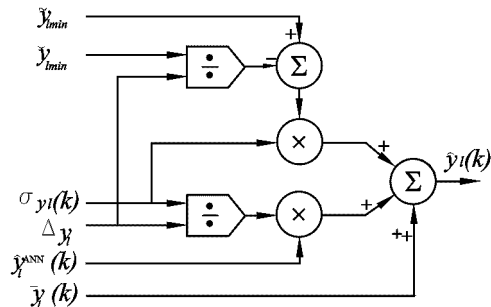


Рисунок 4 - Схема нейрона-декодера выходного сигнала

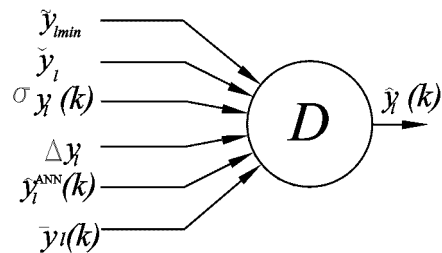


Рисунок 5 - Условное обозначение нейрона-декодера

На рисунках $\hat{y}_l^{ANN}(k)$ обозначает сигнал на выходе нейроэмулятора $\check{y}_{l\min} \leq \hat{y}_l^{ANN}(k) \leq \check{y}_{l\max}$, а $\hat{y}_l(k)$ - его декодированное значение, являющееся оценкой l -го выходного сигнала $y_l(k)$ реального объекта.

ЗАКЛЮЧЕНИЕ

Применение подобных кодирующих и декодирующих нейронов открывает широкие возможности в создании адаптивных нейросетевых систем, решающих широкий круг задач идентификации, управления, контроля и прогнозирования. Поскольку для обучения искусственной нейронной сети необходимо достаточно большое число итераций настройки, в таких задачах сеть взаимодействует с физическим объектом в реальном времени. Размерность векторов входных и выходных переменных для реальных объектов, как правило, превышает единицу, а интервалы изменения переменных могут различаться на несколько порядков. Поэтому в нейроэмулятор целесообразно включить дополнительный входной слой, который состоит из кодирующих нейронов, и выходной слой декодирующих нейронов, выполняющих обратное преобразование. Схема моделирования, включающая кодирующие и декодирующие нейроны, показана на рис.б.

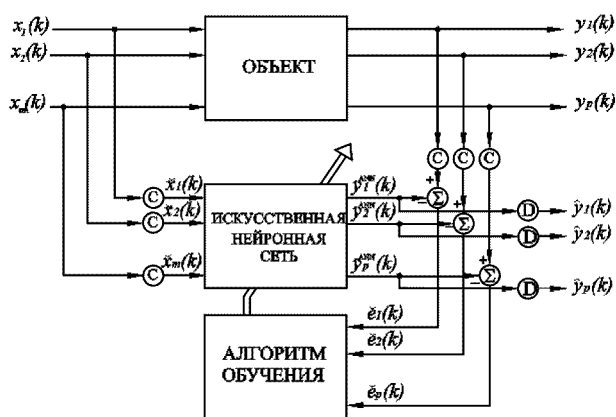


Рисунок б - Схема нейроэмуляции объекта управления

В такой системе входной вектор реального объекта $x_1(k), x_2(k), \dots, x_m(k)$ через слой кодирующих нейронов C поступает на входы нейроэмулятора, будучи преобразованным в $\check{x}_1(k), \check{x}_2(k), \dots, \check{x}_m(k)$. Алгоритм обучения с использованием выходного вектора нейронной сети $\hat{y}_1^{ANN}(k), \hat{y}_2^{ANN}(k), \dots, \hat{y}_p^{ANN}(k)$ и выходных переменных объекта $y_1(k), y_2(k), \dots, y_p(k)$, проходящих через кодирующие нейроны C , производит настройку сети. После обучения нейроэмулятор будет представлять собой модель реального объекта, при этом выходной вектор преобразуется с учетом реального масштаба переменных в $\hat{y}_1(k), \hat{y}_2(k), \dots, \hat{y}_p(k)$ с помощью декодирующих нейронов D .

Предложенная схема нейроэмулятора весьма удобна при создании интеллектуальных систем управления и является инвариантной к конкретным типам моделируемых объектов.

ПЕРЕЧЕНЬ ССЫЛОК

1. Райбман Н.С., Чадеев В.М. Построение моделей процессов производства. - М.: Энергия, 1975. - 376 с.
2. Pao Y.-H., Sobajic D.J. Nonlinear process control with neural nets// Neurocomputing. - 1990. - 2. - P. 51 - 59.
3. Neural Networks for Control/ Ed. by T. Miller. - Cambridge: MIT Press, 1991. - 524 p.
4. Neural Networks for Control and Systems/ Ed. by K. Warwick. - London: Peregrinus, 1992. - 260 p.
5. Advances in Intelligent Control/ Ed. by C.J. Harris. - London: Taylor and Francis, 1994. - 373 p.
6. Pham D.J., Liu X. Neural Networks for Identification, Prediction and Control. - London: Springer-Verlag, 1995. - 238 p.
7. Omatu S., Khalid M., Yusov R., Neurocontrol and its Applications. - London: Springer-Verlag, 1995. - 255 p.
8. Zbikowski R., Hunt K.J. Neural Adaptive Control Technology. - Singapore: World Scientific, 1996. - 347 p.
9. Noergaard M. Neural Networks for Modelling and Control of Dynamic Systems: A Practitioner's Handbook. - London: Springer, 2000. - 246 p.
10. Narendra K.S., Parthasarathy K. Identification and control of dynamical systems using neural networks// IEEE Trans. of Neural Networks. - 1990. - 1. - P. 4 - 27.
11. Zell A. Simulation Neuronale Netze. - Bonn: Addison Wesley, 1994. - 624 S.
12. Nelles O., Ernst S., Isermann R. Neuronale Netze zur Identifikation nichtlinearer dynamischer Systems: Ein Ueberblick// Automatisierungstechnik. - 1997. - 45. - № 6. - S. 251 - 262.
13. Crooks T. Care and feeding of neural networks // AI Expert. - 1992. - 7. - № 7. - P. 36 - 41.
14. Masters T. Practical Neural Network Recipes in C++. - San Diego: Academic Press, 1993. - 493 p.
15. Tsoukalas L.H., Uhrig R.E. Fuzzy and Neural Approaches in Engineering. - N.Y.: John Wiley & Sons, Inc., 1997. - 587 p.
16. Бодянский Е.В., Кулишова Н. Е., Руденко О.Г. Рекуррентна прогнозуоча штучна нейронна мережа: архітектура та алгоритми навчання// Адаптивні системи автоматичного управління. - Дніпропетровськ: Системні технології, 1999. - Вип. 2 (22). - С. 129 - 137.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ТЕХНОЛОГИЧЕСКОЙ СХЕМЫ СБОРКИ И ПРОЦЕДУРА ЕЕ ПОСТРОЕНИЯ

И.П.Гамаюн

Предлагаемая математическая модель воплощает одну из важных концепций системного анализа, которая заключается в процессуальном рассмотрении морфологического среза объекта сборки. Модель создается на основе сетей Петри и позволяет получить оценки продолжительностей цикла сборки для различных вариантов структуры собираемого объекта. Предлагается процедура построения модели.

Математична модель, що пропонується, втілює одну з важливих концепцій системного аналізу, яка полягає в процесуальному розгляді морфологічного зрізу об'єкту складання. Модель створюється на основі сіток Петрі та дозволяє одержати оцінки тривалості циклу складання для різних варіантів структури складаного об'єкту. Пропонується процедура створення моделі.

The offered mathematical model makes one from important system analysis concepts consisting in consideration of a morphological view of system as process. Model is created on Petri nets basis and allows an estimation the duration of assembly cycle for different variants of structure of assembled unit to be got. The procedure of the model creating is offered.

ВВЕДЕНИЕ

К сборочным процессам предъявляются требования высокой производительности, точности и надежности. Основными путями выполнения этих требований являются технические, конструкционные и организационные.

Технические пути являются достаточно очевидными и реализуются обычно в рамках технологического комплекса сборочного производства за счет механизации и автоматизации производства, внедрения новой техники и технологий.

Конструкционные получили широкое распространение благодаря внедрению системы DFMA, в которой основными факторами, определяющими эффективность сборки, считаются общее число деталей в конструкции изделия и легкость (простота) захвата и закрепления деталей при сборке [1].

Организационные пути в основном используются для обеспечения высокой производительности и реализуются в рамках автоматизированной подсистемы технологической подготовки сборочного производства. При этом высокая производительность достигается минимизацией производственного цикла сборки изделий. Особая актуальность минимизации производственного цикла сборки обуславливается тем, что сборка, являясь завершающей стадией производственного процесса, концентрирует значительные оборотные средства, определяемые размером незавершенного производства.

Организационные пути связывают с реализацией таких

организационных принципов как параллельность, прямая направленность, непрерывность, пропорциональность и ритмичность [2]. Анализ содержания этих принципов показывает, что их реализация достигается в основном соответствующим расчленением системного технического объекта (СТО) на целостные части, сборка которых может осуществляться независимо друг от друга различными элементами производственной системы путем выполнения согласованных, упорядоченных действий, направленных на последовательное формирование подсистем все более высокого иерархического уровня вплоть до образования собираемого СТО.

Степень реализации указанных принципов отражается различными вариантами технологической схемы сборки (ТСС), создаваемой на первом этапе разработки технологического процесса сборки. Эти варианты предопределяются вариантами структуры СТО и в рамках каждой структуры вариантами последовательности ввода в сборочный процесс различных структурных элементов СТО - подсистем СТО (сборочных единиц изделия) и элементов СТО (деталей изделия).

Однако как любая схема ТСС является всего лишь своеобразным переходом от содержательного описания процесса или объекта к их формальному описанию, исключающему неоднозначность представления и позволяющему дать количественную оценку свойствам изучаемого объекта. Поэтому актуальным является разработка математической модели, позволяющей оценить продолжительность цикла сборки для каждого варианта ТСС.

ПРЕДПОСЫЛКИ И ДОПУЩЕНИЯ

Традиционно рассматривают два типа ТСС - веерного типа и с базовым элементом.

Схемы веерного типа представляют собой древовидную структуру, элементами которой являются элементы и подсистемы СТО, распределенные по иерархическим уровням. Связи между элементами структуры отражают вхождение одних в состав других. При этом подсистемы третьего и более высоких уровней могут включать в свой состав как подсистемы любых более низких уровней, так и отдельные элементы СТО. Таким образом, схема веерного типа аналогична структуре СТО, отражающей его строение или морфологию. Процессуальной характеристикой схемы веерного типа является ступень сборки, которая соответствует уровню иерархии в структуре СТО. Переход с данной ступени сборки к следующей, более высокой,

соответствует образованию подсистем более высокого уровня из подсистем данного и всех предшествующих уровней, а также из элементов СТО. Возможность такого перехода обуславливается возможностью одновременного сопряжения с последующей его фиксацией всех составляющих частей каждой подсистемы структуры. Такая возможность представляет собой редкий случай и соответствует максимальной параллельности процесса сборки в рамках данной структуры СТО.

Достоинством схемы веерного типа является то, что она числом ступеней сборки выражает степень расчлененности или степень сложности сборки. Существенный ее недостаток заключается в том, не показывается последовательность ввода в сборочный процесс элементов структуры СТО.

Схема с базовым элементом отражает последовательность ввода в сборочный процесс отдельных элементов и подсистем СТО, а также технологических материальных элементов (компаунд, флюс и др.). При этом в ней выделяют ветви общей и узловых сборок. Ветвь общей сборки соответствует последовательности образования СТО из непосредственно составляющих его согласно структуре подсистем и отдельных элементов, а ветви узловых сборок соответствуют последовательностям образования подсистем всех уровней структуры. Ветвь общей сборки начинается с базового элемента СТО, а ветви узловых сборок начинаются с базовых элементов соответствующих подсистем.

Недостатком схемы с базовым элементом является то, что в ее описании не указываются условия, определяющие последовательность ввода в сборочный процесс структурных элементов СТО. Такими основными условиями являются конструктивно-технологические условия доступа и базирования. С учетом их можно дать следующее определение ТСС СТО.

ТСС является наглядным средством для отображения последовательности ввода в сборочный процесс структурных элементов СТО, определяемой отношениями предшествования на основе конструктивно-технологических условий доступа и базирования.

Данным определением подтверждается двойственный характер ТСС. С одной стороны, она отражает структуру СТО, то есть его морфологию, а с другой стороны, отражает процесс сборки, определяемый отношениями предшествования между элементами структуры СТО. Поэтому ТСС можно считать воплощением одной из ведущих концепций системного анализа - процессуального рассмотрения морфологического среза СТО [3]. При этом необходимо иметь в виду именно процесс сборки как процесс образования подсистем СТО и СТО в целом.

Для процесса сборки характерными являются такие особенности его реализации как параллельность, дискретность и асинхронность. Поэтому наиболее подходящей математической схемой для построения математической модели такого процесса является сеть Петри, элементы и механизмы выполнения которой предназначены для отражения

указанных особенностей [4].

Основными элементами сети Петри являются переходы и позиции. Переходы соответствуют выполняемым действиям, к числу которых в моделируемом процессе относятся действия по сопряжению элементов, фиксации сопряжения и контролю выполненного сопряжения. Позиции соответствуют условиям, предопределяющим возможность реализации действий, а также подтверждающим факт выполнения действий. Условия считаются выполненными, если в соответствующих позициях появляется такой динамический элемент как фишка (метка). Возникновение и уничтожение фишек осуществляется согласно правилам выполнения сети Петри.

Связи сети Петри определяются между ее разнотипными элементами. Они направлены от позиций к переходам и от переходов к позициям. К каждому переходу должны быть направлены связи от позиций, соответствующих условиям вхождения сопрягаемых элементов в определенную подсистему структуры СТО, и для которых в данный момент сборки выполняются условия базирования и доступа. В этом выражается учет морфологии СТО, необходимость учета которой обуславливается двойственным характером ТСС. От перехода должна быть направлена связь к позиции, соответствующей условию, которое подтверждает факт выполненного действия.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И ЕЕ ГРАФИЧЕСКАЯ ИЛЛЮСТРАЦИЯ

Иерархичность структуры СТО предопределяет иерархичность модели ТСС на основе сети Петри с указанным определением ее позиций, переходов и связей между ними. В иерархической модели ТСС позиции по уровням распределяются следующим образом:

- на нулевом уровне располагаются позиции, соответствующие условиям "имеется элемент $d_i (i \in \overline{1, n})$ СТО", которые принимают значение "истина" при наличии фишки в позиции и значение "ложь" при отсутствии фишки;

- на уровнях от первого до предпоследнего располагаются позиции, соответствующие условиям "выполнено сопряжение подмножества элементов $\{d_{i_1}, d_{i_2}, \dots, d_{i_k}\}$ ", где состав подмножества определяется составом элементов $\{d_{i_1}, d_{i_2}, \dots, d_{i_k}\}$ в подсистемах структуры СТО, для которых выполняются условия доступа и базирования;

- на последнем уровне располагается единственная позиция, соответствующая условию "выполнено сопряжение всех элементов СТО множества $D = \{d_i | (i \in \overline{1, n})\}$ ", появление в которой фишки означает завершение процесса сборки.

Между уровнями иерархии располагаются переходы, соответствующие событиям как элементарным действиям по сопряжению элементов подмножеств вида $\{d_{i_1}, d_{i_2}, \dots, d_{i_k}\}$.

Связи, отражая возможность сопряжения элементов подмножества $\{d_{i_1}, d_{i_2}, \dots, d_{i_k}\}$, направляются от позиций некоторых уровней к переходам. Кроме этого, проводится связь от каждого перехода к позиции следующего уровня, которая соответствует условию "выполнено сопряжение элементов подмножества $\{d_{i_1}, d_{i_2}, \dots, d_{i_k}\}$ ".

Например, для изделия на рисунке 1 и одного из вариантов его структуры на рисунке 2 графическая иллюстрация математической модели ТСС на основе сети Петри показывается на рисунке 3.

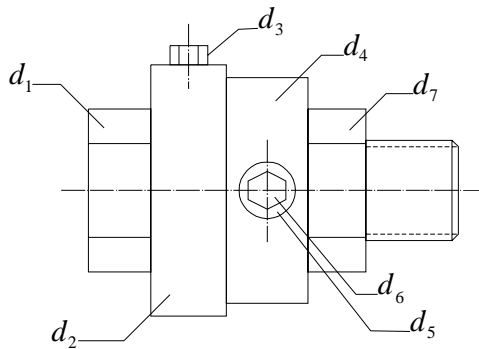


Рисунок 1

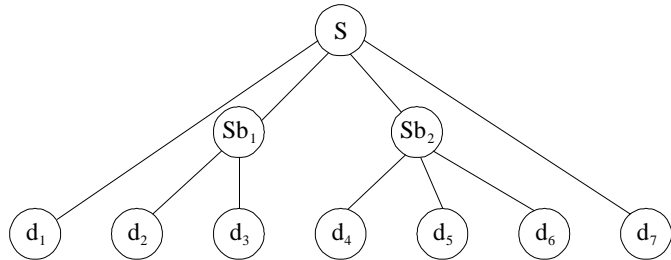


Рисунок 2

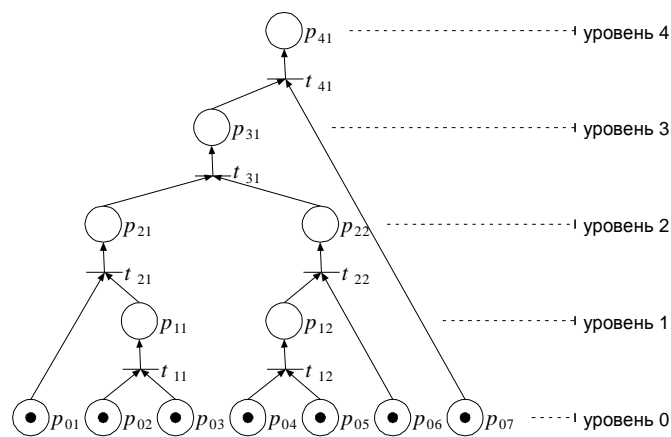


Рисунок 3

Позиции нулевого уровня $p_{0i}(i \in \overline{1,7})$ соответствуют условиям "имеется элемент d_i ". Наличие фишек в этих позициях означает, что соответствующие условия принимают значение "истина".

Определим состояние сборочного процесса распределением фишек по позициям сети. Поэтому наличие фишек в позициях нулевого уровня следует рассматривать как начальное состояние этого процесса или как готовность к реализации процесса сборки.

Связи $(p_{02}, t_{11}), (p_{03}, t_{11})$ означают возможность сопряжения элементов d_2 и d_3 согласно структуре СТО на рисунке 2 и условиям доступа и базирования, а связи $(p_{04}, t_{12}), (p_{05}, t_{12})$ означают возможность сопряжения элементов d_4 и d_5 опять согласно структуре СТО и условиям доступа и базирования.

Наличие фишек в позициях p_{02}, p_{03} предопределяет возможность запуска перехода t_{11} , а в позициях p_{04}, p_{05} - перехода t_{12} . Переходы могут быть запущены одновременно, что и демонстрирует параллельность выполнения сборочного процесса. Запуск переходов t_{11}, t_{12} , то есть выполнение действий по сопряжению элементов d_2, d_3 и d_4, d_5 приводит к удалению фишек из позиций $p_{02}, p_{03}, p_{04}, p_{05}$ и к появлению фишек в позициях p_{11}, p_{12} .

Появление фишек в позициях p_{11}, p_{12} означает, что выполнены соответственно условия "реализовано сопряжение элементов d_2, d_3 " и "реализовано сопряжение элементов d_4, d_5 ". В результате сопряжения элементов d_2, d_3 образуется подсистема Sb_1 . Таким образом, часть сети Петри, представленная позициями p_{02}, p_{03}, p_{11} и переходом t_{11} , соответствует узловой сборке подсистемы Sb_1 в ТСС с базовым элементом.

Наличие фишек в p_{11}, p_{12} означает, что сборочный процесс перешел в другое состояние, в котором, согласно правилам выполнения сети, разрешен запуск переходов t_{21}, t_{22} . Кроме этого, если в указанной на рисунке 3 сети предусмотреть механизм возвращения фишек в позиции нулевого уровня путем проведения связей от переходов, запуск которых удаляет фишки из этих позиций, то рассмотренный ранее запуск переходов t_{21}, t_{22} не только удалит бы фишки из позиций p_{02}, p_{03} и p_{04}, p_{05} , но возвратил бы их в эти же позиции. Последнее означает, что в тот момент, когда разрешен запуск переходов t_{21}, t_{22} , будет разрешен запуск еще и переходов t_{11}, t_{12} . Таким образом, в новой маркировке сети, а значит в новом состоянии сборочного процесса, возможно одновременное выполнение сразу уже четырех действий, отражаемых запуском переходов $t_{11}, t_{12}, t_{21}, t_{22}$. Запуск перехода t_{21} соответствует выполнению действия по сопряжению эле-

мента d_1 с ранее сопряженными элементами d_2, d_3 , то есть с подсистемой Sb_1 , а запуск перехода t_{22} соответствует выполнению действия по сопряжению элемента d_6 с ранее сопряженными элементами d_4, d_5 . Запуск переходов t_{11}, t_{12} соответствует повторному выполнению действий по сопряжению элементов d_2, d_3 и d_4, d_5 .

Появление фишек в позициях p_{21}, p_{22} второго уровня в результате запусков переходов t_{21}, t_{22} означает переход сборочного процесса в следующее состояние. Наличие фишки в позиции p_{21} означает, что "выполнено сопряжение d_1 и Sb_1 ", а наличие фишки в p_{22} означает, что "выполнено сопряжение d_4, d_5, d_6 ". Последнее еще означает, что образована подсистема Sb_2 . Часть сети, состоящая из позиций $p_{22}, p_{12}, p_{04}, p_{05}, p_{06}$ и переходов t_{22}, t_{12} , соответствует узловой сборке подсистемы Sb_2 .

В сети на рисунке 3 кроме указанных частей, соответствующих узловым сборкам подсистем Sb_1, Sb_2 , выделяется часть, представленная позициями $p_{01}, p_{07}, p_{21}, p_{31}, p_{41}$ и переходами t_{21}, t_{31}, t_{41} , которая соответствует основной сборке СТО.

Наличие фишек в позициях p_{21}, p_{22} предопределяет возможность запуска перехода t_{31} , реализация которого означает выполнение действия по сопряжению подсистемы Sb_2 с ранее выполненным сопряжением d_1 и Sb_1 . Запуск t_{31} приводит к образованию фишки в позиции p_{31} , что означает выполнение условия "реализовано сопряжение d_1, Sb_1, Sb_2 ". Наконец, наличие фишки в позициях p_{31}, p_{07} предопределяет возможность запуска перехода t_{41} , реализация которого приводит к образованию СТО, что подтверждается появлением фишки в позиции p_{41} . Цикл сборки СТО завершен.

Модель, графическая иллюстрация которой показана на рисунке 3, отражает двойственный характер ТСС и реализацию таких организационных принципов, как параллельность, прямая направленность, ритмичность. Параллельность обеспечивается возможностью одновременного запуска разрешенных переходов. Прямая направленность обеспечивается отсутствием обратных связей от позиций к переходам, являющимся входными для позиций более низких уровней иерархии. Ритмичность обеспечивается восстановлением маркировки позиций нулевого уровня, что позволяет возобновлять цикл сборки.

Реализация других организационных принципов - пропорциональности и непрерывности, которые связываются с согласованием во времени определенных действий, а также оценка в целом продолжительности цикла сборки, возможны на основе другой модели, являющейся модификацией рассмотренной. Такой модификацией является модель на основе временной сети Петри, которая в явном виде учитывает время выполнения действий. Графическая иллю-

страция модели на основе временной сети Петри, соответствующая структуре СТО на рисунке 2, показывается на рисунке 4.

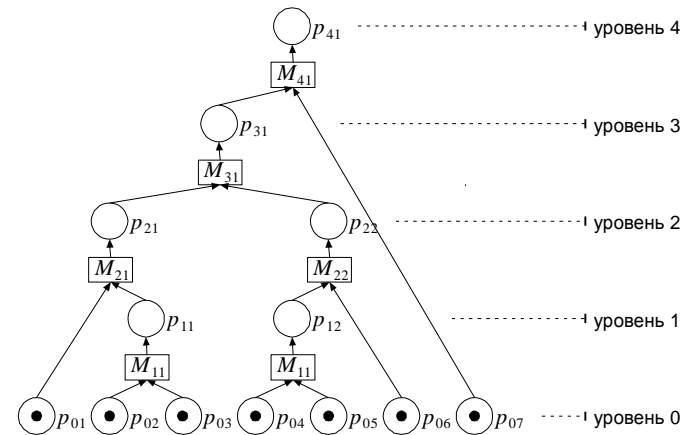


Рисунок 4

В модифицированной модели переходы представляются модулями, которые обеспечивают учет продолжительности действий с фиксацией моментов их начала и окончания. Такие модули создаются на основе обыкновенной сети Петри [5].

Моменты поступления фишек в позиции представляют важную информацию как для оценки продолжительности цикла сборки, так и для оценки степени реализации принципов пропорциональности и непрерывности.

Так момент поступления фишки в позицию, соответствующую условию "выполнено сопряжение всех элементов СТО множества $D = \{d_i | i \in \overline{1, n}\}$ " (позиция p_{41} в рассматриваемом примере), при известном моменте начала процесса сборки (момент размещения фишек в позициях нулевого уровня) позволяет определить продолжительность цикла сборки.

Моменты поступления фишек в позиции, являющиеся входными для каждого из переходов, позволяют судить о согласованности всех предшествующих действий. Если все указанные моменты совпадают, то это свидетельствует о согласованном характере предшествующих действий. В противном случае - о несогласованном характере предшествующих действий.

Степень реализации принципов пропорциональности и непрерывности может быть оценена как отношение числа переходов, во входных позициях которых фишки появляются одновременно, к общему числу переходов сети.

Предлагаемая модель относится к классу имитационных, поскольку получаемые на ее основе результаты не вычисляются как в случае аналитических моделей, а являются следствием эксперимента (имитации), заключающегося в выполнении сети Петри согласно определенным правилам.

ПРОЦЕДУРА ПОСТРОЕНИЯ МОДЕЛИ

Для реальных СТО, число элементов которых может исчисляться десятками и сотнями единиц, составление предлагаемой модели их ТСС является трудоемким процессом. Такая трудоемкость возрастает в несколько раз с учетом различных вариантов структур СТО и последовательностей ввода в сборочный процесс элементов и подсистем в рамках каждой структуры. Поэтому разработать всевозможные варианты моделей ТСС и получить на их основе оценки продолжительностей циклов сборки в приемлемые сроки возможно лишь путем реализуемых с помощью компьютера процедур. К числу таких относится процедура, основанная на последовательном применении алгоритма сборки-разборки *AA* к составным элементам структуры СТО, то есть к СТО в целом и его подсистемам. Применение *AA* к СТО и его подсистемам позволяет сформировать цепочки основной и узловых сборок в виде цепочек отношений предшествований между непосредственно их составляющих структурных элементов с учетом условий базирования и доступа [6]. Затем на каждой итерации процедуры осуществляется объединение элементов СТО, разделенных в цепочках первым слева знаком предшествования. Это те элементы, сопряжение которых допускается рассматриваемым вариантом структуры СТО и условиями базирования и доступа. Объединенные элементы заключаются в скобки и считаются новым элементом, то есть понятию элемент в цепочках придается рекурсивный характер.

Пусть, например, в результате применения алгоритма *AA*, рассматриваемого как некоторое преобразование, к подсистеме Sb_k получена цепочка отношений предшествований следующего вида:

$$AA:Sb_k \rightarrow d_{i_1} \prec d_{i_2}, d_{i_3} \prec d_{i_4} \prec \dots \prec d_{i_k}.$$

Заметим, что элементы d_{i_2} и d_{i_3} , разделенные в цепочке запятой, могут одновременно сопрягаться с элементом d_{i_1} .

На первой итерации процедуры осуществляется объединение элементов, разделенных первым слева знаком предшествования, то есть элементов $d_{i_1}, d_{i_2}, d_{i_3}$. Это объединение определяется как новый элемент и обозначается $(d_{i_1}, d_{i_2}, d_{i_3})$. Затем осуществляется преобразование цепочки так, что участок цепочки $d_{i_1} \prec d_{i_2}, d_{i_3}$ заменяется элементом $(d_{i_1}, d_{i_2}, d_{i_3})$. В результате цепочка приобретает вид

$$Sb_k \rightarrow (d_{i_1}, d_{i_2}, d_{i_3}) \prec d_{i_4} \prec \dots \prec d_{i_k}.$$

На следующей итерации осуществляется объединение элементов, разделенных первым слева знаком предшествования преобразованной цепочки. Это объединение опять определяется как новый элемент и обозначается $((d_{i_1}, d_{i_2}, d_{i_3}), d_{i_4})$.

Такие объединения и преобразования цепочки осуществляются до тех пор, пока все элементы цепочки не будут определены как новый элемент, который обозначается как $(\dots((d_{i_1}, d_{i_2}, d_{i_3}), d_{i_4}), \dots d_{i_k})$.

В результате такого объединения статус элемента получает подсистема Sb_k , что обозначается $\overline{Sb_k}$. Это позволяет осуществлять аналогичные преобразования тех цепочек, в которые входит данная подсистема.

В целом процедура построения модели реализуется как последовательность следующих действий.

1. Путем применения алгоритма *AA* к структуре СТО создаются цепочки основной и узловых сборок в виде цепочек отношений предшествований.

2. Реализуется операция объединения элементов СТО в цепочках отношений предшествований и элементы, вошедшие в объединения, определяются как новый элемент.

3. От элементов, вошедших в объединение, проводится связь к переходу, соответствующему действию по сопряжению этих элементов.

4. От каждого перехода проводится связь к позиции, соответствующей условию "выполнено сопряжение элементов объединения".

5. Проводится преобразование цепочек отношений предшествований с учетом полученных новых элементов.

6. Проверяется, получил ли статус нового элемента СТО? Если да, то процедура завершена. В противном случае, осуществляется переход к действию 2.

Реализация указанных действий на примере структуры СТО, показанной на рисунке 2, приводит к следующему результату.

Выполнение первого действия позволяет получить цепочки основной и узловых сборок в виде таких отношений предшествований:

$$AA:S \rightarrow d_1 \prec Sb_1 \prec Sb_2 \prec d_7,$$

$$AA:Sb_1 \rightarrow d_2 \prec d_3,$$

$$AA:Sb_2 \rightarrow d_4 \prec d_5 \prec d_6.$$

В первой цепочке объединение d_1 и Sb_1 не является возможным, так как Sb_1 не имеет статуса элемента. Возможно объединение элементов d_2, d_3 второй цепочки и элементов d_4, d_5 третьей цепочки. В результате объединения элементов d_2, d_3 статус нового элемента получает подсистема Sb_1 , что обозначается $\overline{Sb_1}$. Далее образуется переход t_{11} , к которому проводятся связи от позиций p_{02}, p_{03} соответствующих условиям "имеется элемент d_2 " и "имеется элемент d_3 ", а также образуется переход t_{12} , к которому проводятся связи от позиций p_{04}, p_{05} , соответствующих условиям "имеется элемент d_4 " и "имеется элемент d_5 ". Затем осуществляется преобразование цепочек, в результате которого они приобретают вид

$$AA:S \rightarrow d_1 \prec \overline{Sb}_1 \prec Sb_2 \prec d_7, \\ \overline{Sb}_1 \sim (d_2, d_3), \\ AA:Sb_2 \rightarrow (d_4, d_5) \prec d_6.$$

На следующей итерации процедуры возможно объединение элементов d_1, \overline{Sb}_1 первой цепочки, так как подсистема Sb_1 рассматривается как элемент, и элементов $(d_4, d_5), d_6$ второй цепочки. В результате объединения $(d_4, d_5), d_6$ статус нового элемента получает подсистема Sb_2 , что обозначается \overline{Sb}_2 . Образуются переходы t_{21}, t_{22} , к которым и от которых проводятся связи. Преобразованные цепочки на второй итерации приобретают вид

$$AA:S \rightarrow (d_1, \overline{Sb}_1) \prec \overline{Sb}_2 \prec d_7, \\ \overline{Sb}_1 \sim (d_2, d_3), \\ \overline{Sb}_2 \sim ((d_4, d_5), d_6).$$

На третьей итерации возможно объединение элементов только первой цепочки. Объединяются элементы (d_1, \overline{Sb}_1) и \overline{Sb}_2 . В результате объединения и преобразования цепочки приобретают вид

$$AA:S \rightarrow (d_1, \overline{Sb}_1, \overline{Sb}_2) \prec d_7, \\ \overline{Sb}_1 \sim (d_2, d_3), \\ \overline{Sb}_2 \sim ((d_4, d_5), d_6).$$

УДК 681.32:007.52

АЛГОРИТМ КЛАССИФИКАЦИИ С ОЦЕНКОЙ ЗНАЧИМОСТИ ПРИЗНАКОВ

В.И.Дубровин, С.А.Субботин

Разработан алгоритм, позволяющий осуществлять классификацию объектов и оценивать значимость их признаков. Описана нейросетевая интерпретация разработанного алгоритма. Приведены результаты экспериментов по апробации рассмотренного алгоритма.

Розроблено алгоритм, що дозволяє здійснювати класифікацію об'єктів та оцінювати значимість їхніх ознак. Описано нейронетеву інтерпретацію розробленого алгоритму. Наведені результати експериментів з апробації розглянутого алгоритму.

The algorithm permitting to realize a classification of objects and to evaluate a significance of their features is developed. The neural network interpretation of the developed algorithm is described. The results of experiments on approbation of the considered algorithm are considered.

Наконец, на четвертой итерации в результате объединения элементов $(d_1, \overline{Sb}_1, \overline{Sb}_2)$ и d_7 статус нового элемента получает СТО, что обозначается как \overline{S} . Таким образом, преобразованные цепочки на четвертой итерации приобретают вид

$$\overline{S} \sim (((d_1, \overline{Sb}_1), \overline{Sb}_2), d_7), \\ \overline{Sb}_1 \sim (d_2, d_3), \\ \overline{Sb}_2 \sim ((d_4, d_5), d_6),$$

что позволяет завершить процедуру построения математической модели ТСС, графическая иллюстрация которой была ранее показана на рисунке 3.

ПЕРЕЧЕНЬ ССЫЛОК

1. Product Design for Assembly. USA, 1991.
2. Технология и автоматизация производства радиоэлектронной аппаратуры: Учебник для вузов/ И.П. Бушминский, О.Ш. Даутов, А.П. Достанко и др.; Под ред. А.П. Достанко, И.М. Чабдурова. - М.: Радио и связь, 1989. - 624 с.
3. Николаев В.И., Брук В.М. Системотехника: методы и приложения.- Л.: Машиностроение, Ленингр. отд-ние, 1985. - 199 с.
4. Питерсон Дж. Теория сетей Петри и моделирование систем.- М., 1984. - 264 с.
5. Гамаюн И.П. Имитационное моделирование динамики технологического процесса сборки // Электронное моделирование. - 2000. - №1. - С.100-106.
6. Гамаюн И.П. Эвристический алгоритм разборки-сборки сложной машиностроительной конструкции // Механіка та машинобудування. - 1998. - №1. - С.146-149.

ВВЕДЕНИЕ

В связи с требованиями к повышению качества приборов и механизмов и все возрастающим значением сложных и дорогостоящих систем весьма актуальной в настоящее время становится задача оценки состояния сложных систем в процессе их изготовления и эксплуатации.

Требования, предъявляемые к изделиям в современном производстве, обеспечиваются не только созданием высоконадежных изделий, но и организацией надлежащего контроля, осуществляемого при их эксплуатации. Основной задачей эксплуатационного контроля является поддержание работоспособного состояния контролируемых систем в течение заданного промежутка времени.

При разработке, изготовлении и эксплуатации объекта

принимаются разнообразные меры, направленные на обеспечение его безотказности. Однако полностью устранить возможность появления отказов не удастся. Поэтому задача поддержания работоспособного состояния объекта заключается, прежде всего, в своевременном распознавании отказов и устранении причин их появления. Распознавание отказов необходимо и для того, чтобы предупредить аварийные ситуации, которые могут повлечь за собой тяжелые последствия.

Наиболее действенными методами оценки состояния объектов и прогнозирования его изменения во времени являются методы технической диагностики, которая решает задачи распознавания состояний системы, определения причин нарушения работоспособности, а также установления вида и места дефекта.

В настоящее время разработан большой арсенал методов технической диагностики, включающий статистические, вероятностные, логические [1], нечеткие и нейросетевые методы [2].

Статистические, вероятностные и логические методы характеризуются относительной простотой реализации, однако требуют достаточно большого объема экспериментальных данных и не позволяют получать высокоточные модели многомерных объектов и процессов.

В свою очередь нейросетевые методы и методы нечеткой логики позволяют строить более точные модели в условиях ограниченного набора обучающих данных, но при этом характеризуются относительной сложностью практической реализации и высокими требованиями к ресурсам ЭВМ.

Поэтому достаточно актуальной является разработка методов, обладающих относительной простотой реализации, характеризующихся относительно невысокими требованиями к ресурсам ЭВМ и при этом позволяющих строить достаточно точные модели многомерных нелинейных объектов и процессов. Подмножеством таких методов являются эвристические методы - методы не имеющие строгого теоретического обоснования, основанные на опыте и интуиции разработчика.

В настоящей работе для решения задачи диагностики предлагается использовать эвристический алгоритм, позволяющий не только осуществлять классификацию объектов, но и оценивать значимость признаков, характеризующих данные объекты.

1. АЛГОРИТМ ОЦЕНКИ ЗНАЧИМОСТИ ПРИЗНАКОВ И РАСЧЕТА ПАРАМЕТРОВ РЕШАЮЩЕГО ПРАВИЛА

Признаки, характеризующие моделируемый объект (процесс), будем условно делить на значимые и незначимые. К значимым будем относить те признаки, которые тесно связаны с выходным параметром и пренебрежение которыми может существенно ухудшить модель, к незначимым - те признаки, которые слабо связаны (или вообще не связаны) с выходным параметром и пренебрежение которыми не ухудшает модель, либо ухудшает, но не на много.

Традиционно для оценки степени связи параметров используют различные критерии, наиболее известным представителем которых является коэффициент корреляции. Однако коэффициент корреляции в основном применим лишь для оценки взаимосвязи вещественных параметров, в то время как во многих задачах диагностики необходимо получать оценку взаимосвязи вещественного (признак) и дискретного (номер класса) параметров. С другой стороны, при построении диагностических моделей желательно заранее знать, сколько потребуется разделяющих плоскостей для осуществления классификации, что нельзя оценить, исходя из коэффициента корреляции.

Объединяя вышеизложенные соображения, в качестве меры связи признака и выходного параметра (меры влияния признака на выходной параметр) будем использовать количество интервалов, на которые разбивается диапазон значений признака, таких, что экземпляры, со значением признака, попавшим в один интервал, относятся к одному и тому же классу, а экземпляры смежных интервалов относятся к разным классам. Очевидно, что такая мера позволит не только оценить значимость признака (чем меньше количество интервалов, тем больше значимость и наоборот), но и оценить необходимое количество разделяющих прямых для классификации по данному признаку. Одновременно с оценкой значимости признаков представляется возможным для каждого интервала найти граничные значения признака для экземпляров обучающей выборки, которые можно будет использовать при классификации.

Обобщая вышеизложенное, сформулируем алгоритм оценки значимости признаков и расчета параметров решающего правила

Шаг 1. Инициализация. Задать обучающую выборку экземпляров, представленную в виде массива данных p , в котором признаки линеаризованы по строкам, а экземпляры - по столбцам, а также соответствующий массив t , содержащий номера классов, сопоставленные экземплярам обучающей выборки (0 или 1). Создать массив nx , равный по размеру количеству признаков N , элементы которого будут содержать число интервалов для каждого признака. Установить $nx(i) = 0, i = 1, \dots, N$, где i - номер текущего признака. Занести количество экземпляров обучающей выборки в переменную S . Установить номер текущего признака $i=1$.

Шаг 2. Если $i \leq N$, тогда перейти на шаг 3, в противном случае - перейти на шаг 12.

Шаг 3. Занести в буфер признака x вектор значений i -го признака из обучающей выборки: $x(j) = p(i, j), j=1, \dots, S$; занести в буфер класса y копию массива t : $y(j) = t(j), j=1, \dots, S$.

Шаг 4. Отсортировать массивы x и y в порядке возрастания массива x (шаги 4.1-4.7 реализуют простейший алгоритм пузырьковой сортировки, который можно заменить на практике более быстродействующим алгоритмом).

Шаг 4.1 Установить номер текущего экземпляра обучающей выборки $j=1$.

Шаг 4.2 Если $j \leq S$, тогда перейти на шаг 4.3, в противном случае - перейти на шаг 5.

Шаг 4.3 Установить номер текущего экземпляра: $k = j + 1$.

Шаг 4.4 Если $k \leq S$, тогда перейти на шаг 4.5, в противном случае - перейти на шаг 4.7.

Шаг 4.5 Если $x(j) > x(k)$, тогда установить: $tmpx = x(j)$, $x(j) = x(k)$, $x(k) = tmpx$, $tmpy = y(j)$, $y(j) = y(k)$, $y(k) = tmpy$, где $tmpx$ и $tmpy$ - буферные переменные.

Шаг 4.6 Установить: $k=k+1$. Перейти на шаг 4.4.

Шаг 4.7 Установить: $j=j+1$. Перейти на шаг 4.2.

Шаг 5. Установить: $s = 1$, $k = 1$.

Шаг 6. Если $s \leq S$, тогда установить $tempa=x(s)$, где $tempa$ - буфер для хранения левой границы k -го интервала i -го признака, и перейти на шаг 7, в противном случае - перейти на шаг 11.

Шаг 7. Пока $(s < S)$ и $(y(s) = y(s+1))$ выполнять: $s=s+1$.

Шаг 8. Если $(s = S)$ и $(y(s) = y(s-1))$, тогда установить: $Kx(i, k) = y(s)$, $Ax(i, k) = tempa$, $Bx(i, k) = x(s)$, $k = k+1$, $s = s+1$, перейти на шаг 10. Здесь $Kx(i, k)$ - номер класса сопоставленный экземплярам обучающей выборки, значение i -го признака которых попадает внутрь k -го интервала; $Ax(i, k)$ и $Bx(i, k)$ - левая и правая границы k -го интервала i -го признака, соответственно.

Шаг 9. Если $(s < S)$ и $(y(s) \neq y(s+1))$, тогда установить: $Kx(i, k) = y(s)$, $Ax(i, k) = tempa$, $Bx(i, k) = x(s)$, $k = k+1$, $s = s+1$, $nx(i) = nx(i)+1$, в противном случае - установить: $Kx(i, k) = y(s)$, $Ax(i, k) = x(s)$, $Bx(i, k) = x(s)$, $k = k+1$, $s = s+1$.

Шаг 10 Перейти на шаг 6.

Шаг 11 Установить: $i=i+1$, перейти на шаг 2.

Шаг 12 Останов.

В результате выполнения шагов 1-12 для обучающей пары $\{p, t\}$ мы получим массив nx , содержащий для каждого признака количество интервалов на которые он разбивается (для оценки информативности признаков необходимо принять $NNx(i) = \min(nx) / nx(i)$, $i=1, \dots, N$), а также массивы Ax , Bx и Kx , содержащие информацию о границах интервалов и номерах классов, сопоставленных им для всех признаков. На основе этих массивов будем осуществлять классификацию.

2. АЛГОРИТМ КЛАССИФИКАЦИИ

Одномерную классификацию по i -му признаку будем осуществлять следующим образом. Найдем интервал, в который попадает значение признака и отнесем экземпляр по данному признаку к классу, номер которого сопоставлен интервалу, в который попало значение признака. Если

значение признака не попадает ни в один интервал из определенных в массивах Ax и Bx , тогда отнесем экземпляр по данному признаку к классу, сопоставленному экземплярам ближайшего интервала.

Произведя одномерные классификации экземпляра по всем признакам, отобразим результаты классификаций с интервала $[0,1]$ на интервал $[-1,1]$ и найдем их сумму, взвешенную с помощью коэффициентов $NNx(i)$. Очевидно, что результаты одномерной классификации для значимых признаков в этом случае будут вносить больший вклад в сумму, чем результаты классификации по малозначимым признакам. Рассчитав взвешенную сумму, отобразим ее на интервал $[0,1]$, что и будет итоговым результатом классификации экземпляра по всем признакам.

Для оценки относительной надежности классификации для нескольких экземпляров разделим модуль взвешенной суммы результатов классификации (без отображения на интервал $[0,1]$) на максимальное по модулю значение взвешенной суммы для данных экземпляров.

Обобщая вышесказанное, запишем алгоритм классификации.

Шаг 1. Инициализация. Задать массивы p , nx , Ax , Bx и Kx .

Шаг 2. Найти оценки значимости для признаков: $NNx(i) = \min(nx(i)) / nx(i)$, $i=1, \dots, N$.

Шаг 3. Установить номер текущего экземпляра $j=1$.

Шаг 4. Если $j \leq S$, тогда перейти на шаг 5, в противном случае - перейти на шаг 15

Шаг 5. Установить значение взвешенной суммы результатов одномерных классификаций j -го экземпляра $rj = 0$, номер текущего признака j -го экземпляра: $i=1$.

Шаг 6. Если $i \leq N$, тогда перейти на шаг 7, в противном случае - перейти на шаг 12

Шаг 7. Установить результат классификации для j -го экземпляра по i -му признаку $r(i)=0$;

Шаг 8. Определить интервал, к которому относится j -ый экземпляр по i -му признаку и номер класса, сопоставленный данному интервалу.

Шаг 8.1. Если $p(i, j) < Ax(i, 1)$, тогда $r(i)=Kx(i, 1)$, перейти на шаг 9, в противном случае - перейти на шаг 8.2.

Шаг 8.2. Если $p(i, j) > Bx(i, nx(i)+1)$, тогда установить $r(i) = Kx(i, nx(i)+1)$ и перейти на шаг 9, в противном случае - перейти на шаг 8.3.

Шаг 8.3. Установить: $k=1$.

Шаг 8.4. Если $k \leq nx(i)+1$, тогда перейти на шаг 8.5, в противном случае - перейти на шаг 9.

Шаг 8.5. Если $(p(i, j) \geq Ax(i, k))$ и $(p(i, j) \leq Bx(i, k))$, тогда установить $r(i)=Kx(i, k)$ и перейти на шаг 9, в противном случае - перейти на шаг 8.6.

Шаг 8.6. Если $(k < nx(i)+1)$ и $(p(i, j) > Bx(i, k))$ и $(p(i, j) < Ax(i, k+1))$, тогда перейти на шаг 8.7, в противном случае - перейти на шаг 8.8.

Шаг 8.7. Если $(Ax(i, k+1)-p(i, j)) < (p(i, j)-Ax(i, k+1))$, тог-

да установить $r(i) = Kx(i, k+1)$, в противном случае - установить $r(i) = Kx(i, k)$.

Шаг 8.8. Установить: $k=k+1$, перейти на шаг 8.4

Шаг 9. Если $r(i)>0$, тогда установить $r(i) = 1$, в противном случае установить $r(i) = -1$.

Шаг 10. Установить: $rj = rj+(r(i).*NNx(i))$.

Шаг 11. $i=i+1$, перейти на шаг 6.

Шаг 12. Установить $pr(j)=rj$, где $pr(j)$ - массив, содержащий оценки относительной уверенности (надежности) классификации.

Шаг 13. Если $rj>0$, тогда установить $t(j)=1$, в противном случае - установить $t(j)=0$. Здесь t - массив результатов классификации.

Шаг 14. Установить: $j=j+1$, перейти на шаг 4.

Шаг 15. Установить: $pr = |pr|/\max(|pr|)$.

3. НЕЙРОСЕТЕВАЯ ИНТЕРПРЕТАЦИЯ АЛГОРИТМА КЛАССИФИКАЦИИ

Рассмотренный алгоритм классификации является неитеративным и может быть использован для построения и настройки весов нейронной сети прямого распространения - многослойного персептрона [2].

Для этого функции активации для всех нейронов сети

следует задать как $\Psi(x) = \begin{cases} 0, & x \leq 0; \\ 1, & x > 0. \end{cases}$

Весовой коэффициент q -го входа p -го нейрона μ -го слоя установить в соответствии с формулой:

$$w_q^{(\mu, p)} = \begin{cases} NNx(i), \mu = 6, p = 1, q = 2i - 1; \\ -NNx(i), \mu = 6, p = 1, q = 2i; \\ 0, \mu = 6, p = 1, q = 0; \\ 0, \mu = 5, \forall p, q = 0; \\ 1, \mu = 5, p = 2i - 1, q = 1; \\ -1, \mu = 5, p = 2i, q = 1; \\ 0, \mu = 4, \forall p, q = 0; \\ 1, \mu = 4, \forall p, q > 0; \\ 1, \mu = 3, \forall p, q = 1; \\ Kx(i, k), \mu = 3, \forall p, q = 0; \\ -1, \mu = 2, \forall p, q = 0; \\ 1, \mu = 2, \forall p, q > 0; \\ 1, \mu = 1, p = 2i - 1, q = 1; \\ -1, \mu = 1, p = 2i, q = 1; \\ -Ax(i, k), \mu = 1, p = 2i - 1, q = 0; \\ Bx(i, k), \mu = 1, p = 2i, q = 0. \end{cases}$$

$k=1, \dots, nx(i); i=1, \dots, N$.

Схема персептрона, веса которого настроены по предложенной формуле представлена на рисунке 1.

Такая схема будет обеспечивать только жесткую классификацию, т.е. экземпляр может быть классифицирован только в том случае, если значения всех его признаков попадают в интервалы, значения классов для которых сопоставлены алгоритмом расчета параметров решающего правила.

Для того, чтобы осуществлять более гибкую классификацию перед созданием ее нейросетевой реализации для каждого признака следует найти все те интервалы, которым не сопоставлены номера классов, разделить эти интервалы пополам и изменить значения правой границы предыдущего и левой границы последующего смежных интервалов, которым сопоставлены номера классов, чтобы поглотить этот интервал. Например, если правая граница предыдущего интервала равна b , а левая граница последующего интервала равна a , тогда установить: $b=b+(a-b)/2$, $a=a-(a-b)/2$. После чего произвести построение нейросетевой реализации.

4. ЭКСПЕРИМЕНТЫ И РЕЗУЛЬТАТЫ

Для предложенного алгоритма были разработаны процедуры на языке программного пакета Matlab5.2 фирмы MathWorks Inc. Для исследования практической применимости разработанного метода на его основе решались различные задачи технической и медицинской диагностики, распознавания образов и прогнозирования. В частности, решались: задача диагностики изделий электронной техники по 15 признакам [3-5], задача классификации сельскохозяйственных растений на культурные и сорные по результатам дистанционного зондирования (55 признаков) [6-13], задача прогнозирования спортивного потенциала спортсменов по результатам биохимического контроля (18 признаков) [14,15], задача диагностики газотурбинных авиадвигателей (в разных экспериментах от 10000 до 100 признаков) [16-28], задача прогнозирования качества аудиохарактеристик магнитных головок (5 признаков) [29,30].

Результаты проведенных экспериментов представлены в таблице 1.

Как видно из таблицы, разработанный алгоритм позволяет с достаточно высокой точностью строить модели многомерных объектов и процессов. Достаточно высокие средние показатели надежности классификации показывают, что классификация имеет неслучайный характер.

Выгодным отличием предложенного алгоритма является то, что он позволяет не только производить классификацию, но и оценивать ее относительную надежность. Также следует отметить, что важной особенностью данного алгоритма является то, что после построения модели можно не только оценить информативность признаков, но и исключить малозначимые признаки и при этом модель не нужно будет перестраивать.

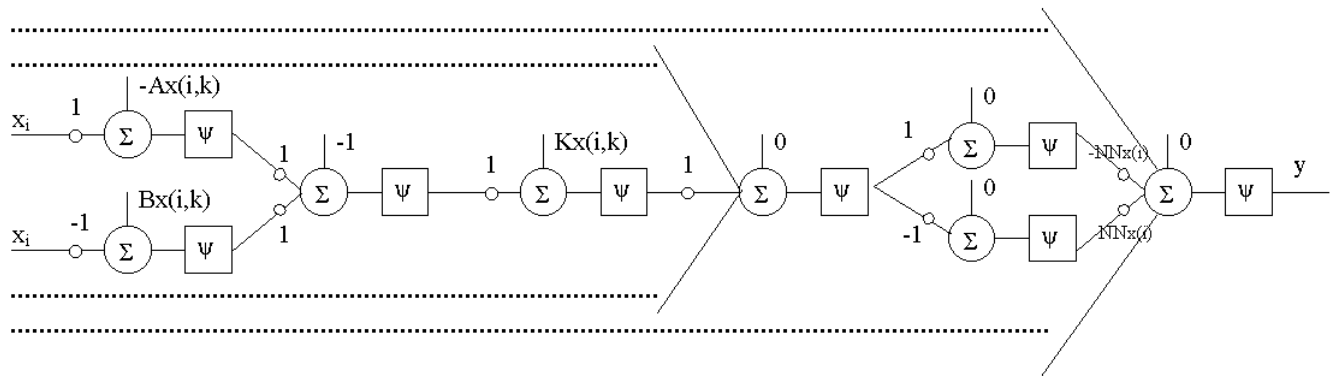


Рисунок 1 - Схема персептрона

Таблица 1 - Результаты экспериментов

Задача распознавания	Вероятность ошибки распознавания для обучающей выборки	Уровень достоверности классификации для обучающей выборки		
		минимальный	максимальный	средний
диагностика изделий электронной техники	0	0.52	1	0.82
классификация растений	0	0.9849	1	0.9995
прогнозирование спортивного потенциала	0.035	0.075	1	0.553
диагностика авиадвигателей	0	1	1	1
прогнозирование качества аудиохарактеристик магнитных головок	0.045	0.15	1	0.633

ПЕРЕЧЕНЬ ССЫЛОК

- Биргер И.А. Техническая диагностика.-М.: Машиностроение, 1978.-240 с.
- Neural Network Toolbox User Guide / Beale M., Demuth H. - Natick: Mathworks, 1997. - 700 p.
- Дубровин В.И., Субботин С.А.. Построение адаптивных систем классификации на основе нейронных сетей с латеральным торможением. // Радиоэлектроника. Информатика. Управление, 1999, №2,С. 110-114.
- Дубровин В.И., Субботин С.А. Построение систем диагностики на основе карт самоорганизации Кохонена // Нейрокомпьютеры и их применение / Сборник докладов 6-ой Всероссийской конференции, Москва 16-18 февраля 2000.-М.:Издательское предприятие журнала "Радиотехника", 2000, С. 464-467.
- Дубровин В.И., Корецкий Н.Х., Субботин С.А. Индивидуальное прогнозирование надежности изделий электронной техники на основе нейронных сетей // Труды VII Всероссийской конференции "Нейрокомпьютеры и их применение" НКП-2001 с международным участием, Москва, 14-16 февраля, 2001 г.- М.: ИПУ РАН, С. 228-231.
- Морщавка С.В., Субботин С.А., Дубровин В.И., Пиза Д.М. Нейросетевая классификация растений по результатам дистанционного зондирования // 5-й Международный молодежный форум "Радиоэлектроника и молодежь в XXI веке". Сб. научных трудов. Ч.2. - Харьков: ХТУРЭ, 2001. - С.324-325.
- Дубровин В.И., Морщавка С.В., Пиза Д. М., Субботин С.А. Применение радиально-базисных нейронных сетей для обработки данных дистанционного зондирования растений // Цифровая обработка сигналов и ее применение: 3-я Международная конференция и выставка. - М.: РНТОРЭС им.А.С. Попова, 2000.-С.48-53.
- 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.
- Дубровин В.И., Морщавка С.В., Субботин С.А., Пиза Д.М. Нейросетевая идентификация растений // Современные проблемы радиоэлектроники / Сборник научных трудов / Под ред. А.В. Сарафанова. - Красноярск: ИПЦ КГТУ, 2000.- С. 198.
- Dubrov 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. Dagli et al.-Missouri-Rolla:ASME Press, 2000, vol.10, P. 697-701.
- 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).
- Dubrov 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.
- Дубровин В.И., Морщавка С.В., Пиза Д.М., Субботин С.А. Нейросетевая классификация растений по коэффициентам спектральной яркости // Нейроинформатика и ее приложения // Материалы VIII Всероссийского семинара 6-8 октября 2000 года, Красноярск / Под общей ред. А.Н. Горбаня; Отв. за вып. Г.М. Цыбульский. - Красноярск:ИПЦ КГТУ, 2000.- С. 61-62.
- Дубровин В.И., Колесник Н.В., Субботин С.А. Нейросетевое прогнозирование состояния функционально-метаболических систем у спортсменов высшей квалификации // Моделирование неравновесных систем-2000: Материалы III Всероссийского семинара / Под.ред. А.Н. Горбаня.-Красноярск: ИПЦ КГТУ, 2000.-С. 82-83.
- Дубровин В.И., Колесник Н.В., Субботин С.А. Нейросетевая модель зависимости спортивного потенциала от данных биохимического контроля тренировочных нагрузок // Вестник новых медицинских технологий, 2000, № 3-4, С.39.
- Адаменко В.А., Дубровин В.И., Субботин С.А. Диагностика

- лопаток авиадвигателей по спектрам затухающих колебаний после ударного возбуждения на основе нейронных сетей прямого распространения // Нові матеріали і технології в металургії та машинобудуванні, 2000, № 1, С.91-96.
17. Дубровин В.И., Субботин С.А. Нейросетевая диагностика лопаток 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.
 18. Дубровин В.И., Субботин С.А. Нейросетевая диагностика газотурбинных лопаток // Оптические, радиоволновые и тепловые методы и средства контроля качества материалов, промышленных изделий и окружающей среды / Тезисы докладов VIII международной научно-технической конференции.-Ульяновск: УЛГТУ, 2000.-С.121-124.
 19. Дубровин В.И., Субботин С.А. Диагностика на основе эвристических алгоритмов в условиях ограниченного объема обучающей выборки // Proceedings of International conference "Soft computing and measurement" SCM-2000, 27-30 June 2000.-Saint-Petersburg: Saint-Petersburg State Electrotechnical University (LETI), 2000.-CD-ROM.
 20. 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-ми томах.-Т. 5.-Львів: Державний НДІ інформаційної інфраструктури, 2000.- С. 7-13.
 21. Адаменко В.А., Дубровин В.И., Субботин С.А. Нейросетевая диагностика деталей энергетических установок, работающих при циклических нагрузках // Новые технологии, методы обработки и упрочнения деталей энергетических установок: Тез. докл. Международной конференции "Новые технологии, методы обработки и упрочнения деталей энергетических установок" / Отв. ред. В.К. Яценко.-Запорожье: ЗГТУ, 2000.-С. 4-6.
 22. Дубровин В.И., Субботин С.А. Выбор информативных признаков при диагностике лопаток ГТД // Новые технологии, методы обработки и упрочнения деталей энергетических установок: Тез. докл. Международной конференции "Новые технологии, методы обработки и упрочнения деталей энергетических установок" / Отв. ред. В.К. Яценко.-Запорожье: ЗГТУ, 2000.-С.25-27.
 23. Дубровин В.И., Субботин С.А., Согорин А.А. Радиально-базисные нейронные сети в задачах технической диагностики // Интернет, освіта, наука / Друга міжнародна конференція ІОН-2000. Збірник матеріалів конференції.-Вінниця:Універсум-Вінниця, 2000.-С. 303-306.
 24. Дубровин В.И., Субботин С.А. Нейросетевая подсистема диагностического программного комплекса // Нейрокомпьютеры: разработка и применение, 2001, №2, С. 55-62.
 25. 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.
 26. Адаменко В.А., Дубровин В.И., Жеманюк П.Д., Субботин С.А. Диагностика усталостных трещин в деталях газотурбинных авиадвигателей // Надійність машин та прогнозування їх ресурсу / Доповіді міжнародної науково-технічної конференції.-В 2-х томах. Том 1.-Івано-Франківськ: ІФДТУНГ-Факел, 2000.-С. 151 - 158.
 27. Дубровин В.И., Субботин С.А. Нейросетевая диагностика в управлении качеством // Управление в технических системах - XXI век: сборник научных трудов III Международной научно-технической конференции. - Ковров: КГТА, 2000. - С.136-138.
 28. Дубровин В.И., Субботин С.А. Прогнозирование отказов деталей ГТД в процессе эксплуатации // Моделирование неравновесных систем-2000: Материалы III Всероссийского семинара / Под.ред. А.Н. Горбаня.-Красноярск: ИПЦ КГТУ, 2000.-С. 84.
 29. Dubrovin V., Subbotin S. Model of Magnetic Heads Audio Characteristics // The Experience of Designing and Application of CAD Systems in Microelectronics: Proceedings of the VI International Conference CADSM 2001.-Lviv: Publishing House of LPNU, 2001.-P. 232-233.
 30. Дубровин В.И., Субботин С.А. Нейросетевая модель зависимости акустических характеристик магнитных головок от параметров сендастового сердечника // Вестник Хакасского государственного университета им. Н.Ф. Катанова. Вып. 4. Серия 1: Информатика.-Абакан: Издательство ХГУ им. Н.Ф. Катанова, 2001.-С. 25-28.

УДК 65.012.45:620.22

АВТОМАТИЗАЦИЯ ПРОЦЕССА ИДЕНТИФИКАЦИИ СТАНДАРТНЫХ ДЕТАЛЕЙ, ПОКУПНЫХ КОМПЛЕКТУЮЩИХ ИЗДЕЛИЙ И МАТЕРИАЛОВ

Ю.З.Житников, А.А.Фаткин, Т.В.Рулева

В работе рассматривается подход автоматизированной идентификации стандартных деталей, покупных комплектующих изделий и материалов на основе параметрического описания нормативной документации, регламентирующей их условное обозначение.

The approach of the automated identification of standard details, purchased products and materials is considered on the basis of the parametrical description of the normative documentation regulating their conditional designation.

В условиях мелкосерийного, многономенклатурного производства сдерживающим фактором быстрого перехода от выпуска одного изделия к другому является подготовка производства, время которой может превосходить время выпуска данного изделия.

Такие характерные особенности как:

- многономенклатурность изделий;
- разнообразие применяемых конструкционных материалов;
- новизна и сложность изготавливаемой продукции;
- сжатые сроки изготовления, а так же высокие требования

к качеству изделий в целом и к его отдельным элементам, что накладывает специфичные ограничения на внедрение следующих прогрессивных методов управления подготовкой производства.

Прогрессивные методы управления подготовкой производства представлены в интегрированных системах (ИСУП), разработанных как западными, так и отечественными производителями. Среди этих систем имеются как настроенные под определенный тип производства, так и универсальные. И в том и в другом случае требуется дополнительная подготовка предприятия-потребителя в части технического оснащения, формализации основной исходной информации, квалификации персонала, психологической подготовки, а также заинтересованности первого руководителя.

Проблематична, особенно в условиях единичного производства, задача по формализации исходной информации. Специфика производства, которую нужно сохранить для

перспективного планирования, накладывает свой отпечаток на характер исходной информации.

Формализовать информацию - значит подготовить ее к заведению в базу данных (БД) внедряемой системы. Исходная информация должна соответствовать требованиям внедряемой системы по содержанию, формированию, актуализации, хранению, а также является основной для решения комплекса задач, повышающих эффективность управления производством.

В предложенных на рынке программных продуктах по управлению производством включены подсистемы конструкторской и технологической подготовки как средство ввода исходных данных и оперативно-диспетчерского контроля как средство интерпретации полученных результатов (рис. 1).

Однако для внедрения и поддержки исходных массивов системы, в частности, состава изделия и базы материалов необходимо закодировать стандартные единицы производства (стандартные ДСЕ, ПКИ, материалы) для оптимизации хранения и обработки информации.

Цифровой код, присваиваемый ДСЕ, ПКИ, материалу позволяет однозначно идентифицировать объект, исключая дублирование и содержит информацию о параметрах закодированного объекта (геометрических характеристиках, марке материала, покрытии и т.д.).

Процессу кодирования в условиях единичного производства отводится одно из ведущих мест на этапе технической подготовки производства. Это связано с трудоем-

костью данного процесса, обусловленной в первую очередь, многономенклатурностью заказов. Во-вторых, представление информации о составе изделия на бумажных носителях не формализовано и требует уточнения, что в конечном итоге сопряжено с использованием нормативной документации (НД). Положение усугубляется такой специфической особенностью единичного производства - как неповторяемость заказа, то есть работы по формализации данных проводятся периодически для каждого заказа с максимальным привлечением трудовых людских, информационных ресурсов. К тому же работы по формализации данных необходимо выполнять в сжатые сроки, диктуемые договором на поставку заказа, что является еще одной уникальной особенностью индивидуального производства. Решить проблему формализации данных при таких ограничениях можно только за счет оптимизации работ на данном этапе.

Анализ процесса кодирования стандартных единиц производства показал, что в основу наиболее эффективного и перспективного средства оптимизации необходимо заложить моделирование системы. А так как объект кодирования или обозначение стандартных единиц производства регламентируется нормативной документацией (НД), то предметную область модели составляют все нормативные документы, действующие на предприятии и регулирующие качество соответствующей продукции на всех этапах жизненного цикла.

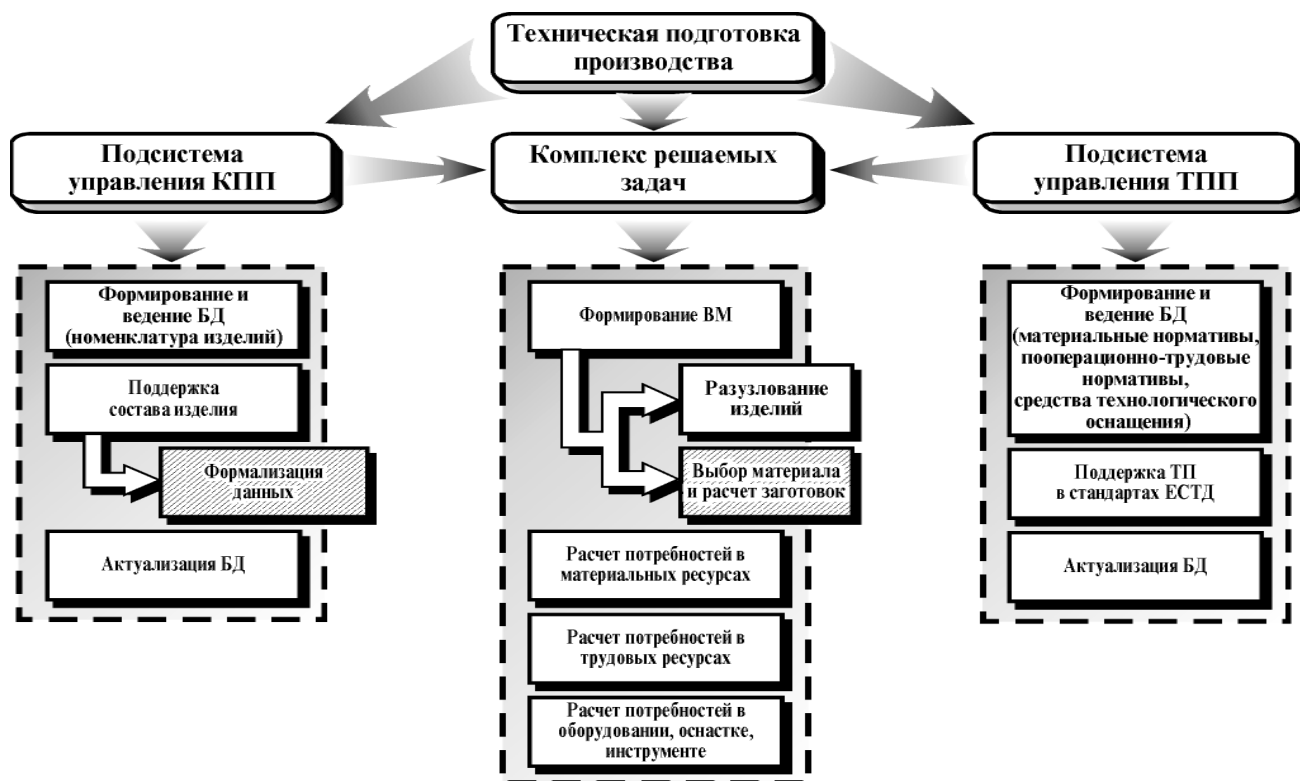


Рисунок 1 - Схема выполнения работ на этапе ТПП

Методологическая платформа модели заключается в представлении информации НД в виде набора параметров, присутствующих в условном обозначении стандартных единиц производства, с закреплением за каждым параметром своего индивидуального места в этой последовательности.

Предлагается такую последовательность определить как формулу обозначения и представить в виде:

$$\Phi_{об_j} = [\underset{i=1}{\overset{n}{\Phi}} B_{iz}]_j,$$

где j - порядковый номер НД; Φ - последовательность; B_i - параметр; i - порядковый номер параметра; z - шаг нумерации параметров; n - количество параметров.

Такой функциональной зависимостью описывается каждый регламентируемый объект НД. Под объектом НД понимается тип стандартной единицы, имеющей структурированное условное обозначение. Например, ГОСТ 24705 - "Резьба метрическая" регламентирует гайки, винты, болты и другой крепеж. Таким образом, данная учетная единица (ГОСТ 24705) в системе будет описана несколькими формулами обозначения соответственно.

Каждый элемент формулы обозначения (параметр) обладает набором свойств: наименованием, порядковым номером, диапазоном допустимых значений, а так же нормативными ссылками, соответствующими структуре стандарта. Графическое изображение параметра представлено на рисунке 2.

Графически каждая грань описывает группу свойств.

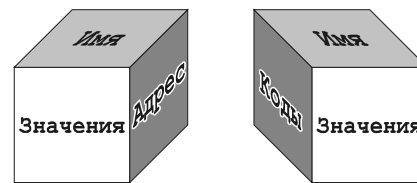


Рисунок 2 - Графическое изображение параметра

Верхняя грань - имя параметра (номер, наименование, тип). Передняя грань - группа значений параметра. Правая грань - адрес параметра (нормативные ссылки). Левая грань - группа кодов значений параметра.

Связь между НД, как объектами модели, осуществляется посредством нормативных ссылок. Организация ссылок с параметра одного НД на параметр этого же или другого НД можно считать самым распространенным способом организации ссылок. В этом случае в качестве значений исследуемого параметра используются значения ссылочного. Менее распространен способ организации ссылки с параметра одного НД на несколько параметров разных НД. Параметры ссылочных ГОСТов равноценны, и использование какого-либо из них для замены исследуемого определяется ограничениями, накладываемыми на допустимые сочетания значений исследуемого параметра и других параметров формулы обозначения. В остальных случаях значение исследуемого параметра является составным из значений ссылочных параметров.

Пример организации ссылок представлен на рисунке 3.

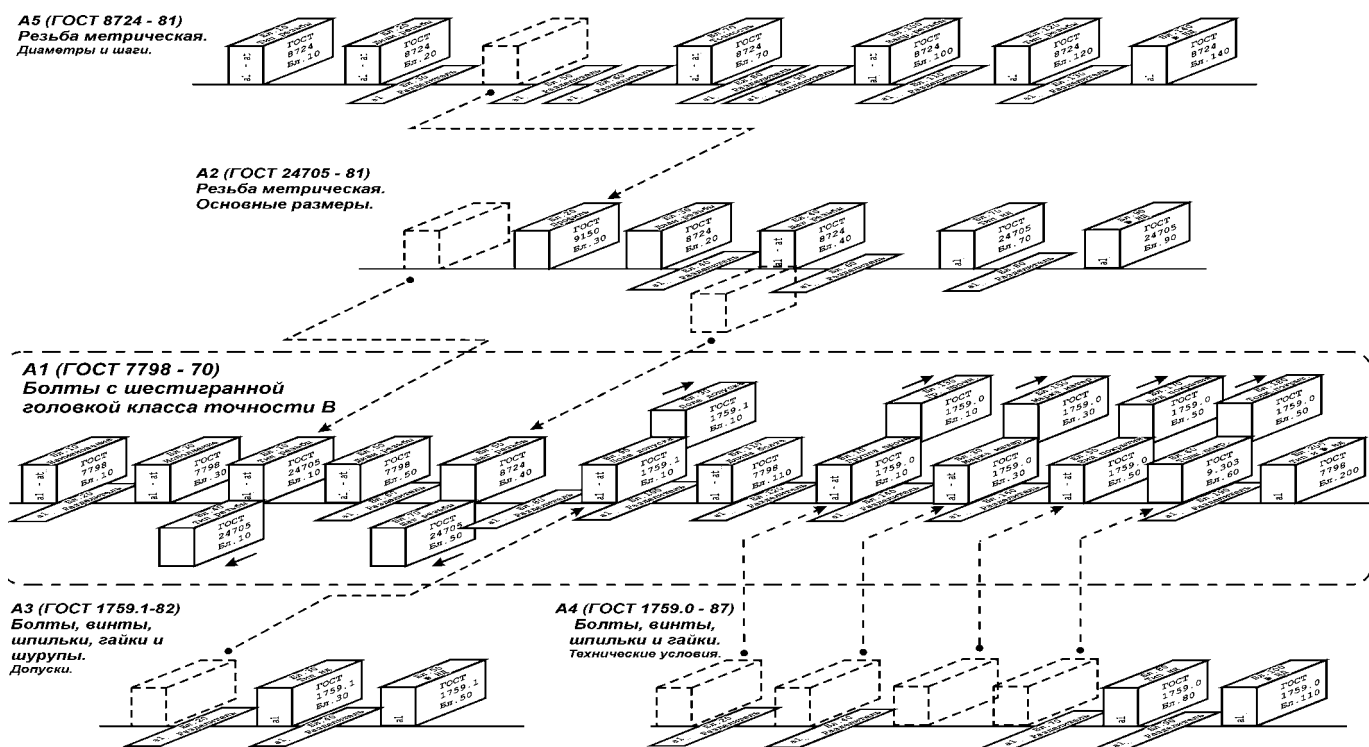


Рисунок 3 - Графическое изображение ссылок ГОСТа 7798-70

Таким образом, в электронном архиве системы стандарт будет храниться в виртуальном виде, то есть представлять собой последовательность параметров, присутствующих в условном обозначении стандартной единицы производства, регламентированной данным стандартом. Набор параметров строго соответствует условному обозначению данной стандартной единицы, но значения параметров могут храниться в ссылочных НД, если таковые имеются.

Результатом построения системы является решение проблемы формализации данных по стандартным единицам производства, а именно, автоматизированное кодирование.

Располагая информацией о структуре (наборе параметров) условного обозначения стандартной единицы и о порядке следования этих параметров в коде, ПЭВМ запись распознается и упорядочивается в соответствии с формулой кодирования.

$$\Phi_{\text{ок}} = \prod_{i=1}^n [X(N)]_{iz},$$

где Φ - последовательность параметров; X - десятичный цифровой разряд кодового обозначения; N - число разрядов кодового обозначения (длина кода); i - порядковый номер параметра в $\Phi_{\text{об}}$; z - шаг нумерации параметров; n - количество параметров в $\Phi_{\text{об}}$; $[X(N)]_{10}, [X(N)]_{20}, [X(N)]_{30}, \dots, [X(N)]_i$ - однозначные коды значений параметров в $\Phi_{\text{ок}}$.

УДК 681.32

СИНТЕЗ ОПТИМАЛЬНОГО НЕЧЕТКОГО РЕГУЛЯТОРА ДЛЯ НЕЛИНЕЙНОЙ ДИНАМИЧЕСКОЙ СИСТЕМЫ

М.Б.Коломейцева, Д.Л.Хо

В работе рассматривается построение схемы нечеткого адаптивного управления для нелинейной динамической системы. Адаптивный нечеткий регулятор строится на основе анализа знаний об объекте управления с помощью людей-операторов в виде If-Then правил. Параметры функции принадлежности, характеризующие лингвистические переменные для входных и выходных сигналов регулятора, определяются по условию обеспечения устойчивости системы управления.

The robust fuzzy adaptive control scheme is developed for a class of unknown nonlinear SIMO systems. An adaptive fuzzy controller is synthesized from a collection of fuzzy If-Then rules. The parameters of the membership functions characterizing the linguistic terms in the If-Then rules change according to some adaptive law for the purpose of controlling a plant to track a reference trajectory. These fuzzy If-Then rules are collected from experienced operators.

Современный этап развития индустриального общества характеризуется внедрением во все отрасли промышленности и техники средств автоматизации, начиная от простых

полный цифровой код обозначения стандартной единицы производства автоматически составляется из однозначных кодов значений параметров в соответствии с установленным порядком в $\Phi_{\text{ок}}$.

Параллельно с процессом кодирования ПЭВМ осуществляется контроль над корректностью написания обозначения.

Автоматизация процесса кодирования обеспечит однозначную идентификацию стандартных единиц производства, в том числе стандартных деталей и, в конечном итоге, позволит эффективно решать задачу унификации, то есть объединения однотипных деталей в партии.

Унификация стандартных деталей в условиях индивидуального производства позволит оптимизировать выполнение работ на многих этапах жизненного цикла изделия, в том числе на этапах проектирования конструкторской и технологической документации, расчета заготовок, нормирования операций техпроцесса, проектирования средств технологического оснащения, а также механической обработки.

Таким образом, предлагаемый подход к организации массива обеспечит эффективное решение формализации данных на этапе ТПП и позволит, в конечном итоге, сократить сроки создания и стоимость выпускаемой продукции.

локальных устройств до сложных управляющих комплексов. Одной из центральных проблем при создании такого рода производств являются проектирование и реализация на современной технической базе эффективных систем автоматического управления, обеспечивающих высокое качество функционирования управляемых объектов и технологических процессов. На первом этапе проектирования указанных систем основной задачей является синтез регуляторов, которые гарантируют асимптотическую устойчивость замкнутых систем и удовлетворяют определенной совокупности инженерных требований к качественным свойствам систем в переходных процессах и установившихся режимах. Однако синтез детерминированных регуляторов при отсутствии адекватного математического описания объекта управления, что часто имеет место на практике, представляет сложную задачу. В данной работе предложен подход построения управления нелинейным динамическим

процессом на основе нечеткого регулятора.

Рассмотрим класс динамических систем, имеющих описание в форме

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = x_3, \\ \dots \\ \dot{x}_n = f(\underline{x}) + u, \end{cases} \quad (1)$$

где $\underline{x} = (x_1, x_2, \dots, x_n)^T = (x, \dot{x}, \dots, x^{(n-1)}) \in R^n$ - вектор состояния; u - управляющий сигнал; f - нелинейная или линейная функция. Цель синтеза состоит в определении такого управления $u(\underline{x})$, которое сначала переводит объект из произвольного начального состояния \underline{x}_0 в окрестность задаваемого многообразия $\Psi(\underline{x}) = 0$, а затем обеспечивает дальнейшее его движение вдоль $\Psi(\underline{x}) = 0$ к началу координат пространства состояний $\underline{x}(t_k) = 0$. При этом обеспечивается минимизация функционала вида

$$J = \int_0^{\infty} [m^2 \varphi^2(\Psi) + c^2 \dot{\Psi}^2(t)] dt, \quad (2)$$

где m, c - постоянные коэффициенты; $\Psi(\underline{x})$ - агрегированная макропеременная, представляющая собой некоторую произвольную дифференцируемую или кусочно-непрерывную функцию и $\Psi(0) = 0$; $\varphi(\Psi)$ - некоторая функция, удовлетворяющаяся следующим условиям:

- а) однозначности, непрерывности и дифференцируемости при всех значениях Ψ ;
- б) $\varphi(0) = 0$;
- в) $\varphi(\Psi) \cdot \Psi < 0$ при любых $\Psi \neq 0$.

Это условие может быть реализовано, если управляющий сигнал определить согласно выражению [1]

$$\hat{u} = - \left[\frac{\partial \Psi(\underline{x})}{\partial x_n} \right]^{-1} \left[\frac{1}{T} \varphi(\Psi) + \sum_{k=1}^n \frac{\partial \Psi(\underline{x})}{\partial x_k} f_k(\underline{x}) \right], \quad (3)$$

где $T = \frac{c}{m} > 0$; $f_k = x_{k+1}$, $k = \overline{1, n-1}$; $f_n = f$.

Однако при отсутствии достаточной информации о функциях $f_i(\underline{x})$ реализовать управляющий сигнал в виде (3) невозможно. В этом случае возможным вариантом решения может быть применение нечеткого регулятора.

Нечеткий регулятор состоит из фазификатора, механизма нечеткого вывода и дефазификатора. Правила управления R^j записываются в виде

$$\begin{aligned} R^j: & \text{ If } (x_1 \text{ is } A_1^j) \text{ and } (x_2 \text{ is } A_2^j) \text{ and } \dots \text{ and } (x_n \text{ is } A_n^j) \\ & \text{ Then } (u \text{ is } B^j), \end{aligned} \quad (4)$$

где $x_i, i = \overline{1, n}$ - входные переменные регулятора; u -

управляющий сигнал (выход регулятора); A_i^j, B^j - лингвистические переменные, характеризующиеся функциями принадлежности $\mu_{A_i^j}(x_i), \mu_{B^j}(u)$; $j = \overline{1, L}$ - количество правил. При использовании метода центральной площади для решения задачи дефазификации выход u имеет вид [2]

$$u = \frac{\sum_{j=1}^L \bar{\lambda}^j \left(\prod_{i=1}^n \mu_{A_i^j}(x_i) \right)}{\sum_{j=1}^L \left(\prod_{i=1}^n \mu_{A_i^j}(x_i) \right)}, \quad (5)$$

где $\bar{\lambda}^j$ - численное значение управляющего сигнала, при котором $\mu_{B^j}(\bar{\lambda}^j) = 1$. При решении задачи синтеза они подлежат определению. Вводим обозначение нечеткой векторной функции $\underline{\zeta}(\underline{x}) = (\zeta^1(\underline{x}), \zeta^2(\underline{x}), \dots, \zeta^L(\underline{x}))^T$, где $\zeta^j(\underline{x})$ имеет вид

$$\zeta^j(\underline{x}) = \frac{\prod_{i=1}^n \mu_{A_i^j}(x_i)}{\sum_{j=1}^L \left(\prod_{i=1}^n \mu_{A_i^j}(x_i) \right)}. \quad (6)$$

Тогда выражение (5) с учетом (6) примет вид

$$u = u(\underline{x}, \underline{\lambda}) = \underline{\lambda}^T \underline{\zeta}(\underline{x}), \quad (7)$$

где $\underline{\lambda} = (\bar{\lambda}^1, \bar{\lambda}^2, \dots, \bar{\lambda}^L)^T$.

Выражение (3) с учетом выбора $\Psi(\underline{x}) = \sum_{i=1}^n \alpha_i x_i$ и

$\varphi(\Psi) = \Psi$, где $\alpha_i, i = \overline{1, n}$ - постоянные коэффициенты, примет вид

$$\hat{u} = - \frac{1}{\alpha_n} \left[\frac{1}{T} \Psi + \sum_{k=1}^n \alpha_k f_k(\underline{x}) \right]. \quad (8)$$

Перепишем (1) в виде:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = x_3, \\ \dots \\ \dot{x}_n = f(\underline{x}) + u + \hat{u} - \hat{u}. \end{cases} \quad (9)$$

Управляющий сигнал \hat{u} может быть аппроксимирован выражением (7) с коэффициентами $\underline{\lambda}^*$. Система дифференциальных уравнений (9) с учетом (7) и (8) после некото-

рого преобразования примет вид

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = x_3, \\ \dots \\ \dot{x}_n = -\frac{1}{\alpha_n T} \Psi - \frac{1}{\alpha_n} \sum_{k=1}^{n-1} \alpha_k f_k + \theta^T \zeta(\underline{x}), \end{cases} \quad (10)$$

где $\theta = \lambda - \lambda^*$. Построим функцию Ляпунова для системы (10) в виде

$$V = \frac{1}{2} \Psi^2 + \frac{1}{2\gamma} \theta^T \theta, \quad (11)$$

где γ - положительная постоянная. Полная производная по времени от функции Ляпунова в виде (11) в силу (10)

$$\begin{aligned} \dot{V} &= \Psi \dot{\Psi} + \frac{1}{\gamma} \theta^T \dot{\theta} = \Psi \left(\sum_{i=1}^n \alpha_i \dot{x}_i \right) + \frac{1}{\gamma} \theta^T \dot{\theta} = \\ &= \Psi \left(\sum_{i=1}^{n-1} \alpha_i \dot{x}_i + \alpha_n \dot{x}_n \right) + \frac{1}{\gamma} \theta^T \dot{\theta} = \\ &= \Psi \left(-\sum_{k=1}^{n-1} \alpha_k f_k + \alpha_n \left(-\frac{1}{\alpha_n T} \Psi - \sum_{k=1}^{n-1} \alpha_k f_k + \theta^T \zeta(\underline{x}) \right) \right) + \frac{1}{\gamma} \theta^T \dot{\theta} = \end{aligned}$$

Отметим, что $\dot{\theta} = \underline{\lambda}$. Тогда \dot{V} примет вид

$$\dot{V} = -\frac{1}{T} \Psi^2 + \theta^T \left[\alpha_n \Psi \zeta(\underline{x}) + \frac{1}{\gamma} \underline{\lambda} \right]. \quad (12)$$

Одним из возможных выборов $\underline{\lambda}$ для обеспечения устойчивости замкнутой системы (т.е. $\dot{V} < 0$) является $\underline{\lambda} = -\gamma \alpha_n \Psi \zeta(\underline{x})$.

УДК 621.313.333

ОБЗОР СОВРЕМЕННЫХ МЕТОДОВ УПРАВЛЕНИЯ СИНХРОННЫМИ ДВИГАТЕЛЯМИ С ПОСТОЯННЫМИ МАГНИТАМИ

Д.В.Корельский, Е.М.Потапенко, Е.В.Васильева

Виконано обзор літератури, касаючоїся питань керування синхронними двигунами з постійними магнітами. Розглянуто 38 джерел, котрі відносяться до періоду 1993-2000 р., та віддзеркалюють стан та основні тенденції вказаної області.

Выполнен обзор литературы, касающейся вопросов управления синхронным двигателем с постоянными магнитами. Рассмотрено 38 источников, которые относятся к периоду 1993-2000 г., и отражают состояние и основные тенденции в указанной области.

В качестве численного примера проводим синтез нечеткого регулятора для нелинейной динамической системы:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = -0,1x_2 + \text{sign}(x_1) + u. \end{cases} \quad (13)$$

Результаты моделирования с выбранным многообразием $\Psi(\underline{x}) = x_1 + x_2$ для четырех разных начальных условий $\underline{x}(0) = [0,5 \ 0,5]^T$, $[1 \ 0,5]^T$, $[-0,5 \ 0,5]^T$, $[-1 \ 0,5]^T$ представлены на рисунке 1. Как показывают исследования, выбор величины γ существенно влияет на форму переходного процесса. На рисунке 2 показаны кривые переходного процесса для $\gamma = 10$, $\gamma = 20$, $\gamma = 100$ при $\underline{x}(0) = [1 \ 1]^T$.

В заключении можно делать следующие выводы:

1. Предложенный метод построения управления нелинейным динамическим процессом на основе адаптивного нечеткого регулятора обеспечивает асимптотическую устойчивость замкнутой системы управления.
2. Использование идеи аппроксимировать управляющий сигнал нечеткой системой с последующей адаптацией позволяет решать задачу синтеза регулятора при недостаточной информации о математическом описании управляемого процесса (объекта управления).
3. Качество переходного процесса может быть задано выбором многообразия $\Psi(\underline{x})$ и коэффициентов γ , α_i .
4. Полученный адаптивный нечеткий регулятор прост в реализации, что может существенно повысить быстродействие системы управления.

ПЕРЕЧЕНЬ ССЫЛОК

1. Колесников А. А. Синергетическая теория управления. - М.: Энергоатомиздат, 1994.
2. Ching-Teng Lin, C. S. George Lee. Neural fuzzy systems.// Prentice-Hall International, Inc., 1996//.

галузі.

The survey of the literature, concerned to the questions of permanent magnet synchronous motor control is carried out. There were included 38 sources, since 1993 till 2000, which reflected the state and main trends in this field.

Синхронный двигатель с постоянными магнитами (СДПМ), благодаря своим высоким эксплуатационным характеристикам, является наиболее перспективной машиной

в диапазоне малых и средних мощностей. СДПМ простой по конструкции, не имеет потерь на возбуждение и обладает высокой стабильностью скорости ротора. Эти качества выделяют его из ряда всех остальных машин и обеспечивают ему применение в системах автоматики, приводах подачи станков, прецизионных системах слежения, а также системах, где стабильность скорости является первостепенным требованием, предъявляемым к технологическому процессу.

Постоянное удешевление магнитных материалов, в частности, внедрение сплавов редкоземельных металлов, совершенствование аппаратной базы управления делают возможным использование этого типа двигателей в тех областях, где традиционно применялись двигатели постоянного тока (ДПТ) или асинхронные двигатели (АД). Использование СДПМ как альтернативы ДПТ в регулируемом приводе стало возможным с появлением соответствующей преобразовательной и цифровой вычислительной техники. Прежде всего, это связано с внедрением преобразователей на транзисторах с изолированным затвором (Insulated Gate Bipolar Transistor (IGBT)), использованием принципов широтно-импульсной модуляции (ШИМ, Pulse Width Modulation (PWM)), преобразования переменного тока в переменный (АС-АС) [1-2, 23].

Направление развития принципов управления обуславливается развитием аппаратной базы привода: параметрами самого двигателя и его качествами, наличием датчиков угла поворота и (или) скорости, типом преобразователя, вычислительной мощностью контроллера. Реализация управления СДПМ с учетом этих требований позволяет судить об эффективности используемых алгоритмов.

В литературе часто встречается понятие вентильного двигателя (ВД) или бесколлекторного двигателя постоянного тока с возбуждением от постоянных магнитов (Brushless Direct Current Motor with Permanent Magnets). Эта машина по своей конструкции ничем не отличается от классического СДПМ, изменен лишь принцип питания. СДПМ питается от источника переменного тока или напряжения, как правило, формируемого при помощи ШИМ, а ВД - трапециидальным напряжением, питающим соответствующие фазы двигателя, и переключаемым по мере поворота ротора.

На протяжении последних 20 лет представление уравнений СДПМ в роторной системе координат стало основным способом описания его работы. Уравнения машины во вращающихся координатах обеспечивают большую наглядность протекающих в обмотках статора процессах. Действительные токи и напряжения статора в приведенной двухфазной неподвижной системе координат связаны с роторными величинами однозначным преобразованием. Эти преобразования основаны на предположении о симметричности электрических и магнитных цепей всех обмоток. Кроме роторной системы координат иногда применяется статорная, при этом значение индуктивности обмоток статора связано тригонометрическими зависимостями с углом

поворота ротора, что не так удобно.

Для реализации управления СДПМ по стандартным схемам (с использованием ПИ-регулятора) необходим тахогенератор и (или) датчик положения ротора. Вычисляя производную от положения, можно определить скорость. Эта величина используется в уравнениях двигателя во вращающейся системе координат, а также в уравнениях перехода из статорной системы в роторную систему. При питании вентильного двигателя угол поворота необходим для определения времени коммутации транзисторов или тиристоров. В современной литературе в ряде работ [3-6, 11, 12, 22, 28, 29] отчетливо просматривается тенденция к отказу от использования датчиков механических величин (Sensorless Control) и восстановления вектора состояния системы с помощью различных оценивающих моделей. Это обусловлено тем, что эти датчики снижают надежность всей системы, увеличивают стоимость привода, ухудшают массогабаритные показатели. Кроме этого, при определении скорости путем дифференцирования угла поворота помехи могут существенно повлиять на процесс управления. Несмотря на то, что современные коммерчески распространяемые приводы выпускаются со встроенными датчиками положения ротора, перспектива управления без датчиков механических величин остается очень заманчивой. В работе [3] предлагается интересный алгоритм в дискретной форме, который производит оценку положения, скорости и константы ЭДС двигателя (потока постоянного магнита). Контроллер вычисляет ток, положение и скорость модели, а затем использует погрешность действительного тока и тока модели для оценки действительной скорости. Для этого производится ряд упрощений в уравнениях модели: для небольших углов ошибки синус берется равным своему аргументу, а косинус - единице. Это дает возможность рассматривать уравнения контуров как независимые и путем подстановки в них измеряемых значений токов идентифицировать параметры системы. По сути, строится наблюдатель константы ЭДС.

Оценка угла поворота с помощью расширенного фильтра Калмана (РФК, Extended Kalman Filter (ЕКФ)) применяется в работах [4, 5, 6, 20]. РФК позволяет посредством оптимальной фильтрации сигналов выполнить оценку параметров стохастической системы и ее переменных. Однако расчет алгоритма фильтра Калмана связан с большим объемом вычислений, что вызывает определенные трудности с реализацией его в режиме реального времени. При реализации этого алгоритма за счет большого времени расчета растет фазовая задержка управления.

Адаптивные системы являются другим направлением развития алгоритмов управления, при этом реализованы как алгоритмы с датчиками, так и без них [6,8-10, 34]. Основным способом адаптации является использование адаптивной системы с настраиваемой моделью (Model Reference Adaptive System (MRAS)). Способ реализуется путем изменения параметров модели. Недостатком этих систем также является большая вычислительная нагрузка на

контроллер.

Применение наблюдателей состояния (State Observers) [7, 11, 12] является естественным способом восстановления значения вектора состояния с использованием матричного представления системы. Так, в [7] выполняется разложение вектора объекта на наблюдаемую и измеряемую части, и соответствующее ему разложение матрицы состояния объекта с целью построения наблюдателя О'Рейли.

В работе [11] наблюдатель используется для предсказания значения токов на 1 цикл вперед, что решает проблему устранения запаздывания управления. Однако предлагаемая система построена без учета магнитного насыщения, потерь в меди и стали, анизотропности ротора двигателя (индуктивности по осям d и q приняты равными), что снижает оптимальность предлагаемого метода.

Другим способом управления является применение систем со скользящими режимами (Sliding Mode Systems) [13-15]. Особенность этих систем, принадлежащих к системам с переменной структурой, состоит в том, что знак управления меняется при пересечении так называемой поверхности скольжения, представляющей собой специальным образом составленное уравнение, в которое в качестве переменных входят переменные состояния объекта. Недостатками такого метода являются возможная потеря устойчивости на участке достижения поверхности переключения; высокочастотные переключения, которые ведут к быстрому износу механических и электрических частей привода; высокие требования к преобразователю.

Прямое управление моментом (Direct Torque Control (DTC)) [16, 17, 18, 19, 21, 33] может быть осуществлено при питании синхронного двигателя от инвертора тока. Такие системы обладают рядом преимуществ: достигается робастность по отношению к разбросу параметров, упрощается алгоритм управления за счет отсутствия токового контура регулирования, обеспечивается высокое быстродействие системы. Однако метод не избавлен от серьезного недостатка: при малых углах нагрузки возникают пульсации момента и колебания скорости ротора.

Нейронные сети (Neuron Network), использованные для построения алгоритма управления СДПМ от инвертора переменный ток - переменный ток (АС-АС) [23] представляют собой самообучающиеся сети, включающие в себя три слоя нейронов - входной, скрытый и выходной. Обучение влияет на значение весовых коэффициентов в передаточных функциях нейронов. Алгоритмы расчета не относятся к простым алгоритмам по вычислительной нагрузке.

Качественно новым методом управления СДПМ является использование наблюдателя неопределенностей (Uncertainties Observer). Это направление в данной области получило развитие в самое последнее время [24, 25, 26]. Суть систем, построенных по такому принципу, состоит в сведении неопределенностей параметров и внешних возмущений в единый вектор неопределенностей, его оценка и компенсация с помощью наблюдателя, а также задание степени робастности путем его настройки. Привлекатель-

ность метода состоит в том, что в качестве неопределенностей могут рассматриваться нелинейности системы, обусловленные насыщением магнитной цепи, которые прямо не отражены в уравнениях машины, но оказывают влияние на динамические процессы в приводе. Кроме того, предлагаемые алгоритмы отличаются относительной вычислительной простотой.

Интересная методика предлагается в [32]; процессы в двигателе разбиваются на две группы: быстрые и медленные. К медленным процессам относятся электромеханические, к быстрым - электромагнитные. Разделение системы на две подсистемы и решение их с помощью итерационных алгоритмов позволяет построить бездатчиковый алгоритм, более выгодный с точки зрения вычислительной нагрузки, чем алгоритмы с фильтром Калмана.

Кроме описанных подходов, построенных для качественной отработки программного задания, существует ряд методов, построенных на основе оптимизации энергопотребления с учетом параметров двигателя.

В большинстве работ, авторы преследуют стратегию поддержания тока $I_d = 0$ с целью избавления от "вредного" возмущающего слагаемого в уравнении момента двигателя. Однако такой режим работы привода не является оптимальным с точки зрения потерь в меди, т.к. развиваемый двигателем момент можно достичь при меньших реальных токах. По сути дела, СДПМ с радиальным расположением магнитов $L_d \neq L_q$ является комбинацией "чистого" СДПМ и синхронного реактивного двигателя. Ясно, что оптимум работы такого двигателя не может быть реализован стратегией $I_d = 0$. Исследования поведения СДПМ при ненулевых размагничивающих токах I_d проведено в работах [35, 36, 37]. Полученные результаты свидетельствуют о возможности значительного снижения энергопотребления СДПМ (на 30% в [36]).

Подход, целью которого является выявление неисправностей ВДПМ во время работы, предложен в [38]. Этот метод основан на определении значений сигнальных функций. При неисправности двигателя с помощью разности показаний оценителей (уравнений модели системы, где по нескольким данным параметрам вычисляются остальные) и датчиков строится таблица идентификации поломок.

Выполненный обзор литературы позволяет выделить следующие основные приоритеты развития алгоритмов управления.

1. Точность отработки программного задания (скорости или положения ротора). В отношении СДПМ это требование актуально в смысле возможных подсинхронных колебаний скорости ротора. Кроме этого, существуют ограничения на динамические режимы машины - с использованием существующих робастных методов можно сформировать практически любой переходный процесс, но скачки токов могут не пройти безвредно для магнитной системы ротора. Определение допустимой границы динамических режимов

является большим полем для развития методов управления.

2. Нечувствительность к неопределенностям параметров СДПМ и помехам. При наличии больших коэффициентов усиления регуляторов, обеспечивающих робастность системы, последняя становится чувствительной к высокочастотным помехам.

3. Минимизация электропотребления. Разработка алгоритмов, оптимизирующих потребляемую приводом мощность как в статических, так и в динамических режимах, может быть осуществлена за счет размагничивающих токов продольной оси.

4. Расширение диапазона регулирования в сторону малых скоростей. Проблема связана с влиянием пульсаций момента на малых скоростях. При больших скоростях пульсации момента не оказывают существенного влияния, т.к. они фильтруются за счет инерционности ротора [34].

5. Уменьшение количества датчиков. Очевидно, что датчики механических величин значительно ухудшают эксплуатационные и стоимостные параметры привода. Устранение датчиков повышает надежность электропривода. Использование встроенных оптических датчиков положения в современных приводах и получение скорости путем дифференцирования положения является приемлемым путем решения данной проблемы. Однако, следует отметить, что бездатчиковые алгоритмы более чувствительны к разбросу параметров.

6. Простота алгоритмов в аспекте вычислительной нагрузки. Несмотря на стремительный рост мощности процессоров, более простые алгоритмы обеспечивают меньшее время расчета, а значит, и фазовую задержку управления. Кроме того, более простые алгоритмы позволяют использовать более дешевые микропроцессоры.

7. Исследование работоспособности СДПМ с многомасовыми объектами управления. СДПМ часто применяется как серводвигатель в роботизированных системах, динамика которых очень сложна. Такие объекты относятся к классу неопределенных. В то же время, требования к точности управления роботами очень высоки, что ставит соответствующую задачу.

Обилие публикаций, разнообразие применяемых методов, их новизна и качественное отличие друг от друга свидетельствуют об интенсивности исследований в данной области, подчеркивают перспективность данного класса электрических машин, совместивших в себе простоту двигателя переменного тока, а по возможностям управления превзошедших ДПТ. Из наблюдаемой динамики роста популярности СДПМ можно сделать вывод о необходимости разработки алгоритмов управления, связанных с использованием прямого управления моментом, наблюдателей состояния и неопределенностей, адаптивных алгоритмов и нейронных сетей, алгоритмов управления, обеспечивающих оптимизацию потребляемой мощности и показателей качества.

ПЕРЕЧЕНЬ ССЫЛОК

1. D.H. Kim, J.H. Kang, S. Kim. Full Digital Controller of Permanent Magnet AC Servo Motor for Industrial Robot and CNC Machine Tool. // IECON-94 -1994 - V. 3 - P. 61-67.
2. Н. Обухов, В. Горбунов, П. Чуев, А. Анучин. Высокопроизводительные встраиваемые системы управления двигателями на базе сигнального микроконтроллера TMS320F241. // Chip news. - 2000. - Май. - С. 28-32.
3. T. Takeshita, N. Matsui. Sensorless Brushless DC Motor Drive with EMF Constant Identifier. // IECON-94.-1994.-V.1-P.8-13.
4. S. Bolognani, R. Oboe, M. Zigliotto. DSP-based Extended Kalman Filter Estimation of Speed and Rotor Position of a PM Synchronous Motor. // IECON-94. - 1994.-Vol.3, №3.-P.85-90.
5. Yoon-Ho Kim, Yoon-Sang Kook. High Performance IPMSM Drives without Rotational Position Sensors Using Reduced-Order EKF. // IEEE Trans. on Energy Conversion.-1999. - Vol.14, №4. - P. 868 - 873.
6. L. Salvatore, S. Stasi. Adaptive Position Control of PMSM Drive. // IECON-94. - 1994. - V. 3. - P. 78-84.
7. K.W. Lim, K.S. Low, M.F. Rahnan. A Position Observer for Permanent Magnet Synchronous Motor Drive. // IECON-94. - 1994. - V. 3. - P. 49-61.
8. A. Kaddouri, O.Achrif, H. Le-Huy. Adaptive Nonlinear Control for Speed Regulation of a Permanent Magnet Synchronous Motor. // IECON-99. - SP-4.
9. Jong Sun Ko, Sung Koo Youn, Bimal K. Bose. A Study on Adaptive Load Torque Observer for Robust Precision Position Control of BLDC Motor.// IECON-99. - PE-16.
10. J-J Chen, K-P Chin. Reduced Order Control of Permanent Magnet Synchronous Motors.// IECON-99. - SP-7.
11. J.F. Moynihan, M.G. Egan, J.M.D. Murphy. The Application of State Observers in Current Regulated PM Synchronous Motor Drives // IECON-94. - 1994. - V. 1. - P. 14-20.
12. J. Solsona, M. I. Valla, C. Muravchik. A Nonlinear Reduced Order Observer for Permanent Magnet Synchronous Motors. // IECON-94. -1994 - V. 1. - P. 32-37.
13. A. Glumineau, M. Hami, C.Lanier, C. H. Moog. Robust Control of a Brushless Servo Motor via Sliding Mode Techniques. // Int. J. Control. - 1993. - Mol.58. - №5. - P. 979-990.
14. O.S. Bogosyan, M. Gokasan, E.M. Jafarov. A Sliding Mode Position Controller for a Nonlinear Time-Varying Motion Control System. // IECON-99. - MT-4.
15. Y. S. Lu, J. S. Chen. Design of a Global Sliding Mode Controller for a Motor Drive with Bounded Control. // Int. J. Control. - 1995. - Vol.62, №5. - P. 1001-1019.
16. L. Zhong, M. Rahman, W. Hu, K. Lim. A Direct Torque Controller for Permanent Magnet Synchronous Motor Drives. // IEEE Trans. on Energy Conversion. - 1999. - Vol. 14, №3. - P. 637 - 643.
17. M.F. Rahman and L. Zhong. Voltage Switching Tables for DTC Controlled Interior Permanent Magnet Motor. // IECON-99. - PE-20.
18. M.F. Rahman and L. Zhong. Comparison of Torque Responses of the Interior Permanent Magnet Motor under PWM Current and Direct Torque Controls. // IECON-99. - PE-20.
19. T.Aihara, A. Toba, T. Yanase, A Mashimo, K. Endo. Sensorless Torque Control of Salient Pole Synchronous Motor at Zero-Speed Operation.// IEEE Trans. on Power Electronics. Vol. 14, №1. - 1999.
20. K. Rajashekara, A. Kawamura. Sensorless Control of Permanent Magnet AC Motors. // IECON-94.- 1994.-V.3.-P.106-111.
21. D. Greiner, R. Mende, J.P. Louis. Comparison of Several Control Strategies for D.C. Brushless Drives. // IECON-94. - 1994. - V. 3. - P. 20-25.
22. J. Hu, B. Wu. New Integration Algorithms for Estimating Motor Flux over a Wide Speed Range. // IEEE Trans. on Power Electronics. - 1998. - Vol. 13, №5. - P. 969 - 978.
23. Der-Fa Chen, Tian-Hua Liu. Design and Implementation for a Novel Matrix PMSM Drive System. // IECON-99. - PE-16.
24. Kyeong-Hwa Kim, In-Cheol Baik, Gun-Woo Moon, Myung-Joong Youn. A Current Control for a Permanent Magnet Synchronous Motor with a Simple Disturbance Estimation Scheme. // IEEE Trans. on Control System technology. - 1997. - Vol. 7, №5.-P.630 - 634.
25. Faa-Jeng Lin, Yueh-Shan Lin. A Robust PM Synchronous Motor Drive with Adaptive Uncertainty Observer. // IEEE Trans. on Energy Conversion. - 1999. - Vol. 14, №4. - P. 959 - 995.
26. Потепенко Е.М., Корельский Д.В., Васильева Е.В. Робастное управление электроприводом с вентильным двигателем. // Радиоэлектроника, информатика, управління. - 2000. - №1. - С. 161-166.
27. U. Shnaible, B. Szabados. Dynamic Motor Parameter Identification for High Speed Flux Weakening Operation of Brushless Permanent Magnet Synchronous Motor. // IEEE Trans. on Energy Conversion. - 1999. - Vol. 14, №3.- P. 486 - 493.
28. R. Marino, S Peresada, P. Tomei. Nonlinear Adaptive Control of Permanent Magnet Synchronous Motor. // Automatica. - 1995. - Vol. 31, №11. - P. 1595-1604.

29. D. Yousefi, M. Azizi, A. Saad. Position and Speed Estimation with Improved Integrator for Synchronous Motor. // IECON-99. - PE-16.
30. J.C. Moreira. Indirect Sensing for Rotor Flux Position of Permanent Magnet AC Motors Operating Over a Wide Speed Range. // IEEE Trans. on Industry Applications. - 1996. -Vol. 32, №6.- P.1394 - 1402.
31. S. Cstlund. M. Brokemper. Sensorless Rotor-Position Detection from Zero to Rated Speed for an Integrated PM Synchronous Motor Drive. // IEEE Trans. on Industry Applications. - 1996. - Vol. 32, №5. - P. 1158 -1164.
32. J.K. Seok, J.S. Kim, S.K. Sul. Overmodulation Strategy for High-Performance Torque Control. // IEEE Trans. on Power Electronics. - 1998. - Vol. 13, №4. - P. 1 - 7.
33. T. Song, M.F. Rahman, K. W. Lim, M.A. Rahman. A Singular Perturbation Approach to Sensorless Control of a Permanent Magnet Synchronous Motor Drive. // IEEE Trans. on Energy Conversion. - 1999. - Vol. 14, №4. - P. 1359 - 1365.
34. B.H. Lam, S.K. Panda, J.-X. Xu, K.W. Lim. Torque Ripple Minimization in PM Synchronous Motor Using Iterative Learning Control. // IECON-99. - PE-20.
35. P.S. Fredericson, J. Birk, F. Blaabjerg. Comparison of Two Energy Optimizing Techniques for PM- Machines. // IECON-94. - 1994 - V. 1. - P. 26-31.
36. S.Vaez, M.A. Rahman. An On-line Minimization Controller for Interior Permanent Magnet Motor Drives. // IEEE Trans. on Energy Conversion. - 1999. - Vol. 14, №4.- P. 1435 - 1440.
37. R.Monajemy, R. Krishnan. Control and Dynamics of Constant Power Loss Based Operation of Permanent Magnet Synchronous Motor Drive System. // IECON-99. - PE-20.
38. O.Moseler, T. Heller, R. Isermann. Model - Based Fault Detection for an Actuator Driven by a Brushless DC Motor. \ 14th World Congress of IFAC. - 1999. - P-7e-08-3.

УДК 681.51.0.15.26

КРИТЕРИИ СИНТЕЗА НОМИНАЛЬНЫХ МОДЕЛЕЙ В ЗАДАЧАХ НАБЛЮДЕНИЯ И УПРАВЛЕНИЯ

В.И.Кортунов

Анализируются критерии выбора номинальных моделей в задачах управления и наблюдения, доказываются условия существования таких моделей в схемах комбинированного управления. Произведены постановки задач параметрического синтеза номинальных моделей по множественным априорным данным исходной модели объекта.

Аналізуються критерії вибору номінальних моделей в задачах керування та спостереження, доводяться умови існування таких моделей в схемах комбінованого керування. Здійснені постановки задач параметричного синтезу номінальних моделей з багатьма априорними даними вихідної моделі об'єкту.

Criteria of choice of nominal models for control and observation are analysed. Existence conditions of such models are proved for combined control schemes. The statements of parametrical synthesis of nominal models using multiple priori dates of object origin model are produced.

ВВЕДЕНИЕ

При проектировании систем управления, реализующих робастное управление с внутренними моделями как самого объекта, так и внешних возмущений [1,2], предполагается знание номинальной модели (НМ) с фиксированными параметрами. Качество таких методов управления зависит от диапазонов изменения параметров объекта и структуры выбранных моделей. Модель может функционировать параллельно с объектом управления или с замкнутой системой в целом. В последнем случае номинальную модель считают эталонной моделью, как отражение требуемых показателей качества системы управления. В большинстве случаев задача синтеза номинальной модели решается интуитивно на основе опыта разработчиков, хотя выбор номинальной модели влияет на параметры законов управления и качество системы в целом.

В статье анализируются критерии выбора НМ по энергетическому критерию и доказываются условия суще-

ствования таких моделей в схемах комбинированного управления.

1. КРИТЕРИИ СИНТЕЗА НОМИНАЛЬНЫХ МОДЕЛЕЙ

На выбор НМ оказывает влияние используемый метод управления, в котором могут по-разному реализовываться внутренние модели. Соответствие номинальной модели объекту, целям и задачам управления, имеющимся ограничениям, техническим средствам реализации может иметь место при определенных ее свойствах. Так, применительно к адаптивным системам использование эталонных моделей приводит к выводам [3], что порядки модели объекта и номинальной модели должны совпадать для обеспечения сходимости алгоритмов адаптации. Для робастных методов управления важным является сохранение свойств устойчивости управления при множественной параметрической неопределенности объекта [4]. В этом случае критерием синтеза следует выбрать максимальную робастность для обеспечения устойчивости системы управления при наиболее широких отклонениях параметров объекта от номинальных [8]. Для систем с ограниченными ресурсами на управление необходимо выбирать НМ по энергетическому критерию. Таким образом, синтез НМ, как процедура упрощения или редукции исходной модели, необходимо решать на основе некоторого подчиненного критерия.

Если при синтезе НМ ставится цель упрощения или редукции модели, то можно воспользоваться следующими известными методами получения упрощенных или номинальных моделей: принцип сравнения, метод сингулярных возмущений, метод весовых функций, метод матричных неравенств, метод аппроксимации или редукции.

Идея метода принципа сравнения [5] состоит в замене

некоторой сложной модели более простой моделью по некоторому показателю так, что на всем временном интервале выбранный показатель для упрощенной модели был не меньше (не больше), чем для полной модели. Применительно к динамическим процессам конструктивные решения получены на основе функций Ляпунова, как некоторые мажоранты динамических процессов или модели сравнения [5]. Эти модели мажорируют норму вектора состояния полной модели и решают задачу выбора номинальной модели в виде линейного стационарного динамического звена первого порядка. Для сложных моделей, которые можно декомпозировать на несколько простых, применяется принцип сравнения на основе векторных функций Ляпунова.

Метод сингулярных возмущений основан на декомпозиции - агрегировании и относится к классу методов со структурными возмущениями. Метод сингулярных возмущений позволяет декомпозировать модель на составные части с медленной и быстрой динамикой. Применительно к задаче выбора номинальной модели конструктивность метода ограничивается разделением модели на элементы с отличающимися масштабами времени.

В методе весовых функций [5], если собственные значения полной линейной модели различны, то передаточная функция модели сравнения представляется аperiodическим звеном первого порядка. В этом звене постоянная времени выбирается обратной к величине минимального по модулю собственного значения, а коэффициент усиления как среднее геометрическое значение по всем входам и выходам.

В методе матричных неравенств [5] модели сравнения можно получить на основе оценки решения уравнения модели по норме, а различие номинальных моделей будет определяться различием используемых матричных норм. Конструктивные решения для мажорант и минорант получены для логарифмической матричной нормы.

Методы получения моделей сравнения верхнего и нижнего пределов параметров показывают возможный диапазон свойств номинальных моделей, но не показывают способ выбора номинальной модели из полученного диапазона.

В статье ставится задача синтеза номинальной модели из условий реализации управления по энергетическому критерию. Сформулируем требования к номинальной модели и обоснуем критерии синтеза такой модели.

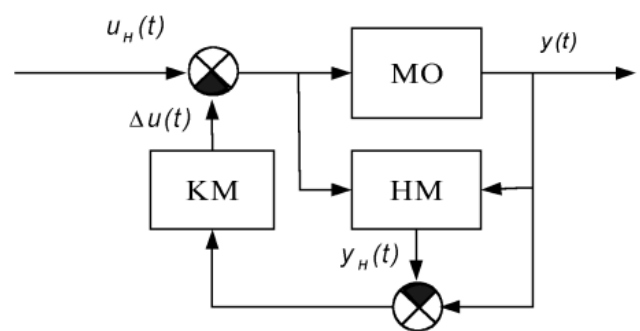
Номинальная модель должна удовлетворять следующим требованиям:

- энергетической достижимости (энергетической ограниченности сигналов управления);
- физическим ограничениям на фазовые переменные и параметры;
- соответствию, принятому уровню адекватности;
- конструктивности в решении задач синтеза алгоритмов контроля и управления;
- технической реализуемости в системах управления.

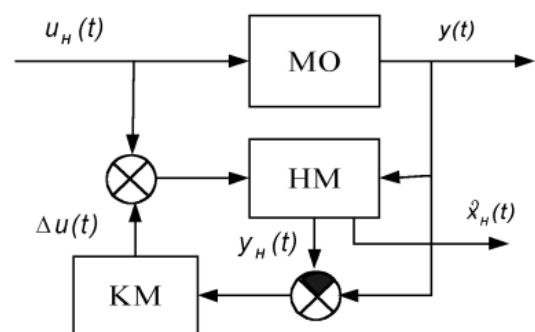
Энергетическая ограниченность сигналов управления вызвана существующими ограничениями исполнительных органов, как по амплитуде, так и по мощности. Упрощенная номинальная модель может повысить реализуемость алгоритмов управления, но может не обеспечить энергетическую ограниченность, что приведет к дополнительным затратам на управление.

Известные схемы формирования дополнительного сигнала управления по внутренней модели принципиально сводятся к двум типам: "дифференциальной вилке" [6] и схеме с внутренней моделью [2,7].

Далее рассматриваем задачу синтеза НМ для схемы, показанной на рисунке 1а, где обозначено: КМ - корректирующая модель; МО - модель объекта; НМ - номинальная модель.



а)



б)

Рисунок 1 - Схемы взаимосвязи двухвходовой НМ в задаче управления (а) и наблюдения (б)

2. ПОСТАНОВКА ЗАДАЧИ

Объект управления описывается системой нелинейных уравнений

$$\dot{x}(t) = f(x, u, \Theta, t), \quad x(t_0) = x_0, \quad y(t) = g(x, u, \Theta, t), \quad (1)$$

где $x(t) \in R^{n_x}$ - вектор состояния, $u(t) \in R^{n_u}$ - вектор управления, $y(t) \in R^{n_y}$ - вектор выхода, $\Theta(t) \in \Omega_\Theta$ - изменяющийся вектор параметров объекта из некоторого ограниченного множества Ω_Θ , R^n - конечномерное евклидово пространство размерности n .

Для системы (1) предполагаем существование линейной модели, достаточно близкой к объекту по норме $\|y(t) - y_H(t)\|_{L_2} < \Delta_y$ при сравнительно малой величине Δ_y , и представляемой системой

$$\dot{x}_H(t) = A^H x_H(t) + B_u^H u_H(t), \quad y_H(t) = C^H x_H(t) + D_u^H u_H(t), \quad (2)$$

где $x_H(t) \in R^{n_x}$, $u(t) \in R^{n_u}$, $y_H(t) \in R^{n_y}$, $n_x \leq n_x^0$ с соответствующими размерностями матриц A^H , B_u^H , C^H , D_u^H состояния, управления и наблюдения.

Определим линейную модель объекта через номинальную

$$\begin{aligned} \dot{x}(t) &= A^H x(t) + B_u^H(u_H(t) + \Delta u(t)) + B_v^H v(t), \\ y(t) &= C^H x(t) + D_u^H(u_H(t) + \Delta u(t)) + D_v^H v(t), \end{aligned} \quad (3)$$

где $v(t) \in R^{n_v}$ - вектор эквивалентных возмущений, учитывающий внутренние и внешние возмущения, $\Delta u(t)$ - сигнал дополнительного управления из схемы компенсации, B_v^H , D_v^H - матрицы управления и наблюдения по возмущению.

Введем следующие определения.

Определение 1. Линейная номинальная модель (2) называется достижимой по выходу, если для любого ограниченного вектора эквивалентного возмущения $v(t) \in \Omega_v$, некоторого номинального вектора параметров $\Theta_H \in \Omega_\Theta$ существует ограниченный вектор управления $u_H(t) + \Delta u(t) \in L_2$ такой, что для моделей (2) и (3) выполняется $\|y(t) - y_H(t)\|_{R^{n_y}} \rightarrow 0$ при $t \rightarrow \infty$.

Определение 2. Линейная номинальная модель (2) называется частично достижимой, если для любого ограниченного вектора эквивалентного возмущения $v(t) \in \Omega_v$, некоторого номинального вектора параметров $\Theta_H \in \Omega_\Theta$ и для некоторого уровня согласованности Δ_y моделей (2) и (3) существует ограниченный вектор управления $u_H(t) + \Delta u(t) \in \Omega_u$ такой, что выполняется неравенство $\|y(t) - y_H(t)\|_{L_2} < \Delta_y$.

Представим зависимость входа и выхода объекта через линейную модель (3) в операторной форме

$$y(t) = W_0^u(p)u(t) + W_0^v(p)v(t) + y_0(t, x_0), \quad (4)$$

где $W_0^u(p)$, $W_0^v(p)$ - динамические операторы модели объекта по управлению и возмущению, а p - оператор дифференцирования, $y_0(t, x_0) = W_0^{x_0}(p)x_0$ - составляющая выхода собственного движения, определяемая вектором начальных условий x_0 . Двухходовую номинальную модель запишем в форме наблюдателя состояния

$$\hat{y}_e(t) = W_e^u(p)u(t) + W_e^y(p)y(t) + \hat{y}_0(t, x_0), \quad (5)$$

где $W_e^u(p)$, $W_e^y(p)$ - динамические операторы наблюдателя НМ по управлению и наблюдению, $\hat{y}_0(t, x_0) = W_e^{x_0}(p)\hat{x}_0$ - составляющая выхода собственного движения НМ. При параметризации НМ предполагаем, что $\|\hat{y}_0(t, x_0)\|_{R^{n_y}} \rightarrow 0$ при $t \rightarrow \infty$.

Для оператора модели объекта предполагаем аддитивное выделение НМ

$$W_0^u(p) = W_H^u(p) + \Delta W_0^u(p), \quad (6)$$

где $\Delta W_0^u(p)$ - аддитивные непараметрические неопределенности, $W_H^u(p) = C^H(sI - A^H + LC^H)^{-1}B^H + D^H$ - номинальная часть модели.

Дополнительное управление в соответствии со схемой на рис.1а запишем $\Delta u(t) = -W_K(p)(y(t) - \hat{y}_e(t))$, где $W_K(p)$ - оператор корректирующей модели.

Ставится задача определения условий сохранения свойств достижимости НМ с множественными значениями параметров объекта $\Theta \in \Omega_\Theta$.

3. УСЛОВИЯ ДОСТИЖИМОСТИ НОМИНАЛЬНОЙ МОДЕЛИ

НМ выбираем по соображениям ограниченности дополнительного сигнала управления, тогда соответствующая замкнутая система по свойству достижимости является робастной. Сохранение свойств робастности управления по модели, как и необходимого условия достижимости НМ, возможно при выполнении условий, которые даются следующей теоремой.

Приведем без доказательства следующее важное утверждение для линейных систем.

Лемма. Для линейной системы вида (2) справедливы тождества матричных передаточных функций

$$\begin{aligned} [I + C^H(sI - A^H)^{-1}L]^{-1} &= I - C^H(sI - A^H + LC^H)^{-1}L, \\ (I - W_e^y(p))W_H^u(p) &= W_e^u(p). \end{aligned}$$

Теорема 1 (Q-критерий достижимости по выходу НМ). Номинальная линейная модель, определяемая вектором

параметров $\Theta_H \in \Omega_\Theta$, будет полностью достижимой по выходу для всех видов неопределенностей объекта и всех возможных значений параметров $\Theta \in \Omega_\Theta$, если:

а) вектор эквивалентного возмущения является ограниченным $v(t) \in \Omega_v$ и выполняется условие согласованности управления и возмущения;

б) матрица $A^H - LC^H$ гурвицева;

с) $n_u = n_y$ и вектор дополнительного управления определяется оператором КМ $W_K(p) = (W_e^u(p))^{-1}$, а $W_e^u(p)$ - обратимый оператор;

д) выполняется условие $\|(W_H^u(p))^{-1} \Delta W_0^u(p, \Theta)\|_{H_\infty} = q < 1$ на всем множестве значений параметров $\Theta \in \Omega_\Theta$, тогда $\|y(t) - y_H(t)\|_{R^{n_y}} \rightarrow 0$ при $t \rightarrow \infty$ для ограниченного сигнала управления.

Доказательство.

Запишем выход объекта и сигнал дополнительного управления в соответствии со схемой на рис.1.а

$$\begin{aligned} y(t) &= W_0^u(p)(u_H(t) + \Delta u(t)) + W_0^v(p)v(t) + y_0(t, x_0), \\ \Delta u(t) &= -W_K(p)(y(t) - W_e^u(p)(u_H(t) + \Delta u(t)) - \\ &\quad - W_e^v(p)y(t) - \hat{y}_0(t, \hat{x}_0)). \end{aligned}$$

Дополнительное управление после преобразования можно записать через операторное выражение

$$\begin{aligned} [W_K^{-1}(p) - W_e^u(p) + (I - W_e^v(p))W_0^u(p)]\Delta u(t) &= \\ = [- (I - W_e^v(p))W_0^u(p) + W_e^u(p)]u_H(t) - \\ - (I - W_e^v(p))[W_0^v(p)v(t) + y_0(t, x_0)] + \hat{y}_0(t, \hat{x}_0). \end{aligned}$$

Из предположения теоремы, что $W_K(p) = (W_e^u(p))^{-1}$ дополнительное управление упрощается

$$\begin{aligned} \Delta u(t) &= -[I - (W_0^u(p))^{-1}(I - W_e^v(p))^{-1}W_e^u(p)]u_H(t) - \\ - (W_0^u(p))^{-1}[W_0^v(p)v(t) + y_0(t, x_0)] + \\ + (W_0^u(p))^{-1}(I - W_e^v(p))^{-1}\hat{y}_0(t, \hat{x}_0). \end{aligned}$$

На основании операторного тождества леммы $(I - W_e^v(p))^{-1}W_e^u(p) = W_H^u(p)$ следует упрощенное выражение дополнительного управления

$$\begin{aligned} \Delta u(t) &= -[I - (W_0^u(p))^{-1}W_H^u(p)]u_H(t) - \\ - (W_0^u(p))^{-1}[W_0^v(p)v(t) + y_0(t, x_0)] + \\ + (W_0^u(p))^{-1}(I - W_e^v(p))^{-1}\hat{y}_0(t, \hat{x}_0). \end{aligned} \quad (7)$$

Подстановка вектора дополнительного управления $\Delta u(t)$ в операторное уравнение выхода МО дает

$$\begin{aligned} y(t) &= W_0^u(p)[I - (I - (W_0^u(p))^{-1}W_H^u(p))]u_H(t) + \\ + W_0^u(p)[- (W_0^u(p))^{-1}[W_0^v(p)v(t) + y_0(t, x_0)]] + \\ + W_0^v(p)v(t) + y_0(t, x_0) = \\ = W_H^u(p)u_H(t) + (I - W_e^v(p))^{-1}\hat{y}_0(t, \hat{x}_0). \end{aligned}$$

Последнее слагаемое выражения можно упростить на основе матричных тождеств леммы и

$$\begin{aligned} (I - W_e^v(p))^{-1}\hat{y}_0(t, \hat{x}_0) &= (I - W_e^v(p))^{-1}W_e^{\hat{x}_0}(p)\hat{x}_0 = \\ = C^H(pI - A^H + LC^H)^{-1}\hat{x}_0. \end{aligned}$$

Разность выходов $y(t) - y_H(t) = C^H(pI - A^H + LC^H)^{-1}\hat{x}_0$ можно оценить по норме и $\|y(t) - y_H(t)\|_{R^{n_y}} \rightarrow 0$ вследствие

предположения устойчивости НМ, что доказывает асимптотическое совпадение выходов модели объекта и НМ.

Доказательство ограниченности сигнала $\Delta u(t)$ можно получить оценкой по норме выражения (7).

Конструктивный подход в синтезе НМ можно получить для оптимально достижимой НМ.

Определение 3. Номинальную модель, определяемую вектором параметров $\Theta_H^* \in \Omega_\Theta$, назовем оптимально достижимой, если для ограниченного вектора эквивалентного возмущения $v(t) \in \Omega_v$ существует ограниченный вектор управления $u_H(t) + \Delta u(t) \in \Omega_u$, выполняется условие достижимости по выходу и параметры номинальной модели являются оптимальными по критерию энергетических затрат вектора дополнительного управления $\Delta u(t)$

$$\Theta_H^* = \arg \min_{\Theta_H \in \Omega_\Theta} \int_{\Omega_\Theta} \|\Delta u(t, \Theta, \Theta_H)\|_{L_2}^2 d\Theta, \quad (8)$$

на множестве возможных значений номинальных параметров.

Оптимальная достижимость номинальной модели определяется способом оценки эквивалентного возмущения, который накладывает особенности при выполнении достижимости. Часто важно обеспечить достижимость только для некоторой рабочей полосы частот и, соответственно, необходимо модифицировать и критерии достижимости.

Некоторым ограничением выполнения критерия достижимости модели является обратимость оператора $W_e^u(p)$.

Если ограничить частотную полосу восстанавливаемых возмущений, то для обращения можно использовать эффективные итерационно инверсные фильтры [9,10] и выполнить тем самым условие частичной достижимости.

4. КРИТЕРИИ ПАРАМЕТРИЧЕСКОГО СИНТЕЗА НМ

Критерием выбора номинальной модели считаем интегральную характеристику на множестве возможных значений номинальной модели от нормы вектора дополнительного управления

$$J(\Theta_H) = \int_{\Omega_\Theta} \|\Delta u(t, \Theta, \Theta_H)\|_{L_2}^2 d\Theta. \quad (9)$$

Из схемы включения НМ, представленной на рис.1а, запишем вектор $\Delta u(t, \Theta, \Theta_H)$ в виде

$$\Delta u(t, \Theta, \Theta_H) = -(W_0^u(p, \Theta))^{-1} \times [(W_0^u(p, \Theta) - W_H^u(p, \Theta_H))u_H(t) - W_0^v(p, \Theta)v(t)]. \quad (10)$$

Выражение (10) содержит разность выходов полной и номинальной моделей на один и тот же входной сигнал $u_H(t)$, который добавляет множитель в энергетический критерий и не зависит от синтезируемых параметров. Для решения задачи синтеза НМ по критерию (9) этот сигнал должен удовлетворять требованиям широкополосного сигнала и ограниченного по энергии. Таким требованиям отвечает сигнал "единичной ступеньки" или "белого шума" с ограниченной полосой, который в пределах полосы пропускания системы управления можно считать "белым шумом". При непосредственном использовании выражения сигнала дополнительного управления (10) в критерии (9) характеристики возмущений не оказывают влияние на оптимальные значения номинальных параметров, поэтому второе слагаемое в (10) можно опустить и критерий параметрического синтеза запишется

$$J(\Theta_H) = \int_{\Omega_\Theta} \|(W_0^u(p, \Theta))^{-1}(y(t, \Theta) - y_H(t, \Theta_H))\|_{L_2}^2 d\Theta. \quad (11)$$

Если сравнить выходные сигналы полной и номинальной моделей при рассмотренном случайном входном сигнале, критерием сравнения следует считать энергетическую мощность разности выходов, к которым применен оператор $(W_0^u(p))^{-1}$.

Критерий (11) можно модифицировать на основе имеющейся статистической информации о плотности распределения параметров, тогда критерий среднего риска запишем

$$J(\Theta_H) = \int_{\Omega_\Theta} \|\Delta u(t, \Theta, \Theta_H)\|_{L_2}^2 f_\Theta(\Theta) d\Theta, \quad (12)$$

где $f_\Theta(\Theta)$ - функция плотности распределения параметров.

При известной структуре номинальной модели параметрический синтез такой модели заключается в выборе оптимальных по критерию (12) параметров с учетом имеющихся ограничений

$$\Theta_H^* = \arg \min_{\Theta_H \in \Omega_\Theta} J(\Theta_H), J_i(\Theta_H) \leq J_i^*, i = \overline{1, n_J}. \quad (13)$$

Решение оптимизационной задачи в общем виде получить затруднительно, но для конкретных линейных моделей с интервальными характеристиками возможны аналитические решения.

ПЕРЕЧЕНЬ ССЫЛОК

1. Цыпкин Я.З. Синтез робастно-оптимальных систем управления объектами в условиях ограниченной неопределенности // АИТ. - 1992. - № 9. - С. 139-159.
2. Garsia С.Е., Morari М. Internal model control/ A unified review and some new results // Ind. Eng. Chem. Process Des. and Dev. -1982.-V.21.№3. -PP. 308-323.
3. Громыко В.Д., Санковский Е.А. Самонастраивающиеся системы с моделью. - М.: Энергия.,1974. - 80 с.
4. Кунцевич В.М. О редуцированных моделях дискретных динамических объектов и их гарантированных оценках в задачах управления // Проблемы управления и информатики. -2001. -№ 1. -С.42-50.
5. Воронов А.А. Введение в динамику сложных управляемых систем. - М.: Наука, 1985. -362 с.
6. Ивахненко А.Г. Кибернетические системы с комбинированным управлением. - Киев: "Техніка", 1966. -512 с.
7. Чаки Ф. Современная теория управления. Нелинейные оптимальные и адаптивные системы. -М.:Мир,1975. - 424 с.
8. Вишняков А.Н. Синтез максимально-робастной системы управления дискретным объектом с непараметрической неопределенностью // АИТ. -1999. -№3. -С. 71-77.
9. Кортунов В.И. Восстановление возмущений в динамической системе с заданной точностью // Системы обработки информации. - Харків: НАНУ, ПАНМ, ХВУ. - 2000. - Вип.3(9). - С.55-60.
10. Кортунов В.И. Оценивание возмущений в динамических системах на основе итерационно-инверсных моделей// Системы обработки информации.- Харків: НАНУ, ПАНМ, ХВУ.- 2001.- Вип.3(13). - С. 56-60.

УДК 621.317

ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНАЯ СИСТЕМА КОНТРОЛЯ КАЧЕСТВА ФИЛЬТРОВ ГАЗОДИНАМИЧЕСКИМ МЕТОДОМ

А.В.Печенов, А.Л.Симаков, Ю.Н.Матросова

Разработана информационно-измерительная система, реализующая измерение проницаемости материалов. Получено аналитическое решение системы нестационарных дифференциальных уравнений движения газа через проницаемый материал, описывающих данную систему.

The informationally-metrical system realizing the material permeability measurement is developed. The analytical solution of

the non-steady differential equations of gas motion through a permeable material for the system has been obtained.

При изготовлении и использовании фильтрующих элементов неизбежно встает вопрос контроля их качества. В зависимости от конкретной области использования фильтра для контроля выбираются те или иные параметры.

Существует несколько методов определения параметров пористого материала, из которого выполнен фильтр, но наиболее перспективным из них является газодинамический метод [1]. Он основан на том, что динамика течения газа через пористый материал определяется структурой его пор. Реализация этого метода требует проведения многократных высокоточных измерений и трудоемкой статистической обработки экспериментальных данных. Автоматизация измерения параметров фильтров позволит сократить время контроля, повысить объективность получаемых данных, исключить грубые ошибки измерений.

Таким образом, задачей работы является, во-первых, получение зависимостей, связывающих динамику газовых процессов с параметрами, описывающими структуру материала фильтра, во-вторых, создание программного и аппаратного обеспечения для информационно-измерительной системы, алгоритм работы которой основан на полученных уравнениях. Конечным результатом применения данной системы является определение активной пористости, коэффициента проницаемости, эквивалентного диаметра пор материала фильтра.

Предлагаемая структура информационно-измерительной системы разработана на основе физической модели, приведенной на рисунке 1. Модель можно разделить на две функциональные части: систему наполнения и измерительный блок.

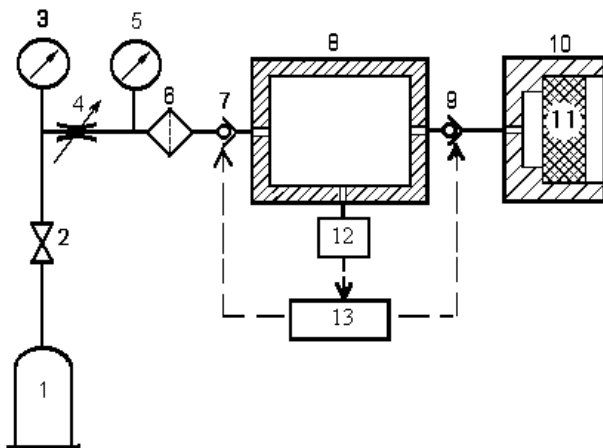


Рисунок 1 - Физическая модель информационно-измерительной системы

Система наполнения предназначена для подачи в измерительный блок газов или газовых смесей. Она состоит из баллона высокого давления 1, снабженного вентилем 2, манометрами 3 и 5, дросселем 4. Для очистки подаваемого газа предусмотрен фильтр-осушитель 6. Система наполнения и емкость 8 соединяются трубопроводом, имеющим электропневмоклапан 7.

Измерительный блок служит для получения и обработки данных о давлении в емкости 8, для управления работой электропневмоклапанов 7 и 9. Он состоит из датчика

давления, электронной системы обработки информации и измерительной головки.

В качестве датчика давления 12 используется преобразователь абсолютного давления в унифицированный токовый сигнал "Сапфир-22".

Система обработки информации состоит из ЭВМ 13, которая через параллельный порт (LPT1) связана с датчиком давления и электропневмо-клапанами 7 и 9.

Измерительная головка 10 предназначена для закрепления исследуемого фильтра 11 и создания условий для течения через него газа.

В калиброванную емкость подается газ, начальное давление, температура и плотность которого равны соответственно P_0 , T_0 , ρ_0 . При этом выполняется условие $P_0 > P_{\text{АТМ}}$. Для сохранения перепада давлений между калиброванной емкостью и атмосферой электропневмо-клапан 9 закрыт.

При определении параметров материала клапан 9 открывается и под действием избыточного давления газ из калиброванной емкости выходит в атмосферу, при этом фильтруясь через испытываемый материал. Параметры материала: активная пористость, проницаемость и эквивалентный диаметр пор определяются в результате измерения изменения давления в калиброванной емкости.

Исходными зависимостями для получения расчетных уравнений являются законы сохранения массы и количества движения газа [2], уравнения связывающие параметры газа (давление, плотность, скорость фильтрации) и уравнения Вейсбаха, учитывающие потери давления на границе пористого тела:

$$\varepsilon \frac{\partial \rho}{\partial t} + \text{div}(\rho \vec{u}) = 0; \quad (1)$$

$$\varepsilon \rho \frac{\partial u_i}{\partial t} + \rho (\vec{u} \nabla) u_i + F_i + \frac{\partial p}{\partial x_i} = 0; \quad (2)$$

$$u = K_{\text{И}} \frac{V}{S} \frac{dP}{dt}; \quad (3)$$

$$\rho = \frac{P}{K_{\text{И}} R_0 T}; \quad (4)$$

$$\Delta P_{\text{ВХ}} = \zeta_{\text{ВХ}} \left(\frac{\rho u^2}{2} \right)_{\text{ВХ}}; \quad (5)$$

$$\Delta P_{\text{ВЫХ}} = \zeta_{\text{ВЫХ}} \left(\frac{\rho u^2}{2} \right)_{\text{ВЫХ}}. \quad (6)$$

Здесь ε - активная пористость материала; ρ - плотность газа; u - скорость течения газа; F_i - межфазная сила трения; V - объем калиброванной емкости; S - площадь поперечного сечения исследуемого образца; P - давление газа в емкости; R_0 - удельная газовая постоянная; T - температура газа; $K_{\text{И}} = P_{\text{ИЗБ}} / (P_{\text{АТМ}} + P_{\text{ИЗБ}})$ - коэффициент избыточности давления; $\Delta P_{\text{ВХ}}$, $\Delta P_{\text{ВЫХ}}$ - потери давления на входной и выходной границе пористого тела соответствен-

но; $\zeta_{ВХ}$, $\zeta_{ВЫХ}$ - коэффициенты потерь соответственно на входе и выходе пористого тела; $(\rho u^2/2)_{ВХ}$, $(\rho u^2/2)_{ВЫХ}$ скоростной напор газа на входе и выходе пористого тела соответственно; $P_{АТМ}$ - атмосферное давление, $P_{ИЗБ}$ - избыточное давление в калиброванной емкости, t - время; x - координата.

Решение для давления газа в калиброванной емкости будем искать методом разделения переменных в виде

$$P = P_{АТМ} + P_{ИЗБ} = P_{АТМ} + T(t)X(x). \quad (7)$$

Для решения полученной системы сначала подставим в выражение (1) уравнения (3, 4)

$$\begin{aligned} \varepsilon \frac{\partial p}{\partial t} + u \frac{\partial p}{\partial x} &= \frac{\varepsilon}{K_{И} R_0 T'} X(x) \frac{\partial T}{\partial t} + \\ &+ \frac{V}{S} \frac{1}{T(t)} \frac{\partial T(t)}{\partial t} K_{И} \frac{1}{K_{И} R_0 T'} T(t) \frac{\partial X(x)}{\partial x} = \\ &= \varepsilon X(x) \frac{\partial T}{\partial t} + \frac{V}{S} K_{И} \frac{\partial T \partial X(x)}{\partial t \partial x} = \\ &= \frac{\partial T}{\partial t} \left[\varepsilon X(x) + \frac{V}{S} K_{И} \frac{\partial X(x)}{\partial x} \right] = 0. \end{aligned} \quad (8)$$

В результате получим выражение

$$X = A \exp\left(-\frac{\varepsilon S L}{V K_{И}} x^*\right). \quad (9)$$

Для определения конкретного значения коэффициента A рассмотрим граничные условия. При $x^*=0$ $X=P_0$, следовательно, $A=P_0$. При этом при $x^*=1$ $X=P_0 \exp\left(-\frac{\varepsilon S L}{V K_{И}}\right)$, но по граничным условиям необходимо, чтобы при $x^*=1$ $X=0$.

Видим, что при данном решении уравнения (8) невозможно получить решение, удовлетворяющее его граничным условиям.

Для того чтобы описать реальное распределение давления по толщине материала предположим, что функция X имеет вид

$$X = A \exp(-kx^*) + B. \quad (10)$$

Для удовлетворения граничных условий необходимо использовать следующее выражение

$$X = \frac{P_0}{[\exp(-k) - 1]} [\exp(-kx) - 1]. \quad (11)$$

Значение k является коэффициентом при переменной x^* в выражении (9). Анализируя формулу (11) видим, что выражение $P_0 \exp(-kx)/[\exp(-k) - 1]$ определяет форму распределения давления по толщине материала, а выражение $-P_0/[\exp(-k) - 1]$ является постоянной величиной.

Однако, выражение (11) не является решением уравне-

ния (8). Скорректируем уравнение (11), используя формулу (7). Не изменяя характера распределения давления по толщине материала увеличим его начальный уровень на величину $P_0/[\exp(-k) - 1]$, а затем вычтем эту же величину из значения давления окружающей среды. Возникающие вследствие этого погрешности учтем введением корректирующей функции $\Phi(t)$.

В итоге уравнение для X принимает вид

$$X = \frac{P_0 \exp(-\varepsilon S L x / V K_{И})}{\exp(-\varepsilon S L / V K_{И}) - 1}, \quad (12)$$

а выражение (7) перепишем в виде

$$P = P_{А} + P_{ИЗБ} = P_{А} - \frac{P_0}{\exp\left(-\frac{\varepsilon S L}{V K_{И}}\right) - 1} \Phi(t) + T(t)X(x). \quad (13)$$

Далее решим уравнение (2), учитывая зависимости (3,4) и задавая начальные условия.

Межфазная сила трения F_i определится выражением

$$F_i = \frac{\mu u}{k_{ПП}} + \frac{\lambda \rho u^2}{k_{ПП}}, \quad (14)$$

Для медленных течений, к которым относится фильтрационное, можно пренебречь вторым членом уравнения (14). Выразим все переменные величины в этой формуле через давление, при этом заметим, что динамическая вязкость μ зависит от давления газа. Чтобы избавиться от этой зависимости перейдем от динамической вязкости μ к кинематической вязкости ν согласно выражения $\nu = \mu/\rho$.

$$\begin{aligned} F &= \frac{\mu u}{k_{ПП}} = \frac{\nu \rho}{k_{ПП}} \frac{V K_{И}}{S} \frac{1}{T(t)} \frac{dT(t)}{dt} = \\ &= \frac{\nu T(t) X(x)}{k_{ПП} R_0 T' K_{И}} \frac{V K_{И}}{S} \frac{1}{T(t)} \frac{dT(t)}{dt} = \frac{\nu X(x)}{k_{ПП} R_0 T' S} \frac{dT(t)}{dt}. \end{aligned} \quad (15)$$

Подставим выражения (3),(4),(15) в дифференциальное уравнение (2)

$$\begin{aligned} \varepsilon \frac{T(t) X(x)}{R_0 T'} \frac{V}{S} \left[\frac{1}{T(t)} \frac{d^2 T(t)}{dt^2} + \frac{dT(t)}{dt} \frac{dT^{-1}(t)}{dt} \right] + \\ + \frac{\nu X(x)}{k_{ПП} R_0 T' S} \frac{V K_{И}}{S} \frac{dT(t)}{dt} + T(t) \frac{dX(x)}{dx} = 0. \end{aligned} \quad (16)$$

Решая последнее уравнение, получим выражение для определения T

$$T = \exp\left[-\frac{\varepsilon S^2 L}{V^2 K_{И}} \frac{k_{ПП} R_0 T'}{\nu} t\right]. \quad (17)$$

Подставим полученные выражения в формулу (13)

$$P_{\text{АБС}} = P_{\text{А}} - \frac{P_0 \Phi(t)}{\exp\left(-\frac{\varepsilon SL}{VK_{\text{И}}}\right) - 1} + \frac{P_0}{\exp\left(-\frac{\varepsilon SL}{VK_{\text{И}}}\right) - 1} \exp\left[-\frac{\varepsilon SL}{VK_{\text{И}}}x - \frac{\varepsilon S^2 L k_{\text{ПП}} R_0 T'}{V^2 K_{\text{И}} v} t\right]. \quad (18)$$

Анализируя выражение (18), видим, что корректировочная функция $\Phi(t)$ должна удовлетворять условиям $\Phi(0) = 1$, $\Phi(\infty) = 0$. Другими словами, корректировочная функция, учитывающая отклонение высоты профиля распределения давления от значения давления окружающей среды, с увеличением t должна уменьшаться аналогично уменьшению функции T (17). Поэтому принимаем

$$\Phi(t) = T = \exp\left[-\frac{\varepsilon S^2 L k_{\text{ПП}} R_0 T'}{V^2 K_{\text{И}} v} t\right]. \quad (19)$$

Проведя преобразования, окончательно перепишем выражение (18) в виде:

$$P = P_{\text{А}} + P_0 \exp\left[-\frac{\varepsilon S^2 L k_{\text{ПП}} R_0 T'}{V^2 K_{\text{И}} v} t\right] \frac{1 - \exp\left(-\frac{\varepsilon SL}{VK_{\text{И}}}x\right)}{1 - \exp\left(-\frac{\varepsilon SL}{VK_{\text{И}}}\right)}. \quad (20)$$

Для описания фильтрации газа в реальных системах необходимо ввести поправку на перепад давлений при входе и выходе газа из пористого материала. Известно, что входные и выходные потери давления газа в пористой среде рекомендуется определять по формулам Вейсбаха (5,6).

Подставляя выражения (3), (4) в уравнение для скоростного напора получим

$$\frac{\rho u^2}{2} = \frac{R_0 T'}{K_{\text{И}}} \left[\frac{\varepsilon SL k_{\text{ПП}}}{V v}\right]^2 \frac{P_0 \exp\left(-\frac{\varepsilon SL}{VK_{\text{И}}}x\right)}{\exp\left(-\frac{\varepsilon SL}{VK_{\text{И}}}\right) - 1} \times \exp\left[-\frac{\varepsilon S^2 L k_{\text{ПП}} R_0 T'}{V^2 K_{\text{И}} v} t\right]. \quad (21)$$

Таким образом, выражения (5), (6) можно записать в виде

$$\Delta P_{\text{ВХ}} = \zeta_{\text{ВХ}} \frac{R_0 T'}{K_{\text{И}}} \left[\frac{\varepsilon SL k_{\text{ПП}}}{V v}\right]^2 \frac{P_0 \exp\left(-\frac{\varepsilon SL}{VK_{\text{И}}}x\right)}{\exp\left(-\frac{\varepsilon SL}{VK_{\text{И}}}\right) - 1} \times \exp\left[-\frac{\varepsilon S^2 L k_{\text{ПП}} R_0 T'}{V^2 K_{\text{И}} v} t\right], \quad (22)$$

для входного сечения $x=1$;

$$\Delta P_{\text{ВЫХ}} = \zeta_{\text{ВЫХ}} \frac{R_0 T'}{K_{\text{И}}} \left[\frac{\varepsilon SL k_{\text{ПП}}}{V v}\right]^2 \frac{P_0}{\exp\left(-\frac{\varepsilon SL}{VK_{\text{И}}}\right) - 1} \times \exp\left[-\frac{\varepsilon S^2 L k_{\text{ПП}} R_0 T'}{V^2 K_{\text{И}} v} t\right], \quad (23)$$

для выходного сечения $x=0$;

С учетом потерь давления на входе и выходе пористого материала выражение (20) примет вид

$$P = P_{\text{А}} + \frac{R_0 T'}{K_{\text{И}}} \left[\frac{\varepsilon SL k_{\text{ПП}}}{V v}\right]^2 \frac{P_0 \left(\zeta_{\text{ВЫХ}} + \zeta_{\text{ВХ}} \exp\left(-\frac{\varepsilon SL}{VK_{\text{И}}}\right)\right)}{\exp\left(-\frac{\varepsilon SL}{VK_{\text{И}}}\right) - 1} \times \exp\left[-\frac{\varepsilon S^2 L k_{\text{ПП}} R_0 T'}{V^2 K_{\text{И}} v} t\right] + \left(P_0 - \frac{R_0 T'}{K_{\text{И}}} \left[\frac{\varepsilon SL k_{\text{ПП}}}{V v}\right]^2 \times \frac{P_0 \left(\zeta_{\text{ВЫХ}} + \zeta_{\text{ВХ}} \exp\left(-\frac{\varepsilon SL}{VK_{\text{И}}}\right)\right)}{\exp\left(-\frac{\varepsilon SL}{VK_{\text{И}}}\right) - 1} \exp\left[-\frac{\varepsilon S^2 L k_{\text{ПП}} R_0 T'}{V^2 K_{\text{И}} v} t\right]\right) \times \exp\left[-\frac{\varepsilon S^2 L k_{\text{ПП}} R_0 T'}{V^2 K_{\text{И}} v} t\right] \frac{1 - \exp\left(-\frac{\varepsilon SL}{VK_{\text{И}}}x\right)}{1 - \exp\left(-\frac{\varepsilon SL}{VK_{\text{И}}}\right)}. \quad (24)$$

На основе данного уравнения реализовано программное обеспечение информационно-измерительной системы. Программный продукт создан в среде Turbo Vision, алгоритм работы программы приведен на рисунке 2.



В результате работы получены зависимости, связывающие динамику фильтрационного течения газа с внутренним строением пористого материала. Разработана информационно-измерительная система для контроля качества фильтров газодинамическим методом. Система может применяться для практического использования в технологическом процессе контроля качества фильтрующих элементов.

ПЕРЕЧЕНЬ ССЫЛОК

1. Можегов Н.А. Автоматические средства измерения объема, уровня и пористости материалов. - М.: Энергоатомиздат, 1990.-116с.
2. Кудряшов Н.А. Автомодельные задачи движения газа в пористой среде.- М.: Изд. МИФИ, 1984.

Рисунок 2 - Алгоритм работы программы