

Науковий журнал
Радіоелектроніка
Інформатика
Управління



2000
1 (3)

Видається двічі на рік

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

Засновник і видавник -
Запорізький державний технічний університет

Передплатний індекс - 22914

З М І С Т

ЗАПОРІЗЬКОМУ ДЕРЖАВНОМУ ТЕХНІЧНОМУ УНІВЕРСИТЕТУ - 100 РОКІВ	
<i>С. Б. Беліков</i>	4
РАДІОЕЛЕКТРОНІКА 6	
МЕТОД ЧИСЕЛЬНОЇ ПОБУДОВИ АНАЛІТИЧНИХ МОДЕЛЕЙ СКЛАДНИХ РАДІОЕЛЕКТРОННИХ СИСТЕМ НА ОСНОВІ БАГАТОКРИТЕРІАЛЬНОГО СТРУКТУРНОГО СИНТЕЗУ <i>Б. Л. Бутвін</i>	6
ОПРЕДЕЛЕНИЕ ФУНКЦИЙ ЧУВСТВИТЕЛЬНОСТИ В СТАТИКЕ <i>А. И. Вершина, Л. В. Кузьмина</i>	9
ЛОКАЛЬНОЕ ОСАЖДЕНИЕ ПЛЁНОК В ТЕХОЛОГИИ ИЗГОТОВЛЕНИЯ ИС <i>А. Н. Горбань, В. В. Кравчина</i>	12
ОПТИМИЗАЦИЯ ПЕРЕДАТОЧНЫХ ФУНКЦИЙ ЦИФРОВЫХ РЕГУЛЯТОРОВ СИСТЕМ ФАП ПРИ ПРОИЗВОЛЬНЫХ ВХОДНЫХ ВОЗДЕЙСТВИЯХ <i>В. И. Гостев, А. П. Хорев, О. А. Побийнеч</i>	16
АЛГОРИТМ КВАЗИСТАТИЧЕСКОГО МОДЕЛИРОВАНИЯ ПОЛОСКОВЫХ СТРУКТУР В	
МНОГОСЛОЙНОЙ АНИЗОТРОПНОЙ СРЕДЕ <i>Л. М. Карпуков</i>	18
ОЦЕНКА ДОСТОВЕРНОСТИ ДОПУСКОВОГО КОНТРОЛЯ <i>К. Н. Касьян</i>	24
ПРОВЕРКА ДИАГНОСТИРУЕМОСТИ СХЕМЫ ПУТЕМ ОПРЕДЕЛЕНИЯ РАНГА ТЕСТОВОЙ МАТРИЦЫ <i>Н. Н. Касьян</i>	26
МЕТОД СИНТЕЗА ТЕПЛОВЫХ И МЕХАНИЧЕСКИХ МОДЕЛЕЙ РАДИОЭЛЕКТРОННЫХ СРЕДСТВ ПРИ ПОМОЩИ ЭКСПЕРТНОЙ СИСТЕМЫ <i>Ю. Н. Кофанов, А. В. Сарафанов</i>	29
РАСЧЕТ КОЭФФИЦИЕНТА ЗАТУХАНИЯ В НАПРАВЛЯЮЩИХ СИСТЕМАХ С ФРАКТАЛЬНЫМИ СВОЙСТВАМИ ГРАНИЦ <i>В. М. Онуфриенко, П. А. Самолчев, Т. И. Слюсарова</i>	31

АВТОМАТИЗАЦІЯ ПРОЦЕСУ ТЕПЛООВОГО ПРОЕКТУВАННЯ МІКРОЕЛЕКТРОННИХ ПРИСТРОЇВ ЗА ДОПОМОГОЮ ПРОГРАМИ CIRCUIT MAKER 6 PRO <i>А. В. Пархоменко, Н. А. Коноваленко, Л. В. Поставець</i>	ВРАХУВАННЯ ПАРАМЕТРІВ ЕЛЕМЕНТІВ СХЕМИ В РЕЖИМІ ПІДВИЩЕННЯ НАПРУГИ ІМПУЛЬСНОГО РЕГУЛЯТОРА <i>В. Д. Флора</i>	35	38
ИНФОРМАТИКА		42	
NEUROINFORMATICS: WHAT ARE US, WHERE ARE WE GOING, HOW TO MEASURE OUR WAY? <i>А. N. Gorban</i>	СИНТЕЗ СТРУКТУРИРОВАННЫХ МОДЕЛЕЙ МНОГОСВЯЗНЫХ СИСТЕМ <i>Е. Н. Литвинов, А. С. Лихоузов, Т. А. Лихоузова</i>	42	89
NEURAL NETWORK MODELING OF DATA WITH GAPS <i>А. N. Gorban, А. А. Rossiev, D. C. Wunsch II</i>	НАСТРАИВАЕМЫЕ ЛОГИЧЕСКИЕ СЕТИ НА НЕЙРОПОДОБНЫХ ЭЛЕМЕНТАХ <i>В. Н. Лопин</i>	47	93
ПРИНЦИП ВАРЬИРОВАНИЯ ДАННЫХ В ПРИКЛАДНЫХ ЗАДАЧАХ ИДЕНТИФИКАЦИИ <i>А.Е. Архипов, С.А. Архипова</i>	ИСПОЛЬЗОВАНИЕ JAVA-ТЕХНОЛОГИЙ В ИНТЕРНЕТ-ПРИЛОЖЕНИЯХ <i>С. А. Онищенко</i>	56	96
ПРОГНОЗИРУЮЩАЯ НЕЙРОННАЯ СЕТЬ И АЛГОРИТМЫ ЕЕ ОБУЧЕНИЯ <i>Е. В. Бодянский, С. В. Попов</i>	БЫСТРОЕ ДИФФЕРЕНЦИРОВАНИЕ СЛОЖНЫХ ФУНКЦИЙ МНОГИХ ПЕРЕМЕННЫХ <i>М. Ю. Сенашова</i>	60	101
МЕТОД ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ГРАНИЧНОГО ВЫБОРОЧНОГО КОНТРОЛЯ <i>А. Ю. Долгов</i>	АНАЛИЗ И СИНТЕЗ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ <i>С. Н. Сердюк, А. А. Поздняков</i>	64	106
АНАЛИЗ ПАРАМЕТРИЧЕСКОЙ ПЛАСТИЧНОСТИ МНОГОСЛОЙНЫХ НЕЙРОННЫХ СЕТЕЙ <i>А. Ю. Дорогов</i>	РОЗРОБКА ТА ДОСЛІДЖЕННЯ СИСТЕМИ ІДЕНТИФІКАЦІЇ КОРИСТУВАЧА НА ОСНОВІ НЕЙРОННОЇ МЕРЕЖІ <i>В. В. Сокирук, Є. В. Яремчук, В. І. Месюра</i>	66	112
ЭВРИСТИЧЕСКИЙ АЛГОРИТМ КЛАССИФИКАЦИИ И ЕГО НЕЙРОСЕТЕВАЯ ИНТЕРПРЕТАЦИЯ <i>В. И. Дубровин, С. А. Субботин</i>	ПОДХОД К МОДЕЛИРОВАНИЮ ЭЛЕМЕНТОВ МАГИСТРАЛЬНОГО ГАЗОПРОВОДА НА ОСНОВЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ <i>С. А. Терехов, Т. А. Мухамадиева, Н. Н. Федорова, Е. В. Диянкова, С. А. Диянкова, А. В. Квичанский, Р. Ф. Мухамадиев</i>	72	116
ВИЗУАЛИЗАЦИЯ ПРОИЗВОЛЬНЫХ ДАННЫХ МЕТОДОМ УПРУГИХ КАРТ <i>А. Ю. Зиновьев, А. А. Питенко</i>	НЕЙРОКОМПЬЮТЕР "ЭМБРИОН" КАК ГЕНЕРАТОР И ПРЕОБРАЗОВАТЕЛЬ СМЫСЛОВ В ВЕРОЯТНОСТНОЙ МОДЕЛИ ЛИЧНОСТИ <i>В. Д. Цыганков</i>	76	123
МНОГОСТУПЕНЧАТЫЕ ИНТЕРВАЛЫ ПРИ РАСЧЕТЕ ЭКСПЛУАТАЦИОННЫХ ДОПУСКОВ <i>В. Н. Кришук, Г. Н. Шило, Н. П. Гапоненко</i>		85	
УПРАВЛІННЯ		132	
НЕЛИНЕЙНЫЕ ОБОБЩЕНИЯ ДИНАМИЧЕСКИХ УРАВНЕНИЙ ЛЕОНТЬЕВА ДЛЯ ДВУХСЕКТОРНОЙ ЭКОНОМИКИ <i>Б. Ю. Базылев, Е. Р. Саенко, А. Д. Шамровский</i>	ЭНЕРГОСБЕРЕГАЮЩЕЕ УПРАВЛЕНИЕ ДВИГАТЕЛЕМ ПОСТОЯННОГО ТОКА <i>С. Г. Деев, В. И. Левыкина</i>	132	139
АНАЛИЗ ЭНЕРГЕТИКИ СОВРЕМЕННЫХ СИСТЕМ ЭЛЕКТРОПИТАНИЯ НА ОСНОВЕ УПРАВЛЯЕМЫХ ВЕНТИЛЬНЫХ ПРЕОБРАЗОВАТЕЛЕЙ <i>М. Н. Горбачев</i>	СИСТЕМА УПРАВЛЕНИЯ КАЧЕСТВОМ ПРОМЫШЛЕННОГО ПРЕДПРИЯТИЯ <i>Ю. Н. Дорошенко, В. И. Дубровин</i>	136	142

ПРОЕКТИРОВАНИЕ КАЧЕСТВА: ДОПУСКОВЫЙ АНАЛИЗ <i>В. И. Дубровин, Г. В. Табунчик 149</i>	РОБАСТНОЕ УПРАВЛЕНИЕ ЭЛЕКТРОПРИВОДОМ С ВЕНТИЛЬНЫМ ДВИГАТЕЛЕМ <i>Е. М. Потапенко, Д. В. Корельский, Е. В. Васильева . 161</i>
УНИФИЦИРОВАННЫЙ МЕТОД РАСЧЕТА ХАРАКТЕРИСТИК РЕАКЦИИ ОБЪЕКТА НА СЛУЧАЙНЫЕ ВОЗДЕЙСТВИЯ МАЛОЙ ИНТЕНСИВНОСТИ <i>Э. Г. Куренный, Ю. И. Четкасов, Л. В. Черникова . . 152</i>	ТЕРМИНАЛЬНОЕ УПРАВЛЕНИЕ СИСТЕМАМИ С ДИФФЕРЕНЦИАЛЬНО-АЛГЕБРАИЧЕСКИМИ ОГРАНИЧЕНИЯМИ МЕТОДОМ КОНЕЧНЫХ СОСТОЯНИЙ <i>Е. А. Шушлятин 166</i>
ФОРМУВАННЯ КЕРУЮЧИХ ВПЛИВІВ ЕЛЕКТРОМЕХАНІЧНИХ СИСТЕМ НА ОСНОВІ ПРИНЦИПІВ НЕЧІТКОЇ ЛОГІКИ <i>А. О. Лозинський 156</i>	АВТОРЫ НОМЕРА 173

I. РАДІОЕЛЕКТРОНІКА

УДК 355. 422:001. 57. 001.63

МЕТОД ЧИСЕЛЬНОЇ ПОБУДОВИ АНАЛІТИЧНИХ МОДЕЛЕЙ СКЛАДНИХ РАДІОЕЛЕКТРОННИХ СИСТЕМ НА ОСНОВІ БАГАТОКРИТЕРІАЛЬНОГО СТРУКТУРНОГО СИНТЕЗУ

Б. Л. Бутвін

Рассматриваются пути и результаты решения задачи построения аналитических моделей сложных радиоэлектронных систем с использованием численного многокритериального подхода к оценке адекватности экспериментальных и расчетных данных.

Розглядаються шляхи та результати вирішення задачі побудови аналітичних моделей складних радіоелектронних систем із використанням численного многокритеріального підходу щодо оцінки адекватності експериментальних та розрахункових даних.

The ways and the results of decision of the task of construction of analytical models of complex radioelectronic systems with using of numerical multicriteria approach to appraisal of adequacy of experimental calculated dates are considered.

Сучасна методологія оптимізації складних радіоелектронних систем (СРС) заснована на знанні їх математичних моделей. Побудова аналітичної моделі СРС, у якій всі критерії якості одержані з прийнятною точністю, являє собою досить складну математичну багатокритеріальну задачу. При цьому деякі критерії можуть розраховуватися досить надійно, а інші з помилкою, яка може досягати десятків відсотків.

Саме така ситуація найбільш типова і правдоподібна. Тому край важливо знати кількісну оцінку адекватності аналітичної моделі експериментальним даним, щоб нікого не вводили в оману успіхами розв'язання задачі оптимізації СРС.

Вирішення питання наскільки повно аналітична модель адекватна експериментальним даним може дати розв'язання задачі багатокритеріального синтезу структури аналітичної моделі. Параметричний синтез [1] розв'язує досить важливу, але тільки одну задачу вибору параметрів вже обраної структури моделі, відповідно до одного критерію (наприклад, мінімуму квадрата відхилення експериментальних даних від тих, що розраховуються за допомогою, наприклад, поліноміальної моделі).

Для багатьох складних радіоелектронних систем, збіг аналітичної моделі з експериментальними даними за одним з показників не дає підстави для довіри до моделі за іншими показниками [2]. Розглянемо основні особливості багатокритеріальної постановки задачі синтезу

аналітичних моделей СРС.

Такий багатокритеріальний підхід необхідний, насамперед, для об'єктивного, багатокритеріального визначення ступеня відповідності аналітичної моделі реальній СРС. І те, що сьогодні не розв'язана задача структурного синтезу аналітичних моделей складних систем говорить, у першу чергу, про труднощі формалізованої постановки даної задачі.

У загальному випадку аналітична модель СРС характеризується двома видами параметрів: структурними і внутрішньоструктурними параметрами.

Позначимо через $G_i = |R_s, A_p|$ вектор параметрів, який цілком описує аналітичну модель СРС у K_m -му класі аналітичних моделей для i -го варіанта моделі, де $R_{si} = |r1_{si}, r2_{si}, \dots, rr_{si}|$ - вектор структурних параметрів i -го варіанта аналітичної моделі; $A_{psi} = |\alpha_{1psi}, \alpha_{2psi}, \dots, \alpha_{rsi}|$ - вектор внутрішньоструктурних параметрів i -го варіанта аналітичної моделі для s -ї структури аналітичної моделі СРС, які розраховуються за допомогою метода найменших квадратів [3,4].

Нехай $F_k(G_i)$ - експериментальне значення k -го критерію ефективності застосування СРС, значення якого вимірені експериментально або за допомогою математичного моделювання, а $F_{ak}(G_i)$ - значення k -го критерію застосування СРС, значення якого розраховані за допомогою i -го варіанта аналітичної моделі.

Тоді $|S_1[F_1(G_i) - F_{a1}(G_i)], S_2[F_2(G_i) - F_{a2}(G_i)], \dots, S_k[F_k(G_i) - F_{ak}(G_i)]|$ - векторний критерій адекватності критеріїв застосування СРС, одержаних експериментально або розрахунковим шляхом на i -му варіанті аналітичної моделі, розрахований за допомогою деякої функції $S_k(\dots)$. Функціями $S_k(\dots)$ можуть виступати як статистичні, так і евристичні функції, що характеризують ступінь збігу експериментальних і розрахункових критеріїв ефективності [4].

Математична постановка задачі багатокритеріального

структурного синтезу аналітичної моделі СРС має такий вигляд:

знайти вектор параметрів $G = |R_s, A_p|$, для якого виконуються умови

$$R_s \text{ opt} = \arg\{extr\{\psi[fS_k[F_k(G_i) - F_{ak}(G_i)]]\}\}; \quad (1)$$

$$A_p \text{ opt} = \arg\left\{ \sum_{t_1=0}^{N-1} \left[\sum_{r=0}^k F_k(G_{t_1}) - F_k^d(G_{t_1}) \right]^2 \right\} \Rightarrow \min; \quad (2)$$

$$S_k[F_k(G_i) - F_{ak}(G_i)] < \Phi^*_u, \quad u = 1, 2, \dots, U; \quad (3)$$

$$\alpha_s^* \alpha_{si} < \alpha_s^{**}; \quad (4)$$

$$r^*_a < R_{si} < r^{**}_a, \quad (5)$$

де α_s^* , α_s^{**} - обмеження на структурні параметри аналітичної моделі; Φ^*_u - припустимі критеріальні обмеження, що призначаються в процесі синтезу аналітичної моделі СРС і визначають адекватність експериментальних даних і результатів, одержаних за допомогою аналітичної моделі; r^*_a , r^{**}_a - функціональні обмеження на структурні параметри; ψ - суб'єктивна функція вибору оптимального структурного варіанту аналітичної моделі.

Аналітичне рішення задачі у постановці (1)-(5) має значні труднощі. Більш перспективним є чисельний метод рішення цієї задачі, який використовує систематичне зондування простору структурних параметрів за допомогою псевдовипадкових послідовностей чисел [2]. Множина структурних варіантів аналітичної моделі, що задовольняє умови (3) - (5), і складає припустиму множину структурних варіантів D_Δ .

Відповідно до принципів багатокритеріальної оптимізації та частина множини структурних параметрів $R_{si} = |r_{1si}, r_{2si}, \dots, r_{rsi}|$, для якої не можуть бути поліпшені критерії адекватності без погіршення хоча б одного критерію, називається областю паретооптимальних структурних векторів аналітичної моделі D_n .

Істотна відмінність задачі структурного синтезу від задач оптимізації полягає в тому, що при постановці задачі структурного синтезу найчастіше немає інформації про межі варіації для багатьох параметрів аналітичної моделі. Тому при невдалому завданні меж (навіть при наявності найкращої моделі) може знадобитися проведення складного експерименту на ПЕОМ, однак результати синтезу все одно можуть виявитися незадовільними або неповними. Якщо розв'язання задачі багатокритеріального синтезу визначено коректно, то в результаті повинна бути одержана непушта множина D_n .

У протилежному разі D_n - порожня множина. Це може бути в тому разі, коли синтезована модель не відбиває основних властивостей СРС, тобто за одним або кількома основними критеріями адекватності значення вектора $S_k(G_i)$ неприпустимо більше вектора Φ^*_u , що дозволяє відкинути результат синтезу як незадовільний.

Тому найважливішим завданням векторного структурного синтезу є визначення моделей СРС, для яких множина D_n непушта, а також вибір такого варіанту синтезу, адекватність якого в деякому (суб'єктивному) змісті найкраще відповідає експериментальним даним. Вектор параметрів моделі $G_{i \text{ opt}} = |R_s, A_p| \in D_n$, що належить паретооптимальній множині і визначений за допомогою деякого вирішального (суб'єктивного) правила ψ як найкращий, визначає оптимальний варіант структурних параметрів аналітичної моделі СРС.

Слід зазначити, що неоднозначність синтезу параметрів моделі варто розглядати, як плату за бажання синтезувати аналітичну модель, якість якої оцінюється різноманітними критеріями, а також за неповноту уявлення СРС у вигляді спрощеної аналітичної моделі і неповноту планування експерименту з реальним СРС (або з його імітаційною математичною моделлю). Загальна структурна схема вирішення задачі чисельної побудови аналітичних моделей складних радіоелектронних засобів на основі багатокритеріального структурного синтезу наведена на рис.1. яка містить у собі:

Блок 1. Реалізує генерацію псевдовипадкової послідовності чисел. Найбільш привабливою для вирішення цієї задачі є $ЛП_t$ - послідовність чисел Соболя [2], що генерує послідовність точок, рівномірно розподілених в одиничному n -мірному кубі, де n - число структурних параметрів аналітичної моделі.

Блок 2. Організує перерахунок псевдовипадкової послідовності чисел у простір структурних параметрів аналітичної моделі.

Блок 3. Реалізує розрахунок вектора параметрів α_{qij} за допомогою метода найменших квадратів [3].

Блок 4. Здійснює розрахунок вектора адекватності $\Delta f_u(A_i) = |s_{1i}, s_{2i}, \dots, s_{ui}|$ для i -го структурного варіанту побудови аналітичної моделі СРС. У якості s_{ui} -х показників адекватності можуть використовуватися критерій Фішера, коефіцієнт детермінації, максимальна помилка, середньоквадратична помилка та ін. [4].

Блок 5. Реалізує перевірку належності вектора структурних параметрів множини Парето $A_i = |q_{si}, \alpha_{pi}| \in D_n$.

Блок 6. Реалізує вибір раціональної структури моделі за допомогою суб'єктивного критерію ухвалення $extr(\bullet)$.

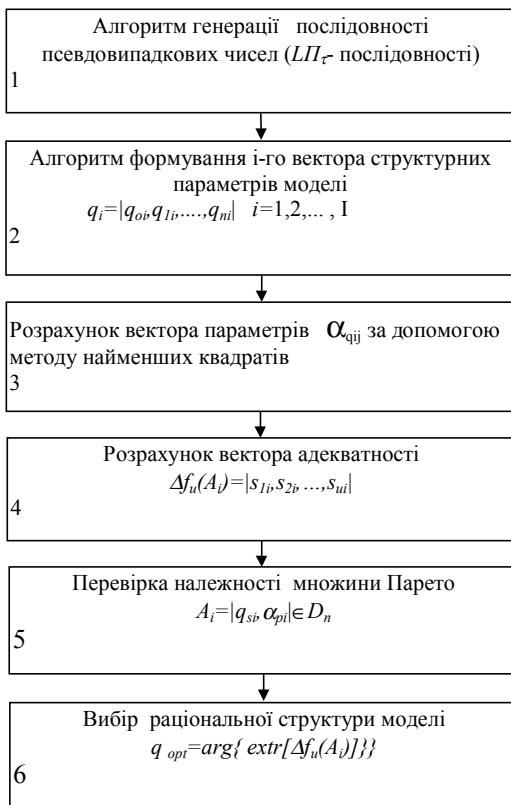


Рисунок 1 - Загальна структурна схема вирішення задачі чисельної побудови аналітичних моделей складних радіоелектронних систем на основі багатокритеріального структурного синтезу

У якості прикладу на рис.2 наведено результати генерації 2048 структурних варіантів аналітичної моделі у площині $\sigma_{скн} - \sigma_{max}$. Найближчим аналогом чисельного методу побудови аналітичних моделей є метод регресійного аналізу. Результати відносного порівняння їх ефективності по показникам σ_{max} - максимальної помилки та $\sigma_{скн}$ - середньоквадратичної помилки наведені на рис.3, аналіз яких дозволяє зробити висновок, що чисельний метод побудови аналітичних моделей складних радіоелектронних систем забезпечує у 2-5 разів менші помилки ніж метод регресійного аналізу.

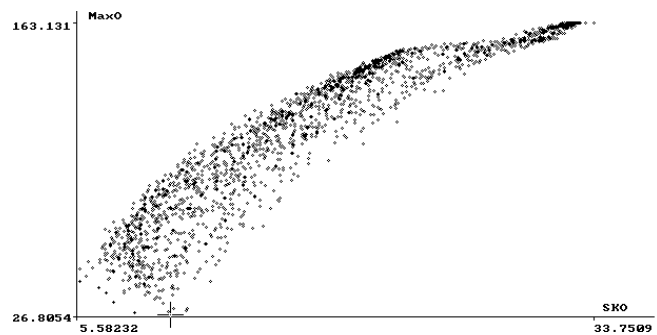


Рисунок 2 - Результати генерації 2048 структурних варіантів в площині $\sigma_{скн} - \sigma_{max}$

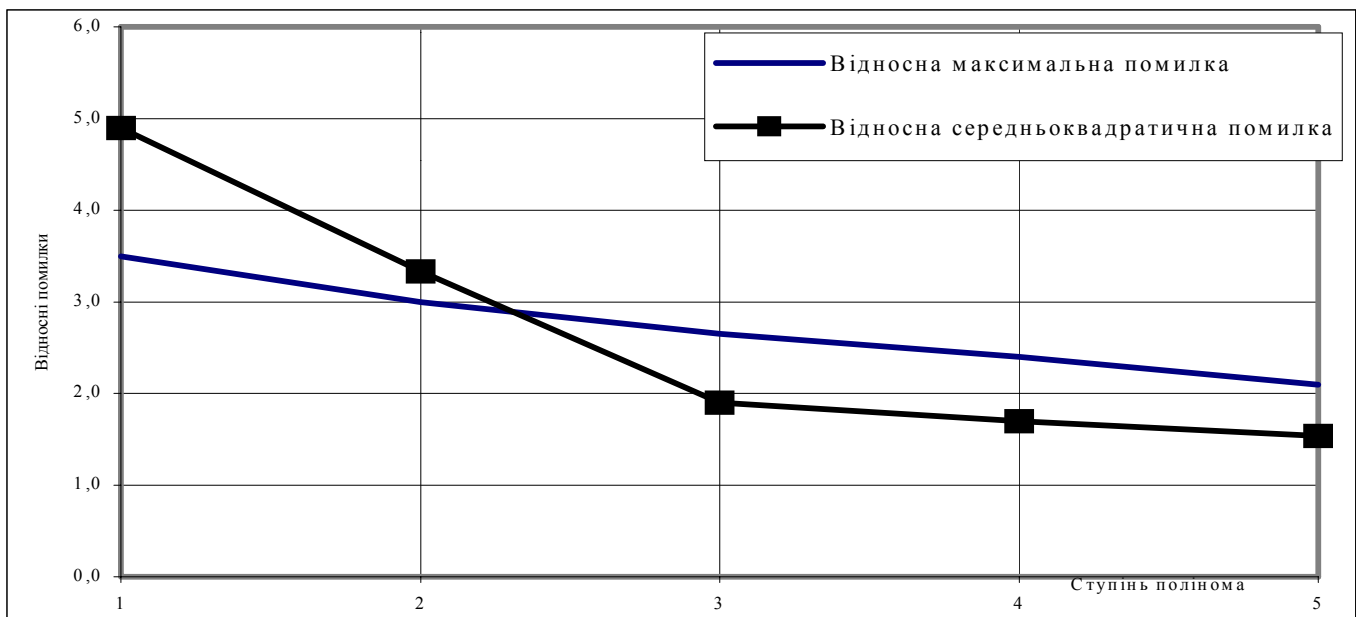


Рисунок 3 - Результати відносного порівняння ефективності чисельного методу побудови аналітичної моделі СРС та метода регресійного аналізу по показниках максимальної та середньоквадратичної помилок

ПЕРЕЛІК ПОСИЛАНЬ

1. Черноруцкий И.Г. Оптимальный параметрический синтез.- Л.: Энергоатомиздат, 1987.-128с.
 2. Соболев И.М., Статников Р.Б. Выбор оптимальных параметров в задачах со многими критериями. - М.: Наука, 1981.-108с.

3. Жданюк Б.Ф. Основы статистической обработки траекторных измерений.-М.:Сов. Радио,1978.-384с.
 4. Крамер Г. Математические методы статистики./Пер. с англ. -М.:Мир,1975.-648с.

Надійшла 17.02.2000

УДК 621.38

ОПРЕДЕЛЕНИЕ ФУНКЦИЙ ЧУВСТВИТЕЛЬНОСТИ В СТАТИКЕ

А. И. Вершина, Л. В. Кузьмина

Рассматриваются пути повышения эффективности определения функций чувствительности при моделировании радиоэлектронных схем в статике. Учет закономерностей, которые имеют место в описывающих их уравнениях, позволяют предложить новые алгоритмы решения этой задачи.

Розглядаються шляхи підвищення ефективності визначення функції чутливості при моделюванні радіоелектронних мереж у статистиці. Врахування закономірностей, які мають місце в рівняннях, що їх описують, дозволяють запропонувати нові алгоритми рішення цієї задачі.

Presented are the ways for increasing the efficiency of determining functions of sensitivity while simulating radio-electronic networks in static. Consideration of relationships occurring in equations describing the latter permits to propose new algorithms for solving this problem.

ВВЕДЕНИЕ

Одной из важнейших задач моделирования радиоэлектронных схем является определение функций чувствительности, под которыми понимаются значения частных производных выходных переменных по параметрам элементов, образующих эту схему [1-3]. Функции чувствительности лежат в основе различных градиентных методов оптимизации, могут использоваться для оценки статистических характеристик и для решения задач синтеза радиоэлектронных схем. При моделировании относительно сложных радиоэлектронных схем определение функций чувствительности представляет собой довольно трудоемкую задачу, связанную с многократным решением систем уравнений высокого порядка. В связи с этим поиск методов повышения эффективности нахождения функций чувствительности является актуальной задачей.

ПОСТАНОВКА ЗАДАЧИ

В общем случае радиоэлектронная схема описывается системой нелинейных алгебро-дифференциальных уравнений

$$F\left(\frac{\partial x}{\partial t}, x, p\right) = 0, \tag{1}$$

где x и $\frac{\partial x}{\partial t}$ - значения неизвестных переменных и их производных по времени;

p - параметры элементов радиоэлектронной схемы.

Для статических режимов имеем $\frac{\partial x}{\partial t} = 0$ и уравнение имеет вид

$$F(x, p) = 0. \tag{2}$$

Производная по параметрам запишется следующим образом:

$$\frac{\partial F}{\partial x} \cdot \frac{\partial x}{\partial p} + \frac{\partial F}{\partial p} = 0, \tag{3}$$

отсюда

$$\frac{\partial F}{\partial p} = -\left[\frac{\partial F}{\partial x}\right]^{-1} \cdot \frac{\partial F}{\partial p}. \tag{4}$$

Набор выходных переменных можно представить как матрицу-столбец

$$\varphi(x, p), \tag{5}$$

для которой значение функции чувствительности запишем в виде

$$\frac{d}{dp}\varphi(x, p) = \frac{\partial \varphi}{\partial x} \cdot \frac{\partial x}{\partial p} + \frac{\partial \varphi}{\partial p}. \tag{6}$$

После подстановки выражения для $\frac{\partial x}{\partial p}$, получим

$$\frac{d}{dp}\varphi(x, p) = \frac{\partial \varphi}{\partial x} \cdot \left[\frac{\partial F}{\partial x}\right]^{-1} \cdot \frac{\partial F}{\partial p} + \frac{\partial \varphi}{\partial p}, \tag{7}$$

где $\frac{d}{dp}\varphi(x, p)$ - матрица производных выходных переменных по варьируемому параметрам;

$\frac{\partial \varphi}{\partial x}$ - матрица частных производных выходных переменных по неизвестным;

$\frac{\partial F}{\partial x}$ - матрица Якоби;

$\frac{\partial F}{\partial p}$ - матрица частных производных уравнений по варьируемому параметрам;

$\frac{\partial \varphi}{\partial p}$ - матрица частных производных выходных переменных по варьируемому параметрам.

Выражение (7) лежит в основе двух основных подходов к определению чувствительности: метода моделей и метода присоединенных схем. Оба этих метода могут рассматриваться с позиции последовательности выполнения операций данного выражения. Так, если выполнять сначала операции умножения

матрицы $\frac{\partial \varphi}{\partial x}$ на $\left[\frac{\partial F}{\partial x}\right]^{-1}$, то такой подход может быть интерпретирован как метод присоединенных схем. В случае, если выполнить сначала операции умножения матрицы $\left[\frac{\partial F}{\partial x}\right]^{-1}$ на $\frac{\partial F}{\partial p}$, то такой подход интерпретируется как метод моделей. Целесообразность использования того или иного подхода определяется объемами вычислений, которые зависят от количества выходных переменных и количества варьируемых параметров. Эта проблема имеет место, когда моделируемая схема достаточно сложна и описывается системой уравнений высокого порядка, что не позволяет непосредственно находить обратную матрицу $\left[\frac{\partial F}{\partial x}\right]^{-1}$, а все необходимые вычисления производить путем решения систем линейных уравнений с учетом разреженности матрицы $\frac{\partial F}{\partial x}$ [4]. На этапе решения матрица $\frac{\partial F}{\partial x}$ подвергается LU-разложению. При этом

в методе моделей последовательно решаются две системы уравнений с полученными в результате разложения треугольными матрицами. В методе присоединенных схем матрица, описывающая присоединенную схему, может рассматриваться как транспонированная по отношению к исходной матрице $\frac{\partial F}{\partial x}$ и также решаются системы уравнений с ранее полученными, но транспонированными треугольными матрицами.

Следует отметить, что LU-разложение осуществляется на каждой итерации решения системы нелинейных уравнений (2) и для последующего нахождения функций чувствительности используется это разложение на последней итерации. Объемы операций с транспонированными и исходными треугольными матрицами примерно одинаковы. Поэтому целесообразность того

или иного подхода к определению функций чувствительности в основном зависит от количества выходных переменных и от количества варьируемых параметров.

Если обозначить через m - количество выходных переменных, а через l - число варьируемых параметров, то при $m > l$ целесообразно применять метод моделей. В случае, когда $m < l$, применяется метод присоединенных схем. При этом, как m , так и l должны быть меньше количества неизвестных n иначе объем вычислений превысит объем операций для нахождения всей обратной матрицы, после нахождения которой трудоемкость определения функций чувствительности любых выходных переменных по любым варьируемым параметрам незначительна.

Предлагается решение задачи дальнейшего повышения эффективности определения функций чувствительности путем учета закономерностей, которые являются следствием представления моделей радиоэлектронных схем взаимосвязью двухполюсных элементов. Эти закономерности заключены в матрицах $\frac{\partial F}{\partial p}$ и $\frac{\partial \varphi}{\partial x}$.

МЕТОДЫ И АЛГОРИТМЫ РЕШЕНИЯ

Рассмотрим матрицу $\frac{\partial \varphi}{\partial x}$. Как правило, для таких уравнений функция $\varphi(x)$ может быть представлена в виде $\varphi_k(a)$ где $a = (x_i - x_j)$. Отсюда $\frac{\partial \varphi_k}{\partial x_i} = \frac{\partial \varphi_k}{\partial x_j} = \frac{\partial \varphi_k}{\partial a}$.

Тогда матрица $\frac{\partial \varphi}{\partial x}$ может быть записана в виде произведения диагональной матрицы $\frac{\partial \varphi}{\partial a}$ на прямоугольную матрицу M_a , размерностью $m \times n$, элементы которой принимают значения 0, 1, -1:

$$\frac{\partial \varphi}{\partial x} = \frac{\partial \varphi}{\partial a} \cdot M_a = \begin{bmatrix} \frac{\partial \varphi_1}{\partial a_1} & 0 & \dots & 0 \\ 0 & \frac{\partial \varphi_2}{\partial a_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \frac{\partial \varphi}{\partial a_l} \end{bmatrix} \cdot \begin{bmatrix} \dots & 1 & \dots & -1 & \dots & \dots \\ \dots & \dots & 1 & \dots & -1 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & -1 & \dots & \dots & 1 & \dots \end{bmatrix}. \quad (8)$$

Легко убедиться, что выражение не изменится, если его умножить справа на диагональную матрицу E_a , размерностью $n \times n$, содержащую единицы в тех элементах, для которых столбцы матрицы M_a имеют ненуле-

вые элементы, и нули в остальных элементах:

$$\frac{\partial \varphi}{\partial x} = \frac{\partial \varphi}{\partial a} \cdot M_a = \frac{\partial \varphi}{\partial a} \cdot M_a \cdot E_a. \quad (9)$$

Выражение (7) принимает вид

$$\frac{\partial}{\partial p} \varphi(x, p) = \frac{\partial \varphi}{\partial a} \cdot M_a \cdot E_a \cdot \left[\frac{\partial F}{\partial x} \right]^{-1} \cdot \frac{\partial F}{\partial p} + \frac{\partial \varphi}{\partial p}. \quad (10)$$

Рассмотрим свойство произведения $E_a \cdot \left[\frac{\partial F}{\partial x} \right]^{-1}$. Его особенностью является то, что произведение строк, содержащих ненулевые элементы, соответствует нахождению соответствующих строк обратной матрицы. Количество возможных аргументов равно числу сочетаний $s = C_K^2$, где K - количество ненулевых элементов матрицы E_a . При этом может оказаться, что K значительно меньше количества выходных переменных m . Это приводит к существенному выигрышу.

Рассмотрим особенности матрицы $\frac{\partial F}{\partial p}$, связанные с тем, что функция описывает электрическую схему. Представление модели электрической схемы в виде взаимосвязи двухполюсников, содержащих зависимые от одного аргумента, источники токов и напряжений, индуктивности, емкости и проводимости приводит к тому, что варьируемый параметр входит в два уравнения, причем в каждом уравнении двухполюсник в общем случае представлен в виде произведения некоторой нелинейной функции $f(x, p)$ на разность аргументов $(x_i - x_j)$ либо в виде функции $g(p)$. Произведение соответствует, как правило, пассивным элементам схемы, а функция $g(p)$ является составляющей правой части уравнений, ей соответствуют независимые источники тока и напряжения. Матрица $\frac{\partial F}{\partial p}$ может быть записана в виде произведения прямоугольной матрицы M_p на диагональную матрицу $\left[\frac{\partial f}{\partial p} \right]$:

$$\frac{\partial F}{\partial p} = M_p \cdot \left[\frac{\partial f}{\partial p} \right] = \begin{bmatrix} \dots & \dots & \dots & \dots & \dots \\ \dots & (x_i - x_j) & \dots & \dots & \dots \\ \dots & \dots & \dots & 1 & \dots \\ \dots & -(x_i - x_j) & \dots & \dots & \dots \\ \dots & \dots & \dots & -1 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial f_1}{\partial p_1} & \dots & \dots & \dots & \dots \\ \dots & \frac{\partial f_2}{\partial p_2} & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \frac{\partial g}{\partial p} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}. \quad (11)$$

Нетрудно видеть, что умножение слева на диагональную матрицу E_p , размерностью $n \times n$, содержащей в диагонали единицы, для которых в матрице M_p , соответствует ненулевая строка, а в остальных элементах диагонали будут нули, не изменит результата:

$$\frac{\partial F}{\partial p} = E_p \cdot M_p \cdot \left[\frac{\partial f}{\partial p} \right]. \quad (12)$$

Рассмотрим свойства произведения $\left[\frac{\partial F}{\partial x} \right]^{-1} \cdot E_p$. Его особенностью является то, что произведение столбцов, содержащих ненулевые элементы, соответствует нахождению столбцов обратной матрицы. Количество возможных варьируемых параметров определяется числом сочетаний $s = C_K^2$, где K - количество ненулевых элементов матрицы E_p . При этом может оказаться, что K значительно меньше количества варьируемых параметров l . Это также может привести к существенному выигрышу.

Выражение (7) может быть представлено следующим образом:

$$\frac{\partial}{\partial p} \varphi(x, p) = - \frac{\partial \varphi}{\partial a} \cdot \left[\frac{\partial F}{\partial x} \right]^{-1} \cdot E_p \cdot M_p \cdot \left[\frac{\partial f}{\partial p} \right] + \frac{\partial \varphi}{\partial p}, \quad (13)$$

а с учетом (10) -

$$\frac{\partial}{\partial p} \varphi(x, p) = - \frac{\partial \varphi}{\partial a} \cdot M_a \cdot E_a \cdot \left[\frac{\partial F}{\partial x} \right]^{-1} \cdot E_p \cdot M_p \cdot \left[\frac{\partial f}{\partial p} \right] + \frac{\partial \varphi}{\partial p}. \quad (14)$$

С точки зрения объема выполняемых операций, наиболее трудоемкой частью данного выражения является произведение $E_a \cdot \left[\frac{\partial F}{\partial x} \right]^{-1} \cdot E_p$. Оно связано с многократным решением системы линейных уравнений. При этом, как уже отмечалось выше, LU-разложение выполняется при решении системы нелинейных уравнений (1) и соответствует разложению на последней итерации метода Ньютона. В то же время введение матриц E_a и E_p позволяет повысить эффективность определения коэффициентов чувствительности за счет учета особенностей уравнений, описывающих электронные схемы, модели которых представлены двухполюсными элементами.

РЕЗУЛЬТАТЫ

Пусть N_a и N_p количество ненулевых элементов матриц E_a и E_p соответственно. Тогда условия эффективного расчета функций чувствительности в статике можно представить следующим алгоритмом:

при $\min\{N_a, m\} > \min\{N_p, l\}$ и $N_a > m$, то расчет нужно производить, используя метод моделей;

при $\min\{N_a, m\} > \min\{N_p, l\}$ и $N_a < m$ - использовать (13);

при $\min\{N_a, m\} < \min\{N_p, l\}$ и $N_a > 1$ - использовать метод присоединенных схем;

при $\min\{N_a, m\} < \min\{N_p, l\}$ и $N_a < 1$ - использовать (10).

ВЫВОДЫ

Моделирование радиоэлектронных схем, представленных в виде взаимосвязей двухполюсных элементов,

находит свое отражение в свойствах систем уравнений, которые описывают их работу. Получены выражения и предложены новые алгоритмы расчета функций чувствительности, которые учитывают эти свойства и дают возможность получить дополнительный выигрыш.

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. - 195 с.
4. Тьюарсон Р. Разреженные матрицы. - М.: Мир, 1977. - 185с.

Надійшла 15.03.2000
Після доробки 22.03.2000

УДК621.3.049.77

ЛОКАЛЬНОЕ ОСАЖДЕНИЕ ПЛЁНОК В ТЕХНОЛОГИИ ИЗГОТОВЛЕНИЯ ИС

А. Н. Горбань, В. В. Кравчина

Электрическое смещение, режим и конструкция установки подбираются таким образом, что происходит селективное плазмохимическое травление кремния и осаждение продуктов травления в области над SiO₂. Последующее их плазмохимическое взаимодействие с кислородом приводит к образованию локального плазмохимического SiO₂ на термическом окисле кремния. Исследования структуры локальных плёнок SiO₂ показали, что полученные плазменные плёнки имеют поликристаллическую структуру из монокристаллов α-кристаллита и α-кварца.

Електричну напругу, режим та конструкцію установки налагоджують таким чином, що відбувається селективне плазмохімічне травлення Si відносно SiO₂ і осадження продуктів травлення на півці високотемпературного окису кремнію. Наступна плазмохімічна взаємодія осаджених продуктів травлення з киснем спричиняє утворення локального SiO₂. Дослідження побудови локальних плівок показали, що утворенні плівки мають полікристалічну структуру з монокристалів α-кристаліта і α-кварцу.

The characterisation of plasma etch process selected thus what silicon was etched and plasma products make set on the areas over silica. These plasma products are oxidised in plasma oxygen. There are created local films self-aligned with silica. The structure of creating local films is polycrystalline silicon oxide.

Важность вопросов разработки высокотемпературных и низкотемпературных методов формирования диэлектрических плёнок для полупроводниковых приборов общеизвестна [1,2]. Диэлектрические слои, полученные методами плазмохимического осаждения, нашли применение для пассивации и межуровневой изоляции интегральных схем. А локальное формирование слоев, самосовмещаемое с определённым топологическим ри-

сунком, например, со сформированными на поверхности подложки островками SiO₂, всегда рассматривается с интересом и часто предпочтительнее других технологий получения диэлектрических плёнок для многих применений.

На плазмохимическое, ионное травление материалов оказывает влияние, как структура материалов, так и фазовое состояние. В работе [4] исследовалось влияние фазового состояния на вторичную эмиссию и другие эффекты, сопровождающие ионную бомбардировку материала.

Особенностям травления структур кремния и окисла кремния посвящено достаточное количество работ [4]. Энергия активации травления значительно меньше для кремния чем для его окислов. Поэтому травление Si атомами и радикалами фтора может протекать спонтанно, а для травления SiO₂ необходима стимуляция ионной бомбардировкой.

Известно использование локального формирования плёнок SiO₂ на поверхности фоторезиста для формирования маскирующих покрытий с элементами субмикронных размеров [5]. В этом случае производится экспонирование поверхностных слоёв фоторезиста ультрафиолетовым облучением с последующими прививкой к экспонированным областям кремнийорганических соединений из газовой фазы и проявлением не привитых областей. Проявление проводится в плазме кислорода с одновременным превращением кремнийорганических соединений на привитых областях в двуокись кремния.

В данной работе определяются условия для протек-

кания плазмохимических реакций локального травления резиста и кремния и плазмохимических реакций локального роста диэлектрических плёнок с использованием продуктов травления Si. Проводятся исследования состава и структуры формируемых плёнок, необходимые для описания механизмов протекания процессов.

МЕТОДИКА ЭКСПЕРИМЕНТА

Процессы плазмохимического травления проводились в реакторе диодного типа. Для реализации осаждения в процессе травления первоначально находились конструктивные особенности реактора и режимы, позволяющие травить кремний с максимальной скоростью, а окисел кремния с минимальной. Процессы селективного плазменного травления пленок поли-Si*, Si₃N₄ относительно SiO₂ исследовались в газовой смеси фреонов "Хладон-14"(CF₄) и "Хладон-113" (CCl₃FCF₃). Травление фоторезиста проводилось в режиме селективного травления пленок поли-Si*, Si₃N₄ относительно SiO₂. Особенностью такого травления является травление резиста в области кремния или поликремния и замедление травления резиста с последующим осаждением продуктов травления кремния в области над SiO₂.

Оже-спектры осаждаемых плёнок исследовались на спектрометре "Шхуна", с компенсацией зарядов, образующихся в диэлектрических плёнках, при помощи пушки нейтрализации. Кроме Оже-спектров ПХЛМ пленки снимались и масс-спектры в растре первичного пучка Ag⁺ 200 × 200 мкм на ионном микронзонде ВИМС "Шиповник". Структура плёнок исследовалась при помощи просвечивающего микроскопа ПРЭМ-200. При этом реплики изготавливались в виде окружности, путём вытравливания кремния подложки.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТА

При исследовании селективности травления пленок поли-Si*, Si₃N₄ относительно SiO₂ было установлено, что максимум селективности поли-Si*, Si₃N₄ наблюдается при мощности 1.8 кВт и давлении 26Па. Селективность плазменного травления увеличивается при увеличении межэлектродного расстояния до 30-40 мм. При уменьшении межэлектродного расстояния до 10-20 мм скорость травления SiO₂ увеличивается.

Процессы травления и осаждения проводились на установке, имеющей регулировку скорости откачки, что совместно с нагрузкой пластин Si, позволяет понизить отношение F/Si, F/C и увеличить вероятность осаждения продуктов травления Si. Мощность и межэлектродное расстояние подбираются таким образом, чтобы происходило селективное травление поли-Si*. Было замечено, что при охлаждении электрода с пластинами до

температуры меньшей температуры окружающей среды, при разгерметизации реактора, в реакторе происходит осаждение паров воды, которое значительно ускоряет модификацию пленок резиста и рост плазменных пленок SiO₂ при последующем продолжении ПХО. Поэтому, процесс локальной модификации, показанный на рис.1, проводится в два этапа.

На первом этапе проводится разрушение фоторезиста в областях над поликремнием или кремнием. При этом происходит травление кремния с образованием летучих кремнийсодержащих частиц типа SiF₂. Частицы SiF₂ осаждаются в области фоторезиста над SiO₂, происходит образование связей Si-C и Si-C-Si на месте оборванных связей C-H, C-C, C=C. На втором этапе, при подаче кислорода (паров воды), проводится локальная, самосовмещенная с островками SiO₂, модификация и в результате образуются пленки окисла кремния. При помощи жидкостной обработки в серной кислоте снимаются остатки резиста над поли-Si* или кремнием.

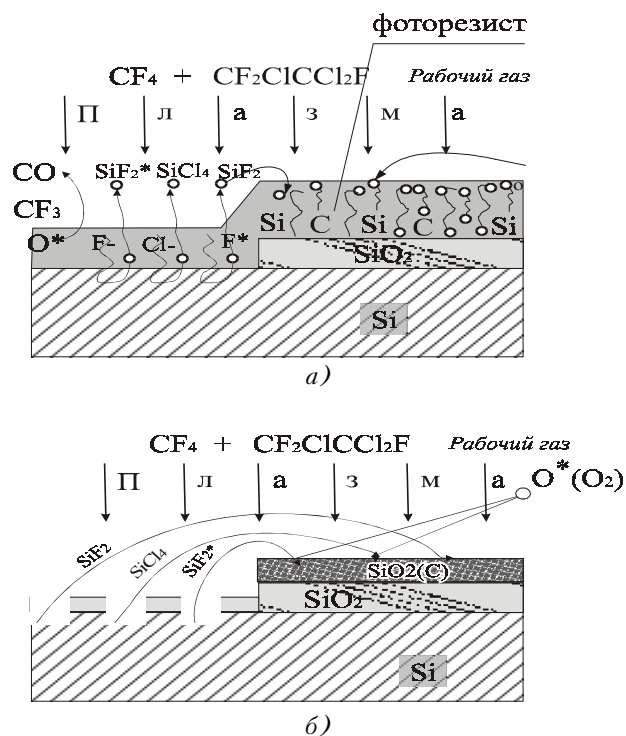


Рисунок 1 - Механизм образования плазмохимических локально-модифицированных (ПХЛМ) плёнок на SiO₂ при ПХТ кремния:

- а) первый этап - разрушение фоторезиста в области Si и травление Si, образование летучих соединений кремния и их осаждение на плёнках SiO₂ с фоторезистом;
- б) второй этап - окисление силилированных плёнок и образование локальных плёнок SiO₂

Ионы и радикалы F и Cl путём электронно- и ионно-ускоренной диффузии в резисте проникают к поверхности Si и взаимодействуют с Si с образованием летучих соединений SiF₂. В процессе разрушения фоторезиста в области Si происходят активное силилирование пленки

фоторезиста на SiO_2 , образование кремний углеродных связей.

При локальном плазменном осаждении используются режимы селективного травления Si_3N_4 относительно SiO_2 . В процессе ПХТ топологического рисунка, содержащего островки SiO_2 с фоторезистивной маской или сплошной фоторезистивной плёнкой, происходит перераспределение напряжения смещения в областях с меньшей толщиной изолятора. Что увеличивает поток ионов и усиливает процесс локального травления кремния или плёнки фоторезиста и кремния (рис.1) и уменьшает поток ионов, ослабляет травление и усиливает осаждение продуктов травления в областях маскирующей пленки SiO_2 .

Возможность создания самосовмещаемых, локальных островков пленок ПХЛМ субмикронных размеров и толщиной 0,1 мкм - 2,0 мкм имеет широкие возможности и делает необходимым изучение строения пленок ПХЛМ.

Анализ с помощью ИК-спектроскопии пленки ПХЛМ толщиной 0,45 мкм, 1,4 мкм показал, рис.2, что основные пики соответствуют связям Si-O, в областях 1060-1070 cm^{-1} , 785-795 cm^{-1} , Si-O-Si в области 450 cm^{-1} , и Si-Si, в области 6,00 cm^{-1} [1,2]. Кроме того для некоторых процессов наблюдаются пики в области 2100-2300 cm^{-1} , соответствующий связям Si-H и в области 840 cm^{-1} , соответствующий связям Si-N. В области кремния где создание пленки ПХЛМ не происходило наблюдается пик, соответствующий связям Si-Si, и незначительные пики соответствующие связям Si-O, Si-O-Si и в области 1500-1700 cm^{-1} предположительно соответствующие связям C-C, C-H, C-OH.

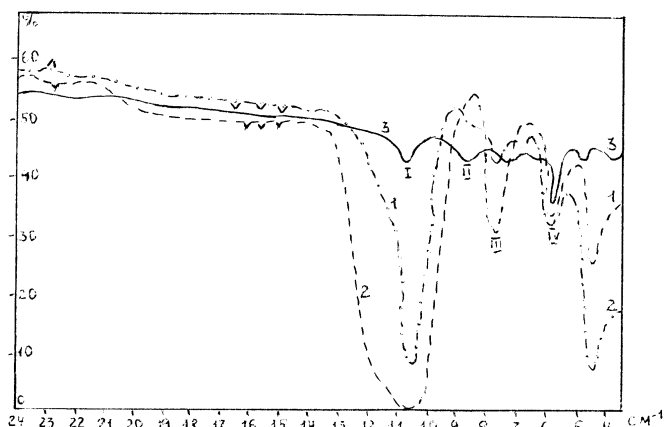


Рисунок 2 - ИК спектры плазмохимической локально модифицированной (ПХЛМ) пленки осаждаемой в области фоторезиста на термическом SiO_2 толщиной 0,08 мкм с осажденными пленками толщиной 1-0,45 мкм и 2 - 1,4 мкм, и в области фоторезиста на Si - 3

С помощью полученных Оже-спектров, рис.3, рассчитаны атомные концентрации Si, C, O пленки ПХЛМ толщиной 1,0 мкм (см. таблицу 1). По сравнению с контрольной пленкой термического окисла кремния в пленке ПХЛМ возрастает концентрация атомов кремния и наблюдается пик соответствующий атомам азота (концентрация азота ~ 1,5%).

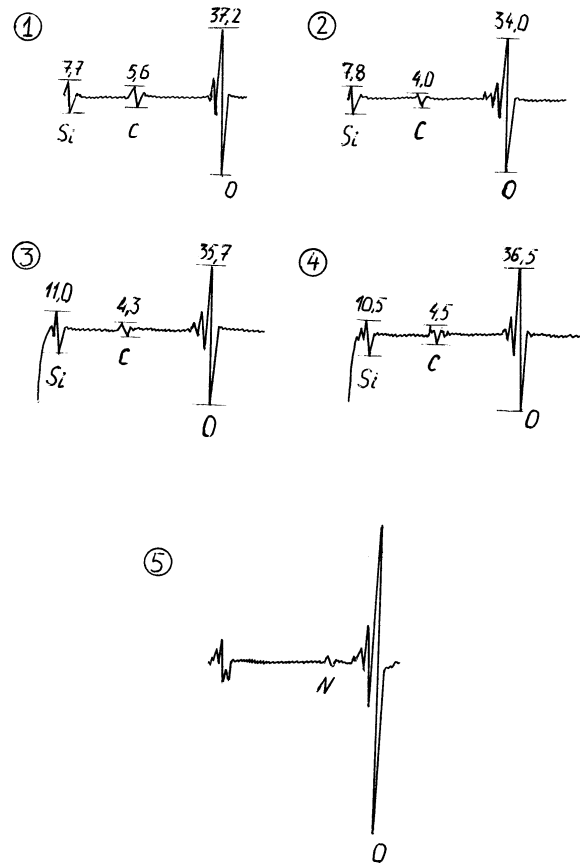


Рисунок 3 - Оже-спектры пленки термического окисла SiO_2 толщиной 0,025 мкм после травления резиста в фреоне-14 - 1, ПХЛМ пленки, осажденной на SiO_2 с использованием продуктов травления Si, в условиях положительной зарядки - 2, при перекомпенсации заряда, создаваемого первичными электронами - 3, в условиях точной компенсации заряда первичных электронов - 4 и при увеличении тока первичного пучка, позволяющего заметить наличие в пленке следов азота - 5

Увеличение концентрации атомов кремния происходит за счет образования соединений с большей концентрацией атомов кремния в единице объема, например, за счет образования карбида кремния или нитрида кремния.

Спектральные данные ВИМС подтверждают, что основой модифицированной пленки является SiO_2 . Максимальная интенсивность в масс спектре принадлежит ионам кремния, кислорода их соединениям,

интенсивность которых на 2 - 3 порядка превышает интенсивность других ионов. В масс-спектр пленки также входят углерод, водород, примеси фтора, азота, калия, комплексы С-Н, Si-Н. Из сравнения масс-спектров плёнок ПХЛМ и термического окисла можно сделать вывод, что плёнки ПХЛМ отличаются повышенным содержанием углерода (примерно на порядок) щелочных, щелочноземельных металлов и имеют существенную особенность - содержат примесь фтора.

Таблица 1 - Расчет атомных процентов по методу чистых стандартов

№ спектра рис. 3	Атомная концентрация, % Si	Атомная концентрация, % С	Атомная концентрация, % О
1	18	22	60
4	24	18	58

Согласно ВИМС содержание объемного углерода в ПХЛМ пленке выше, чем в термическом окисле.

Анализ с помощью просвечивающей микроскопии показал, что пленка имеет вид поликристаллической структуры. Средний линейный размер зерна составляет 0.33 мкм. Зёрна SiO₂ имеют характерные включения, предположительно соответствующие центрам роста и состоящие из SiC с радиусом зерна поликристаллической фазы R_{SiC} ~ 0.4R_{SiO₂}.

Электроннограмме пленки ПХЛМ, полученной на просвечивающем микроскопе ПРЭМ-200, соответствуют кристаллические решетки с межплоскостными расстояниями 3.91 Å, 4.3 Å, 4.48-4.91 Å. Этим межплоскостным расстояниям близки межплоскостные расстояния (101) α-кristобаллита SiO₂ (4,04 Å) и (100) α-кварца SiO₂ (4.25 Å). Межплоскостным расстояниям 4.5 - 4.9 Å близки межплоскостные расстояния (110) SiC (4.62 Å).

Проведенные исследования позволяют представить механизм селективного плазменного роста SiO₂. Режим плазменной обработки подбирают таким образом, чтобы равновесие травление-осаждение в области Si сдвигалось в сторону травления, а в области SiO₂ - в сторону осаждения. Травление пленок резиста происходит по-разному. При их плазменном травлении от 1-2 мкм до 0.4 мкм пленки пористые и "прозрачные" для жидкостного травителя, а над SiO₂ пленки фоторезиста сохраняют маскирующие свойства. Большее ускоряющее напряжение для заряженных частиц над Si, экзотермическое самопроизвольное взаимодействие с Si продиффундировавших через резист радикалов и ионов F и Cl, протекание реакций травления и десорбция продуктов травления приводят к нарушению резиста в области Si. В области SiO₂ самопроизвольного травления не

происходит. Отсутствуют локальное ускорение заряженных частиц, десорбция продуктов травления, что обуславливает замедление деградации резиста в области SiO₂, и создаются условия для локальной модификации и роста пленок селективных пленок SiO₂.

При локальном плазменном упрочнении пленки фоторезиста, перед травлением пленок алюминия или сплавов алюминия с кремнием, модифицированный слой располагается в приповерхностном слое фоторезиста. При плазменном создании локальной пленки окисла кремния процесс силилирования и деструкции протекает по всей толщине пленки фоторезиста.

Рассмотрим поверхностную планаризацию, сосомовещаемую с рисунком металлизации, при формировании ИС с несколькими слоями металлизации.

Формирование второго уровня металлизации ME2 проводится на рельефе, образованном металлизацией первого уровня ME1. Наличие указанного рельефа приводит к ухудшению и обрывам металлизации второго уровня ME2 на ступеньках боковых поверхностей шин ME1 из-за уменьшения коэффициента запыления рельефа. Поэтому, после формирования ME1 с АК-1 толщиной 0,6 мкм и поликремния толщиной 0,1 мкм, методом центрофугирования наносят плёнку резиста с планарной поверхностью. Плёнку подвергают дублированию, после чего проводят ПХТ плёнок резиста и поликремния, кремния под резистом и с подачей кислорода производят формирование плёнок ПХЛМ.

Формирование плёнок ПХЛМ происходит в областях SiO₂ между шинами ME1. После процесса проводят контроль плёнки резиста. Её толщина на пленках металла и поликремния составляет менее 0,1 мкм и легко удаляется при помощи растворителей, а толщина на пленках SiO₂ составляет 0,55 мкм и удаление в растворителях не происходит, так как происходит образование ПХЛМ плёнки SiO₂. Плёнку резиста и поликремния легко удаляют обычными химобработками и получают планарную поверхность ИС. Толщина ПХЛМ SiO₂ d_{SiO_2} , площадью S_{SiO_2} , зависит от площади S_{Si} подвергаемого ПХТ Si. При этом глубину H травления Si можно определить как $H = d_{SiO_2} \cdot S_{SiO_2} / 2kS_{Si}$, где $k < 1$ - коэффициент, показывающий отношение количества переосаждённых атомов Si к количеству атомов травления Si, S_{SiO_2} , S_{Si} - площади ПХЛМ SiO₂ и ПХТ Si.

ВЫВОДЫ

Путём создания на поверхности пластин монокристаллического кремния рисунка из окисла кремния, и подбора параметров ВЧ разряда подобраны условия травления фоторезиста и кремния в областях кремния и частичной деструкции фоторезиста и переосаждения

кремнийсодержащих продуктов травления в областях над окислом кремния. Последующая подача кислорода позволила сформировать локальные диэлектрические плёнки толщиной 0,45 мкм - 2,0 мкм в областях над окислом кремния. Проведено комплексное исследование состава и структуры плёнок, образующихся в системе плазмохимических реакций травление - осаждение, которые представляют собой поликристаллические плёнки окисла кремния из монокристаллов α -квартцита и α -кварца с зародышами из кристаллов карбида кремния.

ПЕРЕЧЕНЬ ССЫЛОК

1. Горбань О.М., Кравчина В.В. Процеси травлення, окислювання і дифузії в монокристалічному кремнії під плівками нітриду і окису кремнію. Вісник ЗДУ. - 1999. - №2. - С.181 - 187.
2. Голиков Ю.А., Смирнова Т.П., Соловьёв А.П. Плазмохимические методы синтеза диэлектрических слоёв в полупроводниковой технологии // Электронная техника. Сер.3 Микроэлектроника 1986. - Вып.5.-С.1-28.
3. Евдокимов И.Н., Юрасова В.Е. Влияние структурных и магнитных фазовых переходов в твёрдых телах на процессы, вызываемые ионной бомбардировкой. Поверхность. Физика, химия, механика №9 1988г.
4. Рябов С.Н., Кутолин С.А., Бойкин Н.И. Физико-химические особенности процессов плазмохимического травления// Электронная техника. Сер.7 Технология, организация производства и оборудование.-1981.- Вып.20(844). -С.9-10,59.
5. Боков Ю.С., Киреев В.Ю., Фролов В.М. Проявление скрытого силированного изображения в плёнке резиста с помощью процессов сухого травления// Электронная техника. Сер.3 Микроэлектроника.-1990г.-Вып.2.-С.51-53.

Надійшла 14.03.2000

Після доробки 04.04.2000

УДК 621.382.3

ОПТИМИЗАЦИЯ ПЕРЕДАТОЧНЫХ ФУНКЦИЙ ЦИФРОВЫХ РЕГУЛЯТОРОВ СИСТЕМ ФАП ПРИ ПРОИЗВОЛЬНЫХ ВХОДНЫХ ВОЗДЕЙСТВИЯХ

В. И. Гостев, А. П. Хорев, О. А. Побийпеч

Изложен метод расчета коэффициентов передаточной функции цифровых регуляторов в системах фазовой автоподстройки (ФАП). Расчет производится с учетом инерционности фильтра дискриминатора. Коэффициенты передаточной функции цифрового регулятора находятся в результате решения оптимизационной задачи методом Хука-Дживса с использованием квадратичного критерия качества для заданной аналоговой части системы.

The method of factors of the transfer function of digital regulators design for the Phase Lock Loop Systems (PLL) is shown. The design is made with allowance for inertia of the filter of the discriminator. The factors of the transfer function of the digital regulator are calculated as an outcome of the solution of an optimisation problem by Hook-Jivs method with use of a square criterion of quality for the preset analog part of the system.

ВВЕДЕНИЕ

Описание систем ФАП, их функциональные и структурные схемы подробно изложены в работе [1]. Обобщенная структурная схема систем ФАП приведена на рис.1,а. При использовании цифрового регулятора, обеспечивающего требуемую динамику системы, структурная схема систем ФАП преобразуется к виду, приведенному на рис.1,б, или, для линейного режима, к виду, приведенному на рис.1,в. Ниже предложен один из вариантов оптимизации передаточных функций цифровых регуляторов систем ФАП при произвольных

входных воздействиях.

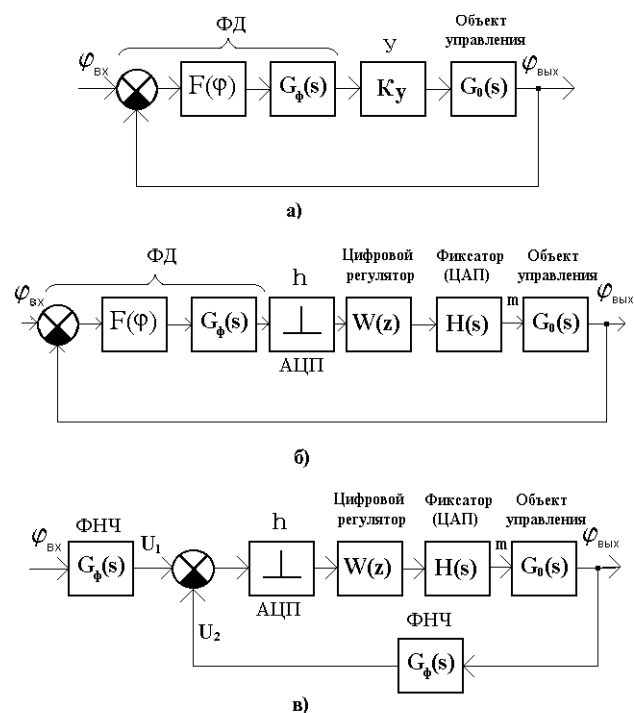


Рисунок 1

РЕШЕНИЕ ЗАДАЧИ

Если на входе аналоговой системы автоматического управления известны максимальная скорость ω_{\max} и максимальное ускорение ϵ_{\max} произвольного входного воздействия, то можно подобрать эквивалентный режим гармонического входного воздействия $u_{\vartheta}(t) = u_{\vartheta \max} \sin \omega_{\vartheta} t$, которому соответствует максимальная скорость, равная заданному значению $u_{\vartheta \max} \omega_{\vartheta} = \omega_{\max}$, и максимальное ускорение, равное заданному значению $u_{\vartheta \max} \omega_{\vartheta}^2 = \epsilon_{\max}$. При этом максимальная ошибка определяется по формулам:

$$\theta_{\max} = \omega_{\max} / K_{\omega} \text{ или } \theta_{\max} = \epsilon_{\max} / K_{\epsilon},$$

где K_{ω} - добротность по скорости системы с астатизмом первого порядка, а K_{ϵ} - добротность по ускорению системы с астатизмом второго порядка. При работе системы с ограниченными значениями скоростей и ускорений, не превышающих максимальных, ошибка системы не будет превышать θ_{\max} [2].

Если известны максимальная скорость ω_{\max} и максимальное ускорение ϵ_{\max} произвольного входного воздействия, то целесообразно оптимизировать коэффициенты передаточной функции цифрового регулятора для эквивалентного гармонического входного воздействия $u_{\vartheta}(t) = u_{\vartheta \max} \sin \omega_{\vartheta} t$.

Выберем интегральный критерий оптимальности

$$J = \sum_{n=0}^M \theta_n^2,$$

где θ_n - ошибка системы на интервалах моделирования, M - число, показывающее сколько интервалов моделирования укладывается во времени затухания переходных процессов в системе. Этот критерий используется наиболее часто и пригоден для монотонных и колебательных переходных процессов. Ошибку системы можно определять при любых детерминированных входных воздействиях. Ниже для эквивалентного гармонического воздействия на входе системы модифицированным методом условной оптимизации Хука-Дживса определяются коэффициенты передаточной функции цифрового регулятора, при которых функционал J достигает минимального значения. Метод Хука-Дживса подробно изложен в работе [3].

Рассмотрим конкретную систему фазовой автоподстройки с электронным интегратором и функциональными преобразователями [1]. Передаточные функции фильтра и объекта управления соответственно определя-

ются как

$$G_{\Phi}(s) = \frac{K_d b}{s + b}; G_0(s) = \frac{\alpha_0}{s}.$$

Таким образом, передаточная функция непрерывной части системы определяется как

$$G(s) = G_0(s)G_{\Phi}(s) = \frac{\alpha}{s(s + b)}.$$

Для такой передаточной функции непрерывной части системы передаточная функция цифрового регулятора записывается в виде [3]

$$W(z) = \frac{b_0 z + b_1}{z + a_1}.$$

Зададим параметры непрерывной части системы: $b = 1/0,016 \text{ с}^{-1}$; $\alpha = 6250 \text{ с}^{-2}$. Шаг квантования в цифровом регуляторе $h = 0,02 \text{ с}$. Шаг моделирования $h_0 = 0,05h = 0,001 \text{ с}$. Оптимизируем коэффициенты передаточной функции цифрового регулятора для эквивалентного гармонического входного воздействия $u_{\vartheta}(t) = \sin 12,5t$.

При моделировании непрерывных частей системы рис.1, в используем рекуррентные формулы по методу трапеций. Так, для интегрирующего и апериодического звеньев соответственно имеем:

$$x_n = x_{n-1} + \frac{h_0}{2}(u_n + u_{n-1});$$

$$x_n = \frac{2 - bh_0}{2 + bh_0} x_{n-1} + \frac{h_0}{2 + bh_0} (u_n + u_{n-1}),$$

где u - входная, x - выходная величины звеньев, b - параметр апериодического звена.

Разностное уравнение цифрового регулятора:

$$m(k) = \sum_{i=0}^1 b_i \theta(k-i) - a_1 m(k-1),$$

где $\theta(k) = u_1(k) - u_2(k)$.

Отметим, что временной параметр k меняется через шаг квантования $h = 0,02 \text{ с}$, а временной параметр меняется через шаг моделирования $h_0 = 0,05h = 0,001 \text{ с}$.

В результате решения оптимизационной программы при $M = 600$ (время наблюдения 0,6с) получаем следующие значения коэффициентов передаточной функции цифрового регулятора: $b_0 = 2,87$; $b_1 = 0,4$; $a_1 = 0,75$.

Таким образом, передаточная функция цифрового регулятора определяется как

$$W(z) = \frac{2,87z + 0,4}{z + 0,75}$$

На рис.2 изображены входное воздействие и выход (а) системы фазовой автоподстройки с электронным интегратором и цифровым регулятором с рассчитанной передаточной функцией, ошибка системы θ (б) и управляющее воздействие на входе объекта управления m_v (в). Система является оптимальной в смысле указанного критерия для рассмотренного гармонического воздействия с частотой $\omega_3 = 12,5$.

ЗАКЛЮЧЕНИЕ

Изложенный подход к расчету коэффициентов передаточной функции цифрового регулятора является достаточно простым, универсальным и гибким. Если цифровой регулятор выполняется на микропроцессоре, то можно произвести расчет коэффициентов для ступенчатого входного воздействия и произвольного входного воздействия и организовать автоматическое переключение из режима отработки скачкообразного воздействия в режим слежения за произвольным входным воздействием, и наоборот. Цифровой регулятор будет при этом иметь постоянную структуру, но разные коэффициенты, соответствующие каждому режиму.

ПЕРЕЧЕНЬ ССЫЛОК

1. Зайцев Г.Ф., Стеклов В.К. Радиотехнические системы автоматического управления высокой точности. - К.: Техника, 1988. - 208 с.
2. Бесекинский И.А. Динамический синтез систем автоматического регулирования. - М.: Наука, 1970. - 576 с.

3. Гостев В.И., Стеклов В.К., Скляренко С.Н. Оптимальные системы управления с цифровыми регуляторами: Справ. - К.: КИРЦ "Сенс", 1995. - 484 с.

Надійшла 16.12.1999
Після доробки 14.03.2000

УДК 621.372.8.01

АЛГОРИТМ КВАЗИСТАТИЧЕСКОГО МОДЕЛИРОВАНИЯ ПОЛОСКОВЫХ СТРУКТУР В МНОГОСЛОЙНОЙ АНИЗОТРОПНОЙ СРЕДЕ

Л. М. Карпуков

Предложена элементная база для построения декомпозиционных схем, моделирующих структуру многослойных анизотропных подложек в задачах квазистатического анализа полосковых линий передачи. Разработан метод расчёта функций Грина по декомпозиционным схемам исследуемых структур. Приведены примеры анализа полосковых линий передачи с учётом анизотропных свойств многослойных подложек.

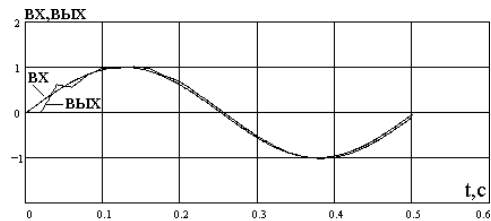
Запропоновано елементну базу для побудови декомпозиційних схем, що моделюють структуру багатослойних анізотропних підкладок у задачах квазістатичного аналізу смужкових ліній передачі. Розроблено метод розрахунку функцій Гріна за декомпозиційними схемами досліджуваних структур. Приведено приклади аналізу смужкових ліній передачі з урахуванням анізотропних властивостей багатослойних

підкладок.

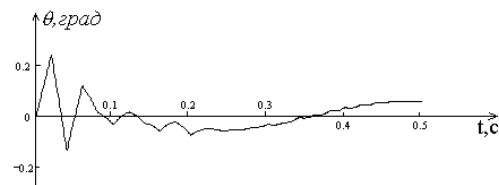
The elementary base for building decomposition schemes which model the structure of the multilayered anisotropic substrates in the problems of quasistatic analysis of strip lines is proposed. The method of counting the Green's functions by using the decomposition schemes of the investigated structures is developed. Examples of the strip lines analysis taking into account the anisotropic properties of multilayered substrates are given.

ВВЕДЕНИЕ

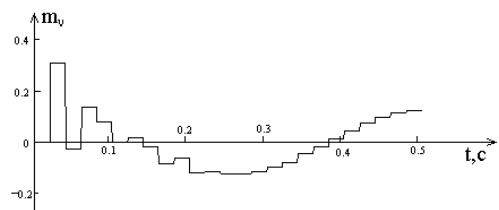
При конструировании современных интегральных схем СВЧ широко используются многослойные комби-



а).



б).



в).

Рисунок 2

нированные подложки с расположением токонесущих проводников линий передач в различных её слоях, материал которых может отличаться физическими свойствами и степенью их анизотропии. Эффективность проектирования указанных конструкций определяется качеством алгоритмов, используемых для моделирования плоскостной анизотропной среды.

В квазистатическом приближении задача моделирования полосковых линий передачи формулируется в виде краевой задачи электростатики. Её решение в случае анализа линии на однослойной анизотропной подложке выполнялось различными методами: конечных разностей [1], моментов [2], конечных элементов [3], преобразования координат [4]. Среди перечисленных методов наиболее полно требованию построения универсальных алгоритмов, обеспечивающих анализ полосковых линий на многослойных анизотропных подложках, удовлетворяет метод моментов. Основной процедурой этого метода является нахождение функции Грина рассматриваемой краевой задачи. В [5] предложен эффективный метод расчёта функций Грина для многослойных изотропных подложек, который позволяет заменить решение сложных краевых задач простой и наглядной процедурой построения и анализа декомпозиционной модели слоистой диэлектрической среды.

Целью настоящей работы является дальнейшее развитие этого метода с разработкой элементной базы для построения декомпозиционных моделей многослойных анизотропных диэлектрических структур и с созданием экономичных и простых в реализации алгоритмов расчёта функций Грина по декомпозиционным моделям соответствующих краевых задач.

ПОСТАНОВКА ЗАДАЧИ

На рис.1 представлено поперечное сечение исследуемой полосковой передающей линии в виде нескольких плоских проводников, расположенных на поверхности диэлектрических пластин. Продольной осью линий является ось z . Сверху и снизу структуры могут располагаться экраны. Пластины и экраны имеют бесконечную протяженность вдоль оси x .

Пусть тензор относительной диэлектрической проницаемости i -й пластины имеет вид:

$$\|\epsilon_i\| = \begin{vmatrix} \epsilon_{xx_i} & \epsilon_{xy_i} & 0 \\ \epsilon_{xy_i} & \epsilon_{yy_i} & 0 \\ 0 & 0 & \epsilon_{zz_i} \end{vmatrix}. \quad (1)$$

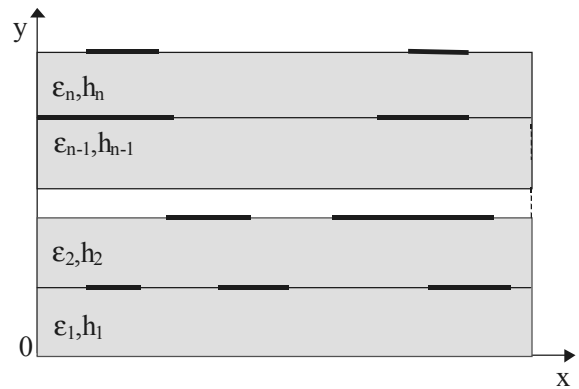


Рисунок 1 - Поперечное сечение подложки

Распределение потенциала в рассматриваемой структуре подчиняется уравнению Лапласа:

$$\nabla \cdot (\|\epsilon_i\| \cdot \nabla \varphi_i) = 0. \quad (2)$$

На поверхности раздела i -й и k -й диэлектрических пластин имеет место:

$$\left. \begin{aligned} \bar{n} \cdot (\|\epsilon_i\| \cdot \nabla \varphi_i - \|\epsilon_k\| \cdot \nabla \varphi_k) &= 0, \\ \varphi_i &= \varphi_k, \end{aligned} \right\} \quad (3)$$

где \bar{n} - нормаль к поверхности, совпадающая по направлению с осью y .

Необходимо определить по (1) - (3) матрицы рассеяния типичных элементов рассматриваемой структуры, составляющих элементную базу для построения декомпозиционных моделей многослойных подложек.

ЭЛЕМЕНТНАЯ БАЗА МОДЕЛИРОВАНИЯ

На основании (2) уравнение для функции Грина моделируемой двумерной структуры при расположении единичного линейного заряда в i -й пластине в точке с координатами x_0, y_0 приобретает следующий вид:

$$\begin{aligned} \epsilon_{yy_i} \frac{\partial^2 G_i}{\partial y^2} + 2\epsilon_{xy_i} \frac{\partial^2 G_i}{\partial x \partial y} + \epsilon_{xx_i} \frac{\partial^2 G_i}{\partial x^2} &= \\ &= \frac{1}{\epsilon_0} \delta(y - y_0) \delta(x - x_0), \end{aligned} \quad (4)$$

где ϵ_0 - электрическая постоянная, δ - дельта-функция Дирака.

Из (3) вытекают граничные условия для функций Грина на поверхности раздела сред:

$$\left. \begin{aligned} & \left(\varepsilon_{yy_i} \frac{\partial G_i}{\partial y} + \varepsilon_{xy_i} \frac{\partial G_i}{\partial x} \right) - \left(\varepsilon_{yy_k} \frac{\partial G_k}{\partial y} + \varepsilon_{xy_k} \frac{\partial G_k}{\partial x} \right) = 0, \\ & G_i = G_k. \end{aligned} \right\} \quad (5)$$

Преобразование Фурье функций G_i , G_k по координате x даёт представление (4), (5) в спектральной области:

$$\begin{aligned} & \frac{\partial^2}{\partial y^2} G_i(k_x) - 2jk_x \alpha_i \frac{\partial}{\partial y} G_i(k_x) - k_x^2 \beta_i G_i(k_x) = \\ & = -\frac{e^{jk_x x_0}}{2\pi \varepsilon_0 \varepsilon_{yy_i}} \delta(y - y_0), \end{aligned} \quad (6)$$

$$\left. \begin{aligned} & \varepsilon_{yy_i} \left(\frac{\partial}{\partial y} G_i(k_x) - jk_x \alpha_i G_i(k_x) \right) = \\ & = \varepsilon_{yy_k} \left(\frac{\partial}{\partial y} G_k(k_x) - jk_x \alpha_k G_k(k_x) \right), \\ & G_i(k_x) = G_k(k_x). \end{aligned} \right\} \quad (7)$$

Здесь k_x - переменная преобразования Фурье, j - мнимая единица, $\alpha_m = \varepsilon_{xy_m} / \varepsilon_{yy_m}$, $\beta_m = \varepsilon_{xx_m} / \varepsilon_{yy_m}$, $m = i, k$.

Решение однородного уравнения, соответствующего (6), запишем в виде суперпозиции прямой и обратной волн, распространяющихся вдоль и против оси y :

$$G_i(k_x) = u_i^+ e^{-k_x(\xi_i - j\alpha_i)y} + u_i^- e^{k_x(\xi_i + j\alpha_i)y}, \quad (8)$$

где $\xi_i = (\beta_i - \alpha_i^2)^{1/2}$.

Выделив граничными сечениями y_1 и $y_2 = y_1 + h$ в i -й пластине слой диэлектрика, толщиной h , получим на основании (8) соотношение для матрицы рассеяния слоя:

$$\left. \begin{aligned} u_1^- &= u_2^+ e^{-k_x(\xi_i + j\alpha_i)h}, \\ u_2^- &= u_1^+ e^{-k_x(\xi_i - j\alpha_i)h}. \end{aligned} \right\} \quad (9)$$

Здесь и далее знаком плюс обозначены амплитуды волн, падающих на граничное сечение, а знаком минус - отраженные от него.

Для выделенной граничными сечениями $y_1 = y_{ik} - 0$ и $y_2 = y_{ik} + 0$ поверхности $y = y_{ik}$ раздела i -го и k -го диэлектриков, в соответствии с (7), (8) будет иметь место:

$$\left. \begin{aligned} u_1^- &= \Gamma_1 u_1^+ + (1 - \Gamma_1) u_2^+, \\ u_2^- &= (1 + \Gamma_1) u_1^+ - \Gamma_1 u_2^+. \end{aligned} \right\} \quad (10)$$

Здесь $\Gamma_1 = (\varepsilon_{\varepsilon_{k_i}} - \varepsilon_{\varepsilon_{k_k}}) / (\varepsilon_{\varepsilon_{k_i}} + \varepsilon_{\varepsilon_{k_k}})$ - коэффициент отражения от слоя; $\varepsilon_{\varepsilon_{k_m}} = (\varepsilon_{xx_m} \varepsilon_{yy_m} - \varepsilon_{xy_m}^2)^{1/2}$, $m = i, k$.

В сечении $y = y_0$, где расположен источник поля, граничные условия в соответствии с (6), (7) приобретают следующий вид:

$$\left. \begin{aligned} & \left(\frac{\partial}{\partial y} G_i(k_x, y_0 - 0) - jk_x \alpha_i G_i(k_x, y_0 - 0) \right) - \\ & - \left(\frac{\partial}{\partial y} G_i(k_x, y_0 + 0) - jk_x \alpha_i G_i(k_x, y_0 + 0) \right) = \\ & = \frac{e^{jk_x x_0}}{2\pi \varepsilon_0 \varepsilon_{yy_i}} \delta(y - y_0), \\ & G_i(k_x, y_0 + 0) = G_k(k_x, y_0 + 0). \end{aligned} \right\} \quad (11)$$

Отсюда с учётом (8) вытекают соотношения, определяющие относительно граничных сечений $y_1 = y_0 - 0$ и $y_2 = y_0 + 0$ матрицу рассеяния источника поля:

$$\left. \begin{aligned} u_1^- &= u_2^+ + e_i, \\ u_2^- &= u_1^+ + e_i, \end{aligned} \right\} \quad (12)$$

где $e_i = \exp(jk_x x_0) / (2\pi \varepsilon_0 \varepsilon_{\varepsilon_{k_i}})$ - Фурье-изображение источника.

МОДЕЛЬ ОДНОСЛОЙНОЙ СТРУКТУРЫ

В спектральной области функция Грина определяется в результате анализа декомпозиционной схемы моделируемой структуры, составляемой из базовых элементов. При этом для простых структур коэффициенты передачи, характеризующие функции Грина, удобно определять по ориентированному графу декомпозиционной схемы. Для примера на рис. 2 приведён ориентированный граф декомпозиционной схемы диэлектрической пластины из анизотропного материала (область 1). Пластина, имеющая толщину h_i , расположена в свободном пространстве (область 2, $\varepsilon_2 = 1$). Нижняя поверхность пластины металлизирована. В точке с координатами x_0, y_0 расположен единичный заряд. Точки с координатами x_0, y_1 и x_0, y_2 являются точками наблюдения.

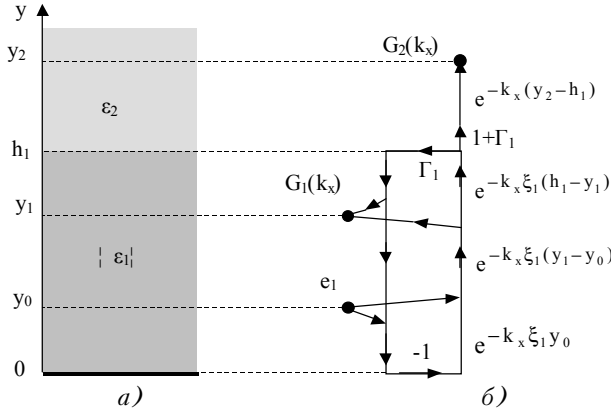


Рисунок 2 - Модель однослойной подложки:
а) структура подложки, б) ориентированный граф её декомпозиционной схемы

Из анализа графа вытекают соотношения для фурье-изображения функций Грина в области 1 и 2:

$$G_1(k_x, x_0, y_1, y_0) = \frac{e^{jk_x(x_0 + \alpha_1(y_1 - y_0)) - k_x \xi_1(y_1 - y_0)}}{4\pi\epsilon_0\epsilon_{\text{ЭК}1} k_x (1 + \Gamma_1 e^{-2k_x \xi_1 h_1})} \cdot [1 - e^{-2k_x \xi_1 y_0} + \Gamma_1 e^{-2k_x \xi_1(h_1 - y_1)} - \Gamma_1 e^{-2k_x \xi_1(h_1 - y_1 + y_0)}], \quad (13)$$

$$G_2(k_x, x_0, y_2, y_0) = \frac{(1 + \Gamma_1) e^{jk_x(x_0 + \alpha_1(h_1 - y_0)) - k_x \xi_1(h_1 - y_0) - k_x(y_1 - h_1)}}{4\pi\epsilon_0\epsilon_{\text{ЭК}1} k_x (1 + \Gamma_1 e^{-2k_x \xi_1 h_1})}. \quad (14)$$

Оригиналы для функций Грина могут быть получены в результате разложения знаменателя этих соотношений в геометрическую прогрессию с последующим применением обратного преобразования Фурье к членам образующегося ряда. Связь между изображением некоторого члена ряда и его оригиналом имеет вид:

$$\int_{-\infty}^{\infty} \frac{e^{-jk_x(x - x_{\text{ЭК}0}) - k_x \Delta y_{\text{ЭК}}}}{4\pi\epsilon_0\epsilon_{\text{ЭК}1} k_x} dk_x = \frac{-1}{4\pi\epsilon_0\epsilon_{\text{ЭК}1}} \ln[(x + x_{\text{ЭК}0})^2 + \Delta y_{\text{ЭК}}^2], \quad (15)$$

где $x_{\text{ЭК}0} = x_0 + \alpha_1(y - y_0)$,

$$\Delta y_{\text{ЭК}} = \xi_1 \Delta y_1 + \Delta y_2.$$

Выполнив указанные преобразования с соотношениями (13), (14), получим искомые зависимости для оригиналов:

$$G_1(x, x_0, y_1, y_0) = \frac{-1}{4\pi\epsilon_0\epsilon_{\text{ЭК}1}} \sum_{n=0}^{\infty} (-\Gamma_1)^n \{ \ln[(x - x_{\text{ЭК}10})^2 + \xi_1^2(y_1 - y_0 + 2nh_1)^2] - \ln[(x - x_{\text{ЭК}10})^2 + \xi_1^2(y_1 + y_0 - 2nh_1)^2] + \Gamma_1 \ln[(x - x_{\text{ЭК}10})^2 + \xi_1^2(2(n+1)h_1 - y_1 - y_0)^2] - \Gamma_1 \ln[(x - x_{\text{ЭК}10})^2 + \xi_1^2(2(n+1)h_1 - y_1 + y_0)^2] \}, \quad (16)$$

$$G_2(x, x_0, y_2, y_0) = \frac{-(1 + \Gamma_1)}{4\pi\epsilon_0\epsilon_{\text{ЭК}1}} \sum_{n=0}^{\infty} (-\Gamma_1)^n \left\{ \ln[(x - x_{\text{ЭК}20})^2 + (\xi_1(h_1 - y_0 + 2nh_1) + y_2 - h_1)^2] - \ln[(x - x_{\text{ЭК}20})^2 + (\xi_1(h_1 + y_0 + 2nh_1) + y_2 - h_1)^2] \right\}, \quad (17)$$

где $x_{\text{ЭК}10} = x_0 + \alpha_1(y_1 - y_0)$, $x_{\text{ЭК}20} = x_0 + \alpha_1(h_1 - y_0)$.

Отметим, что при $y_1 = y_0 = h$ эти зависимости совпадают с соотношением, приведенным в [6].

МОДЕЛИРОВАНИЕ МНОГОСЛОЙНЫХ СТРУКТУР

Для алгоритмизации процесса вычисления функций Грина представим математическую модель декомпозиционной схемы n -слойной структуры следующим образом:

$$U^-(k_x) = S \cdot I \cdot P(k_x) U^-(k_x) + (E + S) V(k_x), \quad (18)$$

$$\left. \begin{aligned} U^+(k_x) &= I \cdot P(k_x) U^-(k_x), \\ U(k_x) &= U^+(k_x) + U^-(k_x). \end{aligned} \right\} \quad (19)$$

Здесь $U^-(k_x)$, $U^+(k_x)$ - векторы, составленные из амплитуд отраженных и падающих волн в граничных сечениях, совпадающих с поверхностью раздела слоёв; $S = \text{diag}(-\Gamma_1, S_2, \dots, S_n, \Gamma_{n+1})$ - блочно-диагональная матрица, составленная из матриц рассеяния разделов слоёв, определяемых по (10); $P(k_x) = \text{diag}[\exp(-k_x h_{\text{ЭК}1}), \dots, \exp(-k_x h_{\text{ЭК}n})]$ - диагональная матрица, составленная по (9) из коэффициентов передачи слоёв, $h_{\text{ЭК}i} = \xi_i h_i$; I - матрица перестановок, состоящая из нулей и единиц и

характеризующая связи между амплитудами падающих и отраженных волн; $V(k_x)$ - вектор, составляемый в соответствии с (12) из фурье-изображений источников e_i , вводимых в i -тое граничное сечение; $U(k_x)$ - вектор искомых реакций, совпадающих с изображениями функций Грина; E - единичная матрица.

В соответствии с [5] для нахождения оригиналов функций Грина введём шаг расчёта h , кратный эквивалентным толщинам $h_{эк_i}$ слоёв. Умножим обе части (18) на $\exp(-k_x mh)$ и перейдем в (18), (19) от фурье-изображений переменных к оригиналам. В результате получим:

$$U^-(m) = S_I U^-(m+i) + S_E V(m), \quad (20)$$

$$U(0) = U^-(0) + I U^-(0+i). \quad (21)$$

Здесь $S_I = SI$, $S_E = E + S$, $U^-(m+i) = [u_1^-(m+i_1), \dots, \dots, u_n^-(m+i_n)]^t$, i_k - число шагов h , укладываемое по ширине $h_{эк_k}$ k -го слоя, t - знак транспонирования.

Элементы вектора $V(m)$ в соответствии с (12), (15) записываются в виде:

$$V_i(m) = \frac{-1}{4\pi\epsilon_0\epsilon_{эк_i}} \ln[(x+x_{эк_0})^2 + (mh)^2]. \quad (22)$$

Соотношение (20) представляет собой рекуррентную формулу. Для начала вычислений используется условие для потенциала на бесконечности:

$$U^-(m+1) = 0, \quad m \rightarrow \infty. \quad (23)$$

Искомая функция Грина в соответствии с (21) определяется по вектору $U^-(0)$, получаемому на последнем шаге расчёта, и по предшествующим значениям этого вектора.

Если расчленить слои вспомогательными граничными сечениями на отрезки длиной h и дополнить уравнения (18), (19) переменными, относящимися к вспомогательным сечениям, то соотношения (20), (21) приобретут наиболее простой вид:

$$U^-(m) = S_I U^-(m+1) + S_E V(m), \quad (24)$$

$$U(0) = U^-(0) + I U^-(1). \quad (25)$$

Вычисления по этим формулам могут быть представлены в явной форме:

$$U = \sum_{m=0}^{\infty} S_I^m \cdot S_E V(m) + I \cdot S_I^m \cdot S_E V(m+1). \quad (26)$$

Составление формул вида (20) или (26) по декомпозиционным схемам анализируемых структур не вызывает затруднений. Например, при возбуждении границы раздела $y = h$ однослойной структуры на рис.2, а и шаге расчёта $h = h_{эк_1}$ функция Грина $G(x, x_0, h, h) = u_2$ определяется формулой:

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \frac{-(1+\Gamma_1)}{4\pi\epsilon_0\epsilon_{эк_1}} \sum_{n=0}^{\infty} \begin{bmatrix} 0 & -1 \\ \Gamma_1 & 0 \end{bmatrix}^n \cdot \begin{bmatrix} 0 \\ \ln(\Delta x^2 + (m+1)^2 h_{эк_1}^2) \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & -1 \\ \Gamma_1 & 0 \end{bmatrix}^m \cdot \begin{bmatrix} 0 \\ \ln(\Delta x^2 + (m+1)^2 h_{эк_1}^2) \end{bmatrix}, \quad (27)$$

где $\Delta x = x - x_0$, $h_{эк_1} = \xi_1 h_1$.

ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Расчёт характеристик полосковых линий передачи производится по результатам решения интегрального уравнения

$$\varphi(x, y) = \int_{s_0} G(x, x_0, y, y_0) \sigma(x_0, y_0) ds_0, \quad (28)$$

где $\sigma(x_0, y_0)$ - искомое распределение заряда на поверхности s_0 проводников, $\varphi(x, y)$ - заданное значение потенциала на проводниках.

Для решения использовался метод моментов с базисом в виде кусочно-постоянных функций. Функции Грина рассчитывались по соотношениям (20), (21), кратность толщин слоёв шагу расчёта учитывалась с помощью линейной интерполяции. По распределению зарядов определялись погонные ёмкости, а затем по известным соотношениям вычислялись волновые сопротивления Z_0 и эффективные диэлектрические проницаемости $\epsilon_{эф}$ линий передачи.

В качестве примера выполнен расчёт эффективной диэлектрической проницаемости микрополосковой линии на комбинированной подложке, структуры которой представлены на рис. 3. Слои подложки составлялись из изотропного диэлектрика с относительной проницаемостью $\epsilon_r = 9,9$, а также из сапфира с проницаемостями: $\epsilon_{xx} = 11,4$, $\epsilon_{yy} = 13,2$ и $\epsilon_{xy} = 0$ [2].

На рис. 4,а представлены зависимости $\epsilon_{эф}$ от отношения ширины полоска w к толщине $h = h_1$ при $h_2 = 1$ мм для линии, изображенной на рис. 3,а. Аналогичные

зависимости для линии на рис. 3,б приведены на рис. 4,б при $h_1 = h_2 = 0,5$ мм и нормировке ширины полоска w к $h = h_1 + h_2$. На рисунках кривая 1 - оба слоя подложки из сапфира, кривая 2 - первый слой из сапфира, второй из изотропного диэлектрика, кривая 3 - первый слой из изотропного диэлектрика, второй из сапфира, кривая 4 - оба слоя из изотропного диэлектрика.

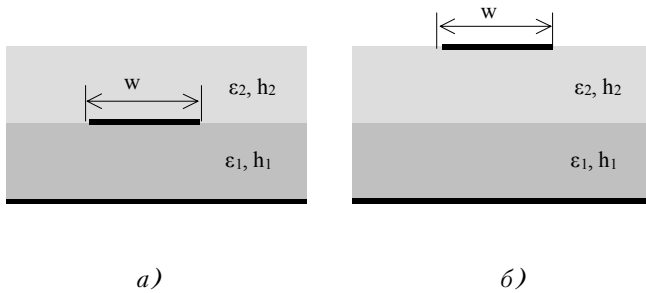
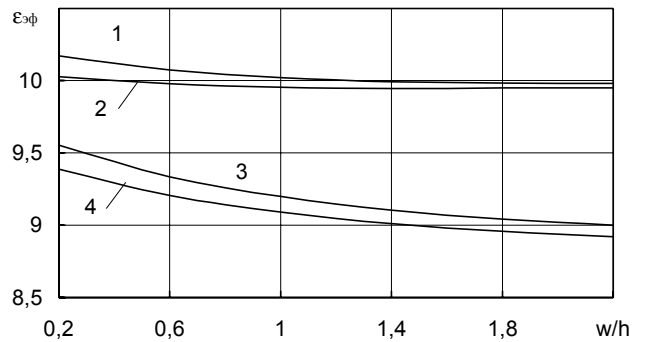


Рисунок 3 - Микрополосковая линия на комбинированной подложке

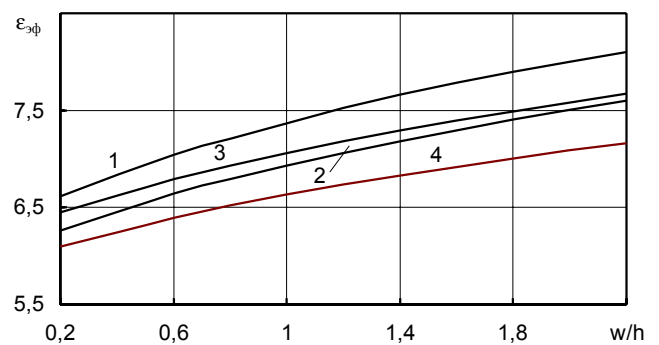
ВЫВОДЫ

Осуществлена алгоритмизация вычислительного процесса, имеющего место при анализе в квазистатическом приближении полосковых линий передачи на многослойной комбинированной подложке, содержащей анизотропные слои. С использованием принципа декомпозиции сложные краевые задачи по расчёту распределения потенциала в слоистой анизотропной среде представлены в виде простой и легко формализуемой процедуры анализа декомпозиционных схем.

Разработанная элементная база для построения декомпозиционных схем многослойных подложек и алгоритмы расчёта функций Грина позволят расширить возможности и повысить эффективность существующих математических средств квазистатического моделирования в системах автоматизированного проектирования микроэлектронных устройств СВЧ.



а)



б)

Рисунок 4 - Зависимость эффективной диэлектрической проницаемости от отношения w/h : а) для линии на рис. 3,а; б) для линии на рис. 3,б

ПЕРЕЧЕНЬ ССЫЛОК

- Owens P., Aitken J.E., Edwards T.C. Quasi-static characteristics of microstrip on an anisotropic sapphire substrate // IEEE Trans. MTT.-1976.-V.24.-№8.-P.499-505.
- Alexopoulos N.G. Krowne C.M. Characteristics of single and coupled microstrips on anisotropic substrates // IEEE Trans. MTT.-1978.-V.26.-№6.-P.387-393.
- El Said M., Ahmed A.A. Microstrip analysis on anisotropic and inhomogeneous substrate with the finite element method // Proc. IEEE Int. Microwave Symp.-1980.-P.465-467.
- Szentkuti B.T. Simple analysis of anisotropic microstrip lines by a transform methods // Electron. Lett. -1976.-V.25.-№12.-P.672-673.
- Карпуков Л. М. Построение и анализ декомпозиционных моделей микрополосковых структур // Радиозлектроника. - 1984. - Т. 27. - №9. - С. 32 - 36. (Изв. высш. учеб. заведений).
- Kobayashi M. Green's function technique for solving anisotropic electrostatic field problems // IEEE Trans. MTT.-1978.-V.26.-№7.-P.510-512.

Надійшла 24.02.2000
Після доробки 27.03.2000

ОЦЕНКА ДОСТОВЕРНОСТИ ДОПУСКОВОГО КОНТРОЛЯ

К. Н. Касьян

Рассматриваются оценки качества определения технического состояния объекта посредством допускового контроля с использованием методов теории вероятности и математической статистики.

Розглядаються оцінки якості визначення технічного стану об'єкту шляхом допускового контролю з використанням методів теорії імовірності і математичної статистики.

The evaluations of quality of definition of availability index of product of the object by means of diagnostic control with use of methods of a probability theory and mathematical statistics are considered.

Основным критерием оценки качества допускового контроля является его достоверность. Достоверность контроля - это показатель степени объективного отображения результатами контроля действительного технического состояния объекта контроля [1].

Для допускового контроля объект контроля может находиться в двух состояниях: исправном (ис) с вероятностью $P_{ис}$ и неисправном (нс) с вероятностью $Q_{нс}$ (рис-сунк 1).

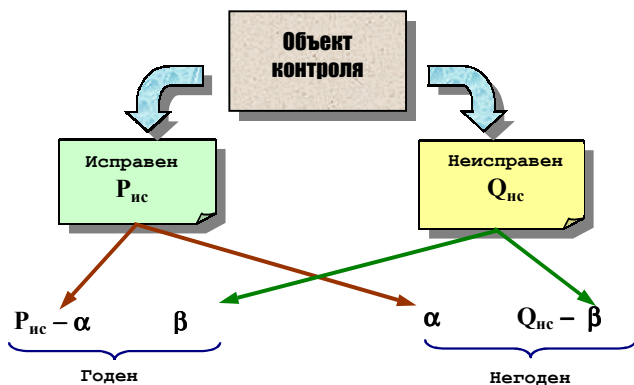


Рисунок 1 - Диаграмма формирования результатов допускового контроля

По результатам контроля проводится оценка "годен" или "негоден" контролируемого параметра, определяющего техническое состояние объекта контроля "исправен" или "неисправен". В результате появления возможных ошибок результатов контроля типа "ложный отказ" с вероятностью α и "необнаруженный отказ" с вероятностью β полные вероятности правильных результатов контроля будут соответствовать:

- оценка "годен" по результатам контроля соответствует исправному техническому состоянию объекта контро-

ля с полной вероятностью $P_{ис} - \alpha$;

- оценка "годен" по результатам контроля соответствует неисправному техническому состоянию объекта контроля с вероятностью β ;

- оценка "негоден" по результатам контроля соответствует исправному техническому состоянию объекта контроля с вероятностью α ;

- оценка "негоден" по результатам контроля соответствует неисправному техническому состоянию объекта контроля с полной вероятностью $Q_{нс} - \beta$.

Для количественной оценки достоверности допускового контроля применяется критерий условной вероятности Бейеса [2], соответствующий отношению искомой вероятности к полной вероятности получения соответствующей информации:

- при соответствии оценки "годен" (Γ) контролируемого параметра исправному техническому состоянию объекта контроля

$$P(\Gamma/ис) = \frac{P_{ис} - \alpha}{(P_{ис} - \alpha) + \beta}; \quad (1)$$

- при соответствии оценки "негоден" ($\Pi\Gamma$) контролируемого параметра неисправному техническому состоянию объекта контроля

$$P(\Pi\Gamma/нс) = \frac{Q_{нс} - \beta}{(Q_{нс} - \beta) + \alpha}. \quad (2)$$

Теоретический анализ уравнений (1) и (2) показывает, что при повышении надежности объекта контроля, т.е. вероятности безотказной работы $P_{ис}$, и при уменьшении вероятностей появления ошибок α и β критерий условной вероятности соответствия оценки "годен" исправному техническому состоянию объекта контроля значительно быстрее приближается к единице $P(\Gamma/ис) \rightarrow 1$, чем критерий условной вероятности соответствия оценки "негоден" неисправному техническому состоянию объекта контроля. Это объясняется соотношениями значений $P_{ис} - \alpha \gg \beta$ и $Q_{нс} - \beta \approx \alpha$ в знаменателях уравнений (1) и (2).

При заданных значениях $P_{ис}$ вероятность $P(\Gamma/ис)$ определяется в основном значением β , а вероятность $P(\Pi\Gamma/нс)$ определяется в основном значением α (при условии соизмеримости значений α и β). При много-

кратном повторении операций контроля одного и того же параметра q , что с точки зрения теории надежности соответствует параллельной структуре объекта контроля, достоверность полученного результата значительно повышается и может быть определена по уравнениям

$$P_n(\Gamma/\text{ис}) = 1 - [1 - P_i(\Gamma/\text{ис})]^n; \quad (3)$$

$$P_n(\text{нг}/\text{ис}) = 1 - [1 - P_i(\text{нг}/\text{ис})]^n, \quad (4)$$

где n - количество операций контроля параметра q .

Окончательная оценка технического состояния объекта контроля проводится по результатам многопараметрического допускового контроля нескольких m параметров q_1, q_2, \dots, q_m , определяющих функциональную работоспособность объекта контроля. Такая структура объекта контроля с точки зрения теории надежности является последовательной.

Общий результат "годен" совокупности всех m контролируемых параметров q_1, q_2, \dots, q_m соответствует только при условии получения "годен" всех m результатов контроля каждого в отдельности контролируемого параметра, т.е. значения всех контролируемых параметров находятся в пределах соответствующих полей допусков. Достоверность допускового контроля, выраженная через условную общую вероятность соответствия результата "годен" совокупности всех m контролируемых параметров q_1, q_2, \dots, q_m исправному техническому состоянию объекта контроля будет равна

$$P_m(\Gamma/\text{ис}) = \frac{\prod_{j=1}^m (P_{\text{ис } j} - \alpha_j)}{\prod_{j=1}^m [(P_{\text{ис } j} - \alpha_j) + \beta_j]} = \prod_{j=1}^m P_j(\Gamma/\text{ис}), \quad (5)$$

где $P_m(\Gamma/\text{ис})$ - условная вероятность соответствия оценки "годен" j -го контролируемого параметра q_j исправному техническому состоянию объекта контроля.

Общий результат "негоден" совокупности всех m контролируемых параметров q_1, q_2, \dots, q_m соответствует при условии получения результата "негоден" любого одного из контролируемых параметров, т.е. значение хотя бы одного из контролируемых параметров находится за пределами соответствующего поля допуска. В этом случае достоверность допускового контроля, выраженная через условную общую вероятность соответствия результата "негоден" неисправному техническому состоянию объекта контроля будет определяться уравнением

$$P_m(\text{нг}/\text{ис}) = \frac{1 - \prod_{j=1}^m (P_{\text{ис } j} + \beta_j)}{1 - \prod_{j=1}^m [(P_{\text{ис } j} - \alpha_j) + \beta_j]}. \quad (6)$$

Результат "негоден", полученный при любой проверке любого j -го контролируемого параметра, распространяется на весь объект контроля. Считается, что в одном цикле контроля результат "негоден" может встретиться только один раз, так как после обнаружения результата "негоден" весь цикл контроля прерывается с целью проведения анализа для локализации и устранения неисправности объекта контроля. С учетом возможного появления ошибок при определении технического состояния объекта контроля целесообразно ввести дополнительный коэффициент запаса, характеризуемый вероятностью P_d повышения надежности безотказной работы объекта контроля и вероятностью P_m методической составляющей достоверности применяемого метода контроля (вероятность безотказной работы средств контроля). Полная вероятность $P_{\Pi}(\Gamma/\text{ис})$, характеризующая коэффициент запаса достоверности контроля, будет равна $P_{\Pi}(\Gamma/\text{ис}) = P_m P_m(\Gamma/\text{ис})$.

Нижним допустимым пределом значения достоверности допускового контроля соответствия результата "годен" исправному техническому состоянию объекта контроля устанавливается заданное значение вероятности P_3 безотказной работы объекта контроля, которое указывается в техническом задании на разработку объекта контроля. Для обеспечения высокого качества допускового контроля результирующее значение достоверности допускового контроля не должно быть меньше нижнего допустимого предела

$$P_{\Pi}(\Gamma/\text{ис}) P_d \geq P_3. \quad (7)$$

В противном случае нецелесообразно проводить допусковый контроль, так как по результатам допускового контроля с достоверностью ниже чем P_3 информация о техническом состоянии объекта контроля будет ошибочной.

Зависимость параметров $P_d, P_{\Pi}(\Gamma/\text{ис}), P_3$ можно представить в виде

$$P_{\Pi}(\Gamma/\text{ис}) = k + 1/\sqrt{P_3}, \quad (8)$$

$$P_d = k + 1/\sqrt{P_3^k}, \quad (9)$$

где $k = \lg P_d / \lg P_{\Pi}(\Gamma/\text{ис})$ может изменяться в диапазоне $0 < k < \infty$.

При значениях $k \rightarrow \infty$ показатель достоверности результата контроля $P_{\Pi}(\Gamma/\text{ис}) \rightarrow 1$, а вероятность P_d практически будет совпадать с величиной P_3 . При $k = 0$ величина $P_d \rightarrow 1$, а показатель достоверности результата контроля $P_{\Pi}(\Gamma/\text{ис})$ не должен быть меньше P_3 . При $k = 1$ $P_{\Pi}(\Gamma/\text{ис}) = P_d = \sqrt{P_3}$. На практике значение ко-эфицента k выбирается в пределах $k = 5 \dots 10$, так как в большинстве случаев легче повысить достоверность $P_{\Pi}(\Gamma/\text{ис})$ результатов

контроля (многократными операциями контроля, расширением пределов допуска, уменьшением величины α и β и др.), чем увеличить вероятность P_3 безотказной работы объекта контроля.

ПЕРЕЧЕНЬ ССЫЛОК

1. ГОСТ 199919-74. Контроль автоматизированный технического состояния изделий авиационной техники. Термины и определения.
2. Гмурман В.Е. Теория вероятностей и математическая статистика. - М.: Высш. шк., 1977.- 392 с.

Надійшла 18.02.2000

Після доробки 02.03.2000

УДК 621.396.6.004

ПРОВЕРКА ДИАГНОСТИРУЕМОСТИ СХЕМЫ ПУТЕМ ОПРЕДЕЛЕНИЯ РАНГА ТЕСТОВОЙ МАТРИЦЫ

Н. Н. Касьян

Предлагается в качестве меры тестируемости аналоговых схем на этапе их проектирования использовать ранг тестовой матрицы цепи, составленной для постоянного тока, временной и частотной областей с целью выбора контрольных точек и обеспечения диагностируемости схемы.

Пропонується в якості міри тестуємості аналогових схем на етапі їхнього проектування використовувати ранг тестової матриці кола, складеної для постійного струму, часової та частотної областей з метою вибору контрольних точок та забезпечення діагностуємості схеми.

It is offered as a measure testing of the analogue circuits use at a stage of their designing of a rank of a test matrix of a chain composed for a direct current, temporary and frequent areas with the purpose of a choice of breakpoints and support diagnosing of the circuit.

В методе выявления неисправностей, основанном на параметрической идентификации, результаты измерений используются для идентификации значений параметров всех элементов схемы. При этом выявление неисправных элементов производится путем классификации элементов по признаку попадания или непадения значений их параметров в пределы, устанавливаемые границами зон допусков. Считается, что топология проверяемой схемы и номинальные значения параметров известны. Схема считается диагностируемой только в том случае, если по ее характеристикам, снимаемым с внешних выводов или с доступных узлов схемы, можно однозначно определить значения каждого параметра элементов схемы.

Предполагается, что диагностируемая, в общем случае, нелинейная схема описывается следующей моделью "вход-выход":

$$\Phi = f(\mathbf{X}, \mathbf{P}), \quad (1)$$

где $\mathbf{X} = (x_1, x_2, \dots, x_m)$ - вектор входных сигналов,

$\Phi = (\varphi_1, \varphi_2, \dots, \varphi_m)$ - вектор выходных сигналов,

$\mathbf{P} = (p_1, p_2, \dots, p_n)$ - вектор параметров схемы.

Так как схема является нелинейной, любое изменение входных сигналов \mathbf{X} будет порождать новое множество выходных сигналов Φ , которое в общем случае не зависит от значений выходных сигналов до изменения. Чтобы охарактеризовать диагностируемость схемы, вводится тестовая матрица, которая зависит только от параметров схемы \mathbf{P} . Пусть $v(\mathbf{X})$ - положительная весовая функция относительно входного сигнала \mathbf{X} . Тогда тестовая матрица определяется выражением

$$\mathbf{S}(\mathbf{P}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v(\mathbf{X}) [\nabla_{\mathbf{P}} f(\mathbf{X}, \mathbf{P})] [\nabla_{\mathbf{P}} f(\mathbf{X}, \mathbf{P})]^T dx_1 dx_2 \dots dx_n, \quad (2)$$

где $\nabla_{\mathbf{P}}$ - градиенты функции f по параметрам \mathbf{P} . При этом $\mathbf{S}(\mathbf{P})$ является симметричной положительно полуопределенной матрицей. Элементы матрицы $\mathbf{S}(\mathbf{P})$ - это непрерывные функции от \mathbf{P} . Значение параметра \mathbf{P}^* является регулярной точкой матрицы $\mathbf{S}(\mathbf{P})$ [1], если существует такая открытая окрестность точки \mathbf{P}^* , в которой ранг матрицы $\mathbf{S}(\mathbf{P})$ не меняется. Значение параметра \mathbf{P}^* является диагностируемым только тогда,

когда матрица $\mathbf{S}(\mathbf{P}^*)$ положительно определена [2]. Если ранг $\mathbf{S}(\mathbf{P}^*)$ равен ρ , а $n_{\mathbf{P}}$ - число диагностируемых параметров, то мера разрешимости относительно параметрической точки \mathbf{P}^* определяется как

$$\mu(\mathbf{P}^*) = n_{\mathbf{P}} - \rho. \quad (3)$$

Пусть

$$\rho^* = \max_{\mathbf{P} \in \mathbf{R}^{n_{\mathbf{P}}}} \text{rank} \mathbf{S}(\mathbf{P}) \quad (4)$$

является общим рангом матрицы $\mathbf{S}(\mathbf{P})$. Тогда почти все \mathbf{P} , принадлежащие множеству параметров $\mathbf{R}^{n_{\mathbf{P}}}$, являются регулярными точками матрицы $\mathbf{S}(\mathbf{P})$; схема диагностируема только в том случае, когда $\rho^* = n_{\mathbf{P}}$.

Если функция $f(\mathbf{X}, \mathbf{P})$ аналитическая по \mathbf{P} , то величина ρ^* , определенная по формуле (4), является общим рангом матрицы $\mathbf{S}(\mathbf{P})$, а $\mu(\mathbf{P}^*)$ из формулы (3) имеет общее значение μ^* , представляющее собой меру диагностируемости схемы. Эта мера определяется структурой, а также расположением входов и выходов схемы. Она не зависит от выбора тестовых сигналов, алгоритма диагностирования и от значений параметров. Поэтому она может использоваться как инструмент проверки диагностируемости схемы на этапе ее проектирования для выбора контрольных точек и обеспечения диагностируемости и контролепригодности схемы.

При условии, что функция $f(\mathbf{X}, \mathbf{P})$ аналитична по \mathbf{X} и \mathbf{P} , для произвольно выбранного входного сигнала \mathbf{X}_i , $i = 1, 2, \dots, n_{\mathbf{P}}$, и случайно выбранного значения параметра \mathbf{P}^* практически достаточно проверить ранг матрицы

$$\mathbf{S}(\mathbf{P}) = \sum_{i=1}^{n_{\mathbf{P}}} [\nabla_{\mathbf{P}} f(\mathbf{X}_i, \mathbf{P})][\nabla_{\mathbf{P}} f(\mathbf{X}_i, \mathbf{P})]^T. \quad (5)$$

Если ранг матрицы (5) равен $n_{\mathbf{P}}$, то схема является диагностируемой и любые $n_{\mathbf{P}}$ произвольно выбранных входов можно использовать для однозначного решения уравнений относительно параметров элементов схемы.

Диагностирование нелинейных схем во временной области основано на том, что входные и выходные сигналы могут рассматриваться как элементы бесконечномерного гильбертова пространства, и являются кусочно-непрерывными функциями времени [1]. Тогда тестовую матрицу можно определить как

$$\mathbf{S}(\mathbf{P}) = \sum_{i=10}^{n_x T} \int [\nabla_{\mathbf{P}} f(\mathbf{X}_i(t), \mathbf{P})][\nabla_{\mathbf{P}} f(\mathbf{X}_i(t), \mathbf{P})]^T dt, \quad (6)$$

где n_x - количество входных (выходных) временных характеристик.

В качестве примера рассмотрим схему, приведенную на рисунке 1. Она содержит нелинейный и линейные элементы. Токи в ветвях схемы: $i_{\text{д}}(t) = \mathbf{I}_S(e^{\alpha u(t)} - 1)$, где \mathbf{I}_S - ток насыщения диода, α - диодная постоянная,

$$i_{\text{R}}(t) = \frac{1}{R}u(t), \quad i_{\text{L}}(t) = \frac{1}{L} \int u(t) dt, \quad i_{\text{C}}(t) = C \frac{d}{dt} u(t).$$

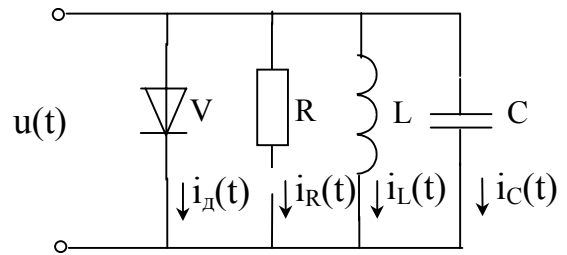


Рисунок 1 - Нелинейная цепь

Причем начальные условия - нулевые, а характеристика вход-выход имеет вид

$$i(t) = \mathbf{I}_S(e^{\alpha u(t)} - 1) + \frac{1}{R}u(t) + \frac{1}{L} \int u(t) dt + C \frac{d}{dt} u(t).$$

При $u(t) = t, \forall t \in [0, 1]$ и значениях параметров $\mathbf{P}^* = [\mathbf{I}_S, \alpha, R, L, C]^T = [1, 1, 1, 1, 1]^T$ находим

$$\begin{aligned} [\nabla_{\mathbf{P}} i(t)]^T &= \left[\frac{d}{d\mathbf{I}_S} i(t), \frac{d}{d\alpha} i(t), \frac{d}{dR} i(t), \frac{d}{dL} i(t), \frac{d}{dC} i(t) \right] = \\ &= \left[e^{\alpha u(t)} - 1, \mathbf{I}_S u(t) e^{\alpha u(t)}, -\frac{1}{R^2} u(t), -\frac{1}{L^2} \int u(t) dt, \frac{d}{dt} u(t) \right] = \\ &= \left[e^t - 1, te^t, -t, -\frac{t^2}{2}, 1 \right]. \end{aligned}$$

Таким образом выражение (6) приобретает вид

$$\mathbf{S}(\mathbf{P}, \mathbf{X}(t)) = \int_0^1 \begin{bmatrix} (e^t - 1)^2 & te^t(e^t - 1) - t(e^t - 1) \left(-\frac{t^2}{2}\right)(e^t - 1) & e^t - 1 \\ te^t(e^t - 1) & t^2 e^{2t} & -t^2 e^t & \left(-\frac{t^3}{2}\right)e^t & te^t \\ -t(e^t - 1) & -t^2 e^t & t^2 & \frac{t^3}{2} & -t \\ \left(-\frac{t^2}{2}\right)(e^t - 1) & \left(-\frac{t^3}{2}\right)e^t & \frac{t^3}{2} & \frac{t^4}{4} & -\frac{t^2}{2} \\ e^t - 1 & te^t & -t & -\frac{t^2}{2} & 1 \end{bmatrix} dt =$$

$$= \begin{bmatrix} 0,75786 & 1,09726 & -0,5 & -0,19247 & 0,718 \\ 1,09726 & 1,59726 & -0,7182 & 0,28172 & 1 \\ -0,5 & -0,7182 & 0,3333 & 0,125 & -0,5 \\ -0,19247 & 0,28172 & 0,125 & 0,05 & -0,167 \\ 0,718 & 1 & -0,5 & -0,167 & 1 \end{bmatrix}.$$

Отсюда следует, что матрица не вырождена. Так как функция вход-выход аналитична по \mathbf{P} , то цепь диагностируема.

Для цепей размера, представляющего интерес для практики, трудно получить явную форму записи вход-выход, поэтому каждая характеристика представляется определенным числом отсчетов во временной области (например, в соответствии с теоремой Котельникова), с учетом чего тестовую матрицу можно представить в виде

$$\mathbf{S}(\mathbf{P}, \mathbf{X}(t_i)) = \sum_{i=1}^{m_0} \sum_{j=1}^{n_x} [\nabla_{\mathbf{P}} f(\mathbf{X}_j(t_i), \mathbf{P})][\nabla_{\mathbf{P}} f(\mathbf{X}_j(t_i), \mathbf{P})]^T, \quad (7)$$

где n_x - количество характеристик вход-выход,

m_0 - количество отсчетов на j -й временной характеристике.

Такая цепь будет диагностируемой, если ранг матрицы $\mathbf{S}(\mathbf{P}, \mathbf{X}(t_i))$ равен количеству диагностируемых параметров.

Динамическая цепь может быть проверена с помощью меньшего чем число параметров числа выходных характеристик по сравнению с диагностированием по постоянному току, так как в этом случае общее количество выходных характеристик дополняется зависимостями вход-выход в дискретные моменты времени. При этом общее число независимых уравнений должно быть не менее числа диагностируемых параметров схемы. Для избежания трудностей при численных расчетах и уменьшения чувствительности к ошибкам, обусловленным погрешностями измерений выходных характеристик суммарное число характеристик и отсчетов на них необходимо выбирать больше числа диагностируемых параметров.

Для диагностирования линейных схем в частотной области тестовую матрицу можно записать

$$\mathbf{S}(\mathbf{P}, \mathbf{X}(\omega_i)) = \sum_{i=1}^{m_0} \sum_{j=1}^{n_x} [\nabla_{\mathbf{P}} f(\mathbf{X}_j(\omega_i), \mathbf{P})][\nabla_{\mathbf{P}} f(\mathbf{X}_j(\omega_i), \mathbf{P})]^T, \quad (8)$$

где n_x - количество частотных характеристик вход-выход,

m_0 - количество отсчетов на j -й частотной характеристике.

Невырожденность матрицы $\mathbf{S}(\mathbf{P}, \mathbf{X}(\omega_i))$ означает, что схема по рассматриваемым параметрам является диагно-

стируемой при выбранных частотных характеристиках и отсчетах на них.

Например, рассмотрим схему фильтра, приведенную на рисунке 2. Если контролируемым выходным сигналом является напряжение \mathbf{u}_2 , то

$$K(j\omega) = \frac{\dot{\mathbf{u}}_2}{\dot{\mathbf{u}}_1} = \frac{R_2}{R_1 + R_2 + j\omega CR_1 R_2};$$

$$\begin{aligned} \nabla_{\mathbf{P}}^T K(j\omega) &= \left[\frac{\partial}{\partial R_1} K(j\omega), \frac{\partial}{\partial R_2} K(j\omega), \frac{\partial}{\partial C} K(j\omega) \right] = \\ &= \left[\frac{R_2(1 + j\omega CR_2)}{(R_1 + R_2 + j\omega CR_1 R_2)^2}, \frac{R_1}{(R_1 + R_2 + j\omega CR_1 R_2)^2}, \right. \\ &\quad \left. \frac{j\omega R_1 R_2^2}{(R_1 + R_2 + j\omega CR_1 R_2)^2} \right]. \end{aligned}$$

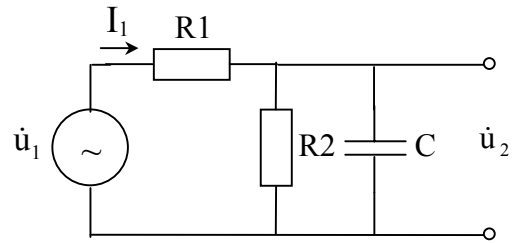


Рисунок 2 - Схема фильтра

Для $\mathbf{P}^* = [1, 1, 1]^T$ получаем матрицу

$$\nabla_{\mathbf{P}}^T K(j\omega) = \left[\frac{1 + j\omega}{(2 + j\omega)^2}, \frac{1}{(2 + j\omega)^2}, -\frac{j\omega}{(2 + j\omega)^2} \right],$$

которая имеет два независимых столбца, и следовательно, мера тестируемости $\mu^* = 1$, т.е. при использовании только выходного напряжения \mathbf{u}_2 цепь не является диагностируемой.

Если измеряемым сигналом служит входной ток $\dot{\mathbf{I}}_1$, то

$$Y_1(j\omega) = \frac{\dot{\mathbf{I}}_1}{\dot{\mathbf{u}}_1} = \frac{j\omega CR_2 + 1}{R_1 + R_2 + j\omega CR_1 R_2};$$

$$\begin{aligned} \nabla_{\mathbf{P}}^T Y_1(j\omega) &= \left[\frac{(1 + j\omega CR_2)^2}{(R_1 + R_2 + j\omega CR_1 R_2)^2}, \right. \\ &\quad \frac{j\omega C(R_1 + R_2 + j\omega CR_1 R_2) - (1 + j\omega CR_1)^2}{(R_1 + R_2 + j\omega CR_1 R_2)^2}, \\ &\quad \left. \frac{j\omega R_2(R_1 + R_2 + j\omega CR_1 R_2) - (1 + j\omega CR_1)j\omega CR_1 R_2}{(R_1 + R_2 + j\omega CR_1 R_2)^2} \right]. \end{aligned}$$

При $\mathbf{P}^* = [1, 1, 1]^T$ получаем

$$\nabla_P^T(Y(j\omega)) = \left[-\frac{(1+j\omega)^2}{(2+j\omega)^2}, \frac{1}{(2+j\omega)^2}, \frac{j\omega}{(2+j\omega)^2} \right].$$

Полученная матрица имеет три независимых столбца над полем комплексных чисел. Следовательно, мера тестируемости $\mu^* = 0$, т.е. при использовании выходного сигнала \dot{I}_1 цепь является диагностируемой. Измеряя значение $Y(j\omega)$ при трех независимых частотах, можно определить значения параметров R_1 , R_2 и C .

Таким образом, рассмотренный критерий оценки диагностируемости аналоговых цепей позволяет выбирать достаточное количество характеристик и отсчетов

на них во временной и частотной областях для однозначного определения значений параметров элементов схемы, а также эффективные тестовые воздействия.

ПЕРЕЧЕНЬ ССЫЛОК

1. R. Saeks, A. Sangiovanni-Vincentelli and V. Visvanathan "Diagnosability of nonlinear circuits and systems - Part II: Dynamical Systems", IEEE Trans. Circuits Syst., vol. CAS-28, pp. 1103-1108, 1981.
2. V. Visvanathan and A. Sangiovanni-Vincentelli, "Diagnosability of nonlinear circuits and systems - Part I: The dc case," IEEE Trans. Circuits Syst., vol. CAS-28, pp. 1093-1102. 1981.

Надійшла 18.02.2000

Після доробки 02.03.2000

УДК 621.37

МЕТОД СИНТЕЗА ТЕПЛОВЫХ И МЕХАНИЧЕСКИХ МОДЕЛЕЙ РАДИОЭЛЕКТРОННЫХ СРЕДСТВ ПРИ ПОМОЩИ ЭКСПЕРТНОЙ СИСТЕМЫ

Ю. Н. Кофанов, А. В. Сарафанов

Рассматривается метод формирования моделей РЭС на основе экспертной системы, осуществляющей формирование карты с управляющей информацией, используемой подсистемами проектирования РЭС.

The method of creation of ratio electronics schemes on the basis of the expert system that formed map with the managing information used by subsystems of ratio electronics schemes designing is considered.

В современных пакетах прикладных программ (ППП) и подсистемах по моделированию тепловых и механических режимов радиоэлектронных средств (РЭС) наряду с вводом непосредственно топологии и параметров моделей применяются специальные модули-предпроцессоры, выполняющие операции по автоматическому синтезу сеточных моделей анализируемых объектов на основе их формализованного описания (множества геометрических и тепло-физикомеханических параметров). Такие модули позволяют автоматически синтезировать модели наиболее типовых конструкций РЭС (стойки, каскетные и этажерочные конструкции, печатные узлы, микросборки и т.п.), описание которых хорошо формализуется и может быть реализовано в специализированных программах графического ввода данных. Как правило, для каждой типовой конструкции разрабатывается свой программный комплекс по графическому вводу и набор модулей-предпроцессоров, осуществляющих синтез структуры и параметров модели. Переключение в ППП или в подсистемах на той или иной тип модели (конструкции РЭС) выполняется пользователем при помощи управляющей информации в

соответствии с используемой им методикой моделирования физических процессов. Описанный выше подход к синтезу моделей имеет ряд существенных недостатков:

1. Модулями-предпроцессорами синтезируются жесткие по своей структуре модели, в которых, в ряде случаев (например, при проектировании самолетной и космической радиоаппаратуры), невозможно отразить специфику конструкторских решений проектировщика, имеющих принципиальное значение с позиций обеспечения нормального того или иного физического процесса.

2. В подавляющем своем большинстве настройка модуля-предпроцессора на особенности той или иной типовой конструкции РЭС может осуществляться путем доработки программного обеспечения, а не внешней настройки.

3. В большинстве случаев отсутствует возможность гибко влиять на размерность математической модели (представляет собой систему нелинейных уравнений 1000-4000 порядка), которая, как правило, определяется целью проводимого моделирования и определенным набором физических параметров моделируемого объекта.

В статье рассматривается метод формирования тепловых и механических моделей РЭС, основанный на применении экспертной системы (ЭС). В предлагаемом методе экспертная система, функционирующая совместно с программой графического описания объекта моделирования, осуществляет формирование специальной

карты с управляющей информацией, которая отдельно вводится и используется математическим ядром модифицированной подсистемы теплового проектирования РЭС [1] и ППП по анализу механических характеристик [2]. Структурная схема метода синтеза моделей приведена на рисунке.

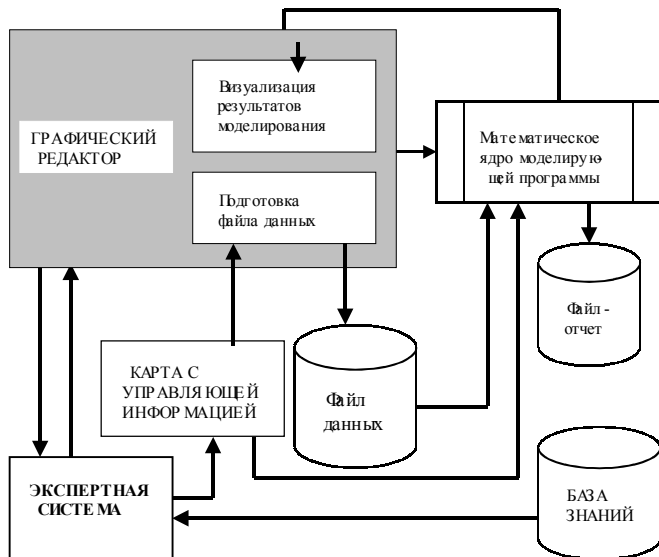


Рисунок - Схема синтеза моделей при помощи ЭС

Карта с управляющей информацией формируется ЭС (на основе специальной базы знаний) в интерактивном режиме, в процессе которого распознаются конструктивные особенности объекта моделирования и цель проведения сеанса моделирования (определение температур активных зон электрорадиоэлементов (ЭРЭ), расчет коэффициентов динамичности ЭРЭ, определение интегральной температуры объекта, вычисление резонансных частот конструкции, вычисление амплитудно-частотных характеристик конструктивных узлов, определение температур определенных ЭРЭ или виброускорений определенных ЭРЭ для дальнейшего их использования в процессе анализа электрических или надежностных характеристик, определение теплового поля конструкции РЭС для его дальнейшего использования в процессе моделирования механических характеристик разрабатываемой конструкции и т.д.). В базах знаний (для наглядности строятся по системе продукций), используемой ЭС декларируются характерные признаки множества типовых узлов, блоков, микроблоков и набора конструкций РЭС в целом. Используя множество характерных признаков, ЭС распознает (в режиме консультации с пользователем) конструкцию анализируемого объекта. В результате этого процесса ЭС определяется набор типовых конструкций РЭС, для которых в программном обеспечении по моделированию физических процессов есть разработанные модуль-предпроцессоры. Кроме этого, через карту с управляющей информацией осуществляется настройка модуль-предпроцессоров на структуру, размерность и парамет-

ры моделей, которые определяются целью проводимого моделирования и конструктивными особенностями объекта. В случае отсутствия необходимых модуль-предпроцессоров (для генерации модели определенного узла) ЭС предлагает пользователю подключить свою типовую модель (или макро модель), которая описывается на специальном формализованном языке и хранится в отдельном файле. При этом графический редактор ввода данных, в специальном режиме работы, отражает такую модель на общей структурной схеме модели объекта в виде многополюсника. ЭС также решается задача (дальнейшая настройка карты с управляющей информацией) по организации иерархического алгоритма моделирования физических процессов объекта или по использованию подхода смешанной иерархии (детальная модель определенного конструктивного узла более низкого уровня конструктивной иерархии интегрируется с макро моделью конструкции более высокого уровня иерархии).

Следует также отметить, что для реализации описанного подхода, в графический редактор по аналогии с математическим ядром встраивается анализатор карты с управляющей информацией (см. рисунок). Анализатор необходим для настройки процесса графического ввода информации, описывающей объект моделирования.

Применяемая в предлагаемом подходе экспертная система может быть снабжена другой базой знаний, в которой, например, может быть реализована специфическая методика математического моделирования физических процессов, протекающих в исследуемом объекте.

Предлагаемый выше подход позволяет при помощи отдельной базы знаний реализовать полностью процесс автоматического синтеза тепловых и механических моделей сложного объекта, используя при этом карту с управляющей информацией для настройки имеющихся модуль-предпроцессоров, а также для объединения, генерируемых ими моделей. В данном случае карта с управляющей информацией содержит определенные команды, при помощи которых карта представляет собой генеральный алгоритм синтеза той или иной сеточной модели.

Рассмотренный в статье подход позволяет достаточно гибко синтезировать тепловые и механические модели, управлять их структурой и параметрами, сведя при этом к минимуму эвристические операции, присутствующие в методиках математического моделирования физических процессов в РЭС.

ПЕРЕЧЕНЬ ССЫЛОК

1. А. В. Сарафанов. Структурная организация подсистемы моделирования тепловых характеристик РЭС // Вестник Красноярского государственного технического университета. Вып. 4: Сборник научных трудов / КГТУ. Красноярск, 1996. С. 37-42.
2. Ю. Н. Кованов, А. С. Шалумов. Обеспечение надежности аппаратуры при механических воздействиях с применением САПР: Учебное пособие. - Ковров: Изд-во КТИ, 1995.-

РАСЧЕТ КОЭФФИЦИЕНТА ЗАТУХАНИЯ В НАПРАВЛЯЮЩИХ СИСТЕМАХ С ФРАКТАЛЬНЫМИ СВОЙСТВАМИ ГРАНИЦ

В. М. Онуфриенко, П. А. Самолчев, Т. И. Слюсарова

Представлены результаты дифференциального моделирования потерь в направляющих системах с фрактальными свойствами. Исследование базируется на фрактальных представлениях о структуре поверхности направляющей системы. Полученная модель расчета потерь в направляющих системах позволяет описать с использованием α -характеристик любую заданную степень фрактальности поверхности. Для 0-характеристик результаты совпадают с классическими данными.

Наводяться результати диференціального моделювання втрат у напрямних системах з фрактальними властивостями меж. Дослідження базується на фрактальному уявленні про структуру поверхні напрямної системи. Отримана модель розрахунку втрат в напрямних системах дозволяє описати з використанням α -характеристик будь-який заданий ступінь фрактальності поверхні. Для 0-характеристик результати збігаються з класичними даними.

The results of differintegrational modelling of losses in directing systems by fractional properties of borders are submitted. The research is based on fractional performances about structure of a surface of directing system. The received model of account of losses in directing systems 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

Хорошо известны результаты классической макроскопической электродинамики о влиянии границы среды с разными свойствами на структуру электромагнитного поля и характер распространения волн. Вследствие взаимодействия волны с поверхностью раздела сред проанализированы возможные явления отражения (частичного и полного), прохождения волн в среду, явления дифракции и т. п. Однако с помощью современных технологий обнаружен ряд несоответствий теоретических моделей относительно взаимодействия электромагнитных волн с реальными поверхностями раздела сред (см., напр., [1],[2]).

Введение α -характеристик компонент электромагнитного поля дает возможность учета неровностей реальных поверхностей и структуры среды [3].

Далее приведено построение интегродифференциальной модели электромагнитного поля, которая используется для решения задач о передаче энергии в направляющих системах с фрактальными свойствами границ.

ПОСТАНОВКА ЗАДАЧИ

Алгоритм ввода α -характеристик электромагнитного поля в условиях задачи о взаимодействии волн с фрактальными границами раздела двух сред рассмотрен в [4].

Геометрическая основа ввода α -характеристик базируется на идее Ф.Хаусдорфа об определении непрерывности отображения в точке через покрытие и измерение окрестности $O_\epsilon: \rho(x, x_0) < \epsilon$ точки.

В связи с этим применяется обобщение меры величины множества, связанное с выбором некоторой пробной степенной функции [5]

$$h(\epsilon) = \gamma(v) \times \epsilon^v$$

(геометрически: отрезок прямой, квадрат, круг, шар, куб) и покрытием рассматриваемого множества точек с конструированием v -меры Хаусдорфа

$$M_v = \sum h(\epsilon).$$

Вместо сложной процедуры геометрического построения фрактального множества каждый раз, нахождения меры Хаусдорфа с последующими граничными переходами, будем использовать аппарат дробного интегрирования и дифференцирования (см., напр., [6], [7])

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

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

где $\Gamma(\alpha)$ - гамма-функция Эйлера, α - порядок дробного интегродифференциала (скейлинговый показатель).

Анализ подынтегрального выражения указывает на возможные случаи описания явлений с помощью α -характеристик: при наличии геометрических особенностей контуров и поверхностей (геометрические сингулярности) или когда физические параметры (проницаемость, проводимость, плотность зарядов, токов и т.п.) имеют фрактальную природу [4].

Введем уравнения Максвелла с учетом α -характеристик электромагнитного поля, с последующей записью теоремы Умова-Пойнтинга. Исходим из уравнений Максвелла в дифференциальной форме:

$$\begin{aligned} \text{rot} D^\alpha \vec{H}^\alpha &= \frac{\partial}{\partial t} \vec{D}^\alpha + \vec{j}_\alpha, \\ \text{rot} D^\alpha \vec{E}^\alpha &= \frac{\partial}{\partial t} \vec{B}^\alpha + \vec{j}_M, \\ \text{div} \vec{D}^\alpha &= \rho_\alpha, \quad \text{div} \vec{B}^\alpha = \rho_M, \end{aligned} \quad (1)$$

где

$$\begin{aligned} \vec{j}_\alpha &= \sigma_\alpha (D^\alpha \vec{E}^\alpha + D^\alpha \vec{E}_{\text{ст}}^\alpha), \\ \vec{j}_M &= \sigma_M (D^\alpha \vec{H}^\alpha + D^\alpha \vec{H}_{\text{ст}}^\alpha), \\ \vec{D}^\alpha &= \epsilon_a \cdot D^\alpha \vec{E}^\alpha, \quad \vec{B}^\alpha = \mu_a \cdot D^\alpha \vec{H}^\alpha, \end{aligned} \quad (2)$$

σ_α и σ_M - удельные проводимости (электрическая и магнитная, соответственно), ϵ_a - относительная диэлектрическая проницаемость, μ_a - относительная магнитная проницаемость, индекс "ст" указывает на наличие сторонних источников тока.

Подставляя (2) в (1), получим уравнения Максвелла в следующем виде:

$$\begin{aligned} \text{rot} D^\alpha \vec{H}^\alpha &= \epsilon_a \frac{\partial}{\partial t} (D^\alpha \vec{E}^\alpha) + \sigma_\alpha D^\alpha \vec{E}^\alpha + D^\alpha \vec{j}_{\text{эст}}^\alpha, \\ \text{rot} D^\alpha \vec{E}^\alpha &= -\mu_a \frac{\partial}{\partial t} (D^\alpha \vec{H}^\alpha) - \sigma_M D^\alpha \vec{H}^\alpha - D^\alpha \vec{j}_{\text{мст}}^\alpha, \\ \text{div} (\epsilon_a \cdot D^\alpha \vec{E}^\alpha) &= \rho_\alpha, \quad \text{div} (\mu_a \cdot D^\alpha \vec{H}^\alpha) = \rho_M. \end{aligned}$$

Для комплексных амплитуд уравнения Максвелла принимают вид:

$$\begin{aligned} \text{rot} D^\alpha \vec{H}^\alpha &= i\omega \epsilon_a \frac{\partial}{\partial t} (D^\alpha \vec{E}^\alpha) + \sigma_\alpha D^\alpha \vec{E}^\alpha + D^\alpha \vec{j}_{\text{эст}}^\alpha, \\ \text{rot} D^\alpha \vec{E}^\alpha &= -i\omega \mu_a \frac{\partial}{\partial t} (D^\alpha \vec{H}^\alpha) - \sigma_M D^\alpha \vec{H}^\alpha - D^\alpha \vec{j}_{\text{мст}}^\alpha, \\ \text{div} (\epsilon_a \cdot D^\alpha \vec{E}^\alpha) &= \dot{\rho}_\alpha, \quad \text{div} (\mu_a \cdot D^\alpha \vec{H}^\alpha) = \dot{\rho}_M. \end{aligned}$$

Теорема Умова-Пойнтинга о балансе мощности электромагнитного поля в терминах α -характеристик выглядит следующим образом:

$$\begin{aligned} &-\frac{1}{2} \int_V \left(D^\alpha \vec{j}_{\text{эст}}^\alpha \cdot D^\alpha \vec{E}^\alpha + D^\alpha \vec{j}_{\text{мст}}^\alpha \cdot D^\alpha \vec{H}^\alpha \right) dV^\alpha = \\ &= \frac{i\omega}{2} \cdot \int_V \left(\mu_a \cdot D^\alpha \vec{H}^\alpha \cdot D^\alpha \vec{H}^\alpha - \epsilon_a \cdot D^\alpha \vec{E}^\alpha \cdot D^\alpha \vec{E}^\alpha \right) dV^\alpha + \\ &+ \frac{1}{2} \int_V \left(\sigma_\alpha \cdot D^\alpha \vec{E}^\alpha \cdot D^\alpha \vec{E}^\alpha + \sigma_M \cdot D^\alpha \vec{H}^\alpha \cdot D^\alpha \vec{H}^\alpha \right) dV^\alpha + \\ &+ \frac{1}{2} \int_S \left(\vec{\Pi}^\alpha, \vec{n} \right) dS^\alpha, \\ &\text{где } \vec{\Pi}^\alpha = \left[D^\alpha \vec{E}^\alpha, D^\alpha \vec{H}^\alpha \right] - \text{вектор Пойнтинга.} \end{aligned}$$

МОДЕЛИРОВАНИЕ ПОТЕРЬ В НАПРАВЛЯЮЩИХ СИСТЕМАХ

Поток среднего вектора Пойнтинга через поперечное фрактальное сечение S_\perp направляющей системы через α -характеристики запишем в виде

$$\bar{P}^\alpha = \int_{S_\perp} \vec{\Pi}^\alpha \cdot d\vec{s}^\alpha = \frac{1}{2} \text{Re} \int_{S_\perp} \left[D^\alpha \vec{E}_m^\alpha, D^\alpha \vec{H}_m^{\alpha*} \right] ds^\alpha \quad (3)$$

(в случае открытой системы область S_\perp бесконечна). Отсюда для систем с однородной фрактальной средой:

$$\begin{aligned} \bar{P}^\alpha &= \frac{1}{2} \text{Re} \dot{W}^{(E,H)} \int_{S_\perp} \left[D^\alpha \vec{H}_{mt}^\alpha \right]^2 ds^\alpha = \\ &= \frac{1}{2} \text{Re} \frac{1}{\dot{W}^{(E,H)*}} \int_{S_\perp} \left[D^\alpha \vec{E}_{mt}^\alpha \right]^2 ds^\alpha, \end{aligned}$$

где $\dot{W}^{(E,H)}$ - волновое сопротивление W , W^E или W^H (в зависимости от типа волны E или H), \vec{H}_{mt}^α и \vec{E}_{mt}^α - поперечные компоненты векторов \vec{E}^α и \vec{H}^α ; при отсутствии поглощения величина $\dot{W}^{(E,H)}$ либо чисто вещественна, и тогда символ Re не нужен, либо - при частотах ниже критической - оказывается чисто мнимой величиной, так что $\bar{P}^\alpha = 0$.

Поглощение энергии в направляющей системе приводит к комплексности постоянной распространения. Передаваемая мощность \bar{P}^α ввиду (3) изменяется вдоль оси z по закону $e^{-2\Gamma^\alpha z}$, следовательно

$$\bar{P}^\alpha(z) = \bar{P}^\alpha(0) \cdot e^{-2\Gamma^\alpha z}.$$

Убывание мощности волны при прохождении весьма малого участка Δz направляющей системы

$$\Delta \bar{P}^\alpha = -\frac{d}{dz} \bar{P}^\alpha \Delta z + \dots = 2\Gamma^\alpha \bar{P}^\alpha \Delta z + \dots \quad (4)$$

(в пренебрежении членами высшего порядка малости). Аналогично выразим мощность потерь в системе на участке Δz

$$\Delta \bar{P}_n^\alpha = \frac{d}{dz} \bar{P}_n^\alpha \Delta z + \dots \quad (5)$$

При $\Delta z \rightarrow 0$ оба равенства (4) и (5) становятся точными, а в силу закона сохранения энергии $\Delta \bar{P}^\alpha = \Delta \bar{P}_n^\alpha$. Отсюда

$$\Gamma^\alpha = \frac{1}{2\bar{P}^\alpha} \frac{d}{dz} \bar{P}_n^\alpha \quad (6)$$

(энергетическая трактовка затухания).

Формулу (6) нетрудно конкретизировать. Знаменатель при этом вычисляется на основании выражения (3). Для определения числителя используем равенство

$$\begin{aligned} \bar{P}_n^\alpha = & \\ = & \frac{\omega}{2} \int_V \left(\epsilon_0 \cdot \epsilon \cdot D^\alpha E_m \cdot D^\alpha E_m^* + \right. \\ & \left. + \mu_0 \cdot \mu \cdot D^\alpha H_m \cdot D^\alpha H_m^* \right) dV^\alpha \quad (7) \end{aligned}$$

рассматривая область $V = S_\perp \Delta z$, где S_\perp - поперечное фрактальное сечение направляющей системы; обе части (7) делятся на Δz и совершается переход к пределу при $\Delta z \rightarrow 0$. В результате получаем

$$\begin{aligned} \Gamma^\alpha = & \\ = & \frac{\omega \int_{S_\perp} \left(\epsilon_0 \cdot \epsilon \cdot D^\alpha E_m \cdot D^\alpha E_m^* + \mu_0 \cdot \mu \cdot D^\alpha H_m \cdot D^\alpha H_m^* \right) ds^\alpha}{2 \operatorname{Re} \int_{S_\perp} \left[D^\alpha E_m, D^\alpha H_m \right] ds^\alpha} \quad (8) \end{aligned}$$

Здесь ϵ , μ - функции координат, принимающие различные значения в областях, соответствующих диэлектрику и проводнику.

Потери энергии в направляющей системе можно представить в виде суммы, слагаемые которой соответствуют разным физическим факторам. Учитывая это в (6), имеем

$$\Gamma^\alpha = \Gamma_{\alpha_1}^\alpha + \Gamma_{\alpha_2}^\alpha + \dots = \frac{1}{2\bar{P}^\alpha} \left(\frac{d}{dz} \bar{P}_{n_1}^\alpha + \frac{d}{dz} \bar{P}_{n_2}^\alpha \right). \quad (9)$$

В частности, оказывается удобным отдельно рассмотреть

затухание в диэлектрике и проводнике. Формула (8), например, может быть использована только для нахождения "частичного" коэффициента затухания $\Gamma_{\text{д}}^\alpha = \Gamma_{\text{д}}^\alpha$, учитывающего потери в диэлектрике с фрактальной структурой

$$\begin{aligned} \Gamma_{\text{д}}^\alpha = & \\ = & \frac{\omega \epsilon_0 \int_{S_\perp \text{д}} \epsilon \cdot D^\alpha E_m \cdot D^\alpha E_m^* ds^\alpha}{2 \operatorname{Re} \int_{S_\perp} \left[D^\alpha E_m, D^\alpha H_m \right] ds^\alpha} \quad (10) \end{aligned}$$

($S_\perp \text{д}$ - поперечное фрактальное сечение диэлектрика системы).

Влияние проводника обычно учитывают с позиций граничного условия Леонтовича. Выделив на проводящих элементах направляющей системы пояс ΔS шириной Δz , с учетом

$$\bar{P}_n^\alpha = \operatorname{Re} \oint_S \vec{H}_{m\tau} \cdot d\vec{s}^\alpha = \frac{1}{2\sigma \Delta^0} \oint_S \left(D^\alpha H_{m\tau} \right)^2 ds^\alpha,$$

где $\vec{H}_{m\tau}$ - тангенциальная компонента вектора \vec{H} , Δ^0 - глубина скин-слоя, σ - удельная проводимость проводника; запишем

$$\frac{d}{dz} \bar{P}_{np}^\alpha = \lim_{\Delta z \rightarrow 0} \frac{1}{\Delta z} \frac{1}{2\sigma \Delta^0} \int_{\Delta S} \left(D^\alpha H_{m\tau} \right)^2 ds^\alpha = \frac{1}{2\sigma \Delta^0} \int_L \left(D^\alpha H_m \right)^2 dl^\alpha.$$

Таким образом, коэффициент $\Gamma_{\text{пр}}^\alpha = \Gamma_{\text{пр}}^\alpha$, отражающий неидеальность проводника (фрактальную структуру поверхности), получаем в виде

$$\Gamma_{\text{пр}}^\alpha = \frac{\int_L \left(D^\alpha H_m \right)^2 dl^\alpha}{2\sigma \Delta^0 \operatorname{Re} \int_{S_\perp} \left[D^\alpha E_m, D^\alpha H_m \right] ds^\alpha} \quad (11)$$

с учетом (9) $\Gamma^\alpha = \Gamma_{\text{д}}^\alpha + \Gamma_{\text{пр}}^\alpha$, где слагаемые находятся по формулам (10) и (11).

Анализ структуры поля в линиях передачи, сделанный в предположении идеальной проводимости стенок волновода, неточен для реальных линий с конечной проводимостью стенок. Однако если проводимость стенок конечна, но весьма велика (что обычно имеет место), то действительная структура поля мало отличается от структуры поля, полученной в предположении

идеальной проводимости стенок. Отличие в основном сводится к тому, что в соответствии с граничным условием Леонтовича-Щукина вдоль стенок волновода появляется некоторая весьма малая тангенциальная составляющая вектора напряженности электрического поля. Например, действительная структура электрического поля волны типа H_{10} в волноводе прямоугольного сечения, благодаря конечной проводимости стенок волновода, имеет вид, при котором электрические силовые линии вблизи стенок волновода несколько наклонены в направлении распространения, в то время как согласно классическому анализу (при $\alpha = 0$) они должны быть строго нормальны к поверхностям стенок. Наклон силовых линий весьма мал, однако учесть деформацию силовых линий можно с помощью α -характеристик [8], [9]. Аналогичные изменения структуры поля происходят и при распространении волн других типов.

Изменению структуры электрических силовых линий соответствует изменение структуры магнитных силовых линий. В частности, у стенок волновода нормальная составляющая вектора \vec{H} не равна нулю. Изменение структуры токов в основном сводится к тому, что они в действительности текут не по поверхности, а проникают на некоторую глубину внутрь стенок. Наличие отличной от нуля тангенциальной составляющей векторов \vec{E} и \vec{H} у стенок волновода означает, что вектор Пойнтинга имеет составляющую, нормальную стенкам волновода, и, следовательно, в них имеются потери на тепло.

Выделим на поверхности металлического проводника направляющей системы участок длиной Δz . Средняя мощность тепловых потерь $\Delta \bar{P}_{n_{cp}}^\alpha$ на отрезке проводника длиной Δz равна

$$\Delta \bar{P}_{n_{cp}}^\alpha = \frac{1}{2\sigma\Delta^0} \int_L^{\Delta z} dz \oint_L |D^\alpha H_{m\tau}|^2 d\alpha = \frac{\Delta z}{2\sigma\Delta^0} \oint_L |D^\alpha H_{m\tau}|^2 d\alpha,$$

т. е. потери на единицу длины

$$\bar{P}_{n_{cp}}^\alpha = \frac{\Delta \bar{P}_{n_{cp}}^\alpha}{\Delta z} = \frac{1}{2\sigma\Delta^0} \oint_L |D^\alpha H_{m\tau}|^2 d\alpha. \quad (12)$$

Подставляя (12) в (9), получаем

$$\Gamma^\alpha = \frac{R_S}{4\bar{P}_{n_{cp}}^\alpha} \oint_L |D^\alpha H_{m\tau}|^2 d\alpha, \quad (13)$$

где через R_S обозначена активная часть поверхностного сопротивления, равная

$$R_S = \frac{1}{\sigma\Delta^0} = \sqrt{\frac{\pi f \mu_a}{\sigma}}.$$

Так как для всех металлов, кроме ферромагнитных, $\mu_a \approx \mu_0$, то

$$R_S \approx 1,987 \sqrt{\frac{f}{\sigma}} \cdot 10^{-3} \text{ ом},$$

где f - частота в герцах.

Как следует из равенства (13), коэффициент затухания убывает прямо пропорционально величине поверхностного сопротивления R_S . Неизбежная шероховатость поверхности металла заметно увеличивает затухание в проводниках. В сантиметровом диапазоне волн глубина проникновения тока в металл не превышает 0,1-0,2 мкм. В этом диапазоне неровности (сингулярности) металлических поверхностей даже после тщательной обработки и шлифовки обычно значительно больше глубины проникновения тока в металл, что приводит к увеличению пути, проходимого током по металлической поверхности, а следовательно, к возрастанию потерь. Поэтому действительные значения коэффициента затухания оказываются несколько выше рассчитанных по формуле (13) в классической трактовке (при $\alpha = 0$), однако применение α -характеристик позволяет описать любую заданную степень фрактальности (шероховатости) поверхности.

ВЫВОДЫ

Введение уравнений Максвелла с учетом α -характеристик электромагнитного поля и с последующей записью теоремы Умова-Пойнтинга позволяют исследовать поток среднего вектора Пойнтинга через поперечное фрактальное сечение S_\perp направляющей системы. В результате моделирования потерь в направляющих системах получены модели расчета потерь с использованием α -характеристик, которые позволяют описать любую заданную степень фрактальности (шероховатости) поверхности, при этом действительные значения коэффициента затухания оказываются несколько выше рассчитанных по формулам при классической трактовке затухания.

Полученные результаты могут применяться для анализа и проектирования волноведущих структур различного типа. Введенные в рассмотрение α -характеристики электромагнитного поля, построенные на основе аппарата дробного дифференцирования, могут применяться для анализа взаимодействия электромагнитных волн миллиметрового и субмиллиметрового диапазона при падении на фрактальную поверхность под любым углом, для решения задач анализа явлений отражения (частичного и полного), прохождения волн в среду, явления дифракции и т. п.

ПЕРЕЧЕНЬ ССЫЛОК

1. Никольский В.В. Электродинамика и распространение радиоволн. - М.: Наука, 1973. - 608 с.
2. Ацюковский В.А. Общая эфиродинамика. - М.: Энергоатомиздат, 1990. - 280с.
3. Onufrienko V. On " α -features" of electrical waves above impedance plane// Conference Proceedings 12 International Conference on Microwaves & Radar. - Vol.1. - Krakov (Poland). - 1998. - P.212-215.
4. Онуфрієнко В.М. Фізико-геометрична інтерпретація α -характеристик електромагнітного поля // Радиофизика и электроника. - Харьков : Ин-т радиофизики и электроники НАН Украины . -1999. - 4, № 1. -С. 7-10.
5. Фракталы в физике// Труды 6-го международного симпозиума по фракталам в физике (Триест, Италия, 9-12 июля, 1985 год):Пер. с англ./ Под ред. А. Пьетронеро, Э. Тозотти. -М.: Мир. - 1988. - 672 с.
6. Oldham K., Spanier J. The fractional calculus. N.York and London: Academic Press. - 1974.
7. 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.
8. Онуфрієнко В.М., Самолчев П.А., Слюсарова Т.И. Деформация однородного поля фрактальным цилиндрическим проводником// Труды 9-ой Международной Крымской конференции "СВЧ-техника и телекоммуникационные технологии". - Севастополь: СевГТУ. - 1999. - С. 229-230.
9. Онуфрієнко В.М., Самолчев П.О., Слюсарова Т.І. Деформація електричного поля циліндричним тілом з фрактальними властивостями поверхні// "Радіоелектроніка, інформатика, управління". - Запоріжжя: ЗДТУ. - 1999. - № 2. - С.33-37.

Надійшла 11.02.2000

Після доробки 13.03.2000

УДК 621.396.6.001.2:681.3

АВТОМАТИЗАЦІЯ ПРОЦЕСУ ТЕПЛОГО ПРОЕКТУВАННЯ МІКРОЕЛЕКТРОННИХ ПРИСТРОЇВ ЗА ДОПОМОГОЮ ПРОГРАМИ CIRCUIT MAKER 6 PRO

А. В. Пархоменко, Н. А. Коноваленко, Л. В. Поставець

В данной работе представлены результаты теплового моделирования интегральной микросхемы с применением программы Circuit Maker 6 PRO. Предложенный подход обеспечивает необходимую точность моделирования.

В цій роботі представлені результати теплового моделювання інтегральної мікросхеми з використанням програми Circuit Maker 6 PRO. Запропонований підхід забезпечує необхідну точність моделювання.

In this paper were introduced the results of thermal simulation of integrated chip using program Circuit Maker 6 PRO. Proposed approach provide the necessary accuracy of modelling.

Головною причиною розробки та використання систем автоматизованого проектування (САПР) є зростаюча невідповідність між вимогами скорочення термінів, підвищення якості, зниження вартості проектних робіт та старими методами проектування. Впровадження в інженерну практику методів автоматизації проектування дозволяє перейти від традиційного макетування проектованої апаратури до її моделювання за допомогою ПЕОМ. Крім цього, застосування ПЕОМ дозволяє здійснити весь цикл проектування, що складається з наступних етапів:

1. Синтез структури та принципової схеми пристрою.
2. Аналіз його характеристик в різних режимах з урахуванням розкиду параметрів компонентів та наявності дестабілізуючих факторів, а також параметричну оптимізацію.
3. Синтез топології, включаючи розміщення елементів на платі або кристалі та розведення міжз'єднань.
4. Верифікацію топології.

5. Випуск конструкторської документації.

Завдяки можливості доступу до інформації в мережі Internet стало можливим ширше охопити і проаналізувати існуючі сучасні програмні системи автоматизованого проектування радіоелектронних та мікроелектронних пристроїв.

На цей час існує багато пакетів автоматизованого схемотехнічного проектування: Circuit Maker 6 PRO, P-CAD, OrCAD, PSpice, Micro-CAP, Micro-Logic, ELECTRONICS WORKBENCH, MAEC-II та інші, кожний з яких має свої особливості і відрізняється універсальністю, економічністю і точністю розрахунків.

Сьогодні розробники програмного забезпечення систем автоматизованого схемотехнічного проектування основну увагу приділяють створенню універсальних моделюючих програм для аналізу характеристик аналогових та цифрових пристроїв. Сучасні програми САПР працюють в діалоговому режимі та мають великий набір сервісних модулів, що значно спрощують роботу на ПЕОМ та роблять їх дружніми по відношенню до користувача.

Для більшості електронних пристроїв їх структура та принципова схема в значній мірі залежать від галузі застосування та вхідних даних на проектування, що створює значні труднощі при синтезі принципової схеми за допомогою ПЕОМ. В цьому випадку первинний варіант схеми складається інженером "вручну" з наступним моделюванням та оптимізацією на ПЕОМ. В сучасних програмах схемотехнічного проектування передбачене введення принципової схеми модельованого

пристрою в графічному вигляді, що значно спрощує підготовку вхідних даних та зменшує кількість помилок. Результати моделювання виводяться в зручній формі для сприйняття проектувальником. Недоліком є те, що зазвичай при моделюванні вибираються спрощені доступні схеми, не повністю адекватні умовам роботи реального пристрою, тому результати моделювання цифрових пристроїв є більш вірогідними.

Однією з відомих та доступних користувачам систем є програма Circuit Maker 6 PRO, призначена для схематичного складання та моделювання електричних кіл. Вона дозволяє виконувати наступні проектні операції:

- створення умовних графічних позначень елементів принципової електричної схеми ;
- графічне введення креслення принципової електричної схеми і конструктивів проєктованого пристрою;
- аналогове і цифрове моделювання електронних пристроїв;
- визначення напруг, струмів, потужності елементів принципової схеми;
- визначення залежності напруг, струмів і миттєвих потужностей елементів принципової схеми від часу, частоти і т.і.

Вхідними даними програми Circuit Maker є електричні принципові схеми модельованих пристроїв, представлені у графічному вигляді, з використанням бібліотечних елементів.

Вихідними даними програми Circuit Maker 6 PRO є:

- зображення електричних принципових схем;
- результати аналізу характеристик схем, наведені в сучасному графічному вигляді.

До найбільш відомих серед впроваджених на підприємствах СНД програмних засобів, які здійснюють теплове проектування, належать АСОНИКА-Т, ПРАМ-9, ТРАП-1, ТЕПЛОТА, ТЕРМ-4. Серед зарубіжних систем теплового моделювання радіоелектронних та мікроелектронних пристроїв можна відзначити: mS-THERMANAL, THERMODEL, TULSOFT, SSSSI, SUNRED, TULTHERM, THERM. Система теплового проектування електронного пристрою повинна дозволити в інтерактивному режимі аналізувати, синтезувати та оптимізувати відповідні теплові характеристики.

Деякі програмні системи теплового проектування дозволяють крім моделювання температурних полів аналізувати і теплоелектричні характеристики, наприклад APLAC, SMODTHERM, ATLAS, DF/Thermax, FIDAP, FloTHERM, Flow-3D, Flunt, Flux-chip, PAFEC, PC-3D, PHOENICS, Sauna, SINDA, TAK III, TULSoft 2, Visula [1].

Однією з відомих вітчизняних систем є програма моделювання теплових процесів методами теплоелектричної аналогії STEA [1]. Система призначена для теплового моделювання конструкцій мікроелектронних пристроїв, забезпечує формування теплоелектричної схеми (ТЕС) прямим та редуційним методами. При

реалізації прямого методу можливі такі види аналізу лінійних теплоелектричних схем: аналіз стаціонарного режиму, аналіз нестаціонарного режиму.

Редуційний метод формування дозволяє виконання більшої кількості видів моделювання як для лінійних, так і для нелінійних схем:

- аналіз стаціонарного режиму;
- аналіз нестаціонарного режиму;
- аналіз параметричної чутливості;
- однопараметричний синтез.

Система містить підсистему формування теплоелектричної схеми, яка виконує формування ТЕС прямим методом на основі заданого опису конструкції пристрою, та підсистему моделювання, яка відповідає за моделювання ТЕС, створених у першій підсистемі, а також забезпечує реалізацію редуційного підходу до побудови ТЕС.

Аналіз особливостей функціонування системи STEA, порівняльний аналіз функціональних можливостей систем Circuit Maker та STEA, а також відомий принцип електротеплової аналогії (табл. 1) дозволив зробити припущення про можливість використання системи схематичного проектування Circuit Maker для моделювання теплових процесів в електронних пристроях.

Для порівняння результатів роботи цих програм використовувались результати теплового моделювання типової конструкції мікроелектронного пристрою - інтегральної схеми (ІС) за допомогою програми STEA, наведені в [1], а також виконувалося її моделювання за допомогою програми Circuit Maker.

Таблиця 1 - Відповідність між характеристиками електричного й теплового полів

Параметр електричного поля	Параметр теплового поля
Потенціал Φ , [В]	Температура T , [К]
Сила струму I , [А]	Тепловий потік P , [Вт]
Напруга $U = \Phi_1 - \Phi_2$, [В]	Різниця температур $T = T_1 - T_2$, [К]

Тестова теплоелектрична схема, що побудована на основі редуційного підходу, наведена на рисунку 1. Основною частиною схеми є кристал (3-й вузол), який закріплений на вивідній рамці, що містить 8 виводів (вузли - 7, 8, 9, 10, 11, 12, 13, 14), які з'єднуються з напівпровідниковим кристалом. Схема є лінійною, містить 18 лінійних теплових опорів, 9 лінійних теплових ємностей та одне динамічне джерело теплового потоку.

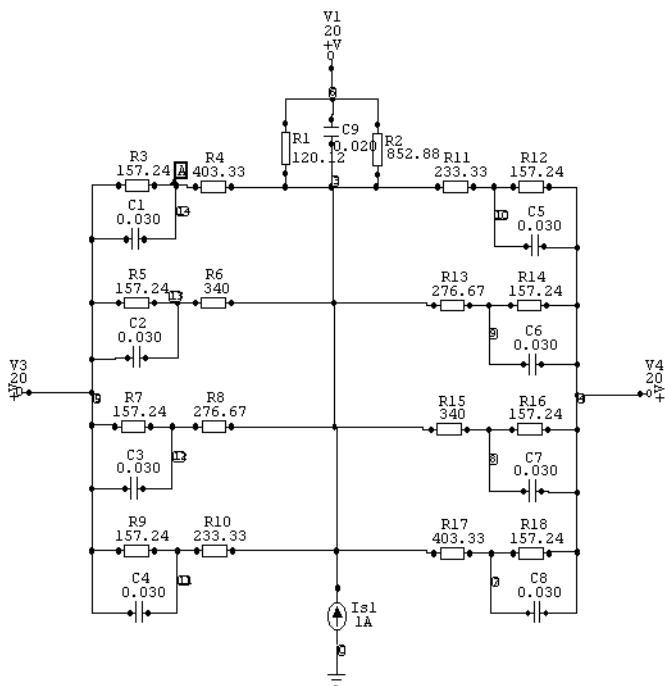


Рисунок 1 - Тестова теплоелектрична схема конструкції ІС

Джерело теплового потоку I_{s1} відповідає тепло-виділенню в кристалі, його величина залежить від розсіюваної кристалом електричної потужності. Для коректного проведення динамічного аналізу це джерело повинне мати нульове значення в першій часовій точці і встановлене стаціонарне значення у всіх наступних точках. Тепловий опір R_1 моделює стік теплоти в оточуюче середовище від кристалу через поверхню мідного тепловідводу. Тепловий опір R_2 відображає відвід тепла від кристалу в оточуюче середовище через передню та бокові поверхні пластмасової оболонки. Теплова ємність C_9 відповідає теплоємності закріпленого на вивідній рамці кристалу. Теплові опори $R_4, R_6, R_8, R_{10}, R_{11}, R_{13}, R_{15}, R_{17}$ відображають стік теплоти в друковану плату від кристалу через всі виводи. Теплові опори $R_3, R_5, R_7, R_9, R_{12}, R_{14}, R_{16}, R_{18}$ враховують відвід теплоти, перенесеної i -им виводом з плати в оточуюче середовище, їх величина залежить від розмірів плати, від товщини й конфігурації провідного шару, від умов охолодження. Теплові ємності $C_1 - C_8$ відповідають теплоємності місця контакту i -го виводу з друкованою платою. Елементи $C_1 - C_9$ введені в схему для якісної ілюстрації теплових перехідних процесів в ІС. Наведена теплоелектрична схема має достатньо

високу адекватність реальній конструкції ІС і може бути використана для теплового моделювання даної ІС.

Для визначення температурного стаціонарного режиму ІС був виконаний статичний аналіз теплоелектричної схеми. За допомогою функції **Multimeter** було виконано розрахунок потенціалів в вузлах схеми та порівняння з результатами статичного аналізу (стаціонарний режим) у STEA. У таблиці 2 наведені результати статичного аналізу - температури в центрі кристалу та в місцях з'єднання виводів з платою. Температура оточуючого середовища була прийнята $20\text{ }^\circ\text{C}$.

Таблиця 2 - Порівняння результатів

STEA		Circuit Maker	
Частина схеми	Температура, $^\circ\text{C}$	Вузол	Напруга, В
Кристал	57.289	3	57.289
Вивід 1	30.460	14	30.460
Вивід 2	31.792	13	31.792
Вивід 3	33.513	12	33.513
Вивід 4	35.012	11	35.012
Вивід 5	35.012	10	35.012
Вивід 6	33.513	9	33.513
Вивід 7	31.792	8	31.792
Вивід 8	30.460	7	30.460

Для дослідження нестационарного теплового режиму ІС був проведений динамічний аналіз за допомогою функції **Transient Analysis** та порівняння з результатами динамічного аналізу (нестационарний тепловий режим) в STEA. Задані наступні параметри аналізу: початковий момент часу - 0.0 с, кінцевий момент часу - 20.0 с, крок - 0.5 с. Результати аналізу перехідного теплового режиму кристалу в графічній формі наведені на рисунку 2. Аналогічні залежності були побудовані для всіх виводів. З аналізу одержаних результатів випливає, що приблизно через 5с після включення кристал переходить в усталений температурний режим. Виконані дослідження свідчать, що відповідність результатів моделювання в Circuit Maker та в STEA досить висока.

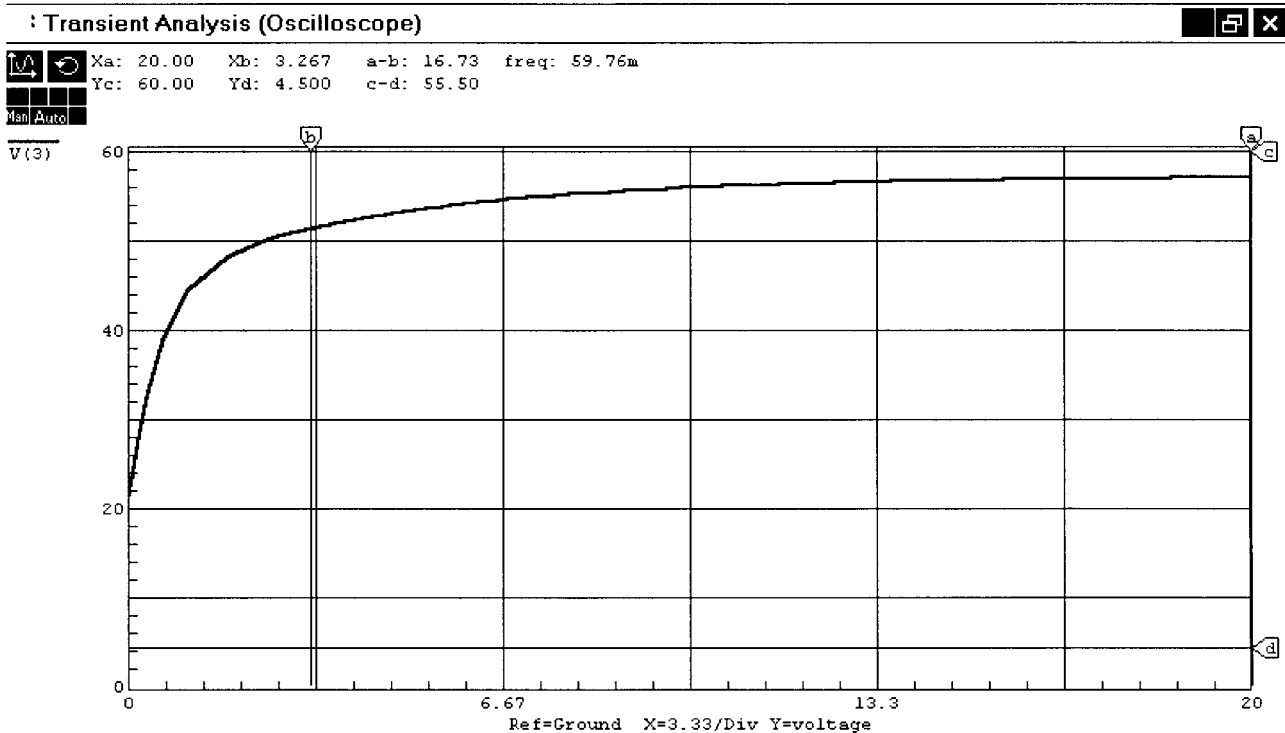


Рисунок 2 - Результати моделювання перехідного теплового режиму кристалу

Таким чином, за допомогою пакета Circuit Maker є можливість моделювати як електричні, так і теплові процеси в схемах та конструкціях мікроелектронних приладів. Значною перевагою пакета є його доступність, зручний інтерфейс, широкі функціональні можливості, в тому числі можливість представлення вхідних та вихідних даних у наочному графічному вигляді.

ПЕРЕЛІК ПОСИЛАНЬ

1. Федасюк Д.В. Методи та засоби теплового проектування мікроелектронних пристроїв. - Львів: Видавництво Державного університету "Львівська політехніка", 1999. - 228с.

Надійшла 10.01.2000
Після доробки 07.02.2000

УДК 621.314.6

ВРАХУВАННЯ ПАРАМЕТРІВ ЕЛЕМЕНТІВ СХЕМИ В РЕЖИМІ ПІДВИЩЕННЯ НАПРУГИ ІМПУЛЬСНОГО РЕГУЛЯТОРА

В. Д. Флора

Предложено учитывать параметры источника и элементов схемы импульсного регулятора в режиме повышения напряжения методом энергетического баланса.

Запропоновано врахування параметрів джерела та елементів схеми імпульсного регулятора в режимі підвищення напруги методом енергетичного балансу.

It is offered to take into account parameters of a source and circuit elements of an impulse regulator in the mode of rise of voltage by a method of energy balance.

Перетворювачі постійної напруги в постійну ефективно застосовуються з такими джерелами первинного електроживлення, як сонячні батареї, паливні елементи, термоелектричні генератори і т. і. Вони є основною складовою частиною більшості систем безперервного електропостачання для живлення виконавчих пристроїв з електродвигунами, для узгодження постійних напруг джерела та навантаження [1,2,3].

Одним з найбільш вживаних у радіоелектронній апаратурі є імпульсний регулятор для підвищення напруги

нерегульованого джерела, який розглянуто без врахування активних опорів та індуктивностей [4].

Більш точні співвідношення для усталеного режиму визначаються з урахуванням активних опорів джерела $R_{дж}$, накопичувального дроселя $R_{нк}$ та навантаження $R_{н}$, а також їхніх індуктивностей $L_{дж}$, $L_{нк}$, $L_{н}$.

Розрахунковою є схема рис. 1, а часові діаграми змінення основних величин наведені на рис. 2.

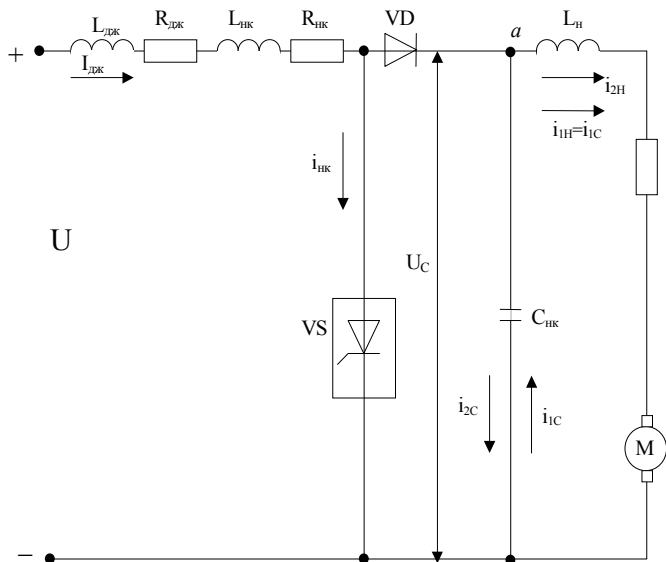


Рисунок 1 - Підвищувач напруги регулятор з паралельним накопичувальним конденсатором

Розмах пульсацій струму $\Delta I_{нк}$ накопичувального дроселя $L_{нк}$ можливо визначити методом енергетичного балансу [5]. Для проміжку часу $0 \leq t \leq T_p$, де T_p - час розряду накопиченої енергії:

$$\Delta W_{L_{нк}} = \frac{L_{дж} + L_{нк}}{2} \cdot (I_{нк \max}^2 - I_{нк \min}^2) = [I_{нк \text{ ср}}(U_{c \text{ ср}} - U) + I_{нк \text{ ср}}^2(R_{дж} + R_{нк}) + I_{нк \text{ ср}}^2 R_{н}] T_p. \quad (1)$$

Лінеаризуємо залежність $i_{нк}(t)$:

$$I_{нк \max} = I_{нк \text{ ср}} + \frac{\Delta I_{нк}}{2}, \quad (2)$$

$$I_{нк \min} = I_{нк \text{ ср}} - \frac{\Delta I_{нк}}{2}, \quad (3)$$

де $I_{нк \max}$, $I_{нк \min}$, $I_{нк \text{ ср}}$ - відповідно максимальний, мінімальний та середній струми накопичувального дроселя;

$U_{c \text{ ср}}$ - середня напруга накопичувального конденсатора.

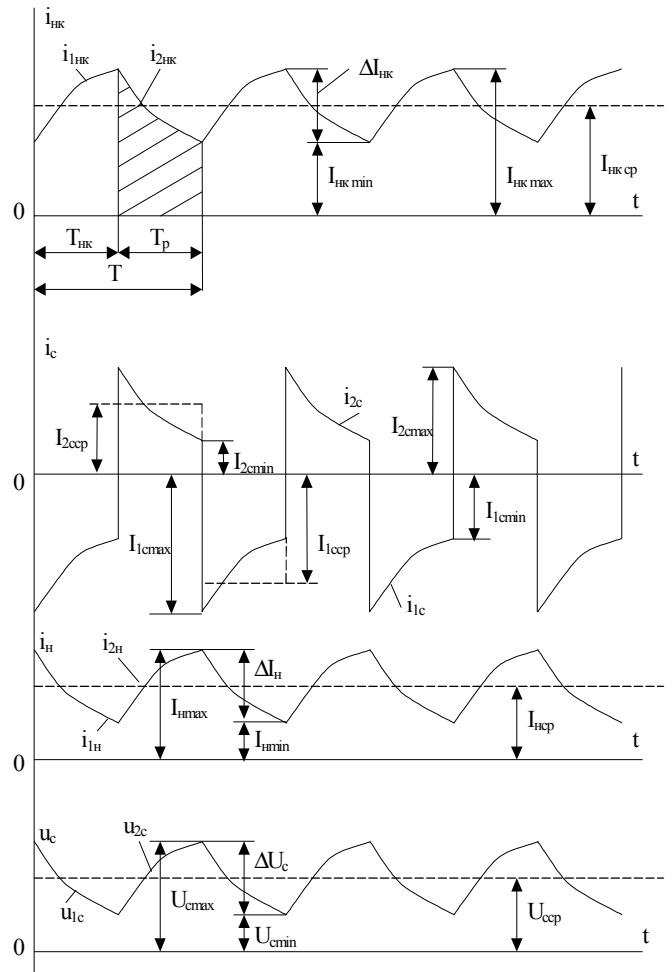


Рисунок 2 - Часові діаграми змінення основних величин

Враховуючи в формулі (1) залежності (2) та (3), після перетворень:

$$\Delta I_{нк} = \frac{\left[U_{c \text{ ср}} - U + I_{нк \text{ ср}}(R_{дж} + R_{нк}) + \frac{I_{нк \text{ ср}}^2 R_{н}}{I_{нк \text{ ср}}} \right] \cdot (1 - \alpha)}{(L_{дж} + L_{нк}) \cdot f}, \quad (4)$$

де

$\alpha = T_{нк}/T$ - відносна тривалість накопичення енергії.

При цьому

$$I_{н2 \text{ ср}} = I_{н \text{ ср}} \cdot (1 - \alpha), \quad (5)$$

а середній струм навантаження (якоря двигуна постійного струму незалежного збудження) $I_{н \text{ ср}}$ визначається зі співвідношень

$$M_c = M_{\text{ср}} = C_M \cdot \Phi \cdot I_{н \text{ ср}};$$

$$I_{\text{н ср}} = \frac{M_c}{C_m \cdot \Phi}, \quad (6)$$

де M_c - статичний момент на валу двигуна.

Для проміжку часу $0 \leq t \leq T_{\text{нк}}$

$$\begin{aligned} \Delta W_{L_{\text{нк}}} &= \frac{L_{\text{дж}} + L_{\text{нк}}}{2} \cdot (I_{\text{нк max}}^2 - I_{\text{нк min}}^2) = \\ &= [I_{\text{нк ср}} U + I_{\text{нк ср}}^2 (R_{\text{дж}} + R_{\text{нк}})] T_{\text{нк}}. \end{aligned}$$

Враховуючи формули (2) та (3), після перетворень:

$$\Delta I_{\text{нк}} = \frac{[U + I_{\text{нк ср}} (R_{\text{дж}} + R_{\text{нк}})] \cdot \alpha}{(L_{\text{дж}} + L_{\text{нк}}) \cdot f}. \quad (7)$$

Середній струм накопичувального дроселя визначається зі співвідношення:

$$\frac{I_{\text{нк ср}}}{I_{2\text{нк ср}}} = \frac{T}{T_p} = \frac{1}{1-\alpha}.$$

Тому

$$I_{\text{нк ср}} = \frac{I_{2\text{нк ср}}}{1-\alpha}. \quad (8)$$

Виходячи з першого закону Кірхгофа, для середніх струмів у точці "а" рис. 1.:

$$I_{2\text{нк ср}} = I_{2\text{н ср}} + I_{2c\text{ ср}}. \quad (9)$$

В усталеному режимі приріст кількості електрики також усталений, тобто

$$\Delta Q_{2c} = \Delta Q_{1c},$$

$$I_{2c\text{ ср}} \cdot T_p = I_{1c\text{ ср}} \cdot T_{\text{нк}},$$

$$I_{2c\text{ ср}} \cdot (1-\alpha) = I_{1c\text{ ср}} \cdot \alpha. \quad (10)$$

Оскільки

$$I_{1c\text{ ср}} = I_{1\text{н ср}} = I_{\text{н ср}} \cdot \alpha, \quad (11)$$

з формул (10) та (11):

$$I_{2\text{н ср}} = I_{\text{н ср}} \frac{\alpha^2}{1-\alpha}. \quad (12)$$

Тоді зі співвідношення (9), враховуючи (5) та (12), одержимо

$$I_{2\text{нк ср}} = I_{\text{н ср}} \frac{(1-\alpha) + \alpha^2}{1-\alpha}, \quad (13)$$

а з формули (8):

$$I_{\text{нк ср}} = I_{\text{н ср}} \left[1 + \frac{\alpha^2}{(1-\alpha)^2} \right]. \quad (14)$$

Враховуючи в (4), (7) формулу (14), після перетворень

$$\Delta I_{\text{нк}} = \frac{b \cdot (1-\alpha)}{(L_{\text{дж}} + L_{\text{нк}}) \cdot f}, \quad (15)$$

де

$$b = U_{c\text{ ср}} - U +$$

$$+ I_{\text{н ср}} \left\{ \left[1 + \frac{\alpha^2}{(1-\alpha)^2} \right] (R_{\text{дж}} + R_{\text{нк}}) + \left[\frac{(1-\alpha)^2}{(1-\alpha)^2 + \alpha^2} \right] R_{\text{н}} \right\},$$

$$\Delta I_{\text{нк}} = \frac{\left\{ U + I_{\text{н ср}} \left[1 + \frac{\alpha^2}{(1-\alpha)^2} \right] (R_{\text{дж}} + R_{\text{нк}}) \right\} \cdot \alpha}{(L_{\text{дж}} + L_{\text{нк}}) \cdot f}. \quad (16)$$

Порівнюючи формули (15) та (16), які визначають одну й ту ж величину, одержимо:

$$U_{c\text{ ср}} = \frac{U}{1-\alpha} +$$

$$+ I_{\text{н ср}} \left\{ \frac{1-2\alpha}{1-\alpha} \left[1 + \frac{\alpha^2}{(1-\alpha)^2} \right] (R_{\text{дж}} + R_{\text{нк}}) - R_{\text{н}} \frac{(1-\alpha)^2}{(1-\alpha)^2 + \alpha^2} \right\}. \quad (17)$$

Із загального співвідношення (17) можливо одержати більш прості формули для різних окремих випадків. Зокрема, якщо $R_{\text{н}} \approx 0$,

$$R_{\text{дж}} + R_{\text{нк}} \approx 0,$$

то

$$U_{c\text{ ср}} = \frac{U}{1-\alpha}, \quad (18)$$

тобто - така ж формула, як і наведена [1, 2, 3, 4].

Розмах пульсацій напруги на конденсаторі $C_{\text{нк}}$ зручніше підраховувати для проміжку часу $0 \leq t \leq T_{\text{нк}}$, оскільки для нього $i_{2\text{нк}} = 0$, $i_{1c} = i_{1\text{н}}$.

В такому разі

$$\Delta W_c = \frac{C}{2} (U_{c\text{ max}}^2 - U_{c\text{ min}}^2) = (I_{\text{н ср}} \cdot U_{c\text{ ср}} + I_{\text{н ср}}^2 \cdot R_{\text{н}}) T_{\text{нк}}. \quad (19)$$

Враховуючи, що

$$U_{c\text{ max}} = U_{c\text{ ср}} + \frac{\Delta U_c}{2}; \quad (20)$$

$$U_{c\text{ min}} = U_{c\text{ ср}} - \frac{\Delta U_c}{2}, \quad (21)$$

після перетворень

$$\Delta U_c = \frac{I_{н\text{ ср}} \left(1 + \frac{I_{н\text{ ср}} R_n}{U_{c\text{ ср}}} \right) \alpha}{C \cdot f}. \quad (22)$$

Для відповідних випадків підставляється відповідне значення $U_{c\text{ ср}}$, підраховане за формулою (17).

Для того ж проміжку часу $0 \leq t \leq T_{нк}$, враховуючи що L_n - індуктивність навантаження, можливо записати:

$$\Delta W_{L_{нк}} = \frac{L_n}{2} (I_{н\text{ max}}^2 - I_{н\text{ min}}^2) = (I_{н\text{ ср}} \cdot U_{c\text{ ср}} + I_{н\text{ ср}}^2 \cdot R_n) T_{нк} \quad (23)$$

Після перетворень з рівняння (23):

$$\Delta I_n = \frac{(U_{c\text{ ср}} + I_{н\text{ ср}} R_n) \alpha}{L_n \cdot f}. \quad (24)$$

Для різних окремих випадків у формулі (24) враховується $U_{c\text{ ср}}$ за формулою, одержаною зі співвідношення (17). Зокрема, якщо активні опори $R_i \approx 0$, то

$$\Delta I_n = \frac{U \cdot \alpha}{L_n \cdot f \cdot (1 - \alpha)}, \quad (25)$$

тобто таке ж співвідношення, як і наведене [4].

У співвідношеннях (23)-(25) індуктивність ланцюга навантаження

$$L_n = L_p + L_{ла}, \quad (26)$$

де L_p - індуктивність згладжуючого реактора,

$$L_{ла} = L_a + L_{псз} + L_k + L_{дп}; \quad (27)$$

індуктивності обмоток:

L_a - якоря;

$L_{псз}$ - послідовного збудження;

L_k - компенсуючої;

$L_{дп}$ - додаткових полюсів.

Приблизно визначити $L_{ла}$ можливо за формулою [6]:

$$L_{ла} = \beta \cdot \frac{U_N}{p \cdot \omega_N \cdot I_N}, \quad (28)$$

де коефіцієнт $\beta = 0,25$ - для двигунів з компенсаційною обмоткою, $\beta = 0,6$ - без компенсаційної обмотки;

U_N, I_N, ω_N - відповідно номінальні напруга, струм якоря та кутова швидкість обертання;

p - число пар полюсів машини.

Таким чином, одержані уточнені формули для основних величин в усталеному режимі імпульсного регулятора з урахуванням параметрів схеми.

ПЕРЕЛІК ПОСИЛАНЬ

1. Севернс Р., Блум Г. Импульсные преобразователи постоянного напряжения для системы вторичного электропитания: Пер. с англ. под ред. Л. Е. Смольникова.-М.: Энергоатомиздат, 1988.-294 с.
2. Четти П. Проектирование ключевых источников электропитания: Пер. с англ.-М.: Энергоатомиздат, 1990.-240 с.
3. Поликарпов А. Г., Сергиенко Е. Ф. Однотактные преобразователи напряжения в устройствах электропитания РЭА.-М.: Радио и связь, 1989.-160 с.
4. Флора В. Д. Дослідження режиму підвищення напруги імпульсних регуляторів з паралельним накопичувальним конденсатором. Електричний журнал, 1997, №2, с. 28-32.
5. Флора В. Д. Використання рівнянь енергетичного балансу при розрахунках усталених режимів імпульсних регуляторів. Електричний журнал, 1998, № 1, с. 28-32.
6. Андреев В. П., Сабинин Ю. А. Основы электропривода.-М.-Л.: ГЭИ, 1963.-772 с.

Надійшла 23.02.99

ІНФОРМАТИКА

УДК 681.32:007

NEUROINFORMATICS: WHAT ARE US, WHERE ARE WE GOING, HOW TO MEASURE OUR WAY?

A. N. Gorban

Дан обзор основных проблем и современных тенденций развития нейроинформатики.

Дан огляд основных проблем та сучасних тенденцій розвитку нейроінформатики.

The review of main problems and modern lines of development of neuroinformatics is given.

What is neuroinformatics? For me here and now neuroinformatics is a direction of science and information technology, dealing with development and study of the methods for solution of problems by means of neural networks. A base example of artificial neural network, which will be referred to below, is a feed-forward network from standard neurons (fig. 1)

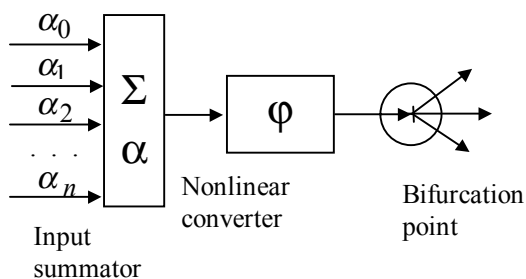


Figure 1 - Standard neuron

A field of science cannot be determined only by fixing what it is "dealing with". The main component, actually constituting a scientific direction, is "THE GREAT PROBLEM", around which the efforts are concentrated. One may state even categorically: if there is no a great problem, there is no a field of science, but only more or less skilful imitation (the question, who and what for needs imitations, is an amusing one, but it can lead us far away). Just main problems impart sense to each particular study.

What is "THE GREAT PROBLEM" for neuroinformat-

ics? What imparts sense to this vast flow of studies? The answer will be as we state it - it is impossible to separate a reflexive exploration of sense of an activity from generation and ascription of the sense "post factum".

There are two directions of search for the sense - first, study of brain (solution of mysteries of thinking), and, second, problems of efficiency of computations. Many people are inspired by the banner of "brain-like" computers, but for me it is alien banner. I lack the determinacy of problem here - the determinacy of gap in activity: what we cannot do now and what we will be able to do after obtaining the solution? Brain-like computer is rather a metaphor than designation of the problem. However, each has its own symbol of belief, though those are not numerous.

The problem of effective parallelism pretends to be the central problem which is being solved by the whole neuroinformatics. Long ago sank into oblivion the naive idea: let take more processors, and the efficiency of computer will increase proportionally. There is a well known "Minsky hypothesis" (fig. 2): efficiency of a parallel system increases (approximately) proportionally to logarithm of the number of processors; at least, it is a convex upwards (i.e. concave) function (discussion of these problems can be found in [2]).

Nowadays more and more often as a way to overcome this limitation the following approach is used: for different classes of problems maximum parallel algorithms of solution are constructed which use some abstract architecture (paradigm) of fine grain parallelism, and for concrete parallel computers means are developed for realization of parallel processes of a given abstract architecture. As a result, an efficient technique of production of parallel programs is created. Neuroinformatics supplies universal fine grain parallel architectures for solution of different classes of problems. Thus, it is used for solving the problem of efficient parallelism according to the following scheme (fig. 3).

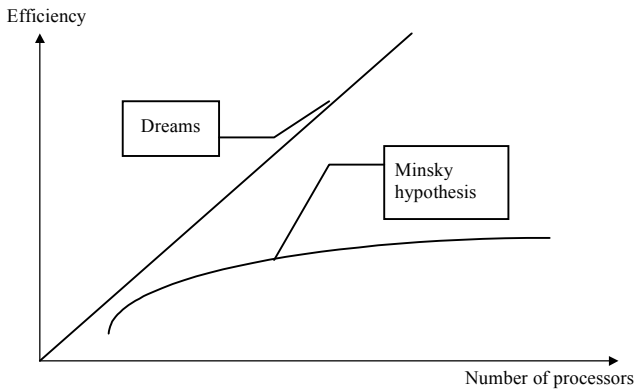


Figure 2 - Minsky hypothesis

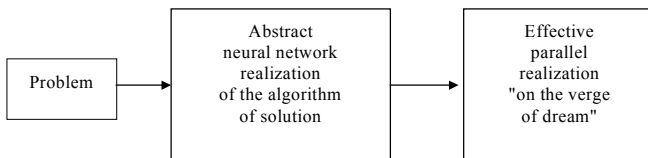


Figure 3 - Way to effective parallelism

One of main problems of brain studies and creation of "brain-like" devices is the same problem of effective parallelism: how the brain uses its billion of neurons for effective parallel operation? So, contradictions between the different ways of search for sense are not so large.

In addition to "the great problem" of effective parallelism, let suggest a great applied problem, which can be solved on this way. **Controlled thermonuclear synthesis** demands for new tools for retaining hot plasma, and, in general, for controlling its state. It seems quite probable that the possibilities of physical traps are exhausted, and any essential increase of stability is unreachable on this way. May be, time has come for creation of a cybernetic system which recognizes instabilities and effectively suppresses them. There exist neuro-equilibrist - a cart driven by self-training neural network, which supports a thin rod in unstable vertical position. Can neural network recognize changes in the state of plasma and rise of instabilities, elaborate and realize the control action over time of the order of 10^{-8} s? It is very likely that this is possible. The time of operation of one layer of network can be of the order of $0,3 \times 10^{-8}$ s, and three - five layers can turn out to be quite sufficient. Then such quick-action is attainable, if there are no extra devices between the network and plasma which cause considerable delay, the information is perceived by the network directly through a system of sensors connected to it, and the signals of output neurons act upon the plasma directly (most probably,

through electromagnetic field) - as in the magic "thinking is action". The number of neurons can be estimated as the number of unstable degrees of freedom which must be suppressed. On the whole, such a project seems to be quite realizable by contemporary tools, and not too expensive in comparison with other ways of retaining plasma.

Yet, why we speak only about great problems? Generally, not everything useful and intellectual is a science. For instance, most important in personal computers is neither scientific character nor progress in solution of great problems, but a system of services which they propose to the user. What if to consider neuroinformatics as **a services sector**, what that will result in?

In the system of services rendered by neuroinformatics to actual customer the main place now belongs to data processing, including adaptive signal processing [3]. The central place at the services market is occupied by financial applications [4]. Further are situated military-industrial (and here one cannot find exhaustive review of literature because of understandable reasons), and then medical applications (some examples are given in [5]). In my opinion, the most important achievement of neuroinformatics over the last decade in the sector of intellectual services is systematic use of neural networks **for extraction of knowledge from data**. In the second place I would put the large-scale circulation of *neural net methods of data visualization*, based on self-organizing Kohonen maps [6]. As to **pattern recognition and non-linear regression problems** of various kinds, nowadays exist tens (or even hundreds) neural network program products which give such opportunity to user, and there are thousands of works in which this opportunity was used.

Neural network extraction of knowledge from data is based on two circumstances: firstly, neural network usually is very redundant - there are too many connections for solving the given concrete problem, and it is possible rarefy the network, constantly training it, so that only the necessary connections would remain. Secondly, rarefied network can be read as explicit algorithm of solution of the problem (for instance, in the form of tree of solutions), - and this is, in essence, explicit knowledge. That is especially simple if one follows certain rules of rarefaction ensuring easiness of reading. For instance, at first one can strive for minimization of the number of input connections of all neurons, then decrease the number of neurons and the number of entries of network, and finally, do away with as many connections as possible. Such simplification is done either on the basis of analysis of indices of sensitivity or due to introduction into the evaluation of the quality of training a penalty for complexity of structure. Both these methods were introduced and studied in monograph [7], but other variants of algorithms [8,9]¹ are widely spread now.

During training artificial neural network creates some internal rules, but these rules are hidden in the structure

of a network and not clear to the user. Moreover, these rules are so difficult for interpretation and understanding. Also discovery of explicit rules for solving of a problem is often more significant than the solution itself. Thus, there is a problem of refining of the hidden knowledge and translating them to natural language.

The main ideas of generalized technology of extraction of explicit knowledge from data are:

1) maximal reduction of network complexity (not only removal of neurons or synapses, but removal all the unnecessary elements and signals and reduction of the complexity of elements),

2) using of adjustable and flexible pruning process (the pruning sequence shouldn't be predetermined - the user should have a possibility to prune network on his own way in order to achieve a desired network structure for the purpose of extraction of rules of desired type and form),

3) extraction of rules not in predetermined but any desired form. Some considerations and notes about network architecture and training process and applicability of currently developed pruning techniques and rule extraction algorithms are discussed. This technology, being developed by us for more than 10 years, allowed us to create dozens of knowledge-based expert systems.

The eight-year experience of application of this technology to problems in various fields allows us to estimate some propositions and draw some conclusions. The brief description is following.

First, neural network that can solve the problem with desired degree of accuracy has to be trained. Instead of use of mean square error as a estimation function it's better to use modified estimation functions which allow to control accuracy of decision in order to reach more simple resulting structure of the network at the second step of proposed technology. Also, in some cases it would be better to use multilayer neural networks, not only three-layered, in order to achieve more hierarchical set of rules with more simple rules at each level of hierarchy. The successful training of the network creates a hidden complex set of decision rules - the implicit knowledge.

Second, it's necessary to remove superfluous elements and inputs from the network. The pruning should lead to more simple interpretation of hidden rules, therefore it's not enough to remove only neurons or connections and it's necessary to introduce a methodology of complex pruning. So, we present the following set of available pruning operations: removing of inputs, neurons, synapses, biases, uniform simplification of a network (when the maximum number of synapses connected to neuron is decreased over the network). The user can establish the execution order of these operations itself. Then reduction of synaptic weights to values from a finite set of fixed values should

be done. All these pruning operations are based on sensitivity analysis, only first-order derivatives are used [1]. Pruning is carried out by consecutive removal of inputs or elements and fails when it's impossible to reach zero of estimation function by retraining. The last step is replacement of sigmoid nonlinear transfer functions by threshold or piecewise-linear functions. Now the hidden implicit knowledge is refined and simplified, so it's possible to understand the meaning and generate explicit knowledge.

The third step is to write down the explicit knowledge in natural language. We propose to carry out such process by consecutive analysis of network structure manually. Because the meaning of input signals is already known, it's possible to substantially name the output signals of the first layer neurons, then second layer neurons and so on. The introduced requirement of network's uniform simplicity is proposed only for simplification of analysis phase, because the less number of input synapses of the neuron the easier to interpret and name its output signal. Also it's possible to extract rules from the network not manually, but automatically in fuzzy form of If-Then form, or in another form. Here we present a complete set of available types of rules that can be extracted from the network. The rule of a certain type may be interpreted in a different ways as a fuzzy, probabilistic or logical statement, but such interpretation involves expert knowledge about problem area and should be done manually by user.

It should be noted that several networks of the same initial architecture and for the same data often lead to different decision rules (different explicit knowledge). It is not a lack.

A particular case of rarefication of neural networks is selection of minimal sets of input signals sufficient for solving a problem. Such sets are not unique, as any knowledge based on data is not, though - other variants always are possible (except for really trivial and uninteresting cases). **Minimization of description** is one more sphere of services rendered by neuroinformatics.

One most important feature makes similar the expansion of neuroinformatics and the information revolution performed by personal computers - transition **from elite marks to profane techniques** (to the techniques which can be mastered by every commonly educated person, i.e. profane but not professional). Profanation not always is bad. Sacral oriental martial arts are being transformed into a quite profane occidental sport - and that is all right. Machinery evolves in the direction of increasing profanity, and that is normal.

Numerous population of diverse neural net algorithms inhabits the natural habitat of data processing, substituting the statistical provability [10] with neural net verisimilitude. That is sufficient in many fields, especially for

1. Now as before I wonder why in the algorithms of rarefication widely spread in the West the indices of sensitivity are used which are based on analysis of the second derivatives, whereas it would be much faster and cheaper to use average values of modules of the first derivatives in the course of training [7]. The other used methods of "zero order" which consist in pruning of connections with small weights are less efficient.

preliminary conclusions and advices. Alchemy of financial decisions is the best example. But also inverse process is necessary - from verisimilitude to provability. As one of urgent problem which is vital for further advance of neuroinformatics I would indicated **the problem of provability of neurocomputations**. Regular tools of verification of trained neural networks with exact evaluation of reliability are needed. A new field of science, "theory of reliability of artificial neural networks" should be created, otherwise many grave applications will be inaccessible for us.

An important argument for substantiation of expansion of artificial neural networks is the result that "neural networks can do anything", that by means of neural networks it is possible to compute any function of many variables up to any accuracy [11]. Yet there is such a widespread confusion in this issue that I should dwell upon it. As a symbol of belief many authors use the famous Kolmogorov-Arnold theorem on representation of continuous functions of several variables as superposition of continuous functions of one variable [13, 14]. However, this theorem states the possibility of *exact* representation of a function of many variables by means of *a very special set of functions* of one variable. These functions are very exotic, in particular, they are nowhere differentiable. On the contrary, by means of neural networks *approximate* representations of function of many variables are constructed. At that, *any limitations on the functions of one variable are absent*, one function is sufficient if only it is nonlinear [12]. Corresponding theorems do not have any concern to the outstanding Kolmogorov-Arnold theorem, but they are a generalization of as much outstanding Stone theorem [15], which, in turn, generalizes the Weierstrass theorem on approximation of functions by polynomials. Let X be a compact set, $C(X)$ be the space of continuous functions on X , $M \subset C(X)$ be linear subspace, elements of M separate points in X (for any x, y from X there exists such $f \in M$ that $f(x) \neq f(y)$) and $1 \in M$.

Stone theorem. If M is subalgebra in $C(X)$ (i.e. for any $f, g \in M$ $fg \in M$), then M is dense in $C(X)$.

Generalized approximation theorem. If there exists continuous nonlinear function of one variable φ , for which under any $f \in M$ also $\varphi(f) \in M$, then M is dense in $C(X)$.

Corollary. For any continuous nonlinear activation function of standard neuron (fig. 1) φ by means of neural networks it is possible to approximate uniformly any continuous function of many variables up to any accuracy on any closed bounded set.

There exist simple regular procedure of obtaining neural net approximation of continuous functions on bounded sets (as well as on finite ones - it is important for solution of problems of regression). That is training of so called $1\frac{1}{2}$ -

layer approximator with addition of neurons one-by-one. In this network there are only one layer of standard neurons and one output summator receiving all the output signals of neurons. Each new $(n+1)$ -th neuron is trained to minimize mean square error of approximation of the function by the network (under fixed parameters of the first n neurons). The process converges: average square of the error tends to zero.

Are there mathematical achievements generated by neuroinformatics? Positive answer to this question would be a proof of maturity of this field of science. I believe that the answer is "yes". A result has matured inside neuroinformatics, which should come into courses of mathematical analysis and, therefore, is more important than most of technical inventions. Its importance for computational mathematics and all that what is called "computer sciences" exceeds the limits of neuroinformatics. I mean the basis of adaptation capabilities of neural networks - duality and fast differentiation of the functions computed by networks. Now and then, the method of back propagation of error is considered as a rule of training. That is not the most productive point of view: of course, motion against gradient decreases the value of function, but nevertheless it is natural to consider computation of derivatives apart from optimization. The method of back propagation of error is a way of computation of the gradient of estimate. Initially it was formulated for mean square deviations from the right answer, but it is easy to regard it as a method of computation of gradient of any function computable by network.

In order to clarify this issue, let discuss one "obvious" dogma, without destroying of which the effective training of networks would be impossible. Let the computational expenses (evaluated by the time spent by an universal computational device) for computation of one value of function of n variables $H(x_1, \dots, x_n)$ approximately equal to T . How much time it will take for the same device to compute $\text{grad}H$ (if the program is rationally composed)? Most of mathematicians with university diploma would answer:

$$T_{\text{grad}H} \sim nT_H.$$

Incorrect! The right answer is:

$$T_{\text{grad}H} \sim CT_H,$$

where C is a constant which does not depend on the dimensionality n (in most cases $C \sim 3$).

For all functions of many variables occurring in practice, including those computable by neural networks, necessary computational expenses on finding gradient are only two-three times larger than the expenses on computation of one value of function. That is astonishing, because the coordinates of the vector of gradient are n partial deriva-

tives, and expenses on computation of one such derivative are in general case approximately the same as on computation of the value of function. Why it is less expensive to compute them together than separately?

The "miracle" is explained simply enough: it is necessary to organize computation of derivatives of function of many variables efficiently, avoiding duplication. For that, it is necessary to represent the computation of function in more details, in order to deal later not with a "black box" transforming the vector of arguments into the value of function, but with the graph of computations described in detail.

It is convenient to represent the search of $\text{grad}H$ as a dual process over the structure of computation of H . Intermediate results coming from computation of gradient are nothing else than Lagrangian multipliers. It appears that if we represent H as composite function being a superposition of functions of small number of variables (and we are unable to compute functions of many variables in other way) and then accurately use the rule of differentiation of composite function, avoiding unnecessary computations and keeping useful intermediate results, then computation of the totality $\partial H/\partial x_i$ ($i = 1, \dots, n$) is a little bit complicated than computation of one of these functions - they all are constructed from identical blocks.

I do not know who was the first to think all this out. In neuroinformatics, the disputes on the priority are lasting so far. An end to re-discoveries was put by two works in 1986: Rumelhart, Hinton and Williams [16] and Bartsev and Okhonin [17]. However, the first publications belong to 70-th and even 60-th of our century. In the opinion of V.A. Okhonin, Lagrange and Legendre also have right to pretend to authorship of the method. Detailed presentation of mathematical content of these methods is given in chapter 3 of the book [18].

In order to represent the field of action in modern neuroinformatics as a whole, it is useful to draw a bird's-eye view of the main directions of investigations. A variant of such a map is given in fig. 4. The largest wave of works is devoted to "**neuronization**" - construction and application of neural net methods for solution of diverse problems amenable to the methods, but the other directions are of importance as well. In **user neural net programs** in the form of complete services neural network methods of solution of problems are realized, **modelling of real neural networks** is an inexhaustible source of ideas and analogies, and successful **new models of neurons and networks** though rarely appear but can have strong influence upon all other directions. As regards **realization of neural networks** either on usual parallel computing structures or in the form of specialized devices and neurocomputers, the future, undoubtedly, belongs to them. It should be noted that when constructing neurocomputers the main complication is not the realization of neural networks - that is simple, but representation of the processes

of preconditioning of data in more or less standard form and their interpretation, formalization of standard formulations of problems and methods of training e.c. [19].

In conclusion, let me try to formulate what gives neural network realization of solution of a problem:

1. Simple parallel realization on any parallel structures.
2. Realization on special, fast and inexpensive neuro-(co)processors.
3. Realization on reliable systems from unreliable (including analogous) elements.
4. Simple completion of adaptation blocks.
5. Possibility of very fast realizations.
6. Possibility of "explicitation" of implicit knowledge by means of rarefiable neural networks.
7. Minimization of description.

.....

What else give neural networks?

.....

N+1. Universality - standard methods for solution of many problems, including non-standard ones. ...

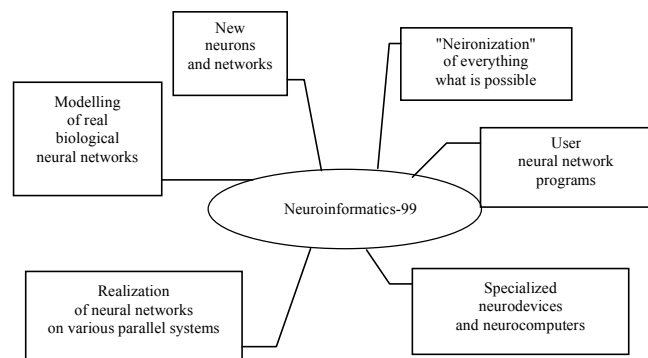


Figure 4 - Neuroinformatics-99. Main directions of works

REFERENCES

1. Rosenblatt F. Principles of neurodynamics. Perceptron and theory of mechanisms of brain. Moscow: Mir, 1965. 480 p. [In Russian]
2. Systems of parallel processing. Ed. D. Ivans. Moscow: Mir, 1985.
3. Widrow B., Stearns S. Adaptive signal processing. Moscow: Mir, 1989.
4. Ezhov A.A., Shumsky S.A. Neurocomputing and its applications in economics and business. Moscow: Izd-vo MIFI, 1998.
5. Gorban A.N., Rossiev D.A. Neural networks on personal computer. Novosibirsk: Nauka (Sib. otделение), 1996. 276 p.
6. Kohonen T. Self-Organizing Maps. Springer, 1997.
7. Gorban A.N. Training neural networks. Moscow: izd. SSSR-USA SP "ParaGraf", 1990. 160 p.
8. Le Cun Y., Denker J.S., Solla S.A. Optimal Brain Damage. Advances in Neural Information Processing Systems II (Denver 1989). San Mateo, Morgan Kaufman, 1990. PP. 598-605.
9. Prechelt L. Comparing Adaptive and Non-Adaptive Connection Pruning With Pure Early Stopping. Progress in Neural Information Processing (Hong Kong, September 24-27, 1996), Springer, Vol. 1. PP. 46-52.
10. Kendall M., Stuart A. Statistical inferences and relations. - Moscow: Nauka, 1973.- 900 p.
11. Cybenko G. Approximation by superposition of a sigmoidal function. Mathematics of Control, Signals, and Systems, 1989. Vol. 2. PP. 303 - 314.
12. Gorban A.N. Approximation of continuous functions of several variables by an arbitrary nonlinear continuous functions of one

- variable, linear functions and their superpositions. Appl. Math. Lett., 1998. V. 11, No. 3, pp. 45-49. Gorban A.N., Wunsch D. The General Approximation Theorem. Proceedings IJCNN'98, IEEE, 1998. PP. 1271-1274.
13. Kolmogorov A.N. On representation of continuous functions of several variables in the form of superposition of continuous functions of one variable. Dokl. AN SSSR, 1957. T. 114, No. 5. P. 953-956.
 14. Arnold V.I. On representation of functions of several variables in the form of superposition of functions of smaller number of variables // Matematicheskoe prosveshchenie, 1958. Vyp. 3. P. 41-61.
 15. Stone M.N. The generalized Weierstrass approximation theorem, Math. Mag., 1948. V.21. PP. 167-183, 237-254.
 16. Rumelhart D.E., Hinton G.E., Williams R.J. Learning internal representations by error propagation, Parallel Distributed Processing: Exploration in the Microstructure of Cognition, D.E.Rumelhart and J.L.McClelland (Eds.), vol. 1, Cambridge, MA: MIT Press, 1986. PP. 318 - 362.
 17. Bartsev S.I., Okhonin V.A. Adaptive networks of information processing. Preprint IF SO AN SSSR, Krasnoyarsk, 1986, ?59B, 20 p.
 18. Neuroinformatics / A.N. Gorban, V.L. Dunin-Barkovsky, E.M. Mirkes i dr. Novosibirsk: Nauka (Sib. otделение), 1998.
 19. Mirkes E.M. Neurocomputer: a project of standard. Novosibirsk: Nauka (Sib. otделение), 1998.

Надійшла 12.02.2000

Після доробки 01.03.2000

УДК 681.32:007

NEURAL NETWORK MODELING OF DATA WITH GAPS

A. N. Gorban, A. A. Rossiev, D. C. Wunsch II

A method of modeling data with gaps by a sequence of curves has been developed. It is a generalization of iterative construction of singular expansion of matrices with gaps. The derived dependencies are extrapolated by Carleman's formulas. The method is interpreted as a construction of neural network conveyor.

Предложен метод моделирования данных с пробелами последовательностью кривых. Он является обобщением сингулярного разложения матриц с пробелами. Полученные зависимости экстраполируются с использованием формул Карлемана. Метод интерпретируется как конвейер нейронов.

Запропоновано метод моделювання даних з пропусками послідовністю кривих. Він є узагальненням сингулярного розкладання матриць з пропусками. Отримані залежності екстраполюються з використанням формул Карлемана. Метод інтерпретується як конвеєр нейронів.

INTRODUCTION

Information about an object under study often carries gaps and erroneous values. Generally, data need preliminary processing, e.g. to fill gaps and correct distorted values. Explicit or implicit such a preprocessing takes place almost every time.

Let there be a table of data (a_{ij}) , its rows correspond to objects, and its columns correspond to features. Then, let a part of information in the table be missing - there are gaps (some $a_{ij} = @$, symbol @ is used to denote gaps in the data). The main problem that arises in this connection is to fill the existing gaps plausibly. There exists an associated problem - to "repair" the table: to distinguish the data with unfeasible values and correct them. In addition, it is useful to construct a calculator associated with the table, to fill the gaps in the incoming data about new objects and repair these new data (under the assumption that the data are interrelated in the same way as in the initial table). Such a calculator implies these data about new

objects to be interrelated as in the initial table.

It should be emphasized that these problems deny discussion of either true data values or statistical provability, they discuss plausibility only.

The described problems are especially difficult (and simultaneously - attractive) in the cases when the density of gaps is high, their location is irregular, and the amount of data is small, for instance, the number of rows is approximately equal to the number of columns.

Ordinary regression algorithms - construction of empirical dependences of some data on other - are inapplicable here. If the gaps are located irregularly, this would require to construct dependencies of unknown data with respect to all their known locations in the table. This would actually mean to construct 2^{n-1} dependences, where n is the number of features. To recover any unknown set of data at least something should be known. In this connection one needs to use the method of modeling data by manifolds of small dimension.

The essence of the method is as follows. The vector of data x with k gaps is represented as the k -dimensional linear manifold L_x , parallel to k coordinate axes corresponding to the missing data. Under a priori restrictions on the missing values instead of L_x we have a rectangular parallelepiped $P_x \subset L_x$. The manifold M of a given small dimension (in most cases - a curve) approximating the data in the best way and satisfying certain regularity conditions) is searched for. For the complete vectors of data the accuracy of approximation is determined as a regular distance from a point to a set (the lower bound of the distances to the points of the set). For the incomplete data in its stead use is made of the lower bound of the distances between the points of M and L_x (or, accordingly, P_x). From the data the closest to them points of M are subtracted - we

obtain a residual, and the process is repeated until the residuals are close enough to zero. Proximity of the linear manifold L_x or parallelepiped P_x to zero means that the distance from zero to the point of L_x (accordingly, P_x) closest to it is small. To further specify the method is to determine how the manifold M is constructed.

The idea of modeling data by manifolds of small dimension has been conceived long ago.

Its most widespread, old and feasible implementation for the data without gaps is the classical method of principal components. The method calls for modeling the data by their orthogonal projections over "principal components" - eigen vectors of the correlation matrix with corresponding largest eigen values. Another algebraic interpretation of the principal component method is singular expansion of the data table. Generally, to present data with sufficient accuracy requires relatively few principal components and dimension is reduced by a factor of tens.

Its generalization for the nonlinear case - the method of principal curves - was proposed in 1988. There are generalizations of the classical method of principal components for data with gaps as well.

The work describes a technique of constructing a system of models for incomplete data. In the simplest case they are a generalization of the classical (linear) method of principal components for data with gaps. The quasilinear method below is superstructured over the linear and employs its results. Finally, the formalism of self-organizing curves is employed to construct an essentially non-linear method.

Each method is accompanied with a physical interpretation illustrating similarity of the methods and their sequential development.

In the general case the method of modeling data with gaps by manifolds (linear and non-linear) of small dimension appears to be more efficient as compared to ordinary regression equations.

The algorithms being developed can be applied when the data matrix cannot be reduced by rearrangement of lines and columns to the following box-diagonal form:

$$A = \begin{bmatrix} A_1 & @ & \dots & @ \\ @ & A_2 & \dots & @ \\ \dots & \dots & \dots & \dots \\ @ & @ & \dots & A_n \end{bmatrix},$$

where @ are the rectangular matrices with unknown elements. To establish connections between different A_i boxes in such tables is impossible.

STATEMENT OF THE PROBLEM

Let there be a rectangular table $A=(a_{ij})$ the cells of which are filled with real numbers or symbol @ denoting

absence of data.

It is required to construct models that make possible to solve the following three problems related to recovery of the missing data:

1. To fill the gaps in data plausibly.
2. To repair the data, i.e. to correct their values in such a way as to make the models constructed work best.
3. To construct by the table available a calculator that would fill the gaps in data and repair them as they arrive (assuming the data in the line arriving at the input to be connected by the same relations as in the initial table).

This raises the question: how (*in which metric*) to evaluate the error of the model? To choose the measure of the error is required both to construct the models and to test them. From the viewpoint of simplicity of calculations the most attractive is the least squares method (LSM). By this method the error is calculated as the sum of squares of deviations over all known data (Mean Square Error - MSE). Yet, arbitrary rule associated with the choice of scale, i.e. normalization of data is present here, too.

1 THE LINEAR MODEL

1.1 Singular Expansion of Tables with Gaps

The material of this section is not immediately employed to process data, yet it provides a simplest example and a prototype for further constructions.

Let there be a rectangular table $A=(a_{ij})$ the cells of which are filled with real numbers or symbol @ denoting absence of data.

The problem is to approximate A best by a matrix of form $x_i y_j$ by the least squares method.

$$\Phi = \sum_{\substack{i,j \\ a_{ij} \neq @}} (a_{ij} - x_i y_j)^2 \rightarrow \min. \tag{1}$$

The problem is solved by successive iterations by explicit formulas.

The halting criterion is the smallness of relative improvement $\Delta\Phi/\Phi$, where $\Delta\Phi$ is the decrease of value Φ obtained in one cycle, and Φ is the current value. The second criterion is the smallness of the value Φ per se. The procedure comes to a halt when $\Delta\Phi/\Phi < \epsilon$ or $\Phi < \delta$ for certain $\epsilon, \delta < 0$. As a result for the given matrix A find the best approximation by the matrix P_1 of the form $x_i y_j$. Further on, look for $A-P_1$ for the best approximation of the same form P_2 and so on, until, e.g. the norm A does not sufficiently approach zero.

Thus, the initial matrix A is presented in the form of a sum of matrices of rank 1, i.e. $A=P_1+P_2+\dots+P_q$.

1.2 Principal Component Method for Tables with Gaps

Following the method described in the previous section derive straight lines crossing the origin of coordinates. Such homogeneous models are not always required. Expand the initial table not over the matrices of form $P=x_i y_j$ but over the matrices of form $P=x_i y_j + b_j$. This brings us to the next problem:

$$\Phi = \sum_{\substack{i,j \\ a_{ij} \neq @}} (a_{ij} - x_i y_j - b_j)^2 \rightarrow \min. \quad (2)$$

Solution of the problem (2) yields models of data by the linear manifolds not necessarily crossing the origin of coordinates.

The basic procedure is to find the best approximation of the table with gaps by matrix of form $x_i y_j + b_j$ [4, 5].

With the vectors y_j and b_j fixed the values x_i , providing minimum to the form (2) unambiguously and simply are defined from equalities $\frac{\partial \Phi}{\partial x_i} = 0$.

In analogy, with the vector x_i fixed the values y_j и b_j , providing minimum to the form (2) are explicitly defined from two equalities $\frac{\partial \Phi}{\partial y_j} = 0$ and $\frac{\partial \Phi}{\partial b_j} = 0$.

Initial values:

y is random, normalized to 1 (i.e. $\sum_j y_j^2 = 1$)

$b_j = \frac{1}{n_j} \sum_{\substack{i \\ a_{ij} \neq @}} a_{ij}$, where $n_j = \sum_{\substack{i \\ a_{ij} \neq @}} 1$ (number of known data), i.e. b_j is determined as mean value in a column.

Setting practically arbitrary initial approximations for y_j and b_j , find value x_i then assuming y_j and b_j unknown find their values for fixed x_i and so on - these simple iterations converge.

The halting criterion is the same as for the problem (1).

Successive exhaustion of the matrix A. Find for a given matrix A the best approximation by the matrix P_1 of the form $x_i y_j + b_j$. Then find for $A - P_1$ the best approximation of the same form P_2 and so on. The checking can be done, e.g. by the residual dispersion of columns.

In the case of no gaps the method described brings to ordinary principal components - singular expansion of centered initial table of data. In this case, starting with $q=2$, $P_q = x_i^q y_j^q$ ($b=0$). In the general case this is not necessarily so. *It should be emphasized that centering (transition to zero mean values) is not applicable to the data with*

gaps.

Q-factor filling of gaps is their definition from the sum of Q of obtained matrices of the form $x_i y_j + b_j$,

Q-factor repairing of a table is its substitution with the Q of obtained matrices of the form $x_i y_j + b_j$.

Describe the procedure of recovery of data in the line b_j with gaps (some $a_j = @$) arriving for processing. Let there be constructed a sequence of matrices P_q of the form $x_i y_j + b_j$ ($P_q = x_i^q y_j^q + b_j^q$), exhausting the initial matrix A with a preset accuracy. For each q determine the number $x^q(a)$ and vector a_j^q with a given line:

$$a_j^0 = a_j \quad (a_j \neq @);$$

$$x^q(a) = \frac{\sum_{\substack{j \\ a_j \neq @}} (a_j^{q-1} - b_j^q) y_j^q}{\sum_{\substack{j \\ a_j \neq @}} (y_j^q)^2}; \quad (3)$$

$$a_j^q = a_j^{q-1} - b_j^q - x^q(a) y_j^q, \quad (a_j \neq @);$$

here the manifold M is a straight line, the coordinates of points on M are assigned by the parametric equation $z_j = t y_j + b_j$ and the projection $\text{Pr}_M(a)$ is defined by (2):

$$\text{Pr}(a) = t(a) y_j + b_j;$$

$$t(a) = \frac{\sum_{\substack{i,j \\ a_{ij} \neq @}} (a_j - b_j) y_j}{\sum_{\substack{i,j \\ a_{ij} \neq @}} (y_j)^2}.$$

For recovery of the Q - factor assume

$$\bar{a}_j = \sum_{q=1}^Q x^q(a) y_j^q + b_j^q, \quad (a_j \neq @). \quad (4)$$

When there are no gaps the derived straight lines are orthogonal and we have an orthogonal system of factors. For the incomplete data this is not the case, but the derived system of factors can be orthogonalized, the process consists in recovery of the initial table by the derived system of factors following which the system is again recalculated, now on complemented data.

1.3 Mechanical Interpretation

Consider mechanical concepts underlying construction of the modeling manifold.

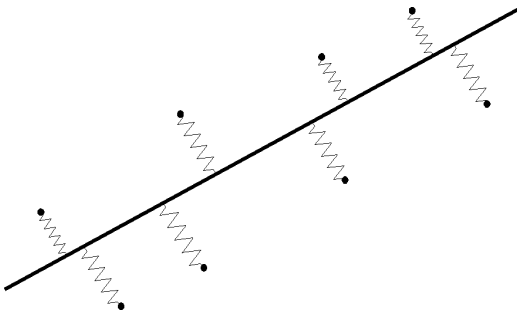


Figure 1

Let the space of data have a straight rigid beam (Fig.1). Let every datum be connected with the beam by a spring with the end of the spring moving along the beam. Fix the initial position of the beam and find such a position of springs that corresponds to the minimum of elastic energy. Then fix the position of spring ends on the beam, release the beam and let it regain mechanical equilibrium. Then fix the beam in a new position and again release the spring ends.

The system monotonically approaches equilibrium - minimum of its energy, as at every stage the strain energy decreases. The obtained beam is nothing but the first principal component. The projection of the datum is, at this, determined by the site where the spring corresponding to it is fixed on the beam.

The iteration finding of the coefficients x_i , y_j and b_j described in the previous section is fully analogous to the described mechanical process: the position of the beam is determined by the coefficients y_j , b_j and the points where the springs are fixed - by the coefficients x_i .

The data with gaps are modeled by hard rods (one gap), planes, etc. A corresponding spring can freely move along these objects. This means that the spring efficiently connects the shadow (projection) of the beam on the subspace of known data with the point in this subspace.

2 THE QUASILINEAR MODEL

2.1 The Principal Curve Method

Take the principal curve method as a prototype to construct the nonlinear models [1-3].

Definition 1: (Principal curve). Let X denote a random vector in the subspace R^d and let there be n data vectors. The principal curve $f \in R^d$ is a smooth (C^∞) curve (of dimension 1) in R^d parametrized by the parameter $\lambda \in A \subset R^1$, going through the middle of d - dimensional data described by X ,

$$f(\lambda) = \begin{bmatrix} f_1(\lambda) \\ \dots \\ f_d(\lambda) \end{bmatrix} = E\{X | \lambda_f(X) = \lambda\}, \quad (5)$$

where

$$\lambda_f(x) = \sup_{\lambda} \{\lambda: \|x - f(\lambda)\| = \inf_{\mu} \|x - f(\mu)\|\} \quad (6)$$

is the function of projection on the curve.

In brief, every point on the principal curve is mean value of all data points projected on it. The mean square error between the data points and their closest projections on the principal curve in the j -th iteration is as follows:

$$MSE^{(j)} = E\left\{\|X - f^{(j)}[\lambda_{f^{(j)}}(X)]\|^2\right\}. \quad (7)$$

The algorithm of constructing the principal curve is iteration involving calculation of mathematical expectation (5) and projection at each iteration.

2.2 A Method of Constructing Quasilinear Models

A simplest version of nonlinear factor analysis is super-structured over the linear one. It is proposed to use quasilinear models admitting simple explicit formulas for data processing based on the described algorithm of constructing linear models.

Let, as in the case of linear models, there be a table with gaps A . Quasilinear model is constructed in several stages:

- 1) To construct a linear model is to solve the problem (2). For the sake of certainty assume $(y, b)=0$, $(y, y)=1$ - this can be achieved always.
- 2) Interpolate (smoothen). Construct a vector-function $f(t)$ minimizing the functional

$$\Phi = \sum_{\substack{i,j \\ i \neq @}} \left(a_{i,j} - f_j \left(\sum_k a_{ik} y_k \right) \right)^2 + \alpha \int_{-\infty}^{+\infty} (f''(t))^2 dt, \quad (8)$$

where $\alpha > 0$ is the smoothening parameter.

The problem is solved by cubic splines whose coefficients are found from the equality to zero of corresponding partial derivatives Φ (8) on a certain uniform network.

The problem can also be solved with a polynomial of a small degree, however, even though such a solution makes possible to obtain a satisfactory interpolation with small calculation efforts, it cannot yield good extrapolation.

- 3) Extrapolation. The obtained function is extrapolated over the entire real axis.

2.2.1 Interpolation with a polynomial of small degree

In the case of approximating with a polynomial of degree n the problem is to best approximate the matrix A by the polynomials of the form $f_j(x) = f_n^j x^n + f_{n-1}^j x^{n-1} + \dots + f_1^j x + f_0^j$, i.e.:

$$\Phi = \sum_{\substack{i,j \\ a_{ij} \neq 0}} \left(a_{i,j} - f_j \left(\sum_k a_{ik} y_k \right) \right)^2 + \alpha \int_{-\infty}^{+\infty} (f''(t))^2 dt \rightarrow \min,$$

where $\alpha > 0$ - is the smoothening parameter.

2.2.2 Interpolation by cubic splines

The following smoothening problem is to be solved by cubic splines:

$$\Phi = \sum_{\substack{i,j \\ a_{ij} \neq 0}} \left(a_{i,j} - f_j \left(\sum_k a_{ik} y_k \right) \right)^2 + \alpha \int_{-\infty}^{+\infty} (f''(t))^2 dt \rightarrow \min,$$

where $\alpha > 0$ is the smoothening parameter, and f - the arbitrary smooth function.

2.3 The Problem of Extrapolation, Optimum Analytical Continuation and Carleman's Formula

The problem of extrapolating the available data beyond the limits of dispersion is well known. Its solution cannot be omitted - there is no guarantee that the data to follow will get exactly into the variation range of those available and it is not possible always limit the scope beyond this range. The necessity to construct new formulas for all possible data values is called by two circumstances more: first - the smoothened dependence constructed at the first stage is fundamentally interpolation and cannot be extrapolated, second - actually it carries in explicit form the information about each line of the data matrix. Smoothening, e.g. by merely a polynomial of a small degree by the least squares method is free from the second drawback (the information is "reduced" to several coefficients), but does not yield a good extrapolation.

Common (but fairly rough) extrapolation by the straight lines is: the resulting function $f(t)$ is extrapolated from the segment (e.g. $[a, b]$) over the entire real axis owing to the first approximation constructed at the ends of the segment $f(t) = f(a) + f'(a)(t-a)$ with $t < a$ и $f(t) = f(b) + f'(b)(t-b)$ with $t > b$.

Optimal extrapolation is of more interest. Its more strict statement calls for involvement of the problem of analytical extension of the function (from the finite set of points over a straight line or space). It is also convenient to pass over from the real variable t to a band on the plane

of complex variables.

So, under consideration is the problem of analytical extension of the function, set on an infinite sequence of points $\{t_k\}$ ($k=1,2,\dots$). It is required to construct a formula of extension from a *finite* set best in the following sense: the sequence of functions $f_n(t)$ obtained by extension from the set $\{t_k\}$ ($k=1,2,\dots,n$), converges faster than for all other formulas of this class. Of course, it takes to additionally define the following concepts: what are "formulas of this class", what convergence is meant, etc. All this has been done in relevant mathematical literature [6].

The smoothened vector-function $f(t)$ can be optimally extrapolated from a certain finite set $\{t_k\}$ (not necessarily related with projections over the straight line $z_j = ty_j + b_j$ of initial lines of data)

$$f(t) \approx ty + b + \sum_{k=1}^m (f(t_k) - t_k y - b) \times \\ \times \frac{2(e^{\lambda t} - e^{\lambda t_k})}{\lambda(e^{\lambda t} + e^{\lambda t_k})(t - t_k)} \prod_{\substack{j=1 \\ j \neq k}}^m \frac{(e^{\lambda t_k} + e^{\lambda t_j})(e^{\lambda t} - e^{\lambda t_j})}{(e^{\lambda t_k} - e^{\lambda t_j})(e^{\lambda t} + e^{\lambda t_j})}, \quad (9)$$

where λ is the parameter of the methods specifying the width of the band on the plane of complex numbers, where the extrapolated function is holomorphic with a guarantee (the width is π/λ).

Generally, to extrapolate according to Carleman for a set of nodes $\{t_k\}$ we take points uniformly placed on the segment, but not the initial experimental data. The values of $f(t)$ at these points we find by the interpolation formulas.

By Carleman's formulas we extrapolate the deviation of the curve $f(t)$ from the straight line $ty+b$. The Carleman's formulas provide for good extrapolation of analytical functions over the entire straight line (it cannot be guaranteed, of course, that in every specific case it is the formula (9) that will assure the best extrapolation, yet, there are several theorems that the formula (9) and relevant formulas yield the best approximation for different classes of analytical functions [6]).

Smoothing and extrapolation by Carleman's formula can be combined into a single process, i.e. interpolation and extrapolation can be done simultaneously.

2.4 Mechanical Interpretation

Assuming the beam to be able to deviate elastically from its straight form we have the following picture (Fig. 2)

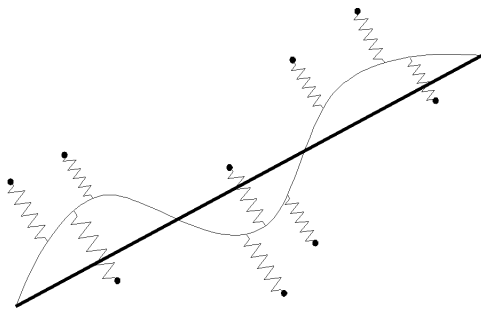


Figure 2

The points where the springs are fixed on the beam are determined by the projections on the straight beam (as the model is quasilinear).

The problem arises to determine behavior of the beam ends beyond the boundaries of the data range - the above described extrapolation problem.

2.5 Application of Quasilinear Models

A point on the constructed curve $f(t)$, corresponding to the complete ("complete") data vector a is constructed as $f(a, y)$, this is the quasilinearity of the method: First we find the projection of the data vector over the straight line $Pr(a)=ty+b, t=(a, y)$ after that we construct a point on the curve $f(t)$. This is also true for the incomplete data vectors - first we find on the straight line the nearest point $t(a)$, then - respective point on the curve $f(t)$ with $t=t(a)$.

After the curve $f(t)$ is constructed the data matrix is displaced with the matrix of deviation from the model. Then again we find the best approximation of the form $x_i y_j + b_j$ for the matrix of deviations, again construct smoothening, following which we extrapolate by Carleman and so on, until the deviations sufficiently approach zero.

As a result the initial table takes the form of the Q-factor model:

$$a_{ij} \cong \sum_q f_j^q(t_i^q). \tag{10}$$

If $a_{ij} \neq @$, this formula approximates the initial data, otherwise it yields a method of data recovery.

3 NEURAL CONVEYOR

The constructed algorithm makes possible neural network interpretation. Connected to each curve $f_q(t)$ is one summator (its weights are the coordinates of the vector y^q), a set of n free summands ("thresholds") - coordinates of the vector b^q and n nonlinear converters each of which

calculates one coordinate of a point on the curve by the formula (9). Such a "neuron" acts on the vector a of input signals (with gaps) as follows: $t(a)$ is calculated by the formula (1.6) (operation of the summator), then the nonlinear elements calculate $f^q(t(a))$ after that the difference $f_j^q(t(a))$ ($a_j \neq @$) is transmitted to the following neuron. With a traveling along this conveyor the sum of values of $f_j^q(t(a))$ ($a_j = @$) builds up. It is these sums that form the vector of output signals - the proposed values of the missing data. Should the need arise to make repairs of data the sum of values of $f_j^q(t(a))$ is built up for each coordinate j .

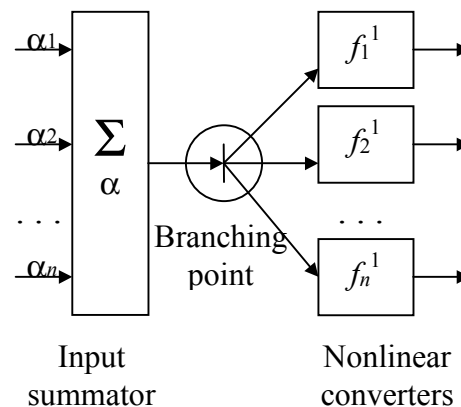


Figure 3

The structure of neurons is not standard (Fig. 3) - it has one input summator and n nonlinear converters (in compliance with the dimension of the data vector).

Operation of the summator is not quite ordinary either - for incomplete data vectors it calculates the scalar product with available data and performs additional normalization. This additional normalization of the input weights of the summator takes into account only those with known respective values of the input vector coordinates).

4 THE SELF-ORGANIZING CURVES (SOC)

4.1 The Idea of Self-Organizing Curves

The quasilinear factors are good for the "medium nonlinear problems" where the contribution of the linear part is relatively substantial (about 50% or more). For essentially nonlinear problems it is but natural to use some approximation of principal curves instead of linear principal components.

Instead of linear manifolds of small dimension we propose to use corresponding self-organizing maps (SOM) or, in other words - self-organizing curves (SOC). The method we use is somewhat different from the method of Kohonen maps [7] by more transparent physical interpretation and

explicit form of the variation principle.

Let SOC be defined by a set of points (kernels) successively placed on a curve (in the first approximation let SOC be simply a broken line Y) and it is required to map on it a set of points of data $X=\{x_i\}$. Introduce operator Π , which for every vector $x \in X$ associates the nearest to it point from Y :

$$x \xrightarrow{\Pi} y_j, \|y_j - x\|^2 \rightarrow \min, \quad (11)$$

each kernel y_j is associated with its taxon

$$K_j = \{x \in X | x \xrightarrow{\Pi} y_j\}. \quad (12)$$

The method of constructing SOC resembles the methods of dynamic kernels except for the additional restrictions on connectivity and elasticity. The minimized value is constructed from the following summands: the measure of data approximation -

$$D_1 = \sum_j \sum_{x \in K_j} \|x - y_j\|^2, \quad (13)$$

the measure of connectivity (the points close on the curve must get close on the data subspace) -

$$D_2 = \sum_j \|y_j - y_{j+1}\|^2, \quad (14)$$

the measure of nonlinearity -

$$D_3 = \sum_j \|2y_j - y_{j-1} - y_{j+1}\|^2. \quad (15)$$

So, to construct SOC we have to minimize the functional:

$$D = \frac{D_1}{|X|} + \lambda \cdot \frac{D_2}{m} + \mu \cdot \frac{D_3}{m} \rightarrow \min, \quad (16)$$

where λ , μ are the parameters of connectivity and nonlinearity - "moduli of elasticity" (division by the number of points $|X|$ and the number of kernels m means normalization "in one summand" and makes possible to use identical methods of varying λ and μ for samplings of different size).

When division of a set of data into taxons is fixed, SOC are constructed unambiguously - a simple linear problem is solved. When the position of kernels is also fixed the taxons can also be easily constructed by the formulas (11), (12). Dividing the problem into successive search of kernels - taxons - kernels we have an algorithm whose convergence is ensured by the fact that the D criterion decreases with its each step (16).

4.2 The Local Minimum Problem

As opposed to the linear and quasilinear cases this problem of minimizing the functional (4.6.) is not convex and there arise difficulties associated with getting to the local minimum range. This can result in unsatisfactory solution of the problem.

Even though there are numerous methods to solve this problem, we would rather dwell upon the multigrid method and the "annealing" method.

4.3 The Smoothing Problem

The obtained broken line $\{y_j\}$ ($j=1..m$) can be smoothed by different methods, e.g. by cubic splines. This, however, brings forth certain difficulties associated with finding projections of the data over the smoothed curve, as this requires to solve algebraic equations of the 5th degree.

This called for construction of the initial broken line of a continuous projector. Corresponding to the ends of the broken line are values (-1) and 1, respectively, the projections of nodes onto the segment [-1, 1] are determined by the nodes of uniform (considered is the number of the nodes of the broken line only) or non-uniform (account is also taken of the distance between the nodes of the broken line) grid.

Final smoothing is done by Carleman's formulas. The broken line $\{y_j\}$ ($j=1..m$) replaces, at this, the principal component of the classical method, while the smoothing process is analogous to construction of the quasilinear model.

4.4 Mechanical Interpretation

The above described linear and quasilinear models have a very strong restriction - the beam is either rigid or it is flexible, but is built along a straight line, which is essential e.g. in the case when the data are placed not along some straight line, but along a circumference (or along a strongly curved arc, at least).

To circumvent it the beam should be elastic (to be determined not by a straight line, but by a curve). This presents difficulties in defining the distance from the point to the curve in space (even more so, when it is not a point but a linear manifold).

In the above described method close to the method of Kohonen's self-organizing maps the sought for elastic beam is presented in the form of a broken line whose nodes are freely connected with the data (Fig. 4).

In analogy to the rigid case in several iterations the system regains its equilibrium. Their number is, at this, finite, as at every step the springs decrease their total energy, and the number of possible states (methods to fix the nodes of the broken line by the springs to the data) is finite.

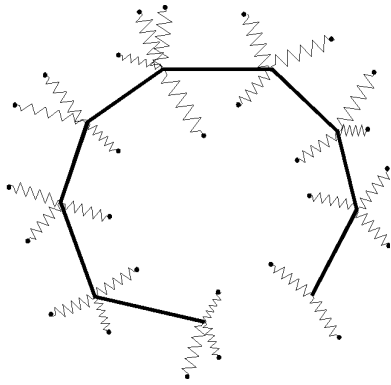


Figure 4

The introduced moduli of elasticity represent the degree of attraction of the nodes of the broken line to each other and the degree of resistance to bending at the nodes, respectively.

5 EXPERIMENTAL RESULTS

Illustrate the process of modeling the data with gaps on the basis of the table of presidential elections in the US with 31 election situation (from 1860 to 1980). Each election in the table contains data of 12 binary features [10]:

1. Has the presidential party (P-party) been in power for more than one term? (More1)
2. Did the P-party receive more than 50% of the popular vote in the last election? (More50)
3. Was there significant activity of a third party during the election year? (Third)
4. Was there serious competition in the P-party primaries? (Conc)
5. Was the P-party candidate the president at the time of the election? (Prez)
6. Was there a depression or recession in the election year? (Depr)
7. Was there an average annual growth in the gross national product of more than 2.1% in the last term? (Val2.1)
8. Did the P-party president make any substantial political changes during his term? (Chan)
9. Did significant social tension exist during the term of the P-party? (Wave)
10. Was the P-party administration guilty of any serious mistakes or scandals? (Mist)
11. Was the P-party candidate a national hero? (R.Hero)
12. Was the O-party candidate a national hero? (O.Hero)

The table also contains information about results of elections (victory of the presidential or opposition party).

The values of the binary features are equal to -1 (answer "no" for the input feature or the victory of the presidential party) and to 1 (answer "yes" for the input feature or the victory of the opposition).

The models constructed by this table confidently predicted results of the second election of Reagan, victory of Bush over Ducacis, both victories of Clinton [11].

The degrees of approximation (in per cent of the initial value) of the table by several factors according to the model (Table 1) were as follows (Table 1).

If the error in the calculated value of the feature is less than 50%, this means that it is the exact value (the features are qualitative, therefore, with the error less than 50% the sign of prediction defines exact value).

For satisfactory prediction by linear models suffice are 4 factors, which indicates that this is a 4-factor problem (in the ordinary meaning of this word).

The quasilinear models, SOC-based models, in particular, usually do with only one nonlinear factor to predict satisfactorily.

The produced sets of factors were tested as follows:

1. A model was constructed on a complete table, then gaps were randomly added into the table, following which the procedure of filling the gaps was launched. As a result the obtained values were compared to the initial values.

The testing demonstrated that up to 25% of gaps (of the total number of initial data) were satisfactorily filled by linear and quasilinear models. For SOC-based models this index is 50%. I.e. even every other datum is eliminated from the table, still it can be recovered with satisfactory accuracy.

2. Gaps were randomly introduced into the table, then a model was constructed on the basis of "gap" table, following which the procedure of filling the gaps was launched. As a result the obtained values were compared to the initial values. The filling was satisfactory when the gaps amounted to 10% of the total number of initial data.

The method was actually tested on the problem of predicting complications in myocardial infarction. The table of data on complications in myocardial infarction presents observation of 1700 patients by 126 parameters. Experiments demonstrated that the first 15-20 quasilinear factors are sufficient to satisfactorily repair most values of features in the table 1.

DISCUSSION

A method applicable to fill gaps and repair data with gaps has been developed. Three different versions of the method - from simplest linear models to the method of principal curves for the data with gaps have been presented. The neural network implementation of the method makes possible easy construction of its parallel implementations.

Table 1

№	Feature	Approximation (%) depending on the number of factors								
		Linear model			Quasilinear model			SOC		
		1	4	10	1	4	10	1	4	10
1	Morel	11.88	59.98	77.36	25.37	63.75	95.91	53.49	80.81	96.85
2	More50	9.07	61.10	79.43	14.99	73.69	95.18	30.77	75.89	95.12
3	Third	29.66	4489	91.56	31.73	6697	97.45	32.94	76.83	96.93
4	Conc	62.30	63.28	77.84	69.51	77.12	90.24	72.63	78.86	95.42
5	Prez	31.72	59.68	80.27	45.58	68.01	93.03	56.08	74.85	95.18
6	Depr	32.17	52.43	93.38	37.95	71.08	95.56	58.86	80.53	95.62
7	Val2_1	4.12	37.67	94.19	6.27	69.22	96.53	28.80	72.23	95.44
8	Changes	2.33	49.87	86.19	16.81	61.15	94.95	13.01	72.01	93.77
9	Wave	25.13	62.33	80.34	33.18	66.82	95.65	32.68	63.96	96.71
10	Mist	50.34	61.05	86.17	64.83	70.52	97.55	60.80	81.35	96.90
11	R_Hero	33.35	48.12	90.69	54.86	66.27	97.52	27.30	83.67	95.76
12	O_Hero	36.55	50.07	92.03	45.69	68.41	97.55	52.22	76.42	96.17
13	Answer	69.22	69.96	81.78	92.27	9285	96.72	97.82	98.43	99.50

The given algorithm of filling gaps - as opposed to many other algorithms designed for this purpose - does not involve its a priori filling. However, it calls for preliminary normalization ("dedimensionalizing") of the data - transition in each column of the table to a "natural" unit of measurement. It is noteworthy that data centering cannot turn the problem of processing data with gaps into a homogeneous problem.

Of great interest is the question: how many summands (principal curves) should be taken to process the data? There are several versions of answers, yet most of them are subject to the heuristic formula: *the number of summands must be minimal among those that provide for satisfactory (tolerant) testing of the method with the known data*. Such a principle of "minimum sufficiency" is specific for many neural network applications [8,9].

The method developed manifests itself in the form of an "anzatz" - a sentence, but not a series of theorems. This is not incidental - we propose a technology of constructing plausible evaluations of missing data, but not of their unknown real value.

REFERENCES

1. *Hastie T., Stuetzle W.* Principal curves. Journal of the American Statistical Association. 1988, Jun. V. 84, No. 406. PP.502-516.
2. *LeBlank M., Tibshorany N.* Adaptive principal surfaces. Journal of the American Statistical Association. 1994, Mar. V. 89, No. 425. PP. 53-64.
3. *Kramer M.A.* Nonlinear principal component analysis using autoassociative neural networks. AIChE Journal. 1991. V.37, No. 2. PP. 233-243.
4. *Gorban A.N., Makarov S.V., Rossiev A.A.* Neural conveyor to recover gaps in tables and construct regression by small samplings with incomplete data // *Matematika. Computer. Obrazovanie. Vyp. 5. Part II. Selected Transactions / Ed. G.Yu. Riznichenko. M.: Progress-Traditsiya Publishers, 1998. PP. 27-32.*
5. *Rossiev A.A.* Modeling data by curves to recover gaps in tables // *Neuroinformatics methods / Ed. A.N. Gorban, Krasnoyarsk: KSTU Press, 1998. PP.6-22.*
6. *Aizenberg L.A.* Carleman's Formulas in Complex Analysis. Theory and Applications / Ed. M. Hazewinkel - Kluwer Academic Publishers-Dordrecht/Boston/London, 1993, 300 p.
7. *Kohonen T.* Self-Organizing Maps. Springer: Berlin - Heidelberg, 1997.
8. *Gorban A.N., Rossiev A.A.* Neural networks on PC. Novosibirsk:Nauka, 1996. 276 p.
9. *Neuroinformatics / A.N. Gorban, V.L. Dunin-Barkovsky, E.M. Mirkes et al.* Novosibirsk: Nauka (Sib. Otd-nie), 1995. 256 p.
10. *Lichtman A.J., Keilis-Borok V.I.,* Pattern Recognition as Applied to Presidential Elections in U.S.A., 1860-1980; Role of Integral Social, Economic and Political Traits, Contribution N 3760. 1981, Division of Geological and Planetary Sciences, California Institute of Technology.
11. *Gorban A.N., Waxman C.* Neural Networks for Political Forecast. Proceedings of the WCNN'95 (World Congress on Neural Networks'95, Washington DC, July 1995). P.176- 178.

Надійшла 08.02.2000

Після доробки 01.03.2000

ПРИНЦИП ВАРЬИРОВАНИЯ ДАННЫХ В ПРИКЛАДНЫХ ЗАДАЧАХ ИДЕНТИФИКАЦИИ

А.Е. Архипов, С.А. Архипова

Рассмотрен подход к решению ряда проблемных вопросов структурно-параметрической идентификации в условиях минимума априорной информации о свойствах исходных данных путем генерации однородных с имеющимся набором псевдоданных. Получаемое на них множество псевдорешений позволяет радикально восполнить и использовать недостающую информацию для практического решения задачи.

Розглянуто підхід до розв'язку ряду проблемних питань структурно-параметричної ідентифікації в умовах мінімуму априорної інформації про властивості вихідних даних шляхом генерації однорідних із наявним набором псевдоданих. Одержувано на них множина псевдорішень дозволяє радикально заповнити і використати відсутню інформацію для практичного розв'язку задачі.

The approach to a solution of a series of problems is considered in structural - parametrical identifications in conditions of a minimum of an a priori information about properties of input data by generation homogeneous with an available gang of pseudodata. The set, obtained them, of pseudosolutions allows radically to fill and to use a missing information for a practical solution of a task.

ВВЕДЕНИЕ

В научно-исследовательской практике часто встречаются задачи, относящиеся к направлению, называемому анализом данным [1, 2].

Его содержание - приемы и методы обработки данных наблюдений многомерных систем на ранних стадиях их исследования, когда имеющаяся у обработчика информация исчерпывается только сведениями, зафиксированными в этих данных, без наличия каких-либо дополнительных знаний о механизме функционирования изучаемых систем и характеристиках данных. Последнее, в частности, статистическое описание погрешностей данных, резко сужает возможности применения традиционных математико-статистических методов обработки.

Например, пусть для отображения связей между признаками (информационными переменными), характеризующими состояние системы, применяется распространенная в практических приложениях аппроксимативная регрессионная модель общего вида [3]

$$Z = \mu(X, A) + E, \quad (1)$$

где $X = [X_1, \dots, X_p]$ - вектор входных переменных модели $\mu(X, A)$, $Z = Y + E$ - фактически наблюдаемая выходная координата, Y - выход модели, E - случайная составляющая с математическим ожиданием $M\{E\} = 0$

и дисперсией $D\{E\} = \sigma_E^2$. Модель $\mu(X, A)$ аппроксимирует неизвестное отображение $\varphi: X \rightarrow Y$, реализуемое исследуемой системой, $A = [a_0, \dots, a_k]^T$ - вектор параметров (коэффициентов) модели.

Для уточнения структуры модели применительно к конкретно решаемой задаче следует в соответствии с методами многомерного регрессионного анализа [3,4,8], используя информацию о виде распределения F_E , построить критериальные статистики, позволяющие судить о степени адекватности модели исходным данным и о значимости коэффициентов. Наиболее разработанный аппарат линейного регрессионного анализа, соответствующий модели вида

$$\mu(X, A) = a_0 + a_1 x_1 + \dots + a_p x_p, \quad (2)$$

ориентирован на нормальность погрешности E и связанные с ней статистики, имеющие распределение Фишера и Стьюдента [3, 8]. Очевидно, что отсутствие информации о распределении F_E исключает применение каких-либо методов оценивания структуры модели, опирающихся на знание закона F_E .

Уязвимыми в этом случае оказываются и основные процедуры оценивания параметров модели, построенные с применением принципа максимума правдоподобия, также использующего информацию о виде распределения F_E [6, 8].

ПОСТАНОВКА ЗАДАЧИ

Возможным выходом в сложившейся ситуации является применение так называемого "прагматического" подхода [9] к задаче параметрической идентификации (ПИ), суть которого в оценивании вектора A параметров модели на одном и том же наборе исходных данных $[z_i, x_i, \dots, x_{ip}]$, $i = \overline{1, n}$ несколькими разными методами $\omega_1, \omega_2, \dots$ с последующим определением лучшего метода ω^+ путем анализа качества полученных оценок $\tilde{A}(\omega_1), \tilde{A}(\omega_2), \dots$ с использованием критериальных статистик, свободных от вида распределения F_E . Множество методов $\{\omega_1, \omega_2, \dots\} = \Omega$ составляется из апробированных

методов ПИ, эффективность которых доказана практикой.

Аналогичный подход применим и для селекции лучшей структуры модели $\mu^+(X, A)$ из множества моделей-претендентов $\{\mu_j(X, A_j)\}$ [9,11], с той лишь разницей, что в этом случае выбор модели должен производиться по результатам интегральной оценки качества решения задачи структурно-параметрической идентификации (СПИ), учитывающий как точность модели, так и качество ПИ [9-12].

Таким образом, актуальной является разработка методик оценивания качества ПИ и СПИ, которые, с учетом изложенного выше подхода к решению задач ПИ и СПИ, условий минимума априорных сведений об объекте исследования и характеристиках исходных данных, должны быть:

- а) свободны от вида закона распределения F_E ;
- б) применимы для широкого класса модельных структур;
- в) внешние по отношению к используемым методам ПИ, т.е. не должны зависеть от применяемого способа оценивания параметров модели;
- г) практически реализуемы и эффективны на выборках исходных данных конечного объема.

ПРИНЦИП ВАРЬИРОВАНИЯ ДАННЫХ

Один из подходов к решению поставленной выше проблемы - применение принципа варьирования данных [9, 10] для получения возможно более полного набора статистических характеристик результатов решения конкретной задачи ПИ, СПИ, включая распределения оценок параметров $F(a_0), F(a_1), \dots$, моменты соответствующих распределений, показатели точности и адекватности моделей исходным данным, т.п.

Суть принципа варьирования данных состоит в использовании полученных в ходе реального эксперимента исходных данных, заданных матрицей $[z_i, x_i] = [z_i, x_{i1}, \dots, x_{ip}]$, $i = \overline{1, n}$, для имитации результатов выполнения серии из L экспериментов, обработке "полученной" в каждом из них матрицы псевдоданных $[z_i, x_i]^{(l)}$, $l = \overline{1, L}$ каким либо одним выбранным методом ПИ и вычислении на найденном множестве векторных оценок $\{\tilde{A}_l\} = \{\tilde{a}_{l0}, \dots, \tilde{a}_{lk}\}$, $l = \overline{1, L}$ эмпирических функций распределения $\tilde{F}_n(\tilde{a}_0), \tilde{F}_n(\tilde{a}_1), \dots, \tilde{F}_n(\tilde{a}_k)$, выборочных дисперсий $\tilde{D}_n\{\tilde{a}_i\}$ и ковариаций $\tilde{\text{cov}}_n\{\tilde{a}_j, \tilde{a}_g\}$, $j \neq g$, а также других статистик, характеризующих качество ПИ и СПИ.

Один из вариантов практической реализации варьирования данных состоит в том, что выборка исходных дан-

ных $[z_i, x_{i1}, \dots, x_{ip}]$, $i = \overline{1, n}$ рассматривается как генеральная совокупность, из которой путем равновероятного выбора строк с возвращением формируется требуемое количество L матриц псевдоданных. Случайный выбор очередного номера строки осуществляется генератором псевдослучайных целых чисел, равномерно распределенных в интервале $[1, n]$. В соответствие с очередным выпавшим номером строки за строкой формируются матрицы псевдоданных.

Согласно обобщению на многомерный случай теоремы Гливленко-Кантелли [7] о сходимости эмпирической функции распределения к истинной и ряда следствий из этой теоремы, устанавливающих существование аналогичной сходимости широкого класса статистик (выборочных оценок) к своим "теоретическим" значениям, можно утверждать, что с ростом L происходит неограниченное сближение эмпирических распределений $\tilde{F}_n(\tilde{a}_0), \dots, \tilde{F}_n(\tilde{a}_k)$ с теоретическими распределениями $F_n(\tilde{a}_0), \dots, F_n(\tilde{a}_k)$, полностью определяемыми многомерным распределением $F_n(z, X)$ исходной выборки $[z_i, x_{i1}, \dots, x_{ip}]$, $i = \overline{1, n}$: $P\{\sup_n |F_n(\tilde{a}_j) - \tilde{F}_n(\tilde{a}_j)| \rightarrow 0\} = 1$, $j = \overline{0, k}$. Отсюда следует сходимость выборочных дисперсий $\tilde{D}_n\{\tilde{a}_j\}$ и ковариаций $\tilde{\text{cov}}_n\{\tilde{a}_j, \tilde{a}_g\}$, $j \neq g$ к теоретическим значениям $D_n\{\tilde{a}_j\}$, $\text{cov}_n\{\tilde{a}_j, \tilde{a}_g\}$. В реальных задачах теоретическое распределение $F_n(a_j)$ и соответствующие теоретические значения моментных характеристик обычно не поддаются определению (исключение составляет математические ожидания для распределений $F_n(a_j)$, $j = \overline{0, k}$, равные рассчитанным по исходным данным оценкам $\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_k$). Поэтому метод варьирования является единственным способом получения информации о статистических характеристиках оценок $\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_k$.

При $n \rightarrow \infty$ на основании той же теоремы Гливленко-Кантелли и упомянутых следствий существует сходимость распределений $F_n(a_j)$ к теоретическим распределениям $F(a_j)$, полностью определяемым неизвестным исходным распределением $F(z, X)$, а также сходимость соответствующих моментных характеристик и статистик, т.е. для этого случая рассчитанные по псевдоданным статистические характеристики оценок $\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_k$ являются состоятельными и асимптотически несмещенными оценками неизвестных распределений $F(a_j)$, $j = \overline{0, k}$ и их моментов.

Не в асимптотике, при конечных значениях n , нельзя сказать что-либо определенное о надежности результа-

тов, получаемых с привлечением варьирования данных, если оставаться в рамках традиционной схемы исследований, в которой за опорные приняты результаты, соответствующие некоторой гипотетической совокупности, нахождение фактических значений которых не представляется возможным. С другой стороны, единственный реальный источник информации о свойствах данных - это только сама исходная выборка (по условиям решаемых задач, любая дополнительная информация отсутствует), и вопрос в том, как обеспечить извлечение этой информации в максимально полном объеме и в форме, удобной для последующего использования. С этих позиций метод варьирования данных можно рассматривать как выборочный вариант метода максимального правдоподобия, позволяющий оптимизировать процедуру СПИ (в частности, выбрать метод ПИ, обеспечивающий наименьшую дисперсию оценок параметров на варьированных данных) без домысливания характеристик исходных данных путем описания их с помощью параметризованных моделей, приводящих к искажению фактической информации.

Для конечных значений $n \geq 10$ исследование применимости метода варьирования данных на тестовых примерах, имитирующих решение задач ПИ и СПИ, показывает достаточно хорошие результаты, зависящие однако от используемых методов варьирования данных.

Кроме описанного выше метода случайного варьирования строк матрицы исходных данных (ВСМ), можно предложить и другие методы. Один из них - метод случайного варьирования значений шума (ВЗШ) в исходных данных, где последовательность невязок $[\epsilon_1, \dots, \epsilon_n]$, $\epsilon_i = z_i - \mu(x, \tilde{A})$, выступает в качестве генеральной совокупности, из которой генерируются псевдо-выборки шума $[e_1, \dots, e_n]^{(1)}, \dots, [e_1, \dots, e_n]^{(L)}$ а затем формируются псевдовыборки зависимой переменной $[z_i]^{(l)} = [\tilde{y}_1 + \tilde{e}_1^{(l)}, \dots, \tilde{y}_n + \tilde{e}_n^{(l)}]$, где $\tilde{y}_i = \mu(x_i, \tilde{A})$, $i = \overline{1, n}$, $l = \overline{1, L}$ и матрицы псевдоданных $[z_i, x_i]^{(l)} = [[z_i]^{(l)}, x_i]$.

В методе генерации псевдощумов (ГПШ), в отличие от ВЗШ, генерация значений шума реализуется специальным генератором псевдослучайных чисел в соответствие с непараметрически заданным распределением F_ϵ , оцененным по выборке $[\epsilon_1, \dots, \epsilon_n]$.

Методы ВЗШ и ГПШ при варьировании данных сохраняют неизменными матрицы плана $[x_i]$ и обеспечивают намного более точное, по сравнению с ВСМ, оценивание распределений $F(\tilde{a}_j)$, $j = \overline{0, k}$ их моментных характеристик, однако при этом они весьма чувствительны к загроблению модели. Поэтому при выполнении подбора модели более надежным является применение ВСМ.

При задании исходных данных в виде временных

рядов варьирование данных осуществляется методом ГПШ путем разделения трендовой и случайной составляющих, генерации множества псевдореализаций последней с помощью специальных генераторов, в которые в непараметрической форме вводятся описания характеристик случайных составляющих (включая их автокорреляционные функции), и наложения псевдореализаций случайной составляющей на выделенный тренд [12].

ПРИМЕНЕНИЕ ВАРЬИРОВАНИЯ ДАННЫХ ДЛЯ ОПТИМИЗАЦИИ РЕШЕНИЯ ЗАДАЧИ ИДЕНТИФИКАЦИИ

Полученные методом варьирования исходных данных статистические характеристики оценок $\tilde{A}(\omega_1), \tilde{A}(\omega_2), \dots$, найденных соответственно методами $\omega_1, (\omega_2), \dots$, в частности, вектор дисперсий оценок параметров $\sum(\omega_r) =$

$$= [\sigma_r^2\{a_0\}, \dots, \sigma_r^2\{a_k\}],$$

позволяют реализовать селекцию лучшего метода ω^+ путем, например, введения обобщенного показателя эффективности методов ПИ. При работе с линейными моделями вида (2) для свертывания вектора $\sum(\omega_r)$ в скаляр в [9,12] предложен показатель

$$V^2(\omega_r) = \frac{1}{k} \sum_{j=0}^k \frac{\tilde{\sigma}_r^2\{\tilde{a}_j\}}{\tilde{a}_{jr}^2}, \quad (3)$$

где $\tilde{\sigma}_r^2\{\tilde{a}_j\} = \frac{1}{L-1} \sum_{l=1}^L (\tilde{a}_{ljr} - \tilde{a}_{jr})^2$ оценка a_{jr} рассчитана

методом ω_r по исходной выборке, а оценки a_{1jr}, \dots, a_{Ljr} - по псевдоданным.

Для моделей $\mu(X, A)$ более сложной структуры наряду с (3) возможно применение показателя вида

$$W(\omega_r) = \sum_{j=0}^k w_j \tilde{\sigma}_r^2\{\tilde{a}_j\} / S_z^2, \quad (4)$$

где $w_j = (\partial \mu(X, A) / \partial a_j)^2 \Big|_{\tilde{A}, \bar{X}}$, $S_z^2 = \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2$,

черта сверху обозначает средние выборочные значения соответствующих переменных. Селекция лучшего метода ω^+ осуществляется по минимуму соответствующего показателя.

Для выбора лучшей модели μ^+ на множестве моделей-претендентов $\{\mu_t\}$, $t = \overline{1, T}$ в общем случае можно использовать эвристический показатель качества СПИ, комплексно учитывающий точность модели и степень устойчивости процедуры ПИ [9, 12]:

$$Q_t = \gamma I_{nt} / S_Z^2 + (1 - \gamma) V_t^2(\omega_r), \quad t = \overline{1, T}, \quad (5)$$

где $I_{nt} = \sum_{i=1}^n [z_i - \mu_t(x_i, \tilde{A}_t)]^2 / n$, γ - весовой коэффициент,

определяющий приоритет требований к точности модели либо к устойчивости процедуры оценивания параметров. Исходя из рассчитанного для T моделей множества $\{I_{nt}\}$ один из вариантов задания γ можно описать выражениями [12]: $\gamma = 1 - I_{n\min} / I_{n\max}$, $I_{n\min} = \min\{I_{nt}\}$, $I_{n\max} = \max\{I_{nt}\}$, $t = \overline{1, T}$. При примерно одинаковой точности аппроксимации моделями исходных данных $\gamma \rightarrow 0$ и приоритет в выборе модели принадлежит показателю V^2 . При существенном разбросе значений I_n основную роль в выборе модели играет показатель I_n .

Возможно построение итеративной процедуры селекции с поэтапным удалением наиболее неудачных вариантов модели.

Для линейных регрессионных моделей выбор лучшей можно реализовать более строго, используя для этой цели средний риск $I = M\{[z_i - \mu_t(x, A)]^2\}$. В [10, 11] показано, что при оценивании параметров модели по методу наименьших квадратов справедливо соотношение

$$I = M(I_n) + \frac{1}{n} M\{(\tilde{A} - A)^T x_i^T x_i (\tilde{A} - A)\}. \quad (6)$$

Используя варьирование данных, на базе (6) строится показатель качества СПИ вида [10, 11]:

$$\tilde{I} = \frac{1}{L} \sum_{l=1}^L I_n^{(l)} + \frac{1}{n(L-1)} \sum_{l=1}^L (\tilde{A}^{(l)} - \tilde{A})^T (x_i^T x_i)^{(l)} (\tilde{A}^{(l)} - \tilde{A}). \quad (7)$$

Обозначая второе слагаемое через Δ и принимая во внимание, что с ростом L первое слагаемое стремится к I_n , рассчитанному по реальным исходным данным, получаем:

$$\tilde{I} = I_n + \Delta. \quad (8)$$

Также как и Q , показатель \tilde{I} комплексно учитывает требования к точности модели (составляющая I_n) и устойчивости процедуры ПИ (Δ), причем в Δ интегрировано три фактора, влияющие на качество ПИ:

- а) сложность модели;
- б) характер информационной матрицы $x_i^T x_i$ (степень коллинеарности исходных данных);
- в) устойчивость используемых алгоритмов ПИ.

Поэтому показатель Δ целесообразно использовать и

самостоятельно при подборе метода ω^+ . В условиях обычной регрессионной задачи "эксплуатационных" свойства Δ гораздо выше, чем у V^2 , однако при наличии таких особенностей данных, как случайные погрешности в результатах наблюдения входных переменных, результативность селекции по показателю Δ резко падает [11].

Кроме того, зашумление входных переменных может приводить к появлению смещения оценок параметров, отсутствующего при точном измерении значений X [12]. Применение варьирования данных позволяет построить достаточно простые, но эффективные процедуры выявления и коррекции смещения оценок параметров моделей [12, 13].

ВЫВОДЫ

Применение варьирования данных к решению прикладных задач идентификации в условиях минимума априорной информации о свойствах обрабатываемых данных позволяет получить объективные результаты при рассмотрении ряда ключевых проблем структурно-параметрической идентификации в ситуации, когда иные средства решения практически отсутствуют.

Принцип варьирования данных является внешним по отношению к методам и алгоритмам оценивания структуры и параметров модели, так как не ориентирован на какие-либо элементы применяемого для решения этих задач инструментария.

Использование исключительно непараметрических методов варьирования данных сводит к минимуму возможности субъективно влияния исследователя на полученные результаты, исключает искажения исходной информации вследствие формального задания параметризованной модели случайной составляющей в реальных данных.

ПЕРЕЧЕНЬ ССЫЛОК

1. Статистический словарь/Гл. ред. М.А. Королев. - 2-е изд., перераб. и доп. - М.: Финансы и статистика. -1989. - 623с.
2. Прикладная статистика: Классификация и снижение размерности: Справ. изд. / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. - М.: Финансы и статистика, 1989. - 607 с.
3. Прикладная статистика: Основы моделирования и первичная обработка данных. Справочное изд. / С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин. - М.: Финансы и статистика, 1983. - 471 с.
4. Демиденко Е.З. Линейная и нелинейная регрессия. - М.: Финансы и статистика, 1981. -302 с.
5. Демиденко Е.З. Оптимизация и регрессия. - М.: Наука, 1989. - 296 с.
6. Мудров В.И., Кушко В.Л. Методы обработки измерений. -М.: Сов. радио. 1976. -190 с.
7. Боровков А.А. Математическая статистика. - Учебник. - М.: Наука. Главная редакция физ.-мат. литературы, 1984. - 472с.
8. Пугачев В.С. Теория вероятностей и математическая статистика. - М.: Наука, 1979. - 496 с.
9. Архипов А.Е., Архипова С.А. Анализ и оптимизация качества решения задачи идентификации. // Праці П'ятої Української конференції з автоматичного управління "Автоматика-98": Київ, 13-16 травня 1998р. - ч. III, - Київ: видавництво

- тво НТУУ "Київський політехнічний інститут", 1998, - с. 9-15.
10. Архипов А.Е., Архипова С.А. Идентификация аппроксимативных моделей методом варьирования данных. // Адаптивные системы автоматического управления. // Межвидом. научно-техн. зб. - Дніпропетровськ: Системні технології, 1998. - Вип. 1 (21). с. 81-86.
 11. Архипов А.Е., Архипова С.А. Селекция структуры линейной регрессии путем случайного варьирования исходных данных. // Системні технології. Дослідження динаміки і оптимізація параметрів технологічних процесів: Збір. наук. праць. - Вип.2 - Дніпропетровськ: "Системні технології", 1998. - с. 3-9.
 12. Архипов А.Е. Выбор, оценка качества и методы повышения точности аппроксимативных моделей в задачах идентификации. - К.: УМК ВО, 1992. - 56 с.
 13. Архипов А.Е. Коррекция смещений коэффициентов регрессии, обусловленных ошибками в независимых переменных. Тези доповідей 1-ої Української конференції з автоматичного керування. АВТОМАТИКА-94 (Київ, 12-23 травня 1994 р.) - Київ, 1994, с.123.

Надійшла 02.03.2000

Після доробки 10.03.2000

УДК 681.513.6

ПРОГНОЗИРУЮЩАЯ НЕЙРОННАЯ СЕТЬ И АЛГОРИТМЫ ЕЕ ОБУЧЕНИЯ

Е. В. Бодянский, С. В. Попов

Предлагается подход к решению задач прогнозирования и раннего обнаружения изменения свойств нестационарных стохастических последовательностей, описываемых нелинейными разностными уравнениями авторегрессии - скользящего среднего (NARMA-модели). Предполагается, что прогнозируемая последовательность является нестационарной в широком смысле, причем изменения структуры, свидетельствующие о возникшей разладке, происходят в заранее неизвестные моменты времени. Предлагается архитектура рекуррентной искусственной нейронной сети и новые алгоритмы ее обучения, позволяющие с высокой точностью и быстродействием получать прогнозы контролируемого сигнала и обнаруживать моменты изменения его свойств.

Пропонується підхід до вирішення задачі прогнозування і раннього виявлення змін властивостей нестационарних стохастичних послідовностей, що описуються нелінійними різнице-вими рівняннями авторегресії - ковзного середнього (NARMA-моделі). Припускається, що прогнозована послідовність є нестационарною в широкому розумінні, при чому зміни структури, що свідчать про виникаюче розладання, відбуваються у заздалегідь невідомі моменти часу. Запропоновано архітектуру штучної нейронної мережі і нові алгоритми її навчання, що дозволяють з високою точністю та швидкодією отримувати прогнози контрольованого сигналу та виявляти моменти зміни його властивостей.

An approach to prediction and early properties changes detection problems solution for nonstationary stochastic sequences that are described by nonlinear difference autoregression - moving average equations (NARMA-models) is proposed. The forecasted sequence is assumed to be nonstationary in a wide sense, and structure changes pointing to emerging faults are assumed to occur in moments, unknown a priori. Architecture of an artificial neural network and its learning algorithms are proposed that enable obtaining high quality predictions of the monitored signal and detection of its properties changes moments.

ВВЕДЕНИЕ

Задача раннего обнаружения разладок достаточно часто встречается на практике и, прежде всего, в технической диагностике механизмов и машин и тесно связана с проблемой адаптивного прогнозирования нестационарных стохастических последовательностей. Для решения этой проблемы предложено множество подходов, свя-

занных в значительной мере с методами математической статистики, теории случайных процессов, адаптивной идентификации, однако жесткие предположения о стохастических свойствах рядов и их структуре, как правило, линейной, ограничивают возможности традиционного адаптивного подхода.

Последние годы характеризуются всплеском исследований в области теории и практики искусственных нейронных сетей, в том числе и для решения задач прогнозирования [1-9]. Предложенные нейронные сети, обеспечивая высокое качество прогнозирования, тем не менее, не способны обнаруживать и сигнализировать об изменениях свойств контролируемого сигнала, что совершенно необходимо в задачах априорной диагностики в реальном времени.

В данной работе предложена архитектура прогнозирующей рекуррентной искусственной нейронной сети и алгоритмы ее обучения, позволяющие не только получать прогноз нелинейной стохастической последовательности, но и обнаруживать моменты изменения ее свойств по ходу процесса обучения и прогнозирования. Предлагаемая сеть обладает высокой скоростью обучения и обеспечивает оптимальную точность прогноза на принятом классе моделей.

1 АРХИТЕКТУРА ПРОГНОЗИРУЮЩЕЙ НЕЙРОННОЙ СЕТИ

Архитектура предлагаемой прогнозирующей искусственной нейронной сети приведена на рис. 1 и представляет собой сеть элементарных нейронов, отличающихся видом функций активации и алгоритмами обучения, являющимися в общем случае рекуррентными процедурами безусловной или условной оптимизации.

Прогнозируемая стохастическая последовательность $\{z(k)\}$, $k = 1, 2, 3, \dots$ предварительно обрабатывается входным нейроном С таким образом, что все ее значения (прошлые и будущие) принадлежат интервалу $[-1, +1]$.

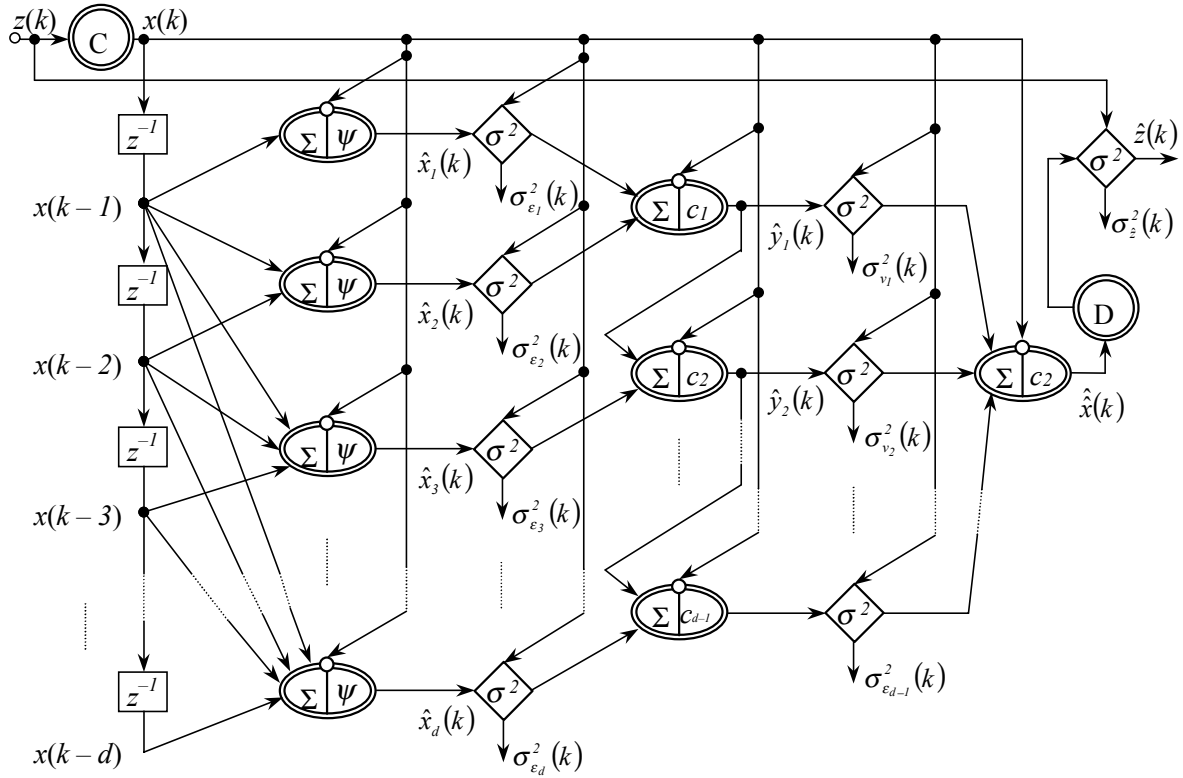


Рисунок 1 - Рекуррентная прогнозирующая нейронная сеть

Для этого в реальном времени вычисляются экстремальные значения последовательности $\{z(k)\}$.

$$\begin{cases} z^{\max}(k) = \max\{z^{\max}(k-1), z(k)\}, \\ z^{\min}(k) = \min\{z^{\max}(k-1), z(k)\}, \end{cases} \quad (1)$$

после чего сигнал кодируется так, что

$$x(k) = \frac{2z(k) - z^{\max}(k) - z^{\min}(k)}{z^{\max}(k) - z^{\min}(k)}, \quad (2)$$

при этом всегда $-1 \leq x(k) \leq 1$.

Сигнал $\{x(k)\}$ далее подается на входной слой сети, образованный элементами чистого запаздывания z^{-1} : $z^{-1}x(k) = x(k-1)$, в результате чего на выходе этого слоя формируется набор задержанных значений $x(k-1), x(k-2), \dots, x(k-d)$. При этом, чем больше значение d , тем более широкими возможностями обладает сеть.

Первый скрытый слой образован стандартными нейронами с нелинейной функцией активации, на входы которых подаются задержанные значения сигнала $x(k)$ и по цепи обратной связи каждого нейрона - задержанные значения прогноза $\hat{x}_j(k), j = 1, 2, \dots, d$. Входы нейронов, обозначенные \circ , соответствуют входам обучающего сигнала, а ψ описывает функцию активации.

В результате обработки последовательности $\{x(k)\}$ нейронами первого скрытого слоя на их выходах появляются одношаговые прогнозы

$$\begin{cases} \hat{x}_1(k) = \psi(x(k-1), \hat{x}_1(k-1)) \\ \hat{x}_2(k) = \psi(x(k-1), x(k-2), \hat{x}_2(k-1), \hat{x}_2(k-2)) \\ \vdots \\ \hat{x}_d(k) = \psi(x(k-1), \dots, x(k-d), \hat{x}_d(k-1), \dots, \hat{x}_d(k-d)), \end{cases} \quad (3)$$

соответствующие нелинейному процессу авторегрессии - скользящего среднего (NARMA) порядка от 1 до d .

Второй скрытый слой образован однотипными элементарными нейронами, вычисляющими оценки точности прогнозов $\hat{x}_j(k)$ в виде дисперсий обновляющих последовательностей $\sigma_{e_j}^2(k)$.

Нейронами третьего скрытого слоя производится попарное объединение сигналов из второго слоя с целью получения оценок $y_j(k), j = 1, 2, \dots, d-1$ таких, что

$$\begin{cases} y_j(k) = c_j(k)y_{j-1}(k) + (1-c_j(k))x_{j+1}(k), \\ \hat{y}_0(k) \equiv \hat{x}_1(k), \end{cases} \quad (4)$$

при этом весовые коэффициенты $c_j(k)$ характеризуют

оценку точности прогноза $y_j(k)$ по сравнению с $\hat{y}_{j-1}(k)$ и $x_{j+1}(k)$. Вектор весов $c(k) = (c_1(k), c_2(k), \dots, c_{d-1}(k))$ описывает качество прогнозов, получаемых в третьем слое, при этом оно, естественно, должно быть выше, чем в первом слое. Изменение соотношений между элементами этого вектора свидетельствует об изменении структуры модели сигнала, т.е. о возникшей разладке.

Четвертый слой полностью аналогичен второму и предназначен для оценки точности прогнозов $y_1(k), y_2(k), \dots, y_{d-1}(k)$, которые затем подаются на единственный нейрон пятого слоя, являющийся по структуре линейным ассоциатором и вычисляющий объединенный оптимальный прогноз $\hat{x}(k)$.

Нейрон, обозначенный D , предназначен для декодирования сигнала $\hat{x}(k)$ к исходной шкале

$$\hat{z}(k) = 0,5(z^{\max}(k)(\hat{x}(k) + 1) - z^{\min}(k)(\hat{x}(k) - 1)). \quad (5)$$

И, наконец, последний нейрон сети вычисляет точность окончательного прогноза.

2 АЛГОРИТМЫ ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ

Выходной сигнал j -го нейрона первого скрытого слоя может быть представлен в виде

$$\hat{x}_j(k) = \psi \left(\sum_{i=1}^j w_{ji}(k)x_i(k-1) + \sum_{i=1}^j \hat{w}_{ji}(k)\hat{x}_i(k-1) + w_{j0}(k) \right) = \psi(w_j^T(k)X_j(k)) = \psi(u_j(k)), \quad (6)$$

где $\psi(\cdot)$ - нелинейная функция активации, обычно сигмоидальная или гиперболического тангенса, а в нашем случае более простые в вычислительном отношении, но тем не менее удовлетворяющие всем необходимым условиям

$$\psi(u_i(k)) = \begin{cases} 1, 5u_i(k) - 0, 5u_i^3(k), \text{ или} \\ \sin \frac{\pi}{2}u_i(k), \text{ или} \\ \frac{2}{\pi} \arctan \gamma u_i(k), \end{cases} \quad (7)$$

$w_j(k) = (w_{j0}(k), w_{j1}(k), \dots, w_{jj}(k), \hat{w}_{j1}(k), \dots, \hat{w}_{jj}(k))^T - (2j+1) \times 1$ - вектор настраиваемых синаптических весов, $X_j(k) = (1, x(k-1), \dots, x(k-j), \hat{x}_j(k-1), \dots, \hat{x}_j(k-j))^T$ - вектор входов j -го нейрона, $u_j(k) = w_j^T(k)X_j(k)$, $j = 1,$

$2, \dots, d, k = 1, 2, 3, \dots$ - текущее дискретное время.

Вводя в рассмотрение ошибку прогнозирования j -го нейрона первого скрытого слоя

$$\epsilon_j(k) = x(k) - \hat{x}_j(k) = x(k) - \psi(u_j(k)), \quad (8)$$

запишем дельта-правило настройки синаптических весов

$$\begin{aligned} w_j(k+1) &= w_j(k) + \eta_j(k)\epsilon_j(k)\nabla_{w_j}\psi(u_j(k)) = \\ &= w_j(k) + \eta_j(k)\epsilon_j(k)G_j(k), \end{aligned} \quad (9)$$

где $\eta_j(k)$ - параметр шага поиска, принимаемый обычно постоянным, $\nabla_{w_j}\psi(u_j(k)) = G_j(k)$ - градиент функции активации по настраиваемым весам.

Сходимость алгоритма (9) обеспечивается в достаточно широком интервале варьирования параметра шага $\eta_j(k)$, однако естественным представляется выбор шага, обеспечивающего максимальную скорость. Можно показать [10], что в детерминированном случае максимальное быстроедействие обучения обеспечивается алгоритмом

$$w_j(k+1) = w_j(k) + \frac{x(k) - \hat{x}_j(k)}{\|G_j(k)\|^2} G_j(k) =$$

$$= \begin{cases} w_j k + \frac{xk - \hat{x}_j k}{1, 51 - u_j^2 k \|X_j k\|^2} X_j k, \text{ или} \\ w_j k + \frac{xk - \hat{x}_j k}{\frac{\pi}{2} \cos \frac{\pi}{2} u_j k \|X_j k\|^2} X_j k, \text{ или} \\ w_j k + \frac{xk - \hat{x}_j k}{\gamma} X_j k. \end{cases} \quad (10)$$

Несложно видеть, что в линейном случае (10) совпадает с алгоритмом обучения Уидроу-Хоффа.

В случае, если контролируемый сигнал наблюдается на фоне помех, алгоритм обучения должен обладать как следящими, так и фильтрующими свойствами, для чего может быть использована процедура [11]

$$\begin{cases} w_j(k+1) = w_j(k) + r_j^{-1}(k)(x(k) - \hat{x}_j(k))G_j(k), \\ r_j(k) = \alpha r_j(k-1) + \|G_j(k)\|^2, 0 \leq \alpha \leq 1, r_j(0) = 0, \end{cases} \quad (11)$$

близкая по свойствам при $\alpha = 1$ к алгоритму стохастической аппроксимации Гудвина-Рэмеджа-Кэйнеса.

После каждого такта обучения нейронов первого скрытого слоя несложно получить многошаговый прогноз в виде

$$x_j(k+l) = \psi(w_j^T(k+l)X_j(k+l)), \quad (12)$$

где

$$\left\{ \begin{array}{l} X_j(k+1) = (1, x(k), \dots, x(k-j+1), \hat{x}_j(k), \dots, \hat{x}_j(k-j+1))^T, \\ X_j(k+2) = (1, \hat{x}_j(k+1), x(k), \dots, x(k-j+2), \hat{x}_j(k+1), \dots, \\ \dots, \hat{x}_j(k-j+2))^T, \\ \vdots \\ X_j(k+j) = (1, \hat{x}_j(k+j-1), \dots, x(k), \hat{x}_j(k+j-1), \dots, \hat{x}_j(k))^T, \\ \vdots \\ X_j(k+l) = (1, \hat{x}_j(k+l-1), \dots, \hat{x}_j(k-l+1), \hat{x}_j(k+l-1), \dots, \\ \dots, \hat{x}_j(k-j+1))^T, \end{array} \right. \quad (13)$$

$$\hat{x}(k) = \sum_{j=1}^{d-1} c_j^0(k) \hat{y}_j(k) = \hat{y}^T(k) c^0(k), \quad (19)$$

где вектор весов $c^0(k) = (c_1^0(k), c_2^0(k), \dots, c_{d-1}^0(k))^T$ отве-

чает условию несмещенности, $\sum_{j=1}^{d-1} c_j^0(k) = E^T c^0(k) = 1$,

$y(k) = (y_1(k), y_2(k), \dots, y_{d-1}(k))^T$, $E = (1, 1, \dots, 1)^T$. Тре-

бование оптимальности прогноза $\hat{x}(k)$ приводит к алго-

ритму обучения вида [13]:

Во втором скрытом слое с помощью рекуррентных процедур

$$\sigma_{\varepsilon_j}^2(k) = \frac{k-1}{k} \sigma_{\varepsilon_j}^2(k-1) + \frac{1}{k} \varepsilon_j^2(k) \quad (14)$$

для стационарного и

$$\sigma_{\varepsilon_j}^2(k) = \alpha \sigma_{\varepsilon_j}^2(k-1) + (1-\alpha) \varepsilon_j^2(k), 0 \leq \alpha < 1 \quad (15)$$

для нестационарных сигналов вычисляется точность полученных прогнозов.

Попарное объединение прогнозов в третьем скрытом слое в виде (4) призвано обеспечить более высокое качество прогнозирования, при этом обучение нейронов этого слоя, являющихся двухвходовыми линейными ассоциаторами, осуществляется с помощью процедуры [12]

$$\left\{ \begin{array}{l} c_j(k+1) = \frac{s_j(k)}{s_j(k+1)} c_j(k) + \frac{\varepsilon_j(k+1)(\hat{y}_{j-1}(k+1) - \hat{x}_{j+1}(k+1))}{s_j(k+1)}, \\ s_j(k+1) = s_j(k) + (\hat{y}_{j-1}(k+1) - \hat{x}_{j+1}(k+1))^2, \\ c_j(0) = 0, 5. \end{array} \right. \quad (16)$$

В четвертом скрытом слое на основе ошибки $v_j(k) = x(k) - y_j(k)$ и формул типа (14), (15) вычисляются дисперсии $\sigma_{v_j}^2(k)$, при этом можно показать [12], что

$$\sigma_{v_j}^2(k) \leq \min\{\sigma_{v_{j-1}}^2(k), \sigma_{\varepsilon_{j+1}}^2(k)\}, \quad (17)$$

т.е. прогнозы, получаемые в третьем скрытом слое по точности не уступают лучшим прогнозам первого скрытого слоя.

Многошаговый прогноз аналогично (12) может быть записан в виде

$$y_j(k+1) = c_j(k+1) y_{j-1}(k+l) + (1-c_j(k+1)) x_j(k+l). \quad (18)$$

В пятом скрытом слое сети, образованном $(d-1)$ -входовой адальиной, производится объединение прогнозов в виде

$$\left\{ \begin{array}{l} P(k+1) = P(k) - \\ - (P(k) \hat{y}(k+1))(P(k) \hat{y}(k+1))^T \times \\ \times (1 + \hat{y}^T(k+1) P(k) \hat{y}(k+1))^{-1}, \\ c^*(k+1) = c^*(k) + \\ + P(k+1)(x(k+1) - \hat{y}^T(k+1) c^*(k)) \hat{y}(k+1), \\ c^0(k+1) = c^0(k) - \\ - P(k+1)(E^T P(k+1) E)^{-1} (E^T c^*(k+1) - 1) E, \\ c_j^* = (d-1)^{-1}, \end{array} \right. \quad (20)$$

где $c^*(k+1)$ - оценка стандартного рекуррентного метода наименьших квадратов.

Многошаговый прогноз этого слоя имеет вид

$$\hat{x}(k+1) = \hat{y}^T(k+l) c^0(k+1) \quad (21)$$

и после декодирования нейроном D может быть преобразован в прогноз исходной последовательности $z(k+l)$.

На выходе сети производится оценка точности объединенного прогноза в виде

$$\sigma_z^2(k) = \frac{k-1}{k} \sigma_z^2(k-1) + \frac{1}{k} (z(k) - \hat{z}(k))^2 \quad (22)$$

или

$$\sigma_z^2(k) = \alpha \sigma_z^2(k-1) + (1-\alpha) (z(k) - \hat{z}(k))^2, 0 \leq \alpha < 1, \quad (23)$$

при этом можно показать [13], что

$$\hat{\sigma}_z^2(k) \leq \min\{\min \sigma_{\varepsilon_j}^2(k), \min \sigma_{v_j}^2(k)\}, \forall j, \quad (24)$$

т.е. прогноз на выходе сети является действительно оптимальным на классе всех NARMA-моделей порядка от 1 до d .

ЗАКЛЮЧЕНИЕ

Предлагаемая искусственная нейронная сеть является специализированной вычислительной структурой, предназначенной для решения задач прогнозирования и раннего обнаружения изменения свойств стохастических по-

следовательностей, описываемых различными NARMA-уравнениями с неизвестными параметрами. В результате обучения сеть обеспечивает оптимальное на классе данных моделей прогнозирование, при этом, если вместо нейронов первого скрытого слоя использовать стандартные структуры прогнозирующих сетей типа MLP, RBFN, SOM и т.п., можно говорить об оптимальном на ансамбле нейросетей [14] прогнозе. С вычислительной точки зрения подобная метанейросеть не намного сложнее рассмотренной выше, поскольку программы, реализующие стандартные архитектуры, входят в состав многих коммерческих пакетов прикладных программ [15].

ПЕРЕЧЕНЬ ССЫЛОК

1. Wong F.S. Time series forecasting using backpropagation neural networks // Neurocomputing. - 1990/91. - 2. - P. 147-159.
2. Groot de C., Wuertz D. Analysis of univariate time series with connectionist nets: A case study of two classical examples // Neurocomputing. - 1991. - 3. - P. 177-192.
3. Connor J.T., Martin R.D., Atlas L.E. Recurrent neural networks and robust time series prediction // IEEE Trans. Neural Networks. - 1994. - 5. - N2. - P. 240-254.
4. Saxen H. Nonlinear time series analysis by neural networks. A case study // Int. J. Neural Systems. - 1996. - 7. - N2. - P.

5. 195-201.
5. Madhavan P.G. A new recurrent neural network learning algorithm for time series prediction // J. of Intelligent Systems. - 1997. - 7. - N1, 2. - P. 103-116.
6. Yu H.-Y., Bang S.-Ya. An improved time series prediction by applying the layer-by-layer learning method to FIR neural networks // Neural Networks. - 1997. - 10. - N9. - P. 1717-1729.
7. Nie J. Nonlinear time-series forecasting: A fuzzy-neural approach // Neurocomputing. - 1997. - 16. - P. 63-76.
8. Billings S.A., Hong X. Dual-orthogonal radial basis function networks for nonlinear time series prediction // Neural Networks. - 1998. - 11. - P. 479-493.
9. Conway A.J., Macpherson K.P., Brown J.C. Delayed time series prediction with neural networks // Neurocomputing. - 1998. - 18. - P. 81-89.
10. Бодянский Е.В. Адаптивные алгоритмы идентификации нелинейных объектов управления // АСУ и приборы автоматики. - 1987. - Вып. 81. - С. 43-46.
11. Бодянский Е.В. Обнаружение разладок в нелинейных стохастических последовательностях с помощью рекуррентных искусственных нейронных сетей // Проблемы бионики. - 1998. - Вып. 49. - С. 23-33.
12. Bodyanskiy Ye.V., Vorobyov S.A., Stephan A. Detection of NARMA-sequence order using recurrent artificial neural networks // Proc. of European Control Conference ECC'99. - Karlsruhe, Germany, 1999. - CD-ROM.
13. Бодянский Е.В. Автоматическое обнаружение разладок с помощью искусственной нейронной метасети // Проблемы бионики. - 1998. - Вып. 49. - С. 34-38.
14. Sharkey A.J.C. On combining artificial neural nets // Connection Science. - 1996. - 8. - N3, 4. - P. 299-313.
15. Braun H., Feulner J., Malaka R. Praktikum Neuronale Netze. Berlin: Springer-Verlag, 1996. - 242 P.

Надійшла 06.03.2000

Після доробки 10.03.2000

УДК 621.3.049.77.002:519.24

МЕТОД ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ГРАНИЧНОГО ВЫБОРОЧНОГО КОНТРОЛЯ

А. Ю. Долгов

Предлагается метод повышения эффективности прогнозирования (оценок) величины брака в принятых пластинах при граничном выборочном контроле качества кристаллов интегрированных микросхем.

The method of increase of a forecasting effectiveness (ratings) of a spoilage size in the accepted plates at boundary selective quality surveillance of integrated microcircuits crystals is offered.

Граничный контроль [1] является основным методом контроля качества при производстве кристаллов интегральных микросхем. Основой метода является выборка малого объема (обычно $n=5$, иногда $n=10$), по которой следует судить о качестве кристаллов на пластине, содержащей от 400 до 4000 потенциально годных изделий. Такой малый объем выборки связан с количеством тестовых ячеек на пластине, которые имеют другую структуру, чем рабочие ячейки, и количество которых не может быть увеличено без ущерба для выхода годных. Но этот малый объем контрольной выборки исключает применение классических методов статистического контроля по количественному признаку [2], хотя сам по себе является достаточно эффективным. Так при реша-

ющем правиле "5 из 5" (т.е. при попадании всех пяти измеренных величин в норму по конструкторской документации) точность прогноза выхода годных, т.е. величина возможного брака на принятых пластинах, колеблется от 0 до 69 %, а при решающем правиле "3 из 5" - от 15 до 85% [1]. Такие точности не могут удовлетворить производство, однако метод остается по-прежнему востребованным, так как не имеется другого столь же простого в применении на практике, а для классических методов контроля, как уже упоминалось, не хватает объема выборки.

Классический метод прогнозирования состоит в том, что определяются квантили граничных значений по норме конструкторской документации относительно среднего арифметического контрольной выборки, нормированные среднеквадратическим отклонением этой же выборки. По квантилям с помощью закона распределения Стьюдента определяются вероятности, лежащие вне границ нормы, которые затем складываются в прогнозируемый брак. Анализ этого метода прогнозирования брака показал, что слишком большой разброс между минимальным и максимальным значениями брака при одних

Таблица 1 - Величина прогнозируемого брака (%) на пластине с пятью тестовыми ячейками при СКО повышенной эффективности

Прог. брак	Расчет СКО	Число измерений, не выходящих за пределы нормы, m										
		$m = 5$			$m = 4$			$m = 3$				
		$v=1,0$	$v=0,7$	$v=0,5$	$v=1,3$	$v=1,0$	$v=0,7$	$v=0,5$	$v=2,0$	$v=1,0$	$v=0,7$	$v=0,5$
мин.	(1)	5,3	1,1	0,1	12,0	5,5	2,9	2,8	31,2	16,0	15,7	15,6
	(2)	15,4	5,2	1,1	23,8	15,7	8,7	8,0	47,0	25,8	23,8	29,6
	(3)	24,2	10,0	2,8	31,9	24,5	14,4	12,5	55,0	32,2	28,4	27,8
	(4)	32,2	15,4	6,4	39,1	32,5	19,8	16,4	61,8	37,9	32,2	30,9
	(5)	67,7	55,2	41,0	71,6	67,8	56,5	44,6	83,6	68,4	58,9	49,7
Наивероятнейший	(1)	-	2,4	2,1	12,9	7,9	7,0	6,9	37,4	30,6	30,5	30,5
	(2)	-	6,0	3,8	27,3	18,1	14,6	14,2	50,1	36,9	36,0	35,1
	(3)	-	11,1	6,9	36,2	26,2	19,8	18,7	57,0	40,8	38,6	38,3
	(4)	-	17,0	10,7	43,8	33,2	25,0	22,9	62,3	44,4	40,8	40,0
	(5)	-	55,5	42,9	73,2	67,9	58,2	48,6	83,8	69,2	60,2	52,6
Макс	(1)	-	2,8	2,7	17,9	15,9	15,6	15,5	52,7	50,0	50,0	50,0
	(2)	-	8,5	7,8	30,7	25,6	23,7	23,5	58,1	50,5	50,0	50,0
	(3)	-	14,1	12,3	38,9	31,9	28,2	27,6	61,9	51,4	50,2	50,0
	(4)	-	19,6	16,2	45,4	37,5	31,9	30,7	66,9	52,9	50,5	50,0
	(5)	-	56,2	44,1	75,2	68,1	58,5	49,1	84,3	70,6	62,2	55,5

и тех же условиях его выявления сложился из двух составляющих: объективно (реально) существующего брака (может быть выявлен при стопроцентном контроле) и субъективной неточностью (ошибкой, разбросом) методов расчета, главный вклад в которую вносит ошибка определения среднеквадратичного отклонения.

Согласно [3] ошибка в определении среднеквадратического отклонения при объеме выборки $n=5$ может составить свыше 34 %, что приводит к дополнительной ошибке в определении прогнозируемого брака только за счет неточностей расчета по формулам до 40 % от полученной цифры прогноза. Для уменьшения этой ошибки предлагается воспользоваться идеей определения параметров выборки малого объема более эффективной, чем классические методы расчета [4]. Она была доработана в [5] и превратилась в алгоритм расчета, при котором относительная ошибка определения среднеквадратического определения уменьшилась до 18 %, а ошибка вычисления прогнозируемого брака соответственно до 21 %. Результаты расчетов по предложенной схеме представлены в таблице 1.

В этой таблице графа "Прогнозируемый брак" предусматривает три случая расположения чисел выборки относительно границ нормы: когда крайнее число находится вблизи границы с внутренней стороны нормы, когда оно находится с внешней стороны нормы и промежуточное (наивероятнейшее) расположение. Графа "Расчет СКО" показывает пять вариантов расчета величины среднеквадратического отклонения по различным фор-

мулам (минимальное, максимальное и наивероятнейшее значения СКО, вычисленные через размах выборки, СКО, вычисленное по формуле Лежандра, и СКО, вычисленное по методике [5]).

Анализ таблицы показывает, что прогнозируемый брак в случае решающего правила "5 из 5" лежит в пределах от 0 до 56 %, а в случае решающего правила "3 из 5" от 15 до 85 %. Сравнение с результатами расчетов по классическим методам позволяет сделать вывод, что уменьшение величины прогнозируемого брака для решающего правила "5 из 5" произошло за счет повышения эффективности расчетных формул, в то время как доля реального брака осталась неизменной. Такое же сравнение для решающего правила "3 из 5" показывает совпадение результатов, что означает пренебрежимо малую долю ошибки за счет неточности расчетных формул и основную долю в виде реально существующего брака.

Интересно также отметить существенную разницу в величине прогнозируемого брака при расчетах по разным формулам СКО (соотношение максимального и минимального брака достигает 50 раз). Это свидетельствует о минимаксных границах прогнозирования, которые по теории вероятности достигаются лишь в трех-пяти случаях из тысячи. Из практических соображений наиболее удобно пользоваться вероятностными формулами расчета СКО, из которых наибольший интерес представляет собой формула (3), взятая из работы [5]. На наш взгляд, она приводит к

результатам прогноза, наиболее близко совпадающих с реальным браком. Считаем целесообразным рекомендовать для практических целей пользоваться результатами расчетов именно по этой формуле.

ПЕРЕЧЕНЬ ССЫЛОК

1. Долгов Ю.А. Анализ граничного метода контроля качества кристаллов ИМС при выборке малого объема // Технология и конструирование в электронной аппаратуре. - 1992. -

- Вып. 4. - С. 26-30.
2. Статистический приемочный контроль по количественному признаку. Планы контроля: ГОСТ 20736-75 (СТ СЭФ 1672-79). - М.: Изд-во стандартов, 1982. - 120 с.
3. Шор Я.Б. Статистические методы анализа и контроля качества и надежности. - М.: Сов. радио, 1962. - 552 с.
4. Гаскаров Д.В., Шаповалов В.И. Малая выборка. -М.: Статистика, 1978. - 248 с.
5. Долгов Ю.А. Статистический контроль качества продукции при выборках малого объема // Технология и конструирование в электронной аппаратуре. - 1993. - Вып.2. - С.17-21. Надійшла 12.04.99

Після доробки 17.05.99

УДК 007:159.955:519.72

АНАЛИЗ ПАРАМЕТРИЧЕСКОЙ ПЛАСТИЧНОСТИ МНОГОСЛОЙНЫХ НЕЙРОННЫХ СЕТЕЙ

А. Ю. Дорогов

В статье рассматриваются многослойные нейронные сети прямого распространения. Пластичность нейронной сети оценивается числом степеней свободы нелинейного оператора. Вводится понятие структурной модели сети, как ориентированного графа, для которого определены весовые функции вершин и дуг. На основе структурной модели получены аналитические выражения для расчета числа степеней свободы. Приведены примеры расчета степени пластичности для двухслойных нейронных сетей.

Multilayer feedforward neural nets are researched on the paper. Plasticity of neural net is estimated by freedom degrees of nonlinear operator. Structure models of neural net are suggested. Model is represented by oriented graph with weighted arcs and nodes. Analytic expressions for calculation of freedom degrees are resulted on base of structure model. Examples of two-layers neural nets are given.

ВВЕДЕНИЕ

Многослойные сети прямого распространения широко используются в задачах распознавания образов, аппроксимации функций, системах адаптивного управления [1,2]. Для сетей подобного вида предложен ряд эффективных алгоритмов обучения, среди которых наибольшую известность получил алгоритм Error Back propagation [3], основанный на рекуррентном использовании градиентного метода поиска экстремума. При практическом использовании нейронных сетей, неизбежно, возникает вопрос выбора структуры сети. Для многослойных сетей прямого распространения необходимо определить количество слоев в сети и количество нейронов в каждом слое. Существует ряд рекомендаций по выбору структуры, основанных на эмпирическом исследовании сетей. Но любое экспериментальное исследование неразрывно связано с конкретной задачей, поэтому трудно ожидать, что рекомендации будут пригодными как для задач распознавания, так и, например, для задач аппроксимации функций. По-види-

мому, для каждой прикладной области существует свой набор критериев, которые следует использовать при выборе структуры. Тем не менее, можно выделить критерии, общие для различных приложений. Одним из таких критериев может служить способность нейронной сети к обучению. Интуитивно можно предположить, что чем больше "знаний" способна поглотить сеть, тем лучшими свойствами она будет обладать при использовании в конкретной задаче. Способность к обучению (в литературе часто используется удачный термин "пластичность") целесообразно оценивать числом независимых настроек, существующих в сети. Это значение, как правило, меньше чем полное количество синаптических весов, подвергающихся изменению в процессе обучения сети. (Исключением является однослойный перцептрон, для которого соблюдается равенство.) В механике для оценки числа независимых координат используется понятие "число степеней свободы". Близкую аналогию можно провести и для нейронных сетей. В самом деле, нейронную сеть можно представить как нелинейный оператор, осуществляющий преобразование входного вектора в выходной. Полное множество операторов образует многомерное пространство, в котором каждый оператор можно рассматривать как некоторую материальную точку. Изменение синаптических весов нейронной сети приводит к перемещению точки-оператора в пространстве операторов. Следуя далее механической аналогии, будем называть число независимых координат, необходимое и достаточное для однозначного определения местоположения точки-оператора в пространстве операторов, числом степеней свободы нейронной сети. Класс операторов, порождаемых изменением синаптических весов, образует некоторую поверхность (многообразие) в пространстве операторов. Число степеней свободы, по существу, определяется размерностью минимального про-

странства, в которое можно погрузить некоторую малую окрестность любой точки многообразия операторов.

На этапе обучения нейронная сеть обычно линеаризуется в окрестности некоторых точек нелинейных функций активации нейронов. В пределах этой окрестности все операторы можно рассматривать как линейные. Поэтому задача вычисления степеней свободы состоит в том, чтобы определить размерность линейной оболочки класса операторов, порождаемого нейронной сетью. Решению поставленной задачи и будет посвящена данная работа.

1 СТРУКТУРНАЯ МОДЕЛЬ НЕЙРОННОЙ СЕТИ

На рис.1 показана многослойная нейронная сеть прямого распространения. В полностью связанной сети каждый из нейронов f_i связан со всеми нейронами предшествующего слоя. Размерность входного вектора для данной сети равна N , а выходного M . Поставим в соответствие каждому i -му нейронному слою пару векторных градуированных [4] пространств (E^i, D^i) с размерностями (p^i, g^i) . Указанные пространства будем называть терминальными пространствами нейронного слоя. Условие градуировки задается фиксированным разложением каждого пространства в прямую сумму одномерных подпространств, которые назовем базовыми направлениями.

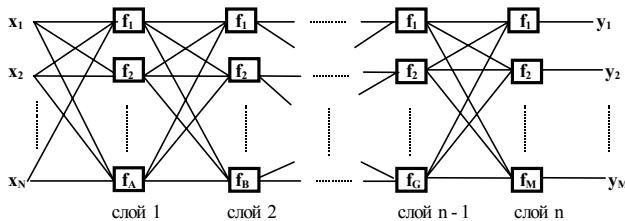


Рисунок 1 - Многослойная полностью связанная нейронная сеть

Каждый нейронный слой можно описать нелинейным оператором A_i . Изменение синаптических весов в слое i порождает класс операторов A_i , осуществляющих отображение пространств так, что $(E^i)A_i \rightarrow D^i$. Как уже было отмечено, для задачи оценки пластичности можно считать, что класс A_i состоит из линейных операторов. Связь между смежными слоями определим операторами проектирования $P_{i,i+1}$ так, что $(D^i)P_{i,i+1}c \rightarrow E^{i+1}$. Операторы $P_{i,i+1}$ будем считать фиксированными операторами, осуществляющими не более чем перестановку

базисных векторов в смежных терминальных пространствах, размерности которых, очевидно, совпадают. Базисные вектора всегда будем выбирать так, чтобы они принадлежали базовым направлениям. С учетом введенных определений класс операторов, соответствующий всей нейронной сети будет определяться произведением:

$$W = A_1 P_{12} A_2 P_{23} \dots A_{n-1} P_{n-1,n} A_n. \quad (1)$$

Выражению (1) соответствует линейный граф, показанный на рис.2.

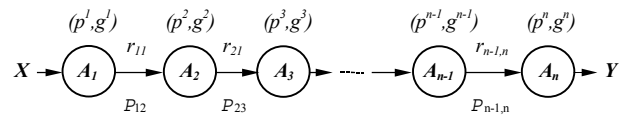


Рисунок 2 - Структурная модель многослойной нейронной сети

Каждой вершине графа отвечает класс операторов одного нейронного слоя, а дуги соответствуют операторам проектирования $P_{i,i+1}$. На данном графе определены функции весов вершин и дуг. Весом i -ой вершины будем считать пару чисел (p^i, g^i) , а весом дуги ранг оператора связи $r_{i,i+1} = \text{rank} P_{i,i+1}$. Максимальный ранг операторов в классе A_i назовем рангом класса и обозначим r_i . Очевидно,

$$r_i = \max_{A \in A_i} (\text{rank} A) = \min(p^i, g^i) = p^i \circ g^i.$$

Вычисление минимума в многозначной логике [5] рассматривается как операция логического умножения, в последнем выражении и далее для обозначения этой операции используется символ " \circ ".

Поскольку $g^i = p^{i+1}$, то $r_{i,i+1} = g^i = p^{i+1}$, и нетрудно показать, что в этом случае

$$r_{i,i+1} \geq \min(r_i, r_{i+1}) = r_i \circ r_{i+1}. \quad (2)$$

Взвешенный граф, приведенный на рис.2, отражает структурные свойства нейронной сети. В дальнейшем такой граф будем называть структурной моделью нейронной сети.

2 РАЗМЕРНОСТЬ ПРОСТРАНСТВА ОПЕРАТОРОВ

Будем считать класс операторов A_i полным, т.е. образующим пространство операторов. Последнее означает, что полный класс операторов изоморфен тензор-

ному произведению [6] пространств (E^i, D^i) и, следовательно, его размерность равна

$$\dim A_i = \dim(E^i \times D^i) = p^i g^i.$$

Обозначим через W_i подкласс операторов класса W (см. выражение (1)), который порождается классом операторов A_i , при фиксированных в общем положении операторов для остальных классов A_j $j \neq i$. Легко проверить, что класс W_i будет линейным подпространством. Условие "общее положение" [4] означает, что фиксация операторов в классах A_j осуществляется таким образом, чтобы обеспечить наибольшую размерность подпространства W_i . Для класса операторов W линейной оболочкой будет объединение операторных подпространств W_i , т.е.:

$$L(W) = W_1 \cup W_2 \cup W_3 \cup \dots \cup W_{n-1} \cup W_n. \quad (3)$$

Размерность этого линейного пространства по определению будет равна числу степеней свободы. Обозначим эту величину через S , полагая

$$S = \dim(L(W)) = \dim W.$$

Как известно [4], размерность объединения любой пары пространств B' и B'' определяется следующим выражением:

$$\dim(B' \cup B'') = \dim(B') + \dim(B'') - \dim(B' \cap B'').$$

В Приложении показано, что градуированные подпространства образуют дистрибутивную решетку по операциям " \cup " и " \cap ". В этом случае допустимо к выражению (3) многократно применить последнюю формулу, в результате получим:

$$\begin{aligned} \dim W &= \dim W_1 + \dim W_2 + \dim W_3 + \dots + \dim W_{n-1} + \dim W_n - \\ &- \dim W_1 \cap (W_2 \cup W_3 \cup \dots \cup W_{n-1} \cup W_n) - \\ &- \dim W_2 \cap (W_3 \cup W_4 \cup \dots \cup W_{n-1} \cup W_n) - \\ &- \dots - \dim W_{n-2} \cap (W_{n-1} \cup W_n) - \dim(W_{n-1} \cup W_n). \end{aligned}$$

В этом выражении обозначим через $S^{(+)}$ все положительные компоненты, а через $S^{(-)}$ все отрицательные, тогда можно записать:

$$\dim W = S^{(+)} - S^{(-)}.$$

Рассмотрим компоненту $S^{(+)}$. Обозначим через r_i^s ранг произведения последовательности операторов, предшествующих классу A_i в выражении (1), а через r_i^c ранг произведения последовательности операторов следующих за этим классом. Поскольку ранг произведения операторов не превосходит минимального ранга сомножителей, то учитывая условие (2) и следуя принципу "общего положения", можно записать:

$$r_i^s = r_1 \circ r_2 \circ \dots \circ r_{i-2} \circ r_{i-1}, \text{ где } r_1^s = N, \quad (4)$$

$$r_i^c = r_{i+1} \circ r_{i+2} \circ \dots \circ r_{n-1} \circ r_n, \text{ где } r_n^c = M. \quad (5)$$

Величину r_i^s будем называть входным рангом для класса A_i , а величину r_i^c выходным рангом. Входной и выходной ранги, по существу, определяют "степень участия" класса A_i в формировании операторного подпространства W_i . На рис.3. показана мнемоническая схема порождения операторного подпространства W_i .

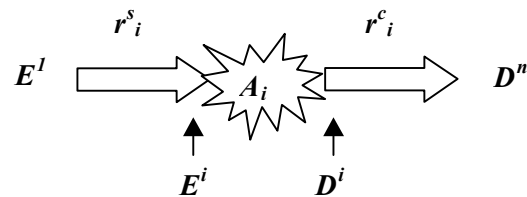


Рисунок 3 - Схема порождения операторного подпространства W_i

По теореме о структуре линейного отображения [4], для цепочки операторов, предшествующих A_i , существуют такие прямые разложения пространств E^1 и E^i

$$E^1 = E_0^1 \oplus E_1^1 \rightarrow E_1^i \oplus E_2^i = E^i,$$

что E_1^1 изоморфно E_1^i и имеет размерность r_i^s , а E_0^1 составляет нулевое пространство (ядро линейного отображения). Аналогично, для цепочки операторов, следующих за A_i существуют прямые разложения

$$D^i = D_0^i \oplus D_1^i \rightarrow D_1^n \oplus D_2^n = D^n$$

такие, что D_1^n изоморфно D_1^i и имеет размерность r_i^c , а

D_0^i составляет нулевое пространство.

Нулевое пространство при отображении переходит в нулевой вектор, поэтому операторное подпространство W_i изоморфно тензорному произведению $E^i \times D_1^i$. Обозначим изоморфное отображение символом " \cong ", тогда можно записать

$$W_i \cong E^i \times D_1^i \cong E_1^i \times D_1^n, \quad (6)$$

откуда следует, что

$$\dim W_i = r_i^s r_i^c \text{ и, следовательно } S^{(+)} = \sum_{i=1}^n r_i^s r_i^c.$$

На рис.4 приведена графическая интерпретация выражения (6). Внешний прямоугольник соответствует максимально-возможному операторному пространству нейронной сети при заданных размерностях входного и выходного векторов, а внутренний - отвечает операторному подпространству W_i .

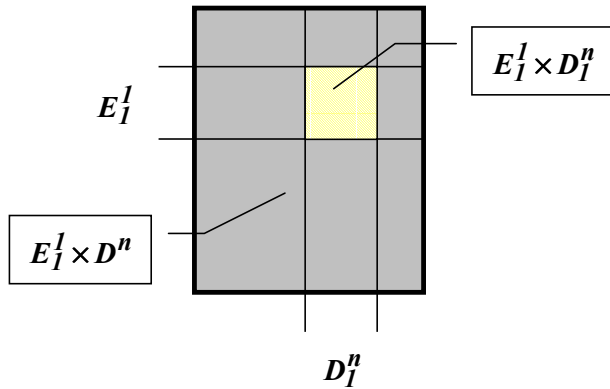


Рисунок 4 - Графическая интерпретация операторного подпространства W_i

Рассмотрим теперь компоненту $S^{(-)}$. Общий член для $S^{(-)}$ имеет вид:

$$\dim W_i \cap (W_{i+1} \cup W_{i+2} \cup \dots \cup W_{n-1} \cup W_n). \quad (7)$$

Для цепочки операторов, предшествующих классу A_{i+1} существуют прямые разложения пространств

$$E^1 = \tilde{E}_0^1 \oplus \tilde{E}_1^1 \rightarrow E_1^{i+1} \oplus E_2^{i+1} = E^{i+1}.$$

Подпространства \tilde{E}_1^1 , E_1^{i+1} изоморфны и имеют

размерность r_{i+1}^s , а \tilde{E}_0^1 составляет нулевое пространство. Класс операторов, образованных объединением $W_{i+1} \cup W_{i+2} \cup \dots \cup W_{n-1} \cup W_n$, осуществляет отображение пространства E^1 в D_n , при этом пространство \tilde{E}_0^1 переходит в нулевой вектор. Таким образом, существует изоморфизм:

$$W_{i+1} \cup W_{i+2} \cup \dots \cup W_{n-1} \cup W_n \cong \tilde{E}_1^1 \times D^n. \quad (8)$$

Искомое операторное многообразие образуется пересечением объединения (8) с операторным подпространством W_i . На рис.5 показана мнемоническая схема порождения операторного подпространства.

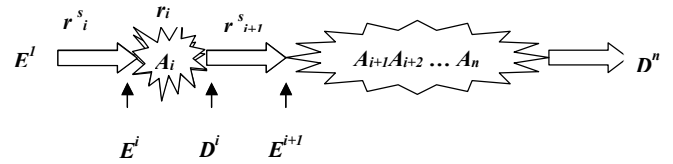


Рисунок 5 - Схема порождения операторного многообразия

Поскольку $r_{i+1}^s = r_i^s \circ r_i \leq r_i^s$, то $\tilde{E}_1^1 \subseteq E_1^1$. Подставляя (6) и (8) в (7) и учитывая, что $D_1^n \subseteq D^n$ получим:

$$\begin{aligned} \dim W_i \cap (W_{i+1} \cup W_{i+2} \cup \dots \cup W_{n-1} \cup W_n) &= \\ &= \dim(E_1^1 \times D^n) \cap (\tilde{E}_1^1 \times D^n) = \\ &= \dim(E_1^1 \cap \tilde{E}_1^1) \times (D_1^n \cap D^n) = \dim(\tilde{E}_1^1 \times D_1^n) = r_{i+1}^s r_i^c. \end{aligned}$$

На рис.6 приведена геометрическая интерпретация последнего выражения.

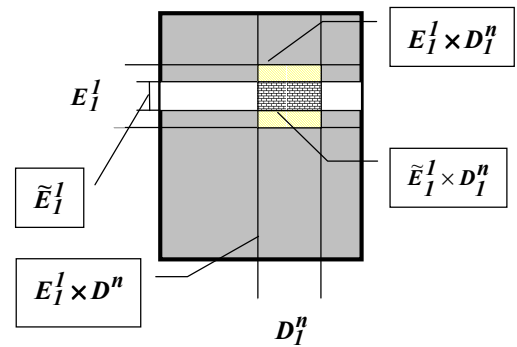


Рисунок 6 - Графическая интерпретация образования операторного подпространства

Подводя итог получим $S^{(-)} = \sum_{i=1}^{n-1} r_{i+1}^s r_i^c$, и оконча-

тельно будем иметь

$$S = S^{(+)} - S^{(-)} = \sum_{i=1}^n r_i^s r_i^c - \sum_{i=1}^{n-1} r_{i+1}^s r_i^c. \quad (9)$$

3 ПРИМЕРЫ

Рассмотрим двухслойную нейронную сеть с параметрами

$$n = 2, p^1 = N, g^1 = p^2 = D, g^2 = M.$$

На основании (9) можно записать $S = r_1^s r_1^c + r_2^s r_2^c - r_2^s r_1^c$.

Из выражений (4) и (5) получим

$$r_1^s = N, r_2^s = r_1 = N \circ g^1 = N \circ D,$$

$$r_1^c = r_2 = p^2 \circ M = D \circ M, r_2^c = M.$$

В результате будем иметь следующую расчетную формулу

$$S = N(D \circ M) + (N \circ D)M - (N \circ D)(D \circ M). \quad (10)$$

Рассмотрим различные варианты.

1) Пусть $D < \min(N, M)$, тогда

$$S = ND + DM - D^2. \quad (11)$$

Из этой формулы, очевидно, что наилучшей пластичностью обладает сеть со структурой типа "двухсторонний узел" (см. рис.7). Для этого типа сети значение $D = 1$ и, следовательно, $S = N + M$.

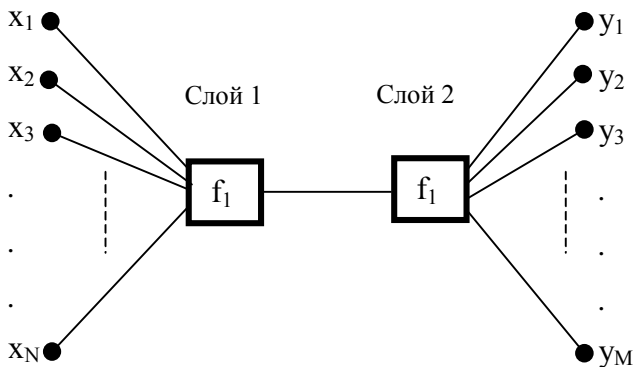


Рисунок 7 - Нейронная сеть "двухсторонний узел"

2) Пусть $D \geq \max(N, M)$, тогда из формулы (10) следует

$$S = NM + NM - NM = NM.$$

Это значение является максимально возможным для класса операторов, действующих из пространства размерности N в пространство размерности M , поэтому с точки зрения пластичности сети нет смысла увеличивать число нейронов в первом слое больше, чем $\max(N, M)$. Нетрудно показать, что этот вывод справедлив для всех скрытых слоев в многослойных сетях. Отметим, что значение NM совпадает с числом степеней свободы однослойного персептрона.

3) Структуры типа "односторонний узел" (см. рис.8)) часто используются для реализации аппроксиматоров функций.

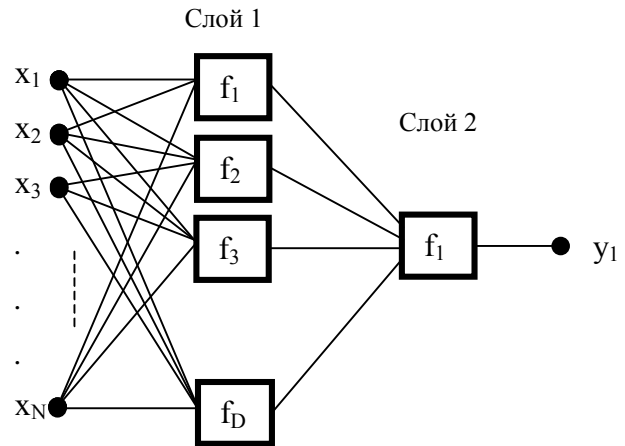


Рисунок 8 - Нейронная сеть "односторонний узел"

Рассмотрим число степеней свободы для такой сети. Из формулы (10) при условии $M = 1$, следует

$$S = N + (N \circ D) - (N \circ D) = N.$$

т.е. пластичность данной сети не зависит от числа нейронов в первом слое, вполне достаточно наличие только одного нейрона. Но с другой стороны для задач аппроксимации нелинейных функций необходимо обеспечить достаточный "запас" по элементарным нелинейностям [7], реализуемым функциями активации нейронов. Это еще раз подтверждает предположение, что для каждой задачи существуют собственные критерии качества нейронной сети, определяющие ее структуру, и в целом задача синтеза структуры является многокритериальной.

ЗАКЛЮЧЕНИЕ

Пластичность нейронных сетей характеризует дифференциальную способность нейронной сети к обучению. Можно сказать, что качество одной нейронной сети лучше другой, если первая при прочих равных условиях обладает лучшей пластичностью. В данной работе получены формулы расчета степени пластичности многослойных сетей прямого распространения. Приведенные выражения позволяют определить точное значение степени пластичности на основе структурных характеристик нейронной сети. Доказано, что пластичность многослойных сетей не может быть выше, чем пластичность однослойного персептрона. Этот вывод не отрицает целесообразности использования многослойных сетей, а напротив подтверждает необходимость многокритериального подхода к задаче структурного синтеза. В работах [9,10,11] приведены дополнительные результаты по оценке пластичности нейронных сетей прямого распространения с более сложной структурной организацией.

Данная работа поддержана грантом Минобразования РФ.

ПРИЛОЖЕНИЕ

ГРАДУИРОВАННЫЕ МОДУЛЯРНЫЕ РЕШЕТКИ

Пусть E конечномерное пространство размерности N . Как известно [8], совокупность всех подпространств E пространства образует модулярную решетку U по включению с операциями объединения и пересечения. Если A, B, C - подпространства пространства E , тогда для модулярной решетки имеет место

$$A \cap ((A \cup B) \cap C) = (A \cap B) \cup (A \cap C).$$

Откуда следует $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Условие градуировки в терминологии теории решеток означает, что фиксируется множество независимых элементов $e_i \in E$, образующих прямое разложение пространства E .

$$E = e_1 \oplus e_2 \oplus \dots \oplus e_N. \tag{П1}$$

Неразложимые далее независимые элементы называются атомами. В решетке подпространств атомы, по существу, предоставляют собой одномерные подпространства пространства E . Среди всех подпространств выделим подмножество подпространств которые имеют атомные разложения в системе атомов $\{e_1, e_2, \dots, e_N\}$. Такие подпространства будем называть градуированными с условиями градуировки (П1). Множество градуированных подпространств замкнуто относительно объединений и пересечений и образуют подрешетку W моду-

лярной решетки U . Покажем, что эта решетка дистрибутивна. Пусть A, B, C градуированные подпространства в системе атомов $\{e_1, e_2, \dots, e_N\}$, тогда

$$A = \sum_{i \in I} e_i, B = \sum_{j \in J} e_j, C = \sum_{k \in K} e_k,$$

где I, J, K - множества индексов. Поскольку элементы e_i независимы то $e_i \cap e_j = 0$ если $i \neq j$ и напротив $e_i \cap e_i = e_i$, кроме того $e_i \cup e_i = e_i$. Учитывая эти условия, получим:

$$A \cap B = \sum_{i \in I \cap J} e_i, A \cap C = \sum_{i \in I \cap K} e_i, B \cap C = \sum_{i \in J \cap K} e_i,$$

откуда следует:

$$\begin{aligned} A \cap (B \cup C) &= \sum_{i \in I \cap (J \cup K)} e_i = \sum_{i \in I \cap J} e_i \cup \sum_{i \in I \cap K} e_i = \\ &= (A \cap B) \cup (A \cap C). \end{aligned}$$

Дистрибутивность решетки W следует из дистрибутивности операций " \cup ", " \cap " над числовыми множествами I, J, K .

ПЕРЕЧЕНЬ ССЫЛОК

1. Mills Peter M., Zomaya Albert Y, Tade Moses O. Neuro-Adaptive Process Control. Practical Approach. - England: JOHN WILEY&SONS,- 1995. - 212p.
2. Cichocki A., Unbernhauen R. Neural Networks for optimization and signal processing. - Stuttgart: - JOHN WILEY&SONS, - 1994. - 526p.
3. Уоссерман Ф. Нейрокомпьютерная техника. Теория и практика. М.: Мир, 1992. - 240с.
4. Кострикин А.И., Манин Ю.М. Линейная алгебра и геометрия.- М.: Наука, - 1986. - 304с.
5. Яблонский С.В. Введение в дискретную математику. - М.: Наука, - 1986. - 384с.
6. Ефимов Н.В, Розендорн Э.Р. Линейная алгебра и многомерная геометрия. - М.: Наука, 1970. - 528с.
7. Fundamentals of neural Networks, Architectures algorithms and application. Lourence Fauselt, USA, Florida, 1994. -461p.
8. Скорняков Л.А. Элементы теории структур. М.: Наука, - 1982. - 160с.
9. Дорогов А.Ю., Алексеев А.А. Анализ пластичности двухслойных быстрых нейронных сетей //Труды международной научно-технической конференции "Нейронные, реляторные сети и модели" - Ульяновск, 19-21 мая 1998г. Т.1. С.49-51.
10. Дорогов А.Ю. Оптимизация структуры двухслойной ядерной нейронной сети по критерию параметрической пластичности// Сборник докладов 5 Всероссийской конференции "Нейрокомпьютеры и их применение" НКП-99. Москва, 17-19 февраля 1999г.- С.368-371.
11. Алексеев А.А., Дорогов А.Ю. Пластичность двухслойных быстрых нейронных сетей// Известия АН. Теория и системы управления. №5, 1999.-С.121-126.

Надійшла 15.03.2000

ЭВРИСТИЧЕСКИЙ АЛГОРИТМ КЛАССИФИКАЦИИ И ЕГО НЕЙРОСЕТЕВАЯ ИНТЕРПРЕТАЦИЯ

В. И. Дубровин, С. А. Субботин

Реализован эвристический алгоритм многомерной классификации по признакам, основанный на объединении результатов одномерной классификации с учетом значимости признаков, а также дана его нейросетевая интерпретация. Проведены эксперименты по сравнению различных модификаций данного алгоритма на примере решения практических задач.

Реалізовано евристичний алгоритм багатомірної класифікації за ознаками, заснований на поєднанні результатів одномірної класифікації з урахуванням значимості ознак, а також дана його нейронетєва інтерпретація. Проведені експерименти з порівняння різноманітних модифікацій даного алгоритму на прикладі розв'язання практичних задач.

The heuristic algorithm of many-dimensional classification on the features is realised. This algorithm based on the summation of one-dimensional classification results with allowance for features significances. The neural network interpretation of developed algorithm is given. The experiments on matching of various implementations of this algorithm on an examples of the practical problems solution are conducted.

1 ВВЕДЕНИЕ

Возможности прогнозирования привлекают в настоящее время широкие круги специалистов как в области исследования и разработки методов прогнозирования, так и в области их практического применения в производстве и эксплуатации.

Использование прогнозирования в производстве позволяет устранить потенциально ненадежные изделия из готовой продукции, что само по себе очень важно, так как способствует повышению эффективности эксплуатации. Кроме того, анализ причин выпуска потенциально ненадежных изделий позволяет оперативно воздействовать на производство, так как появляется возможность управления качеством выпускаемой продукции за счет введения обратной связи от прогнозирования к производству.

Цель прогнозирования в эксплуатации - предотвращение отказов и увеличение сроков между профилактическими работами путем выявления и исключения из эксплуатации потенциально ненадежных экземпляров с ухудшенными значениями параметров и интенсивным старением.

При прогнозировании по признакам с классификацией задача состоит в разделении исследуемой совокупности изделий на классы и нет необходимости в оценке конкретного значения прогнозируемого параметра. В большинстве практических приложений этого метода число классов равно двум. Так бывает, например, когда

исследуемую совокупность необходимо по заданному правилу разделить на класс годных и дефектных изделий.

Для решения задачи прогнозирования методами теории статистической классификации необходимо располагать условными многомерными плотностями распределения признаков для каждого класса, тогда задача заключается в отыскании способа принятия оптимального решения о принадлежности проверяемого экземпляра к тому или иному классу в условиях неопределенности, т.е. в условиях действия случайных факторов, маскирующих связь между признаками и классом экземпляра.

Классические статистические методы [1] дают оптимальное решение задачи прогнозирования. Однако практическое применение этих методов возможно, если проведен специальный эксперимент по сбору и такой обработке статистических данных о прогнозируемом параметре и признаках, в результате которой найдены подходящие аналитические модели условных многомерных плотностей распределения прогнозируемого параметра и признаков. Однако в реальных задачах исследователь сталкивается здесь с рядом проблем, поэтому реализовать классические статистические методы не всегда возможно.

Во-первых, для реальных изделий даже при известной совокупности информативных признаков (выявление которых представляет весьма трудоемкую задачу) не всегда изучены многомерные условные плотности распределения признаков и прогнозируемого параметра.

Во-вторых, получение аналитических моделей этих условных плотностей распределения представляет трудоемкий процесс и может быть поставлено только отдельной самостоятельной задачей для каждого типа изделий и для данных условий эксплуатации.

В-третьих, даже если такие аналитические модели получены, необходимые при этих методах прогнозирования аналитические преобразования достаточно сложны. Задача относительно легко решается аналитически, если многомерные условные плотности подчиняются нормальному закону, что в действительности имеет место далеко не всегда.

В связи со сказанным выше представляет интерес применение методов решения задач прогнозирования, основанных на эвристических алгоритмах. Смысл понятия "эвристический алгоритм" состоит в том, что в этом случае алгоритм прогнозирования не вытекает из строгих положений теории, а в значительной степени осно-

ван на интуиции и опыте исследователя.

Такие методы могут давать удовлетворительные результаты и при ограниченной исходной информации о вероятностных характеристиках признаков и прогнозируемого параметра. Так, для применения этих методов для прогнозирования по признакам необходимо иметь набор признаков, сильно коррелированных с прогнозируемым параметром, и необязательно знать вид их условных плотностей распределения.

Следует сказать, что методы прогнозирования, основанные на использовании эвристических алгоритмов, не всегда приводят к оптимальным решениям. Однако для их применения на практике достаточно, чтобы ошибка прогнозирования не превышала допустимого значения, а этого можно добиться, например, подбором более информативных признаков, применением соответствующих способов улучшения оператора прогнозирования.

2 ЭВРИСТИЧЕСКИЙ АЛГОРИТМ КЛАССИФИКАЦИИ

В настоящей работе решение задачи классификации осуществляется по данным обучающего эксперимента. При этом нет необходимости прибегать к сложным аналитическим преобразованиям и статистическому эксперименту по определению оценки совместной плотности, если она неизвестна.

Пусть мы имеем обучающую выборку $x = (x^1, x^2, \dots, x^S)$, состоящую из S экземпляров, характеризующихся N признаками x_i^q , где q - номер экземпляра обучающей выборки, i - номер признака. Каждому экземпляру обучающей выборки сопоставлен номер класса y^q , $\forall y^q \in \{K_0, K_1\}$, где K_0 и K_1 - условные обозначения разных классов экземпляров. Условимся, что проверяемый экземпляр принадлежит к классу K_0 , если значение прогнозируемого параметра к моменту времени прогнозирования будет больше некоторого граничного значения; будем считать такие изделия годными. В противном случае экземпляр принадлежит к классу K_1 (дефектных).

В общем случае, экземпляры характеризуются достаточно большим количеством признаков, имеющих разную (как правило, относительно небольшую) информативность. Признаки зачастую связаны с прогнозируемым номером класса экземпляра нелинейными связями, что не позволяет в большинстве случаев построить линейную модель для классификации по одному признаку, удовлетворяющую заданным требованиям достоверности прогнозирования. Поэтому метод классификации должен быть многомерным (учитывать все признаки) и должен учитывать информативность (значимость) признаков.

Наиболее простым способом реализации многомерной классификации, очевидно, будет объединение результа-

тов одномерных классификаций с учетом значимости признаков.

Если $\psi(x_i)$ - результат одномерной классификации по i -му признаку, то y - номер класса экземпляра можно представить как округленную взвешенную сумму:

$$y = \text{round} \left(\sum_{i=1}^N \alpha_i \psi(x_i) \right), \quad (1)$$

где α_i - коэффициент, учитывающий значимость результата одномерной классификации по i -му признаку (доля i -го признака в формировании значения y).

Если значение α_i известно, то задача сводится к разработке правила одномерной классификации для i -го признака относительно порога классификации θ_i . Это правило должно быть нелинейным, а также должно учитывать степень близости экземпляров к центрам классов. В качестве такого правила предлагается использовать сигмоидную функцию:

$$\psi(x_i) = \frac{1}{1 + e^{(x_i - \theta_i)}},$$

которая будет тем ближе к 1, чем ближе экземпляр к центру класса K_1 , и наоборот, тем ближе к 0, чем ближе экземпляр к центру класса K_0 . Это правило предполагает, что центр класса K_0 находится левее центра класса K_1 , что на практике происходит далеко не всегда, поэтому введем в правило одномерной классификации параметр β_i , учитывающий наиболее вероятное размещение центров классов относительно порога θ_i :

$$\psi(x_i) = \frac{1}{1 + e^{\beta_i(x_i - \theta_i)}}. \quad (2)$$

Подставив (2) в выражение (1), получим правило классификации:

$$y = \text{round} \left(\sum_{i=1}^N \alpha_i \frac{1}{1 + e^{\beta_i(x_i - \theta_i)}} \right). \quad (3)$$

Для нахождения значений параметров α_i , β_i и θ_i предлагается использовать следующий алгоритм.

Шаг 1. Вычислить:

M_{x_i} - математическое ожидание i -го признака x_i :

$$M_{x_i} = \frac{1}{S} \sum_{q=1}^S x_i^q, \quad i = 1, 2, \dots, N,$$

где S - количество экземпляров обучающей выборки, x_i^q - значение i -го признака q -го экземпляра обучающей выборки.

$M_{x_i}^{K_1}$ - математическое ожидание i -го признака для экземпляров обучающей выборки, принадлежащих к классу K_1 :

$$M_{x_i}^{K_1} = \frac{1}{S^{K_1}} \sum_{q=1}^{S^{K_1}} x_i^q, \quad x_i^q \in K_1, \quad i = 1, 2, \dots, N,$$

где S^{K_1} - количество экземпляров обучающей выборки, принадлежащих к классу K_1 .

$M_{x_i}^{K_0}$ - математическое ожидание i -го признака для экземпляров обучающей выборки, принадлежащих к классу K_0 :

$$M_{x_i}^{K_0} = \frac{1}{S^{K_0}} \sum_{q=1}^{S^{K_0}} x_i^q, \quad x_i^q \in K_0, \quad i = 1, 2, \dots, N,$$

где S^{K_0} - количество экземпляров обучающей выборки, принадлежащих к классу K_0 .

M_y - математическое ожидание номера класса:

$$M_y = \frac{1}{S} \sum_{q=1}^S y^q,$$

где y^q - номер класса q -го экземпляра обучающей выборки.

Для нахождения порога θ_i , вычислить:

Дисперсии признаков:

$$D_{x_i} = \frac{1}{S} \sum_{q=1}^S (x_i^q - M_{x_i})^2, \quad i = 1, 2, \dots, N.$$

Дисперсии признаков экземпляров, относящихся к классу K_1 :

$$D_{x_i}^{K_1} = \frac{1}{S^{K_1}} \sum_{q=1}^{S^{K_1}} (x_i^q - M_{x_i}^{K_1})^2, \quad x_i^q \in K_1, \quad i = 1, 2, \dots, N.$$

Дисперсии признаков экземпляров, относящихся к классу K_0 :

$$D_{x_i}^{K_0} = \frac{1}{S^{K_0}} \sum_{q=1}^{S^{K_0}} (x_i^q - M_{x_i}^{K_0})^2, \quad x_i^q \in K_0, \quad i = 1, 2, \dots, N,$$

Дисперсию номера класса:

$$D_y = \frac{1}{S} \sum_{q=1}^S (y^q - M_y)^2.$$

Шаг 2. Вычислить коэффициенты корреляции каждого i -го признака и номера класса .

$$r_{x_i y} = \frac{\sum_{q=1}^S (x_i^q - M_{x_i})(y^q - M_y)}{\sqrt{\sum_{q=1}^S (x_i^q - M_{x_i})^2 \sum_{q=1}^S (y^q - M_y)^2}}.$$

Шаг 3. Вычислить степень (долю) влияния i -го признака на номер класса экземпляра

$$\alpha_i = \frac{|r_{x_i y}|}{\sum_{j=1}^N |r_{x_j y}|}, \quad i = 1, 2, \dots, N,$$

где j - номер текущего признака .

Вычислить коэффициент β_i , учитывающий наиболее вероятное размещение полюсов (центров сосредоточения экземпляров) классов при одномерной классификации по i -му признаку:

$$\beta_i = \text{sign}(M_{x_i}^{K_1} - M_{x_i}^{K_0}).$$

Этот коэффициент будет равен:

+1, если полюс класса K_0 расположен левее полюса класса K_1 по оси значений i -го признака;

-1, если полюс класса K_0 расположен правее полюса класса K_1 по оси значений i -го признака;

0, если полюса классов совпадают.

Вычислить значение порога, относительно которого будем осуществлять одномерную классификацию экземпляров по i -му признаку. Для нахождения значения порога можно предложить достаточно много различных способов. Рассмотрим наиболее быстрые и простые.

$$\theta_i = \frac{|M_{x_i}^{K_1} - M_{x_i}^{K_0}|}{2} + \min(M_{x_i}^{K_1}, M_{x_i}^{K_0}), \quad i = 1, 2, \dots, N, \quad (4)$$

$$\theta_i = M_{x_i} + \frac{D_{x_i(0,5-M_y)}}{r_{x_i y} D_y}, \quad i = 1, 2, \dots, N, \quad (5)$$

$$\theta_i = \max(M_{x_i}^{K_1}, M_{x_i}^{K_0}) - \frac{|M_{x_i}^{K_1} - M_{x_i}^{K_0}|}{1 + \frac{D_{x_i}^{K_0}}{D_{x_i}^{K_1}}}, \quad i = 1, 2, \dots, N, \quad (6)$$

Шаг 4. Оценить вероятности ошибки переименования классов для экземпляров обучающей и (или) контрольной выборок и сделать вывод о применимости данного алгоритма для решения поставленной задачи. Для этого следует найти значения номеров классов для экземпляров обучающей и (или) контрольной выборок по правилу классификации (3), после чего определить количество неверных решений $S_{\text{ош}}$ и оценить вероятность принятия ошибочных решений $P_{\text{ош}} = S_{\text{ош}} / (S + S_k)$, где S и S_k - размер обучающей и контрольной выборок, соответственно.

3 НЕЙРОСЕТЕВАЯ ИНТЕРПРЕТАЦИЯ ЭВРИСТИЧЕСКОГО АЛГОРИТМА

Из приведенного эвристического алгоритма можно видеть, что его правило классификации делится на два этапа: на первом этапе осуществляется одномерная классификация на основе N признаков, на втором этапе осуществляется объединение результатов, полученных на первом этапе, с учетом значимости признаков. Такой метод подобен классификации на основе двухслойного перцептрона, являющегося частным случаем многослойной нейронной сети (МНС) [2]. Применение нейронных сетей (НС) для решения задачи классификации по признакам целесообразно, поскольку они обладают высокими адаптивными способностями, способны обучаться аппроксимировать многомерные функции, то есть, могут извлекать, правда, в неявном для пользователя виде, знания из исследуемой предметной области [3 - 6].

Целью процесса обучения НС является настройка множества весов НС таким образом, чтобы удовлетворить некоторому критерию обучения, например, минимуму среднеквадратической ошибки сети. Для обучения МНС применяют градиентные алгоритмы, наиболее быстрым среди которых на сегодняшний день является алгоритм Левенберга-Марквардта [5,7].

Процесс обучения МНС является итерационным и, в общем случае, достаточно длительным, поскольку заранее нельзя определить количество итераций, необходимых для обучения НС. Поэтому разработка методов, позволяющих обучать НС в безытерационном режиме путем проекции обучающих данных на множество весов НС, представляется весьма актуальной и важной.

Одним из таких методов настройки весов для частного случая МНС - двухслойного перцептрона может служить разработанный нами эвристический алгоритм. Нейросетевая интерпретация эвристического алгоритма классификации представлена на рис.1.

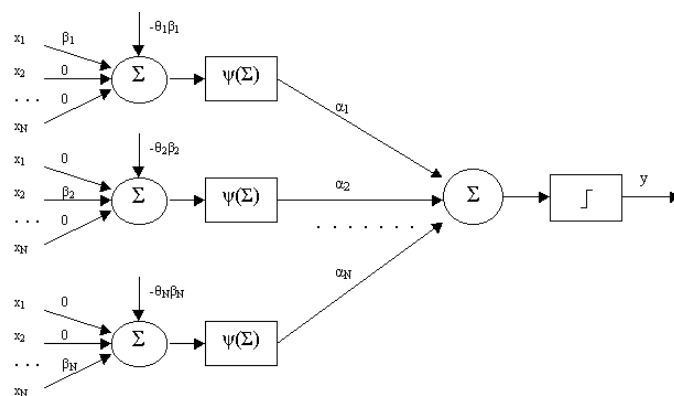


Рисунок 1 - Нейросетевая интерпретация эвристического алгоритма классификации

Правила вычисления параметров эвристического алгоритма в этом случае останутся неизменными, а параметры и функции активации НС необходимо будет определить на их основе по следующим правилам.

Функция активации $\psi^{(\mu, i)}$ i -го нейрона μ -го слоя:

$$\psi^{(1, i)}(\vartheta) = \frac{1}{1 + e^{-\vartheta}}, \quad \forall i = \overline{1, N},$$

$$\psi^{(2, i)}(\vartheta) = \begin{cases} 0, & \vartheta \leq 0, \\ 1, & \vartheta > 0. \end{cases}$$

Весовой коэффициент $w_j^{(\mu, i)}$ j -го входа i -го нейрона μ -го слоя:

$$w_j^{(\mu, i)} = \begin{cases} 0, & i \neq j, j > 0, \mu = 1, \\ \beta_i, & i = j, j > 0, \mu = 1, \\ -\theta_i \beta_i, & j = 0, \mu = 1, \\ \alpha_i, & \mu = 2, \end{cases}$$

$$i = 1, 2, \dots, N, \quad j = 1, 2, \dots, N$$

4 РЕЗУЛЬТАТЫ И ВЫВОДЫ

Для проверки работоспособности эвристического алгоритма на его основе осуществлялось решение ряда практических задач классификации: задача диагностики лопаток газотурбинных авиадвигателей по спектрам свободных затухающих колебаний после ударного возбуждения, задача распознавания сельскохозяйственных растений по результатам дистанционного зондирования, задача прогнозирования состояния спортивной формы спортсменов высшей квалификации на основе данных биохимического контроля тренировочных нагрузок.

Для проведения экспериментов по решению поставленных задач была выполнена программная реализация алгоритма эвристической классификации на языке мак-

росов пакета MATLAB 5.2. Для сравнения разработанного алгоритма с алгоритмом обучения Левенберга-Марквардта использовался модуль Neural Toolbox пакета MATLAB 5.2. При этом в качестве модели НС, обучавшейся на основе алгоритма Левенберга-Марквардта, использовался двухслойный перцептрон, содержащий на первом слое столько нейронов, сколько и признаков, а на втором слое - 1 нейрон. В качестве цели обучения была задана среднеквадратическая ошибка 0.01, максимальное количество циклов обучения - 500.

Результаты экспериментов показывают, что время обучения НС на основе алгоритма Левенберга-Марквардта, в целом, существенно больше чем время обучения НС на основе алгоритма эвристической классификации, но при этом алгоритм Левенберга-Марквардта обеспечивает меньшие вероятности принятия ошибочных решений. Поэтому на практике применять разработанный метод эвристической классификации следует тогда, когда вероятность принятия ошибочных решений не будет превышать заданное значение.

Эвристический алгоритм, предложенный в настоящей работе, является приемлемым для решения многих практических задач (обладает универсальностью), хотя его конкретные реализации, зависящие от правила вычисления значения порога θ_i , могут иметь более узкие области применения. Можно сделать вывод о том, что

эвристический алгоритм целесообразно применять в тех случаях, когда признаки достаточно информативны и граница между классами не очень сложная.

ПЕРЕЧЕНЬ ССЫЛОК

1. Дубровин В.И. Идентификация и оптимизация сложных технических процессов и объектов. - Запорожье: ЗГТУ, 1997. - 92 с.
2. Субботін С.О. Нейронні мережі керують якістю // Пульсар, 1999, № 12, С. 8 - 10
3. Дубровин В.И., Субботин С.А. Построение адаптивных систем диагностики на основе нейронных сетей с латеральным торможением // Радиоелектроніка. Інформатика. Управління, 1999, № 2, С. 110 -114
4. Дубровин В.И., Субботин С.А. Построение систем диагностики на основе карт самоорганизации Кохонена / 6-я Всероссийская конференция "Нейрокомпьютеры и их применение", Москва 16 - 18 февраля 2000: Сборник докладов. - М.: Издательское предприятие журнала "Радиотехника", 2000. - С. 464 - 467
5. Дубровин В.И., Субботин С.А. Нейросетевое моделирование и оценка параметров нелинейных регрессий / 6-я Всероссийская конференция "Нейрокомпьютеры и их применение", Москва 16 - 18 февраля 2000: Сборник докладов. - М.: Издательское предприятие журнала "Радиотехника", 2000. - С. 118 - 120
6. Dubrovин V.I., Subbotin S.A. Choice of neuron transfer functions and research their influence for learning quality of neural networks / Proceedings of International Conference on Modern Problems of Telecommunications, Computer Science and Engineers Training TCSET'2000. - Lviv - Slavsko, 2000. - pp. 114 - 115
7. Neural Network Toolbox User Guide / Beale M., Demuth H. - Natick: Mathworks, 1997. - 700 p.

Надійшла 03.03.2000

Після доробки 21.04.2000

УДК 681.32:007

ВИЗУАЛІЗАЦІЯ ПРОИЗВОЛЬНИХ ДАННИХ МЕТОДОМ УПРУГИХ КАРТ

А. Ю. Зиновьев, А. А. Питенко

Описується технологія візуалізації произвольних даних. Розглядається представлення даних і алгоритм побудови пружної карти, що моделює дані. Запропоновано різні способи проектування багатомірних даних на двумірну сітку. Продемонстровано застосування методів візуалізації будь-яких даних на прикладі картографування економічних таблиць.

Описано технологію візуалізації будь-яких даних. Розглядається уявлення даних та алгоритм побудови пружної карти, що моделює дані. Запропоновано різні способи проектування багатомірних даних на двумірну сітку. Продемонстровано застосування методів візуалізації будь-яких даних на прикладі картографування економічних таблиць.

The technology of any data visualization is described. The data representation and algorithm of construction of an elastic map simulating the data is considered. The various ways of projecting multidimensional data on a two-dimensional grid are offered. The application of methods to mapping of economical tables is discussed.

1 ВВЕДЕНИЕ

В самых разных областях человеческой деятельности (в медицине, биологии, экономике и т.д.) исследователи сталкиваются с необходимостью осмысления больших таблиц данных, собранных в результате наблюдения за свойствами объектов той или иной природы. Как правило, такие таблицы содержат информацию о состоянии нескольких сотен или тысяч объектов, по каждому из которых известны значения определенного набора интересующих исследователя свойств. Число таких свойств (*признаков*) может также достигать нескольких сотен. Естественно, среди исследуемых объектов могут найтись такие, некоторые свойства которых неизвестны или недоступны для измерения. Такие объекты называются неполными данными или *данными с пробелами*.

Традиционным приемом при анализе таких таблиц является их представление, когда каждому исследуемому объекту сопоставляется точка в некотором аб-

страктом *многомерном пространстве данных*. Размерность этого пространства равна числу свойств-признаков в наборе, который характеризует состояние каждого из объектов. Значение каждой из координат точки в пространстве данных равно значению соответствующего признака, так что близким (в той или иной метрике) точкам в пространстве данных соответствуют объекты со сходными свойствами.

Таким образом, таблица данных представлена как облако точек в многомерном пространстве. Следует заметить, что такая картина не совсем верна в случае неполных данных. Так, например, если для объекта неизвестно значение только одной из координат, то его правильнее представлять как прямую, параллельную соответствующей координатной оси. Если на значение отсутствующего признака наложены априорные ограничения, то прямая превращается в отрезок. В случае когда число неизвестных признаков более одного, то объект представляется в виде m -мерной плоскости (m - число неизвестных признаков), параллельной m координатным осям, или, соответственно, многомерным прямоугольным параллелепипедом.

Об анализе (осмыслении) набора данных можно говорить как о красивом, наглядном и компактном их описании. На сегодняшний день известно два основных способа такого описания. Первый из них заключается в том, что точки данных разбиваются на несколько больших классов (кластеров), затем в классах ищутся более мелкие подклассы и т.д. В конечном итоге таблица данных описывается как иерархическая система кластеров в облаке точек.

Второй способ - сокращение размерности описания данных. Дело в том, что человеческий мозг неспособен к эффективному анализу объектов размерности более

трех, поэтому возникает необходимость в построении различных сечений многомерных данных, причем эти сечения малой размерности, как правило, выбираются так, чтобы максимально полно сохранить информацию об имеющихся в наборе данных закономерностях. Такой подход, например, характерен для традиционного линейного факторного анализа, когда в качестве такого сечения выбирается пространство, натянутое на несколько главных собственных векторов корреляционной матрицы (*главных компонент*).

Под *визуализацией данных* можно понимать такой способ описания данных, когда размерность их описания сокращается до *двух* измерений. В этом случае данные можно изображать, например, в виде точек на экране монитора компьютера. Если в процессе сокращения размерности будет сохранена существенная часть закономерностей, присущих данным (разбиение на классы, отношения соседства), то исследователь получает возможность наглядно представить себе исходный набор многомерных данных, сделать выводы об их распределении.

Для визуализации многомерных данных применяется метод, заключающийся в построении вложенного в многомерное пространство данных двумерного многообразия, называемого *картой*, которое определенным способом *моделирует* или аппроксимирует данные (то есть большая часть точек данных лежит в окрестности карты). После этого точки данных с помощью определенной процедуры переносятся или *проецируются* на карту (см. рис. 1). После того, как найдены положения всех образов точек данных на карте, каждому объекту из набора данных можно поставить в соответствие пару координат, характеризующих положение образа на двумерной карте.

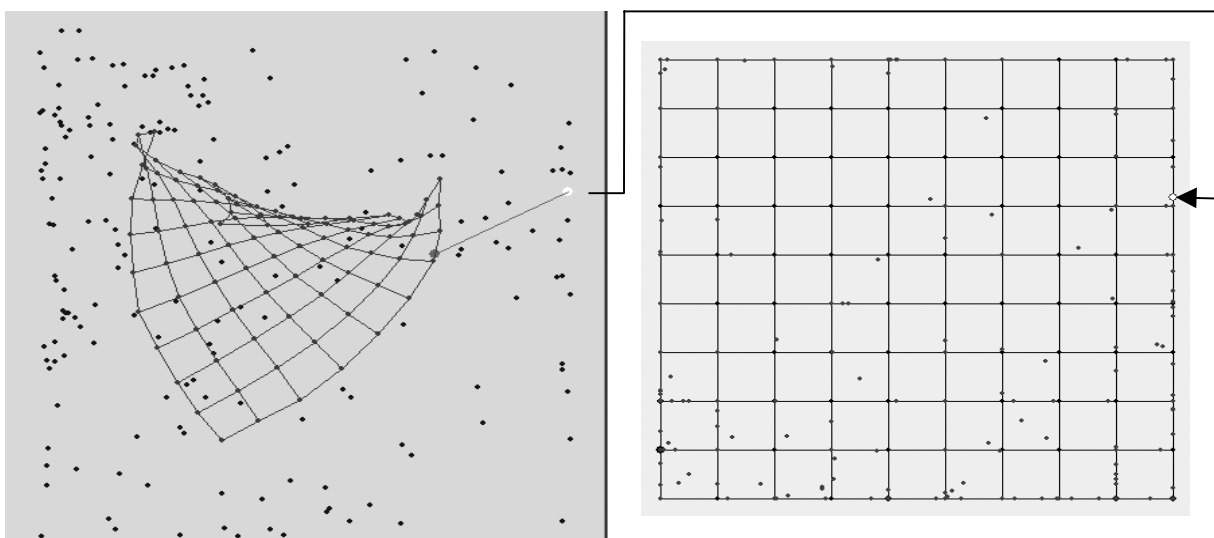


Рисунок 1 - Процедура проекции данных на карту

Слева на рисунке 1 показано облако точек данных в пятимерном пространстве (показана проекция на первые две координатные оси), и карта, моделирующая эти данные. Белый кружок с линией отображает процесс проектирования многомерных данных на двумерное многообразие. Справа - двумерная карта с нанесенными на нее образами данных "развернута" и каждому объекту соответствует теперь только две координаты образа на карте.

Сама идеология построения моделирующих карт впервые была предложена Кохоненом [1]. Им же был разработан алгоритм построения самоорганизующихся карт (SOM), который с успехом применяется на практике. В дальнейшем алгоритм SOM совершенствовался в нескольких направлениях и на данный момент известно большое количество его модификаций. В данной статье предлагается принципиально новый алгоритм построения карт, названных *упругими*, свойства которых отличны от SOM. Кроме этого, описываются различные способы непрерывного проектирования данных на карту (а не только в узлы, как это было предложено Кохоненом), что позволяет посмотреть на принципы визуализации данных в ином ракурсе. Также приведен пример картографирования экономических показателей двухста крупнейших российских предприятий по данным журнала "Эксперт".

2 ПОСТАНОВКА ЗАДАЧИ

Проблема визуализации многомерных данных может быть разбита на несколько задач. Сформулируем эти задачи на математическом языке.

Рассматривается пространство R^n , в котором задано конечное множество точек X , которое интерпретируется как набор произвольных данных, полученных в результате наблюдений за состоянием исследуемой системы. Результат одного измерения (или набор признаков для одного объекта) изображаются точкой x^i ($i = 1 \dots N$, N - количество точек в X) в R^n . Расстояние между точками - обычное евклидово.

1) Построение карты

В целях исследования множества X оно аппроксимируется двумерным многообразием. Можно указать два основных способа построения такого моделирующего многообразия. Во-первых, можно построить какое-либо стандартное линейное многообразие (например, плоскость двух главных компонент) и искать малые отклонения от него. Такой подход можно назвать *квазилинейным*.

Нами используется принципиально нелинейный способ. В R^n размещается двумерная прямоугольная сетка G , расположение узлов которой r^{ij} , удовлетворяет опре-

деленным требованиям. Для упругой карты этими требованиями являются: а) близость узлов сетки к данным; б) не слишком сильная "растянутость" сетки; в) не слишком сильная изогнутость сетки.

Будем считать, что на прямоугольной сетке изначально задана "внутренняя" ортонормированная система координат u, v так, что целым положительным значениям координат соответствуют вершины сетки. Эти целые значения и используются в качестве значений индексов i, j для r^{ij} ($1 \leq i \leq n_1, 1 \leq j \leq n_2$, где n_1, n_2 - число узлов сетки по горизонтали и вертикали).

После того, как сетка G размещена в многомерном пространстве, она *доопределяется* с помощью той или иной процедуры интерполяции между узлами до многообразия, которое обозначим через M . Самой простой такой процедурой является какой-либо вариант триангуляции, в результате чего получается кусочно-линейное многообразие M .

2) Проектирование данных

Для того, чтобы можно было бы представлять X с помощью двумерного многообразия M , необходимо построить отображение $P: x \in X \rightarrow r \in M$, которое будем называть правилом проектирования или *проектором*. Для рассматриваемых задач желательно, чтобы проектор обладал следующими качествами:

- а) проектор должен сохранять отношения соседства, то есть желательно, чтобы близким точкам в R^n соответствовали близкие точки на карте;
- б) проектор, по крайней мере, в некоторой конечной окрестности сетки должен быть однозначным;
- в) проектор должен быть по возможности непрерывным, чтобы плавным изменениям состояния системы в X соответствовали непрерывные изменения положения образа в M .
- г) проекция должна не слишком сильно отличаться от ближайшей вершины сетки G или, если доопределена карта, то от ближайшей точки карты.

Создателем SOM Кохоненом был применен самый очевидный и в некотором смысле естественный вариант проектирования - кусочно-постоянный. При этом каждой точке из X сопоставляется та вершина сетки, которая является ближайшей к этой точке:

$$P: x^k \rightarrow r^{ij}, x^k \in X, r^{ij} \in G, (x^k - r^{ij})^2 \rightarrow \min.$$

Достоинством такого проектирования являются его логическая прозрачность и простота, очевидный его недостаток - разрывность, что не позволяет подробно изобразить картой структуру X . Тем не менее, такой вид проектирования успешно применяется на практике для раскраски карт Кохонена и может изображать разбиение данных по кластерам, размеры кластеров и их взаимное расположение.

Для применяемой нами технологии построения упругих карт кусочно-постоянный способ проектирования малоприменим, поскольку, в отличие от SOM, каждый из узлов сетки G , вообще говоря, не располагается в центре локального сгущения точек данных. Напротив, упругая карта представляет собой более-менее равномерно натянутую на данные сетку, и поэтому существенная часть данных может быть расположена в промежутках между узлами. В этой статье рассмотрены два варианта кусочно-линейного непрерывного проектирования многомерных данных на двумерную карту.

3 ПОСТРОЕНИЕ УПРУГОЙ КАРТЫ

Основой для построения упругой карты является двумерная прямоугольная сетка G , вложенная в многомерное пространство, аппроксимирующая данные и обладающая регулируемыми свойствами упругости по отношению к растяжению и изгибу.

Расположение узлов сетки ищется в результате решения вариационной задачи нахождение минимума следующего функционала:

$$D = \frac{D_1}{|X|} + \lambda \frac{D_2}{m} + \mu \frac{D_3}{m} \rightarrow \min, \quad (1)$$

где $|X|$ - число точек в X ; m - число узлов сетки (в случае прямоугольной сетки $m = n_1 \cdot n_2$); λ, μ - коэффициенты упругости, отвечающие за растяжение и изогнутость сетки соответственно; D_1, D_2, D_3 - слагаемые, отвечающие за свойства сетки, именно:

$$D_1 = \sum_{ij} \sum_{x \in K_{ij}} \|x - r^{ij}\|^2$$

- является мерой близости расположения узлов сетки к данным. Здесь K_{ij} - подмножество точек из X , для которых узел сетки r^{ij} является ближайшим (*таксоны*):

$$x \rightarrow r^{ij}, \quad \|r^{ij} - x\|^2 \rightarrow \min, \quad K_{ij} = \left\{ x \in X \mid x \rightarrow r^{ij} \right\};$$

$$D_2 = \sum_{ij} \|r^{ij} - r^{i,j+1}\|^2 + \sum_{ij} \|r^{ij} - r^{i+1,j}\|^2$$

- мера растянутости сетки;

$$D_3 = \sum_{ij} \|2r^{ij} - r^{i,j-1} - r^{i,j+1}\|^2 + \sum_{ij} \|2r^{ij} - r^{i-1,j} - r^{i+1,j}\|^2$$

- мера изогнутости (кривизны) сетки.

Алгоритм решения вариационной задачи напоминает метод динамических ядер. На каждом шаге алгоритма:

1) Квадратичный функционал D минимизируется при заданном разбиении множества X на подмножества то-

чек K_{ij} , ближайших к данному узлу сетки. При фиксированном разбиении на таксоны, функционал D квадратичен по переменным r^{ij} , поэтому задача нахождения его минимума сводится к решению системы линейных уравнений с матрицей размерами $m \times m$, причем коэффициенты самой матрицы системы зависят только от положения узлов r^{ij} ;

2) Производится новое разбиение множества X на таксоны, соответствующее новым положениям узлов сетки.

Общий критерий D уменьшается при каждой итерации алгоритма, он естественно ограничен снизу нулем, а число разбиений конечного множества X по таксонам конечно. Следовательно, алгоритм сходится.

При построении карты возникает общая для всех оптимизационных задач проблема нахождения глобального минимума функционала D . Предлагается следующий подход для ее решения. Карта строится следующим образом:

1) Первоначально карта располагается в плоскости первых двух главных компонент. Ее размеры выбираются равными порядка $1/2 - 1/3$ от длины двух главных полуосей эллипсоида рассеяния.

2) К положениям узлов добавляется малый случайный шум для того, чтобы вывести расположение узлов из одной плоскости и избежать возникновения различных вырождений матрицы системы;

3) При нахождении минимума функционала D параметры λ, μ меняются от больших значений (порядка 10) к малым (порядка 0,01). В результате карта в начале расчета очень "жесткая", ее кривизна невелика, в результате чего моделируются самые общие особенности распределения точек данных. При уменьшении коэффициентов упругости карта становится "мягкой" и ее узлы располагаются к точкам данных гораздо ближе. Такой метод можно назвать методом *отжига*.

В результате в пространстве данных располагается более-менее равномерная (в отличие от SOM, где расположение узлов соответствует локальным сгущениям данных) сетка G , аппроксимирующая данные. После проектирования данных на сетку, расстояния между образами и их скоплениями на карте отражают соответствующие особенности распределения данных в многомерном пространстве.

4 ПРОЕКТИРОВАНИЕ МНОГОМЕРНЫХ ДАННЫХ НА ДВУМЕРНУЮ СЕТКУ

После построения сетки G , которая является точечной аппроксимацией множества X , она должна быть доопределена до многообразия. Самым простым способом

является выбор какого-либо способа триангуляции, один из вариантов которой показан на рисунке 2.

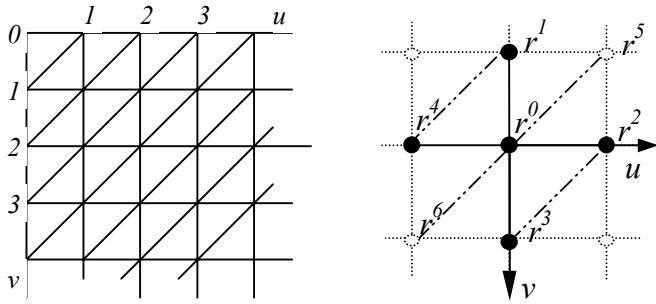


Рисунок 2 - Вариант триангуляции сетки

Слева на рисунке 2 показан один из вариантов триангуляции сетки. Справа - выбранный способ нумерации узлов, окружающих ближайший к данной точке данных узел.

В результате в многомерном пространстве данных располагается двумерное кусочно-линейное многообразие, на котором соседние вершины соединяются ребрами, а ребра ограничивают плоские грани сетки.

В качестве одного из самых очевидных способов проектирования многомерных данных на двумерную сетку можно предложить нахождение ближайшей точки карты (а не только ближайшей вершины, как в случае SOM). Задача состоит в нахождении ближайшей к данной точке данных x точки кусочно-линейного многообразия M . Заметим, что ближайшая точка карты может оказаться на грани сетки, на ребре или в вершине сетки. Таким образом, пространство вокруг карты разбивается на области, где ближайшей точкой карты является точка грани, ребра или вершина.

Прежде всего, определим ближайшую к x вершину сетки и обозначим ее радиус-вектор как r^0 , а окружающие ее вершины перенумеруем изображенным на рисунке 2 образом.

Помимо ближайшей вершины r^0 , можно выбрать еще две наиболее близкие к x вершины r^k и r^l из ближайшего окружения r^0 , причем таким образом, чтобы вершины r^0, r^k, r^l являлись вершинами грани карты при выбранном способе триангуляции. Таким образом мы определяем *ближайшее ребро* и *ближайший треугольник* (грань) карты.

Алгоритм нахождения ближайшей точки карты заключается в следующем:

1. Выполняется ортогональное проектирование x на

плоскость, натянутую на вектора $m = r^k - r^0$ и $n = r^l - r^0$.

Образом x окажется точка $r^* = r^0 + um + vn$, где

$$u = \frac{(x - r^0, m) \cdot n^2 - (n, m)(x - r^0, m)}{m^2 n^2 - (m, n)^2},$$

$$v = \frac{(x - r^0, n) \cdot m^2 - (m, n)(x - r^0, n)}{m^2 n^2 - (m, n)^2}.$$

В случае, если $u \geq 0, v \geq 0, u + v \leq 1$ то ближайшей точкой карты является точка r^* . Иначе

2. Выполняется ортогональное проектирование на ближайшее ребро, натянутое на вектор $m = r^k - r^0$. При этом

$$x \rightarrow r^{**} = r^0 + um, \quad u = \frac{(x - r^0, m)}{m^2},$$

и если $0 \leq u \leq 1$ - ближайшей точкой карты является r^{**} . Иначе ближайшей точкой карты является вершина r^0 .

Другой вариант кусочно-линейного проектора. Для случая моделирования конечного множества точек данных с помощью самоорганизующейся кривой (SOC) [2] был предложен вариант центральной проекции на одномерную сетку. Точки данных сначала ортогонально проектируются в плоскость трех вершин на сетке - ближайшей к заданной точке данных и двух ее соседей, после этого на плоскости выполняется центральное проектирование, в качестве центра проекции O выбирается точка пересечения серединных перпендикуляров к ребрам ломанной, непосредственно прилегающих к ближайшей к точке данных x вершине (см. рис. 3).

Можно перенести этот вариант проектирования на случай двумерной карты, если считать, что проектирование поочередно выполняется сначала для одной, а потом для другой координатной линии двумерной сетки. Таким образом, можно определить значения внутренних координат u, v для точки проекции.

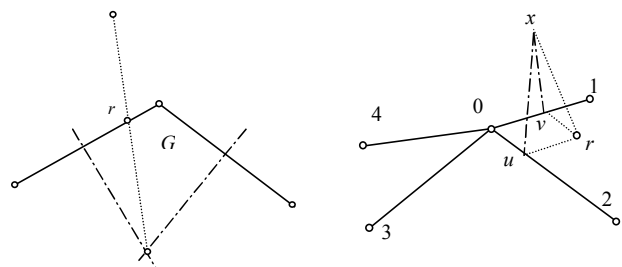


Рисунок 3

Справа на рисунке 3 показан способ нахождения центральной проекции точки x на ломаную, которая является одной из координатных линий u , v двумерного многообразия M . Слева на рисунке 3 изображен способ нахождения координат на карте образа точки x .

Следует заметить, что для описанных выше алгоритмов (как при построении карты, так и при проектировании) не является принципиальным наличие в данных пробелов. В случае, если по какой-либо точке данных известен не полный набор признаков, то вычисление всех необходимых расстояний от точки данных до карты выполняется в пространстве меньшей размерности, где для выбранной точки данных известны все координаты. Таким образом, на карте находят свое положение также и неполные данные, что позволяет использовать изложенные алгоритмы в качестве средства восстановления пробелов в данных. Хотя практика показывает, что точность восстановления данных при таком способе невысока, тем не менее пользователь может получить представление о месте данных с пробелами среди всей совокупности данных.

5 ИСПОЛЬЗОВАНИЕ КАРТЫ ДЛЯ АНАЛИЗА РАСПРЕДЕЛЕНИЯ ДАННЫХ

После того как карта построена, ее можно "развернуть" (см. рис. 1) и наносить разного рода слои информации. В результате карта приобретает разные раскраски, дающие представление об исходном распределении данных. Такие раскраски могут быть построены с использованием богатого арсенала средств и методов ГИС. Таким образом, открывается новое широкое поле деятельности для использования ГИС-технологий для картирования информации самого разного происхождения.

Во-первых, на карте можно изобразить сами данные. При этом можно отображать различные разбиения на подмножества данных, в соответствии со значением того или иного признака.

Во-вторых, на карте можно изобразить произвольные функции координат, поскольку каждой точке с координатами u , v на двумерной карте соответствует точка в многомерном пространстве данных. Самыми простыми раскрасками являются раскраски, отражающие значение той или иной координаты в точках размещения карты. Кроме этого, на карте можно отображать такие координатные функции, как плотность распределения данных в пространстве или плотность того или иного подмножества данных. Сама плотность может быть рассчитана с помощью какой-либо непараметрической оценки. Кроме собственно плотностей подмножеств интерес могут представлять раскраски, отвечающие значению относи-

тельных плотностей подмножеств на фоне общего распределения.

В-третьих, на карте можно изображать различные свойства самого моделирующего многообразия. Например, можно получить раскраски, отражающие значения метрических коэффициентов карты или ее кривизны, свидетельствующие о мере ее растянутости и изогнутости на разных участках. Также интерес представляет такая раскраска карты, на которой цветом отражены расстояния от точки карты до ближайшей точки данных в многомерном пространстве. Такая раскраска позволяет оценить степень близости моделирующей карты к точкам данных в разных областях пространства.

Наконец, весьма плодотворной может оказаться идея составления по множеству X нескольких карт (которые можно по аналогии с начертательной геометрией образно назвать "главными проекциями множества"). Первая из таких карт визуализирует сами данные, вторая "главная проекция" визуализирует погрешности аппроксимации данных (то есть разности между положениями точек данных и их образами на карте) и позволяет оценить точность описания данных первой картой, и так далее. Последовательность таких "главных проекций" позволяет моделировать данные с высокой точностью, что, в случае неполных данных, позволяет правдоподобно восстанавливать пропущенные или ремонтировать недостоверные данные.

6 ПРИМЕНЕНИЕ МЕТОДОВ ВИЗУАЛИЗАЦИИ ПРОИЗВОЛЬНЫХ ДАННЫХ К КАРТОГРАФИРОВАНИЮ ЭКОНОМИЧЕСКИХ ТАБЛИЦ

В качестве примера применения описанной технологии нами была предпринята попытка применить методы визуализации произвольных данных к картографированию таблицы крупнейших российских предприятий, взятой из журнала "Эксперт-200" [3]. Файлы исходных данных были получены с официального сайта журнала <http://www.expert.com>.

Исходная таблица содержала информацию об экономическом положении двухста крупнейших российских предприятий, ранжированную в порядке убывания валового объема производства продукции. Изначально таблица содержала следующие поля-признаки (часть из них является независимыми признаками, часть рассчитывается по явным формулам):

- 1) Название предприятия;
- 2) Регион местонахождения предприятия;
- 3) Отрасль, к которой относится предприятие;
- 4) Валовый объем производства продукции в 1998 году;
- 5) Валовый объем производства продукции в 1997 году;
- 6) Темпы роста предприятия
- 7) Валовый объем производства в 1998 году, выра-

женный в долларовом эквиваленте;

- 8) Балансовая прибыль предприятия;
- 9) Прибыль предприятия после налогообложения;
- 10) Прибыльность предприятия;
- 11) Число работающих на предприятии;
- 12) Производительность труда.

Шумским [4] уже была предпринята попытка визуализации таблицы предприятий, взятой из журнала "Эксперт" за 1997 год. В этой работе были использованы традиционные самоорганизующиеся карты Кохонена и диаграммы Хинтона. Там же было предложено использовать в качестве координат пространства данных отношения некоторых независимых признаков из таблицы. Было предложено четыре таких координаты.

Нами было решено расширить пространство исходных данных еще одним измерением, в результате чего был получен набор независимых факторов (табл. 1).

В результате была составлена таблица из двухсот записей с пятью полями. Часть записей содержала неполную информацию (по отдельным признакам информация отсутствовала).

Данные были предварительно нормированы по формуле $\tilde{x}_i = \text{th}\left(\frac{x_i - M}{\sqrt{D}}\right)$, где \tilde{x}_i , x_i , M , D - новое, старое значения фактора, среднее значение и дисперсия признака соответственно.

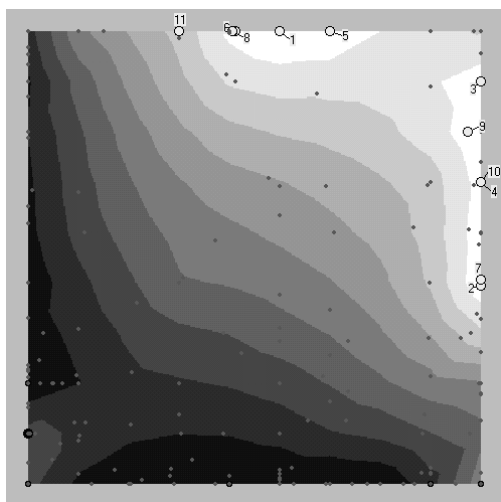
Карта, с помощью которой осуществлялась визуализация множества данных, была построена по описанному выше алгоритму построения упругих карт. Первоначальная сека содержала 10 узлов по вертикали и 10 по горизонтали. Для нахождения локального минимума функционала применялся описанный выше метод отжига. Параметры μ и λ медленно (так чтобы при каждом изменении карта успевала перейти в близлежащий локальный минимум) менялись от значений $\mu = 5$, $\lambda = 5$ до $\mu = 0, 1$, $\lambda = 0, 1$.

После построения упругой карты данные из пространства факторов были спроектированы на карту с помощью описанной выше процедуры нахождения ближайшей точки карты в случае кусочно-линейной интерполяции между узлами.

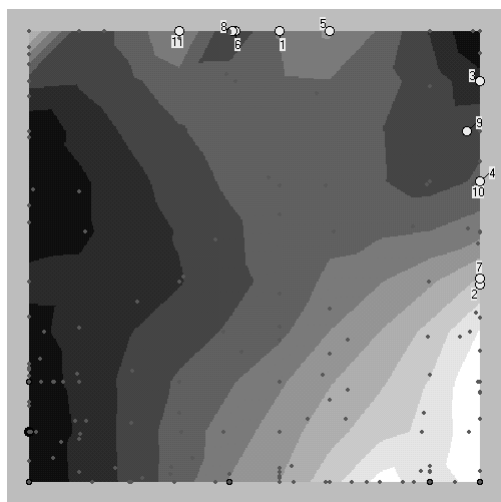
В качестве иллюстрации анализа экономических данных ниже приведены раскраски полученной карты по координатным полям, а также слой рассчитанной плотности данных в точках карты. На раскрасках большими точками с номерами выделена группа предприятий, принадлежащих нефтегазовой промышленности. Такое выделение позволяет проанализировать место той или иной отрасли промышленности среди других предприятий.

Таблица 1

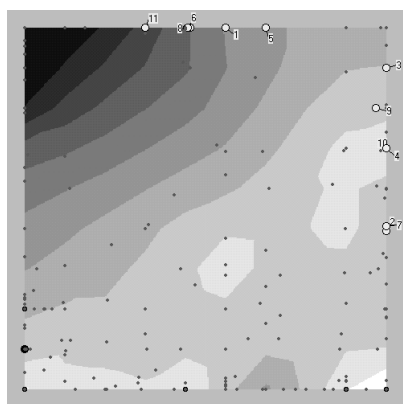
N	Обозначение фактора	Значение	Условное название
1	LG_VO1998	Логарифм валового объема производства продукции в 1998 году	Размер
2	TEMP	Валовый объем производства продукции в 1998 году Валовый объем производства продукции в 1997 году	Темп роста
3	PROFIT_BAL	Балансовая прибыль предприятия Валовый объем производства продукции в 1998 году	Балансовая прибыль
4	PROFIT_NAL	Прибыль предприятия после налогообложения Валовый объем производства продукции в 1998 году	Чистая прибыль
5	PRODUCTIV	Прибыль предприятия после налогообложения Число работающих на предприятии	Производительность



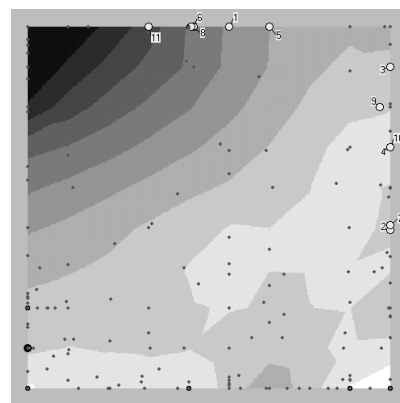
а)



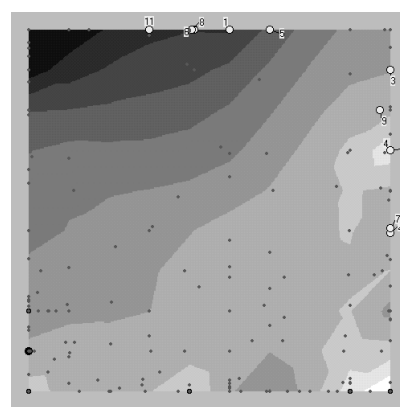
б)



в)



г)



д)

Рисунок 4

1) Раскраска по факторам

На рисунке 4а изображено значение фактора LG_VO1998 в точках карты. При этом более светлым участкам соответствуют более высокие показатели фактора. Самый яркий цвет соответствует первым 10% предприятий с самым большим валовым объемом производства. Для примера кружками с цифрами выделены предприятия нефтегазовой промышленности. Цифрам соответствуют следующие названия предприятий:

1 - ОАО "Газпром"; 2 - Нефтяная компания "ЛУКОЙЛ"; 3 - Башкирская топливная компания; 4 - Нефтяная компания "Сургутнефтегаз"; 5 - Тюменская нефтяная компания; 6 - "Татнефть"; 7 - Нефтяная компания "Славнефть"; 8 - Нефтяная компания "Роснефть"; 9 - Оренбургская нефтяная компания "Онако"; 10 - Центральная топливная компания; 11 - Нефтяная компания "КомитЭК".

Рисунок 4б изображает раскраску по показателю TEMP. Как видно из рисунка 4б, область крупнейших предприятий не пересекается с областью наиболее высоких темпов роста. В правом нижнем углу карты,

например, располагаются предприятия пищевой промышленности, цветной металлургии и другие быстро развивающиеся отрасли.

На рисунках 4в, 4г, 4д показаны раскраски по факторам PROFIT_BAL, PROFIT_NAL, PRODUCTIV. Раскраски этих факторов схожи, что указывает на корреляцию последних трех признаков. Вместе с этим различия в раскраске позволяют выделить предприятия, которые выпадают из корреляционной зависимости.

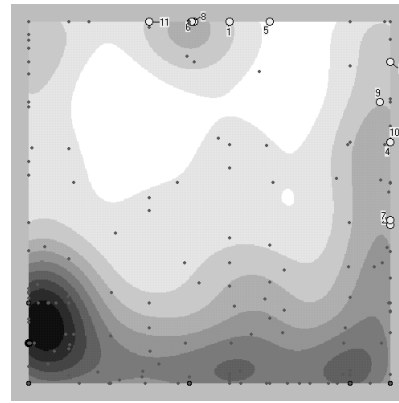
2) Раскраска по плотности данных

На рисунках 5а, 5б, 5в показана раскраска карты по плотности данных, оцененной с помощью какой-либо непараметрической оценки. Существует два способа оценить плотность данных. Во-первых, можно рассматривать двумерное распределение точек на карте. Во-вторых, можно рассчитать плотность точек в исходном n -мерном пространстве, и изображать на карте значения этой плотности в точках расположения карты. На рисунках изображено применение первого способа. Более темным участкам соответствуют более высокие значения плотности.

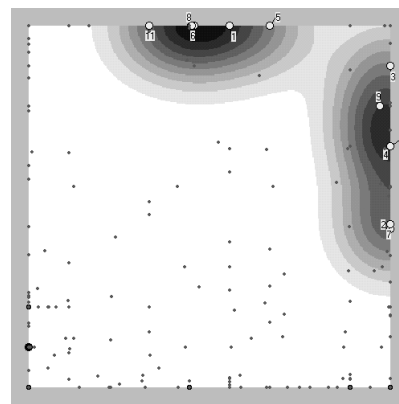
Рисунок 5а изображает двумерное распределение общей плотности данных. На рисунке 5б - распределение плотности предприятий нефтегазовой промышленности. Рисунок 5в отражает удобную для оценок относительную плотность предприятий нефтегазовой промышленности (то есть отношение первых двух плотностей).

На рисунке 6 отражено расстояние от каждой из точек карты до ближайшей точки данных. Более темным участкам соответствуют большие расстояния. Видно, что в целом данные достаточно плотно прилегают к карте, за исключением участка в левом верхнем углу (впрочем, точки данных там отсутствуют и темный цвет указывает на то, что точки в левом верхнем углу карты расположены в многомерном пространстве достаточно далеко от основного массива данных).

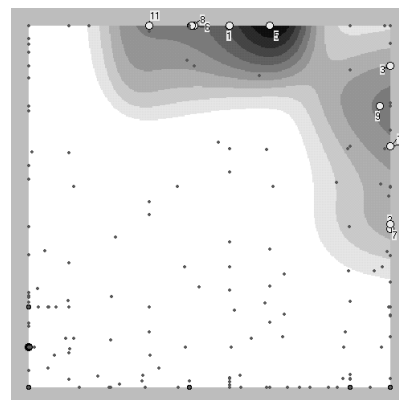
Беглый взгляд на рисунки позволяет сделать, например, такие выводы. Предприятия нефтегазовой промышленности являются лидерами по объему валового производства, но темпы роста этой области промышленности невелики по сравнению, например, с пищевой промышленностью. Предприятия нефтегазовой промышленности распадаются на две группы, которые существенно отличаются по прибыльности производства. В целом, набор таких рисунков могут служить удобным средством анализа для специалистов в макроэкономике.



а)



б)



в)

Рисунок 5

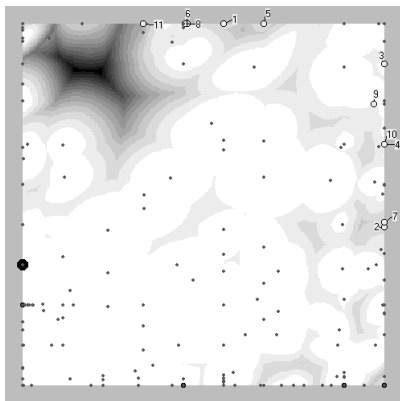


Рисунок 6

7 ЗАКЛЮЧЕНИЕ

Разработана и опробована на практике технология визуализации или картографирования многомерных данных (в которых могут содержаться пробелы) с помощью вложенных в пространство данных двумерных многообразий, названных упругими картами. И алгоритм построения этих многообразий, и общая идеология визуализации данных с их помощью существенно отличается от общепринятой на сегодняшний день технологии SOM. Особенностью технологии также явля-

ется возможность непрерывного проектирования данных на карту, что существенно повышает точность представления данных.

Еще раз стоит отметить, что описанная технология открывает перспективы для использования всего арсенала методов и средств, накопленных в ГИС-технологиях для картирования данных самой различной природы, без привязки к географическим координатам. Можно сказать, что вместо географической карты в описанной технологии используется подложка, образованная структурой самих данных.

Статья выполнена при поддержке гранта для молодых ученых №1M0034 Красноярского краевого фонда науки.

ПЕРЕЧЕНЬ ССЫЛОК

1. Kohonen T. Self-Organizing Maps. Springer Verlag, 1995.
2. Горбань А. Н., Россиев А. А. Итерационный метод главных кривых для данных с пробелами // Проблемы нейрокибернетики: Труды 12 Международной конференции по нейрокибернетике. Ростов-на-Дону: Издательство СКНЦ ВШ, 1999. С. 198-201.
3. "Эксперт-200": ежегодный рейтинг крупнейших компаний России // Журнал "Эксперт". 1999. №36.
4. Шумский С. А., Кочкин А. Н. Самоорганизующиеся карты финансовых индикаторов 200 крупнейших российских предприятий // Материалы Всероссийской научной конференции "Нейроинформатика-99". Москва, 1999. Часть 3. С.122-127.

Надійшла 15.02.2000
Після доробки 21.02.2000

УДК 519.863

МНОГОСТУПЕНЧАТЫЕ ИНТЕРВАЛЫ ПРИ РАСЧЕТЕ ЭКСПЛУАТАЦИОННЫХ ДОПУСКОВ

В. Н. Крищук, Г. Н. Шило, Н. П. Гапоненко

Рассмотрена трехуровневая модель формирования допусков при производстве и эксплуатации электронных аппаратов. Используются эквивалентные параметры многоэлементных моделей. Получены условия компенсации внешних воздействий в многоэлементных моделях.

Розглянута трьохрівнева модель формування допусків при виробництві і експлуатації електронних апаратів. Використовуються еквівалентні параметри багатоелементних моделей. Отримані умови компенсації зовнішніх впливів у багатоелементних моделях.

The three-level model of forming tolerances is considered when producing and maintaining electronic devices. Equivalent parameters of multiple models are used. The conditions of compensation of exposures in multiple models are obtained.

ВВЕДЕНИЕ

В большинстве случаев отклонения параметров электрорадиоэлементов задаются при нормальных условиях окружающей среды. В процессе эксплуатации электронных аппаратов внешние воздействия приводят к изменению номинальных отклонений параметров и ухудшению показателей аппаратуры. Оценки показывают, что изменения отклонений могут достигать 30-50% их номинального значения при использовании электрорадиоэлементов общего назначения. Для прецизионных элементов этот показатель может увеличиваться до 70-150%.

Учет этих особенностей на стадии разработки электронных аппаратов может осуществляться путем представления параметров в виде интервалов, границы которых являются интервальными величинами [1,2].

Эти представления основываются на перемещении границ интервалов при внешних воздействиях. В общем случае можно выделить четыре основных уровня учета отклонений параметров с помощью интервальных величин (рис.1). Нулевому уровню соответствует описание параметров элементов только на основе его номинального значения. Первый уровень позволяет учесть отклонения параметров элементов, возникающие в процессе их изготовления. Наряду с номинальным значением описание параметров модели включает их отклонения. Интервальное представление имеет вид:

$$\mathbf{a} = [\underline{a}; a], \tag{1}$$

где \underline{a} и a - нижняя и верхняя границы интервала \mathbf{a} .

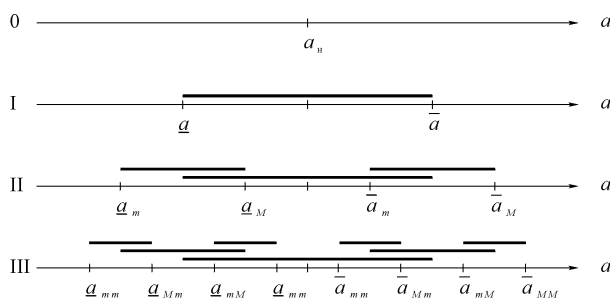


Рисунок 1 - Модели учета отклонений параметров

Второй уровень позволяет учесть отклонения параметров, вызванных внешними воздействиями. Каждая из границ интервала становится интервалом:

$$\underline{\mathbf{a}} = [\underline{a}_m; \underline{a}_M], \mathbf{a} = [a_m; a_M], \tag{2}$$

где индексы m и M соответствуют минимальному и максимальному значению измененных границ интервала.

Параметр при этом становится интервалом с интервальными границами:

$$\mathbf{A} = [\underline{\mathbf{a}}, \mathbf{a}], \tag{3}$$

а граничные значения параметра определяются интервалом:

$$\mathbf{a} = [\underline{a}_m; a_M]. \tag{4}$$

Второму уровню соответствует номинальное значение функции влияния внешних воздействий на параметр. Третий уровень учитывает отклонения функции влияния от ее номинального значения, которые обусловлены особенностями производства электронных аппаратов. Нижнему и верхнему значению функции влияния соответствуют интервалы с интервальными границами (двухступенчатые интервалы):

$$\underline{\mathbf{A}} = [\underline{\mathbf{a}}_m; \underline{\mathbf{a}}_M], \bar{\mathbf{A}} = [\bar{\mathbf{a}}_m; \bar{\mathbf{a}}_M], \tag{5}$$

а параметр становится трехступенчатым интервалом:

$$\mathcal{A} = [\underline{\mathbf{A}}, \bar{\mathbf{A}}]. \tag{6}$$

Граничные значения параметра для третьего уровня определяются интервалами:

$$\mathbf{a} = [\underline{a}_{mm}; a_{MM}]. \tag{7}$$

Вид соотношений (2) - (7) зависит от величины внешних воздействий и функции влияния, связывающей внешние воздействия с изменением параметров элементов. Целью настоящей работы является исследование этого влияния на параметры элементов и на параметры моделей, у которых проявляется взаимодействие различных величин, подверженных внешним воздействиям.

1 РАСЧЕТ ДОПУСКОВ ОДНОЭЛЕМЕНТНЫХ МОДЕЛЕЙ

Перемещение границ интервалов при внешних воздействиях можно представить выражением:

$$a_k(L) = a_k[1 + f_k(L)], \tag{8}$$

где a_k - значение k -ой границы (нижней или верхней) интервала при нормальных условиях окружающей среды;

$L = [l_1, l_2, \dots, l_n]^T$ - вектор n внешних воздействий с амплитудами l_i ;

$f_k(L)$ - функция влияния внешних воздействий на k -ую границу интервала.

При отсутствии внешних воздействий:

$$f(L) = 0, (l_i = 0; i = \overline{1, n}).$$

Если производится оценка границ интервала в диапазоне изменения внешних воздействий, то тогда амплитуды и вектор внешних воздействий становятся интервальными величинами:

$$\mathbf{l}_i = [\underline{l}_i; \bar{l}_i]; \mathbf{L} = [\underline{L}, \bar{L}],$$

а функции $f_k(L)$ соответствует ее интервальное расширение $f_k(\mathbf{L})$ [3,4]. Граничные значения интервала (4) могут тогда определяться с помощью соотношений:

$$\underline{a}_m = \underline{a} \cdot \min[1 + f_n(\mathbf{L})]; a_M = a \cdot \max[1 + f_n(\mathbf{L})], \tag{9}$$

где $f_n(\mathbf{L})$ и $f_n(\mathbf{L})$ - интервальные расширения функции влияния нижней и верхних границ интервала (1).

Учет интервального характера функции влияния пре-

образует (9) в интервальные соотношения и границы интервала (7) определяются выражениями:

$$\underline{a}_{mm} = \underline{a}_m \cdot \min[1 + \mathbf{f}_n(\mathbf{L})]; a_{MM} = a_M \cdot \max[1 + \mathbf{f}_B(\mathbf{L})] \quad (10)$$

где $\mathbf{f}_n(\mathbf{L})$ и $\mathbf{f}_B(\mathbf{L})$ - интервальные функции влияния нижней и верхней границ.

В большинстве практических приложений внешние воздействия проявляются взаимонезависимо. Это позволяет функцию влияния представить в виде сепарабельной функции:

$$f(L) = \sum_{i=1}^n f_i(l_i). \quad (11)$$

Если каждая из функций $f_i(l_i)$ является линейной функцией, то соотношение (11) может быть представлено в виде [2]:

$$f(L) = \sum_{i=1}^n \alpha_i l_i = \alpha L, \quad (12)$$

где α_i - коэффициент i -го внешнего воздействия.

α и L - приведенные коэффициент и величина внешнего воздействия.

С учетом соотношения (12) выражения (9) преобразуются к виду:

$$\underline{a}_m = \underline{a} \cdot \min[1 + \alpha_n \cdot \mathbf{L}]; a_M = a \cdot \max[1 + \alpha_B \cdot \mathbf{L}]. \quad (13)$$

Соотношения (13) позволяют осуществить оценку отклонений при номинальных значениях коэффициентов внешних воздействий. В большинстве случаев такой оценки оказывается достаточно. Но при значительных отклонениях коэффициента внешних воздействий от его номинальной величины возникает необходимость в более точном расчете отклонений верхних и нижних границ

$$\underline{a}_{mm} = \underline{a} \cdot \min[1 + \alpha_n \cdot \mathbf{L}]; a_{MM} = a \cdot \max[1 + \alpha_B \cdot \mathbf{L}], \quad (14)$$

где $\alpha = [\underline{\alpha}; \bar{\alpha}]$ - интервальный коэффициент внешних воздействий.

Соотношения (13) и (14) могут использоваться для расчета предельных отклонений одного параметра. Если эти выражения использовать при расчете допусков моделей с взаимодействующими параметрами, то полученные интервальные оценки не будут учитывать взаимного перемещения границ у различных параметров. В результате могут получиться завышенные значения границ интервалов и, соответственно, допусков.

2 ИССЛЕДОВАНИЕ МНОГОЭЛЕМЕНТНЫХ МОДЕЛЕЙ

В многоэлементных моделях с функциональными зависимостями параметров модели любая из функций может быть представлена в виде ее интервального расширения. После этого анализ внешних воздействий на модели с взаимодействующими параметрами сводится к исследованию арифметических операций над интервалами с подвижными границами.

Внешние воздействия с амплитудой L на интервалы с функциями влияния (12) приводят к одновременному изменению границ интервала:

$$\mathbf{a} = [\underline{a}(1 + \alpha_n L); \bar{a}(1 + \alpha_B L)]. \quad (15)$$

Операция сложения двух таких интервалов приводит к соотношениям:

$$\begin{aligned} \mathbf{a}_1 + \mathbf{a}_2 &= \\ &= [\underline{a}_1(1 + \alpha_{1n} L); \bar{a}_1(1 + \alpha_{1B} L)] + [\underline{a}_2(1 + \alpha_{2n} L); \bar{a}_2(1 + \alpha_{2B} L)] = \\ &= [\underline{a}_1 + \underline{a}_2 + (\alpha_{1n} \underline{a}_1 + \alpha_{2n} \underline{a}_2)L; \bar{a}_1 + \bar{a}_2 + (\alpha_{1B} \bar{a}_1 + \alpha_{2B} \bar{a}_2)L] \end{aligned}$$

Последнее выражение может быть приведено к виду (15) с эквивалентными параметрами:

$$\mathbf{a}_{12} = \mathbf{a}_1 + \mathbf{a}_2 = [\underline{a}_{12}(1 + \alpha_{12n} L); \bar{a}_{12}(1 + \alpha_{12B} L)]. \quad (16)$$

где a_{12} и α_{12} - эквивалентные границы интервалов и коэффициенты внешних воздействий.

Величины a_{12} и α_{12} определяются с помощью соотношений:

$$\begin{aligned} \underline{a}_{12} &= \underline{a}_1 + \underline{a}_2; \bar{a}_{12} = \bar{a}_1 + \bar{a}_2; \\ \alpha_{12n} &= \frac{\underline{a}_1 \alpha_{1n} + \underline{a}_2 \alpha_{2n}}{\underline{a}_1 + \underline{a}_2}; \alpha_{12B} = \frac{\bar{a}_1 \alpha_{1B} + \bar{a}_2 \alpha_{2B}}{\bar{a}_1 + \bar{a}_2}. \end{aligned} \quad (17)$$

Как следует из выражений (17), при одинаковых знаках α_1 и α_2 эквивалентные коэффициенты внешнего воздействия не превышают максимального значения коэффициентов слагаемых. Возможна взаимная компенсация перемещения границ суммы при выполнении условий:

$$\left. \begin{aligned} \underline{a}_1 \alpha_{1n} + \underline{a}_2 \alpha_{2n} &= 0 \\ \bar{a}_1 \alpha_{1B} + \bar{a}_2 \alpha_{2B} &= 0 \end{aligned} \right\}. \quad (18)$$

Если условия (18) выполняются, то отклонение параметра \mathbf{a}_{12} не изменяется при изменении обобщенной амплитуды внешнего воздействия.

Операция вычитания двух интервалов с плавающими границами приводит к выражению:

$$\begin{aligned} \mathbf{a}_{12} &= \mathbf{a}_1 - \mathbf{a}_2 = \\ &= [\underline{a}_1 - \bar{a}_2 + (\alpha_{1n} \underline{a}_1 - \alpha_{2n} \bar{a}_2)L; \bar{a}_1 - \underline{a}_2 + (\alpha_{1B} \bar{a}_1 - \alpha_{2B} \underline{a}_2)L]. \end{aligned}$$

Эквивалентные границы интервалов и коэффициенты внешних воздействий для операции вычитания имеют вид:

$$\begin{aligned} \underline{a}_{12} &= \underline{a}_1 - \underline{a}_2; \quad \alpha_{12} = \alpha_1 - \alpha_2; \\ \alpha_{12H} &= \frac{a_1 \alpha_{1H} - a_2 \alpha_{2B}}{\underline{a}_1 - \underline{a}_2}; \quad \alpha_{12B} = \frac{a_1 \alpha_{1B} - a_2 \alpha_{2H}}{\underline{a}_1 - \underline{a}_2}. \end{aligned} \quad (19)$$

Условия компенсации внешнего воздействия на границы разности вытекают из выражений (19):

$$\left. \begin{aligned} \underline{a}_1 \alpha_{1H} - a_2 \alpha_{2B} &= 0 \\ \bar{a}_1 \alpha_{1B} - a_2 \alpha_{2H} &= 0 \end{aligned} \right\}. \quad (20)$$

Операция умножения двух интервалов вида (15) приводит к соотношениям:

$$\begin{aligned} \mathbf{a}_{12} &= \mathbf{a}_1 \cdot \mathbf{a}_2 = \\ &= [\underline{a}_1 \underline{a}_2 (1 + \alpha_{1H} L)(1 + \alpha_{2H} L); \bar{a}_1 \bar{a}_2 (1 + \alpha_{1B} L)(1 + \alpha_{2B} L)] = \\ &= [\underline{a}_1 \underline{a}_2 (1 + (\alpha_{1H} + \alpha_{2H} + \alpha_{1H} \alpha_{2H} L)L); \bar{a}_1 \bar{a}_2 (1 + (\alpha_{1B} + \alpha_{2B} + \alpha_{1B} \alpha_{2B} L)L)]. \end{aligned} \quad (21)$$

Представление выражения (21) в виде (15) приводит к эквивалентным параметрам:

$$\begin{aligned} \underline{a}_{12} &= \underline{a}_1 \underline{a}_2; \quad \bar{a}_{12} = \bar{a}_1 \bar{a}_2; \\ \alpha_{12H} &= \alpha_{1H} + \alpha_{2H} + \alpha_{1H} \alpha_{2H} L; \quad \alpha_{12B} = \alpha_{1B} + \alpha_{2B} + \alpha_{1B} \alpha_{2B} L. \end{aligned} \quad (22)$$

После умножения двух интервалов с плавающими границами эквивалентные коэффициенты внешнего воздействия оказываются зависящими от амплитуды внешнего воздействия.

При выполнении условий:

$$\alpha_{1H} + \alpha_{2H} \ll \alpha_{1H} \alpha_{2H} L; \quad \alpha_{1B} + \alpha_{2B} \ll \alpha_{1B} \alpha_{2B} L$$

соотношения для эквивалентных коэффициентов внешних воздействий можно представить в виде:

$$\alpha_{12H} \approx \alpha_{1H} + \alpha_{2H}; \quad \alpha_{12B} \approx \alpha_{1B} + \alpha_{2B}.$$

Тогда компенсация внешних воздействий оказывается возможной, если выполняются условия:

$$\alpha_{1H} + \alpha_{2H} = 0; \quad \alpha_{1B} + \alpha_{2B} = 0. \quad (23)$$

Операция деления двух интервалов вида (15) приводит к соотношению:

$$\mathbf{a}_{12} = \frac{\mathbf{a}_1}{\mathbf{a}_2} = \left[\frac{a_1 (1 + \alpha_{1H} L)}{a_2 (1 + \alpha_{2B} L)}; \frac{\bar{a}_1 (1 + \alpha_{1B} L)}{\bar{a}_2 (1 + \alpha_{2H} L)} \right]. \quad (24)$$

Выражение (24) преобразуется к виду (15), если принять обозначения:

$$\underline{a}_{12} = \frac{a_1}{a_2}; \quad \bar{a}_{12} = \frac{\bar{a}_1}{\bar{a}_2};$$

$$\alpha_{12H} = \frac{\alpha_{1H} - \alpha_{2B}}{1 + \alpha_{2B} L}; \quad \alpha_{12B} = \frac{\alpha_{1B} - \alpha_{2H}}{1 + \alpha_{2H} L}. \quad (25)$$

Эквивалентные коэффициенты внешнего воздействия частного (25) оказываются зависящими от амплитуды внешнего воздействия. Условия компенсации внешних воздействий при делении двух параметров имеют вид:

$$\alpha_{1H} - \alpha_{2B} = 0; \quad \alpha_{1B} - \alpha_{2H} = 0. \quad (26)$$

При выполнении условий (18), (20), (23) и (26) отклонение параметров многоэлементных моделей не изменяется при изменении обобщений амплитуды внешнего воздействия. Значения границ интервалов многоэлементных моделей тогда остаются равными их значениям при нормальных условиях окружающей среды.

3 АРИФМЕТИКА ИНТЕРВАЛОВ С ПЛАВАЮЩИМИ ГРАНИЦАМИ

Выражения для эквивалентных параметров многоэлементных моделей и условия компенсации отклонений получены при представлении обобщенных коэффициентов и амплитуд внешних воздействий в виде вещественных чисел. Использование классической интервальной арифметики при интервальном характере этих величин приводит к завышенному значению отклонений. Связано это с тем, что в классической интервальной арифметике не учитывается направление движения границ интервалов при взаимодействии интервальных величин. Для учета этого взаимодействия введем операцию связанного умножения:

$$(\alpha \mathbf{L}) = (\alpha | \underline{L}; \bar{L}) = [\alpha \underline{L}; \alpha \bar{L}], \quad (27)$$

где α - вещественное число.

При положительных значениях α выражение (27) совпадает с выражением классической интервальной арифметики, сохраняющим ориентацию интервала. Отрицательному значению α соответствует обратный интервал. Увеличение L в этом случае сопровождается уменьшением интервальной границы параметра, что учитывает направление перемещения границ. Операция связанного умножения обладает свойством:

$$(\alpha_1 \mathbf{L}) + (\alpha_2 \mathbf{L}) = ((\alpha_1 + \alpha_2) \mathbf{L}). \quad (28)$$

Интервальная запись границ интервала (15) с учетом соотношений (27) и (28) имеет вид:

$$\underline{\mathbf{a}} = \underline{a} (1 + (\alpha_H \mathbf{L})) = \underline{a} + \underline{a} (\alpha_H \mathbf{L});$$

$$\mathbf{a} = a(1 + (\alpha_B \mathbf{L})) = a + a(\alpha_B \mathbf{L}). \quad (29)$$

В этом случае результаты интервальных арифметических операций могут быть представлены с использованием полученных ранее эквивалентных коэффициентов внешних воздействий. Остаются в силе и условия компенсации влияния внешних воздействий на границы параметров многоэлементных моделей.

Применение интервальных коэффициентов внешних воздействий не вносит существенных изменений в полученные соотношения. Следует только учитывать, что во всех соотношениях для эквивалентных коэффициентов подставляется максимальное по модулю значение коэффициента внешних воздействий.

ЗАКЛЮЧЕНИЕ

Таким образом, при расчете влияния внешних воздействий на отклонения параметров могут использоваться многоуровневые модели с интервальными коэффициентами и амплитудами внешних воздействий. В большинстве случаев при расчете отклонений могут использоваться вещественные коэффициенты внешних воздействий. Необходимость использования интервальных ко-

эффициентов внешних воздействий возникает при расчете допусков прецизионных устройств.

В многоэлементных моделях электронных устройств возможно использование эквивалентных коэффициентов внешних воздействий. Важной особенностью этих моделей является возможность компенсации внешних воздействий на границы результирующего интервала. Расчет допусков в многоэлементных моделях производится с учетом взаимного перемещения границ интервалов. Для этого введена операция связанного умножения интервалов.

ПЕРЕЧЕНЬ ССЫЛОК

1. *Krischuk V., Shilo G., Gaponenko N.* Interval calculation of the tolerances at external influences.// Proceedings of International Conference on Modern Problems of Telecommunications, Computer Science and Engineers Training, TCSET'2000, Lviv-Slavsko, 2000. С.34-35.
2. *Кришук В., Шило Г., Гапоненко М.* Інтервальний розрахунок допусків при зовнішніх впливах.// Вісник державного університету "Львівська політехніка" "Радіоелектроніка та телекомунікації", Львів, №387, 2000. С.191-196.
3. *Калмыков С.А., Шокин Ю.И., Юлдашев З.Х.* Методы интервального анализа. - Новосибирск: Наука, 1986.
4. *Alefeld, G and Herzberger, J:* Introduction to Interval Computations. Academic Press, New York, 1983.

Надійшла 21.02.2000

Після доробки 21.04.2000

УДК 618.513

СИНТЕЗ СТРУКТУРИРОВАННЫХ МОДЕЛЕЙ МНОГОСВЯЗНЫХ СИСТЕМ

Е. Н. Литвинов, А. С. Лихоузов, Т. А. Лихоузова

Робота стосується проблеми ідентифікації багатозв'язних систем. Розроблено алгоритм побудови структурованої моделі багатозв'язної системи на основі адаптації відомих алгоритмів структурної ідентифікації систем. Для оцінки якості моделей використовується інформаційний критерій.

Робота пов'язана з проблемою ідентифікації багатозв'язних систем. Розроблено алгоритм синтезу структурованої моделі багатозв'язної системи на основі адаптації відомих алгоритмів структурної ідентифікації систем. Для оцінки якості моделей використовується інформаційний критерій.

The work concerns a problem of identification of multicoherent systems. The algorithm of construction of the structured model of multicoherent system is developed on the basis of adaptation of known algorithms of structural identification of systems. For an estimation of quality of models the informational criterion is used.

Задача идентификации систем по наблюдениям является одной из основных задач современной теории автоматического управления. Она имеет место при изучении свойств и особенностей объектов с целью контроля, последующего управления ими, а также при создании адаптивных систем, в которых на основе иден-

тификации объекта формируются оптимальные управляющие воздействия. С задачей идентификации исследователь сталкивается во время анализа различного рода информации, например, физического, экономического, социологического, биологического характера и т.д.

В последние годы в литературе большое внимание уделяется задаче идентификации систем с конечным числом состояний и, особенно, вероятностных систем [1-6]. Как правило, задача идентификации при этом описывается как многоцелевая задача оптимизации. В такой постановке она формулируется следующим образом: необходимо найти систему, из некоторого допустимого класса систем, которая является наилучшей моделью для заданной системы данных. Под термином "наилучшая" подразумевается, что искомая модель способна генерировать данные, наиболее близкие к заданным данным, или, говоря другими словами, имеет наилучшее приближение (наименьшее несоответствие) к данным. Модель также должна быть самой простой из множества доступных моделей. Кроме критериев простоты и точности, могут применяться и другие критерии.

При решении задачи структурной идентификации

многосвязной системы необходимо выполнить ряд этапов прежде, чем перейти собственно к построению модели структуры. На первом этапе нужно задать исходную систему на основании априорной информации, которая имеется о системе, например, количество переменных, количество учитываемых категорий значений этих переменных и количество измерений, необходимое для формирования системы более высокого уровня - системы данных. На втором этапе исходная система наполняется реальными данными, которые необходимо отквантовать с целью удовлетворения вычислительного ограничения и ограничения данных [7]. Только после этого следует переходить к этапу построения модели порождающей системы, а затем модели структуры системы.

Пусть задана система из n переменных

$$X = \{X_i\}, \quad i = \overline{1, n} \quad (1)$$

с Q_i уровнями квантования для каждой переменной, и, следовательно, последовательность N одновременных измерений X переменных. Тогда можно составить матрицу данных

$$D = X \times N, \quad (2)$$

или матрицу активности, если переменные в матрице упорядочены по времени.

Известно [1,5] несколько подходов к построению модели порождающей системы. Если количество переменных в системе достаточно мало (с точки зрения вычислительных ограничений), то для синтеза модели целесообразно использовать алгоритмы, предложенные в [1], в которых маска выбирается одинаковой для всех переменных X системы. Достоинствами этих алгоритмов являются простота и быстроедействие, а недостатком - очень ограниченное количество переменных в системе. Подход, используемый в алгоритмах [5], позволяет несколько расширить их область применимости по сравнению с алгоритмами [1], поскольку маска в них выбирается отдельно для каждой переменной X и её мощность последовательно наращивается. Недостатком этих алгоритмов является большая длительность выполнения.

При обоих подходах основным ограничивающим фактором является размерность системы, которая не позволяет применить алгоритмы, предложенные в [1,5] для идентификации многосвязных систем. Это ограничение можно обойти, если построение модели порождающей системы разбить на два этапа. На первом этапе проводится замена групп наиболее связанных (по критерию взаимной информации) переменных производными переменными, удаление из рассмотрения наименее значимых для определения структуры системы переменных и сокращение числа различных значений оставшихся про-

изводных переменных (для этого наименее вероятные значения группируются в одну категорию) [8]. На втором этапе - выбор наилучшей маски и собственно построение модели при помощи алгоритмов [1,5].

Рассмотрим каждый из этапов. На первом в качестве критерия для выделения групп наиболее связанных переменных будем использовать определение взаимной информации [9]

$$I(X_i; X_j) = \sum_{(X_i, X_j)} p(X_i, X_j) \log_2 \frac{p(X_i, X_j)}{p(X_i)p(X_j)}. \quad (3)$$

Процедура выделения групп наиболее связанных переменных выглядит следующим образом. Находится пара переменных (X_i, X_j) , для которых $I(X_i; X_j)$ является максимальной среди всех значений взаимной информации, рассчитанных для пар переменных в X . Переменная X_i должна зависеть от X_j с определенным уровнем значимости, что может быть проверено при помощи χ^2 -критерия. На основании того, что $1,3863\overline{NI}(X_i; X_j)$ асимптотически распределено как χ^2 [10], значимость этой величины может быть оценена по таблицам χ^2 -распределения для степени свободы df , которая может быть определена в виде

$$df = [Q_i - 1] \times [Q_j - 1]. \quad (4)$$

Если величина $I(X_i; X_j)$ значима, то переменные (X_i, X_j) можно объединить и получить новую систему с $n - 1$ переменной. Далее весь процесс повторяется с новой системой. Если в результате проведённых вычислений не будет найдено значимых величин $I(X_i; X_j)$, это говорит о том, что процесс выделения групп взаимосвязанных переменных закончен и получена производная система с $m \leq n$ переменными.

В том случае, если число m тоже слишком велико, можно убрать из рассмотрения наименее значимые для определения структуры системы переменные. Это можно сделать на основании величины взаимной информации

$$I(X_i; X - X_i) = \sum_{(X_i, X - X_i)} p(X_i, X - X_i) \log_2 \frac{p(X_i, X - X_i)}{p(X_i)p(X - X_i)}, \quad (5)$$

$$i = \overline{1, m}.$$

Задавшись уровнем значимости величины взаимной информации, находим с помощью χ^2 -критерия её пороговое значение и отбрасываем незначимые переменные. Степень свободы df для χ^2 -распределения может быть определена в виде

$$df = \prod_X [Q_i - 1]. \quad (6)$$

В результате получим систему с $k \leq m \leq n$ переменными.

Благодаря замене групп наиболее связанных переменных новыми переменными можно сократить число различных значений этих переменных. Для этого наименее вероятные значения можно сгруппировать в одну категорию [8]. Например, рассмотрим переменную X_i , принимающую значения $x_1, x_2, \dots, x_k, x_{k+1}, \dots, x_n$, и определим потерю информации при объединении значений x_{k+1}, \dots, x_n в категорию x'_{k+1} . Объединение значений создаёт новую переменную X'_i с $k+1$ значением вместо n , где $(k+1)$ -е значение относится ко множеству x_{k+1}, \dots, x_n значений X_i . Потери информации при замене переменной X_i на переменную X'_i составят

$$\Delta I = I(X_i; X - X_i) - I(X'_i; X - X_i). \quad (7)$$

Если эти потери не значимы, то можно производить группировку значений.

После выполнения первого этапа построения модели порождающей системы получаем систему из k переменных

$$V = \{V_i\}, \quad i = \overline{1, k}, \quad (8)$$

с которой могут работать алгоритмы, предложенные в [1,5].

На втором этапе применение алгоритма [1] позволяет получить набор наилучших масок для каждой из мощностей маски от k до максимально допустимой [1]. Чтобы выбрать среди них маску, наиболее пригодную для построения модели структурированной системы, можно сравнить относительное приращение условной энтропии (по которой они оценивались). Например, для системы из двух переменных при максимальной мощности маски 6 в результате применения алгоритма [1] был получен следующий набор из пяти наилучших масок для мощностей $|M| = \overline{2, 6}$ (пример 3.7, с.148, [1]):

Таблица 1

$ M $	2	3	4	5	6
$H(G/\overline{G})$	3,38	1,88	1,07	0,55	0,41
ΔH	0,38	0,43	0,48	0,25	-

Относительное приращение условной энтропии возрастает с ростом мощности маски и убывает, начиная с маски оптимальной мощности. В данном примере такой маской является маска мощности 5, которая и будет выбрана для построения модели структурированной системы.

В настоящее время существует множество методов построения моделей структурированных систем. В работе [6] дается обобщение и сравнительная характеристика алгоритмов структурной идентификации, разработанных в рамках различных подходов к построению структурной теории систем. На основании этого сравнения делается вывод, что наиболее перспективным выглядит клировский подход [1,11 - 13], поскольку он позволяет использовать одни и те же алгоритмы для идентификации систем различных классов. На базе этого подхода разработан ряд алгоритмов структурной идентификации, использующих решетки уточнения моделей структур с оценением их характеристик по специальным критериям [1 - 5, 11 - 15]. Для решения задачи структурной идентификации многосвязной системы можно использовать, например [1,11,15].

Объединяя вышесказанное, можно сформулировать следующий алгоритм идентификации многосвязных систем:

1. Формирование исходной системы на основании априорной информации;
2. Формирование системы данных на основании вычислительных ограничений [7];
3. Выделение и замена групп наиболее связанных (по критерию взаимной информации) переменных производными переменными [8];
4. Сокращение числа рассматриваемых переменных;
5. Сокращение числа различных значений производных переменных;
6. Выбор наилучшей маски и построение модели порождающей системы [1,5];
7. Построение модели структурированной системы [1, 11,15].

С целью определения качественных характеристик предложенного алгоритма был проведен ряд экспериментов имитационного моделирования. В качестве исследуемой системы использовалась система из четырех переменных с двумя уровнями квантования и максимальной мощностью маски 8 (пример 4.3, с.202, [1]). Процедура реализации одного эксперимента следующая. По исходной модели порождающей системы генерируется выборка данных объемом 4000 значений (для удовлетворения вычислительных ограничений). Затем проводится выделение и замена групп наиболее связанных переменных производными переменными и сокращение числа различных значений производных переменных (таблица перехода сохраняется). Исходная выборка данных преобразуется в выборку данных для производных переменных, для которой выбирается

наилучшая маска и строятся модели порождающей и структурированной систем. Для того чтобы определить насколько модель соответствует исходной системе данных, выполняется обратный переход от структурированной модели к модели порождающей системы, по ней генерируется выборка данных для производных переменных и по ней с использованием таблицы перехода восстанавливается выборка данных для исходных переменных (значения переменных, которые были сгруппированы в одну категорию, восстанавливаются с равной вероятностью).

В результате экспериментов были получены следующие зависимости.

Влияние уровня значимости взаимной информации при группировке переменных на долю восстановленных состояний. С уменьшением количества групп переменных, т.е. с уменьшением уровня значимости взаимной информации между переменными, доля восстановленных состояний убывает, за исключением случая, когда не производится группировка значений производных переменных (при этом восстановление 100%). При уровне значимости взаимной информации от 0.05 до 0.99 доля восстановленных состояний составляла не менее 85%.

Влияние уровня значимости взаимной информации при группировке переменных на информационное расстояние между исходной системой и моделью. Поскольку сравниваемые распределения вероятностей являются произвольными, необходима обобщенная мера информационного расстояния, определенная формулой (9).

$$G(1f, 2f) = D\left(1f, \frac{1f+2f}{2}\right) + D\left(2f, \frac{1f+2f}{2}\right), \quad (9)$$

где $1f$ и $2f$ - произвольные распределения вероятностей, определенные на одном и том же множестве состояний; $(1f+2f)/2$ - распределение вероятностей, полученное взятием среднего для каждой пары соответствующих вероятностей из $1f$ и $2f$; D - информационное расстояние, заданное уравнением

$$D(f, f^h) = \frac{1}{\log_2|C|} \sum_{c \in C} f(c) \log_2 \frac{f(c)}{f^h(c)}, \quad (10)$$

где константа $1/\log_2|C|$ - нормирующий коэффициент, обеспечивающий выполнение соотношения

$$0 \leq D(f, f^h) \leq 1. \quad (11)$$

Информационное расстояние между исходной системой и моделью возрастает с уменьшением уровня значимости взаимной информации между переменными при условии, что производится группировка значений производных переменных. При уровне значимости взаимной информа-

ции от 0.05 до 0.99 информационное расстояние не превысило 0.072.

Влияние уровня значимости потерь взаимной информации при сокращении числа различных значений переменных на долю восстановленных состояний. С ростом порогового значения потерь взаимной информации доля восстановленных состояний уменьшается. При уровне значимости взаимной информации от 0.05 до 0.99 доля восстановленных состояний составляла не менее 85%.

Влияние уровня значимости потерь взаимной информации при сокращении числа различных значений переменных на информационное расстояние между исходной системой и моделью. Информационное расстояние между исходной системой и моделью возрастает с ростом порогового значения потерь взаимной информации. При уровне значимости потерь взаимной информации от 0.05 до 0.99 информационное расстояние не превысило 0.086.

При использовании предложенного алгоритма для идентификации системы по текущим данным необходимо производить периодический контроль пригодности модели. Это можно осуществить при помощи наблюдения за изменением взаимной информации между сгруппированными переменными, методика которых описана в [6]. Если приращение взаимной информации для какой-либо из групп переменных превысит пороговое значение, то необходимо провести изменение модели.

ПЕРЕЧЕНЬ ССЫЛОК

1. Клик, Дж. "Системология: автоматизация решения системных задач". М.: Радио и связь, 1990
2. Broekstra, G. "Some comments on the application of the informational measures of activity arrays". International Journal of General Systems, 3, N1, 1976, pp.43-51
3. Gerardy, R. "Probabilistic finite-state system identification". International Journal of General Systems, 8, N4, 1982, pp.229-242
4. Gerardy, R. "Experiments with some methods for the identification of finite-state system". International Journal of General Systems, 9, 1983, pp.229-242
5. Conant, R. "Detection and analysis of dependency structures". International Journal of General Systems, 7, 1981, pp.81-91
6. Литвинов Е. Н., Лихоузов А. С., Лихоузова Т. А. "Проблемы идентификации нестационарных объектов и анализ методов их решения". "Проблемы создания новых машин и технологий" сборник научных трудов КГПИ, выпуск № 1, Кривой Рог, 1999
7. Литвинов Е. Н., Лихоузов А. С., Лихоузова Т. А. "Модель системы данных в анализе реконструируемости сложных динамических систем", Адаптивные системы автоматического управления, Днепропетровск, 1998
8. Литвинов Е. Н., Лихоузов А. С., Лихоузова Т. А. "Методика идентификации многосвязных систем". "Проблемы создания новых машин и технологий" сборник научных трудов КГПИ, выпуск № 1, Кривой Рог, 1999
9. Шеннон К. "Работы по теории информации и кибернетике". М.: Издательство иностранной литературы, 1963
10. Conant R., Valdes-Perez R. "Information loss due the data quantization in reconstructability analysis". International Journal of General Systems, 9, 1983, pp.235-247
11. Klier, G.J. "Identification of generative structure in empirical data". International Journal of General Systems, 3, N2, 1976, pp.89-104
12. Klier, G.J. "On the representation of activity arrays". International Journal of General Systems, 2, N3, 1975, pp.149-168
13. Cavallo R.E., Klier G.J. "Reconstructability analysis of multidimensional relations". International Journal of General Sys-

tems, 5, N3, 1979, pp.143-171
 14. Gerardy, R. "Probabilistic finite-state system identification". International Journal of General Systems, 8, N4, 1982, pp.229-242
 15. Литвинов Е. Н., Лихоузов А. С., Лихоузова Т. А. "Алгоритм структурной идентификации сложных динамических систем", Адаптивные системы автоматического управления, Днепропетровск, 1998

тем", Адаптивные системы автоматического управления, Днепропетровск, 1998

Надійшла 06.03.2000
 Після доробки 15.03.2000

УДК 681.32:007

НАСТРАИВАЕМЫЕ ЛОГИЧЕСКИЕ СЕТИ НА НЕЙРОПОДОБНЫХ ЭЛЕМЕНТАХ

В. Н. Лопин

Предложен метод синтеза настраиваемых логических сетей на нейроподобных элементах - формальных нейронах. Рассматриваемые сети используют однородную структурную организацию и механизм обработки информации, присущие отдельному базисному элементу. Показывается, что наследование сетью свойств формального нейрона существенно упрощает управление сетью с помощью вектора настройки сети и определяет способ повышения ее надежности. Элементарная однородность и структура связей базисных элементов сетей отвечают требованиям микроэлектронной реализации.

Offered method of syntheses adjusted logical networks on neuro-like elements - formal neurons. Considered network use an uniform structured organization and mechanism of information handling, inherent separate base element. Appears that inheritance by the network the characteristics of formal neuron greatly simplifies a management a network by means of the vector of adjusting to network and defines a way of raising its reliability. Element homogeneity and structure of relationships of base network elements meet the demands microelectronics to realization.

1 ВВЕДЕНИЕ

Современное состояние вычислительной техники характеризуется интенсивным поиском новых принципов обработки информации и новых архитектур вычислительных структур. Анализ известных подходов к проектированию вычислительных высокопроизводительных систем показывает, что их основу составляют те или иные формы параллельной обработки информации, ориентированные на технологию СБИС. Известно, что эффективность решения задач логического управления существенно зависит как от методов, так и от выбранной элементной базы. Прогресс в области микроэлектроники позволил синтезировать управляющие логические устройства в базе программируемых логических интегральных схем (ПЛИС), обеспечивающих высокое быстродействие и отличающихся универсальностью на множестве логических алгоритмов управления. Архитектура ПЛИС представляет собой логику, состоящую из элементов И-ИЛИ-НЕ, триггеров и системы коммутации. Наряду с традиционными подходами к синтезу управляющих логических устройств в базе ПЛИС, ориентированными на использование элементов И-ИЛИ-НЕ, продолжают оставаться актуальными вопросы логического

синтеза устройств в нейросетевом базисе [1-4]. Это объясняется, прежде всего, эффективностью обработки информации такими устройствами, их функциональной гибкостью и надежностью. Несмотря на большое количество проведенных исследований этой эффективности, требуется ее дальнейшее системное изучение. К настоящему времени отсутствуют полные сведения о закономерностях взаимосвязи структурной избыточности нейронных сетей с их надежностью и функциональной мощностью. Знания этих закономерностей позволили бы рационально использовать структурную избыточность при создании надежных многофункциональных управляющих устройств.

Настоящая работа рассматривает вопросы синтеза настраиваемых логических сетей, использующих такие закономерности, установленные при выполнении предыдущих исследований в работе [5]. Можно утверждать, что рассматриваемый класс сетей наследует свойства своих базисных элементов - формальных нейронов.

2 СИНТЕЗ СТРУКТУРЫ НАСТРАИВАЕМОЙ ЛОГИЧЕСКОЙ СЕТИ

Рассматриваемый метод синтеза предполагает использование в качестве базисных элементов сети формальных нейронов, исследованных в [5]. Суть такого подхода состоит в следующем.

Очевидно, что любая булева функция многих переменных $y = f(x_1, \dots, x_n)$ всегда может быть представлена некоторой системой из 2^n остаточных функций, являющихся булевыми функциями одной переменной. В этом случае система остаточных функций определится как:

$$\left. \begin{aligned} y_0 &= f_0(x_1 = 0, \dots, x_{n-1} = 0, x_n); \\ y_1 &= f_1(x_1 = 0, \dots, x_{n-1} = 0, x_n); \\ &\dots \dots \dots \\ y_{2^{n-1}-1} &= f_{2^{n-1}-1}(x_1 = 0, \dots, x_{n-1} = 0, x_n). \end{aligned} \right\} \quad (1)$$

Известно, что любая булева функция одной переменной

ной всегда является пороговой. Следовательно, любой $y_i \in D$, $D = \{y_0, y_1, \dots, y_{2^{n-1}-1}\}$ этой системы можно поставить в соответствие некоторую пару вещественных чисел $(B_i, T_i) \in R$. Таким образом, можно записать

$$y_i = \text{Sign}\{B_i x_n + T_i\}, \quad i \in \{0, 1, \dots, 2^{n-1}-1\}. \quad (2)$$

С другой стороны, как следует из (1), любой $y_i \in D$, $D = \{y_0, y_1, \dots, y_{2^{n-1}-1}\}$ можно поставить в соответствие некоторый двоичный набор $(x_1, x_2, \dots, x_{n-1})_i$. Следовательно, если в выражении (2) ввести пару вещественных функций $B^n(x_1, \dots, x_{n-1})$, $T^n(x_1, \dots, x_{n-1})$ определенных соответственно на множествах

$$\begin{aligned} M_B &= \{(B_i): i \in \{0, 1, \dots, 2^{n-1}-1\}\} \\ M_T &= \{(T_i): i \in \{0, 1, \dots, 2^{n-1}-1\}\} \end{aligned}$$

то для любой $y = f(x_1, \dots, x_n)$ можно записать выражение:

$$\begin{aligned} f(x_1, \dots, x_n) &= \\ &= \text{Sign}\{B^{n-1}(x_1, \dots, x_{n-1})x_n + T^{n-1}(x_1, \dots, x_{n-1})\}, \end{aligned} \quad (3)$$

Без потери общности, можно рассматривать функции $B^{n-1}(x_1, \dots, x_{n-1})$, $T^{n-1}(x_1, \dots, x_{n-1})$ в классе целочисленных функций. В этом случае, каждая из этих функций представима в виде некоторой алгебраической суммы логических функций $\alpha_i^{n-1}(x_1, x_2, \dots, x_{n-1})$ в базисе $\{-1, 1\}$, т.е.

$$\left. \begin{aligned} B^{n-1}(x_1, \dots, x_{n-1}) &= \sum_{\substack{i=1 \\ m^{n-1}+1+k^{n-1}}}^{m^{n-1}} \alpha_i^{n-1}(x_1, \dots, x_{n-1}) \\ T^{n-1}(x_1, \dots, x_{n-1}) &= \sum_{i=m^{n-1}+1} \alpha_i^{n-1}(x_1, \dots, x_{n-1}) \end{aligned} \right\}. \quad (4)$$

Целочисленные функции $B^{n-1}(x_1, \dots, x_{n-1})$, $T^{n-1}(x_1, \dots, x_{n-1})$ могут рассматриваться как функции состояний формального нейрона. Эта интерпретация определяет тождественность логических функций, участвующих в формировании $B^{n-1}(x_1, \dots, x_{n-1})$, $T^{n-1}(x_1, \dots, x_{n-1})$, т.е. выполняются условия:

$$\left. \begin{aligned} \alpha_i^{n-1} &= \alpha_j^{n-1} \forall i, j \in \{1, 2, \dots, m^{n-1}\} \\ \alpha_i^{n-1} &= \alpha_j^{n-1} \forall i, j \in \{m^{n-1}+1, \dots, m^{n-1}+1+k^{n-1}\} \end{aligned} \right\}. \quad (5)$$

Таким образом, выражение (3) может быть представлено в виде :

$$\begin{aligned} f(x_1, \dots, x_n) &= \text{Sign} \left\{ \sum_{i=1}^{m^{n-1}} \alpha_i^{n-1}(x_1, \dots, x_{n-1})x_n + \right. \\ &\left. + \sum_{i=m^{n-1}+1}^{m^{n-1}+1+k^{n-1}} \alpha_i^{n-1}(x_1, \dots, x_{n-1}) \right\}. \end{aligned} \quad (6)$$

Условия (5) допускают следующую упрощенную форму разложения (4):

$$\left. \begin{aligned} B^{n-1}(x_1, \dots, x_{n-1}) &= m^{n-1} \alpha_1^{n-1}(x_1, \dots, x_{n-1}) \\ T^{n-1}(x_1, \dots, x_{n-1}) &= k^{n-1} \alpha_2^{n-1}(x_1, \dots, x_{n-1}) \end{aligned} \right\}. \quad (7)$$

Эта форма разложения определяет целочисленную функцию $B^{n-1}(x_1, \dots, x_{n-1})$ в виде суммы m^{n-1} экземпляров функции $\alpha_1^{n-1}(x_1, \dots, x_{n-1})$, а функцию $T^{n-1}(x_1, \dots, x_{n-1})$ в виде суммы k^{n-1} экземпляров функции $\alpha_2^{n-1}(x_1, \dots, x_{n-1})$. В этом случае, выражение (3) может быть представлено как:

$$\begin{aligned} f(x_1, \dots, x_n) &= \\ &= \text{Sign}\{m^{n-1} \alpha_1^{n-1}(x_1, \dots, x_{n-1})x_n + \\ &+ k^{n-1} \alpha_2^{n-1}(x_1, \dots, x_{n-1})\}. \end{aligned} \quad (8)$$

Очевидно, любая булева функция $\alpha_i^{n-1}(x_1, \dots, x_{n-1})$ в выражении (6) может быть представлена через булевы функции $\alpha_i^{n-2}(x_1, \dots, x_{n-2})$, т.е. :

$$\begin{aligned} \alpha_i^{n-1}(x_1, \dots, x_{n-1}) &= \text{Sign} \left\{ \sum_{i=1}^{m^{n-2}} \alpha_i^{n-2}(x_1, \dots, x_{n-2})x_{n-1} + \right. \\ &\left. + \sum_{i=m^{n-2}+1}^{m^{n-2}+k^{n-2}} \alpha_i^{n-2}(x_1, \dots, x_{n-2}) \right\}. \end{aligned} \quad (9)$$

Используя последовательное разложение булевых функций, на n -ом шаге можно получить систему булевых функций одной переменной:

$$\left. \begin{aligned} \alpha_1^1(x_1) &= \text{Sign}\{m^0 \alpha_1^0 x_1 + k^0 \alpha_2^0\}; \\ \alpha_2^1(x_1) &= \text{Sign}\{m^0 \alpha_3^0 x_1 + k^0 \alpha_4^0\}; \\ &\dots \\ \alpha_{2^{n-1}-1}^1(x_1) &= \text{Sign}\{m^0 \alpha_{2^{n-3}}^0 x_1 + k^0 \alpha_{2^{n-2}}^0\}; \\ \alpha_{2^n-1}^1(x_1) &= \text{Sign}\{m^0 \alpha_{2^{n-1}}^0 x_1 + k^0 \alpha_{2^n}^0\}. \end{aligned} \right\} \quad (10)$$

Полученная на этом шаге упорядоченная последовательность двоичных констант α_i^0 , $i \in \{1, 2, \dots, 2^n\}$ определяет вектор настройки сети:

$$\bar{q} = [\alpha_1^0, \alpha_2^0, \dots, \alpha_{2^n}^0]. \quad (11)$$

Таким образом, рассмотренная выше процедура разложения булевой функции $y = f(x_1, \dots, x_n)$ однозначно задает n -уровневую иерархическую структуру настраиваемой логической сети.

3 ОЦЕНКИ СЛОЖНОСТИ СЕТИ

Для характеристики распределения базисных элементов сети по ее уровням целесообразно ввести в рассмотрение матрицу состояний сети:

$$MC = \begin{vmatrix} m^{n-1} & k^{n-1} \\ m^{n-2} & k^{n-2} \\ \cdot & \cdot \\ \cdot & \cdot \\ m^1 & k^1 \end{vmatrix}. \quad (12)$$

Матрица состояний сети MC , по аналогии с состоянием формального нейрона, отражает распределение настроенных входов для базисного элемента уровня $i \in \{1, 2, \dots, n-1\}$ иерархической нейронной сети. Анализ, проведенный в работе [5], позволяет утверждать, что для обеспечения максимальной надежности сети необходимо выполнение условия:

$$MC = \begin{vmatrix} m^{*n-1} & k^{*n-1} \\ m^{*n-2} & k^{*n-2} \\ \cdot & \cdot \\ \cdot & \cdot \\ m^{*1} & k^{*1} \end{vmatrix}, \quad (13)$$

где (m^{*j}, k^{*j}) - начальное состояние базисного элемента j уровня сети.

Таким образом, для сети может быть поставлена задача о нахождении матрицы [13], обеспечивающей максимальную надежность сети при ограничении на сложность и функциональную мощность сети. В общем случае эта задача имеет неоднозначное решение, являющееся предметом отдельного исследования и не входящее в материал настоящей работы.

Для характеристики распределенной сложности сети вводится в рассмотрение вектор сложности сети:

$$\overline{MS} = \begin{vmatrix} S^{n-1} \\ S^{n-2} \\ \cdot \\ \cdot \\ S^1 \end{vmatrix}, \text{ где } S^j = m^{*j} + k^{*j}. \quad (14)$$

В этом случае, общая сложность сети SN , определяемая общим количеством базисных элементов в n уровнях

сети, может быть найдена с помощью выражения:

$$SN = 1 + S^{n-1} + S^{n-1}S^{n-2} + \dots + S^{n-1}S^{n-2}S^{n-3} \dots S^1. \quad (15)$$

Из [5] следует, что минимальной формой разложения целочисленных функций (7) является:

$$\left. \begin{aligned} B^{n-1}(x_1, \dots, x_{n-1}) &= 2\alpha_1^{n-1}(x_1, \dots, x_{n-1}) \\ T^{n-1}(x_1, \dots, x_{n-1}) &= \alpha_2^{n-1}(x_1, \dots, x_{n-1}) \end{aligned} \right\}. \quad (16)$$

В этом случае, выражение (8) может быть представлено в виде:

$$\begin{aligned} f(x_1, \dots, x_n) &= \\ &= \text{Sign}\{2\alpha_1^{n-1}(x_1, \dots, x_{n-1})x_n + \alpha_2^{n-1}(x_1, \dots, x_{n-1})\}. \end{aligned} \quad (17)$$

Применяя последовательное разложение булевых функций, на n -ом шаге можно получить следующую систему булевых функций одной переменной:

$$\left. \begin{aligned} \alpha_1^1(x_1) &= \text{Sign}\{2\alpha_1^0x_1 + \alpha_2^0\}; \\ \alpha_2^1(x_1) &= \text{Sign}\{2\alpha_3^0x_1 + \alpha_4^0\}; \\ &\dots \\ \alpha_{2^{n-1}-1}^1(x_1) &= \text{Sign}\{2\alpha_{2^{n-3}}^0x_1 + \alpha_{2^{n-2}}^0\}; \\ \alpha_{2^n-1}^1(x_1) &= \text{Sign}\{2\alpha_{2^{n-1}}^0x_1 + \alpha_{2^n}^0\}. \end{aligned} \right\} \quad (18)$$

В этом случае вектор настройки сети q определится согласно выражению (11).

Из анализа формального нейрона [5] следует, что начальное состояние базисного элемента

$$E_o^* = (n^*, k^*) = (2, 1) \quad (19)$$

определяет как минимальную сложность нейрона $S = 3$, так и минимальную сложность n -уровневой сети определяемую, согласно (15), выражением

$$SN = 1 + \sum_{i=1}^{n-1} 3^i. \quad (20)$$

Оценка (20), при выполнении условия (19), может быть еще уменьшена до значения

$$SN = 1 + \sum_{i=1}^{n-1} 2^i \quad (21)$$

без снижения уровня надежности, соответствующего сложности (20). Действительно, в последовательном разложении (16) можно использовать лишь один экземпляр функции $\alpha_1^n(\cdot)$, умноженной на постоянный коэффициент 2. Этот подход можно применить для всех последовательных разложений булевых функций при син-

тезе минимальных конфигураций. В этом случае задается топология иерархической сети минимальной сложности. Как следует из рассмотренного, вектор настройки сети $\bar{q} = |\alpha_1^0, \alpha_2^0, \dots, \alpha_{2^n}^0|$ определяется лишь исходной булевой функцией $y = f(x_1, \dots, x_n)$ и не зависит от матрицы состояний MS . Это позволяет использовать вектор настройки сети минимальной конфигурации и для избыточной сети. В таких сетях, для сокращения количества внешних настроечных входов целесообразно в первом уровне использовать компоненты вектора \bar{q} в качестве единичных экземпляров настроечных констант, умноженных на соответствующие значения m^{*0} и k^{*0} .

3 ВЫВОДЫ

Рассмотренный метод синтеза позволяет проектировать многофункциональные управляющие устройства, отличающиеся оперативностью настройки и возможностью работы в реальном масштабе времени. Особенностью организации настраиваемых логических сетей, составляющих основу таких устройств, является насле-

дование структуры и свойств базисных элементов. Это позволяет использовать способ введения структурной избыточности, исследованный в [5], для построения надежных многофункциональных управляющих логических устройств, реализованных в виде СБИС. Предполагается, что вопросы синтеза таких избыточных перестраиваемых логических сетей будут рассмотрены в отдельных публикациях.

ПЕРЕЧЕНЬ ССЫЛОК

1. *Галушкин А.И., Кирсанов Д.В.* Цифровые нейрочипы: (специализированные цифровые БИС для нейрокомпьютеров) // Зарубежная радиоэлектроника, 1999, №1, С. 17-37.
2. *Алексеев А.А., Дорогов А.Ю., Пименов В.Е.* Структурный синтез нейронных сетей для СБИС реализации // Нейроинформатика и ее приложения: Тез. докл. 7 Всерос. семинара, 1-3 окт. 1999 г. - Красноярск: КГТУ, 1999. - С. 12.
3. *Лавренюк А.Ф.* Нейроподобные модули нейросетевых процессоров // Нейроинформатика и ее приложения: Тез. докл. 7 Всерос. семинара, 1-3 окт. 1999 г. - Красноярск: КГТУ, 1999. - С. 89.
4. *Лопин В.Н.* Исследование одного класса нейронных сетей с конвейерной обработкой информации // Автоматика и вычислительная техника, 1999, №5, С.80-84.
5. *Лопин В.Н.* Анализ формально-логической модели нейрона с помощью дискретного пространства состояний // Автоматика и вычислительная техника, 1999, №6, С.54-58.

Надійшла 10.03.2000

Після доробки 21.03.2000

УДК 681.3.06

ИСПОЛЬЗОВАНИЕ JAVA-ТЕХНОЛОГИЙ В ИНТЕРНЕТ-ПРИЛОЖЕНИЯХ

С. А. Онищенко

Рассматриваются Java-технологии как инструмент для создания и поддержки Интернет-проектов. Основное внимание уделяется использованию Java в серверной части программных продуктов. Приведен пример создания Интернет-представительства на основе Java-сервера JServ.

Розглядаються Java-технології як інструмент для побудови й підтримки Інтернет-проектів. Основна увага приділяється використанню Java у серверній частині програмних продуктів. Наведено приклад створення Інтернет-представництва на основі Java-серверу JServ.

The article considers Java technologies as a tool for building and maintaining Internet-based projects. Major impact is done on using Java in server-side software. The article provides an example of Web-site built on Apache JServ Java server.

ВВЕДЕНИЕ

По прогнозам специалистов при существующих темпах развития Интернет рискует обойти по популярности не только печатные издания, но и телевидение. Создание своего представительства в Интернет становится правилом хорошего тона. К сожалению, такой аспект как поддержание веб-проектов, как правило, упускается из виду. Для обеспечения качественного контента (напол-

нения) веб-сайтов используются различные технологии динамического создания страниц - CGI-скрипты, PHP, SSI и т.п. Зачастую модули динамического создания страниц, в свою очередь, обращаются к различным репозиториям (базы данных, базы знаний, экспертные системы). Как утверждают специалисты компании Sun [1], технология Java пригодна не только для создания апплетов, которые исполняются на клиентской стороне (в браузере обозревателя). Гибкий и простой интерфейс сервлетов (Servlet API) позволяет использовать всю мощь Java в серверных приложениях. Все преимущества Java, которые выделяют этот язык среди прочих средств программирования, свидетельствуют в пользу Java как инструмента для создания серверов.

ПОСТАНОВКА ЗАДАЧИ

Представляет интерес проблема реализации программного комплекса для поддержки веб-представительства, который бы одновременно гарантировал высокую надежность работы веб-сайта и простоту его администрирования. Немаловажным аспектом является наличие

базы данных, возможность быстрого поиска и схема разграничения прав доступа.

Самым трудоемким этапом разработчика интернет-приложений является реализация шлюза между интерфейсом клиента (клиентом, как правило, выступает браузер: Netscape Navigator, Opera, Lynx, Internet Explorer) и долговременным хранилищем данных на сервере (файлы, базы данных, базы знаний). В последнее время было разработано несколько схем и методик создания веб-серверов на базе Perl, PHP, ASP, SSI. Все эти языки и методики используют стандартный CGI-интерфейс (Common Gateway Interface) для получения параметров от клиента при динамическом формировании HTML-страниц. Одним из наиболее перспективных серверных средств на данный момент являются Java-сервлеты.

JAVA-СЕРВЛЕТЫ

Сервер, написанный на Java, обречен на высокую надежность и безопасность. Возможности многозадачности, заложенные в Java, повышают производительность серверов, а независимость от платформы позволяет использовать одни и те же модули на любых компьютерах и операционных системах, поддерживающих Java. Это позволяет создавать многоуровневые корпоративные приложения в рамках сетей любой сложности и работающие под управлением любой операционной системы [2]. Ключевыми компонентами Java-сервера являются сервлеты.

Сервлет - это независимая от платформы и протокола серверная компонента, написанная на Java, которая динамически расширяет возможности Java-сервера. Основное назначение сервлетов - это предоставление защищенного доступа к данным через Интернет. Сервлеты позволяют просматривать и модифицировать данные посредством динамически генерируемых HTML-страниц. Так как сервлеты исполняются на сервере, они не нуждаются в графическом интерфейсе, что значительно облегчает задачу программиста. В отличие от апплетов, сервлеты не передаются по сети для выполнения на клиентской машине, что значительно облегчает загрузку каналов связи.

Программы-клиенты, которые обращаются к сервлетам, могут отличаться по сложности. Клиентом может выступать как простая HTML форма, так и Java-апплет. На рис.1 представлена схема взаимодействия HTTP-клиентов (апплеты, HTML-формы) с Java-сервером, в контексте которого выполняются сервлеты.

С точки зрения разработчика сервлет - это Java-класс. Сервлет реализует функции интерфейса Servlet API [1], как правило, расширяя класс GenericServlet или его модификацию - класс HttpServlet. Простейший сервлет переопределяет один метод класса GenericServlet - service:

```
import java.servlet.*;
public class MyServlet extends GenericServlet {
    public void service (
        ServletRequest request,
        ServletResponse response
    ) throws ServletException, IOException
    {
        ...
    }
    ...
}
```

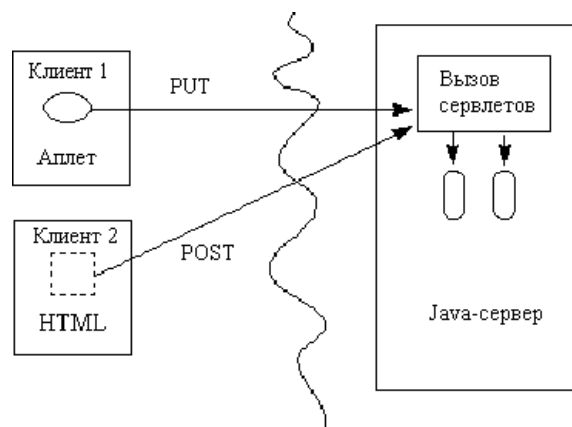


Рисунок 1 - Взаимодействие HTTP-клиентов с сервлетами

Если в качестве базового класса при создании сервлета используется HttpServlet, то необходимо определить реакцию сервлета на HTTP запросы POST, GET, HEAD при помощи переопределения функций doPost(), doGet() и doHead(), соответственно.

Сервлеты всегда загружаются в оперативную память динамически. Администратор сервера также имеет возможность принудительно загружать и инициализировать определенные сервлеты при запуске Java-сервера (startup). Сервлеты загружаются в память машины при помощи обычного загрузчика классов Java. При этом сервлет может располагаться как на локальной файловой системе, так и в удаленных каталогах, например, в каталоге `http://some.host.com/servlets`. Этим достигается гибкость системной архитектуры и обеспечивается распределение сервисов в сети.

ЗАДАЧИ, РЕШАЕМЫЕ ПРИ ПОМОЩИ СЕРВЛЕТОВ

Сервлет, как правило, обрабатывает информацию, которая поступает на сервер по протоколу HTTP в результате отправки данных какой-либо HTML формы. Так как сервлеты могут одновременно обрабатывать несколько запросов, происходит синхронизация запро-

сов. Эта возможность широко используется для создания многопользовательских приложений, таких как онлайн-конференции и Интернет-магазины. При помощи сервлетов можно создать группу активных элементов, которые бы разделяли нагрузку между собой. Один сервлет может перенаправлять запросы разным серверам. Эта методика применяется для балансирования нагрузки между несколькими серверами, которые отображают один и тот же контент. Сервлеты также могут использоваться для распределения сервисов между несколькими серверами, маршрутизируя запросы согласно типу задачи или организационной структуре сервера, сети или предприятия.

Одним из преимуществ сервлетов является то, что они не требуют создания нового системного процесса для каждого приходящего запроса. Практически в любой программной среде множество сервлетов исполняются параллельно как сервер в контексте единого системного процесса (процесс Java-сервера). По этому критерию использование сервлетов является более предпочтительным, чем использование CGI или Fast-CGI (рис.2).

При использовании методики CGI (рис.2а) каждый запрос, приходящий на сервер из мира, обрабатывается отдельным системным процессом, который порождается главным процессом (веб-сервер). Когда применяется технология Fast-CGI (рис.2б), все запросы для CGI-модулей перенаправляются в отдельный системный процесс, который их обрабатывает. Если запросы не поступают, то этот процесс завершается (разрушается системой). С приходом очередного запроса процесс инициализируется (создается) вновь. Операции создания и разрушения процессов, как известно, отбирают время и системные ресурсы. Использование Java-сервера (рис.2в) позволяет избежать этих затрат, так как в системе всегда присутствует процесс, готовый в любую минуту принять запрос и обработать его. Этот подход имеет еще одно преимущество - любой операционной системе удобнее следить за одним процессом, чем за многими. В связи с этим повышается надежность работы сервера в целом.

Java-сервер позволяет реализовать трехуровневую информационную модель. В отличие от классической двухуровневой модели, когда программа-клиент "общается" с одним сервером, в трехуровневой модели серверы могут "общаться" друг с другом.

Первым уровнем трехуровневой Интернет-модели являются браузеры (желательно с поддержкой Java). Сложные задачи, требующие расширенные возможности интерфейса, решаются при помощи Java-апплетов. Более тривиальные - при помощи обыкновенных HTML форм.

Вторым уровнем являются сервлеты, которые реализуют бизнес-правила и логику приложений.

Третий уровень - это репозитории данных. Доступ к элементам (серверам) третьего уровня осуществляется

через интерфейсы реляционных баз данных (JDBC, ODBC) или посредством других интерфейсов и протоколов, таких как ONC RPC, DCE RPC или CORBA/ПОР (рис.3).

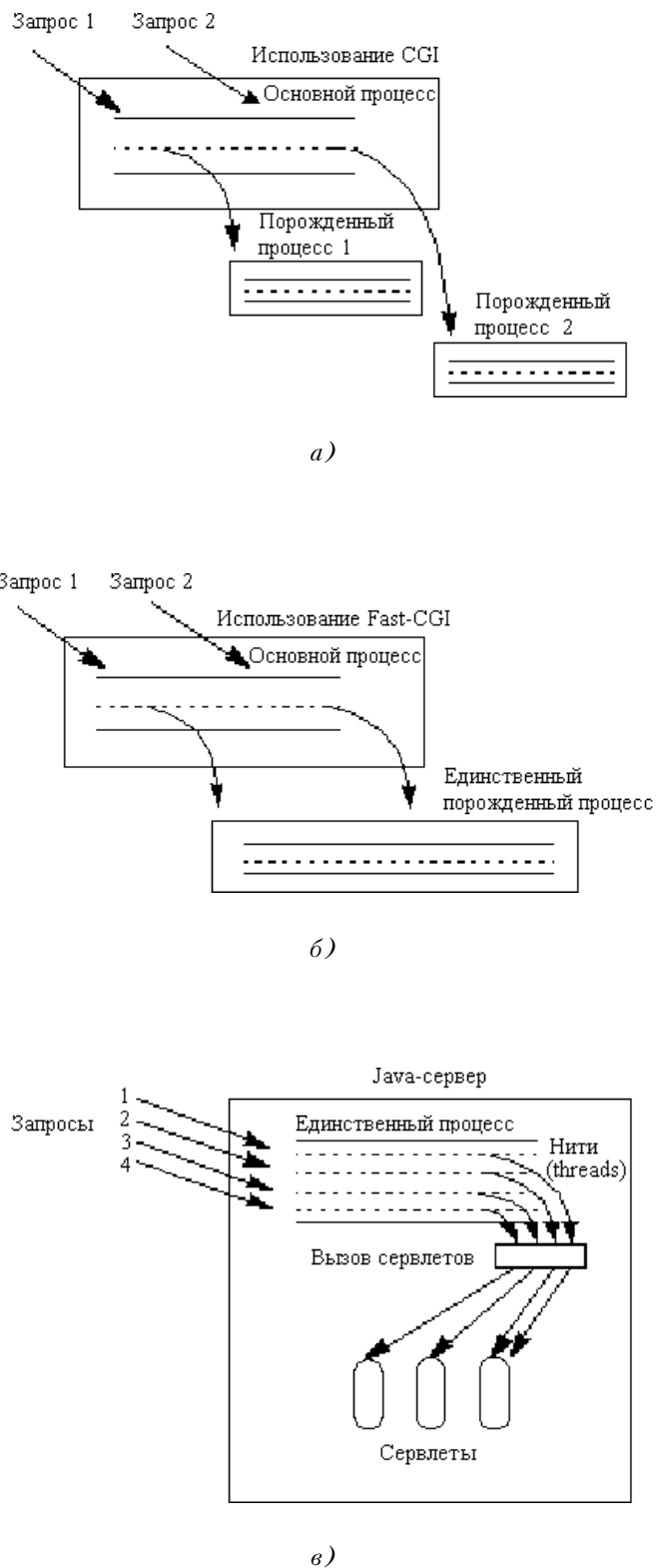


Рисунок 2 - Сравнение работы сервлетов с работой CGI и Fast-CGI

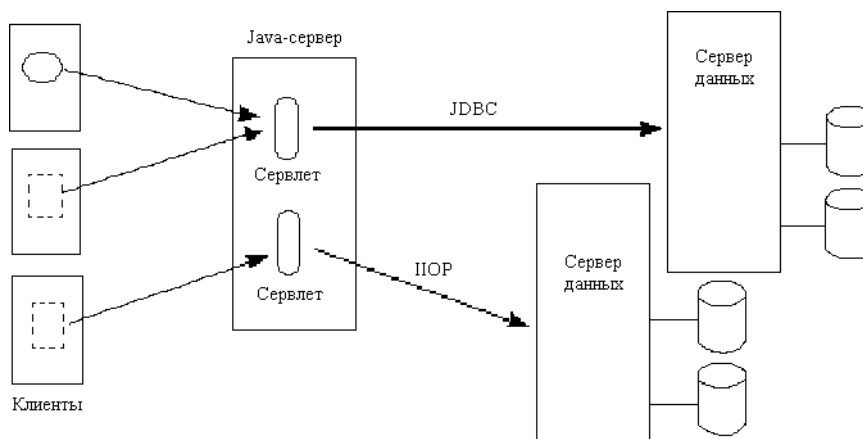


Рисунок 3 - Трехуровневая модель с использованием Java-сервера

Интерфейс сервлетов описывается в стандарте Java Servlet API [1] от компании SUN. Этот интерфейс поддерживается несколькими веб-серверами, среди которых самый популярный в мире [3] веб-сервер Apache.

Apache JServ - это сервер, обеспечивающий работу сервлетов (servlet engine), полностью написанный на Java и полностью совместимый с интерфейсом JavaSoft Java Servlet APIs 2.0 [4]. Благодаря своей портативности сервер JServ 1.0 способен функционировать на всех виртуальных Java-машинах (JVM) версии 1.1 и выше и обслуживать все сервлеты, которые поддерживают интерфейс Java Servlet API версии 2.0. В целях минимизации зависимости от какой-либо определенной платформы, JServ реализован в виде отдельного сервера (процесса-демона), который общается с другими серверами посредством специального сетевого протокола AJP [5]. Apache JServ поддерживает несколько модулей расширения, среди которых Apache JSSI, Cocoon, GNU Java Server Pages, WebMacro Java Servlet Toolkit. Apache JServ способен работать под управлением различных операционных систем.

ПРИНЦИПЫ ПОСТРОЕНИЯ И РАБОТЫ СИСТЕМЫ

В качестве примера веб-сайта, разработанного с применением Java-технологий, предлагается рассмотреть сайт журнала "Радиоэлектроника. Информатика. Управление", который располагается по адресу <http://www.zstu.edu.ua/RIC/>. В качестве серверной ОС был выбран клон UNIX - FreeBSD 4.0-STABLE. В качестве веб-сервера использовался Apache/1.3.6 (Unix) rus/PL28.16. В качестве СУБД - MySQL server version 3.22.25. Все эти программные продукты (включая саму ОС FreeBSD) распространяются бесплатно в исходных текстах по лицензии GPL [6]. Широкое использование этих продуктов профессионалами во всем мире гарантирует их надежность и обеспечивает качественную поддержку.

Так как на сайте присутствует информация на трех языках (украинский, русский, английский), кодовой страницей для хранения и обработки данных на сервере была выбрана KOI8-U [7]. Кодировка, в которой данные отдаются клиенту, выбирается динамически средствами расширения веб-сервера Apache - RA (Russian Apache), исходя из заголовков HTTP-запроса и типа агента (браузера) [8]. Задача динамического формирования веб-страниц была реализована при помощи Java-сервера ApacheJServ/1.0, JDK 1.1.8, JSDK 2.0, JSSI 1.0 [4]. Ограничение доступа к данным было реализовано средствами веб-сервера Apache (файл ограничений access.conf и утилита httpasswd).

Веб-сайт представляет собой набор .jhtml файлов, которые являются прототипами формируемых HTML-страниц. Получая запрос на файл с таким расширением, веб-сервер (в данном случае Apache) формирует свой запрос и посылает его Java-серверу (JServ) по протоколу AJP. Java-сервер инициализирует сервлет org.apache.ssi.SSIServlet, который просматривает исходный код jhtml-файла и исполняет сервлеты, запрашиваемые тэгами `<SERVLET> ... </SERVLET>`. Результат работы вызываемого сервлета подставляется на место его вызова в файл .jhtml. Во всем остальном .jhtml файл сходен с обыкновенным .html файлом, т.е. содержит чистый HTML-код. Сформированная страница отдается веб-серверу, а затем от него к клиенту.

Поведение запрашиваемых в теле .jhtml файла сервлетов зависит от параметров HTTP-запроса, передаваемых от клиента (из браузера), а также от параметров, заданных при помощи директив `<PARAM name='...' value='...'>` в теле исходного файла. Например:

```
<SERVLET code='MagazineSSI.class'>
<PARAM name='mode' value='issue'>
<PARAM name='text' value='${date}'>
<PARAM name='lang' value='rus'>
</SERVLET>
```

В рассматриваемом примере используются два модуля (два сервлета) - MagazineSSI и MagazineAdminSSI, которые обеспечивают динамику на страницах сайта. Упрощенные блок-схемы модулей приведены на рис.4 и 5, соответственно.

Функции простой выборки и поиска (select) и администрирования (update, delete, insert) разнесены по разным модулям из соображений безопасности. Подключение к базе данных из модуля MagazineSSI не дает возможности модифицировать данные. Модуль MagazineAdminSSI, наоборот, обеспечивает практически полный доступ к статьям. Подключение к базе данных (MySQL) осуществляется при помощи драйвера JDBC twz1 версии 1.0.4-GA. В зависимости от параметров, передаваемых сервлету, конструируется SQL-запрос. Он обрабатывается MySQL-сервером, его результаты возвращаются в сервлет, формируются и отдаются в качестве HTML-страницы клиенту.

Модули MagazineSSI и MagazineAdminSSI принимают параметр lang, который определяет язык вывода - украинский, русский или английский. Формат вывода, как правило, определяется параметром text.

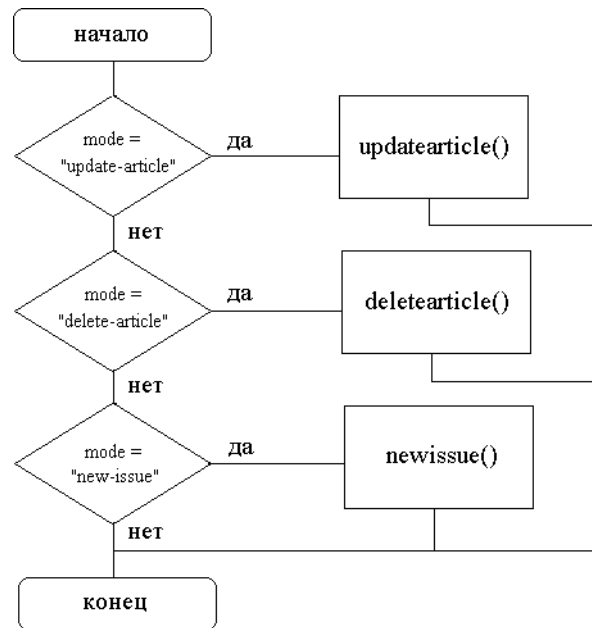


Рисунок 5 - Блок-схема модуля MagazineAdminSSI

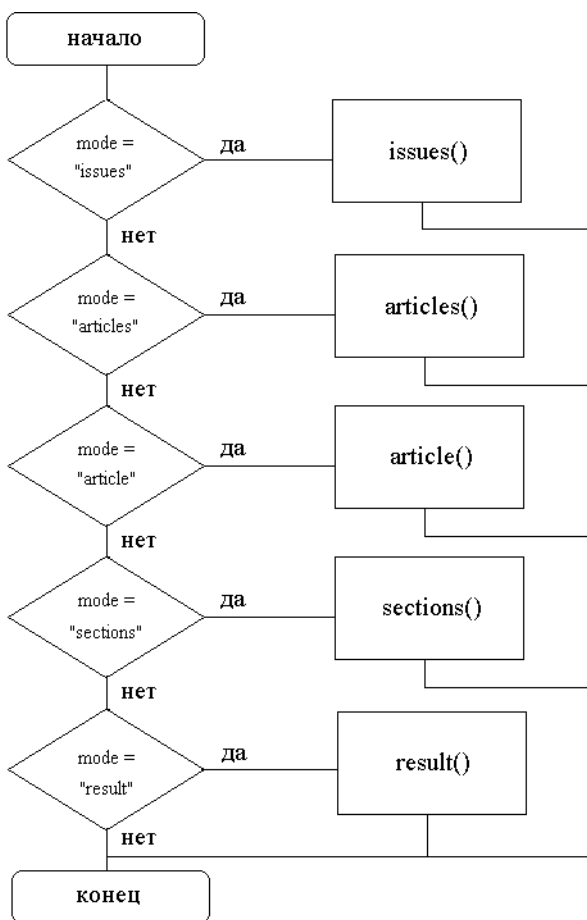


Рисунок 4 - Блок-схема модуля MagazineSSI

Администрирование веб-сайта (модификация данных, ввод новых статей) осуществляется через Интернет. В подкаталоге admin размещаются .jhtml файлы, которые вызывают сервлет MagazineAdminSSI. В модуле MagazineAdminSSI реализованы функции создания нового номера журнала в базе, добавления, редактирования и удаления статей. Доступ к подкаталогу admin ограничивается средствами самого веб-сервера. Каталог может быть доступен только определенному пользователю или группе пользователей (по паролю), или только с определенных адресов или подсетей (по IP-адресу).

На настоящий момент рассматриваемая система работала 6 месяцев. Судя по визуальным наблюдениям и файлам-протоколам, нарушений в работе не было. В целях тестирования на отказ на веб-сайт журнала была проведена небольшая DOS-атака (denial-of-service attack), что, однако, не повлияло на стабильность работы сервера.

РЕЗУЛЬТАТЫ И ВЫВОДЫ

Описанная выше методика позволяет создавать и поддерживать надежные Интернет-приложения любого масштаба и сложности. При помощи Java-технологий сравнительно легко создаются многоуровневые системы, которые успешно интегрируются в Интернет. В качестве примера рассматривалась полнофункциональная система поддержки веб-сайта журнала "Радиоэлектроника. Информатика. Управление", которая за 6 месяцев работы не дала сбой. Использование Java-технологий, удобный унифицированный интерфейс для администрирования, а

также гибкая настройка позволяют применять разра- ботанную систему для поддержки веб-представительства практически любого печатного издания.

ПЕРЕЧЕНЬ ССЫЛОК

1. The Java Servlet API (<http://java.sun.com/products/servlet/whitepaper.html>).
2. Морган М. Java 2. Руководство разработчика.: Пер. с англ. - М.: Издательский дом "Вильямс", 2000. - 720 с.

3. The Netcraft Web Server Survey (<http://www.netcraft.com/survey/>).
4. The Apache JServ Project (<http://java.apache.org/jserv/>).
5. Apache JServ Protocol Version 1.1 (AJPv1.1) (<http://java.apache.org/jserv/protocol/AJPv11.html>).
6. GNU General Public License (GPL) (<http://www.gnu.org/copyleft/gpl.html>).
7. RFC2319 - Ukrainian Character Set KOI8-U (<ftp://ftp.isi.edu/in-notes/rfc2319.txt>)
8. Russian Apache: Как это работает (<http://apache.lexa.ru/internals.html>).

Надійшла 23.02.2000

Після доробки 22.03.2000

УДК 681.32:007

БЫСТРОЕ ДИФФЕРЕНЦИРОВАНИЕ СЛОЖНЫХ ФУНКЦИЙ МНОГИХ ПЕРЕМЕННЫХ

М. Ю. Сенашова

Описаны графы вычисления сложных функций многих переменных, изложен принцип двойственности для вычисления градиента сложной функции и градиента от функционалов градиента, выделен класс таких функций, для которых вычисление градиента наиболее просто (он включает нейронные сети, но не только - существуют экстремально простые схемы более общего вида).

The graphs of calculation of composite functions of many variables are described. Principle of duality for calculation of gradient and second derivative of composite function is stated. Class of functions with the most simple calculation of gradient is determined. (Besides the neural networks it includes the extremely simple circuits of a more general kind).

ВВЕДЕНИЕ

Неверно широко распространенное мнение, что вычислительные затраты для градиента функции n переменных в n раз превышают затраты на вычисление значения самой функции. В действительности, для всех функций многих переменных, встречающихся на практике, необходимые вычислительные затраты на поиск их градиента всего лишь в два-три раза превосходят затраты на вычисление одного значения функции.

Поиск $\text{grad}F$ удобно представить как некоторый двойственный процесс над структурой вычисления F . Оказывается, что если представить F как сложную функцию, являющуюся суперпозицией функций малого числа переменных, и аккуратно воспользоваться правилом дифференцирования сложной функции, не производя по дороге лишних вычислений и сохраняя полезные промежуточные результаты, то вычисление всей совокупности $\frac{\partial F}{\partial x_i}$ ($i = \overline{1, \dots, n}$) немногим сложнее, чем одной из этих функций - они все собраны из одинаковых блоков. Частным случаем такого рационального подхода к вычислению градиента является метод обратного рас-

пространения ошибки [1].

Эти рассуждения справедливы и для производных более высоких порядков.

1 ВЫЧИСЛЕНИЕ ВТОРЫХ ПРОИЗВОДНЫХ СЛОЖНЫХ ФУНКЦИЙ МНОГИХ ПЕРЕМЕННЫХ

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 F требуется вычислить первые и вторые производные. Используемый аппарат позволяет вычислять не всю матрицу вторых производных, что может быть слишком сложно, а некоторый промежуточный вариант - вычислять градиенты от некоторых функцио-

налов градиента F , например от $\Phi(\text{grad}F) = \sum_{i=1}^n \frac{\partial F}{\partial x_i}$.

Рассмотрим процесс вычисления вторых производных функции F . Этот процесс состоит из трех этапов: на первом этапе строится граф прямого функционирования и вычисляется сама функция F , на втором этапе строится граф для вычисления ее градиента (граф первого обратного функционирования), и, наконец, на третьем этапе строится граф для вычисления градиента от функционала градиента ("обратно-обратное" функционирование).

Перейдем к более подробному описанию этапов вычисления производных.

Первый этап

Для построения графа, вычисляющего сложную фун-

кцию 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 [2, 3].

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 рис. 1 изображен граф, соответствующий этой функции. Окружностями изображены вершины графа, S_i - разбиение функций f_i на слои по глубине вхождения в функцию F .

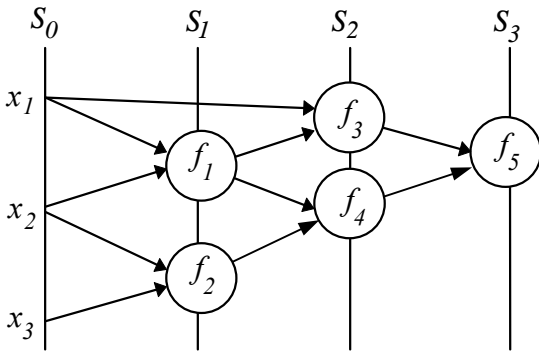


Рисунок 1

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 фун-

кции F при конкретных значениях вершин нулевого слоя.

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 прямого функционирования и соответствующие ей вершины графов обратного функционирования.

Второй этап

Вычисление градиента сложной функции многих переменных представляет собой некоторый вычислительный процесс, в ходе которого сигналы перемещаются в обратном направлении - от выходных элементов графа к входным. При этом каждый элемент графа заменяется на двойственный, с сохранением структуры связей между элементами.

Двойственным элементом к точке ветвления является сумматор. Двойственным элемент к сумматору - точка ветвления. Двойственным к элементу, вычисляющему значение функции, является элемент, умножающий свой входной сигнал на производную функции по входу исходного элемента. Более сложные элементы конструируются из этих элементов.

На рис. 2 изображены вершина графа прямого функционирования и двойственная ей вершина графа первого обратного функционирования. Здесь для простоты функцией f обозначена одна из функций f_i , входящая в суперпозицию функции F . Как показано на рисунке, точка ветвления вершины прямого функционирования заменяется сумматором, а элемент, двойственный вычисляющему функцию f , умножает приходящий к нему сигнал на частные производные функции f по соответствующим входам и посылает их на элементы, двойственные этим входам. Входными сигналами графа прямого функционирования являются компоненты вектора входных сигналов x_{in} . Входные сигналы графа обратного функционирования $\mu_{in} = 1$.

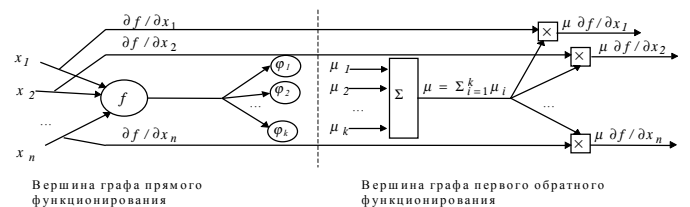


Рисунок 2

Формулы, полученные для вершины графа первого обратного функционирования путем построения двой-

ственных элементов, могут быть получены при помощи формулы для дифференцирования "двухслойной" сложной функции нескольких переменных:

$$\frac{\partial}{\partial x_i} \Phi(f_1(x_1, \dots, x_n), f_2(x_1, \dots, x_n), \dots, f_k(x_1, \dots, x_n)) = \quad (1)$$

$$= \sum_{j=1}^k \frac{\partial \Phi}{\partial f_j} \cdot \frac{\partial f_j}{\partial x_i}.$$

Для вершины графа первого обратного функционирования такие "двухслойные" функции имеют вид: $\Phi_i(f(x_1, \dots, x_n))$. Так как сигнал от вершины f поступает на k вершин Φ_i следующего слоя, рассмотрим сумму $\sum_{i=1}^k \Phi_i(f(x_1, \dots, x_n))$. Вычислим частную производную этой суммы по некоторой переменной x_m :

$$\frac{\partial}{\partial x_m} \left(\sum_{i=1}^k \Phi_i(f(x_1, \dots, x_n)) \right) = \sum_{i=1}^k \frac{\partial}{\partial x_m} \Phi_i(f(x_1, \dots, x_n)) =$$

$$= \sum_{i=1}^k \frac{\partial \Phi_i}{\partial f} \cdot \frac{\partial f}{\partial x_m} = \frac{\partial f}{\partial x_m} \sum_{i=1}^k \frac{\partial \Phi_i}{\partial f}.$$

Обозначив $\frac{\partial \Phi_i}{\partial f}$ через μ_i , получаем формулу для вершины графа первого обратного функционирования.

Третий этап

При вычислении градиента функции F строился граф, с той же системой связей, что и у графа, вычисляющего значение функции F , но все элементы графа заменялись на двойственные. Для того, чтобы вычислить вторые производные, нужно "развернуть" графы, вычисляющие значение функции и ее градиента в один граф, а затем построить граф, двойственный всему "развернутому" графу. Общая схема дважды двойственного функционирования показана на рис. 3.

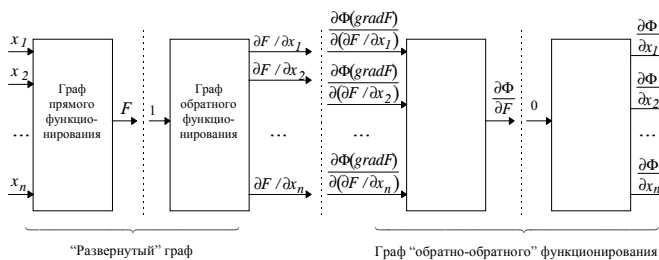


Рисунок 3

На рис.3 все графы изображены как самостоятельные элементы процесса вычисления вторых производных, входные и выходные сигналы которых никак не связаны между собой. Однако, процесс вычисления вторых производных можно рассматривать как единый граф, по которому сигналы проходят насквозь - от входов графа к его выходам. Для того, чтобы представить процесс вычисления в таком виде, необходимо между графами, изображенными на рис.3, вставить дополнительные элементы, обеспечивающие сквозное прохождение сигналов.

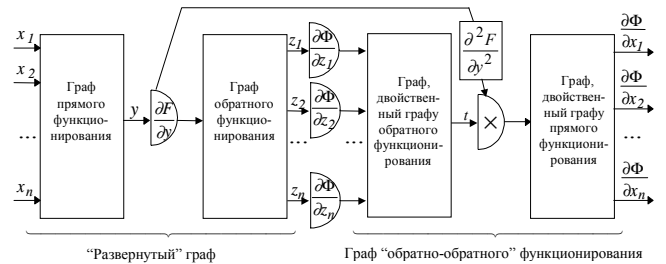


Рисунок 4

На рис.4 изображены графы и соответствующие дополнительные элементы. Здесь, для удобства, введены новые обозначения. Через y обозначено значение сложной функции F , через z_i обозначены частные производные $\frac{\partial F}{\partial x_i}$, а через t - производная $\frac{\partial \Phi}{\partial F}$. Элемент, двой-

ственный к элементу, вычисляющему $\frac{\partial F}{\partial y}$, перемножает свой вход t и производную исходного элемента, т.е. $\frac{\partial^2 F}{\partial y^2}$.

В процессе первого обратного функционирования участвуют элементы, производящие умножение двух входных сигналов. Это элементы, умножающие μ на частные производные $\frac{\partial f}{\partial x_i}$. Двойственные к ним элементы "обратно-обратного" функционирования имеют по одному входу и по два выхода. Если исходный элемент Y перемножил сигналы A, B , то выходы двойственного элемента Y' передают сигналы к элементам, двойственным источникам A и B соответственно. Обозначим здесь эти источники через A' и B' . При этом на A' подается входной сигнал Y' , умноженный на B ($d(AB)/dA=B$), а на B' - входной сигнал Y' , умноженный на A ($d(AB)/dB=A$). Умножающий элемент \times и двойственный ему элемент \times' изображены на рис. 5.

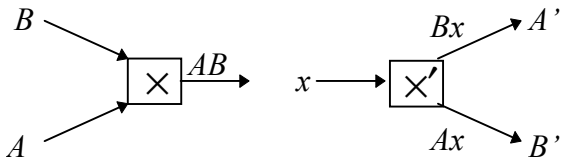


Рисунок 5

Построим вершины графа "обратно-обратного" функционирования.

Покажем, как строится вершина графа "обратно-обратного" функционирования, двойственная вершине графа первого обратного функционирования. Сигналы v_i приходят на входы элементов \times' , двойственных элементам \times вершины графа обратного функционирования, перемножающих сигнал μ и частные производные $\frac{\partial f}{\partial x_i}$. Как было описано выше, элемент, двойственный к перемножающему сигналы, посылает входящий к нему сигнал к элементам, двойственным исходным, умножая его на сигнал противоположного входа. Следовательно, сигналы v_i , умноженные на μ , посылаются к вершине графа, двойственной вершине графа прямого функционирования, а они же, умноженные на частные производные $\frac{\partial f}{\partial x_i}$, посылаются на сумматор, двойственный точке ветвления вершины графа обратного функци-

онирования. Далее полученная сумма $\sum_{i=1}^n v_i \cdot \frac{\partial f}{\partial x_i}$ приходит к точке ветвления, двойственной сумматору вершины графа обратного функционирования.

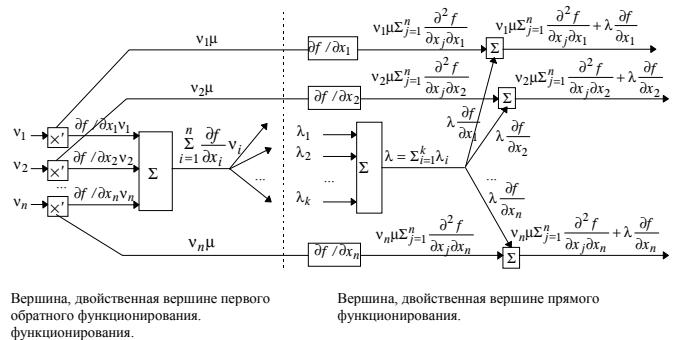
Входными сигналами графа "обратно-обратного" функционирования являются частные производные некоторого функционала от градиента функции F (то есть $v_{in} = \frac{\partial \Phi(\text{grad} F)}{\partial F / \partial x_i}$). В простейшем случае, когда функционал Φ является просто суммой частных производных

функции F ($\Phi(\text{grad} F) = \sum_{i=1}^n \frac{\partial F}{\partial x_i}$), входные сигналы графа v_{in} будут равны единицам.

Перейдем теперь к вершине графа "обратно-обратного" функционирования, двойственной вершине графа прямого функционирования. Сигналы λ_i поступают на входы сумматора, двойственной точке ветвления вершины графа прямого функционирования. После прохождения сумматора, полученный сигнал λ проходит через элемент, двойственный элементу, вычисляющему значе-

ние функции f . При этом сигнал λ умножается на соответствующие частные производные $\frac{\partial f}{\partial x_i}$. Поскольку от входных сигналов x_i элемента, вычисляющего значение функции, ответвлялись сигналы, передающие частные производные $\frac{\partial f}{\partial x_i}$ к элементам \times , то к сигналам $\lambda \frac{\partial f}{\partial x_i}$

добавляются сигналы $v_i \mu \sum_{j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}$. Входные сигналы графа "обратно-обратного" функционирования, двойственного графу прямого функционирования $\lambda_{in} = 0$. Вершины графа "обратно-обратного" функционирования изображены на рис.6.



Вершина, двойственная вершине первого обратного функционирования.

Вершина, двойственная вершине прямого функционирования.

Рисунок 6

Как и для графа первого обратного функционирования, можно показать, что формулы, полученные при помощи двойственных элементов, могут быть получены с использованием формулы дифференцирования сложной функции нескольких переменных. Для вычисления гра-

диента функции мы получили формулу $\frac{\partial f}{\partial x_m} \sum_{i=1}^k \frac{\partial \phi_i}{\partial f}$. Поскольку теперь мы движемся в обратном направлении, нам требуется вычислить частную производную по неко-

торой переменной x_l от суммы $\sum_{j=1}^n \frac{\partial f}{\partial x_j} \sum_{i=1}^k \frac{\partial \phi_i}{\partial f}$:

$$\begin{aligned} \frac{\partial}{\partial x_l} \left(\sum_{j=1}^n \frac{\partial f}{\partial x_j} \sum_{i=1}^k \frac{\partial \phi_i}{\partial f} \right) &= \sum_{j=1}^n \frac{\partial}{\partial x_l \partial x_j} \sum_{i=1}^k \frac{\partial \phi_i}{\partial f} = \\ &= \sum_{j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_l} \sum_{i=1}^k \frac{\partial \phi_i}{\partial f} + \sum_{j=1}^n \frac{\partial f}{\partial x_j} \frac{\partial}{\partial x_l} \sum_{i=1}^k \frac{\partial \phi_i}{\partial f} = \\ &= \sum_{j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_l} \sum_{i=1}^k \frac{\partial \phi_i}{\partial f} + \sum_{j=1}^n \frac{\partial f}{\partial x_j} \sum_{i=1}^k \frac{\partial^2 f}{\partial f^2} \frac{\partial \phi_i}{\partial x_l} \end{aligned}$$

Обозначая $\sum_{i=1}^k \frac{\partial \varphi_i}{\partial f}$ через μ и $\sum_{j=1}^n \frac{\partial f}{\partial x_j} \sum_{i=1}^k \frac{\partial^2 \varphi_i}{\partial f^2}$ через λ ,

получаем формулу для вершины, двойственной вершине графа прямого функционирования, изображенной на рис. 6. Множитель v_i для "двухслойных" сложных формул равен 1. Отличные от единицы значения множителя v_i появляются при большем количестве слоев.

Таким образом, для того, чтобы вычислить градиент от функционала градиента функции F , нужно сначала по графу прямого функционирования провести вычисление самой функции F , затем по графу первого обратного функционирования вычислить ее градиент, а затем провести вычисления по графу "обратно-обратного" функционирования.

2 ОЦЕНКА СЛОЖНОСТИ ВЫЧИСЛЕНИЯ НА ГРАФАХ

Подсчитаем число операций, необходимых для вычисления всех двойственных переменных для вершин и для ребер графа.

Если характерное число переменных у простых функций, соответствующих вершинам графа, равно k , то для вычисляемой по графу сложной функции F можно оценить отношение затрат на вычисление F и $\text{grad}F$ следующим образом:

$$T_{\text{grad}F} \cong c \frac{E}{k} k^2 = ckE, T_F \cong cE, \frac{T_{\text{grad}F}}{T_F} \cong k,$$

где E - число ребер в графе, c - сложность вычисления частной производной для вершины графа.

Если нас интересуют графы с большим числом связей, то для сохранения эффективности вычисления градиентов нужно ограничиваться специальными классами простых функций. Для функций f следующего вида вычисление частных производных вообще не требуется:

$$f(z_1, \dots, z_n) = \sum_{i \in P_1} z_i + \sum_{j \in P_2, k \in P_3} z_j z_k, \quad (2)$$

где множества индексов P_1, P_2, P_3 не пересекаются и, кроме того, P_2, P_3 имеют одинаковое число элементов. Значения всех частных производных таких функций уже известны, а источники (адреса) этих значений указываются связями графа. Какие-то значения z_k во второй сумме могут быть константами (значения нулевого слоя), какие-то - независимыми переменными (также нулевой слой) или значениями, найденными при промежуточных вычислениях. В общем случае функции (2) билинейны. Их частный случай - линейные функции: если индексам из P_2 в (2) соответствуют кон-

станты, то функция f - линейная. И функции вида (2), и соответствующие им вершины будем называть *квазилинейными*.

Пусть в графе вычислений присутствуют только вершины двух сортов: квазилинейные или с одной входной связью (соответствующие простым функциям одного переменного). Обозначим количества этих вершин V_q и V_1 соответственно. Оценка сложности вычисления градиента для таких графов принципиально меняется. Обратное функционирование в этом случае требует следующих затрат:

$$T_{\text{grad}F} = cV_1 + 2E - (V - V_{out}),$$

где V - количество вершин в графе.

Оценим сложность вычислений при прямом распространении:

$$T_F = cV_1 + (E - V - V_q).$$

С ростом числа связей отношение затрат при обратном функционировании к затратам при прямом функционировании стремится к двум.

Графы вычислений, в которых присутствуют только вершины двух сортов: квазилинейные или с одной входной связью (соответствующие простым функциям одного переменного) играют особую роль. Будем называть их *существенно квазилинейными*. Для функций, вычисляемых с помощью таких графов, затраты на вычисление вектора градиента примерно вдвое больше, чем затраты на вычисление значения функции. При этом число связей и отношение E/V могут быть сколь угодно большими. Это достоинство делает использование существенно квазилинейных графов весьма притягательным во всех задачах оптимизации. Их частным случаем являются нейронные сети, для которых роль квазилинейных вершин играют адаптивные линейные сумматоры.

Поэтому для нейронных сетей обратное распространение выглядит особенно просто.

Отличие общего случая от более привычных нейронных сетей состоит в том, что можно использовать билинейные комбинации (2) произвольных уже вычисленных функций, а не только линейные функции от них с постоянными коэффициентами-весами.

ВЫВОДЫ

Изложенные в работе способы вычисления сложных функций многих переменных, а так же градиентов и вторых производных таких функций при помощи графов, позволяют существенно сократить вычислительные затраты. Выделен класс функций, для которых вычисление градиента наиболее просто (нейронные сети являются частным случаем таких функций). Принцип двойственности, описанный в работе, позволяет достаточно

легко конструировать графы, вычисляющие градиент сложной функции и градиент от некоторого функционала градиента этой функции.

Работа поддержана грантом для молодых учёных №1M0034 Красноярского краевого фонда науки.

ПЕРЕЧЕНЬ ССЫЛОК

1. Горбань А.Н. Обучение нейронных сетей. М.: изд. СССР-США СП "ПараГраф", 1990. 160 с.
2. Горбань А.Н., Дунин-Барковский В.Л., Кирдин А.Н. и др. Нейроинформатика. Новосибирск. Наука. 1998.
3. 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.
4. Rummelhart D.E., Hinton G.E., Williams R.J. Learning representations by back-propagating errors. // Nature, 1986. V. 323. Pp. 533-536.
5. Senashova M. Yu., Gorban A. N., Wunsch D. Back-Propagation of Accuracy. // Proc. IEEE/INNS International Conference of Neural Networks, IEEE, USA, 1997. Pp. 1998-2001.

Надійшла 01.03.2000

Після доробки 28.04.2000

УДК 681.5:65.014.1

АНАЛИЗ И СИНТЕЗ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

С. Н. Сердюк, А. А. Поздняков

Рассмотрены проблемы управления информацией в современных организациях. Анализируются различные классы информационных систем и технологии их построения, призванные решить выявленные проблемы и обеспечить поддержку принятия решения управленцами всех уровней. Описан подход к синтезу системы поддержки принятия решений для бизнес-процессов организации, реализованный на базе OLAP-технологии.

Розглянуті проблеми управління інформацією в сучасних організаціях. Аналізуються різні класи інформаційних систем і технології їх побудови, покликані вирішити виявлені проблеми і забезпечити підтримку прийняття рішення керівниками всіх рівнів. Описан підхід до синтезу системи підтримки прийняття рішень для бізнес-процесів організації, реалізований на базі OLAP-технології.

In article consider the management problems by information in contemporary organizations. Analyse the different classes of information systems and technologies of their construction, called to work out exposed problems and to provide decision support by managers of all of levels. Is Described approach to decisions support system synthesis for organization business-processes, realized on the basis of OLAP-technology.

ВВЕДЕНИЕ

В настоящее время, во многих организациях, несмотря на широкое использование различных систем оперативной обработки данных, систем поддержки принятия решений (СППР) сложилась парадоксальная ситуация: информация существует, интенсивно создаётся и используется, но она не структурирована, не согласована, её довольно сложно найти и получить. Такие системы не в состоянии обеспечить хранение исторических данных более чем за несколько месяцев. Массивы данных выгружаются в архивный набор и вычищаются из оперативной базы, следовательно остаются невостребованными. В то же время, высшее звено управляющих, отве-

чающих за определение стратегии развития деятельности организации, вынуждены работать в условиях практически полного отсутствия информации - "вслепую", полагаясь только на собственный опыт и интуицию. Требуемую управленцу нестандартную информацию не удаётся получить быстро вследствие того, что информационные потоки внутри фирмы спланированы заранее, сроки представления отчетности регламентированы. Поэтому, потребность в быстром и нестандартном решении сразу сталкивается с негибкостью механизма доступа к информации. В итоге, возрастает количество отчетов которые готовятся на заказ, по требованию руководства организацией, а это может привести к расширению штата программистов и аналитиков, обеспечивающих генерацию этих отчетов.

Таким образом, в настоящее время существует проблема быстрого поиска требуемой информации и представление её в том виде, в котором она нужна для лиц принимающих решения (ЛПР).

ПОСТАНОВКА ЗАДАЧИ

Рассмотрим подходы, используемые в современных информационных системах (ИС), и определим класс систем, позволяющий решить описанную проблему.

В настоящее время не существует единой классификации ИС, что обусловлено сложностью систем данного типа. В основу дальнейшего рассмотрения положены результаты работы одного из авторов данной статьи по классификации ИС с точки зрения их исторического развития и способности осуществлять поддержку ЛПР [1]. Проведенный анализ показал, что требованию обеспечения интеллектуальной поддержки ЛПР удовлетворяют два класса ИС - экспертные системы и интеллект-

туальные СППР, причем различие между ними заключается только в организации логического вывода. В рамках данной статьи объектом дальнейшего рассмотрения будут интеллектуальные СППР.

СППР, в зависимости от данных, с которыми они работают, можно разделить на оперативные, предназначенные для немедленного реагирования на текущую ситуацию, и стратегические (интеллектуальные) - основанные на анализе большого количества информации из разных источников с привлечением сведений, содержащихся в системах, аккумулирующих опыт решения проблем [2].

СППР первого типа представляют собой конечные наборы отчетов, построенные на основании данных из транзакционной информационной системы организации (OLTP-системы, On-Line Transaction Processing). По мере их развития возникла проблема, связанная с накоплением огромного объема данных, замедляющих процесс построения отчетов настолько, что управленческий состав не успевает готовить на их основе соответствующие решения. В связи с этим, было найдено решение, сформулированное в виде концепции Хранилища Данных (ХД).

Цель ХД - оперативный поиск информации в предварительно подготовленном хранилище. По сути, именно ХД позволяют решить проблему, поставленную в начале данной статьи. В самом простом варианте для ХД используется та модель данных, которая лежит в основе транзакционной системы. Если эта система функционирует на реляционной СУБД (РСУБД), то самой сложной задачей становится выполнение произвольных запросов. Проблема заключается в невозможности заранее оптимизировать структуру БД так, чтобы все запросы работали эффективно. Практика принятия решений показала, что чем более агрегированными являются данные, тем чаще выполняются запросы. Круг пользователей работающих с обобщенными данными шире, чем тот, для которого нужны детальные данные. Это наблюдение легло в основу технологии оперативной аналитической обработки (On-Line Analytical Processing, OLAP), предложенной Е.Коддом в 1993 году.

В основе OLAP лежит понятие многомерности модели данных - гиперкуба (многомерная БД - МБД), в ячейках которого хранятся анализируемые данные. Оперативный анализ данных, реализованный на основе МБД, получил название Multidimensional OLAP (MOLAP).

Помимо MOLAP существует ROLAP технология выполненная на основе реляционной БД (РБД) в которой многомерная структура реализуется реляционными таблицами, а данные в процессе анализа, выбираются из реляционной БД аналитическим инструментом. В настоящее время исторические данные практически во всех организациях хранятся в РБД, поэтому резкий переход к МБД неоправдан, так как его реализация потребует значительных материальных и временных ресурсов.

Данная ситуация послужила поводом к возникновению нового класса аналитических инструментов - HОLAP (Hybrid OLAP), сочетающего в себе оба подхода. Этот тип OLAP является самым эффективным для реализации СППР, так как обеспечивает доступ как к МБД, так и к РБД.

Однако, разработка корпоративного ХД даже на основе HОLAP не позволяет достичь вышеуказанной цели ХД в связи с: сложностью бизнес-процессов организации; ограниченностью технических возможностей; необходимостью в изменении организационной структуры и др. В результате попыток смягчить трудности разработки и внедрения ХД появились витрины данных (ВД).

Витрины данных представляют собой специальные хранилища, состоящие из множества тематических БД, каждая из которых содержит информацию об одном из направлений деятельности организации, что приводит к сокращению времени поиска данных необходимых ЛПР. Однако, вместе с ростом числа ВД, растет сложность организации их взаимодействия, поскольку сделать витрины полностью независимыми не удастся. Кроме того, специализация ВД по направлениям деятельности не всегда эффективна, так как, при этом, не учитываются потребности конкретного ЛПР.

Общепринятого подхода к формализации знаний необходимых для построения ВД, обеспечивающих ЛПР исходными данными для выработки бизнес-решений пока не существует. Поэтому, предлагается один из способов формализации знаний о бизнес процессах организации, основанный на положениях функционально-структурной теории "человеко-машинных систем" (ФСТ "ЧМС") [3].

МЕТОДЫ И АЛГОРИТМЫ РЕШЕНИЙ

Стадия формализации прикладной области (ПРО) для которой создается СППР, её "настройка" на эту область проводимая системным аналитиком организации, является центральной в процессе общего проектирования системы. На этапе системного анализа ПРО ему необходимо: выявить структуру данной области; выделить из всех субъектов, участвующих в бизнес-процессах организации только ЛПР; для каждого l -го ЛПР необходимо определить номенклатуру интеллектуальных задач Z_l функций V_l и процедур P_l характеризующих проблемную ситуацию PS_{li} , которая может возникнуть в процессе его трудовой деятельности; для каждой выявленной интеллектуальной задачи, функции и процедуры определить объем знаний (данных) необходимых для построения соответствующих математических моделей и информационной поддержки ЛПР.

Наиболее общим формализмом, позволяющим описать ПРО является формально-аксиологическая система вида [3]:

$$FS(N) = \langle SnFS, SmFS, AxFS \rangle, \quad (1)$$

где $FS(N)$ - модель информационного отображения организации с именем N ;

$SnFS$ - синтаксическая компонента (данные) информационного отображения;

$SmFS$ - семантическая компонента (знания) информационного отображения;

$AxFS$ - аксиологическая (оценочная) компонента в виде количественных данных и знаний о процедурах над ними.

Для решения стоящих перед ним задач, системный аналитик должен использовать следующее множество структур, являющихся составляющими синтаксической компоненты $SnFS$: компонентно-системных $KSS(N)$; компонентно-элементных $KES(N)$ и компонентно-функциональных $KFS(N)$.

Компонентно-системная структура представляет собой знания об окружающей среде $EES(N)$ и о внутреннем (подсистемном) составе данной организации - $ETS(N)$. Знания $EES(N)$ включают: имя метасистемы, в которую частью входит данная организация ($EES - 0$); множество смежных систем аналогичного уровня, с которыми связана данная организация (но которые не входят в её состав) ($EES - 1$); множество возмущающих систем внешней среды ($EES - 2$), которые организационно не связаны с данной организацией, но оказывающие влияние на качество её функционирования.

Типовая структура знаний о внутреннем подсистемном составе организации представлена на рис.1. В общем

случае она описывается иерархическим деревом, вершинами которого являются: начальная вершина графа - сама организация $ETS(N)$; промежуточные вершины - групповые подсистемы ($GETS - N.X.Y$), где X - номер уровня иерархии данной подсистемы, Y - номер подсистемы на данном уровне (групповые организации, получаемые путем выделения функционально-целевых комплексов вначале из организации - первый уровень иерархии, а затем из групповых организаций первого уровня - второй уровень иерархии и т.д.); конечные вершины - локальные организации ($LETS - N.X.Y$), которыми являются моноэргатические подсистемы (в которых действует только одно ЛПП).

Верхняя часть $KSS(N)$ имеет следующие дуги: метасистема и данная организация связаны отношением "целое-часть" (R_{GP}), а смежные и возмущающие системы с данной организацией - отношением "воздействовать" (R_F).

Для всех локальных подсистем $LETS-N.X.Y$ строится множество KES .

$$KES(N) = \{KES-N.X.Y\}, \quad (2)$$

где $KES-N.X.Y$ - компонентно-элементная структура локальных подсистем организации, описывающая совокупность знаний о её внутреннем элементном составе.

$$KES-N.X.Y = \langle KTK, KDK-N.X.Y \rangle, \quad (3)$$

где KTK - знания о типах компонентов любой $LETS-N.X.Y$;

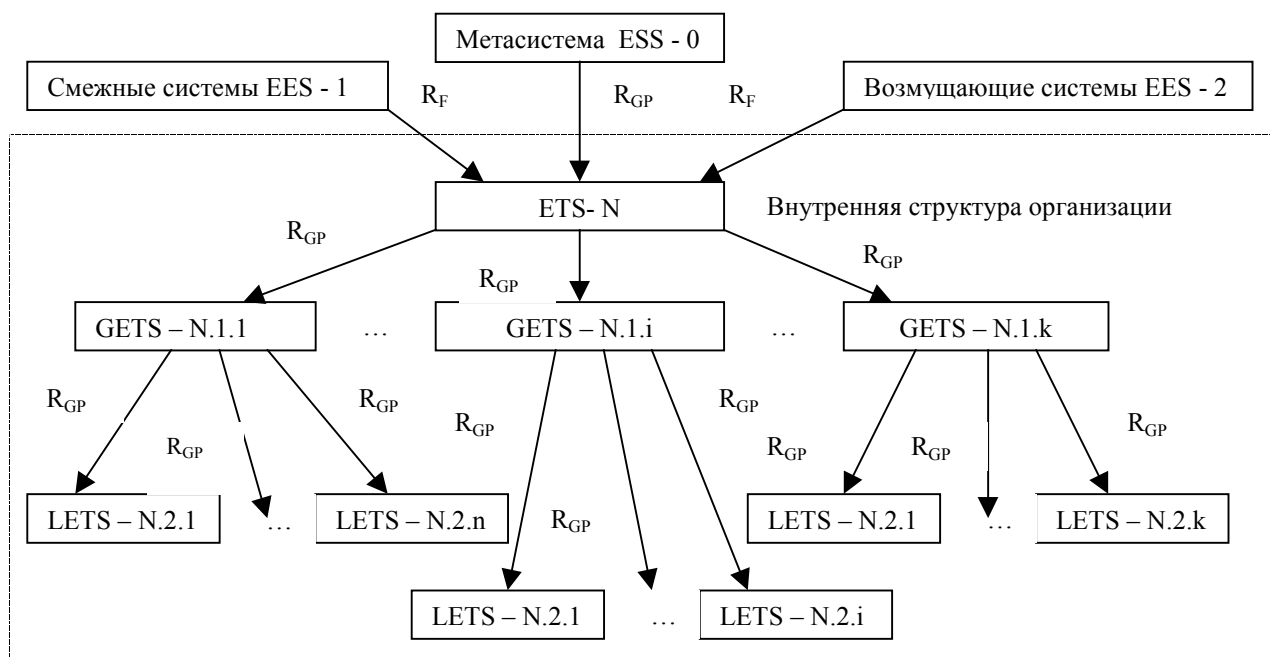


Рисунок 1 - Компонентно-системная структура организации

KDK-N.X.Y - знания о конкретных компонентах, образующих данную локальную подсистему LETS - N.X.Y:

$$KTK = \langle ST, R_{GP}^{ST*ST} \rangle, \tag{4}$$

где $ST = \{LHT-N.X.Y, TK\}$ - множество вершин типовой структуры;

$$TK = \{T_i\} = \{TH, TM, TO, TE_1, TE_2, TE_3, TE_4, TE_5, TE_6\}$$

- множество типов компонентов внутренней структуры организации;

R_{GP}^{ST*ST} - отношение "целое-часть", заданное на декартовом произведении $ST*ST$ в соответствии с рис.2.

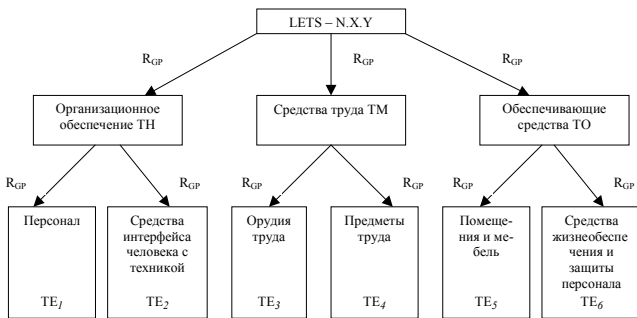


Рисунок 2 - Компонентно-элементная структура локальной организации

Структура KFS (N) представляет собой семантические графы в виде иерархических деревьев, вершинами которых являются имена функциональных единиц (Fi) различного масштаба, а дуги отображают отношения R_{FiFj} - "для реализации функциональной единицы Fi необходимо выполнить функциональную единицу Fj". KFS имеют типовую структуру, показанную на рис.3. Корень дерева - понятие "Процесс функционирования ETS (N)" (PFETS (N)). Первый уровень вершин дерева - название целей - T_{ri} , для которых создана данная организация. Второй уровень - название задач T_{sij} , которые надо решить для достижения целей T_{ri} . Третий уровень - название процедур, которые нужно выполнить для решения задачи T_{sij} , а дуги отображают отношение R_{FiFj} "для реализации функциональной сущности Fi (цели, задачи) необходимо выполнить функциональную сущность Fj (задачу, процедуру, функцию)".

Затем, для всех построенных функциональных сетей проводится декомпозиция составляющих их функций, процедур на эргатические (выполняемых человеком совместно с техникой) и автоматические, которые выполняются автоматическим оборудованием без вмешательства человека. К дальнейшему анализу принимаются только эргатические интеллектуальные функции ЛПП в бизнес-процессах. Считаем, что интеллектуальная функция будет иметь в своем составе в любой форме "выбор из..." (из вариантов действий, решений, планов и т.д.) для которой отсутствует предписанная процедура (правило, алгоритм) реализации однозначного выбора.

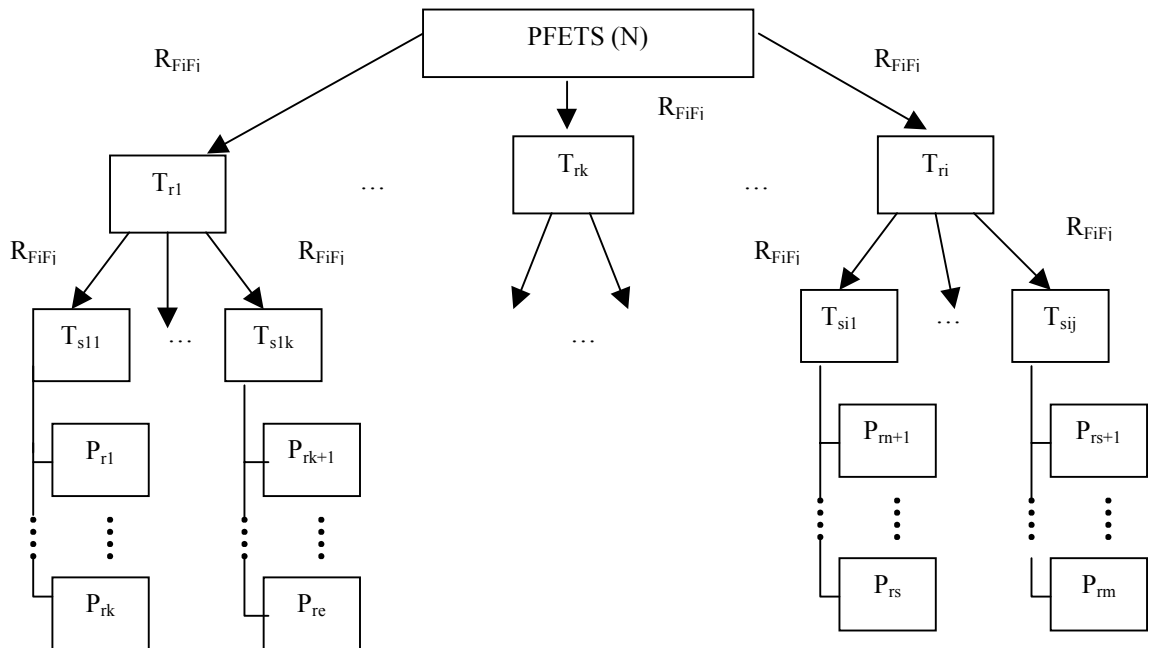


Рисунок 3 - Компонентно-функциональная структура

На следующем этапе, для каждой интеллектуальной функции F_l l -го ЛПР определяется перечень проблемных ситуаций $S_l = \{S_{li}\}$. Имея этот перечень, системный аналитик (или инженер по знаниям) в процессе работы с экспертом по бизнес-процессам определяет структуру знаний, используемых для каждой S_{li} -й проблемной ситуации - в виде перечня имен баз знаний данных (БЗД), которые, используя терминологию OLAP, можно рассматривать как ВД. Процедуры генерации ВД реализуются системным аналитиком и программистом в виде соответствующих управляющих метаправил, расширяющих БЗД СППР. В случае необходимости, системный аналитик для S_{li} -ой проблемной ситуации должен разработать соответствующую математическую модель.

Таким образом, приведенный выше подход позволяет формализовать знания о ПРО "Бизнес-процессы организации" на этапе синтеза СППР.

РЕЗУЛЬТАТЫ

Разработанная методика системного анализа ПРО, является теоретической основой для построения ВД СППР. Проектирование таких сложных систем как СППР необходимо проводить на основе системного подхода. Поэтому, в целях повышения эффективности и качества интеллектуальной поддержки ЛПР необходимо разработать СППР в виде оболочки "настраиваемой" на бизнес-процессы конкретной организации. При этом СППР должна быть направлена не только на конечного пользователя - ЛПР в бизнес-процессах, но и автоматизировать работу системного аналитика, который обеспечивает формализацию бизнес-процессов, т.е. система "системный аналитик - СППР - ЛПР" должна рассматриваться как единое целое.

Такой подход определяет следующие требования к разрабатываемой СППР: в качестве основного источника оперативных данных, должны выступать OLTP- системы; все данные должны быть очищены, преобразованы и собраны в ХД; должны быть разработаны метаправила генераций ВД для каждой интеллектуальной функции ЛПР, математические модели для функций ЛПР требующие поддержки при принятии решений и обеспечен свободный доступ пользователя к данным посредством информационного сервера.

Данные требования, а также технология OLAP определяют базу архитектуры СППР для бизнес-процессов организации, которая условно разделена на восемь уровней и представлена на рис.4.

Первый уровень представлен источниками внутренних и внешних данных. Источниками внутренних данных являются OLTP - системы, представленные в виде сервера оперативной БД. Внешние данные, проходят обработку

в сервере очистки и преобразования данных в автоматическом режиме на основе метаправил, задающих алгоритм очистки. Вся информация, необходимая для реализации этих процедур, хранится в виде единого справочника метаданных.

Второй уровень представлен ХД, которое, размещается на серверах БД детализированных данных. В нём концентрируются данные, поступающие как из внешних, так и из внутренних источников.

Третий уровень представляет собой модуль настройки на ПРО. В данном модуле заполняются: компонентно-системная, компонентно-функциональные и компонентно-элементные структуры, позволяющие системному аналитику структурировать ПРО. Результатом работы данного уровня являются метаправила генерации ВД.

Четвёртый уровень архитектуры СППР представлен модулем настройки ВД. Назначением модуля является разграничение данных ХД на тематические базы и генерация ВД, в соответствии с имеющимися метаправилами. В качестве тематических баз выступают знания о каждой S_{li} -й проблемной ситуации.

Пятый уровень представлен множеством ВД в виде серверов БД агрегированных данных и предназначен для проведения аналитических и прогностических исследований.

Шестой уровень представлен модулем принятия решений содержащим математические модели принятия решений, которые, в случае необходимости, разрабатываются системным аналитиком для S_{li} -ой проблемной ситуации.

Вся информация, предназначенная ЛПР, собирается в информационном концентраторе, представленном седьмым уровнем. Его основной задачей является сбор достоверной информации, для публикации в информационной сети организации. В качестве информационного концентратора выступают WEB-сервера.

Последний уровень архитектуры СППР представлен рабочими местами ЛПР организации оснащенными стандартными навигаторами (Netscape Navigator или Internet Explorer). Необходимую для интеллектуальной поддержки информации они получают от информационного концентратора.

ВЫВОДЫ

Приведенный подход построения СППР является усовершенствованным вариантом типового, использующего OLAP-технологию. Новизна подхода заключается в построении ВД не по тематическим базам для подразделений организации, а по конкретным функциям ЛПР, нуждающимся в интеллектуальной поддержке, что позволяет значительно сократить время поиска необходимой информации за счёт генерации соответствующих ВД.

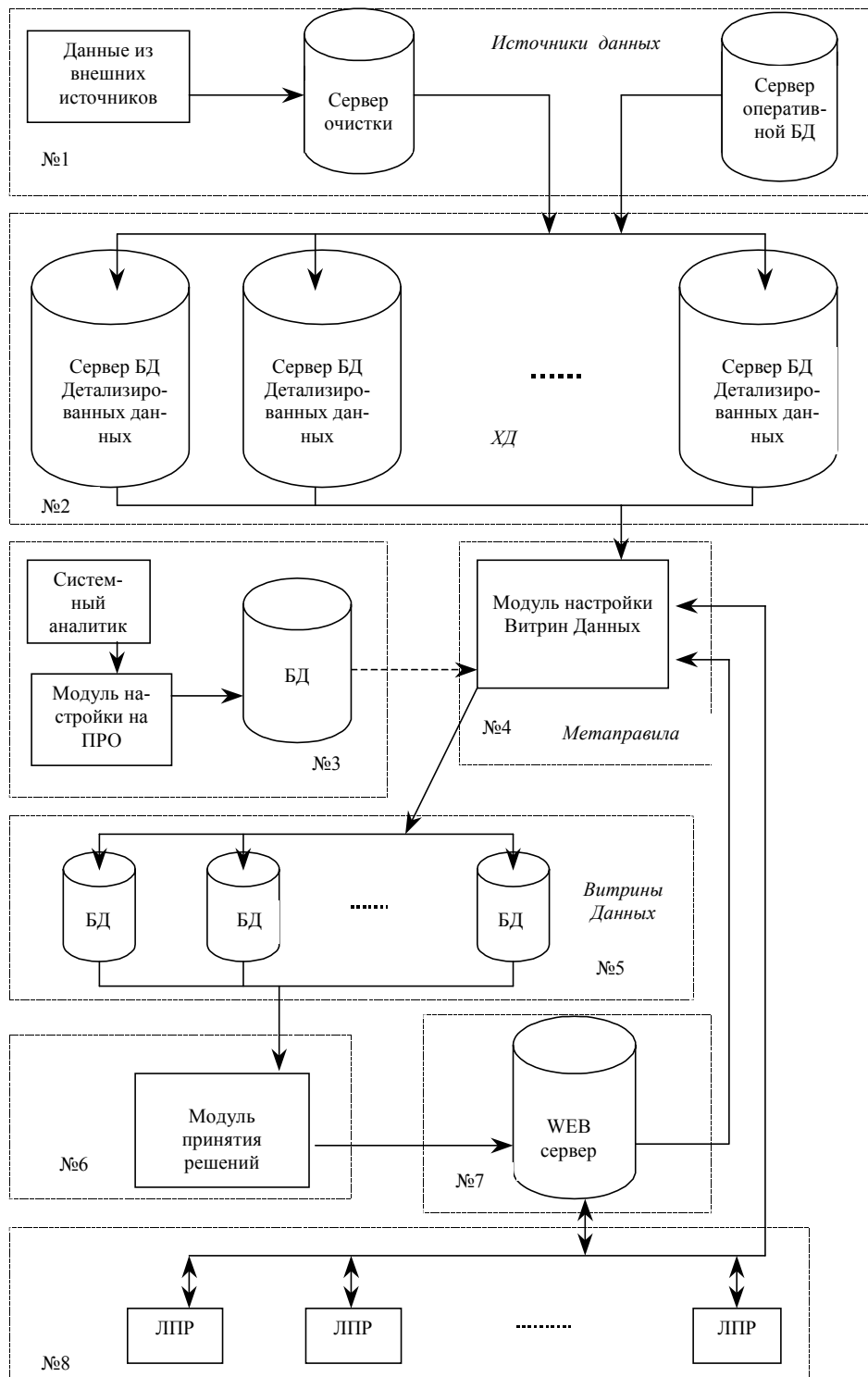


Рисунок 4 - Архитектура СППР

ПЕРЕЧЕНЬ ССЫЛОК

1. Сердюк С.Н. Классификация систем интеллектуальной поддержки эргономического обеспечения проектирования, XVII Межрегиональный семинар "Эргономика и эффективность систем "человек-техника". Тез.докл. - Игналина: Сов. эргон. Ассоц., ЛЭТИ, Академия управления Литвы, 1991. - с.82.
2. Львов В. Создание систем поддержки принятия решений на основе хранилищ данных, Системы Управлениями Базами Данных, #3, 1997, с 30- 40.
3. Попович П.Р., Губинский А.И., Колесников Г.М. Эргономическое обеспечение деятельности космонавтов.-М.: Машиностроение, 1985.- 256с.

Надійшла 24.02.2000
Після доробки 02.03.2000

РОЗРОБКА ТА ДОСЛІДЖЕННЯ СИСТЕМИ ІДЕНТИФІКАЦІЇ КОРИСТУВАЧА НА ОСНОВІ НЕЙРОННОЇ МЕРЕЖІ

В. В. Сокирук, Є. В. Яремчук, В. І. Месюра

Предложен оригинальный подход к решению задачи идентификации в системах защиты информации, которые используют в качестве объекта идентификации пароль пользователя. Этот подход базируется на использовании алгоритма обучения нейронной сети в качестве хеш-функции, которая превращает пароль пользователя в вектор весовых коэффициентов сети. Показана высокая стойкость предложенной системы к известным атакам на одностронность и на коллизии.

Запропоновано оригінальний підхід до вирішення задачі ідентифікації в системах захисту інформації, які використовують в якості об'єкта ідентифікації пароль користувача. Цей підхід базується на використанні алгоритму навчання нейронної мережі в якості хеш-функції, яка перетворює пароль користувача в вектор вагових коефіцієнтів мережі. Показано високу стійкість запропонованої системи до відомих атак на односторонність та на колізії.

We proposed the original decision of a user authentication task in cryptosystems, in which the user password is used as object of identification. This approach is based on use of training algorithm of a neural network as hash-function, which will transform the user password into a vector of network parameters. High resistance of the offered system against known cryptographic attacks is shown.

ВСТУП

Найбільш поширені системи ідентифікації користувача, які використовуються в задачах захисту інформації, побудовані на основі пароля. Вони є найбільш вразливими, тому що виникає необхідність збереження пароля в системі (наприклад, у вигляді файла), тобто останній залишається практично незахищеним від несанкціонованого доступу. Для подолання цього недоліку застосовують хеш-функції [1,2], які перетворюють пароль у деяке числове значення (хеш-значення), яке можна відкрито зберігати.

В даній статті описується система ідентифікації користувача, побудована на основі нейронної мережі, та проводиться дослідження її стійкості.

СТРУКТУРА ТА ОСНОВНІ ЕЛЕМЕНТИ СИСТЕМИ ІДЕНТИФІКАЦІЇ КОРИСТУВАЧА

В системі, що пропонується, в якості такого хеш-значення використовується вектор вагових коефіцієнтів нейронної мережі. Для забезпечення відтворення та збереження останнього використовуються наступні три блоки системи, зображені на рис. 1.

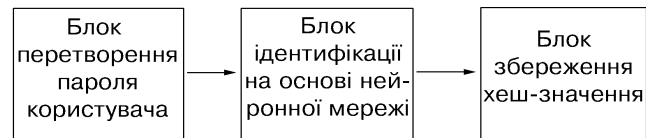


Рисунок 1 - Структура системи ідентифікації користувача

Блок перетворення пароля користувача відповідає за створення вхідних даних для нейронної мережі. Головна умова створення працездатної та коректної нейронної мережі - правильний вибір вибірки, що навчає (далі ВН). ВН будемо називати множини $\{P_n\}$, елементами якої є пари відомих вхідних та вихідних векторів $P_i = (X_i, Y_i)$, які отримуються з початкових даних [3]. В нашому випадку апріорними даними для ВН служить ідентифікаційний пароль S^0 , який задається користувачем на початку експлуатації системи. Для отримання ВН використаємо алгоритм перетворення двійкової послідовності (кодових значень символів вхідного пароля) у множини навчальних прикладів, який відповідає наступним вимогам:

1) унікальність створеної множини $\{P_n^i\}$ для будь-якого вхідного пароля S^i , тобто виконання наступної умови:

$$\{P_n^i\} \neq \{P_n^j\} \text{ при } S^i \neq S^j, i \neq j, S^i, S^j \in \{S^k\}; \quad (1)$$

2) відсутність однакових елементів P_i та елементів з однаковими складовими X_k^* та Y_k^* (елементи ВН повинні відповідати умовам функціональності та ін'єктивності).

Алгоритм А1, який відповідає зазначеним вимогам і забезпечує перетворення бінарної послідовності $\{U_m\}$ у множини $\{P_n\}$, елементами якої є дійсні числа, наведено нижче.

Алгоритм А1 (побудова ВН на основі пароля користувача).

1. Будуємо розширену бінарну послідовність $\{U_s^*\}$ шляхом додавання N-1 перших біт в кінець $\{U_m\}$, де N -

розмір пароля.

2. Для побудови послідовності дійсних чисел $\{V_m\}$ використаємо "метод вікон", який полягає у використанні вікна розміром N біт, яке може пересуватись з певним кроком вздовж розширеної бінарної послідовності. Усі біти, які знаходяться всередині вікна, представляються як деяке N -розрядне двійкове число a_i , де i - поточна позиція вікна. Тоді i -тий елемент послідовності $\{V_m\}$ визначається за формулою

$$v_i = \frac{a_i}{2^{N-1}}, \quad i = 0, 1, \dots, m-1.$$

3. З послідовності дійсних чисел $\{V_m\}$ отримаємо N вхідних векторів X^* .

4. Відповідні вихідні значення Y_j^* знаходяться за формулою

$$Y_j^* = \frac{\sum_{i=0}^j b_i}{N-1}, \quad j = 0, 1, \dots, N-1, \quad (2)$$

$$\sum_{k=0} c_k$$

де b_i - вікно розміру N , яке пересувається по бінарній послідовності з кроком $N-1$;

c_k - вікно розміру N , яке пересувається по бінарній послідовності з кроком N .

Основним елементом системи є блок ідентифікації, побудований на основі багат шарової нейронної мережі прямого розповсюдження [3], що зображена на рис. 2.

Вона складається з трьох шарів, в яких розміщена фіксована кількість нейронів, які утворюють пірамідальну структуру. Вихідний шар мережі представлений одним нейроном, а кількість нейронів першого шару дорівнює кількості бітів, які визначають одну літеру пароля. Представлена мережа забезпечує в найпростішій реалізації, щонайменше, 100 змінних параметрів (приймемо таку розмірність хеш-значення як достатню).

Процес функціонування нейронної мережі полягає у перетворенні вхідного вектора X у деякий вихідний вектор Y . Для M -шарової нейронної мережі його можна описати за допомогою наступного нелінійного рекурентного рівняння [4]:

$$Y = \psi^{(M)}(\omega_0^{(M)} + W^{(M)}\psi^{(M-1)}(\omega_0^{(M-1)} + W^{(M-1)}\psi^{(M-2)} \times (\dots(\omega_0^{(2)} + W^{(2)}\psi^{(1)}(\omega_0^{(1)} + W^{(1)}X)\dots))), \quad (3)$$

де $\omega_0^{(\mu)}$ - N_μ -мірний вектор порогових вагових коефіцієнтів $\omega_0^{(\mu, i)}$;

$W^{(\mu)}$ - $N_\mu \cdot N_{\mu-1}$ -мірна матриця вагових коефіцієнтів μ -го шару;

$\psi^{(\mu)}$ - нелінійний оператор, який перетворює вхідний вектор у відповідності з функцією активації нейрона.

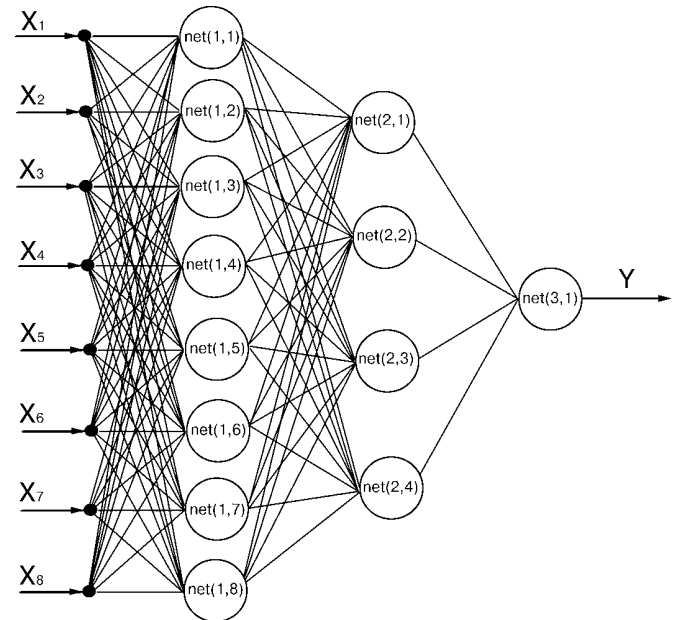


Рисунок 2 - Структура нейронної мережі, покладеної в основу системи

Фактично нейронна мережа реалізує складну нелінійну функцію вхідного вектора X , яка залежить від вектора вагових коефіцієнтів мережі W . Для знаходження останнього застосовують алгоритм навчання - процес знаходження таких вагових коефіцієнтів, при яких мережа відповідає ВН зі ступенем відповідності ξ . Величина ξ носить назву похибки нейронної мережі і визначає інтегральну міру наближення отриманих виходів мережі до бажаних [5]. В якості алгоритму навчання можливе використання будь-якого детермінованого ітеративного алгоритму навчання, який забезпечує необхідну швидкість функціонування системи.

Для підвищення надійності системи ідентифікації при остаточному формуванні хеш-значення використаємо алгоритм перемішування вагових коефіцієнтів (блок 3), в основу якого покладено генератор псевдовипадкових чисел [2]. Останній ініціюється значенням, яке будується на основі пароля, що забезпечує створення унікальних перестановок для різних паролів.

АНАЛІЗ СТІЙКОСТІ СИСТЕМИ

Надійність системи ідентифікації визначається якістю хеш-функції $h(x)$, яка покладена в її основу. Якість хеш-функції, в свою чергу, визначається стійкістю до

криптографічних атак, метою яких є розкриття пароля за відомим хеш-значенням [1]. Найбільш ефективними є наступні:

1) *атака на односторонність*: знаходження вхідного значення x за відомими значеннями $h(x)$ та h ;

2) *атака на колізії*: знаходження такого значення x' , що $h(x') = h(x)$.

Розрізняють пряму та непряму атаки на односторонність [1]. Перша полягає у визначенні оберненої залежності або знаходженні алгоритму отримання невідомого вхідного значення. Якщо пряма атака може досягти успіху у виключних випадках, то непряма атака на односторонність має набагато більше шансів і полягає у перегляді певної множини вхідних значень (взятих випадковим чином або згідно визначеного алгоритму) і порівнянні обчислених для кожного елемента хеш-значень з шуканим.

В нашому випадку хеш-значення представляє собою невпорядкований вектор вагових коефіцієнтів нейронної мережі, який був отриманий після навчання мережі з використанням ВН, що залишається невідомою. Таким чином, за допомогою прямої атаки на односторонність необхідно вирішити дві задачі:

1) відновлення правильного розташування вагових коефіцієнтів нейронної мережі;

2) визначення тих входів з множини можливих, які були задіяні при навчанні (тобто визначення елементів ВН).

Для вирішення першої задачі необхідно здійснити перегляд множини можливих векторів вагових коефіцієнтів, що складається з $N_W!$ елементів, де N_W - розмірність вектора W . Для зменшення цієї множини можливе використання наступних евристичних правил:

1) абсолютні величини вагових коефіцієнтів вищих шарів мережі, як правило, більші за відповідні значення попередніх шарів (особливості градієнтних алгоритмів навчання);

2) сума значень вагових коефіцієнтів для окремого нейрона, як правило, наближується до нуля.

Достовірність отриманого вектора вагових коефіцієнтів можна перевірити лише шляхом розв'язання другої задачі, для чого необхідно виконати наступні дії:

1) для всіх можливих вхідних значень X_i (при фіксованому розмірі пароля N) визначити відповідні їм виходи нейронної мережі Y_i ;

2) використовуючи особливості алгоритму А1 (вихідні значення ВН - детерміновані значення, а виходи нейронної мережі, як правило - випадкові величини), визначити ті елементи отриманої множини, які могли використовуватись при навчанні.

Не дивлячись на простоту реалізації, пряма атака практично не має шансів для успішного завершення,

оскільки в загальному випадку нам необхідно виконати для кожного елемента з множини $N_W!$ (навіть при $N_W = 100$ значення $N_W!$ дуже велике) перегляд множини вхідних значень розмірністю 2^{8N} (якщо розмір пароля невідомий - необхідно перебирати усі можливі значення). Так, при $N = 6$ для перегляду множини 2^{48} за умови обчислення 2^{28} значень за секунду необхідно витратити 2^{20} секунд (приблизно 12 діб).

Непряма атака на односторонність хеш-функції, що побудована на основі нейронної мережі, ускладнюється тим, що обчислення хеш-значення потребує значного часу (у порівнянні зі звичайними хеш-функціями). Якщо припустити, що за одну секунду ми обчислюємо одне хеш-значення, то за один рік ми отримаємо приблизно 5(32 значень, що відповідає множині пошуку для пароля лише з трьох символів.

Атака на колізії, як і непряма атака на односторонність, потребує значної кількості обчислень хеш-значень, і, окрім того, значних обсягів пам'яті для збереження усіх отриманих значень. Так, одна з найрозповсюдженіших атак на колізії на основі так званого "парадоксу днів народжень", *Yuval's birthday attack* [1], потребує перегляду множини, щонайменше, з 2^{32} значень при 64-бітному хеш-значенні.

Зазначимо наступні особливості хеш-функції, побудованої на основі нейронної мережі, які суттєво впливають на її стійкість до криптографічних атак:

1) розмірність хеш-значення значно перевищує розмірність пароля;

2) висока варіативність хеш-значень (його елементи - дійсні числа);

3) відсутність кореляції між паролем та отриманим хеш-значенням.

Навчання нейронної мережі - тривалий ітераційний процес, під час якого вагові коефіцієнти можуть змінюватись непередбачуваним чином. Якщо алгоритм навчання детермінований, справедливе наступне твердження: *при навчанні нейронної мережі з фіксованого початкового стану $W_0^{(i)}$ під дією вибірки $\{P_n^{(i)}\}$ можливо отримати лише єдиний вектор вагових коефіцієнтів $W_k^{(i)}$.*

Цілком можливо припустити, що для різних ВН, які відрізняються між собою, щонайменше, одним елементом, будуть сформовані різні вектори вагових коефіцієнтів (хеш-значення).

Наведене припущення цілком підтверджується експериментально. Так, в таблицях 1 та 2 наведені впорядковані вагові коефіцієнти нейронних мереж, ВН для яких отримані з бітових послідовностей, що розрізняються одним бітом, згідно алгоритму А1.

Таблиця 1

0,51234	0,14607	0,62927	0,85102	0,18205	0,50681	0,24067	0,34077	0,8382	0,90916
0,33231	1,0214	0,74528	0,69843	0,0433	0,82496	-0,05825	0,0601	1,54085	0,33197
-0,13682	-4,19317	0,09542	1,04412	0,2716	0,13562	0,67098	0,3514	0,27134	-1,29619
0,17175	0,97441	0,36254	-0,43472	1,4184	0,03722	0,04031	-2,10465	0,33607	0,99036
0,2935	0,13559	1,27981	0,04752	-0,1001	-3,36166	-0,45762	1,47899	-0,06106	0,58145
1,28335	0,21808	0,5875	0,09391	0,30778	0,09375	-0,25891	-0,283	1,30472	-0,1342
0,36256	-1,66202	0,45412	1,03386	0,19483	-0,08338	-0,33273	0,62097	0,57128	-0,14827
0,31271	0,75703	0,05292	-0,48558	-1,93684	-1,1291	-0,56872	-1,80438	-0,61606	-1,23593
0,9893	1,73147	-3,45193	-0,41368	-1,50529	-2,50271	0,67327	-0,87546	-0,00735	-0,49575
-1,68631	-0,35784	-1,47688	-1,63564	-0,54399	-1,29215	2,06221	-1,27381	-3,15713	-1,29431

Таблиця 2

-0,05237	-0,02091	1,2297	0,14685	0,08947	-0,53944	0,41228	-0,63536	0,06413	-0,28997
-0,66642	0,46552	0,0792	2,44598	-0,65746	0,64578	0,87029	0,26291	0,67656	0,29211
0,63001	0,88069	0,63646	0,32328	0,05403	0,1562	0,72625	0,11947	0,49999	-1,06119
0,48973	-0,67508	0,82057	0,19995	0,91589	0,96654	0,43573	0,45742	0,81283	-0,37392
-0,00736	-0,07402	0,9591	-0,25627	-0,08553	-2,62123	0,84268	-0,79203	0,71734	0,5882
0,42955	0,16828	0,61138	0,68442	0,85652	0,6933	0,83348	-0,11884	0,39016	0,21522
-0,07388	0,97894	0,21065	0,87171	-0,40213	0,2995	0,04214	-0,35679	-0,65878	-1,30128
-0,21696	0,21671	0,40031	-0,33596	-0,64845	-0,28264	0,20248	0,49238	-0,29332	-0,08362
0,96739	-1,48781	-0,38346	1,04628	0,10978	1,93547	-0,36319	-0,60765	-0,50592	2,17804
-0,08373	-1,06018	-0,35104	-1,88595	0,33618	0,59676	-0,91448	0,84773	2,92588	-2,33771

Як бачимо, зміна одного біта вхідного пароля призвела до отримання абсолютно різних вагових коефіцієнтів. Відсутність однакових коефіцієнтів у цих двох таблицях, звичайно, не доводить повну відсутність колізій, але, принаймні, зводить до мінімуму можливість їх існування.

ВИСНОВКИ

У даній статті запропоновано використання нейронних мереж для вирішення задачі ідентифікації в системах захисту інформації, орієнтованих на використання в

якості об'єкта ідентифікації пароля користувача. Вхідний пароль в такій системі за допомогою розробленого алгоритму перетворюється у вибірку, яка навчає нейронну мережу. Хеш-значення отримується шляхом перемішування елементів отриманого вектору вагових коефіцієнтів мережі.

Застосування нейронної мережі в такий спосіб забезпечує наступні переваги запропонованої системи:

1) зменшення імовірності існування колізій за рахунок великої надлишковості хеш-значення (без втрати інформативності) та відсутності кореляції між паролем та хеш-значенням;

2) збільшення стійкості до криптографічних атак на односторонність та колізії завдяки тривалому процесу отримання хеш-значення.

Остання особливість хеш-функції набуває особливо важливого значення в наш час, коли розвиток потужностей обчислювальної техніки відбувається швидше за розробку нових алгоритмів шифрування. Сучасні можливості комп'ютерної техніки дозволяють отримувати і зберігати велику кількість обчислених хеш-значень, що робить подальшу задачу розкриття пароля справою кількох годин (на отримання відповідної кількості хеш-значень для запропонованої хеш-функції необхідно буде витратити кілька років).

Недоліком такої системи є можлива непередбачена

поведінка нейронної мережі під час навчання і відсутність конструктивного доведення наведених вище припущень.

ПЕРЕЧЕНЬ ССЫЛОК

1. Menezes A., van Oorschot P., Vanstone S. Handbook of Applied Cryptography, CRC Press 1996.
2. Месси Дж. М. Введение в современную криптографию // ТИИЭР. - 1998. - №5, С. 24-42.
3. Ф. Уоссермен. Нейрокомпьютерная техника. - М.: Мир, 1992.
4. Аведьян Э.Д. Алгоритмы настройки многослойных нейронных сетей // Автоматика и телемеханика. - 1995. - №4. - С.106-118.
5. Горбань А.Н. Обучение нейронных сетей. М.: СП ПараГраф, 1991.

Надійшла 14.03.2000
Після доробки 21.03.2000

УДК 681.32:007

ПОДХОД К МОДЕЛИРОВАНИЮ ЭЛЕМЕНТОВ МАГИСТРАЛЬНОГО ГАЗОПРОВОДА НА ОСНОВЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

С. А. Терехов, Т. А. Мухамадиева, Н. Н. Федорова, Е. В. Диянкова, С. А. Диянкова, А. В. Квичанский, Р. Ф. Мухамадиев

Предложен подход к построению нейросетевой имитационной модели газоперекачивающего агрегата. Методика иллюстрируется на примере многопараметрической системы, обладающей внутренней нелинейностью и временной инерцией отклика.

This Note is devoted to the neural networks modeling of industrial pipeline elements.

ПРОБЛЕМА

Целью данного исследования является изложение теоретических основ нового подхода к созданию имитационных моделей элементов магистральных газопроводов (на уровне отдельного ГПА, компрессорного цеха и, в перспективе, КС), основанных на искусственных нейронных сетях.

Основной смысл нейросетевых методик моделирования столь же прост, сколь и привлекателен, и может быть сформулирован в виде ряда положений:

- Моделируемая сложная инженерная система описывается набором *входных* (независимых и/или управляющих) и *выходных* (зависимых) переменных, а также набором *внутренних* переменных ее состояния;

- Соотношение между входными и выходными параметрами является некоторой, вообще говоря, не известной вектор-функцией (т.н. *системной функцией*), заключающей в себя *все* внутренние особенности системы (например, трудно учитываемые факторы реальной шероховатости труб и особенности конкретного ГПА);

- Эта функция и ее аргументы является экспериментально наблюдаемыми. При этом внутренние переменные могут быть как наблюдаемыми, так и скрытыми от наблюдателя;

- Изменение в (реальном) времени совокупности всех переменных системы составляет ее динамику, определяемую нелинейными связями в системе и внешними условиями - последние включаются в список входных переменных;

- Нейронная сеть *обучается имитировать* эту динамику, т.е. подменять собой системную функцию. Алгоритм обучения индивидуален для выбранной архитектуры нейросети.

Обучение нейросети может проводиться как в режиме *off-line* (т.е. по заранее подготовленной таблице "входы - выходы", при этом каждая строка в таблице отвечает значениям всех переменных в некоторый момент времени), так и *on-line* (нейросеть по известным входам вычисляет прогнозируемые выходы, далее дожидается экспериментальных значений этих выходов, и использует невязку между ними для своего дообучения).

В *on-line* режиме дообучение (*fine-tuning*) может проводиться только если отклик нейросети, в некотором смысле, близок к экспериментальному (т.е. учет дрейфа параметров), в противном случае система должна выдать сигнал об аварийном несоответствии расчета и эксперимента.

ИМИТАЦИОННАЯ МОДЕЛЬ ПРОТОТИПА КОМПРЕССОРНОГО ЦЕХА

Современные устройства, обеспечивающие процесс транспорта природного газа относятся к классу сложных инженерных систем [1]. Функционирование таких систем складывается на основе множества нелинейных физических факторов, среди которых есть как хорошо исследованные факторы [2] (например, физико-химические характеристики газа, механика кранов и др.), так и явления, трудно поддающиеся формальному физико-математическому описанию вследствие многочисленных неопределенностей (например, фактическая шероховатость труб, истечение газа из разрывов и трещин с учетом их реальной геометрии, особенности конкретных газо-перекачивающих агрегатов (ГПА) и пр.).

Перечисленные факторы вступают во взаимодействие, при этом парадоксально, что иногда уточнение математического описания отдельных факторов приводит к ухудшению описания поведения системы в целом. Примеры таких ситуаций можно найти в многочисленной литературе по сложным системам (см, например, [3-4, 17-18] и дальнейшие ссылки в этих книгах).

Одной из альтернатив формальному математическому моделированию систем является имитационное (информационное, кибернетическое, ситуационное) моделирование [17-18]. При кибернетическом подходе исследуемая система заменяется *черным ящиком* (ЧЯ), внутренняя структура которого не обязательно отвечает структуре системы. ЧЯ имеет те же входные и выходные описательные переменные, что и исходная система, и его главной задачей является внешнее подражание (имитация) работе системы, т.е. правильная передача нелинейной функциональной зависимости входов от выходов.

Имитационное моделирование всегда основывается на экспериментальной информации о системе. Поэтому успешное его применение, прежде всего, возможно в приложениях, богатых экспериментальными данными. К таким отраслям относится и эксплуатация магистральных газопроводов.

Процесс *синтеза* ЧЯ часто называют также *идентификацией системы*. Успех при построении ЧЯ определяется не только наличием наблюдаемых данных, но во многом способностью того математического аппарата¹, на котором основан ЧЯ, представить все многообразие откликов системы на многообразие воздействий.

Весьма перспективным аппаратом для построения имитационных моделей являются *искусственные нейронные сети* (ИНС), свойства которых будут более подробно обсуждаться далее. В литературе имеются публикации об успешном использовании нейронных сетей

для прогноза аварий на компрессорных устройствах [5], а также адаптивного управления [14, 19-20] и оптимизации [6].

В работе будет рассмотрен пример построения имитационной модели одной из сложных систем - упрощенной модели группы трех параллельно соединенных ГПА. Имитационная модель основана на адаптивной аппроксимации многомерной системной функции.

Дифференцируемые нейросетевые аппроксиматоры

Построение аппроксимаций функций многих переменных является одной из важных прикладных задач современной математики и информатики. Особую актуальность при этом приобретают дифференцируемые модели, поскольку они позволяют не только имитировать систему, но и оптимизировать ее по управляющим параметрам [6].

Общая тенденция состоит в использовании суперпозиций простых функций одной переменной и их линейных комбинаций. В работе [7] доказывается возможность приближения непрерывных функций многих переменных при помощи суперпозиций линейной функции и произвольной нелинейной функции одной переменной.

К подобным вычислительным схемам относятся и искусственные нейронные сети (*Artificial neural network*), [7-10]. В простейшем случае нейронная сеть задается выражением:

$$N(\vec{x}, \vec{W}) = f\left(\sum_{m=1, \dots, M} W_m \cdot g\left(W_{m0} \cdot 1 + \sum_{k=1, \dots, K} W_{mk} x_k\right)\right),$$
$$f(x) = x, \quad g(x) = 1/(\exp(-x) + 1). \quad (1)$$

Это так называемая нейронная сеть прямого распространения (*feed-forward neural network*) с K входами, одним выходом, единичным порогом и скрытым слоем из M нейронов. На рис. 1 изображена сеть такого типа с двумя входами ($K = 2$) и двумя нейронами на скрытом слое ($M = 2$). Направление стрелок на схеме отражает поток информации и последовательность действий при вычислении функции N .

Параметры \vec{W} определяются путем обучения, т.е. минимизации некоторого функционала, соответствующего решаемой задаче:

$$\min_{\vec{W}} F. \quad (2)$$

1. В качестве такого аппарата можно было бы рассмотреть, например, многомерные ряды Тейлора и др. Однако широко известны возникающие при этом проблемы, в частности повышение порядка полинома с целью получения уточненного описания приводит лишь к точному запоминанию конечного набора состояний системы с растущими осцилляциями вне точек интерполяции. Ситуация катастрофически ухудшается при значительном шуме в данных.

При построении нейросетевой аппроксимации некоторой функции $U = U(\vec{x})$ по конечному набору ее значений $(\vec{x}^{(p)}, U^{(p)})$, $p = 1, \dots, P$, оптимизируется отклонение нейросети от моделируемой функции, например, в сеточной норме $L2$:

$$F = \sum_{p=1, \dots, P} (N(\vec{x}^{(p)}, \vec{W}) - U^{(p)})^2. \quad (3)$$

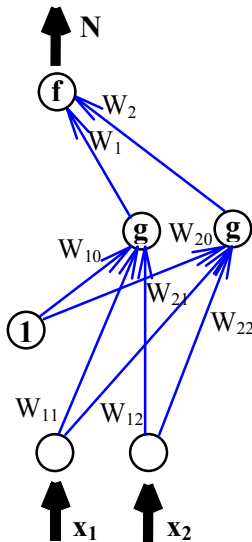


Рисунок 1 - Нейронная сеть прямого распространения со скрытым слоем из двух нелинейных нейронов

Задача (1)-(3) решается при помощи стандартных оптимизирующих алгоритмов. Поскольку производные функции N в (1) по параметрам \vec{W} достаточно просто вычисляются, то для минимизации (3) можно применить одну из градиентных методик, например, метод *BFGS* или алгоритм *Levenberg - Marquardt*, [11-13], а также специализированные методы [15]. При этом вычисление градиента от нейронной сети по ее настраиваемым параметрам - весам связей - требует лишь $O(N)$ операций (конкретно $\sim 3N$) против обычного в таких случаях $O(N^2)$. Это свойство нейронных сетей и определяет всю "скрытую" эффективность нейровычислений.

Важным свойством нейронных сетей прямого распространения (1) является их дифференцируемость по входам, разумеется, при условии достаточной гладкости переходных функций. Это позволяет решать при помощи нейронных сетей задачи совместной аппроксимации функции и ее производных. Для этого в оптимизируемый функционал (2) добавляют соответствующие слагаемые. В последующих разделах данной работы подробно рассматривается один важный прикладной класс аппроксимационных задач этого типа.

Моделируемая система

Рассмотрим некоторую модельную систему, отдаленным прототипом которой служит математическая модель группы ГПА, объединенных по параллельной схеме [2].

Каждый из агрегатов (вместе с системой прилегающих трубопроводов) обладает двумя важными свойствами:

- *Нелинейность отклика*, связанная как с газодинамикой технологических трубопроводов, так и с нелинейностью диаграммы "сжатие-расход" самого ГПА;
- *Запаздывание отклика*, связанное с конечной скоростью распространения сигналов и "инерциальностью" элементов системы и запасенного газа.

В данной схематичной модели подчеркнем лишь эти два свойства, предельно упростив остальную физику процессов, а именно, будем считать постоянными и температуру в системе, и расход газа через каждый ГПА, ограничившись лишь изменениями давления. Более того, из всех нелинейных факторов выделим (весьма условно) лишь один, связанный с зависимостью избыточного давления на выходе ГПА от числа оборотов нагнетателя, для которой примем нелинейный характер, изображенный ниже на рис. 2.

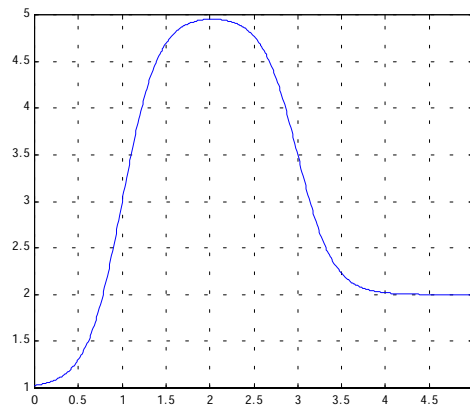


Рисунок 2 - Модельная нелинейная зависимость эффективности ГПА по избыточному давлению от относительной скорости вращения вала

Все ГПА группы будут различаться по параметрам этой зависимости - положению точки максимальной эффективности, положению точки выхода на насыщение (условно связанного с включением петли обратной связи антипомпажной защитой). Эти параметры каждого ГПА будем считать *скрытыми* (внутренними параметрами состояния системы). Единственным управляющим параметром будем считать число оборотов вала (в безразмерных единицах).

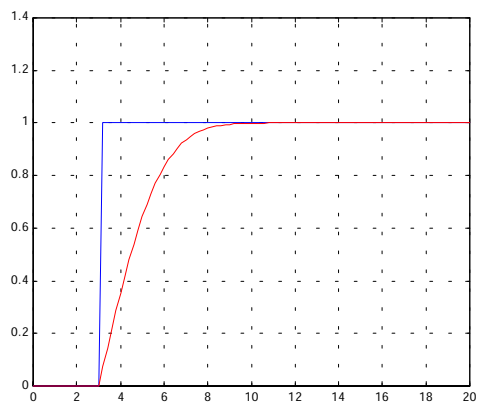


Рисунок 3 - Задержанный отклик с характерным временем $\tau=3$ на ступенчатый сигнал

В качестве модели внутренней инерционности системы "ГПА-технологические трубопроводы" примем функцию отклика в форме линейного фильтра с задержанным временем τ :

$$P_k^{out} = \sum_{m=1}^n b_m P_{k-m}^{in},$$

где b - нормированная Гауссова функция (дискретная) с затуханием τ .

Пример преобразованной с задержкой в 3 временные единицы ступенчатой функции Хевисайда показан на рис. 3.

Таким образом, моделируемый отклик j -го ГПА (выходное давление) на внешнее воздействие (входное давление) дается функцией:

$$P^j(t) = D(P_{in}^j(t) \cdot K(n_j)),$$

где D - функция-фильтр задержки, а K - коэффициент увеличения давления (сжатия), n - обороты нагнетателя.

Еще раз отметим, что функции D и K содержат "скрытые" переменные и зависимости - время и форма задержки, параметры и форма кривой эффективности. Ниже в модельных вычислительных экспериментах эти зависимости предполагается считать "дрейфующими" во времени.

Для параллельной схемы объединения N ГПА (для определенности, следуя [2], положим $N = 3$) и сделанных модельных предположений давление на выходе из системы будет равно среднему значению выходных давлений каждого агрегата (т.е. сумме парциальных

давлений равных поступающих порций газа массой $1/N$ с одинаковой температурой в заданном объеме выходного коллектора).

Данные

Будем далее считать данные, выдаваемые описанной в предыдущем пункте моделью экспериментальными, для чего добавим 5% случайную ошибку к входным и выходным давлениям. При этом скрытые параметры модели будем считать не доступными наблюдателю. Для простоты все наблюдаемые физические величины будем измерять в безразмерных единицах.

Рассмотрим динамический процесс эксплуатации моделируемой системы в течение трех суток. Во времени изменяются:

- *Внешние параметры* - входное давление P_0 . Будем считать, что изменение входного давления является случайным Пуассоновским процессом [16] со средним временем t_p и амплитудой колебаний a_p (т.е. кроме 5% шума давление в среднем раз в t_p секунд принимает новое случайное значение в диапазоне $1 \pm a_p$);

- *Управляющие параметры* - обороты вала n_j для каждого нагнетателя. Будем считать, что управляющие воздействия также не предсказуемы и описываются Пуассоновским процессом со средним временем t_n и амплитудой колебаний a_n около своих номинальных значений (масштаба единицы);

- *Внутренние параметры системы* - времена задержек τ_j и положения точек максимальной эффективности компримирования m_j (т.е. точек максимума на рис.2). Будем считать, что времена задержек также испытывают Пуассоновский дрейф с соответствующими параметрами t_τ , a_τ (дрейф задержек моделирует уменьшение инертности при мелких утечках и не учитываемые в задаче колебания расхода).

В итоге, все величины представляются Пуассоновскими временными рядами, по значениям которых вычисляется выходное давление P , на которое далее налагается 5% шум. Полученное значение и будет рассматриваться, как экспериментально измеренное.

ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ С НЕЙРОСЕТЕВЫМИ МОДЕЛЯМИ

По методике, описанной в предыдущем пункте, были получены временные ряды для параметров модели. Для сокращения времени вычислений при обучении нейросети 3-х суточные ряды описывались с частотой сэмпинга 5 секунд.

Таблица 1 - Значения параметров модели

Номинальное значение входного давления P_0	1
Амплитуда колебаний входного давления a_p	0,2
Пуассоновское время колебаний входного давления t_p	3 часа
Амплитуда шума "измерений" входного давления	0,05
Амплитуда шума "измерений" выходного давления	0,05
<i>Агрегат 1</i>	
Максимум коэффициента компримирования (P_{out}/P_{in})	5
Отн. обороты при максимуме (P_{out}/P_{in})	2
Насыщенное значение (P_{out}/P_{in})	3
Отн. обороты при насыщении (P_{out}/P_{in})	4
Номинальные обороты n	2
Амплитуда управляющих колебаний номинальных оборотов a_n	0,2
Пуассоновское время колебаний n t_n	1 час
Номинальное время инерциального запаздывания τ	10 мин
Амплитуда колебаний времени инерциального запаздывания a_τ	2 мин
Пуассоновское время колебаний τ t_τ	1 час
<i>Агрегат 2</i>	
Максимум коэффициента компримирования (P_{out}/P_{in})	4,5
Отн. обороты при максимуме (P_{out}/P_{in})	3
Насыщенное значение (P_{out}/P_{in})	2,5

Отн. обороты при насыщении (P_{out}/P_{in})	4
Номинальные обороты n	2
Амплитуда управляющих колебаний номинальных оборотов a_n	0,2
Пуассоновское время колебаний n t_n	1 час
Номинальное время инерциального запаздывания τ	12 мин
Амплитуда колебаний времени инерциального запаздывания a_τ	2 мин
Пуассоновское время колебаний τ t_τ	1 час
<i>Агрегат 3</i>	
Максимум коэффициента компримирования (P_{out}/P_{in})	5,5
Отн. обороты при максимуме (P_{out}/P_{in})	2
Насыщенное значение (P_{out}/P_{in})	3
Отн. обороты при насыщении (P_{out}/P_{in})	5
Номинальные обороты n	2
Амплитуда управляющих колебаний номинальных оборотов a_n	0,2
Пуассоновское время колебаний n t_n	1 час
Номинальное время инерциального запаздывания τ	8 мин
Амплитуда колебаний времени инерциального запаздывания a_τ	2 мин
Пуассоновское время колебаний τ t_τ	1 час

Численные значения параметров, использовавшихся при моделировании, приведены в таб. 1. Примеры временных рядов, порождаемых исследуемой моделью, представлены ниже на рис. 4.

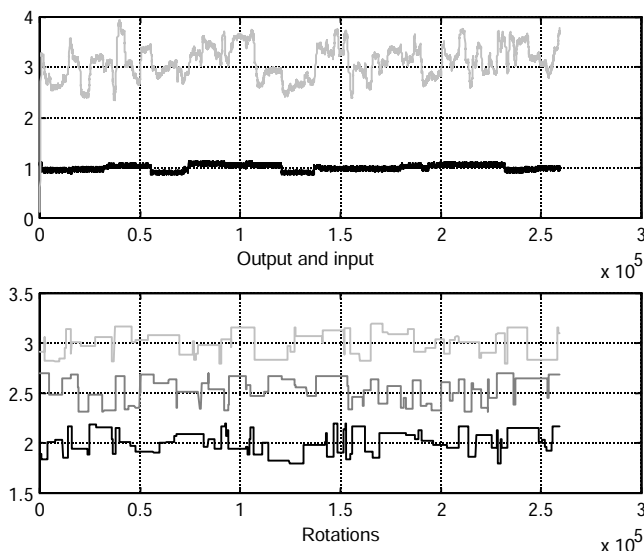


Рисунок 4 - Временные (в секундах) зависимости входного/выходного давления и оборотов валов каждого из трех агрегатов (в условных единицах)

В численных экспериментах отрезок ряда, отвечающий первым двум суткам, использовался для обучения нейросети, а оставшиеся сутки - для тестирования.

Простой нейросетевой предиктор

В первой группе экспериментов использовалась нейронная сеть, имеющая четыре входа (входное давление и три значения оборотов валов нагнетателей) и один выход - выходное давление. Обучение проводилось алгоритмом *Rprop* [15]. Результаты обучения и тестирования обученной сети приведены на рис. 5

Видно, что нейросистема довольно быстро и хорошо "усвоила" обучающие данные (верхний рисунок) и в целом неплохо справилась с прогнозированием выходного давления в течение последних суток (напомним, что данные для третьих суток никак не использовались при обучении). Однако в ряде ситуаций прогнозируемое значение оказывалось ниже или выше экспериментальных наблюдений, уклоняясь на величину до 15-20%, что ощутимо превышает используемую "экспериментальную" погрешность (5%). Кроме того, прогноз сильно зашумлен.

Существует несколько способов повышения точности прогноза¹. Здесь мы остановимся только на одном из них - использовании ранее измеренных экспериментальных значений выходного давления в качестве дополнительных входов в нейросеть.

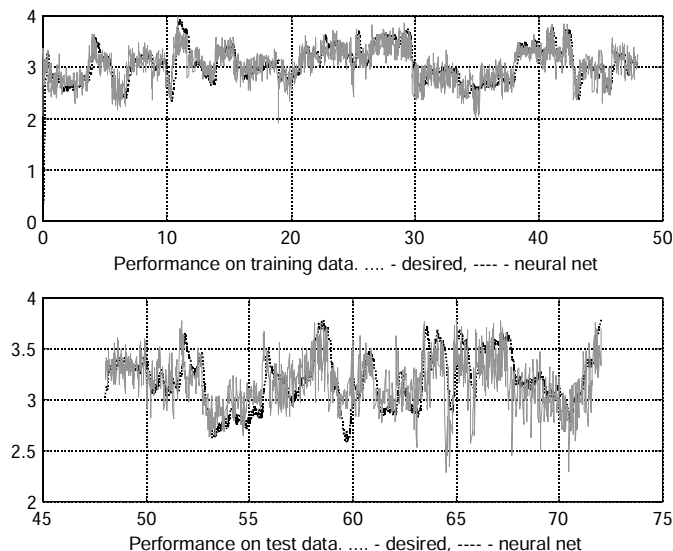


Рисунок 5 - Результаты обучения и тестирования нейронной сети. Сплошные (серые) линии - выходы нейросети, пунктирные (черные) линии - экспериментальные значения

Предиктор, использующий предысторию выходных значений

Рассмотренная в предыдущем пункте нейросетевая модель работала в "слепом" режиме, т.е. поступающая текущая экспериментальная информация никак не учитывалась.

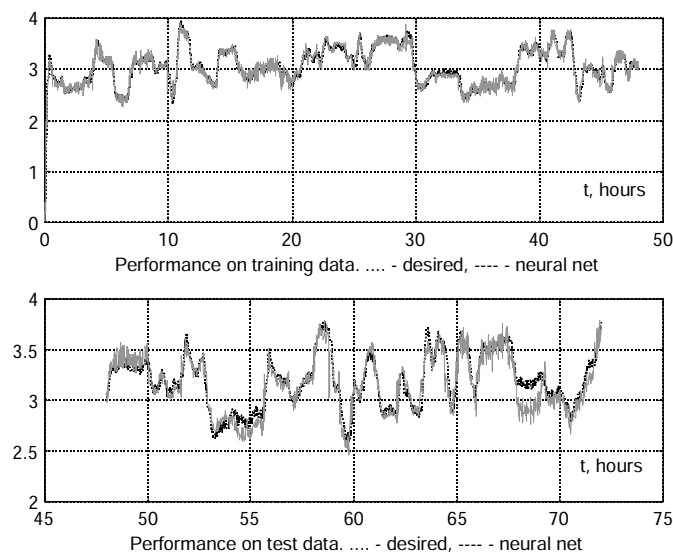


Рисунок 6 - Обучение и тестирование нейронной сети, учитывающей последние наблюдения. Сплошные (серые) линии - выходы нейросети, пунктирные (черные) линии - экспериментальные значения

1.Заметим, что в данном исследовании оптимизация топологии нейросети не производилась.

Однако если в процессе обучения и эксплуатации нейросети подавать на вход не только истинно входные переменные (давление на входе и обороты валов), но и доступные измерения выходного давления в несколько *предыдущих* моментов времени, то адаптивность модели можно значительно повысить.

В проведенных экспериментах использовались 3 предыдущих значения давления, таким образом, нейросеть имела один выход и 7 входов: $P_0(t)$, $n_1(t)$, $n_2(t)$, $n_3(t)$, $P(t-dt)$, $P(t-2dt)$, $P(t-3dt)$.

Итогом явилось не только значительное повышение точности прогноза (до 5%), но и кардинальное сокращение шума в прогнозируемых данных (см. рис. 6).

При обучении и тестировании нейросети замерялось также время выполнения вычислений (для ЭВМ Pentium II-300). На этапе тестирования (т.е. в эксплуатационном режиме) время отклика составило 0.02 сек в расчете на один пример, а при обучении - примерно втрое больше. Это предварительно позволяет говорить о десятикратном запасе по отношению к требуемому времени сбора данных ~0.5 сек.

Обсуждение предварительных результатов

Проведенные вычислительные эксперименты показывают, что нейросетевая имитационная модель справляется с главной задачей - представление функции выходных переменных от входных без информации о скрытых параметрах. В процессе обучения нейроны скрытых слоев нейросети автоматически формируют представление о скрытых переменных модели, поскольку вариации этих скрытых переменных отображаются в изменениях выходных переменных.

Учет предыстории выходов в качестве дополнительных входов в нейросеть значительно улучшает качество ее работы.

Для создания промышленной нейросетевой модели элементов магистрального газопровода потребуются решить две главные группы вопросов:

- Получить экспериментальные данные о реальной системе (КЦ, КС), либо надежные данные из расчетов по реалистичным физико-математическим моделям;
- Провести комплексное исследование по выбору и

оптимизации архитектуры нейросети и наборам учитываемых входных и выходных переменных.

Данные направления составят предмет последующих исследований. Авторы благодарят Селезнева В.Е., Анишимова А.И. и Сергеева Е.А. за полезные обсуждения.

ПЕРЕЧЕНЬ ССЫЛОК

1. С.А.Терехов. Нейросетевые информационные модели сложных инженерных систем. Глава 4 в кн. А.Н.Горбань, В.Л.Дунин-Барковский, А.Н.Кирдин, Е.М.Миркес, А.Ю.Новоходько, Д.А.Россиев, С.А.Терехов, М.Ю.Сенашова, В.Г.Царегородцев. Нейроинформатика. Новосибирск, Наука, 1998.
2. www.licenergy.com
3. А.Ю. Лоскутов, А.С. Михайлов. Введение в синергетику. М. Наука, 1990
4. Г. Шустер. Детерминированный хаос. М. Мир, 1988.
5. W.W. Armstrong, C. Chu, M.M. Thomas. Feasibility of using Adaptive Logic Networks to Predict Compressor Unit Failure. Proc. Battelle Pacific Northwest Laboratories Workshop on Environmental and Energy Applications of Neural Networks, Richland WA, USA, March 30-31, 1995. <http://www.dendronic.com/papers/Alncomp.zip>
6. J. Booker, et al. A Rigorous Framework for Optimization of Expensive Function by Surrogates. NASA ICASE Tech Rep. 98-47, 1998
7. Gorban A.N. Approximation of continuous functions of several variables by an arbitrary nonlinear continuous function of one variable, linear function and their superpositions. Appl. Math. Lett. Vol. 11, No. 3, pp. 45-49, 1998.
8. Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере. Новосибирск: Наука, 1996. - 276 с.
9. Ф.Розенблатт. Принципы нейродинамики. М.: Мир, 1965.
10. Minsky M.L., Papert S. Perceptrons. MIT Press, Cambridge, MA, 1969
11. Дэnnис Дж., Шнабель Р. Численные методы безусловной оптимизации и решения нелинейных уравнений. - М.: Мир, 1988.
12. Levenberg, K. A method for the solution of certain nonlinear problems in least squares. Q. J. Appl. Math., v.2, 1944, p.164-168
13. Marquardt, D.W. An algorithm for the estimation of non-linear parameters, SIAM J., v.11, 1963, p.431-441.
14. Т.Кохонен. Self-organizing Maps. Springer, 1995
15. Программно-аппаратный комплекс NIMFA для нейросетевого моделирования. Отчет РФЯЦ-ВНИИТФ, г. Снежинск, 1997.
16. Е.С.Вентцель, Л.А.Овчаров. Теория случайных процессов и ее инженерные приложения. М. Наука, 1991.
17. Р.Шеннон. Имитационное моделирование систем - искусство и наука. М. Мир, 1978.
18. А.Прицкер. Введение в имитационное моделирование и язык SLAMII. М.Мир, 1987.
19. E.D.Sontag. Neural Nets as System Models and Controllers. Proc. Seventh Yale Workshop on Adaptive and Learning Systems, pp. 73-79, Yale University, 1992. http://www.math.rutgers.edu/~sontag/FTP_DIR/yale.ps.gz
20. E.D.Sontag. Some Topics in Neural Networks and Control. Siemens Corporate Research Tech Rep., 1993. http://www.math.rutgers.edu/~sontag/FTP_DIR/93ecc-nn.ps.gz, http://www.math.rutgers.edu/~sontag/FTP_DIR/neural-nets-siemens.ps.gz

Надійшла 03.03.2000

НЕЙРОКОМПЬЮТЕР "ЭМБРИОН" КАК ГЕНЕРАТОР И ПРЕОБРАЗОВАТЕЛЬ СМЫСЛОВ В ВЕРОЯТНОСТНОЙ МОДЕЛИ ЛИЧНОСТИ

В. Д. Цыганков

Розкрито сутність взаємовідносин моделей свідомості та фізичної реальності Налімова та Цыганкова на основі нейрокомп'ютера "Ембріон".

Раскрыта сущность взаимоотношений моделей сознания и физической реальности Налимова и Цыганкова на основе нейрокомпьютера "Эмбрион".

The essence of relationships between the models of consciousness and physical reality of Nalimov and Cyganov on the basis of neural computer "Embrion" is described.

1 ПОСТАНОВКА ЗАДАЧИ

В работе [1] была показана общность квантовой теории измерений (КТИ) и вероятностных динамических процессов, происходящих в нейрокомпьютере (НК) "ЭМБРИОН" [2].

В свою очередь, ПЕРЕВОЗСКИЙ Т.А. в [3] показал общность вероятностной модели смыслов (ВМС) НАЛИМОВА В.В., или его смысловой модели Личности [4], и квантовой теории измерений в виде таблицы 1.

Здесь в функции $\Psi(q, t)$ время $t = t_0$ - "замороженное время", в течение которого производится измерение, многократные статистические испытания, а астрономическое время t - это эволюционный параметр.

Соответствием вида $p(\mu) = |\Psi(q)|^2$ ПЕРЕВОЗСКИЙ Т.А. показывает возможность квантово-механической интерпретации процессов мышления.

Естественно, остается, в логическом треугольнике (рис.1) имеющих трех моделей Сознания и физической реальности, раскрыть существо третьей стороны этого треугольника, показанной на рисунке пунктиром.

Для этого, во-первых, необходимо более компактно изложить существо вероятностной модели смыслов (ВМС) НАЛИМОВА, или его смысловую модель Личности [4], а во-вторых, показать соответствие между параметрами смысловой вероятностной модели НАЛИМОВА [4] и параметрами модели НК "ЭМБРИОН" [2], который может быть как генератором, так и преобразователем смыслов, что, по НАЛИМОВУ, эквивалентно модели Сознания, т.к. он утверждает и доказывает, что "Сознание - это преобразователь смыслов".

Таблица 1

№ п/п	ВМС	КТИ
1.	μ - множество смыслов, переменная, степень свободы текста	q - координата, степень свободы квантового объекта
2.	$p(\mu/y)$ - текст, состоящие субъекта	$\Psi(q, t)$ - функция состояния объекта
3.	$p(\mu)$ - функция вероятности, спонтанная "распаковка" смыслов, обнаружение значения μ	$ \Psi(q) ^2$ - функция вероятности, спонтанное обнаружение значения q измеряемой величины
4.	y - фильтр, текст, с которым взаимодействует исходный текст $p(\mu)$	y - фильтр, измерительный прибор, дающий измерение q
5.	$p(\mu/y) = kp(\mu)p(y/\mu)$ - преобразование весовой функции смыслов $p(\mu)$	$\Psi_0(q) \Rightarrow \Psi_n(q/y)$ - преобразование функции состояния $\Psi(q)$ в процессе измерения

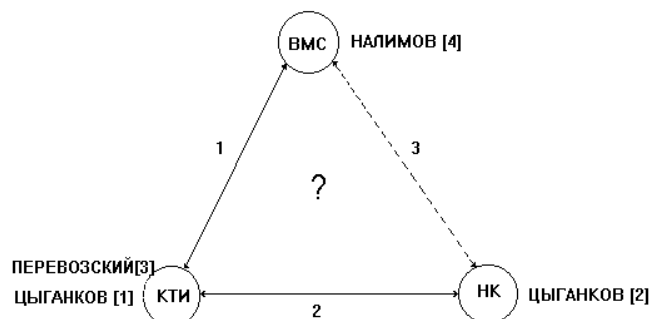


Рисунок 1- Три модели сознания и физической реальности

Здесь мы вплотную соприкасаемся с взаимодействием двух различных по субстрату реальностей: СМЫСЛОВ и МАТЕРИИ, однако, которые имеют общую, единую основу.

НАЛИМОВ В.В.[4], для описания семантических и смысловых (или семиотических) категорий, ввел язык вероятностных образных представлений, опираясь на убежденность А.Н.КОЛМОГорова в том, что "вероятностное мышление - это преимущество человека".

И всё БЫТИЕ наше НАЛИМОВ представляет как ЯЗЫК или ТЕКСТ, это вероятностный поток смыслов, а смыслы порождаются воображением и открываются в языке.

Человек, по НАЛИМОВУ, - это текст.

В [2] показано, что нейрокомпьютер "ЭМБРИОН" можно представить в виде вероятностного конечного марковского автомата, или автомата с переменной структурой, как вероятностного преобразователя множества входных слов (X) во множество выходных слов-реакций (Y) в некотором алфавите.

2 СМЫСЛ РЕАЛЬНОСТИ СОЗНАНИЯ

По НАЛИМОВУ [4], реальность двойственна: она дискретна, как знаковая или семиотическая система, и континуальна, как семантическая система. Он определяет Личность как случайный процесс, при котором ТЕЛО (все атомы) меняется и является условием сохранения ОБРАЗА (мифа, фантома) Личности, инвариантности нашего Сознания. Это образ внутри нашего Сознания и во внешней среде. Этот образ - не телесное начало, но обладает субстанцией и формой. Это есть "ДУША" (b_a обитающая в k_a или "двойник" вычлененной индивидуальности, по "Египетской книге мертвых"). Образ может свободно перемещаться в пространстве, отделяя или соединяя себя с телом.

Человек как Личность, или единство многомерной реальности, согласно воззрениям древнего Египта, состоит из девяти компонент [4]:

- материальное, естественное тело;
- духовное тело;
- сердце;
- двойник (k_a);
- душа ($b_a \in k_a$);
- тень (khaibit)
- эфирная невидимая или светящаяся оболочка-нимб, дух (khu);
- "форма" на небе, сила (sekhem) - "пустота есть форма";
- имя (gen).

НАЛИМОВ отстаивает принцип космизма Сознания по происхождению - это падение в материю СВЕТА. Душа и Дух - части первичной жизни или Света. Жизнь в Космосе направлена не на индивидуальность, а на

ЦЕЛОЕ. "Единое" - это трансцендентное начало [5].

Фундаментальные вопросы природы Человека, на которые многие, и автор, в том числе в [1],[5], пытаются дать ответы:

- являются ли Личность и Мир вечными и конечными?

- являются ли Душа и Тело идентичными?

Наша жизнь - это "самсара" или процесс жизни-умирания (невязка $J \rightarrow 0$ в нейрокомпьютере [2] убывает с момента "рождения" $U = 0$).

"Я" (Душа, Личность, индивидуальность) есть граница организма, его капсула, принадлежащая одновременно и организму и среде в результате их контакта. "Я" существует в момент взаимодействия или, для НК "ЭМБРИОН", она существует при

$$J = P_0 - S_j > 0,$$

где J - вязка-вектор, P_0 - состояние регистра внутренней памяти, S_j - состояние среды, отраженное на j -ой строке сенсорной матрицы (СМ).

НАЛИМОВ постулирует идею сверхединой теории поля, объединяющей две различные реальности: Физический мир и Семантический мир, через ГЕОМЕТРИЮ МИРА, в отличие от космологической теории БОМА, объединяющей Физический космос и Сознание, как реальности поля Бытия и поля Сознания [4].

Сознание - это, по ТАРТУ (1981г.), системный эффект взаимодействия В-системы и М/L-системы, где В - это Тело, материя, сома, М-mind - это разумное начало, Ум, а L-life - это Жизнь, действие.

К.УИЛБЕРГ (1971г.) выделяет пять уровней в спектре Сознания или "философии вечного":

- уровень Ума (Mind), разум;
- трансперсональные полосы;
- уровень ЭГО;
- экзистенциальный уровень;
- уровень тени.

Сознание - это владение смыслами и умение их изменять, умение управлять ими или преобразовывать их.

Сознание, не обладающее смыслами, не имеет языка как семантически насыщенной знаковой системы и называется квазисознанием [4].

3 ВЕРОЯТНОСТНОЕ ИСЧИСЛЕНИЕ СМЫСЛОВ

Напомним еще раз, что Сознание - это текст, а текст - носитель смыслов, гибкий, подвижный, динамичный, изменяющийся, эволюционирующий, создающийся заново.

Сознание - это трансцендирующее устройство, связывающее разные Миры, наш обычный мир и Космос (Мир МЕТАСЕМАНТИКИ).

Создавая новые тексты, Сознание порождает НОВЫЕ

МИРЫ.

Многоуровневая карта или блок-схема функционирования Сознания (по НАЛИМОВУ [4]) приведена на рис.2 и во многом напоминает блок-схему НК "ЭМБРИОН", изображенную на рис.3, где уровни карты Сознания на рис.1 следующим образом соотносятся с уровнями реальности, реализованными в нейрокompьютере,:

I - уровень логического построения выходного распределения вероятностей или уровень текста, как смысловой реакции (Y);

II - уровень виртуального вероятностного поля (Ψ), генерируемого с помощью байесовской логики при задании невязки J , или фильтра $p(y/\mu)$, и гипотезы восприятия (U);

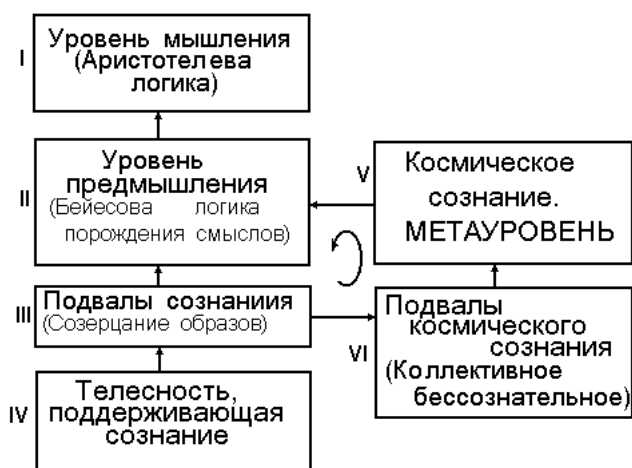


Рисунок 2 - Многоуровневая карта (блок-схема) функционирования трансцендирующего устройства - Сознания (по НАЛИМОВУ)

III - уровень взаимодействия кодов регистров P_0 и S_j -й строки СМ, уровень созерцания образов невязки J ;

IV - уровень тела или физический уровень полей регистров внутренней памяти P_0 и S_{jk} - регистров СМ, генератора шума (ГШ), полей регистров Блока Выдвижения Гипотез (БВГ);

V и VI - уровни формирования вектора ГИПОТЕЗЫ ВОСПРИЯТИЯ (U) или фильтра $p(y/\mu)$.

Вероятностное исчисление смыслов - это язык описания вероятностных процессов на втором уровне предмышления.

Наш эволюционирующий, динамический мир - это множество текстов. Тексты эти также имеют двойственную природу: они дискретны семиотически и континуальны семантически, т.е. имеют вероятностную структуру $p(\mu)$ смыслов, изображенную на рис.4.

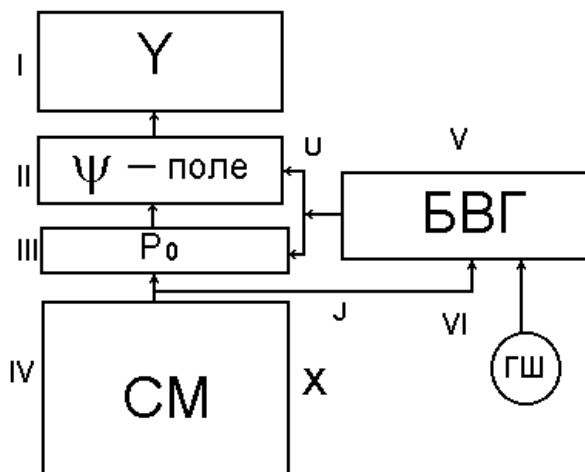


Рисунок 3 - Блок-схема нейрокompьютера "ЭМБРИОН"

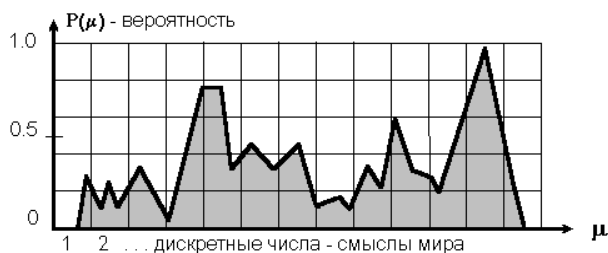


Рисунок 4 - Вероятностная структура смыслов, 0 - семантический вакуум, $p(\mu)$ - функция распределения смыслов μ

Семантический вакуум - это нераспакованный, непроявленный мир, это всё возможное множество $2^n \times 2^n$ слов-текстов. Распаковывание текстов, их проявление или генерация текстов, творчество - это навешивание меры $p(\mu)$ на участки μ смысловой шкалы с заданной на ней метрикой. Функция $p(\mu)$ - это "ОКНО", через которое мы или "ЭМБРИОН" смотрим в Семантический мир.

Изменение текста, его эволюция связана со спонтанным появлением в ситуации y фильтра $p(y/\mu)$ мультипликативно взаимодействующего с исходной ситуацией-текстом $p(\mu)$ в соответствии с байесовой логикой по формуле

$$p(\mu/y) = kp(\mu)p(y/\mu),$$

где $p(\mu/y)$ - функция распределения вероятностей, определяющая семантику нового текста, возникающего после толчка y , появления условия или точки бифур-

кации в нейронной сети НК, а также при смене S_j на S_{j+1} строки сенсорной матрицы (СМ);

$p(y/\mu)$ - фильтр;

k - константа нормировки.

Как видно из вышеизложенного, из двух исходных посылок, $p(\mu)$ и $p(y/\mu)$, следует новая $p(\mu/y)$ - новый текст с новой семантикой. Здесь под $p(\mu)$ следует понимать вероятность кода P_0 в НК "ЭМБРИОН", а код P_0 - есть смысл исходного текста Y_0 ; $p(y/\mu)$ - есть вероятность кода j -й строки СМ. Под j будем понимать номер спонтанного толчка или смены условия y . Это условная функция распределения случайной величины y при заданном μ ; $p(\mu)$ - плотность нормированной в интервале $[0,1]$ вероятностной меры или априорное знание.

Смыслы можно соизмерять, если они упорядочены по оси μ .

Семантическое поле квантуется путем задания на нем распределяемой числовой меры. Каждый семантический квант - это слово, которое содержит весь семантический потенциал, вероятностно взвешенный различным способом. Слово имеет смысловую размытость.

Язык открыт для спонтанной перестройки. Фильтр $p(y/\mu)$ появляется случайно или спонтанно. Он сужает словарный смысл слова $p(\mu)$ в ситуации y . Эволюция текстов данного вероятностного языка не связана с физическим пространством-временем. Семантическое пространство многомерно, чисто логическое, с логическими операциями над смыслами.

Если задано распределение вероятностной меры $p(\mu)$, то имеется семантическое поле.

Сознание - триада: смысл + текст + язык.

Архитектоника Личности является архитектоникой смыслов.

Функция $p(\mu)$ выделяет актуальные смыслы и передает их на первый логический уровень мышления. Его телесной, физической реализацией в технике может быть обычный персональный компьютер (ПК), а второй и третий уровни предмышления и созерцания образов хорошо реализуются на базе НК "ЭМБРИОН".

При изменении ситуации y или образа X на СМ, происходит перераспределение весов с $p_1(\mu)$ на $p_2(\mu)$ значимости смыслов с помощью гипотезы U из БВГ. Новым ситуациям y , новому образу на СМ, соответствует процесс порождения новых фильтров $p(y/\mu)$ и процесс генерации нового текста. Это есть реализация АКЦЕПТОРА ДЕЙСТВИЯ в функциональной системе П.К АНОХИНА [6], забегание вперед себя или "опережающее отражение действительности", разновидность "ясновидения", с помощью выбора ГИПОТЕЗЫ

ВОСПРИЯТИЯ U в НК "ЭМБРИОН".

Происходит распаковка еще скрытых малыми весами $p(\mu)$ смыслов.

$p(\mu)$ - это взаимосвязь органов и их функций в организме, обеспечивающая его целостность и индивидуальность, а порождающие фильтры $p(y/\mu)$ задают его эволюцию.

4 СМЫСЛОВАЯ ПРИРОДА ЛИЧНОСТИ

Личность - это носитель смыслов, генератор смыслов и их преобразователь. Она телесно капсулирована, т.е. смыслы имеют материальные границы. Семантическая модель Личности НАЛИМОВА [4] имеет следующие четыре уровня-представления или проявления себя во вне:

- ЭГО,
- МетаЭГО,
- Многомерная Личность,
- Гиперличность.

ЭГО - вероятностная проявленность $p(\mu)$ семантического поля, т.е. это индивидуальность человека. Его внутренний образ, определенный системой ценностных представлений (U), которая дает распределение плотности вероятностей на семантической шкале m . ЭГО - это индивидуалистическая структура текста, потока слов.

Функция $p(\mu)$ может быть:

- размытой,
- иглоподобной,
- асимметричной,
- многовершинной со стабильным центром рассеяния во "время жизни" ($T_{ж}$), времени генерации текста.

Нам присущ "весь семантический генотип". Мы в настоящем (n) несем в себе как прожитое прошлое (p), так и не реализованное еще будущее (b) - акцептор действия [6] (рис.5).

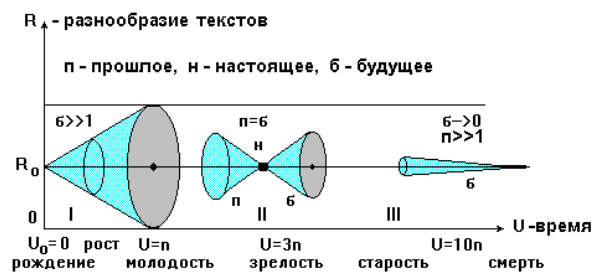


Рисунок 5 - Световые конусы в динамическом семантическом поле НК

Здесь p - прошедшее, прошлое;

n - настоящее в момент U_k собственного, внутренне-

го, "личного времени";

U - собственное, внутреннее время;

b - будущее;

t - астрономическое время.

Геометрическая форма смысла или вероятно заданная структура "Я"- это семантическая капсулированность бесформенного "НИЧТО" Г.И.ШИПОВА [7].

МЕТАЭГО - это ИСТОЧНИК личностных изменений, т.е. гипотеза восприятия (U) БВГ в НК. Он имеет вселенский характер, открыт вселенскому за пределами. МетаЭГО - это способность к генерации фильтров и эволюции (рис.6).

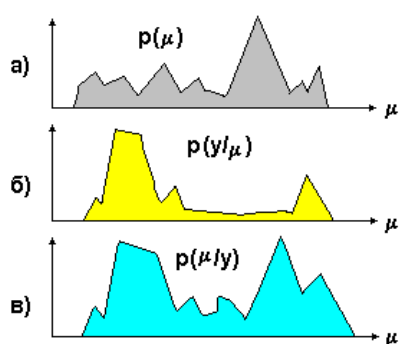


Рисунок 6 - Эволюция индивидуальности

На рисунке а) - это "Я" как исходный семантический портрет;

б) - это $p(y/\mu)$ - фильтр интерпретации и предпочтения в условиях y , эволюционный толчок;

в) - индивидуальный портрет после интерпретации, измененное "Я".

Эволюция под воздействием МетаЭГО протекает следующим образом:

$$p(\mu) \rightarrow p_1(\mu/y) \rightarrow p_2(\mu) \rightarrow p_3(\mu/y) \rightarrow p_4(\mu) \dots$$

Силлогизм БАЙЕСА содержит в себе концепцию времени (ЦЫГАНКОВ [1]). На рис.5 $p(\mu)$ - это прошлое (μ), спонтанный выбор фильтра $p(y/\mu)$ в точках бифуркации в момент U_k - это настоящее (μ), а будущее - это $p(\mu/y)$ в момент U_{k+1} , где П,Н,Б - модусы времени.

Прошлое (μ) свертывается по будущему (b) (АНО-ХИН [6]) для смыслов в настоящем (μ). В этом проявляется себя МетаЭГО.

Время имеет дологический уровень, "оно завихряется", как утверждает НАЛИМОВ. Нет линейного времени (смотри [1]) с причинно-следственными связями. У нас, в нейροкомпьютере "ЭМБРИОН", время U - это

собственное или "личное время" имеет вид спирали [1].

Установка (Y), по УЗНАДЗЕ Д.Н.[8], от которой зависит восприятие, связана с прошлым (μ) опытом (состоянием регистра P_0) и от условий или ситуации в настоящем (μ) (состояние-код j -й строки СМ).

Невязка (J) как вектор - это фильтр содержания Сознания. Стохастический процесс взаимодействия $p(\mu)$ и $p(y/\mu)$ протекает в течение "времени жизни" ($T_{ж}$) возбуждения в нейронной сети нейροкомпьютера.

МНОГОМЕРНОСТЬ ЛИЧНОСТИ - это семантическое пространство вида $[p(\mu) \times \mu]^n$. Двумерное ЭГО в виде поверхности записывается следующим образом:

$$Z = p(\mu_1, \mu_2),$$

где веса при распаковке $p(\mu_1/y, \mu_2/y)$ задаются площадью. Мы "смотрим" на мир через несколько связанных "окон".

"Окнами" в одном НК являются гипотезы U_1, U_2, U_3, \dots по строкам 1,2,3 сенсорной матрицы (СМ).

Можно модель двумерной личности реализовать на двух ортогональных нейροкомпьютерах "ЭМБРИОН" с одной строкой на их сенсорных матрицах. При отсутствии корреляции между S_1 и S_2 , мы имеем две независимые Личности в одном теле или пример раздвоения личности при патологиях мозга или при гипнозе. Здесь имеет место множественная персональность [9].

При наличии корреляции $k(\mu_1, \mu_2)$ не равной нулю между смысловыми осями μ_1 и μ_2 задается вероятностная структура упорядоченности смыслов. При $k(\mu_1, \mu_2) = 1$ имеет место вырождение двумерности.

Личность - это процесс, неустойчивое равновесие (БАУЭР Э.С.[10].

Коэффициент корреляции $k(\mu_1, \mu_2)$ гибко изменяется в процессе жизни, давая возможность Личности адаптироваться в среде.

Саморащепление Личности - это внутренний диалог с самим собою. Он, кроме человека, доступен нейροкомпьютеру, и недоступен обычному ФОН-НЕЙМАНОВСКОМУ компьютеру с аристотелевой логикой. Здесь перед исследователями открываются невообразимые возможности моделирования на НК измененных состояний Сознания (ИСС) - "просветленного сознания", сновидений, дуальных состояний, семантической амнезии и др. [4].

Расширение Сознания происходит за счет перехода к многомерности $p(\mu_1, \mu_2)$, а не за счет размазанности $p(\mu)$.

Мозг - это неравновесная система с переменной структурой [9].

ГИПЕРЛИЧНОСТЬ - это двумерная вероятностная функция $p(\mu_A, \mu_B)$, где А и В - это разные Личности. Функция $p(\mu_A, \mu_B)$ одновременно локализована в двух телах. Имеет место трансперсональное взаимодействие, единство или образование гиперповерхности при коэффициенте корреляции близком к 1. Примером ГИПЕРЛИЧНОСТИ может быть семантическая доминанта А.А.УХТОМСКОГО при гипнозе. Гипнотизер является $p(y/\mu)$ - фильтром с индивидуальностью $p(\mu_B)$, которая вытесняет исходную $p(\mu_A)$. Имеет место вероятностный процесс в семантическом поле.

ГИПЕРЛИЧНОСТЬ - это многомерная Личность, складывающаяся из многих личностей (это культура, парадигма в науке и др.).

Личность - это спонтанность или открытость к Вселенской потенциональности, к самочитаемости, самоизменяемости текстов, к внутреннему самодвижению.

Смыслы связаны со спонтанностью.

Спонтанность текстов и смыслов шире и глубже примитивной случайности, элементарного механического явления, типа бросания игральной кости, а это взаимосвязанность и изменяемость термодинамически неравновесной системы, которой и является Личность.

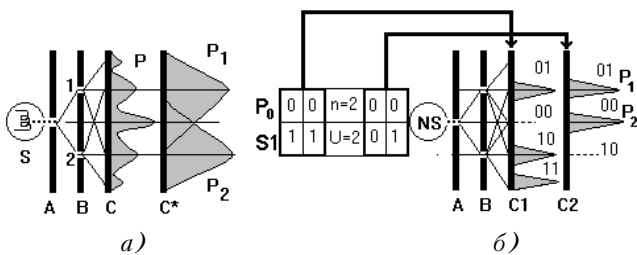
Трансценденция или уровни медитации и нирваны - это освобождение Сознания от семантических оков и выход за пределы капсулы смысла.

Сознание - это генератор ментальной энергии.

"Вселенная скорее великая мысль, чем машина" (Дж. Джинс 1948).

5 НК "ЭМБРИОН" КАК КВАНТОВО-МЕХАНИЧЕСКАЯ СИСТЕМА

Прежде чем сопоставить вероятностную модель смыслов (ВМС) и процессы в нейрокомпьютере "ЭМБРИОН", я хочу напомнить читателю и проиллюстрировать из [1] общность алгоритма или структурного метода измерений в квантовой механике (рис.7), который очень четко и образно изложил В.ФОК в [11], и алгоритмов синтеза и генерации нейронной сети в нейрокомпьютере (рис.8 и рис.9).



а - Опыт Фейнмана; б - Опыт на 2-х разрядном НК

Рисунок 7 - Интерференционные альтернативные траектории частиц на двух щелях

В этой структурной схеме хорошо видно место и роль принципа неопределенности и байесовой логики при исследовании квантового объекта микромира на стадии оценки и интерпретации результата эксперимента.

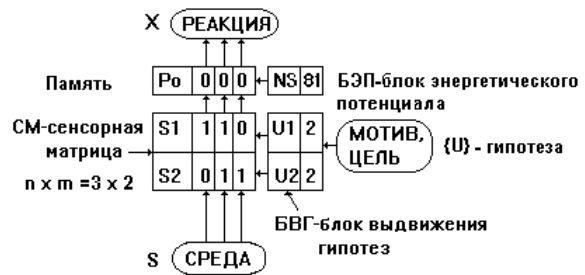


Рисунок 8 - Условия синтеза квазинеуронной сети в нейрокомпьютере "ЭМБРИОН"

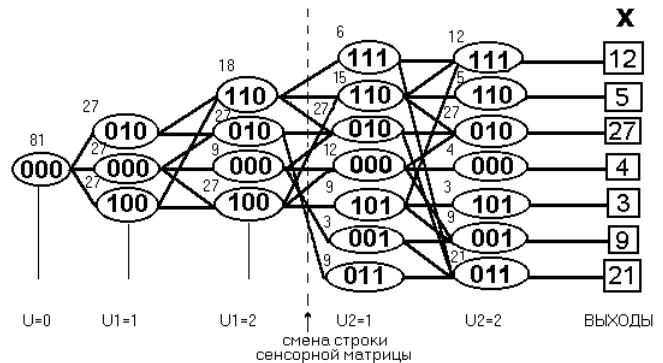


Рисунок 9 - Нейронная сеть, синтезированная НК "ЭМБРИОН" при условии, изображенном на рис. 8

Вот этот алгоритм:

1. Выбирается для исследования микрообъект и макроприбор для регистрации, т.е. выбирается квантово-механическая (КМ) система.

В НК выбирается n, P_0, m СМ, БВГ, БЭП. Квантово-механическим объектом служит квазичастица или виртуальное поле нейронной сети $\Psi(t)$.

2. Задается набор изучаемых и регистрируемых величин, параметров (координаты, импульсы, энергия, частота, спектр и т.д.).

У нас - это коды узлов решетки (X, Y) , спектры (Fx) , P - средняя плотность материи, вероятность смыслов и др.

3. Проводится многократный НАЧАЛЬНЫЙ ОПЫТ, то есть готовится квантово-механическая система и задаются внешние и начальные условия.

Задаются: число NS, коды-образы P_0 и S1 СМ, невязка J , (U)-гипотеза восприятия. Проводится NS-крат-

ное испытание при $U=0$.

4. Делается теоретический вероятностный прогноз (расчет возможных значений исследуемых физических величин).

Определение P , Ψ , $\Psi(t)$ путем решения уравнения ШРЕДИНГЕРА методом марковизации или возведения в U -ю степень марковской матрицы переходных вероятностей, т.е. определяются параметры "полезного результата" в акценторе действия [6].

5. Проводится многократный ПРОВЕРОЧНЫЙ ОПЫТ и регистрация результатов взаимодействия КМ-объекта и макроприбора-регистратора.

При NS-катных испытаниях в течение "времени жизни" ($T_{ж}$) регистрируются P , X , Y .

6. Используя данные результатов испытаний на предыдущем шаге-этапе (память действительных, а не возможных значений исследуемых величин x , p , E и т.д.) и условия ограничения на точность из-за ПРИНЦИПА НЕОПРЕДЕЛЕННОСТИ, сравниваются результаты опыта с порогом ошибки ($\epsilon > \theta$), принимается решение о достоверности опыта. Это есть ЗАВЕРШЕННЫЙ ОПЫТ.

Аналогичные аналитические процедуры выполняются и для НК.

7. Определяется статистика во времени, определяется распределение вероятности и спектры значений, определяется состояние КМ-объекта в виде его волновой функции, образа "волна-частица".

Обрабатываются все строки СМ, проявляется виртуальное поле НК в виде Ψ -волновой функции и распределения вероятностей выходных кодов-слов (X) и кодов-групп (Y) или текстов (рис.10).

6 НК "ЭМБРИОН" КАК ВЕРОЯТНОСТНАЯ МОДЕЛЬ ЛИЧНОСТИ

Читатель, вероятно, уже почувствовал на основании вышеизложенного близость процессов в вероятностной модели Смыслов (ВМС) НАЛИМОВА [4] и процессов в НК "ЭМБРИОН" [2]. Теперь можно построить таблицу 2, аналогичную первой таблице, для сравнения основных параметров обеих моделей - ВМС и НК.

Полную интерпретацию данной таблицы в виде строгой математической теории еще предстоит разработать. Однако, некоторые соображения к будущей "теории Личности" мы, все же, изложим.

Во-первых, разрядность (n) НК и число слоев нейронной сети, равное числу импульсов или гипотез (U_k) Блока Выдвижения Гипотез (БВГ) восприятия j -й строки Сенсорной матрицы (СМ), определяет максимально возможное разнообразие генерируемых текстов и максимальную длину слов в тексте.

Таблица 2

№ п/п	ВМС	НК
1.	μ - множество смыслов, переменная, степень свободы текста	$Y = 2^n$ - множество слов текста из двухбуквенного алфавита, степень свободы координата, смыслы
2.	$p(\mu/y)$ - текст, состояние субъекта	$\Psi(Y, t) = P(Y/J)$ - текст, функция состояния объекта
3.	$p(\mu)$ - функция вероятности, спонтанная "распаковка" смыслов, обнаружение значения μ	$ \Psi(Y) ^2 = P(Y)$ - функция вероятности спонтанного обнаружения Y -величины, "веса" смысла
4.	y - фильтр, текст, с которым взаимодействует исходный текст $p(\mu)$	$P(J/x)$ - фильтр, измерительный прибор, дающий измерение Y
5.	$p(\mu/y) = kp(\mu)p(y/\mu)$ - преобразование весовой функции смыслов $p(\mu)$	$\Psi_0(Y) \Rightarrow \Psi_n(Y/J)$ - преобразование функции состояния $\Psi(Y)$ в процессе измерения. Генерация текста

Во-вторых, NS - параметр, определяющий "время жизни" ($T_{ж}$) возбуждения в нейронной сети. Он определяет длительность текстовой последовательности или размер абзаца как смысловой единицы.

В-третьих, коды слов (X) регистров P_0 и S_{ij} - это смыслы, "распаковка" которых и есть $P(X)$ - распределение вероятностей смыслов слов X или $P(Y)$ - распределение вероятностей или "весов" смыслов групп U .

Определение "слов" и "групп" смотри [2].

В-четвертых, многомерное пространство вида $2^n \times 2^n \times U \times NS$ - это семантическое пространство слов и текстов. Его происхождение, геометрия и метрические свойства определяются ВОЛНОВЫМ ГЕНОМОМ $G = (P_0, S_{ij}, U)$ [5], [13] нейрокомпьютера "ЭМБРИОН" (рис.8) [2], [14], который определяет вид виртуального вероятностного поля $\Psi(t)$, как потока вероятностей, вид квазинеуронной сети (рис.9), как волнового пакета в виде марковской цепи, и вектора невязки J , как фильтра или акцептора действия, предопределяющего будущее - тексты и их смыслы [6].

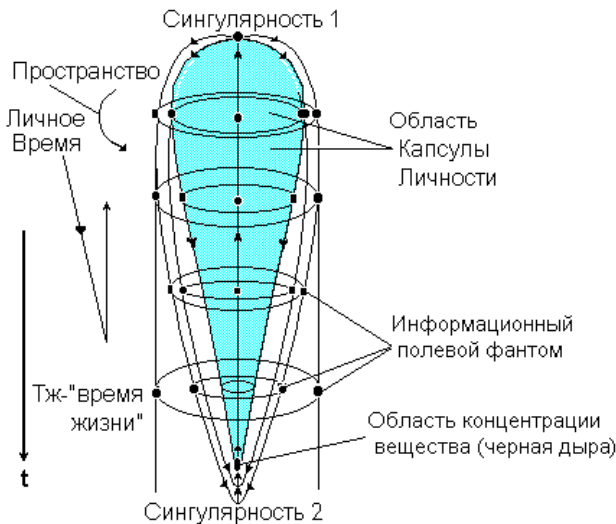


Рисунок 10 - Пространство-время-материя в НК "ЭМБРИОН" как модели Личности

Волновой геном определяет форму и размер Капсулы Личности, ее тела [1] (рис. 10)

Процесс генерации текстов и смыслов Y в НК "ЭМБРИОН" описывается на языке теории вероятностных автоматов следующим образом [2], [15]:

1. НК как стохастический автомат

$$A = \langle S, Y, P, W(P', Y/P, S) \rangle;$$

2. Объем нейронной сети (мозговой массы)

$$V = (2^n + n + 1) \sum_{i=1}^m U;$$

3. Внутренняя память

$$P = aS1 + bS2 + cS3,$$

где коэффициенты

$$a + b + c = 1;$$

4. Невязка

$$J(NS) = A - B = \sum_{i=1}^n a_1^i - \sum_{i=1}^n b_1^i;$$

5. Энтропия возбуждения нейронной сети

$$H = \sum_{i=1}^n d_i / D_Y \text{Ind}_i / D_Y;$$

6. "Время жизни" возбуждения

$$T_{ж} = NS(U1 + U2 + U3 + \dots + Um);$$

7. Вероятность выходного слова X_j

$$P(X_j) = \sum_{i=1}^n W_{ij} USp(P_0)P_0;$$

8. Вероятность слова на U -м шаге

$$P_r U = P_r^0 |W_{ij}| U;$$

9. Уравнение Колмогорова - Чепмена для НК "ЭМБРИОН"

$$P_r U = P_r^{0U-1} |W_{ij}|;$$

10. Матрица переходных вероятностей или "весов" связей между нейронами

$$W_{ij}(U+1) = W_{ij}(U) + / - aW_{ij};$$

11. Выходная реакция нейрокомпьютера на уровне групп Y слов

$$Y_m = \sum_{i=1}^n x_i;$$

12. Вероятность групп слов $P(Y)$ - сечение поля вероятностей на уровне групп Y

$$P(Y) = D \sum_{j=1}^m \sum_{i=1}^n W_{ij} USp(P_0)P_0;$$

13. Реакция реципрокного сжимателя

$$R = k \sum_{i=1}^{n/2} (k_i F_i - k_{-i} F_{-i}).$$

Ниже, на рис.11, приводится пример, подтверждающий вышеизложенное, и иллюстрирующий процесс генерации наследственного текста - последовательностей двухбуквенных слов из четырехбуквенного алфавита при образовании групп крови в двухрядной модели нейрокомпьютера "ЭМБРИОН"[12].

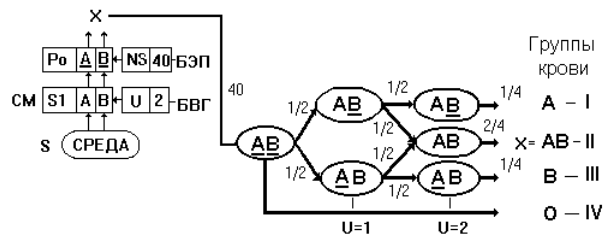


Рисунок 11 - Модель образования четырех групп крови человека

Дроби, стоящие у слов - это вероятности их появления на каждом шаге генерации текста.

ПЕРЕЧЕНЬ ССЫЛОК

1. Цыганков В. Д. Квантовая сингулярность. Вселенная Хокинга и нейрокompьютер. - М., 1997.
2. Цыганков В. Д. Нейрокompьютер и его применение. - М.: Сол Систем, 1993.

3. Перевозский Т. А. Вероятностная концепция смыслов и квантовая теория измерений / Налимов В.В. Спонтанность сознания. - М.: Прометей, 1989.
4. Налимов В. В. Спонтанность сознания. Вероятностная теория смыслов и смысловая архитектура личности. - М.: Прометей, 1989.
5. Цыганков В. Д. Информационный фантом возбуждения квазинейронной сети в нейрокompьютере "ЭМБРИОН". - М., 1997.
6. Анохин П. К. Биология и нейрофизиология условного рефлекса. - М.: Медицина, 1968.
7. Шипов Г. И. Теория физического вакуума. - М.: НТ-Центр, 1993.
8. Узнадзе Д. Н. Психологические исследования. - М.: Наука, 1966.
9. Цыганков В. Д. Живая Вселенная. Нейрокompьютер и мозг. - Нью-Йорк, 1997.
10. Бауэр Э. С. Теоретическая биология. - М.-Л.: ВИЭМ, 1935.
11. Фок В. Квантовая физика и строение материи. - М.: Знание, 1965.
12. Цыганков В. Д. О возможности применения нейрокompьютера "ЭМБРИОН" в генетике //Сознание и физическая реальность, 1997.
13. Горяев П. П. Волновой геном. - М.: Общественная польза, 1994.
14. Цыганков В. Д. Бытовой нейрокompьютер "ЭМБРИОН"// Байтик, 1991, №2, №4.
15. Бухараев Р. Г. Основы теории вероятностных автоматов. - М.: Наука, 1985 .

Надійшла 27.09.99

Після доробки 02.02.2000

ІІІ. УПРАВЛІННЯ

УДК 338.24+517

НЕЛИНЕЙНЫЕ ОБОБЩЕНИЯ ДИНАМИЧЕСКИХ УРАВНЕНИЙ ЛЕОНТЬЕВА ДЛЯ ДВУХСЕКТОРНОЙ ЭКОНОМИКИ

Б. Ю. Базылев, Е. Р. Саенко, А. Д. Шамровский

В работе составляется и анализируется нелинейный вариант двухсекторной динамической модели Леонтьева. Для этого рассматриваются уточненные формы производственных коэффициентов. При проведении анализа, в частности, используются численные методы. По результатам анализа делаются выводы о влиянии значений коэффициентов на устойчивость экономической системы, описываемой моделью. Делается вывод о возможности применения теории катастроф к исследованию данной модели.

У роботі складається та аналізується нелінійний варіант двохсекторної динамічної моделі Леонт'єва. Для цього розглядаються уточнені форми виробничих коефіцієнтів. При проведенні аналізу зокрема використовуються чисельні методи. За результатами аналізу робляться висновки про вплив значень коефіцієнтів на сталість економічної системи, що описується моделлю. Робиться висновок про можливість застосування теорії катастроф до дослідження даної моделі.

In paper the nonlinear variant of a two-sector dynamic Leontief's model is made and analyzed. For this purpose the updated forms of industrial factors are considered. For want of realizations of the analysis the numerical methods are used in particular. By results of the analysis the conclusions about influence of significance of factors to stability of an economic system circumscribed model are made. A possibility of application of the theory of catastrophes to a research of the given model is concluded.

Применение линейных динамических уравнений Леонтьева достаточно широко распространено для анализа функционирования многосекторных экономик [1,2]. Однако во многих случаях более целесообразно ограничиться рассмотрением относительно небольшого количества секторов, оперируя соответствующими интегральными показателями, но при этом уточнить модель, учтя нелинейные эффекты.

Работа посвящена составлению и анализу нелинейного варианта динамических уравнений Леонтьева для двухсекторной экономики.

Рассмотрим следующий вариант динамических уравнений Леонтьева для двухсекторной экономики:

$$x_1 - z_1(x_1) - a_{12}x_2 = v_{11}\frac{dx_1}{dt} + v_{12}\frac{dx_2}{dt} + y_1(x_1), \quad (1)$$

$$-a_{21}x_1 + x_2 - z_2(x_2) = v_{21}\frac{dx_1}{dt} + v_{22}\frac{dx_2}{dt} + y_2(x_2).$$

Здесь x_1, x_2 - валовые объемы продукции, выраженные в одинаковых денежных единицах; $z_1(x_1)$ - объем продукции первого сектора, идущий на потребности этого же сектора; $a_{12}x_2$ - объем продукции первого сектора, расходуемый на потребности второго сектора; $y_1(x_1)$ - объем свободных средств, остающихся в первом секторе; $a_{21}x_1$ - объем продукции второго сектора, расходуемый на потребности первого сектора; $z_2(x_2)$ - объем продукции второго сектора, идущий на потребности этого же сектора; $y_2(x_2)$ - объем свободных средств, остающихся во втором секторе; $v_{11}, v_{12}, v_{21}, v_{22}$ - коэффициенты фондоемкости. Отличие от соответствующих линейных уравнений заключается только в слагаемых $z_1(x_1), y_1(x_1), z_2(x_2), y_2(x_2)$, которые могут быть нелинейными функциями своих аргументов. Иначе говоря, считаем, что взаимные связи двух секторов описываются линейными зависимостями; также линейными являются динамические члены, характеризующие влияние инвестиций в обе отрасли.

Запишем функции $z_1(x_1), z_2(x_2)$ в форме

$$z_i = \beta_{i0} + \beta_{i1}x_i + \beta_{i2}x_i^2 \quad (i = \overline{1,2}). \quad (2)$$

В этой записи постоянные слагаемые β_{i0} показывают затраты, не зависящие от объемов производства в данной отрасли, например, затраты на содержание оборудования. Линейные слагаемые $\beta_{i1}x_i$ имеют обычный в уравнениях Леонтьева смысл и показывают затраты, пропорциональные объему производства. Квадратичные слагаемые $\beta_{i2}x_i^2$ отражают ситуации, когда расходы на производство изменяются не пропорционально объему производства. Например, в случае наличия простаивающего оборудования выполнение дополнительного заказа требует относительно небольших затрат; в этом случае будет $\beta_{i2} < 0$. Наоборот, получение дополнительного заказа при отсутствии необходимого оборудования требует

сравнительно больших расходов на приобретение этого оборудования, непропорциональных объему дополнительного заказа; в этом случае будет $\beta_{i2} > 0$.

В аналогичной форме представим и выражения для $y_1(x_1)$, $y_2(x_2)$

$$y_i = b_{i0} + b_{i1}x_i + b_{i2}x_i^2 \quad (i = \overline{1, 2}). \quad (3)$$

Здесь слагаемые вида $b_{i1}x_i$ отражают обычную для уравнений Леонтьева зависимость между объемом производства и количеством свободных средств. Постоянные слагаемые вида b_{i0} отражают какие-то непроизводственные затраты, не зависящие от объема производства, например, затраты на социальные нужды. Слагаемые вида $b_{i2}x_i^2$ могут отражать, например, прогрессивный налог (при $\beta_{i2} > 0$).

Вводя обозначения

$$f_1(x_1, x_2) = x_1 - z_1(x_1) - y_1(x_1) - a_{12}x_2 = -\beta_{10} - b_{10} + (1 - \beta_{11} - b_{11})x_1 - (\beta_{12} + b_{12})x_1^2 - a_{12}x_2, \quad (4)$$

$$f_2(x_1, x_2) = -a_{21}x_1 + x_2 - z_2(x_2) - y_2(x_2) = -a_{21}x_1 - \beta_{20} - b_{20} + (1 - \beta_{21} - b_{21})x_2 - (\beta_{22} + b_{22})x_2^2,$$

получаем

$$v_{11}\frac{dx_1}{dt} + v_{12}\frac{dx_2}{dt} = f_1(x_1, x_2), \quad (5)$$

$$v_{21}\frac{dx_1}{dt} + v_{22}\frac{dx_2}{dt} = f_2(x_1, x_2).$$

Решая систему уравнений (5) относительно производных $\frac{dx_1}{dt}$, $\frac{dx_2}{dt}$, имеем

$$\frac{dx_1}{dt} = \frac{v_{22}f_1(x_1, x_2) - v_{12}f_2(x_1, x_2)}{v_{11}v_{22} - v_{12}v_{21}} = h_1(x_1, x_2), \quad (6)$$

$$\begin{aligned} \frac{dx_2}{dt} &= \frac{-v_{21}f_1(x_1, x_2) + v_{11}f_2(x_1, x_2)}{v_{11}v_{22} - v_{12}v_{21}} = \\ &= h_2(x_1, x_2)v_{11}v_{22} - v_{12}v_{21} \neq 0. \end{aligned}$$

Нелинейные уравнения (6) можно решить только численным методом с применением компьютера. Однако, несмотря на относительную несложность этой процедуры, она требует предварительного аналитического исследования.

В случае постоянных значений x_1 , x_2 получаем из (6) уравнения равновесия

$$h_1(x_1, x_2) = 0; \quad h_2(x_1, x_2) = 0. \quad (7)$$

Эти уравнения эквивалентны уравнениям

$$f_1(x_1, x_2) = 0; \quad f_2(x_1, x_2) = 0, \quad (8)$$

т.е. статическим уравнениям Леонтьева. Решение уравнений (8) облегчается рассмотрением их графического представления. Рассмотрим плоскость x_1 , x_2 . Первое из уравнений (8) соответствует на этой плоскости параболе с вертикальной осью, а второе - параболе с горизонтальной осью. В зависимости от взаимного расположения этих парабол количество их точек пересечения, которые и соответствуют решениям системы уравнений (8), может быть различным, начиная от отсутствия таких точек (отсутствие действительных решений) и кончая наличием четырех точек пересечения (четыре решения). Случаи двух и четырех решений изображены на рис. 1 и 2.

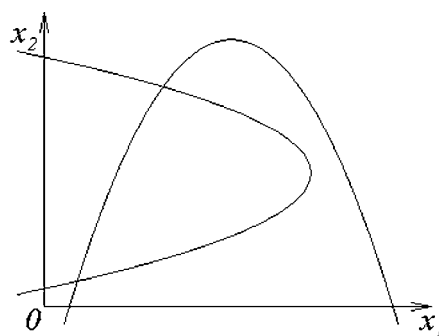


Рисунок 1

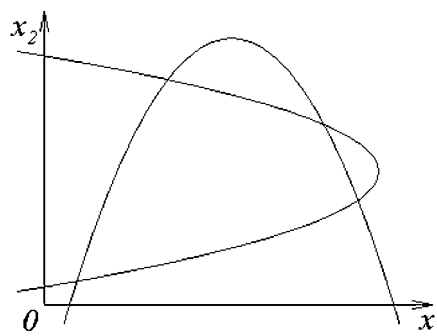


Рисунок 2

Точки, соответствующие состояниям равновесия системы, являются особыми точками на фазовой плоскости x_1 , x_2 . Эти точки оказывают решающее влияние на характер фазовых траекторий.

Рассмотрим ряд характерных примеров. При наличии четырех состояний равновесия системы и условии $v_{11}v_{22} - v_{12}v_{21} > 0$ фазовые кривые могут иметь вид,

изображенный на рис. 3. Эти кривые построены при помощи численного интегрирования уравнений (6) при различных начальных условиях, выбранных в окрестностях особых точек. Рассмотрение фазовых кривых показывает, что точка $(x_{1,1}; x_{2,1})$ является неустойчивым узлом. При любых начальных условиях, выбранных в окрестности этой точки, происходит удаление системы от нее. Точка $(x_{1,2}; x_{2,2})$ является устойчивым узлом. При любых начальных условиях, выбранных в окрестностях этой точки, происходит приближение к ней. Точки $(x_{1,3}; x_{2,3})$, $(x_{1,4}; x_{2,4})$ являются неустойчивыми особыми точками типа седло. Фазовые кривые проходят мимо них.

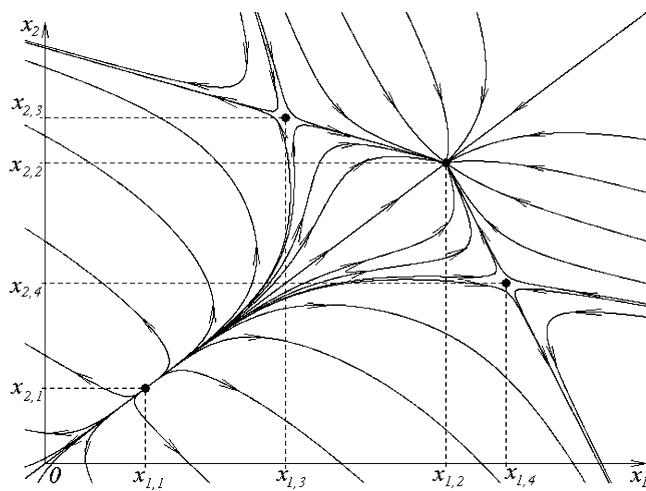


Рисунок 3

Отметим, что локальный анализ особых точек можно было бы произвести и аналитически, но применение компьютера и численного интегрирования позволило получить более полную интегральную картину.

Построенная на рис. 3 картина позволяет изучать варианты эволюции экономической системы. Если система первоначально находится в неустойчивом положении $(x_{1,1}; x_{2,1})$ с малыми объемами производства, то она может, в принципе, перейти в устойчивое положение $(x_{1,2}; x_{2,2})$ с большими объемами производства, но для этого требуется перемещение строго по прямой между соответствующими точками. При очень малых отклонениях от этой прямой фазовые кривые, проходя мимо седловых особых точек, достигают точки $(x_{1,2}; x_{2,2})$, но малейший выход за границы очень узкого "коридора" делает уже невозможным переход из $(x_{1,1}; x_{2,1})$ в $(x_{1,2}; x_{2,2})$, а приводит к отрицательным значениям x_1

или x_2 , т.е. разрушению системы.

В случае $v_{11}v_{22} - v_{12}v_{21} < 0$ положение особых точек остается прежним, однако, их характер существенно изменяется (рис. 4). Точки $(x_{1,1}; x_{2,1})$ и $(x_{1,2}; x_{2,2})$ становятся седлами, а точки $(x_{1,3}; x_{2,3})$ и $(x_{1,4}; x_{2,4})$ - фокусами.

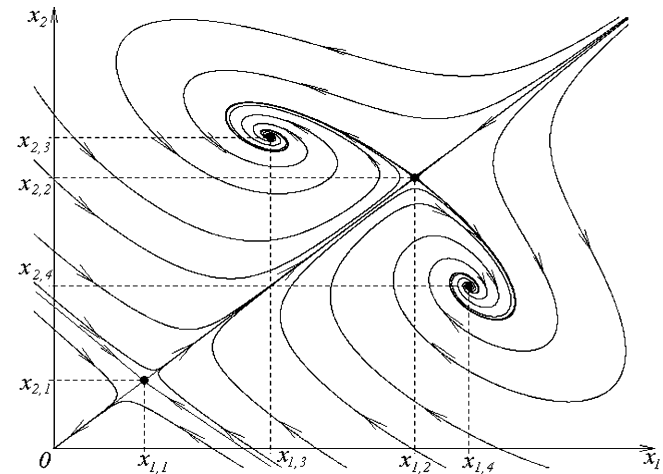


Рисунок 4

Обратим внимание на то, что теперь попадание в одно из двух устойчивых положений равновесия обеспечивается при выходе практически из любой точки фазовой плоскости, за исключением небольшой окрестности начала координат.

Перемена знака выражения $v_{11}v_{22} - v_{12}v_{21}$ имеет очевидный экономический смысл. Большим значениям произведения $v_{11}v_{22}$ отвечает такое состояние экономики, когда секторы плохо воспринимают собственные инвестиции, а лучше воспринимают инвестиции друг от друга, что и соответствует положительному значению $v_{11}v_{22} - v_{12}v_{21}$.

Отрицательным значениям выражения $v_{11}v_{22} - v_{12}v_{21}$ отвечает лучшее усвоение собственных, а не чужих инвестиций. При этом, как мы видели, устойчивость экономической системы резко улучшается.

Перейдем теперь к случаю двух положений равновесия (рис. 2). При $v_{11}v_{22} - v_{12}v_{21} > 0$ численное интегрирование дифференциальных уравнений (6) дает фазовую плоскость, изображенную на рис. 5. Положение равновесия, заданное точкой $(x_{1,1}; x_{2,1})$, является неустойчивым узлом, а положение, заданное точкой $(x_{1,2}; x_{2,2})$, - седлом. Иначе говоря, оба положения равновесия являются неустойчивыми.

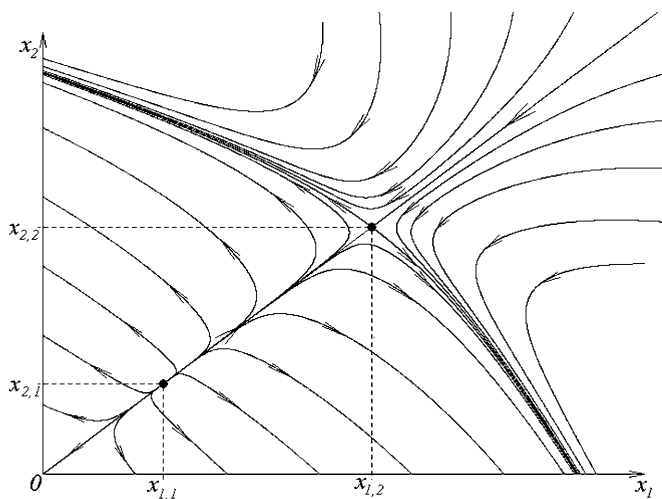


Рисунок 5

При $v_{11}v_{22} - v_{12}v_{21} < 0$ получаем фазовую плоскость, изображенную на рис. 6. Положение $(x_{1,1}; x_{2,1})$ теперь является седлом, а положение $(x_{1,2}; x_{2,2})$ - устойчивым узлом. Поэтому выход системы из положения $(x_{1,1}; x_{2,1})$ с приращением x_1 и x_2 обязательно приводит ее в новое, устойчивое положение равновесия.

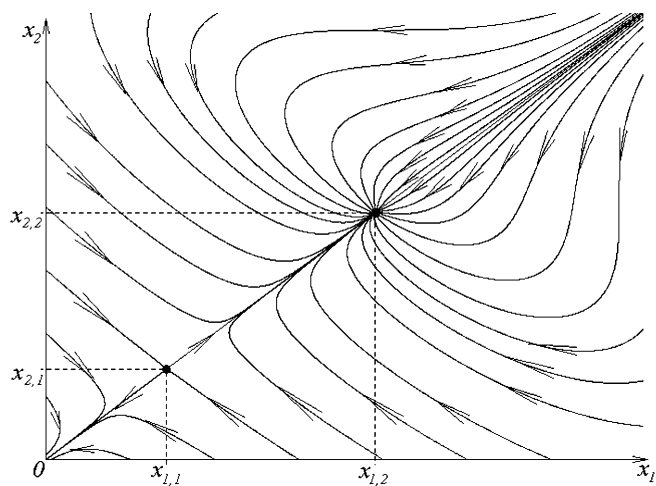


Рисунок 6

Таким образом, видим, что и в этом случае благоприятным для устойчивости системы является случай $v_{11}v_{22} - v_{12}v_{21} < 0$.

Случаи отсутствия состояний равновесия рассматривать не будем, поскольку они приводят к очевидным результатам распада экономической системы.

Уменьшение количества положений равновесия от четырех до двух и затем до нуля связано, в первую очередь, с увеличением коэффициентов $\beta_{10}, \beta_{20}, b_{10}, b_{20}$, характеризующих внепроизводственные затраты, и коэффициентов $\beta_{12}, \beta_{22}, b_{12}, b_{22}$ характеризующих прогрессивный рост производственных затрат и налогов. Также важную роль играют коэффициенты a_{12}, a_{21} характеризующие взаимные производственные вложения экономик друг в друга. Чем меньше величины a_{12}, a_{21} , т.е. чем меньше взаимная технологическая зависимость секторов, тем устойчивее экономическая система.

Обратим внимание на то, что переход системы из состояния с большим числом положений равновесия в состояние с меньшим числом таких положений происходит скачкообразно. Плавное изменение указанных выше параметров до некоторого предела не влияет на количество положений равновесия, но затем при незначительном дополнительном их изменении происходит уменьшение количества этих положений. Такие скачкообразные изменения состояния системы при небольших изменениях ее параметров изучаются в так называемой теории катастроф [3,4,5,6]. В данном случае понятие катастрофы может вполне оправдывать свое наименование, поскольку при уменьшении числа положений равновесия система может потерять устойчивость и разрушиться. Применение изложенной методики с использованием реальных значений параметров может позволить предупреждать нежелательное развитие событий.

ПЕРЕЧЕНЬ ССЫЛОК

1. Пономаренко О. I., Пономаренко В. О. Системні методи в економіці, менеджменті та бізнесі. - Київ: "Либідь", 1995. - 240 с.
2. Моисеев Н.Н. Математические задачи системного анализа. - М.: "Наука", 1981. - 488 с.
3. Труды международной конференции: "Критерии самоорганизации в физических, химических и биологических системах". - М., 1995. - 195 с.
4. Арнольд В.И. Теория катастроф. - М.: "Наука", 1990. - 128с.
5. Касти Дж. Большие системы. Связность, сложность и катастрофы. - М.: "МИР", 1982. - 216 с.
6. Хакен Г. Синергетика. Иерархия неустойчивостей в самоорганизующихся системах и устройствах. - М.: "МИР", 1985. - 420.

Надійшла 15.03.2000
Після доробки 04.05.2000

АНАЛИЗ ЭНЕРГЕТИКИ СОВРЕМЕННЫХ СИСТЕМ ЭЛЕКТРОПИТАНИЯ НА ОСНОВЕ УПРАВЛЯЕМЫХ ВЕНТИЛЬНЫХ ПРЕОБРАЗОВАТЕЛЕЙ

М. Н. Горбачев

Разработан и изложен обобщенный метод анализа энергетики специальных режимов работы современных вторичных систем электропитания на основе вентильных преобразователей и, в частности, управляемых выпрямителей (УВ). На основе этого метода найдены параметризованные уравнения для инженерных расчетов ортогональных составляющих полной мощности на входе стабилизированных УВ.

Розроблено та викладено узагальнюючий метод аналізу енергетики спеціальних режимів роботи сучасних вторинних систем електроживлення на основі вентильних перетворювачів, в тому числі керованих випрямлювачів (КВ). На основі цього методу знайдені параметризовані рівняння для інженерних розрахунків ортогональних компонент повної потужності на вході стабілізованих КВ.

Generalized method of power analysis of specialized modes of modern secondary systems of power sources on the base of fan converters and controlled rectifiers (CR), in particular, has been developed and delivered. Parameterized equations for the engineering calculations of orthogonal constituents of full capacity on the input of stabilized CR.

ВВЕДЕНИЕ

В современных вторичных системах электропитания устройств радиоэлектроники и различных промышленных потребителей электроэнергии наиболее широко применяются на практике вентильные преобразователи (ВП), работающие от сетей переменного напряжения с частотой 50 Гц. Они представляют собой силовые преобразовательные устройства на основе управляемых и неуправляемых полупроводниковых вентилей, например, автономные инверторы и стабилизаторы напряжения и тока, преобразователи частоты, токопараметрические выпрямители, источники реактивной мощности и др. Наиболее многочисленным подклассом преобразовательных устройств, относящихся к классу ВП, являются управляемые выпрямители (УВ), имеющие широкий ряд областей применения и широкий диапазон реализуемых мощностей (от сотен ватт до сотен киловатт в отдельном блоке).

Объектом изучения в настоящей статье являются современные вторичные системы электропитания на основе стабилизированных управляемых выпрямителей и особенности протекающих в них энергетических процессов.

СПЕЦИАЛЬНЫЕ РЕЖИМЫ РАБОТЫ УПРАВЛЯЕМЫХ ВЫПРЯМИТЕЛЕЙ

Многим потребителям электрической энергии для эффективной работы необходим стабилизированный ток,

либо стабилизированное напряжение, либо стабилизированная мощность. Режимы работы управляемых выпрямителей, в частности, управляемого мостового выпрямителя (УМВ), в процессе которых осуществляется стабилизация тех или иных указанных выше выходных параметров, будем называть специальными. Эти режимы реализуются в замкнутых системах автоматического регулирования (САР) с соответствующими отрицательными обратными связями (ООС).

Стабилизацию постоянного (выпрямленного) напряжения требуют, как известно, многие нагрузки, например, различные устройства радиоэлектроники, обмотки возбуждения электрических машин, гальванические и электролизные ванны, тяговые электродвигатели постоянного тока [1-4] и др. Примерами нагрузок, требующих стабилизацию выпрямленного тока, являются различные электротехнологические установки (дуговые сталеплавильные печи, печи электрошлакового переплава, плазмотроны для резки металлов), газоразрядные и лазерные лампы, цепи возбуждения электроприводов экскаваторов на электротяге, сварочная дуга [5-7] и др. Примерами нагрузок, требующих стабилизации потребляемой мощности, являются магнетроны для напыления металлов в вакуумных камерах, печи для выращивания монокристаллов, стекловаренные печи [8, 9] и др. Целью дальнейшего рассмотрения и анализа являются следующие специальные режимы работы УВ:

- 1) режим стабилизации выходного напряжения $U_d = const$;
 - 2) режим стабилизации выходного тока $I_d = const$;
 - 3) режим стабилизации выходной мощности $P_d = const$,
- где U_d , I_d и P_d - средние значения выходных величин.

В некоторых случаях весьма эффективным являются комбинированные специальные режимы, представляющие собой определенные сочетания режимов (1)-(3), что позволяет строить многофункциональные (универсальные) системы электропитания (СЭП) на основе одного или группы УВ.

Специальные режимы работы УВ имеют ряд особенностей. Несмотря на широкое практическое применение УВ, исследованию специальных режимов и их особенностей посвящено малое количество публикаций. И если специфика режима 1 исследована более обстоятельно, то специфика режимов 2 и 3 изучена крайне слабо.

Поэтому основное внимание уделено исследованию специфики именно режимов 2 и 3, которые в последние годы в связи со значительным развитием электротехнологий находят все большее применение.

Условиями существования специальных режимов УВ являются определенные законы управления - законы изменения углов включения α управляемых вентилях - тиристоров. Эти законы формируются с помощью электронных систем управления и реализуются, как правило, в СЭП на основе замкнутых САР за счет соответствующих ООС [3-8].

Специальные режимы УВ и их особенности целесообразно исследовать с единых позиций. Для этого используем энергетический подход, то есть будем изучать энергетические процессы (энергетику) УВ путем анализа полной мощности и ее составляющих как основных интегральных характеристик на входе УВ [1, 2]. При этом в качестве базовой схемы УВ выбираем мостовую симметричную трехфазную схему (УМВ), так как она является оптимальной с точки зрения использования силового согласующего трансформатора и тиристоров [1, 2]. В комплект компенсированного УМВ, схема которого широко известна и поэтому здесь не приводится, обычно входят батареи компенсирующих (косинусных) конденсаторов на стороне переменного тока и простейший сглаживающий фильтр, состоящий из одиночного реактора, - на стороне постоянного тока.

Считая стабилизацию выходных параметров идеальной, при общепринятых допущениях об идеальности питающей сети и элементов УМВ [5-7] используем расчетные формулы для полной мощности S и ее ортогональных составляющих P , Q и T согласно их традиционному определению [1, 2]:

$$P = S \cdot \chi = S \cdot v \cdot \cos \varphi, \quad (1)$$

$$Q = S \cdot v \cdot \sin \varphi, \quad (2)$$

$$T = \sqrt{S^2 - P^2 - Q^2}, \quad (3)$$

где P , Q и T - соответственно активная, реактивная и искажающая мощности на входе УМВ; χ , v , $\cos \varphi$ - соответственно коэффициенты мощности, искажения и сдвига кривой потребляемого тока относительно напряжения сети, имеющего синусоидальную форму.

Исходные соотношения (2) и (3) после математических преобразований удобно представить в следующем виде:

$$Q = S \sqrt{v^2 - \chi^2} = P_d \cdot \operatorname{tg} \varphi, \quad (4)$$

$$T = S \sqrt{1 - v^2} = P_d \cdot \frac{\sqrt{1 - v^2}}{\chi}, \quad (5)$$

где P_d - полезная мощность на стороне выпрямленного

тока ($P_d = P$ при принятых допущениях).

Соотношения (1) - (5) для расчета ортогональных составляющих P , Q и T являются общими и относятся к любым произвольным режимам УВ и УМВ как в замкнутых, так и в разомкнутых системах регулирования. Однако эти соотношения не отражают особенностей специальных режимов работы УВ как замкнутых САР с ООС, поскольку подлежащие определению ортогональные составляющие P , Q и T являются функциями нескольких переменных величин, не связанных условиями стабилизации выходных параметров в исследуемых режимах. Это значительно затрудняет сравнительный анализ специальных режимов с точки зрения энергетических процессов и необходимых устанавливаемых мощностей комплектующих реактивных элементов, входящих в состав силовой части схемы УВ.

ОБОБЩЕННЫЙ МЕТОД АНАЛИЗА СПЕЦИАЛЬНЫХ РЕЖИМОВ РАБОТЫ УПРАВЛЯЕМЫХ ВЫПРЯМИТЕЛЕЙ

Указанные обстоятельства послужили предпосылкой для разработки метода параметризации ортогональных составляющих (МПОС) полной мощности как обобщенного инженерного метода анализа энергетики специальных режимов УВ с учетом их особенностей.

Основой метода являются:

1) введение и использование одного или нескольких характерных параметров, отражающих особенности специальных режимов УВ и условия их реализации;

2) преобразование исходных соотношений (1) - (5) для определения ортогональных составляющих полной мощности на основе введенного параметра (параметров); в результате этих математических преобразований (параметризации) указанные уравнения упрощаются и приобретают вид аналитических функций от введенных параметров.

Рассмотрим применение МПОС на примерах расчета двух специальных режимов работы УМВ как базовой схемы УВ: режима стабилизации выпрямленного тока ($I_d = \text{const}$) и режима стабилизации выходной мощности ($P_d = \text{const}$).

Для исследования в общем виде специальных режимов работы УВ при нагрузке активного и активно-индуктивного характера целесообразно и удобно ввести параметр λ , характеризующий состояние переменной активной нагрузки R_d и ее динамические свойства:

$$\lambda = \frac{R_{d \max}}{R_{d \min}}, \quad (6)$$

где $R_{d \max}$ и $R_{d \min}$ - соответственно максимальное и минимальное возможные значения сопротивления на-

грузки R_d , причём

$$0 < R_{d \min} \leq R_d \leq R_{d \max} < \infty, \lambda \geq 1.$$

В общем случае, когда угол коммутации (угол перекрытия) тиристоров $\gamma \neq 0$, причём $0 < \gamma \leq \gamma_{\max}$, связь между углами φ , α и γ устанавливается из условий стабилизации выходных параметров и известного соотношения для УВМ [1, 2]:

$$\varphi = \alpha + 0,5\gamma, \quad (7)$$

где φ - угол сдвига первой гармоники потребляемого из сети тока относительно напряжения сети, α - угол включения тиристоров, называемый иначе углом регулирования УВ, причём

$$0 \leq \alpha \leq \alpha_{\max}.$$

В простейшем случае, когда угол коммутации тиристоров весьма мал ($\gamma \approx 0$), то получаем следующие соотношения:

а) для режима источника тока

$$\lambda = \frac{1}{\cos \alpha_{\max}}; \quad (8)$$

б) для режима источника мощности

$$\lambda = \frac{1}{\cos^2 \alpha_{\max}}; \quad (9)$$

Выбрав в качестве базовой величины максимально возможное значение мощности $P_{d \max}$ в цепи нагрузки, которое всегда известно и соответствует минимально возможному значению угла регулирования $\alpha_{\min} = 0$, далее находим относительные значения ортогональных составляющих полной мощности из выражений (4) и (5), пронормировав их по базовой величине с учетом соотношений (6) - (7):

$$\bar{Q}(\alpha, \gamma) = \frac{Q(\alpha, \gamma)}{P_{d \max}} = \operatorname{tg}(\alpha + 0,5\gamma), \quad (10)$$

$$\bar{T}(\alpha, \gamma) = \frac{\bar{T}(\alpha, \gamma)}{P_{d \max}} = \frac{\sqrt{\frac{1}{v^2} - 1}}{\cos(\alpha + 0,5\gamma)}, \quad (11)$$

где величина коэффициента искажения v кривой потребляемого УВМ из сети тока находится по известным формулам [1, 2].

Полагая $\alpha = \alpha_{\max}$ и $\gamma = \operatorname{const} > 0$, из соотношений (8) - (11) находим относительные значения $Q(\alpha, \gamma)$ и $T(\alpha, \gamma)$ для изучаемых специальных режимов:

а) для режима источника тока

$$\bar{Q}(\lambda) = \frac{C_1 + \sqrt{\lambda^2 - 1}}{1 - C_1 \cdot \sqrt{\lambda^2 - 1}}; \quad (12)$$

$$\bar{T}(\lambda) = \frac{\sqrt{\frac{1}{v^2} - 1}}{C_2 \cdot \left(\frac{1}{\lambda} - C_1 \sqrt{1 - \frac{1}{v^2}} \right)}; \quad (13)$$

б) для режима источника мощности

$$\bar{Q}(\lambda) = \frac{C_1 + \sqrt{\lambda - 1}}{1 - C_1 \cdot \sqrt{\lambda - 1}}; \quad (14)$$

$$\bar{T}(\lambda) = \frac{\sqrt{\frac{1}{v^2} - 1}}{C_2 \cdot \left(\sqrt{\lambda} - C_1 \sqrt{1 - \frac{1}{v}} \right)}; \quad (15)$$

где $C_1 = \operatorname{tg} 0,5\gamma = \operatorname{const} > 0$ и $C_2 = \cos 0,5\gamma = \operatorname{const} > 0$.

В идеализированном частном случае при $\gamma = 0$, соответствующем мгновенной коммутации тиристор УВМ, расчетные коэффициенты C_1 и C_2 принимают граничные значения: $C_1 = 0$, $C_2 = 1$. Поэтому и расчетные формулы (12) - (15) как функции параметра λ принимают соответственно более простой вид и являются частным случаем, изложенным в работе [8]:

а) для режима источника тока

$$\bar{Q}(\lambda) = \sqrt{\lambda^2 - 1}; \quad (16)$$

$$\bar{T}(\lambda) = \lambda \cdot \sqrt{\frac{1}{v^2} - 1}; \quad (17)$$

б) для режима источника мощности

$$\bar{Q}(\lambda) = \sqrt{\lambda - 1}; \quad (18)$$

$$\bar{T}(\lambda) = \sqrt{\lambda \left(\frac{1}{v^2} - 1 \right)}. \quad (19)$$

Аналитические функции (12) - (19) представляют собой параметризованные уравнения для расчета относительных значений мощностей $\bar{Q}(\lambda)$ и $\bar{T}(\lambda)$, найденные на основе метода МПОС. Они позволяют более точно рассчитать необходимые относительные установленные мощности батарей компенсирующих конденсаторов и сетевых фильтров высших гармоник на входе УВМ. В общем случае для этого в уравнения (12) - (15) нужно подставить максимальное достигаемое значение угла коммутации ($\gamma = \gamma_{\max}$) для рассмотренных специальных режимов.

На основе энергетического подхода и предложенного метода МПОС можно получить аналогичные расчетные соотношения и для других разных схем УВ, работающих в специальных режимах [1, 2].

ЗАКЛЮЧЕНИЕ

Полученные на основе метода МПОС расчетные формулы и функциональные зависимости от обобщенных режимных параметров позволяют глубже исследовать работу УВ, в частности УМВ, в специальных режимах и более точно рассчитывать необходимые устанавливаемые мощности силовых реактивных элементов, входящих в комплект УВ. Достоинством метода МПОС и разработанной на его основе методики расчета являются: простота, доступность, наглядность и удобство при использовании в инженерной практике.

Кроме того, этот метод является универсальным, так как он применим для анализа и расчета специальных режимов не только УВ, но и широкого класса различных вентильных преобразовательных устройств (ВП) силовой полупроводниковой электроники: инверторов, преобразователей частоты, компенсационнопараметрических стабилизаторов тока, регуляторов и стабилизаторов напряжения, симметрирующих устройств, источников реактивной мощности, мощных генераторов импульсов и др.

Предложенный метод является развитием и обобщением методов расчета энергетических режимов широкого

класса схем УВ, изложенных в работах О.А.Маевского, А.В.Поссе, Р.С.Глинтерника, С.В.Свечникова, И.М.Чиженко и других ученых в области преобразовательной техники и теории нелинейных электрических цепей и систем.

ПЕРЕЧЕНЬ ССЫЛОК

1. Полупроводниковые выпрямители // Под ред. Ковалева Ф.И. и Мостковой Г.П. - М.: Энергия, 1967. - 480 с.
2. Руденко В.С., Сенько В.И., Чиженко И.М. // Преобразовательная техника. - Киев: Вища школа, 1978. - 424 с.
3. Горбачев М.Н., Губаревич В.Н., Черная В.Ф. и др. Анализ статического режима стабилизированного тиристорного выпрямителя // Труды Николаевск. кораблестроит. ин-та. - 1976. - Вып.115. - С.66 - 74.
4. Горбачев М.Н. Анализ стабилизированного тиристорного выпрямителя с отрицательной обратной связью по напряжению // Вестник Харьковск.политех.ин-та. Электроэнергетика и автоматизация энергоустановок. - 1974. - № 76. - Вып.3. - С.12 - 16.
5. Горбачев М.Н. Тиристорный стабилизатор тока возбуждения экскаваторных электродвигателей // Системы стабилизированного тока. - Киев: Наук.думка, 1976. - С.149-160.
6. Горбачев М.Н. К вопросу о сопоставлении вторичных источников постоянного тока, питаемых от промышленной сети // Вестник Харьковск.политех.ин-та. Электроэнергетика и автоматизация энергоустановок. - 1986. - № 235. - Вып.13. - С.44 - 48.
7. Горбачев М.Н. Установленные мощности электрооборудования технологических установок стабилизированного тока на основе управляемых выпрямителей // Проблемы энергоснабжения. - 1994. - № 2 - 3. - С. 142 - 149.
8. Грудская В.П., Горбачев М.Н., Бойко Ю.В. Расчет мощностей реактивных элементов управляемого выпрямителя, работающего в режимах стабилизации тока нагрузки и выходной мощности // Вестник Киевск.политех.ин-та. Электроэнергетика. - 1990. - Вып. № 27. - С.40 - 44.

Надійшла 13.03.2000

Після доробки 27.03.2000

УДК 621.313.2.222

ЭНЕРГОСБЕРЕГАЮЩЕЕ УПРАВЛЕНИЕ ДВИГАТЕЛЕМ ПОСТОЯННОГО ТОКА

С. Г. Деев, В. И. Левыкина

В работе рассмотрены вопросы энергосберегающего управления двигателем постоянного тока при одновременном управлении по цепям якоря и возбуждения.

В роботі розглянуті питання енергосберігаючого керування двигуном постійного струму при одночасному керуванні по колах якоря та збудження.

In this article the questions of energy safe control of a direct current driver are considered by simultaneous control in armature and excitation circuits.

ВВЕДЕНИЕ

Приводы с двигателями постоянного тока (ДПТ) составляют сейчас значительную часть находящихся в эксплуатации регулируемых приводов. Подвод энергии в

них осуществляется по двум каналам: через обмотку возбуждения и обмотку якоря. Как правило, через обмотку возбуждения подводится не более 5-10% энергии, но существуют и ДПТ, в которых эта часть достигает 50%. Обычно в таких двигателях обмотка возбуждения включается последовательно с обмоткой якоря. Их характерной особенностью является высокий максимальный момент [1]. В то же время такая схема включения значительно усложняет систему регулирования, что ограничивает область их применения. Целью данной статьи является разработка алгоритма управления, в полной мере использующего возможности двигателей такого типа.

ПОСТАНОВКА ЗАДАЧИ

Уравнения состояния ДПТ имеют следующий вид:

$$M = \kappa_m \cdot \Phi \cdot i_a, \quad (1)$$

$$\omega = \frac{U - R_{яц} \cdot i_a}{\kappa_e \cdot \Phi}, \quad (2)$$

где M - момент двигателя; Φ - магнитный поток; ω - угловая скорость двигателя; i_a - ток якоря; U - напряжение на якорной цепи; $R_{яц}$ - сопротивление якорной цепи; κ_e , κ_m - конструктивные коэффициенты двигателя.

Зависимость потока Φ от тока в обмотке возбуждения i_e носит нелинейный характер, а именно:

$$\Phi = \kappa_e \cdot i_e, \quad (3)$$

где κ_e - переменный коэффициент, зависящий от вида кривой намагничивания (рис.1).

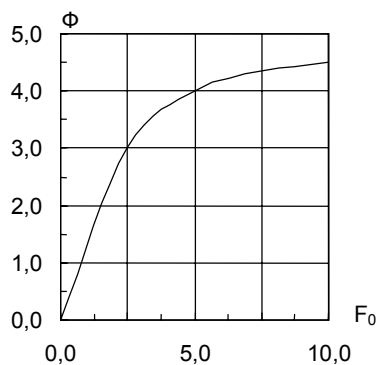


Рисунок 1 - Кривая намагничивания двигателя Д808

Потребляемая двигателем мощность P определяется по зависимости:

$$P = U_e \cdot i_e + U \cdot i_a, \quad (4)$$

где U_e - напряжение на обмотке возбуждения, причём i_e зависит от U_e .

Механическая мощность P_2 и коэффициент полезного действия η при этом равны

$$P_2 = M \cdot \omega, \quad (5)$$

$$\eta = \frac{P_2}{P} \cdot 100\%. \quad (6)$$

При традиционном двухзонном управлении в первой зоне фиксируется $U_e(i_e)$, а управление осуществляется за счёт изменения напряжения на якоре U . Во второй зоне фиксируется U , а управление осуществляется изменением $U_e(i_e)$, а следовательно, потоком Φ , причём, как правило, в ограниченных пределах.

Авторами работы проведены теоретические исследования тихоходного двигателя Д808 (номинальное напряжение 220 В, $R_{яц}=0,034$ Ом, $R_e=0,031$ Ом, $R_{дн}=0,02$ Ом) с последовательным возбуждением при раздельном питании обмоток якоря и возбуждения с учетом обмотки дополнительных полюсов. Они показали, что при постоянных значениях скорости и нагрузки двигателя ($\omega = 55 \text{ с}^{-1}$, $M = 70 \text{ кг} \cdot \text{м}$) оптимальное соотношение токов якоря и возбуждения, обеспечивающее минимум расхода энергии (рис.2), не соответствует типовой схеме включения ($i_a = i_e$).

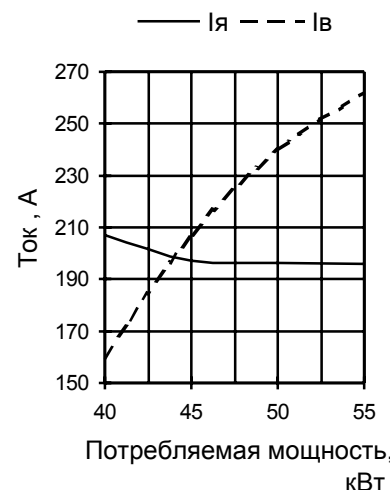


Рисунок 2 - Зависимость потребляемой мощности двигателя Д808 от тока в обмотке возбуждения ($\omega = 55 \text{ с}^{-1}$, $M = 70 \text{ кг} \cdot \text{м}$)

Очевидно, что величина потерь зависит в том числе от степени насыщения магнитной системы. При этом необходимо учесть, что магнитная система, как правило, при работе двигателя в номинальном режиме, находится в зоне, близкой к насыщению. При работе с большими моментами (для двигателя Д808 отношение максимального момента к номинальному равно 5), что характерно для некоторых типов приводов, степень насыщения при включении по типовой схеме значительно увеличивается, что ещё более ухудшает энергетические показатели привода.

Вопросы оптимального соотношения токов якоря и возбуждения рассмотрены в [2].

РАЗРАБОТКА СИСТЕМЫ УПРАВЛЕНИЯ

Учитывая изложенное выше, имеется необходимость разработки системы управления, обеспечивающей минимум потерь энергии в рабочих режимах.

Для обеспечения качественной работы системы необходимо учитывать влияние неопределённостей в параметрах объекта, возникающих в процессе нормальной эксплуатации (например, погрешности определения момента инерции якоря, величины магнитного потока, изменение сопротивления обмоток при изменении температуры и др.), поэтому система управления должна обладать робастными свойствами. Оптимизацию энергопотребления можно проводить путём формирования фазовой траектории и реализующего её управления, при перемещении по которой между двумя точками минимизируется потребляемая энергия. Другим путём оптимизации является минимизация потребляемой мощности в каждой точке заданной фазовой траектории. Такое субоптимальное управление называется экстремальным. Оно может быть реализовано в неопределённых системах, обладающих робастностью.

Закон оптимизации энергетических показателей двигателя привода будет иметь вид

$$P = U_{\theta} \cdot i_{\theta} + U \cdot i_{я} \rightarrow \min. \quad (7)$$

Очевидно, что эффект от оптимизации соотношения токов будет тем выше, чем больше энергии подводится через обмотку возбуждения, поэтому ниже будет рассмотрен алгоритм энергосберегающего оптимального управления ДПТ последовательного возбуждения с независимым управлением по обмоткам якоря и возбуждения.

Воздействие на соотношение токов возможно только изменением U_{θ} (и соответственно i_{θ}), при этом ток якоря из (1) с учётом (3) будет равен:

$$i_{я} = \frac{M}{\kappa_{м} \cdot \kappa_{\theta} \cdot i_{\theta}}. \quad (8)$$

При этом изменяется магнитный поток Φ (3). Так как при изменении потока (3) из (2) видно, что будет меняться и скорость двигателя ω , требуется компенсировать изменение скорости за счет изменения напряжения на якоре U , исходя из требования

$$\omega = \omega_3 = const, \quad (9)$$

где ω_3 - заданная скорость вращения двигателя. Напря-

жение на якоре из (2) при (9)

$$U = \omega_3 \cdot \kappa_e \cdot \Phi + R_{яц} \cdot i_{я}. \quad (10)$$

Точный аналитический расчёт по зависимости (7) при широком изменении рабочих скоростей и моментов затруднителен, так как схема ДПТ, используемая, как правило, для синтеза систем регулирования, учитывает не все процессы, происходящие в двигателе в рабочих режимах, например, искажение формы кривой напряжения в якоре, что в данном случае существенно. Использование более сложных моделей приведёт к значительному усложнению систем регулирования, что нежелательно, поэтому в общем случае необходимо использование в (7) реальных значений $i_{я}$, U , U_{θ} и i_{θ} .

Структурная схема системы, реализующей описанные выше законы регулирования, приведена на рис.3. Она включает экстремальный регулятор (ЭР) тока возбуждения i_{θ} , который, используя измеренные значения $i_{я}$, U , U_{θ} и i_{θ} , вырабатывает сигнал, соответствующий оптимальному значению тока возбуждения i_{θ_0} в соответствии с (7). Вычисленное значение i_{θ_0} подаётся на вход контура тока возбуждения с регулятором тока возбуждения (РТВ). Скорость вращения двигателя поддерживается в соответствии с заданным U_{zc} системой регулирования скорости с регулятором скорости (РС), формирующей напряжение на якоре двигателя в соответствии с (9) и (10). Предложенная структура системы управления позволяет рассматривать регуляторы как независимые друг от друга, что значительно упрощает процесс синтеза.

ВЫВОДЫ

1. Использование ДПТ с последовательным возбуждением с включением по схеме с отдельным питанием обмоток якоря и возбуждения позволяет повысить экономичность его работы.
2. Экономичность работы ДПТ, включенных по такой схеме, зависит от соотношения токов в обмотках якоря и возбуждения.
3. Предложены законы управления, обеспечивающие возможность оптимизации энергетических показателей систем с ДПТ.
4. Предложена общая структура системы управления, обеспечивающая реализацию предложенных законов управления.

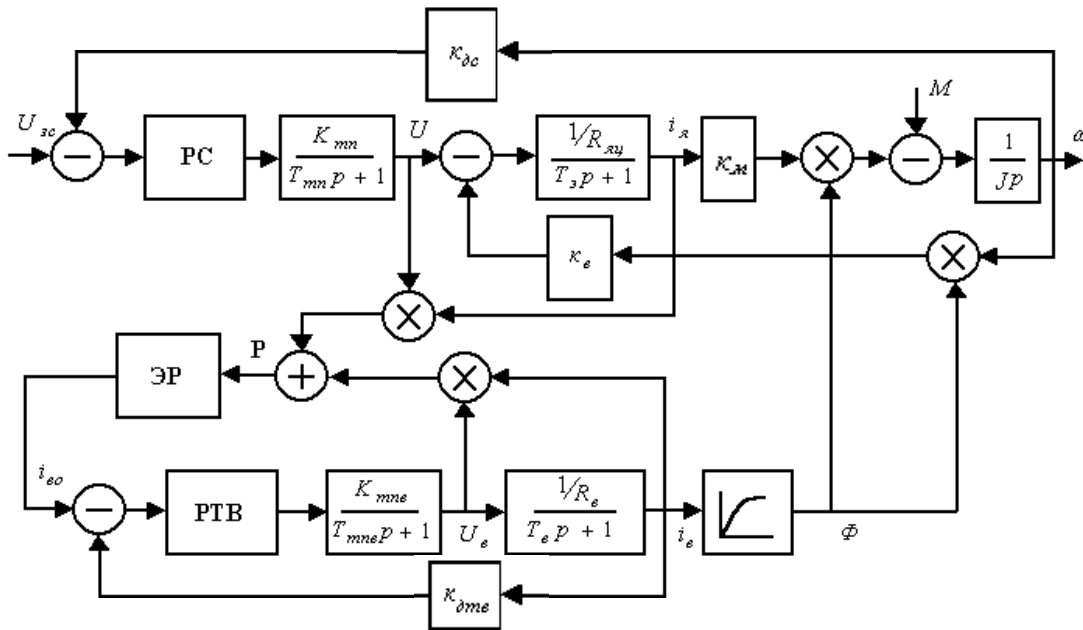


Рисунок 3 - Схема энергосберегающей системы регулирования ДПТ

ПЕРЕЧЕНЬ ССЫЛОК

1. Электротехника СССР. Двигатели постоянного тока металлургические и крановые серии Д мощностью 2.5-185 кВт: Каталог. - Информэлектро, 1978.-27 с.
2. Е.М. Потепенко, С.А. Мережко, А.Н. Юрченко. Оптимальное

управление двигателем постоянного тока.// Вестник Харьковского государственного политехнического университета. - Харьков, ХГПУ. Вып. 70, 1999. - стр. 177-182.

Надійшла 20.03.2000

Після доробки 24.03.2000

УДК 658.562

СИСТЕМА УПРАВЛЕНИЯ КАЧЕСТВОМ ПРОМЫШЛЕННОГО ПРЕДПРИЯТИЯ

Ю. Н. Дорошенко, В. И. Дубровин

Рассмотрены принципы построения системы управления качеством. Исследованы функциональные особенности системы управления качеством на примере модуля Quality Management интегрированной системы управления предприятием.

Розглянуто принципи побудови системи керування якістю. Досліджені функціональні особливості системи керування якістю на прикладі модуля Quality Management інтегрованої системи керування підприємством.

Constrction principles of quality control system are considered. The functional particular features of quality control system are explored. The module of Quality Management of integrated system of enterprise management is used as an example.

ВВЕДЕНИЕ

На сегодняшний день среди поставщиков и производителей продукции на мировом рынке существует жесткая конкуренция. Одним из основных условий, опреде-

ляющих конкурентоспособность продукции, наряду с ценой является качество. Создание эффективного инструментария, позволяющего производить наладку технологических процессов и регулировать качество выпускаемой продукции, является жизненно необходимым для выживания и развития предприятий Украины.

Производство высококачественной продукции невозможно без внедрения соответствующих автоматизированных систем, которые позволяли бы эффективно управлять качеством изделий [1].

1 ФУНКЦИИ СИСТЕМЫ УПРАВЛЕНИЯ КАЧЕСТВОМ ПРОДУКЦИИ

Исходя из того, что управление качеством должно проводиться непрерывно на всех этапах работы предприятия или фирмы, включая весь жизненный цикл

производимой продукции (от проектирования изделия до его обслуживания после продажи), системы управления качеством должны функционировать в рамках интегрированных систем управления предприятием.

На рисунке 1 представлена общая схема управления качеством в процессе проектирования, производства и обслуживания продукции.

Стандарты по управлению качеством [2], предполагают универсальный подход к системам управления, способным обеспечить требуемый уровень качества продукции. Согласно этим стандартам управление качеством достигается, с одной стороны, за счет управления непосредственно самой структурой сети процессов, в которой движется поток продукции или информации, с другой - за счет управления качеством продукции. Таким образом, функции системы управления качеством достаточно широки и обеспечивают решение целого ряда фундаментальных задач, к которым сводится процесс повышения качества продукции.

Основные функции системы управления качеством представлены в таблице 1.

Таблица 1 - Функции системы управления качеством

№	Задача	Основные функции системы
1	Планирование качества	Определение номенклатуры требований к материалам и комплектующим. Определение требований к качеству продукции. Выбор методов контроля и управления качеством. Определение процедур внесения изменений в технологический процесс в зависимости от результатов контроля.
2	Контроль качества	Определение степени соответствия качества материалов и комплектующих установленным требованиям. Определение степени соответствия качества продукции установленным требованиям. Регистрация данных о качестве продукции.
3	Управление качеством	Выбор поставщика, поставляющего материалы и комплектующие, наиболее соответствующие требованиям. Анализ причин дефектов изготавливаемой продукции. Управляющие воздействия, направленные на устранение причин брака и улучшение качества.

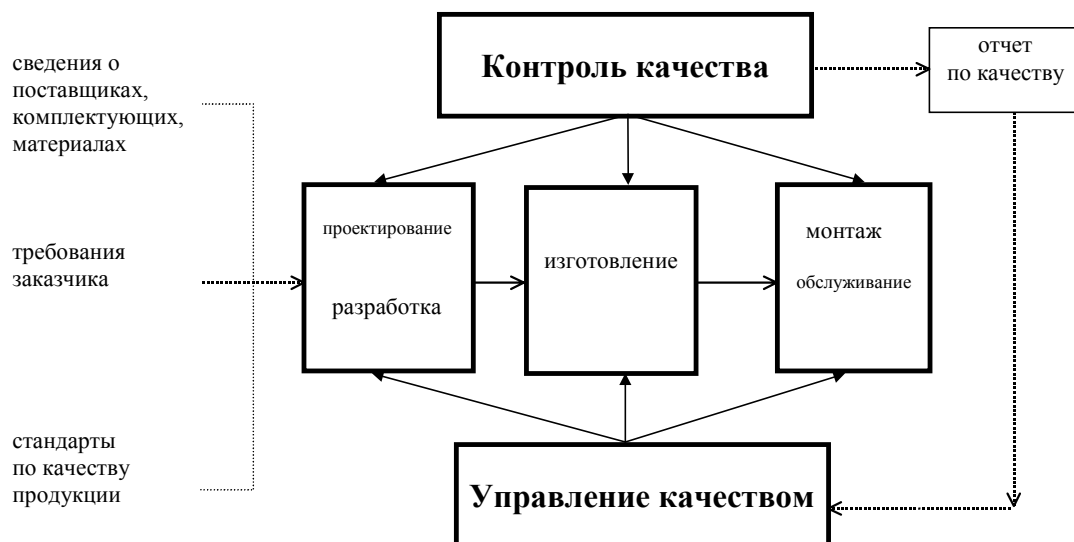


Рисунок 1 - Схема управления качеством на этапах жизненного цикла продукции

2 АНАЛИЗ СУЩЕСТВУЮЩИХ СИСТЕМ УПРАВЛЕНИЯ ПРЕДПРИЯТИЕМ

В настоящее время в мире существуют определенные наработки в области создания систем управления предприятием и систем управления качеством.

Примером отечественных разработок в области управления качеством может служить комплексная система управления качеством продукции (КС УКП) на базе стандартов предприятия. Задачей системы являлось получение максимальных объемов производства продукции высшей категории качества за счет рационального и эффективного использования производственных фондов, материальных, трудовых и финансовых ресурсов [3]. Основным недостатком КС УКП была привязка к командно-административной системе, которая существовала в стране на момент разработки КС УКП, выразившаяся в недостаточной экономической заинтересованности в обеспечении качества и равнодушии к потребителю. Кроме того, КС УКП функционировала несколько обособлено, в отрыве от общей системы управления предприятием, а применение такого подхода не дает возможности достаточно эффективно использовать системные данные и охватывать все уровни управления предприятием. На основе КС УКП, государственных и отраслевых стандартов, а также стандартов предприятия были разработаны КС ПЭП (комплексная система повышения эффективности производства) и КС УКП и ЭИР (комплексная система управления качеством продукции и эффективным использованием ресурсов). Несмотря на большое количество положительных моментов, эти системы также не решили полностью задачу интеграции методов и структур управления качеством продукции с общей структурой производства и управления. Автоматизация управления качеством в этих системах и в ряде других осуществлялась в отрыве от внедряемых систем управления, лишь частично обеспечивая

решение необходимых задач.

В настоящее время на отечественном рынке присутствуют западные системы управления предприятием типа SAP R/3, BAAN IV, Scala, Platinum, CA-PRMS, а также российские разработки "Парус", "Галактика", "ПРО/МИС".

Их появление было вызвано следующими причинами:

- переход к распределенной обработке данных;
- переход к архитектуре клиент-сервер;
- требование обеспечения централизованного хранения данных и отсутствие их дублирования различными службами предприятия;
- требование оперативности получения информации в различных ракурсах и различной степени детализации;
- требование исключения потерь жизненно важной для предприятия информации.

Ниже приведено описание наиболее известных систем управления предприятием.

2.1 Система SAP R/3

Система R/3 немецкой фирмы SAP AG [4] представляет собой интегрированный комплекс, реализующий целый ряд производственных задач по сбыту, учету и непосредственно производству продукции.

Компоненты интегрированной системы представлены на рисунке 2.

Основными достоинствами системы являются:

- универсальность (мощная система настроек позволяет использовать систему для различных типов предприятий, учитывая общие характеристики и специфические особенности технологических процессов);
- открытость (наличие встроенного языка программирования для создания собственных приложений для узких производственных задач, решение которых не предусмотрено в системе);



Рисунок 2 - Компоненты системы SAP R/3

- модульный принцип построения, позволяющий использовать определенную номенклатуру компонентов системы в зависимости от класса решаемых задач;

- удобный пользовательский интерфейс с соблюдением основных требований по стандарту IBM - Microsoft и эргономических характеристик;

- объединение различных классов производственных задач в единый комплекс.

Концепция внедрения системы предполагает приведение предприятия под стандарты R/3 (реинжиниринг), с незначительной настройкой. В связи с этим внедрение системы представляет собой довольно длительный процесс.

2.2 СИСТЕМА CA-PRMS

Система CA-PRMS [5] разработана компанией Computer Associates International, Inc (CA).

В отличие от R/3 система не требует полной реструктуризации бизнес-процессов и подстраивания под новое программное обеспечение. CA выделяет семь уровней фирмы (предприятия), позволяющие наладить управление информационными и материальными потоками:

- сервис для клиентов и менеджмент спроса (прогнозирование, продажи, эксплуатация оборудования, гарантированное обслуживание);

- управление ресурсами предприятия (планирование и потребности в материалах);

- управление закупками и производством при наличии интерфейсов с предприятиями и каналами снабжения;

- планирование и менеджмент ограничений по ресурсам и материалам;

- логистика и менеджмент распределения с применением приложений управления складами и пресылками;

- финансовые, учетные, административные приложения;

- система поддержки принятия решений;

- средства внедрения и стандарты ISO 9000.

2.3 Система BAAN IV

Система управления предприятием BAAN IV [6,7] фирмы BAAN - корпоративная система управления, поддерживающая основные направления деятельности предприятия, включая финансы, производство, сбыт, снабжение, склады, транспортные перевозки и системное обслуживание. В системе поддерживаются методики планирования JIT (ноль запасов), MRP, MRP II (минимум сверхнормативных запасов), Kanban (введение компонентов в производственный процесс в то время, когда они необходимы), планирование по прогнозу потребностей. Процесс производства рассматривается укрупненно, на уровне материально-информационных потоков между подразделениями. Преимущества такого подхода:

- возможность эффективного контроля за состоянием материальных и информационных потоков;

- анализ работы предприятия в целом;

- принятие и быстрая реализация глобальных управленческих решений;

- гибкость производственного процесса;

- эффективный контроль качества продукции.

Однако такой подход вызывает определенные трудности и ограничения, так как не позволяет решать локальные задачи, касающиеся оптимизации работы отдельного цеха, отслеживания изменений внутрицеховых ситуаций.

2.4 Система "Галактика"

Недостатком западных систем управления предприятием является их недостаточная приспособленность к условиям отечественного рынка, таким как: нестабильность системы налогообложения, невыполнение сроков поставки поставщиками, задержка платежей. Эти условия являются трудноформализуемыми в понятиях западных систем. Такого рода затруднения вызвали необходимость отечественных разработок, учитывающих данную специфику [8].

Комплексная система управления деятельностью предприятия "Галактика" российской фирмы "Новый Атлант" и белорусской компании "Топ Софт" предназначена для решения следующих задач:

- анализа финансовой и хозяйственной деятельности;

- календарно-сетевого планирования;

- управления производством;

- управления проектами;

- бухгалтерского учета;

- управления закупками и продажами, материально-техническим снабжением и сбытом;

- учета и управления кадрами;

- управления автотранспортом;

- анализа эффективности маркетинговых мероприятий и рекламных кампаний;

- документооборота.

Достоинствами системы "Галактика" является:

- работа в распределенной сети;

- технология "клиент-сервер";

- модульность и интегрированность;

- ориентация на отечественных производителей.

В качестве недостатка системы "Галактика" можно выделить то, что управлению качеством не уделено должного внимания.

2.5 Система "ПРО/МИС"

"ПРО/МИС" [9] - интегрированная система управления финансовыми и материальными потоками предприятия российской фирмы "Компьютерные системы для бизнеса". Уникальность системы заключается в наличии аналитических возможностей, позволяющих оценить финансово-экономическое состояние предприятия, на основе расчета показателей кредитоспособности, финансовой устойчивости, накопления капитала. Информация представляется в графической форме. В отличии от тра-

диционного подхода, связанного с накоплением и регистрацией информации, ПРО/МИС решает задачи управления предприятием в результате сквозного прохождения документов через все службы.

Однако данная система не является в полной мере интегрированной системой управления предприятием, так как многие функции планирования, контроля и управления в нее не включены.

2.6 Система "Парус"

В 1997 году на рынок был выпущен программный продукт "Парус 97" [10] российской фирмы "Парус". "Парус 97" - модульная масштабируемая система на 32-разрядной платформе, позволяющая автоматизировать учет отдельных рабочих мест и предприятия в целом. Это программный комплекс, на базе которого может быть построена информационная система управления предприятием, охватывающая процессы планирования, снабжения, складирования, сбыта, маркетинга, бухгалтерского учета, начисления заработной платы и учета персонала. Отличительным свойством системы является открытый интерфейс обмена данными с другими приложениями, в том числе и с собственными разработками заказчика. В системе существует возможность качественного взаимодействия с продуктами Microsoft Office - Word, Excel через механизм OLE, DDE. В системе реализована технология "клиент-сервер", в качестве СУБД используется Oracle 7.1, система написана на Delphi. Важным преимуществом системы является возможность внедрять ее поэтапно, начиная с отдельного модуля, и расширять по мере роста потребностей предприятия, приобретая новые модули и переходя на другие СУБД и платформы.

3 ИНТЕГРАЦИЯ СИСТЕМЫ УПРАВЛЕНИЯ КАЧЕСТВОМ В СИСТЕМУ УПРАВЛЕНИЯ ПРЕДПРИЯТИЕМ

Рассмотрим концепцию интеграции системы управления качеством в интегрированную систему управления предприятием на примере программного продукта R/3 немецкой фирмы SAP AG [4].

Функции управления качеством в системе реализуются в виде модуля Quality Management (QM). Данный модуль интегрирован в систему R/3 в последовательной логистической цепочке (получение коммерческого заказа, приобретение материалов и комплектующих, производство продукции, обслуживание клиента) и обрабатывает задачи планирования, контроля и управления качеством. Стандарты Международной организации по стандартизации (ISO) требуют управления качеством на всех этапах работы предприятия, а также анализа всего технологического процесса вместо контроля и устранения отдельных дефектов. В модуле QM эти принципы реализованы через поддержку следующих функций:

1. На этапе снабжения:

1.1 выбор подходящего поставщика материалов и комплектующих через прослеживание динамики цен и качества поставляемых товаров;

1.2 определение порядка закупки товаров, учитывая сроки и объем поставки;

1.3 регистрация полученных товаров и контроль их качества, модификация внутренних баз данных, содержащих сведения о поставленных материалах и характеристиках поставщика;

1.4 дополнительные функции (уведомление поставщика о дефектности товаров и пр.).

2. На этапе производства:

2.1 планирование производственных операций и операций по проверке качества производимой продукции;

2.2 регистрация сведений о результатах контроля качества выборки через использование контрольных карт;

2.3 уведомление о степени стабильности технологического процесса и наличии случайных и общих причин отклонений.

3. На этапе сбыта (продажи и распределении продукции):

3.1 контроль изготовленной продукции, создание контрольной партии;

3.2 сертификация качества поставляемой продукции;

3.3 обработка предложений и жалоб клиентов.

4. На этапе обслуживания:

4.1 документирование операций поставки товаров;

4.2 анализ дефектов товаров, которые были возвращены, генерация корректирующих действий.

В модуле также поддерживается ряд функций создания сообщений по результатам контроля качества, содержащим сведения о дефектах продукции, причинах их возникновения и принятым действиям.

Ниже предлагается обзор методов и моделей, которые целесообразно включить в систему управления качеством, а также приводится анализ недостатков модуля QM системы R/3.

- Инструментальные методы управления качеством [3], в частности классические "семь инструментов качества" представлены не в полном объеме, хотя эффективное управление качеством требует использование их полного набора (различные комбинации на отдельных этапах для решения определенных задач). Для решения задач управления качеством достаточно эффективно использование статистических методов управления качеством, которые позволяют систематизировать, обрабатывать и исследовать большой объем данных, а также выявлять определенные закономерности, которым они подчиняются [11,12]. Применение этих методов может осуществляться непосредственно в ходе изготовления продукции, а не только на этапе приемочного контроля. Полученная информация может использоваться для анализа причин, приводящих к появлению брака, их устранения, а также для настройки технологического

процесса. Использование статистических методов обработки данных должно стать неотъемлемой частью логистического процесса.

- Реализованы только классические контрольные карты (карты Шухарта и приемочные контрольные карты), в то время как другие типы контрольных карт [13] не представлены вообще, хотя эффективное решение ряда производственных задач достигается именно благодаря использованию этих видов контрольных карт [14, 15, 16].

- Методы экспериментального планирования [17, 18, 19] в модуле QM практически не используются.

- Изучаемый процесс может подвергаться воздействию некоторых внутренних и внешних причин, которые не всегда просто определить. Содержательный смысл и количество такого рода факторов часто бывают неизвестны, непосредственное измерение их величин невозможно. Однако влияние неизвестного фактора может проявляться в нескольких измеряемых признаках, коррелированных между собой. Для выявления таких скрытых, но объективно существующих закономерностей целесообразно использовать методы факторного анализа [20, 21] и разработанные на их основе автоматизированные системы [22].

- Управление стоимостью и качеством также предполагает использование офлайн-методов [23], сфокусированных в большей степени на совершенствовании качества, чем на его оценке. В частности, методы японского ученого Генетси Тагути [24, 25] позволяют уменьшить отклонения рабочих характеристик от заданных значений и создать продукт, устойчивый к внешним воздействиям. Применение таких методов в системе управления качеством имеет смысл на этапе проектирования изделия, когда существует возможность учесть все виды шумовых факторов [26, 27].

- Как отмечается в [28], система управления качеством также должна включать в себя экономико-математические модели для выделения приоритетных задач и рационального использования существующих средств. В [28] предлагается рассматривать эти модели в виде задачи линейного программирования. Действительно, качество изготавливаемой продукции в сочетании с ее ценой можно представить в виде целевой функции, а различные показатели (себестоимость, производственные мощности и др.) в виде ограничений [29].

- Чтобы построить современные системы качества, необходимо пройти определенные этапы, а именно, избавиться от такого серьезного недостатка, как высокий уровень дефектности. Эффективным средством против данного недостатка является диагностика процессов и изделий [30, 31, 32]. Несмотря на наличие большого числа разработанных алгоритмов диагностики нерешенным остается ряд задач. Одной из таких задач является задача классификации по признакам в условиях малого размера обучающей выборки. Если размер выборки

достаточно велик, классы легко делимы и число признаков невелико, то, как правило, используются статистические методы классификации [33, 34, 35]. В тех случаях, когда число признаков велико и классы трудно делимы, могут быть применены, в частности, методы дискриминантных и потенциальных функций [36, 37, 38]. Однако, если размер обучающей выборки мал (в случаях, когда из-за уникальности или дороговизны изделий и сложности испытаний невозможно или нежелательно проводить большое число экспериментов), статистические и эвристические [39] алгоритмы, как правило, не позволяют решать задачу классификации с достаточной достоверностью. Альтернативой традиционным используемым методам при построении моделей на основе обучающих выборок могут служить нейронные сети (НС). Одним из важнейших свойств НС является способность к аппроксимации многомерных функций, что позволяет использовать их для моделирования сложных процессов и объектов. Не менее важное свойство НС - обучаемость дает возможность строить на их основе адаптивные системы [40, 41]. Благодаря этому НС могут применяться в задачах диагностики.

- Необходим анализ функционирования действующей системы управления качеством [42, 43], направленный на получение оценки эффективности системы и определение затрат на качество [44]. Целью такого анализа является выявление и устранение недостатков работы системы, ее модернизация и усовершенствование.

- Необходимо также расширение системы отчетов, включающей информацию о качестве материалов, полуфабрикатов, комплектующих и их поставщиках, о результатах проверки функционирования системы качества и затрат на качество, а также информацию для управления технической документацией.

ВЫВОДЫ

Эффективность использования интегрированных систем управления предприятием существенно повышается, если перечисленные выше функции, методы и модели будут реализованы в автоматизированной системе управления качеством продукции. Внедрение таких систем в производство является необходимым условием для конкурентоспособности предприятия.

ПЕРЕЧЕНЬ ССЫЛОК

1. Бухарова И.А., Дубровин В.И., Захаров Г.А., Карпуть М.И., Табунщик Г.В. Автоматизация процессов управления качеством // Управление в технических системах / Материалы научно-технической конференции. - Ковров: КГТА, 1998, С.120-122.
2. ДСТУ ISO 9000-1-95. Стандарти з управління якістю та забезпечення якості. Ч.1. Настанови щодо вибору та застосування. - К.: Держстандарт України, 1995. - 40с.
3. Всеобщее управление качеством / О.П.Глудкин, Н.М.Горбунов, А.И.Гуров, Ю.В.Зорин. - М: Радио и связь, 1999. - 599с.
4. Руководство по внедрению системы R/3 версии 4.0B фирмы SAP AG.

5. Система управления предприятием CA-PRMS // Железнодорожный транспорт, 1998, №6, С.120-121.
6. Бахарев А., Высочин С., Степанов А., Сургучев О., Фролов Е. Интегрированная система технологической подготовки производства на базе BAAN IV // САПР и графика, 1998, №10, С.10-14.
7. Интегрированные системы управления предприятиями: Инструментальные средства разработки: Материалы конференции Москва 16-18 апреля 1997 г. - М.: Институт проблем управления, 1997. - 106с.
8. Старицкий А. Программные продукты для управления предприятиями // Компьютер Пресс, 1997, №12, С. 5-11.
9. Ломовцева О.Н. Интегрированная система "ПРО/МИС" - управление финансовыми и материальными потоками предприятия // Интегрированные системы управления предприятием / Материалы конференции. - М.: 1997, С.26-32.
10. Травин А. Два "Паруса". // Компьютер Пресс, 1997, №12, С. 48-50.
11. Внуков Ю.Н., Дубровин В.И. Концепция всеобщего менеджмента качества и статистические методы // Высокие технологии в машиностроении / Сборник научных трудов ХГПУ. - Харьков, 1998, С.62-64.
12. Внуков Ю.М., Дубровин В.И., Табунщик Г.В. Керування якістю і зниження варіабельності // Стандартизація, сертифікація, якість, 1999, № 4, С.42-43.
13. Дубровин В.И. Применение контрольных карт для управления качеством // Надежность и контроль качества, 1997, №7, С.52-57.
14. Дубровин В.И., Молинчук С.Г. Исследование роботизированной линии с помощью контрольных карт // Совершенствование процессов финишной обработки в машино- и приборостроении, экология и защита окружающей среды / Тезисы докладов международной научно-технической конференции. - Минск: БГПА, 1995, С.39.
15. Дубровин В.И., Колесник А.П. Применение контрольных карт кумулятивных сумм для анализа отказов роботизированных систем // Контроль и управление в технических системах / Тезисы докладов научно-технической конференции стран СНГ. - Винница: ВПИ, 1992, С.187.
16. Дубровин В.И., Рысигов В.П. Анализ работы роботизированной линии с помощью карт долей дефектов для выбора меняющегося объема // Методы и средства повышения надежности приборов, устройств и систем / Тезисы докладов международной научно-технической конференции, часть 2, - Пенза: ПДНТП, 1993, С.7-8.
17. Дубровин В.И. Регрессионные модели с одинаковой точностью прогнозирования отклика в задачах оптимизации технологических процессов производства РЭС // Цифровые модели в проектировании и производстве РЭС / Межвузовский сборник научных трудов. - Пенза: Пензенский государственный технический университет, 1994, №5, С.58-60.
18. Дубровин В.И. Моделирование электромеханических узлов БМЗ с использованием процедуры пошаговой регрессии // Методы и средства повышения надежности приборов, устройств и систем / Тезисы докладов международной научно-технической конференции, часть 1, - Пенза: ПДНТП, 1993, С.48-50.
19. Дубровин В.И. Використання активно-пасивного експерименту для отримання регресійних моделей у промислових умовах // Сучасні проблеми засобів телекомунікації, комп'ютерної інженерії та підготовки спеціалістів / Матеріали міжнародної науково-технічної конференції TCSET'98. - Львів: Держ. ун-т "Львівська політехніка", 1998, С.6-7
20. Дубровин В.И., Емчицкий В.Т. Факторный анализ технологических процессов // Нечеткая логика, интеллектуальные системы и технологии / Материалы международной научно-технической конференции. - Владимир, Владимирский государственный университет, 1998, С.139-140.
21. Дубровин В.И., Емчицкий В.Т., Лунев В.В., Лебедь В.И. Применение факторного анализа для оценки состояния технологии непрерывной выплавки черных металлов и ферросплавов // Пути повышения качества и экономичности литейных процессов. - Одесса: Совпін, 1996, С.57-58.
22. Дубровин В.И., Емчицкий В.Т., Онацко А.Г. Автоматизированная система факторного анализа // Електротехніка та електроенергетика, 1999, № 2, С. 62-67.
23. Дубровин В.И., Табунщик Г.В. Онлайн-методы управления качеством. // Радиоэлектроника. Информатика. Управление, 1999, №1, С.120-123.
24. Дубровин В.И., Табунщик Г.В. Робастное проектирование по Тагути // Надежность и качество. Инновационные технологии производству XXI века / Книга докладов Международного симпозиума, посвященного 275-летию Российской Академии наук. - Пенза, Пенз. гос. университет, 1999, С.149-151.
25. Дубровин В., Табунщик Г. Робастне проектування і керування якістю. // Досвід розробки і застосування САПР в мікроелектроніці / Матеріали п'ятої міжнародної науково-технічної конференції CADSM'99. - Львів, Державний університет "Львівська політехніка", 1999, С.20-21.
26. Внуков Ю.Н., Дубровин В.И., Табунщик Г.В. Робастные методы обеспечения качества при проектировании // Машиностроение и техносфера на рубеже XXI века / Сборник трудов VI Международной научно-технической конференции. - Донецк, ДонГТУ, 1999, С.140-142.
27. Dubrovin V., Tabunshchik G. Loss Function as Tolerance Optimization Criteria // Proceedings of International Conference on Modern Problem of Telecommunications, Computer Science and Engineers Training TCSET'2000, February 14-19, 2000, Lviv-Slavsko, Ukraine, pp. 45-46.
28. Попов В. Підсистема економіко-математичних моделей у системі управління якістю // Стандартизація. Сертифікація. Якість, 1999, №1, С. 62-63.
29. Дубровин В.И. Приведение задачи минимизации коэффициента детонации ЛПМ к задаче линейного программирования // Сучасні проблеми автоматизованої розробки і виробництва радіоелектронних засобів та підготовки інженерних кадрів / Матеріали міжнародної науково-технічної конференції, присвяченої 150-річчю Державного університету "Львівська політехніка", 1 частина - Львів: Держ. ун-т "Львівська політехніка", 1994, С.44-46
30. Дубровин В., Степаненко О., Луценко Д. Диагностика процессов та керування якістю // Досвід розробки і застосування САПР в мікроелектроніці / Матеріали п'ятої міжнародної науково-технічної конференції CADSM'99. - Львів, Державний університет "Львівська політехніка", 1999, С.22-24.
31. Дубровин В.И. Идентификация и оптимизация сложных технических процессов и объектов. - Запорожье: ЗГТУ, 1997.- 92 с.
32. Дубровин В.И., Корецкий Н.Х. Прогнозирование надежности элементов РЭС на основе информативных параметров // Теория и практика обеспечения надежности и качества радиоэлектронных средств / Сборник научных трудов. - УМК ВО, 1992, С.90-95.
33. Внуков Ю.Н., Дубровин В.И. Методики прогнозирования с использованием теории статистических оценок и статистической классификации // Высокие технологии в машиностроении: диагностика процессов и обеспечение качества / Материалы VI международного научно-технического семинара. - Харьков: ХГПУ, 1996, С.26-27.
34. Внуков Ю.Н., Гамов Н.С., Дубровин В.И. Методика применения теории статистических оценок для управления качеством // Методологические проблемы качества обучения и обучения качеству / Материалы научно-технической конференции. - Харьков: ХАИ, 1996, С.85.
35. Дубровин В.И., Корецкий Н.Х. Методика оптимальной классификации // Методологические проблемы качества обучения и обучения качеству / Материалы научно-методической конференции. - Харьков: ХАИ, 1997, С.34-35.
36. Внуков Ю.Н., Дубровин В.И. Алгоритм классификации с использованием дискриминантных функций // Высокие технологии в машиностроении / Сборник научных трудов ХГПУ. - Харьков, 1998, С.64-66.
37. Дубровин В.И., Корецкий Н.Х. Об одном подходе к прогнозированию надежности изделия // Электронное моделирование, 1986, т.8, № 6, С.97-98.
38. Дубровин В.И. О решении задачи обучения распознаванию образов методами потенциальных и дискриминантных функций // Методы прогнозирования надежности проектируемых РЭА и ЭВА / Тезисы докладов к зональной конференции. - Пенза: ПДНТП, 1988, С.47-48.
39. Дубровин В.И. Эвристические алгоритмы классификации // Машиностроитель, 1998, №7, С.6-9.
40. 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.
41. Dubrovin V., Subbotin S. Choice of neuron transfer functions and research of their influence for learning quality of neural networks // Proceedings of International Conference on Modern Problem of Telecommunications, Computer Science and Engineers Training TCSET'2000, February 14-19, 2000, Lviv-Slavsko, Ukraine, pp. 114-115.
42. Дубровин В., Захаров Г., Карпуть М., Табунщик Г. Автоматизована система "Аналіз функціонування елементів системи якості" // Стандартизація. Сертифікація. Якість, 1998, №2, С. 54-55.
43. Дубровин В.И., Захаров Г.А., Карпуть М.И., Табунщик Г.В. Анализ функционирования элементов системы качества предприятия. // Проблемы сертификации и управления качеством: Тезисы докладов всероссийской конференции, часть 1/ под ред. Л.И. Волгина. - Ульяновск: УлГТУ, 1998, С. 41-43
44. Бухарова И., Дубровин В., Захаров Г., Карпуть М., Табунщик Г. Автоматизація аналізу витрат на якість // Стандартизація. Сертифікація. Якість, 1998, №3, С. 59-60.

Надійшла 24.02.2000
Після доробки 28.02.2000

ПРОЕКТИРОВАНИЕ КАЧЕСТВА: ДОПУСКОВЫЙ АНАЛИЗ

В. И. Дубровин, Г. В. Табунщик

Предложен метод решения задачи синтеза точности при проектировании с использованием дисперсионного анализа и стоимостного анализа, основанного на применении функции потерь.

Запропоновано метод вирішення задачі синтезу точності при проектуванні з використанням дисперсійного аналізу і вартісного аналізу, заснованого на застосуванні функції збитків.

The technique of solving the problem of precision synthesis based on the analyze of variance and cost analyze based on Loss Function are offered in the article.

ВВЕДЕНИЕ

Эффективным средством решения проблемы качества отечественной продукции являются офлайн-методы - методы, не встроенные в процесс производства, позволяющие улучшить качество изделий и их технологичность, а также снизить стоимость разработки, изготовления, эксплуатации и ремонта. Эти методы являются техническим средством обеспечения качества на стадии проектирования изделий и разработки технологического процесса [1].

Одной из главных задач, которые должны быть решены для реализации методики офлайн-управления качеством, является задача определения оптимальных значений и допустимых отклонений параметров.

ЗАДАЧА СИНТЕЗА ТОЧНОСТИ ПРИ ПРОЕКТИРОВАНИИ

Проблема установления допусков является очень важной и освещена в значительном количестве публикаций, среди которых большой удельный вес занимают работы, посвященные анализу и расчету допусков определенных типов изделий.

Теория определения и оптимизации допусков возникла при переходе от ремесленного к серийному, а затем и к массовому производству как теоретическая основа взаимозаменяемости [2]. Впервые принцип оценки годности изделия на основе контроля его параметров предельными размерами был разработан Ф. Тейлором и в дальнейшем получил развитие в машиностроении, радиоприборостроении, приборостроении.

Среди фундаментальных работ, посвященных решению задачи анализа и расчета допусков, в первую очередь, следует назвать труды Н.Г. Бруевича [3], Н.А. Бородачева [4], М.Л. Быховского [5].

В отечественных разработках задачи определения и

оптимизации допусков рассматривались как круг задач теории точности. Хотя, возможно, данные задачи должны составлять отдельный круг задач и являться базисом для взаимодействия между конструктором и технологом в деле непрерывного совершенствования создаваемого изделия.

Из проблем, рассматриваемых теорией точности, для решения задач допускового анализа выделим:

- задачу синтеза точности при проектировании
- задачу анализа точности при проектировании.

В задаче синтеза точности при проектировании по заданным значениям предельных отклонений ESy и EIy (ES - верхний предел, EI - нижний предел) выходной характеристики y , требуется найти предельные отклонения ESx_i и EIx_i и, следовательно, $\Delta_{x_i} = ES_{x_i} - EI_{x_i}$ допусков входных параметров x_i .

В задаче анализа точности при проектировании по значениям предельных отклонений входных параметров ESx_i и EIx_i , номинальных значений x_n , y_n необходимо определить ESy и EIy или значение поля допуска Δ_y выходной характеристики.

Запишем аналитический вид зависимости между выходной характеристикой и входными параметрами с помощью выражения:

$$y = f(x_1, x_2, \dots, x_m). \quad (1)$$

Используя классификацию, предложенную в [6], будем рассматривать многомерные множественные зависимости (одна выходная характеристика и несколько входных параметров) по характеру как линейные так и нелинейные, по природе как детерминированные так и стохастические.

При решении задач оптимизации допусков на стадии проектирования возникает ряд трудностей - на данном этапе не известен аналитический вид функции распределения входных параметров. Кроме того, не известны дисперсии и стандартные отклонения. Следовательно, конструкторы оперируют не дисперсиями σ_i , а полями допусков Δ_i . Последние приравниваются к полям рассеивания δ_i , которые в нормально протекающем производстве равны или меньше полей допусков. Зависимость между параметрами δ_i и σ_i выражается следующей зависимостью:

$$\delta_x = t_x \sigma_x, \quad (2)$$

где t_x - положительное число, которое задает вероятность того, что значение x будет вне границ поля рассеяния.

Для решения задач определения допусков на практике широко используются метод максимума-минимума и метод Монте-Карло.

Метод Монте-Карло, при котором с помощью датчика случайных чисел аргументам задаются случайные значения, используется для моделирования распределения случайной величины.

Однако для нелинейных и многомерных задач расчет по принципу максимума-минимума не разработан. Метод Монте-Карло имеет ряд недостатков, достаточно полно описанных в литературе [6, 7].

Для использования вероятностных методов определения допусков необходимо знать законы и параметры распределения как выходной характеристики так и входных параметров, которые при нелинейных исходных зависимостях традиционными методами получить довольно трудно.

Классический метод решения задач анализа и синтеза точности при проектировании можно свести к решению выражения:

$$\Delta_y = \sqrt{\sum_{i=1}^m a_{yi}^2 \Delta_{xi}^2}, \quad (3)$$

где a_{yi} - функция чувствительности, отражающая степень воздействия входных параметров на выходную характеристику. Наиболее широко используемое выражение для определения значения данной функции -

$$a_{yi} = \frac{\partial y}{\partial x_i}.$$

Использование данного метода требует наличия следующих условий:

- а) знания исходной зависимости (1);
- б) только одной выходной характеристики y ;
- с) допустимости линеаризации зависимости (1).

В настоящее время методика формализованного решения задач синтеза точности при проектировании разработана недостаточно [6, 8], и интуиция и опыт конструктора и технолога при решении данного круга задач играет главную роль. Разработанные методы распространяются только на линейную исходную зависимость (1), кроме того возникают трудности при определении количественной оценки влияния каждого входного параметра на выходную характеристику и нет метода, позволяющего выбрать наилучшее решение из множества возможных.

Следовательно, разработка методики решения задачи синтеза точности при проектировании с учетом нелиней-

ных исходных зависимостей между входными параметрами и выходной характеристикой является актуальной.

АНАЛИЗ ЗАВИСИМОСТЕЙ МЕЖДУ ВЫХОДНОЙ ХАРАКТЕРИСТИКОЙ И ВХОДНЫМИ ПАРАМЕТРАМИ

Пусть задана функциональная зависимость (1), известны номинальные значения входных параметров, полученных на этапе проектирования параметров [9], известны допустимые отклонения выходной характеристики y - ESy и EIy или Δ_y . Будем считать, что превышение границ допусков ведет к отказу и, следовательно, требуется ремонт или регулировка параметров, что приводит к потерям A_0 .

Так как в реальных процессах доминируют нелинейные зависимости между входными параметрами и выходными характеристиками, то для решения задачи синтеза точности при проектировании требуется выполнить преобразование исходной зависимости (1).

Для линеаризации нелинейных зависимостей традиционно используется разложение Тейлора, при котором выражение (1) линеаризуется в окрестности некоторой точки $(m+1)$ -мерного пространства. Величина этой области определяется допусками параметров. Однако разложение Тейлора имеет ряд недостатков [6]:

- 1) требуется наличие непрерывных производных в рассматриваемой области;
- 2) при разложении в ряд в окрестности точки с абсциссами $(mx_1, mx_2, \dots, mx_m)$ (где mx_1, mx_2, mx_m - математические ожидания аргументов x_1, x_2, \dots, x_m $(m+1)$ -мерного пространства) ошибка приближения на концах заданных интервалов достигнет максимума;
- 3) затруднен анализ сходимости ряда.

Кроме того, для некоторых задач линеаризация исходной зависимости недопустима. В этом случае используется приближение функций полиномами. Для решения задачи допускового анализа наиболее эффективным инструментом является использование ортогональных многочленов Чебышева, которые в общем виде представляются как полиномиальные функции от уровней варьирования входных параметров. Степень полинома не может быть выше, чем число уровней варьирования минус 1. Разложение Чебышева сходится быстрее, чем разложение Тейлора, т.е. точность аппроксимации полиномами Чебышева выше, чем при любом другом приближении при одинаковом количестве членов [10].

Таким образом, для каждого входного параметра выбирается число уровней его варьирования, которое должно быть на 1 больше максимальной степени полинома, необходимой для описания этим параметром поверхности отклика.

Для определения границ допусков не требуется более трех уровней варьирования, значения которых приведе-

ны в табл. 1, где x_{iH} - номинальное значение i -го входного параметра, j - номер уровня:

Таблица 1 - Уровни варьирования параметров k

j/k	$k=2$	$k=3$
1	$x_{iH} - \sigma_i$	$x_{iH} - \sqrt{3/2} \times \sigma_i$
2	$x_{iH} + \sigma_i$	x_{iH}
3		$x_{iH} + \sqrt{3/2} \times \sigma_i$

ДИСПЕРСИОННЫЙ АНАЛИЗ

В качестве планов для проведения экспериментов предлагается реализовать "Латинский квадрат" или "Ортогональные таблицы" [9,11], что позволяет использовать традиционную схему дисперсионного анализа (ДА), которая состоит из следующих шагов.

Для проведения дисперсионного анализа вычисляем:

1) фактор коррекции

$$CF = \frac{\left(\sum_{u=1}^n y_u \right)^2}{n}, \tag{4}$$

где n - количество опытов, $u = \overline{1, n}$;

2) общую сумму квадратов

$$SS_{общ} = \sum_{u=1}^n y_u^2 - CF; \tag{5}$$

3) суммы квадратов для исследуемых параметров

$$SS_{xi} = b_i \sum_{u=1}^n x_{iu} y_u - CF = b_i^2 n - CF, \tag{6}$$

где $b_i = \sum_{u=1}^n (x_{iu} y_u) / n$, $i = \overline{1, m}$;

4) вычисляем общий средний квадрат $MS_{общ}$

$$MS_{общ} = \frac{SS_{общ}}{df_{общ}}, \tag{7}$$

и, соответственно, MS_{xi} для каждого параметра:

$$MS_{xi} = \frac{SS_{xi}}{df_{xi}}, \tag{8}$$

где df_{xi} - степень свободы;

5) для каждого входного параметра вычисляем степень влияния на выходную характеристику:

$$a_{yi} = \frac{SS_{xi}}{SS_{общ}}. \tag{9}$$

Результаты ДА сводятся в табл. 2.

Таблица 2 - Таблица дисперсионного анализа

Источник дисперсии	df	SS	MS	a_{yi} (%)
x_1	$k-1$	SS_{x1}	$\frac{SS_{x1}}{df_{x1}}$	$\frac{SS_{x1}}{SS_{общ}}$
x_2	$k-1$	SS_{x2}	$\frac{SS_{x2}}{df_{x2}}$	$\frac{SS_{x2}}{SS_{общ}}$
...
x_m	$k-1$	SS_{xm}	$\frac{SS_{xm}}{df_{xm}}$	$\frac{SS_{xm}}{SS_{общ}}$
Общая сумма квадратов	$m-1$	$SS_{общ}$	$\frac{SS_{общ}}{df_{общ}}$	100

В результате ДА мы получаем степень воздействия каждого параметра на выходную характеристику.

СТОИМОСТНОЙ АНАЛИЗ

Так как мы рассматриваем задачу синтеза точности при проектировании, то в качестве исходных данных рассматриваются отклонения ESy и EIy выходного параметра. Обозначим $\Delta_y = ESy - EIy$. Так как при решении задач синтеза существует множество решений, то в качестве критерия оптимальности допускового решения примем затраты. Пусть A_y - потери, когда выходная характеристика y не соответствует установленным допускам.

Для дальнейших вычислений воспользуемся функцией потерь [11]:

$$L = \gamma \Delta^2. \tag{10}$$

Поставив в выражение (10) значения A_y и Δ_y , найдем

значение γ :

$$\gamma = \frac{A_y}{\Delta_y^2}. \quad (11)$$

Для определения потерь для i -го параметра [12] воспользуемся данными, полученными в ходе дисперсионного анализа, в результате получим выражение:

$$A_{xi} = \frac{A_0}{\Delta_y^2} \times a_{yi} \times MS_{\text{общ}}^2. \quad (12)$$

На основании полученных потерь принимается решение о расширении или уменьшении границ допуска. Если принимается решение об изменении границ допуска, то допуск для входных параметров возможно определить из выражения:

$$\Delta_{xi} = \sqrt{\frac{A_{xi}}{A_0}} \times \frac{\Delta_y}{a_{yi}}. \quad (13)$$

Уравнение (13) является модификацией выражения (3) для решения задачи синтеза точности при проектировании.

ВЫВОДЫ

Методика, предложенная в данной работе, предоставляет удобный математический аппарат для решения такого круга задач допускового анализа, как задача синтеза точности при проектировании. Использование дисперсионного анализа и ортогональных полиномов Чебышева позволяет определить степень воздействия входных параметров на выходную характеристику. Применение

теории планирования эксперимента позволяет на данном этапе решения задачи синтеза при проектировании использовать компьютерное моделирование, используя пакеты статистической обработки данных (Statistica, Statgraphics, SPSS, Matlab и т.д.). Использование функции потерь позволяет ввести единый экономический критерий оптимальности допусковых решений - стоимость (потери).

ПЕРЕЧЕНЬ ССЫЛОК

1. Дубровин В.И., Табунщик Г.В. Офлайнные методы управления качеством //Радиоэлектроника. Информатика. Управление, - 1999. - №1. - С.120-123.
2. Карепин П.А. От погрешности к допуску от допуска к погрешности //Надежность и контроль качества, - 1998. - №10. - С.24-34.
3. Бруевич Н.Г., Правоторова Е.А., Сергеев И.В. Основы теории точности механизмов. М.:Наука, 1988. - 238с.
4. Бородачев Н.А. Основные вопросы точности производства. М.:Изд-во АН СССР, 1959. -416с.
5. Быховский М.А. Основы динамической точности электрических и механических цепей. М.: Изд-во АН СССР, 1958. - 315с.
6. Фридендер И.Г. Расчеты точности машин при проектировании. - К.: Вища школа, 1980. - 184с.
7. Brayton, R. K., Hatchet. G. D., Sangiovanni-Vincentelli A. L. 'A survey of optimization techniques for integrated-circuit design'.// Proc. IEEE, 1981, Vol. 69. pp. 1334-1362.
8. Иншаков А.Н., Иншаков С.А. Допусковый анализ при проектировании сложных технических систем //Информационные технологии. - 1997. - №1. - С.34-39.
9. Дубровин В.И., Табунщик Г.В. Робастное проектирование по Тагучи //Надежность и качество. Инновационные технологии производству XXI века/Книга докладов международного симпозиума, посвященного 275-летию Российской Академии наук. - Пенза: ПГУ, 1999, С.149-151
10. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. Изд-е 3-е. - М.: Наука, - 1983. - 416с.
11. Taguchi G. Introduction to Quality Engineering: Designing Quality Into Products And Process. Tokyo. Asian: Productivity Organization, 1986, 191p.
12. Dubrovin V., Tabunshchik G. Loss Function as Tolerance Optimization Criteria //Proceedings of International Conference on Modern Problem of Telecommunications, Computer Science and Engineers Training TCSET'2000, February 14-19, 2000, Lviv-Slavsko, Ukraine, pp. 45-46.

Надійшла 07.02.2000

Після доробки 17.03.2000

УДК 621.3.05: 519.2

УНИФИЦИРОВАННЫЙ МЕТОД РАСЧЕТА ХАРАКТЕРИСТИК РЕАКЦИИ ОБЪЕКТА НА СЛУЧАЙНЫЕ ВОЗДЕЙСТВИЯ МАЛОЙ ИНТЕНСИВНОСТИ

Э. Г. Куренный, Ю. И. Чепкасов, Л. В. Черникова

Рассматривается задача определения реакций при воздействиях случайных помех на объекты. Линеаризацию предлагается осуществлять методом параметрической идентификации. Для расчета характеристик искомой реакции система представляется в виде параллельно включенных инерционных звеньев. Эффективность методов иллюстрируется на примере оценки влияния изменений напряжения, подаваемого на синхронный электродвигатель.

Розглядається задача визначення реакцій у зв'язку з діями випадкових завод на об'єкти. Лінеаризацію пропонується виконувати методом параметричної ідентифікації. Для розрахунку характеристик шуканої реакції система зображається

у вигляді паралельно увімкнених інерційних ланок. Ефективність методів ілюструється на прикладі оцінки впливу змін напруги на синхронному двигуні.

The task of reactions determination by influence of random interferences on objects is considered. It is suggested to accomplish linearization by method of parametric identification. System is presented in the form of connected in parallel linear inertial links for calculation of unknown reaction characteristics. The effectiveness of methods is illustrated on example of estimation of voltage alterations influence on synchronous electric motor.

ПОСТАНОВКА ЗАДАЧИ

В общем случае математические модели радиотехнических устройств и электротехнических комплексов представляют собой систему нелинейных дифференциальных уравнений, решение которых затруднено даже при периодических помехах. При помехах $X(t)$ малой интенсивности возможна линеаризация моделей [1].

Обычно осуществляют прямую линеаризацию путем замены нелинейной модели соответствующей системой линейных дифференциальных уравнений. Однако при этом сохраняется громоздкая структура исходной модели и высокий порядок системы уравнений. Кроме того, даже при одном уравнении с порядком $n \geq 3$ определение характеристик реакции $Y(t)$ объекта на помеху сопряжено с громоздкими выкладками.

В статье рассматриваются две задачи: выбор линеаризованной модели методом параметрической идентификации и унификация линейных моделей с целью существенного упрощения расчетов. Для систем простой структуры параметрическая идентификация может и не потребоваться. Для определенности эффективность методов иллюстрируется на примере оценки электромагнитной совместимости (ЭМС) по отклонениям напряжения электротехнического комплекса "сеть электропитания - синхронный двигатель", в котором затрагиваются все аспекты проблемы. Помеху будем считать стационарной, но все выводы очевидным образом распространяются и на нестационарные помехи.

МЕТОД ПАРАМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ

Среднее значение Y_c и корреляционная функция (КФ) $K_Y(\tau)$ процесса на выходе линейной модели определяются по соответствующим характеристикам X_c и $K_X(\tau)$ помехи через переходную $h(t)$ или весовую $g(t)$ функции модели. Поэтому при линеаризации необходимо воспроизвести именно эту функцию. Сущность предлагаемого метода состоит в подборе простой линейной модели по переходной функции, рассчитанной по исходной модели. Количество переходных функций равно количеству выходов модели. Переходные функции могут быть получены экспериментально или методами имитации.

Для реализации метода вначале доказывается возможность линеаризации. С этой целью по нелинейной модели находятся реакции рассматриваемого объекта на скачкообразные возмущения разной величины и знака. По совокупности реакций устанавливается диапазон изменения помехи, в пределах которого реакции пропорциональны величине скачка. В этом диапазоне допустимо использовать понятие переходной функции.

Достоверность метода обеспечивается тем, что пере-

ходные функции $h_n(t)$ рассчитываются по исходной модели и с заданной точностью аппроксимируются простыми аналитическими выражениями $h(t) \approx h_n(t)$. В первую очередь целесообразно проверить возможность аппроксимации в виде линейной комбинации экспоненциально-синусоидальных и экспоненциальных функций.

По $h(t)$ находится преобразование Лапласа в виде изображения $H(s)$. Это позволяет определить функцию

$$W(s) = sH(s)$$

искомой линейной модели и ее структуру.

Основной эффект упрощения анализа реакций заключается в следующем. При прямой линеаризации получается одна линейная система с несколькими выходами и внутренними связями между ее структурными блоками. Предлагаемый метод приводит к нескольким системам, количество которых равно количеству выходов. Но эти системы просты и не имеют взаимных структурных связей. Это позволяет получать аналитические решения задач, избегая численных методов. Следует отметить, что между выходными процессами сохраняются такие же связи, как и в исходной модели, но они проявляются неявно через параметры линейных систем.

Рассмотрим пример с электротехническим комплексом, ограничившись лишь анализом изменения отклонений активной мощности от заданного значения при изменении напряжения U в сети. Полная модель только одного синхронного электродвигателя с учетом насыщения является довольно сложной [2]. К ней нужно добавить модели электропривода, сети электроснабжения и системы возбуждения. При наличии нескольких синхронных двигателей сложность модели возрастает еще больше.

Расчеты численными методами по исходной нелинейной модели комплекса показали, что в допускаемом стандартом [3] диапазоне изменения напряжения $\pm 5\%$ линеаризация возможна. На рис. 1 показана переходная характеристика $h_n(t)$. Оказалось, что изменение условий протекания переходных процессов (режим возбуждения, сопротивление сети, момент на валу) сказывается только на числовых параметрах переходной функции. Вид же переходной функции не изменяется и определяется выражением

$$h(t) = B e^{-\alpha_1 t} \sin \omega_{\text{ц}} t, \quad (1)$$

в котором параметры B , α_1 и $\omega_{\text{ц}}$ подбираются по $h_n(t)$. Угловая частота $\omega_{\text{ц}} = 2\pi/t_{\text{ц}}$ вычисляется по длительности любого цикла. Остальные два параметра находятся по любым двум точкам $h_n(t)$, которые дают легко решаемую систему двух алгебраических уравнений.

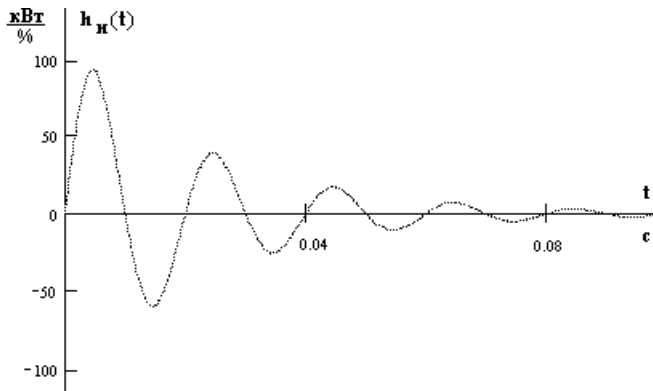


Рисунок 1 - Переходная характеристика

Выражение (1) используется и для аппроксимации переходных функций нескольких синхронных двигателей.

Обозначив $\alpha_2^2 = \alpha_1^2 + \omega_{ц}^2$, $T_1 = 2\alpha_1/\alpha_2^2$, $T_2 = 1/\alpha_2^2$ и $T_3 = \omega_{ц}/\alpha_2^2$, по таблицам преобразований Лапласа найдем передаточную функцию

$$W(s) = \frac{BT_3s}{T_2^2s^2 + T_1s + 1}, \quad (2)$$

которая соответствует функции времени (1).

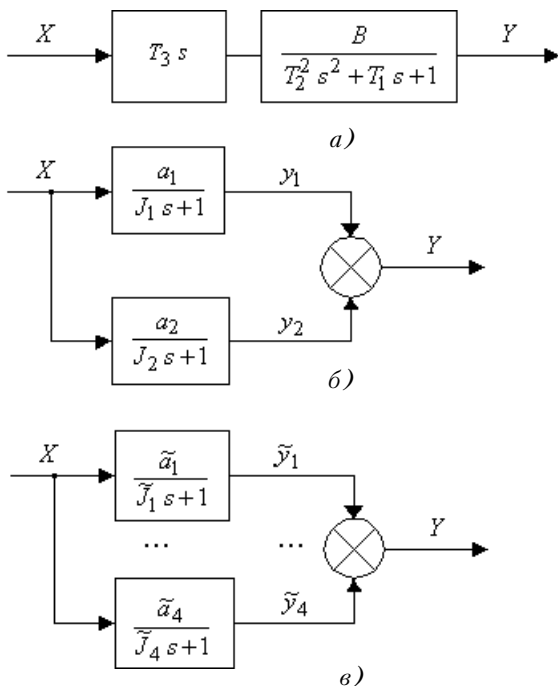


Рисунок 2 - Преобразование исходной системы (а) при частичной (б) и полной (в) унификации

Таким образом, исходная нелинейная система заменена простой линейной системой, состоящей из двух элементарных звеньев: дифференцирующего с постоянной времени T_3 и колебательного с коэффициентом передачи B и постоянными времени T_1 и T_2 (рис. 2, а).

МЕТОД "ПАРЦИАЛЬНЫХ РЕАКЦИЙ"

При стационарной случайной помехе характеристики реакции линейной системы могут быть определены по общим формулам теории случайных процессов [1]

$$Y_c(t) = X_c h(t), \quad (3)$$

$$K_Y(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} g(u)g(v)K_X(\tau + u - v)dudv, \quad (4)$$

где u и v - переменные интегрирования, $\tau = t_2 - t_1$. При большом порядке линейного дифференциального уравнения (например, фильтр фликерметра [4] имеет $n = 4$) применение формулы (4) затруднено.

Упрощение расчетов возможно путем применения метода "парциальных реакций", предложенного в [5] для периодических помех. Идея метода состоит в том, что исходная линейная система (рис. 2, а) представляется в виде n параллельно включенных инерционных (апериодических звеньев) звеньев первого порядка (рис 2, б). Метод применим для систем с передаточной функцией $W(s)$, знаменатель $z(s)$ которой имеет простые корни s_1, \dots, s_n и порядок n , меньший порядка m числителя $q(s)$. Если $m \geq n$ или имеются кратные корни знаменателя, система разделяется на две или более подсистем, а метод применяется последовательно.

Знаменатель при $i = 1, \dots, n$ представляется в виде произведения множителей $s - s_i$ с коэффициентом T_n^n . В этом случае коэффициент передачи и постоянная времени i -го звена определяются выражениями

$$a_i = \frac{-q(s)}{T_n^n s_i \prod_{l=1}^n (s - s_l)} \Bigg|_{s=s_i}, \quad J_i = \frac{1}{s_i}, \quad (5)$$

которые могут быть комплексными. Переходная и весовая функции звена при $\gamma_i = \frac{1}{J_i} = -s_i$ составят

$$h_i(t) = a_i(1 - e^{-\gamma_i t}), \quad g_i(t) = a_i \gamma_i e^{-\gamma_i t}. \quad (6)$$

Искомая реакция равна сумме парциальных реакций $y(t)$ инерционных звеньев. Эти реакции коррелирован-

ны, так как на вход каждого звена поступает одна и та же помеха. Зависимости реакций не сказываются на среднем значении

$$Y_c(t) = X_c \sum_{i=1}^n a_i (1 - e^{-\gamma_i t}), \quad (7)$$

но должны учитываться при определении КФ реакции.

Взаимная КФ между i -ой r -ой реакциями, с учетом (6), определяется по формуле

$$k_{ir}(t_1, t_2) = a_i a_r \gamma_i \gamma_r \int_0^{t_1} \int_0^{t_2} e^{-\gamma_i u} e^{-\gamma_r v} K_X(\tau + u - v) du dv, \quad (8)$$

по структуре аналогичной (4). КФ i -ой реакции найдем согласно (4);

$$k_{y_i}(t_1, t_2) = a_i^2 \gamma_i^2 \int_0^{t_1} \int_0^{t_2} e^{-\gamma_i(u+v)} K_X(\tau + u - v) du dv. \quad (9)$$

Искомые КФ реакции

$$K_Y(t_1, t_2) = \sum_{i=1}^n k_{y_i}(t_1, t_2) + \sum_{i \neq r}^n k_{ir}(t_1, t_2) + \sum_{r \neq i}^n k_{ri}(t_1, t_2). \quad (10)$$

Формулы (8) и (9) для каждого вида КФ помехи дают аналитические выражения, которые являются исходными для применения метода. Тем самым достигается существенное упрощение расчетов, так как пользователю уже не нужно применять формулы (3) и (4).

Унификация расчетов по формулам (7) - (9) не является полной, поскольку каждому виду КФ соответствуют разные выражения. Если рассматривать только стационарные режимы, то можно достичь полной унификации. Для этой цели используется известное положение о том, что стационарная помеха есть результат прохождения белого шума $\xi(t)$ с постоянной спектральной плотностью c через гипотетическую линейную систему, передаточная функция $W_\xi(s)$ которой определяется видом помехи. С учетом этого вместо исходной системы получаем "составную" линейную систему с передаточной функцией

$$W_c(s) = W_\xi(s)W(s),$$

на вход которой поступает не помеха, а белый шум.

Применяя к $W_c(s)$ метод "парциальных реакций", также получим систему параллельно включенных инерционных звеньев (рис. 2, в), но с большим их количе-

ством и другими параметрами a и \tilde{J} (символ \sim) - за счет добавления m звеньев от $W_\xi(s)$.

Для воспроизведения среднего значения помехи необходимо, чтобы белый шум имел среднее значение

$$\xi_c = X_c / h_\xi(\infty),$$

где $h_\xi(\infty)$ - переходная функция гипотетической системы.

Среднее значение реакции определяется по аналогичной (7) формуле

$$Y_c = \xi_c \sum_{i=1}^{n+m} \tilde{a}_i. \quad (11)$$

Так как КФ белого шума равна $\pi c \delta(\tau)$, то при $t \rightarrow \infty$ интегрирование согласно (8) и (9) дает следующие выражения для КФ и взаимных КФ парциальных реакций:

$$\tilde{k}_i(\tau) = \pi c \frac{\tilde{a}_i^2 \tilde{\gamma}_i}{2} e^{-\tilde{\gamma}_i |\tau|}, \quad \tilde{k}_{ir}(\tau) = \pi c \tilde{a}_i \tilde{a}_r \frac{\tilde{\gamma}_i \tilde{\gamma}_r}{\tilde{\gamma}_i + \tilde{\gamma}_r} e^{-\tilde{\gamma}_r |\tau|}. \quad (12)$$

С учетом (10), искомая КФ реакции

$$K_Y(\tau) = \pi c \left[\frac{1}{2} \sum_{i=1}^{n+m} \tilde{a}_i^2 \tilde{\gamma}_i e^{-\tilde{\gamma}_i |\tau|} + \sum_{i \neq r} \tilde{a}_i \tilde{a}_r \frac{\tilde{\gamma}_i \tilde{\gamma}_r}{\tilde{\gamma}_i + \tilde{\gamma}_r} (e^{-\tilde{\gamma}_r |\tau|} + e^{-\tilde{\gamma}_i |\tau|}) \right] \quad (13)$$

В последующих выкладках постоянная c сокращается, поэтому можно сразу положить $c = 1$ с размерностью спектральной плотности.

Простота и универсальность формул (12) и (13) еще больше упрощает расчеты, несмотря на увеличение количества звеньев.

Пример. Оценить дополнительные потери активной мощности в синхронном двигателе от изменений напряжения в сети. Параметры передаточной функции (2) следующие:

$$B = 112, 8 \text{ кВт}/(\%), \quad \alpha_1 = 42, 6 \text{ с}^{-1}, \quad \omega_\omega = 314 \text{ с}^{-1}, \quad T_1 = 8, 485 \cdot 10^{-4} \text{ с}, \quad T_2 = 3, 156 \cdot 10^{-3} \text{ с}, \quad T_3 = 3, 127 \cdot 10^{-3} \text{ с}.$$

Напряжение в сети имеет КФ

$$K_U(\tau) = D_U e^{-\alpha |\tau|} \left(\cos \omega_0 \tau + \frac{\alpha}{\omega_0} \sin \omega_0 |\tau| \right) \quad (14)$$

с параметрами $D_U = \sigma^2_U = 9 (\%)^2$, $\alpha = 0, 8 \text{ с}^{-1}$, $\omega_0 = 2 \text{ с}^{-1}$.

Искомые дополнительные потери в среднем равны стандарту $\sigma_{\Delta P}$ процесса на выходе модели. Решим задачу методом полной унификации. Гипотетическая система, соответствующая КФ вида (14), при $\alpha_3^2 = \alpha^2 + \omega_0^2$

представляє собою коливальне звено з постійними
времін

$$T_{1\xi} = 2\alpha/\alpha_3^2, T_{2\xi} = 1/\alpha_3,$$

коефіцієнтом передачі

$$a_\xi = \frac{2\sigma_U}{\alpha_3} \sqrt{\frac{\alpha C}{\pi}}$$

і корнями знаменателя $s_{1,2} = -\alpha_1 \pm j\omega_0$, где $j = \sqrt{-1}$.

Передаточна функція складної системи

$$W_c(s) = \frac{a_\xi B T_3 s}{(T_2^2 s^2 + T_1 s + 1)(T_2^2 \xi s^2 + T_1 \xi s + 1)}$$

має полюси s_1, s_2 і $s_{3,4} = -\alpha \pm j\omega_0$.

Так як $n = m = 2$ складну систему представимо чотирма інерційними звенями з параметрами $\tilde{\gamma}_{1,2} = -s_{1,2}$, $\tilde{\gamma}_{3,4} = -s_{3,4}$ і коефіцієнтами передачі

$$\tilde{a}_1 = \frac{-a_\xi B T_3}{T_2^2 T_2^2 \xi (s_1 - s_2)(s_1 - s_3)(s_1 - s_4)},$$

...

$$\tilde{a}_4 = \frac{-a_\xi B T_3}{T_2^2 T_2^2 \xi (s_4 - s_1)(s_4 - s_2)(s_4 - s_3)}.$$

Дисперсія реакції вираховується по формулам (12) і

(13) при $\tau = 0$.

При заданих числових значеннях отримав стандарт $\sigma_{\Delta P} = 0,495$ кВт.

ВИВОДИ

1. При розв'язанні задач статистичної динаміки в межах кореляційної теорії лінеаризацію складних моделей цілеспрямовано виконувати методом параметричної ідентифікації з використанням перехідної функції, розрахованої по вихідній моделі.

2. Метод "парціальних реакцій" дозволяє отримати аналітичне рішення задачі визначення характеристик реакції лінійної системи на випадкову похибку при суттєвому спрощенні розрахунків.

ПЕРЕЧЕНЬ ССЫЛОК

1. Тихонов В. И. Статистическая радиотехника. - М.: Радио и связь, 1982. - 624 с.
2. Веников В. А. Переходные электромеханические процессы в электрических системах. - М.: Высшая школа, 1985. - 536 с.
3. ГОСТ 13109-87 "Электрическая энергия. Требования к качеству электрической энергии в электрических сетях общего назначения". - Введ. 01.01.89.
4. Flickermeter. Functional and design specification. - Geneva: IEC Report. - Publication 868. - P. 31.
5. Е. Г. Курінний, О. М. Дмитрієва, Абу С. Самі "Уніфікований аналітичний метод розрахунку процесів на виході лінійних систем під час періодичних діянь" // Тези 1-ї Міжнародної науково-технічної конференції "Математичне моделювання в електротехніці і електроенергетиці". - Львів: Львівська політехніка, 1995. - С. 16-17.

Надійшла 13.03.2000
Після доробки 20.03.2000

УДК 681.513

ФОРМУВАННЯ КЕРУЮЧИХ ВПЛИВІВ ЕЛЕКТРОМЕХАНІЧНИХ СИСТЕМ НА ОСНОВІ ПРИНЦИПІВ НЕЧІТКОЇ ЛОГІКИ

А. О. Лозинський

У статті розглянуто принципи формування керуючих впливів на основі правил нечіткої логіки, описані можливі підходи для синтезу відповідних регуляторів та проведено їх порівняння з існуючими методами керування електромеханічними системами.

В статье рассмотрены принципы формирования управляющих воздействий на основе правил нечеткой логики, описаны возможные подходы для синтеза соответствующих регуляторов и проведено их сравнение с существующими методами управления электромеханическими системами.

In a paper the principles of shaping of control based on rules of fuzzy logic are considered, the possible approaches for synthesis of the appropriate controllers are described and their comparison with existing methods of electromechanical systems control is conducted.

ВСТУП

Будь-яку електромеханічну систему в загальному випадку можна описати системою матрично-векторних рівнянь виду:

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A} \cdot \mathbf{x} + \mathbf{B}_k \cdot \mathbf{u}_k + \mathbf{B}_c \cdot \mathbf{u}_c, \\ \mathbf{y} &= \mathbf{C} \cdot \mathbf{x} + \mathbf{D} \cdot \mathbf{u}, \end{aligned} \quad (1)$$

де \mathbf{x} - вектор змінних стану системи, \mathbf{u}_k - вектор керуючих впливів, \mathbf{u}_c - вектор збурюючих впливів, $\mathbf{u} = [\mathbf{u}_k, \mathbf{u}_c]$, \mathbf{y} - вектор вихідних координат, \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{D} - матриці, що визначаються на основі параметрів і

структури електромеханічної системи. В більшості систем автоматичного керування електроприводом постає задача формування відповідної зміни однієї вихідної координати: швидкості або переміщення.

Для отримання бажаного характеру зміни вихідної координати необхідно відповідним чином сформувати керуючий вплив на систему. На сьогоднішній день в керованому електроприводі для вирішення цієї задачі застосовують принцип, який реалізується за допомогою побудови модального регулятора [1]. Структура такої системи зображена на рис. 1. Сигнал керування, що поступає на об'єкт, визначається наступним чином:

$$u_k = u_z - \sum_i k_i \cdot x_i, \quad (2)$$

де u_z - задане значення сигналу керування, k_i - коефіцієнти зворотніх зв'язків за координатами, що описують роботу системи, x_i - координати стану системи. Коефіцієнти зворотніх зв'язків вибирають з умови забезпечення однієї з стандартних форм, наприклад Баттерворта чи біноміальної.

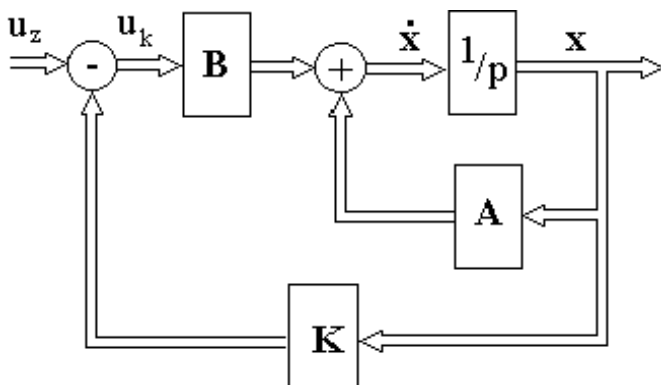


Рисунок 1 - Структура системи модального керування

Проте такий підхід для формування керуючого впливу має ряд недоліків:

- у випадку нелінійної системи розраховані коефіцієнти зворотніх зв'язків забезпечують формування відповідних динамічних характеристик системи тільки для незначної області її роботи, для якої і проведено розрахунок коефіцієнтів;

- при формуванні бажаного характеру зміни вихідної координати не враховується характер зміни проміжних координат, що приводить до їх виходу за допустимі межі. Для забезпечення підтримання проміжних координат в допустимих межах доводиться значно зменшувати швидкість системи.

Для усунення перерахованих недоліків пропонується використовувати в системах автоматичного керування

принципи нечіткої логіки [2].

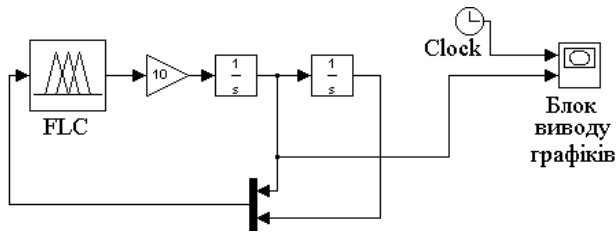
Використання fuzzy-logic регулятора (FLC) типу Такагі-Сугено-Канга дає змогу формувати керуючий вплив для кожної з областей функціонування системи, тобто реалізує адаптивну систему керування. В залежності від значень координат системи активізуються певні лінгвістичні правила регулятора і формуються відповідні сигнали керування. Кожний з сформованих сигналів керування має свій ваговий коефіцієнт, який визначається на основі значень функцій належності координат системи до певної з областей розбиття, отже вибір функцій належності та розбиття системи на області також матиме вплив на формування динамічних характеристик системи. У блоці усунення нечіткості проходить, згідно вибраної методи, формування вихідного сигналу, що поступає на систему.

Вибір коефіцієнтів зворотніх зв'язків, які використовуються при формуванні керуючого впливу окремого правила є досить складною задачею. Найпростішим її вирішенням є знаходження таких значень коефіцієнтів, які б забезпечували одну з відомих форм характеристичного полінома системи [3]. У випадку формування однакового виходу всіх правил, що утворюють FLC, ми повернемося до системи з модальним регулятором. Формування різних виходів окремих правил дозволить нам формувати відповідні динамічні характеристики для різних участків траєкторії. Як приклад, розглянемо систему другого порядку. Використаний в ній FLC в залежності від координат стану системи формуватиме різний сигнал керування. Для області великих відхилень від заданого значення вибрані коефіцієнти зворотніх зв'язків забезпечують налагодження системи згідно стандартної форми Баттерворта при $\omega_0 = 1$, а в області малих відхилень - біноміальну форму при $\omega_0 = 1$. Функції належності мають форму bell-функції. Точка перетину функцій належності області великих і області малих відхилень має координати (7.5, 0.5). Структурна схема такої системи та отримані результати математичного моделювання представлені на рис. 2.

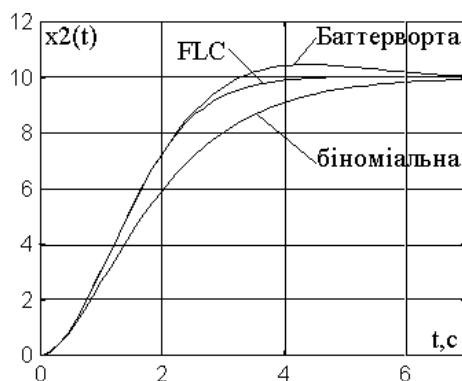
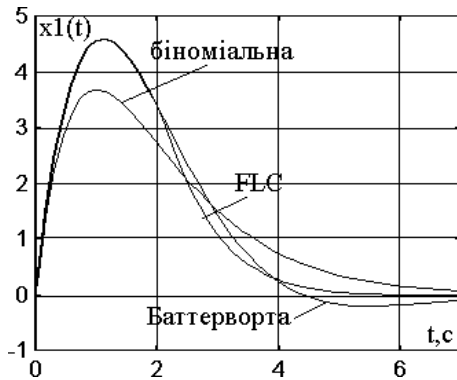
Введення в систему FLC запропонованої структури відповідає формуванню регулятора вихід якого є лінійною комбінацією виходів двох модальних регуляторів:

$$\begin{aligned} u_{FLC} &= \\ &= \frac{\mu_1(u_{\text{задане}} - k_1' \cdot x_1 - k_2' \cdot x_2) + \mu_2(u_{\text{задане}} - k_1'' \cdot x_1 - k_2'' \cdot x_2)}{\mu_1 + \mu_2} = \\ &= u_{\text{задане}} - k_2 \cdot x_2 - k_1, \end{aligned} \quad (3)$$

$$k_2 = k_2' = k_2'', \quad k_1 = \frac{\mu_1 \cdot k_1' + \mu_2 \cdot k_1''}{\mu_1 + \mu_2}.$$



а)



б)

Рисунок 2 - Структурна схема досліджуваної системи а) та її характеристики при виборі стандартних налагоджень та з застосуванням FLC б)

Розподіл полюсів системи, що відповідає зміні коефіцієнта k_1 в межах $[k_1', k_1'']$ представлено на рис. 3. Як видно з отриманого розподілу полюсів системи, введення FLC забезпечує плавний перехід від однієї форми налагодження до іншої.

Застосування запропонованого підходу при заданих параметрах FLC забезпечує зменшення значення функціонала якості системи виду:

$$I = \int_0^T |e(t)| dt, \text{ де } e(t) = x_{\text{задане}} - x_{\text{виходу}} \quad (4)$$

в 1.285 рази у порівнянні з налагодженням системи за біноміальною формою і в 1.035 рази у порівнянні з налагодженням за формою Баттерворта, причому відсутнє перерегулювання вихідної координати.

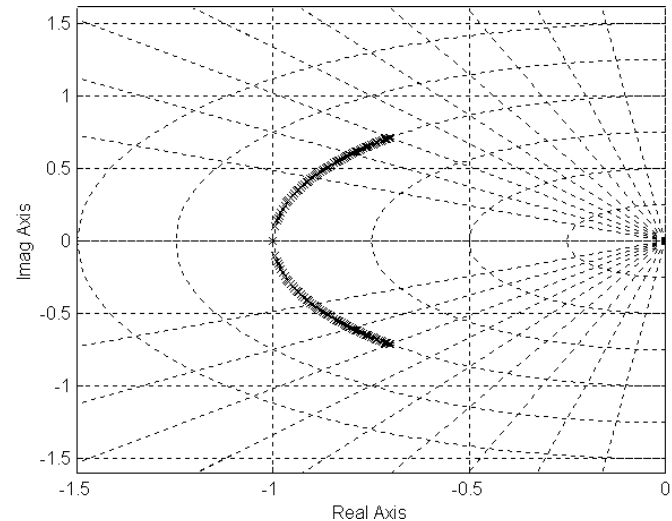


Рисунок 3 - Зміна розташування полюсів системи

Проте використання такого підходу не дозволяє враховувати обмеження, що накладаються на проміжні координати. Тому більш доцільним є застосування для знаходження сигналу керування, що забезпечує формування відповідних динамічних характеристик системи є використання підходів, що відомі з теорії оптимального керування.

У теорії оптимального керування розглядаються методи, які дають змогу сформувати оптимальне керування u , за допомогою якого система переводиться з одного стану в інший так, щоб при цьому мінімізувався функціонал I , дотримуються обмеження на координати і керування, а рівняння динаміки системи і характеристики зовнішніх дій у процесі керування не змінюються [4] з врахуванням (2) задача перетворюється до задачі знаходження відповідних значень коефіцієнтів зворотніх зв'язків за змінними стану системи.

Вимоги, що ставляться до системи в кожній з областей, пов'язують з досягненням екстремуму показника якості роботи системи. Критерієм оптимальності при досягненні максимальної точності роботи системи може бути мінімум функціоналів виду [5]:

$$I = \int_0^T t \cdot |e(t)| dt, \quad I = \int_0^T t \cdot e^2(t) dt, \quad I = \int_0^T e^2(t) dt, \quad (5)$$

де $e(t) = x_{\text{задане}}(t) - x_{\text{виходу}}(t)$.

Обмеження, що накладаються на регульовані координати (змінні стану) та керування, здебільшо подаються у вигляді нерівностей:

$$\begin{aligned} |x_i(t)| &\leq x_{i \max}, \quad i = 1, 2, \dots, n, \\ |u(t)| &\leq u_{\max}, \end{aligned} \quad (6)$$

де $x_{i \max}$ - максимально допустимі значення змінних стану, u_{\max} - максимально допустиме значення сигналу керування.

Накладені обмеження на змінні стану та сигнал керування можуть значно ускладнювати вирішення задачі оптимального керування електромеханічною системою і істотно впливають на її показники якості. В багатьох випадках обмеження є настільки сильними, що знаходження допустимого рішення є не менш трудним ніж знаходження оптимального рішення. Тому більш доцільним при вирішенні поставленої задачі є використання методу функцій покарань. При застосуванні такого методу умови (6) входять у вигляді спеціальних функцій до складу функціоналу якості системи, причому кожне з порушень накладених умов впливатиме на отримане значення функціоналу з заданим коефіцієнтом функції кари. Задача оптимального керування зводиться до знаходження екстремуму функціоналу виду:

$$I = \int_0^T \left(t \cdot e^2(t) + r \cdot \sum_1^n \Phi(h_i(x(t))) \right) dt, \quad (7)$$

де r - коефіцієнт впливу функції покарання, $h_i(x(t)) = |x_i(t)|/x_{i \max}$; $\Phi(h_i(x(t)))$ - функція покарання.

У функціонал якості (7) входить похибка системи за вихідною координатою, для обчислення якої необхідно знати вираз, що описує зміну вихідної координати. Аналітичний розв'язок системи (1) може бути представлений в наступному вигляді [6]:

$$\mathbf{x}(t) = e^{A \cdot t} \cdot \int_{t_0}^t e^{A \cdot \tau} \cdot \mathbf{B} \cdot u(\tau) d\tau + e^{A \cdot (t-t_0)} \cdot \mathbf{x}(t_0), \quad (8.1)$$

де t_0 , t - початкове та біжуче значення часу; або в дискретній формі [7]:

$$\begin{aligned} \mathbf{x}((k+1)h) &= \\ &= e^{A \cdot (k+1) \cdot h} \cdot \int_{kh}^{(k+1) \cdot h} e^{A \cdot \tau} \cdot \mathbf{B} \cdot u(\tau) d\tau + e^{A \cdot h} \cdot \mathbf{x}(kh), \end{aligned} \quad (8.2)$$

де h - приріст часу (крок). Форма запису (8.2) дозволяє знайти значення змінних в дискретні моменти часу

на основі значень змінних з попереднього кроку. Якщо додатково прийняти, що функція керування на даному часовому інтервалі є постійною або змінюється за лінійним законом, то інтеграл можна знайти аналітично і отримати вираз для знаходження значень змінних. Коректність прийнятого допущення впливає з можливості кусково-постійної чи кусково-лінійної апроксимації керуючого впливу, точність якої залежить тільки від величини кроку. У випадку коли $u(t) = const$ вираз (8.2) прийме вигляд:

$$\begin{aligned} \mathbf{x}((k+1)h) &= \\ &= e^{A \cdot h} \cdot \mathbf{x}(kh) + [e^{A \cdot h} - 1] \cdot \mathbf{x}(h) \cdot (A \cdot h)^{-1} \cdot \mathbf{B} \cdot u(\tau). \end{aligned} \quad (9)$$

У вираз (9) входить тільки одна матрична функція $e^{A \cdot h}$, для визначення якої застосуємо апроксимацію Паде порядку (2,2). Вираз для знаходження матричної функції матиме вигляд:

$$e^{A \cdot h} = \mathbf{R}_{22}(A \cdot h) = [\mathbf{Q}_{22}(A \cdot h)]^{-1} \cdot \mathbf{P}_{22}(A \cdot h), \quad (10)$$

де

$$\begin{aligned} \mathbf{P}_{22}(A \cdot h) &= 1 + \frac{1}{2} \cdot (A \cdot h) + \frac{1}{12} \cdot (A \cdot h)^2, \\ \mathbf{Q}_{22}(A \cdot h) &= 1 - \frac{1}{2} \cdot (A \cdot h) + \frac{1}{12} \cdot (A \cdot h)^2. \end{aligned}$$

Проробивши відповідні перетворення ми знайдемо аналітичний вираз для обчислення похибки системи за вихідною координатою в який входять коефіцієнти зворотніх зв'язків за змінними стану системи.

Для вище розглянутої системи другого порядку спрощений вираз (не враховані члени, порядок яких становив приблизно 10^{-6}) для знаходження значень координат системи при $h = 0,01$ матиме вигляд:

$$\begin{aligned} \mathbf{x}((k+1)h) &= \begin{bmatrix} 1 - 0,05 \cdot k_1 & -0,1 \cdot k_2 \\ 1 + 0,05 \cdot k_1 & 1 + 0,05 \cdot k_1 \end{bmatrix} \cdot \mathbf{x}(kh) + \\ &+ \left(\left(\frac{k_2}{1 + 0,05 \cdot k_1} + 10 \right) \cdot \frac{0,001}{k_2} \right) \cdot u(\tau). \end{aligned} \quad (11)$$

Представлений аналітичний вираз (11) дозволяє розраховувати значення координат системи з високою точністю (похибка результатів не перевищує 1.5%).

Сформуємо FLC наступним чином. Діапазон зміни кожної з координат системи розіб'ємо на дві області: область в якій виконуються накладені обмеження і область, в якій координата виходить за накладені обмеження, що відрізняється від традиційних підходів до формування областей належності FLC. В результаті буде виділено чотири області, в яких теоретично можливе

перебування змінних стану системи. Першій області відповідає перебування обох змінних стану в області допустимих значень. Друга і третя області відповідають випадкові, коли порушується одне з обмежень. Четверта область відповідає випадкові одночасного порушення обох обмежень, що в даній системі є неможливим і тому для неї немає необхідності формувати сигнал керування. Для кожної з областей належності змінних стану сформовано відповідний функціонал якості, в якому враховано накладені на координати обмеження ($x_1(t) \leq 4, 25$, $x_2(t) \leq 10$) і знайдено значення коефіцієнтів зворотніх зв'язків, які забезпечують його мінімум (табл. 1).

Таблиця 1

<p>I</p> $I = \int_0^T t \cdot e^2(t) dt$ <p>Значення коефіцієнтів зворотніх зв'язків: $k_1 = -0,123, k_2 = -0,1$</p>	<p>II</p> $I = \int_0^T \left(t \cdot e^2(t) + R_1 \cdot \left(\frac{x_1(t)}{x_{1 \max}} \right)^2 \right) dt$ <p>Значення коефіцієнтів зворотніх зв'язків: $k_1 = -0,162, k_2 = -0,1$</p>
<p>III</p> $I = \int_0^T \left(t \cdot e^2(t) + R_2 \cdot \left(\frac{x_2(t)}{x_{2 \max}} \right)^2 \right) dt$ <p>Значення коефіцієнтів зворотніх зв'язків: $k_1 = -0,2, k_2 = -0,1$</p>	<p>IV</p>

На рис. 4 представлені результати моделювання роботи описаної системи. Вихід координат за допустимі межі в даному випадку не перевищує 5%, при цьому максимальне значення проміжної координати в 1.057 раз, а значення функціоналу якості (3) в 1.012 раз менше ніж при налагодженні згідно стандартної форми Баттерворта. Змінюючи межі областей належності FLC можна добитись ще менших викидів координат за допустимі межі.

Ще одним способом формування динамічних характеристик системи з врахуванням накладених на проміжні координати обмежень є реалізація за допомогою FLC системи зі змінною структурою. Принцип побудови такого регулятора розглянемо на прикладі описаної вище системи другого порядку. Діапазон зміни вихідної координати розбиваємо на дві області: область великих відхилень і область малих відхилень. Реалізовані способи керування для кожної з областей представлені в табл. 2.

На рис. 5 представлені результати моделювання роботи системи зі змінною структурою. Значення проміжної координати при такому підході не перевищує заданого обмеження, а отримане значення функціоналу якості (3) в 1.186 раз менше ніж при налагодженні згідно стандартної форми Баттерворта.

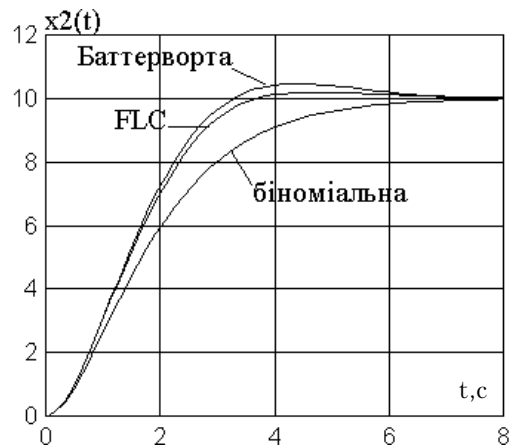
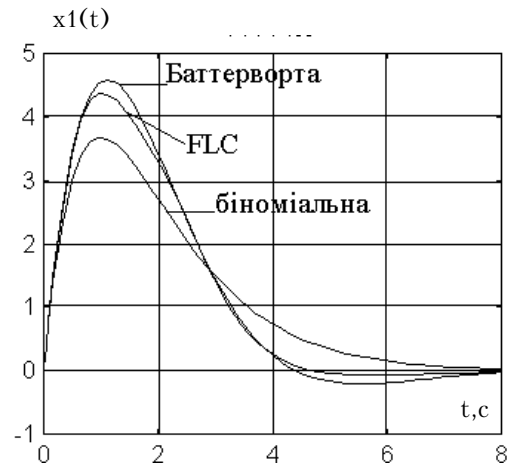


Рисунок 4 - Характеристики досліджуваної системи при виборі стандартних налагоджень та з застосуванням FLC

Таблиця 2

<p>I</p> <p>$k_1 = -0,2381$, $W(p) = 10,5$.</p> <p style="text-align: center;">↓</p> <p>$\frac{u_z}{T_0 p + 1} \cdot \frac{1}{k_1} \cdot \frac{1}{p} \cdot x_1(t) \cdot \frac{1}{p} \cdot x_2(t)$</p> <p>$T_0 = 0,04$</p>	<p>II</p> <p>Реалізоване налагодження системи згідно біноміальної форми: $k_1 = -0,2, k_2 = -0,1$, $W(p) = 1$</p>
--	---

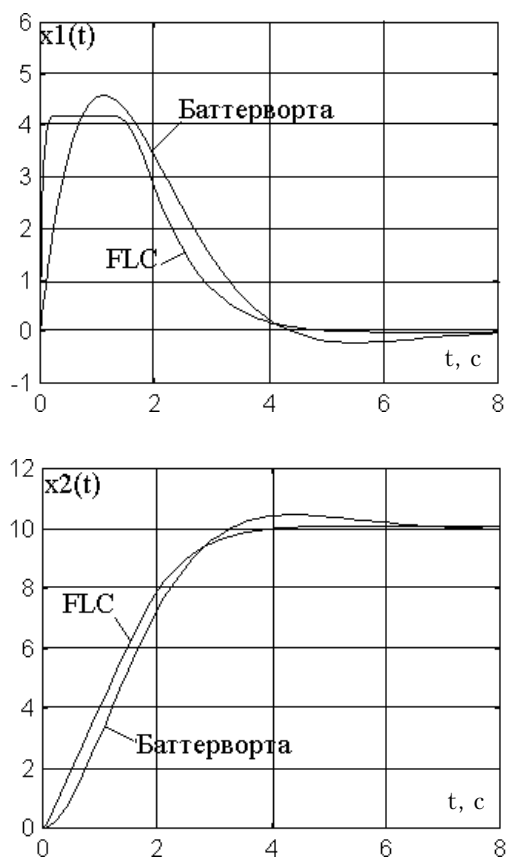


Рисунок 5 - Характеристики досліджуваної системи з FLC у порівнянні з стандартним налагодженням за формою Баттерворта

УДК 681.511.4

РОБАСТНОЕ УПРАВЛЕНИЕ ЭЛЕКТРОПРИВОДОМ С ВЕНТИЛЬНЫМ ДВИГАТЕЛЕМ

Е. М. Потапенко, Д. В. Корельский, Е. В. Васильева

Статья посвящена исследованию проблемы точного управления синхронным двигателем (СД) с учётом неопределённости физических параметров машины и привода в целом. Предлагаемые робастные алгоритмы обеспечивают высокое качество управления СД с постоянными магнитами, что подтверждается анализом цифрового моделирования.

Статья посвящена исследованию проблемы точного управления синхронным двигателем (СД) с учётом неопределённости физических параметров машины и привода в целом. Предлагаемые робастные алгоритмы обеспечивают высокое качество управления СД с постоянными магнитами, что подтверждается анализом цифрового моделирования.

This article is dedicated to the problem of synchronous motor (SM) high accuracy control investigation with the account of machine and drive physical parameter uncertainties. The pro-

posed robust algorithms supply the high quality control of SM with permanent magnets. It is confirmed by the digital simulation analysis.

ВВЕДЕНИЕ

Вентильный двигатель (синхронный двигатель с возбуждением от постоянных магнитов) имеет ряд неоспоримых преимуществ по сравнению с двигателями постоянного тока [1], а во многих случаях, и с асинхронным двигателем (АД) [2]. В связи с этим вентильному электроприводу (ВЭП) в последнее время уделяется повышенное внимание [1-5]. Однако управление ВЭП усложняется наличием ряда неопределённостей в

ВИСНОВКИ

Сучасні системи автоматичного керування побудовані на принципах нечіткої логіки забезпечують реалізацію таких законів керування електроприводами, які дозволяють покращувати їх динамічні характеристики у порівнянні з існуючими, що реалізують принцип модального керування.

Описані підходи формування регуляторів на основі теорії оптимального керування та системи зі змінною структурою забезпечують обмеження проміжних координат системи зберігаючи при цьому високі показники її якості. Застосування різних критеріїв якості на різних ділянках робочої траєкторії забезпечить функціонування системи в оптимальному режимі.

ПЕРЕЛІК ПОСИЛАНЬ

1. Автоматизированные электромеханические системы с модальными регуляторами и наблюдателями состояния. Сборник научн. стат. под. ред. В.Б. Клепикова, Л.В. Акимова, Харьков, 1997 г.
2. Ronald R. Yager, Dimitar P. Filev. Podstawy modelowania i sterowania rozmytego. - Warszawa: WNT, 1995. - 386s.
3. Лозинський А.О. Застосування fuzzy logic регулятора в системах керування за повним вектором стану // Весник ХГПУ "Проблеми автоматизованого електропривода. Теорія і практика". Спец. випуск за матеріалами НТК. - Харків, 1998. - С.388-389.
4. Попович М.Г., Ковальчук О.В. Теорія автоматичного керування. - К: Либідь, 1997. - 544 с.
5. Richard C. Dorf. Modern control systems. - Addison-Wesley Publishing Company, 1990. - 603 p.
6. E. Kreyszig. Advanced engineering mathematics. John Wiley & Sons, 1983. - 1082 p.
7. St. Osowski. Modelowanie ukladow dynamicznych z zastosowaniem jezyka SIMULINK. - Warszawa: OWPW, 1999. - 186 p.

Надійшла 14.03.2000

Після доробки 07.04.2000

его математической модели. Неопределённости обусловлены разбросом сопротивлений статора ($\pm 50\%$), неточностью знания магнитного потока ротора ($\pm 10\%$), насыщением магнитных цепей, неточностью знания момента инерции приводимого в движение механизма (у робота момент инерции может меняться в десятки раз), отсутствием информации о нагрузке и внешних воздействиях, наличием паразитной динамики в виде упругости привода, динамики датчиков и др.

Перечисленные выше проблемы требуют применения робастных методов управления. Среди таковых наибольшее распространение получили алгоритмы с переменной структурой [6,7] и их упрощённый аналог алгоритмы с разрывным управлением (АРУ) [8] (В дальнейшем, для краткости, и те и другие будем называть АРУ).

Характерной особенностью АРУ является наличие в системе управления (СУ) скользящих режимов, которые помимо обеспечения робастности доставляют и недостатки, такие как [1,2,9-12]:

- наличие вибраций, снижающих надёжность механических частей управляемого устройства,
- потеря работоспособности при наличии в структуре управляемого устройства паразитной динамики,
- отсутствие робастности на участке достижения поверхности переключения,
- плохая помехоустойчивость по отношению к высокочастотным помехам,
- АРУ требуют применения в электроприводе мощных полупроводниковых ключей с высокой частотой переключений [13].

Высокая частота переключений напряжения источника приводит к быстрому старению изоляции обмотки двигателя, что снижает надёжность и срок службы электродвигателя.

В работах [1,2,9-12] предлагается ряд мероприятий по устранению отдельных из перечисленных недостатков. Однако ни одно из предлагаемых мероприятий не устраняет сразу все недостатки.

Целью данной статьи является разработка таких алгоритмов управления, которые устраняют все перечисленные недостатки одновременно, обеспечивая при этом робастность системе.

1 ПОСТАНОВКА ЗАДАЧИ

Работа ВЭП во вращающейся системе координат (системе координат, связанной с ротором) описывается уравнениями [1]:

$$\dot{v} = \omega; \quad (1)$$

$$\dot{\omega} = \{(L_d - L_q)i_d + \Phi\}i_q n + M_l + M_d / J; \quad (2)$$

$$\dot{i}_d = (-R_s i_d + nL_q \omega i_q U_q) / L_d; \quad (3)$$

$$\dot{i}_q = (-R_s i_d - nL_q \omega i_d - n\Phi \omega + U_q) / L_q; \quad (4)$$

где

- v, ω - угол поворота и угловая скорость ротора;
- i_d, i_q, U_d, U_q - проекции тока и напряжения статора на оси парковской системы координат с осью, направленной по вектору момента магнита;
- n - число пар полюсов;
- J - приведённый к валу двигателя момент инерции;
- Φ - магнитный поток постоянного магнита;
- M_l, M_d - моменты нагрузки и возмущения, приведённые к валу двигателя;
- R_s, L_d, L_q - сопротивление и составляющие индуктивности статора.

Для работы вентильного двигателя необходим датчик угла поворота ротора относительно статора. С его помощью, осуществляя дифференцирование, можно вычислить угловую скорость. Будет предполагаться, что измеряются $\omega, i_d, i_q, U_d, U_q$ и v (в случае позиционного управления).

Вводится обозначение нижним индексом "o" номинальных значений параметров двигателя, и нижним индексом "δ" отклонений от номинальных значений. Таким образом,

$$J = J_o + J_\delta, R_s = R_{so} + R_{s\delta}, L_q = L_{qo} + L_{q\delta}, \\ L_d = L_{do} + L_{d\delta}, \Phi = \Phi_o + \Phi_\delta. \quad (5)$$

С учётом (5) уравнения (2)-(4) можно представить в виде

$$\dot{\omega} = (\Phi_o i_q n) / J_o + f; \quad (6)$$

$$\dot{i}_d = (-R_{s0} i_d + nL_{q0} \omega i_q + U_d) / L_{d0} + r; \quad (7)$$

$$\dot{i}_q = (-R_{s0} i_q - nL_{d0} \omega i_d - n\Phi \omega + U_q) / L_{q0} + s; \quad (8)$$

где в f, r, s включены все неопределённости. Система (6)-(8) представляет собой точно известную систему, на которую действуют неизвестные возмущения f, r, s .

Суть построения робастной системы управления заключается в оценке с помощью наблюдателя неопределённостей f, r, s и формировании закона управления комбинированного типа, компенсирующего влияние неопределённостей. Система такого типа впервые была предложена и исследована для линейных систем В. Д. Фурасовым [15]. Позднее Е.М. Потапенко был предложен и исследован ряд систем управления комбинированного принципа действия как для линейных, так и для нелинейных объектов управления [16-22]. Как следует из реферативных журналов "Техническая кибернетика" последних нескольких лет, за рубежом с 1998 года

также начались интенсивные исследования робастных комбинированных систем управления. Однако эти исследования пока что ограничиваются линейными системами.

2 АЛГОРИТМЫ РОБАСТНОГО УПРАВЛЕНИЯ

2.1 Управление продольным контуром

Пусть

$$U_d = U_{d1} - U_{d2}, \quad U_{d2} = nL_{q0}\omega i_q. \quad (9)$$

Тогда уравнение (7) принимает вид

$$\dot{i}_d = (-R_{s0}i_d + U_{d1})/L_{d0} + r. \quad (10)$$

Следуя работе [22], будем рассматривать уравнение (10) как измерение. Полагая, что на достаточно малых интервалах времени r меняется мало, т.е. $\dot{r} = 0$, для системы $\dot{r} = 0$ и измерения (10) можно построить наблюдатель

$$\dot{\hat{r}} = K_r(\hat{r} - r) = K_r[-\dot{i}_d - (R_{s0}i_d - U_{d1})/L_{d0}], \quad (11)$$

где K_r - коэффициент закона управления наблюдателя, а \hat{r} - оценка r .

Составляющая тока i_d параллельна оси магнита и не создаёт момента. Однако, в уравнении (2) составляющая электромагнитного момента $(L_{d0} - L_{q0})i_d i_q$ создаёт возмущающий момент. Поэтому надо обеспечить $i_d \equiv 0$. С этой целью полагается

$$U_{d1} = -K_d \dot{i}_d - L_{d0} r. \quad (12)$$

В (11) входит i_d . Для устранения необходимости дифференцирования i_d вводится обозначение

$$\hat{r} + K_r i_d = z_r. \quad (13)$$

Тогда уравнение (11) принимает вид

$$\dot{z}_r = K_r[z_r - K_r i_d - (R_{s0}i_d - U_{d1})/L_{d0}]. \quad (14)$$

Уравнение (14) не содержит \dot{i}_d . Равенство (13) даёт

$$\hat{r} = z_r - K_r i_d. \quad (15)$$

Подстановка (15) в (12), а (12) в (14) позволяет записать

$$\dot{z}_r = -K_r(R_{s0} - K_d)i_d/L_{d0}. \quad (16)$$

Уравнения (16), (15), (12), (9) дают алгоритм формирования управляющего напряжения.

2.2 Управление скоростью ВЭП

Рассмотрим i_q из уравнения (6) как управление и предположим, что имеется оценка \hat{f} неопределённости f . Сформируем закон управления в виде:

$$i_{qr} = (J_0[-K_w(\omega - \omega_r) + \dot{\omega}_r - \hat{f}])/n\Phi_0. \quad (17)$$

Здесь K_w коэффициент закона управления, индекс "r" указывает на программное (желаемое) значение соответствующей переменной. Подстановка i_{qr} из (17) в (6) даёт

$$(\dot{\omega} - \dot{\omega}_r) + K_w(\omega - \omega_r) = f - \hat{f}. \quad (18)$$

При точной оценке \hat{f} уравнение (18) перейдет в уравнение

$$(\dot{\omega} - \dot{\omega}_r) + K_w(\omega - \omega_r) = 0$$

с переходным процессом

$$\omega - \omega_r = [\omega(0) - \omega_r(0)]\exp(-K_w t). \quad (19)$$

В (19) $\omega(0)$ и $\omega_r(0)$ - значения соответствующих переменных в нулевой момент времени. Путём выбора коэффициента $K_w > 0$ можно задавать требуемые переходные процессы, причём $\lim_{t \rightarrow \infty} \omega = \omega_r$.

Построим наблюдатель неопределённости f . Будем рассматривать уравнение

$$f = \dot{\omega} - (n\Phi_0 i_{qr})/J_0 \quad (20)$$

(см. (6)) как измерение f для динамического уравнения $\dot{f} = 0$. Тогда наблюдателю неопределённости f можно придать вид [22]

$$\dot{\hat{f}} = K_f(\hat{f} - f) = K_f[\hat{f} - \dot{\omega} + (n\Phi_0 i_{qr})/J_0], \quad (21)$$

где K_f - коэффициент усиления наблюдателя. Для устранения необходимости дифференцирования ω вводится обозначение

$$\hat{f} + K_f \omega = z_f, \quad \dot{\hat{f}} = z_f - K_f \omega, \quad (22)$$

с учётом которого уравнение (21) принимает вид

$$\dot{z}_f = K_f(\hat{f} + (n\Phi_0 i_{qr})/J_0). \quad (23)$$

Подстановка (17) в (23) даёт

$$\dot{z}_f = K_f[\dot{\omega}_r - K_\omega(\omega - \omega_r)]. \quad (24)$$

Уравнения (24), (22) и (17) формируют программный ток i_{qr} .

В случае позиционного управления i_{qr} должно формироваться по зависимости

$$i_{qr} = J_0[K_v(v - v_r) - K_\omega(\omega - \omega_r) + \dot{\omega}_r - f]/n\Phi_0.$$

Уравнение (22) остаётся в силе, а вместо уравнения (24) должно использоваться уравнение

$$z_f = K_f[\dot{\omega}_r - K_\omega(\omega - \omega_r) - K_v(v - v_r)].$$

При правильном выборе параметров наблюдателя движение ротора будет описываться уравнением

$$(\ddot{v} - \ddot{v}_r) + K_\omega(\dot{v} - \dot{v}_r) - K_v(v - v_r) = 0.$$

С помощью модального управления переходной характеристике можно придать любой желаемый вид.

2.3 Управление поперечным контуром тока

Для уравнения (8) задаётся

$$U_q = U_{q1} + U_{q2}, \quad U_{q2} = n\omega(L_{d0}i_d + \Phi_0). \quad (25)$$

С учётом (25) уравнение (8) принимает вид

$$\dot{i}_q = (-R_{s0}i_q + U_{q1})/L_{q0} + s. \quad (26)$$

Зададим U_{q1} в виде

$$U_{q1} = [-K_q(i_q - i_{qr}) + (R_{s0}/L_{q0})i_{qr} - \hat{s}]L_{q0}, \quad (27)$$

где K_q - коэффициент закона управления, s - оценка неопределённости s , а i_{qr} дано в (17).

Будем рассматривать (26) как измерение неопределённости s , описываемой уравнением $\dot{s} = 0$. Тогда в соответствии с [22] можно получить оценку s с помощью наблюдателя

$$\hat{s} = K_s(\hat{s} - s) = K_s[\hat{s} - \dot{i}_q - (R_{s0}/L_{q0})i_q + U_{q1}/L_{q0}]. \quad (28)$$

Для устранения необходимости получения \dot{i}_q вводится обозначение

$$\hat{s} + K_s i_q = z_s \Rightarrow \dot{\hat{s}} = z_s - K_s i_q, \quad (29)$$

с учётом которого уравнение (28) принимает вид

$$\dot{z}_s = -K_s[s - (R_{s0}/L_{q0})i_q + U_{q1}/L_{q0}]. \quad (30)$$

Подстановка (27) в (30) даёт

$$\dot{z}_s = K_s(R_{s0}/L_{q0} + K_q)(i_q - i_{qr}). \quad (31)$$

Подстановка (27) в (26) даёт

$$\dot{i}_q + (R_{s0}/L_{q0} + K_q)(i_q - i_{qr}) = s - \hat{s}. \quad (32)$$

Поскольку постоянные времени процессов (19), по крайней мере, на порядок больше постоянных времени процессов (32), то можно полагать, что в (32) $\dot{i}_{qr} \equiv 0$.

При точной работе наблюдателя $s \approx \hat{s}$. С учётом сказанного уравнение (32) можно переписать в виде

$$(\dot{i}_q - \dot{i}_{qr}) + (R_{s0}/L_{q0} + K_q)(i_q - i_{qr}) = 0, \quad (33)$$

откуда

$$i_q = i_{qr} + [i_q(0) - i_{qr}(0)] \exp[-(K_q + R_{s0}/L_{q0})t]. \quad (34)$$

Путём выбора коэффициента K_q для i_q можно задать переходный процесс с любым быстродействием, причём при $t \rightarrow \infty$ $\lim(i_q) = i_{qr}$.

3 СТРУКТУРНАЯ СХЕМА ПОЛНОГО АЛГОРИТМА УПРАВЛЕНИЯ

В соответствии с составленными уравнениями, разрабатываются структурные схемы управления синхронным двигателем с постоянными магнитами в системах координат, связанных со статором и ротором. Система, связанная с ротором позволяет легко преобразовывать уравнения машины.

Токи и напряжения во вращающейся системе координат связаны с токами и напряжениями в неподвижной системе следующими уравнениями:

$$u_d = u_\alpha \cos \theta + u_\beta \sin \theta, \quad u_q = u_\alpha \sin \theta - u_\beta \cos \theta,$$

$$i_d = i_\alpha \cos \theta + i_\beta \sin \theta, \quad i_q = i_\alpha \sin \theta - i_\beta \cos \theta,$$

где θ - угол поворота ротора,

$$\theta = \int \omega_p d\tau.$$

Использование вращающейся системы координат позволяет избавиться от тригонометрических функций и упрощает уравнения. Однако, это приводит к необходимости выполнения синусно-косинусных операций в вычислителе.

В качестве аппаратной реализации системы управления может быть использован современный микропроцессорный контроллер [5]. Период вычисления этого алгоритма, например, у контроллера ADSP-2181 фирмы Ана-

log Devices будет составлять $\tau \approx 80 - 100$ мкс.

Схема объекта и системы управления приведена на рис.1. В схеме приняты следующие обозначения:

$$W_1(p) = [K_f K_w + (K_f - K_w)p] / p,$$

$$W_2(p) = [K_s(R_{so} + K_q L_{q0}) - K_q L_{q0} p] / p,$$

$$W_3(p) = [K_r(R_{so} + K_d) + (K_r L_{d0} - K_d)p] / p.$$

В качестве объекта управления выбран двигатель со следующими характеристиками: индуктивности по осям вращающейся системы координат $L_d = 0,0014$ Гн, $L_q = 0,0028$ Гн, сопротивление обмотки $R_0 = 0,6$ Ом, момент инерции ротора $J_0 = 11 \times 10^{-4}$ кгм², поток $\Phi_0 = 0,1194$ Вб, число пар полюсов $n = 4$. Моделирование проводилось для следующих значений коэффициентов:

$K_f = 500$ рад/с², $K_w = 1$ с⁻¹, $K_q = -127$ с⁻¹, $K_s = 3$ с⁻¹, $K_r = -2,2$ с⁻¹, $K_d = 2$ Ом. В качестве инструмента моделирования была использована среда MATLAB-SIMULINK.

На рисунке 2 представлены результаты моделирования: исследованы процессы в приводе при номинальных параметрах и при отклонении параметров от номинальных. На рисунках 2а, 2б, 2в представлены соответственно скорость машины и токи в парковской системе координат: 1 - работа при номинальных параметрах; 2 - отклонение (увеличение) приведённого к валу машины момента инерции в 6 раз, потока возбуждения на 10%, сопротивления обмотки на 50%, индуктивности на 30%; 3 - уменьшение этих параметров.

Для всех случаев также было промоделировано изменение нагрузки на валу.

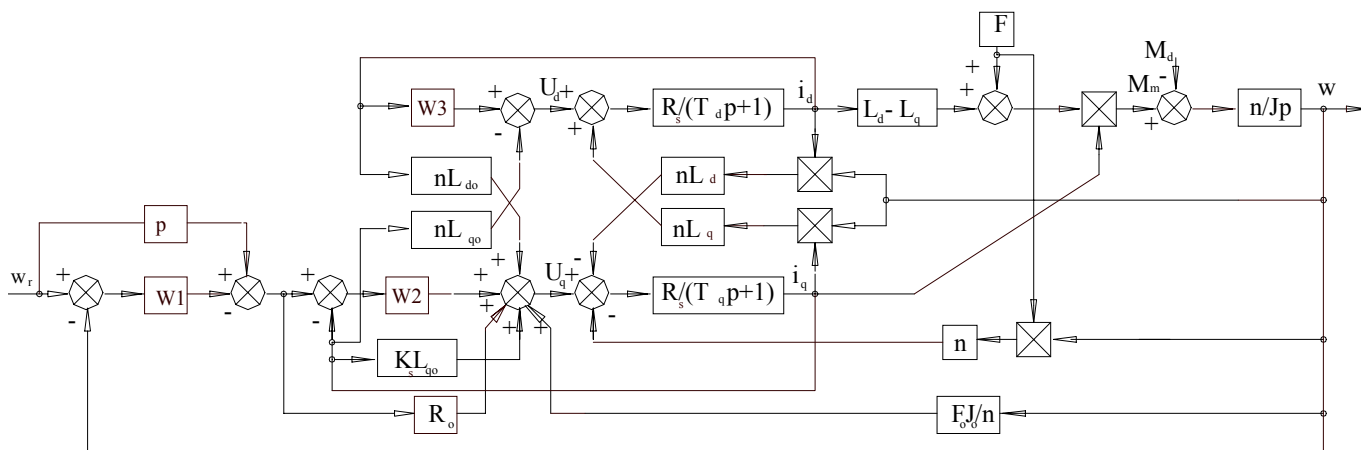


Рисунок 1 - Структурная схема робастного управления синхронным двигателем с постоянными магнитами

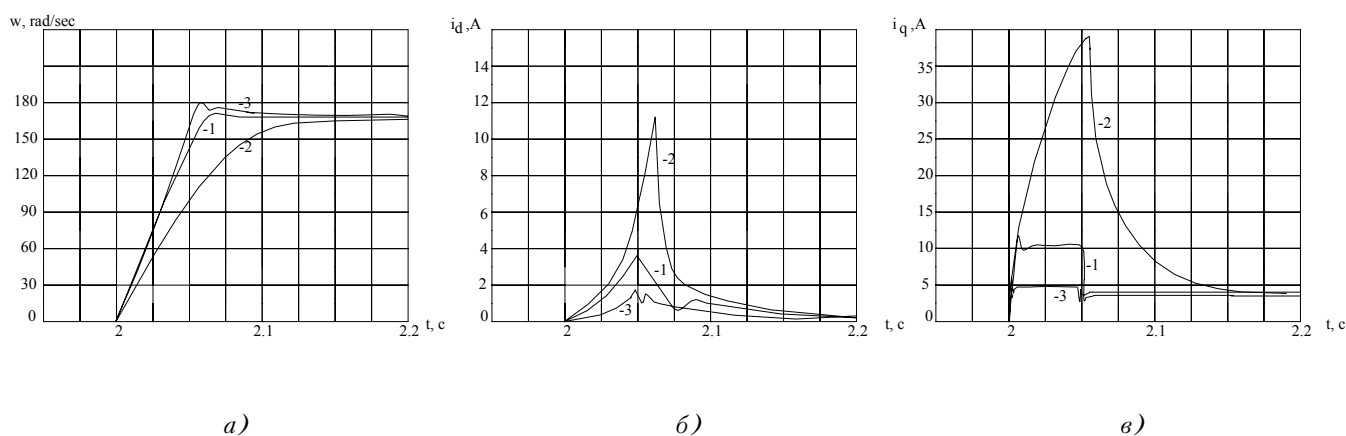


Рисунок 2 - Графики переходных процессов в синхронном двигателе с постоянными магнитами при использовании предложенных алгоритмов управления

ВЫВОДЫ

В статье были разработаны алгоритмы робастного управления синхронным двигателем с постоянными магнитами и на их основе создана математическая модель привода.

Результаты моделирования показывают, что предложенный алгоритм обеспечивает модальную робастность по отношению к неопределёностям параметров двигателя и момента инерции приведённой к валу двигателя системы.

ПЕРЕЧЕНЬ ССЫЛОК

1. *Glumenu A., Hamy M., Lanier C., Moog C.* Robust control of a brushless servo motor via sliding mode techniques. // International journal of control. - 1993.- Vol. 58, №5.- P. 979-990.
2. *Yu-Sheng Lu, Jian-Sshiang Chen.* Design of a global sliding mode controller for a motor drive with bounded control. // International journal of control. - 1995. - Vol. 62, №5. - P.1001-1009.
3. *Козарук А.Е.* Высокоэффективный бесконтактный электропривод с цифровым векторным управлением. // Электротехника. - 1996.- №7.- С. 12-16.
4. *Козаченко В.А.* Основные тенденции развития встроенных систем управления двигателями и требования к микроконтроллерам. // Chip News - 1999, №1 (34).- С.2-9.
5. *Денисов К., Ермилов А., Карпенко Д.* Способы управления машинами переменного тока и их практическая реализация на базе компонентов фирмы Analog devices. // Chip News 1997. - №7 - 8. - С.18-24.
6. *Уткин В.И.* Системы с переменной структурой: состояние, проблемы, перспективы // Автоматика и телемеханика.-1983.- №9.-С. 5-25.
7. *Уткин В.И.* Скользящие режимы в задачах оптимизации и управления. // Москва. - Гл. ред. Физ.-мат. Лит., 1981. - С.368.
8. *Corless M., Leitman G.* Continuous state feedback guaranteed uniform ultimate bounded for uncertain dynamic systems // IEEE Trans. on automatic Control.- 1981. - Vol. 26. - №10. - P. 1139-1144.
9. *Садовой А.В., Сухинин Б. В., Сохина Ю. В.* Системы оптимального управления прецизионными электроприводами./Киев. - ИСИМО, - 1996. - 298с.

10. *Мюдж С.К., Поттон Р. Дж.* Улучшенная оценка робастности регулятора самолёта, работающего в режиме скольжения / /Аэрокосмическая техника. - 1989. - №9. - С.113-122.
11. *Mostafa O., Oz H.* Chatter elimination in variable structure control maneuvering of flexible spacecraft // J. Astronaut. Sciences. - 1989. - Vol. 37, - №4. - P. 529-550.
12. *Hui S., Zak S. H.* Robust control synthesis for uncertain/non-linear dynamical systems // Automatica. - 1992. - Vol.28, - №.2. - P.289-298.
13. *Дацковский Л.Х. и др.* Современное состояние и тенденции в асинхронном частотно-регулируемом приводе (краткий аналитический обзор)// Электротехника. - 1996. - №10. - С.18-28.
14. *Беспалов В.Я., Зверев К.Н.* Импульсные перенапряжения в обмотках асинхронных двигателей при питании от ШИМ - преобразователя // Электротехника. - 1999.- №9.- С.56-59.
15. *Фурасов В. Д.* Устойчивость движения, оценки и стабилизация.// М.- Наука.- 1977.- 248с.
16. *Потапенко Е. М.* Сравнительная оценка робастных систем управления с различными типами наблюдателей // Известия РАН. Теория и системы управления. - 1995. - №1. - С.109-117.
17. *Потапенко Е. М.* Робастные комбинированные системы управления с наблюдателями // Проблемы управления и информатики (Киев). - 1995. - №2. - С.36-44.
18. *Потапенко Е. М.* Исследование робастности систем управления с наблюдателями // Известия РАН. Теория и системы управления. - 1996. - №2.- С.104-108.
19. *Потапенко Е. М.* Синтез и анализ системы управления с переменной структурой // Известия РАН. Теория и системы управления. - 1996. - №3.- С47-50.
20. *Потапенко Е. М.* Робастные системы управления с наблюдателями второго порядка // Автоматика и телемеханика.-1996.- №2.- С.100-108.
21. *Потапенко Е. М.* Синтез и сравнительный анализ робастных компенсаторов пониженного порядка // Автоматика и телемеханика - 1998. - №4. - С.65-74.
22. *Бичай В. Г., Потапенко Е.М.* Об общности альтернативных робастных систем управления // Проблемы управления и информатики. - 1998. - №5. - С.27-30.
23. *Потапенко Е.М., Бичай В.Г.* Синтез и анализ робастной системы управления маневрирующего космического аппарата // Космические исследования. - 1998. - Т.36, - №4. - С. 399-406.
24. *Потапенко Е. М., Бичай В.Г.* Робастное управление маневрирующим упругим космическим аппаратом // Проблемы управления и информатики 1998.- №4.- С. 72-85.
25. *Потапенко Е.М., Савранская А.В.* Синтез и анализ робастной системы управления роботом. // Вісник Київського педагогічного університету.- 1999.- №8.- С. 56.

Надійшла 17.02.2000

Після доробки 21.02.2000

УДК 681.5.015.42

ТЕРМИНАЛЬНОЕ УПРАВЛЕНИЕ СИСТЕМАМИ С ДИФФЕРЕНЦИАЛЬНО-АЛГЕБРАИЧЕСКИМИ ОГРАНИЧЕНИЯМИ МЕТОДОМ КОНЕЧНЫХ СОСТОЯНИЙ

Е. А. Шушляпин

Предложен метод терминального управления системами, представленными совокупностями нелинейных дифференциальных уравнений с гладкими правыми частями и алгебраических уравнений. Метод основан на применении промежуточного описания в виде моделей конечного состояния. Модель конечного состояния строится на основе переменных, отображающих в каждый момент времени прогноз конечного состояния системы в предположении, что от текущего до конечного моментов времени управляющие воздействия отсутствуют. Приведен практический пример из гидромеханики.

Запропоновано метод термінального керування системами,

які представлено сукупностями нелінійних диференціальних рівнянь з гладкими правими частинами та алгебраїчних рівнянь. Метод засновано на застосуванні проміжного опису у вигляді моделей кінцевого стану. Модель кінцевого стану будується на основі змінних, відображаючих у кожний момент часу прогноз кінцевого стану системи у припущенні, що від поточного до кінцевого моментів часу управляючі впливи відсутні. Наведен практичний приклад з гідромеханіки.

The method of terminal control for systems, described by nonlinear differential equations with smooth right parts and algebraic equations, is offered. Method is based on using an intermediate description in the form of terminal state's models.

The terminal state's model is built on the base variables, that reflect a forecast of terminal state at each moment of the time in the suggestion that control action are absent from the current to the terminal moment. The practical example from hydro-mechanics is cited.

ВВЕДЕНИЕ

Задача терминального управления, т.е. приведения всех или части координат системы в заданное положение за заданное время, имеет важное практическое значение. Поскольку в большинстве случаев точное приведение невозможно, часто в качестве цели управления используют минимальное взвешенное расстояние между достигнутой и заданной точками фазового пространства. Это учитывается квадратическим критерием с неотрицательно определенной матрицей весовых коэффициентов. Традиционные подходы к решению данной задачи применительно к нелинейным системам весьма трудоемки как при формировании, так и при компьютерной реализации соответствующих алгоритмов. В статье предлагается метод терминального управления (метод конечных состояний) нелинейными системами, позволяющий при минимальной подготовительной работе, достаточно простым алгоритме и за вполне приемлемое время рассчитать терминальное управление нелинейными системами, описываемыми совокупностями дифференциальных и алгебраических уравнений. Дополнительные ограничения в виде алгебраических уравнений часто имеют место в электротехнических и других моделях. Для сведения таких моделей к чисто дифференциальным используют искусственные приемы, приводящие к увеличению порядка системы дифференциальных уравнений. В предлагаемой статье метод конечного состояния обобщается на системы с дифференциально-алгебраическими ограничениями. Техника применения предложенного метода подробно иллюстрирована на примере практической задачи из гидромеханики.

ПОСТАНОВКА ЗАДАЧИ

Рассмотрим задачу терминального управления, где x^* - желаемое конечное состояние, F - неотрицательно определенная матрица весовых коэффициентов, J^* - некоторое достижимое значение критерия J .

$$\left. \begin{aligned} J &= (x(t_f) - x^*)^T F (x(t_f) - x^*) \rightarrow J^*, \\ \frac{dx}{dt} &= f(t, x, v), \\ \Phi(x, v) &= 0, \\ x, f &= (n \times 1), \quad \Phi = (m \times 1), \quad v = (s \times 1), \quad s \geq m, \\ t &\in [t_0, t_f], \quad x(t_0) = x^0. \end{aligned} \right\} (1)$$

Особенность (1) - смешанная система ограничений в ви-

де дифференциальных и алгебраических уравнений. Алгебраические уравнения при этом отображают неявную связь между вектором переменных состояния x через компоненты вектора управляющих воздействий v . Подобные задачи возникают в электротехнике, гидромеханике и других областях, где используются различные законы сохранения, выраженные в алгебраической форме.

Для определения управления v применим метод конечного состояния, основы которого изложены в [1]. Для применения этого метода требуется, чтобы вектор-функция f была непрерывной вместе со своей частной производной по вектору x , а управляющие воздействия были аддитивными (линейно входящими). Предположим также, что уравнения связи $\Phi(x, v) = 0$ дифференцируемы относительно переменных x и v .

Поскольку между v и x имеется связь, разобьем вектор v на два подвектора - свободный v^c и базисный v^b . Далее свободный подвектор будет представлять независимые управляющие воздействия, а базисный - зависимые. Для v^c введем дополнительные дифференциальные уравнения, определяющие их через аддитивные управляющие воздействия u . Тогда

$$\left. \begin{aligned} J &= (x(t_f) - x^*)^T F (x(t_f) - x^*) \rightarrow J^*, \\ \frac{dx}{dt} &= f(t, x, v^c, v^b), \\ \frac{d}{dt} v^c &= u, \\ \Phi(x, v^c, v^b) &= 0, \\ x, f &= (n \times 1), \quad \Phi = (m \times 1), \quad v = (s \times 1), \\ v^c, u &= (m^c \times 1), \quad v^b = (m \times 1), \quad m^c + m = s, \\ t &\in [t_0, t_f], \quad x(t_0) = x^0. \end{aligned} \right\} (2)$$

Если векторное алгебраическое уравнение связей имеет аналитическое решение

$$v^b = \Phi^{-1}(x, v^c),$$

то (2) сводится к обычному виду, для которого $u(t)$ (или непосредственно $v^c(t)$) может быть найдено любым из известных методов. Однако такой путь может быть применен лишь при достаточно малых m . В общем же случае наличие неизвестной или очень сложной аналитической зависимости Φ^{-1} не позволяет напрямую применять методы, требующие вычисления якобиана от f . Рассматриваемый в данной работе метод конечных состояний как раз использует якобиан, поэтому требуется определенная его модификация.

Схема метода конечных состояний [1] применительно к задаче (2) следующая.

Вначале для (2) записывается промежуточная модель конечного состояния через переменные $x(t_f, t)$, значения которых определяют прогноз неуправляемого конечного состояния системы, находящейся в момент t в состоянии $x(t)$; переменные конечного состояния определяются через матрицу нелинейных весовых функций $W(t_f, t, x(t))$ уравнениями

$$\frac{d}{dt} \bar{x}(t_f, t) = W(t_f, t, x(t))B(t)u(t), \quad (3)$$

$$t \in [t_0, t_f], \quad \bar{x}(t_f, t_0) = \bar{x}^0.$$

Нелинейная весовая матрица определяется как функция первого аргумента уравнениями [2]

$$\left. \begin{aligned} \frac{d}{d\vartheta} W(\vartheta, t, x(t)) &= \left[\frac{\partial \tilde{f}(\vartheta, x)}{\partial x} \right]_{x=\bar{x}(\vartheta, t)} \cdot W(\vartheta, t, x(t)), \\ \vartheta \in [t, t_f], \quad W(t, t, x(t)) &= I, \end{aligned} \right\} (4)$$

где \tilde{f} - вектор правых частей, x - полный вектор состояния системы, I - единичная матрица. Как видно из (4), якобиан $\partial \tilde{f} / \partial x$ при этом вычисляется через значения функций первого аргумента переменных конечного состояния. Последние же определяются через исходную систему с обнуленными управляющими воздействиями. Применительно к (2) совместная система для определения $W(\vartheta, t, x(t))$, $x(\vartheta, t)$ имеет вид

$$\left. \begin{aligned} \frac{d}{d\vartheta} W(\vartheta, t, x(t)) &= \\ &= \left[\begin{array}{cc} \frac{\partial}{\partial x} f(\vartheta, x, v^c, v^6) & \frac{\partial}{\partial v^c} f(\vartheta, x, v^c, v^6) \\ 0_{m^c \times n} & 0_{m^c \times m^c} \end{array} \right]_{\substack{x=\bar{x}(\vartheta, t) \\ v^c=\bar{v}^c(\vartheta, t)}} \cdot W(\vartheta, t, x(t)), \\ \frac{d}{d\vartheta} \bar{x}(\vartheta, t) &= f[\vartheta, \bar{x}(\vartheta, t), \bar{v}^c(\vartheta, t), v^6], \\ v^6 &= \Phi^{-1}(\bar{x}(\vartheta, t), \bar{v}^c(\vartheta, t)), \\ \frac{d}{d\vartheta} \bar{v}^c(\vartheta, t) &= 0, \\ \vartheta \in [t, t_f], \quad W(t_f, t, x(t)) &= I, \quad \bar{x}(t_f, t) = x(t), \quad \bar{v}^c(t_f, t) = v^c(t). \end{aligned} \right\} (5)$$

Система (5) интегрируется многократно для каждого значения t процесса численного интегрирования системы (2). При этом в (3) используются конечные (при $\vartheta = t_f$) значения W .

Матрица $B(t)$ коэффициентов при управляющих воздействиях в (3) применительно к (2) имеет структуру

$$B(t) = \begin{pmatrix} 0_{n \times m^c} \\ I_{m^c \times m^c} \end{pmatrix}. \quad (6)$$

В (5) и (6) $0_{a \times b}$, $I_{a \times b}$ обозначены соответственно нулевая и единичная матрицы размерности $a \times b$.

Для упрощения выкладок исходная модель (1) записана без аддитивно входящих воздействий. Если же таковые на самом деле имеются, нетрудно модифицировать (5), (6) с учетом этого случая.

Следующий шаг применения метода конечных состояний - замена критерия задачи (2) эквивалентным критерием, выраженным через переменные конечного состояния [1].

$$\bar{J}(t_f, t) \Big|_{t=t_f} = (\bar{x}(t_f, t) - x^*)^T F (\bar{x}(t_f, t) - x^*) \Big|_{t=t_f} \rightarrow J^*. \quad (7)$$

Эквивалентность J и $\bar{J}(t_f, t)$ имеет место при $t = t_f$ в силу начальных условий из (5).

Далее вычисляем производную по времени эквивалентного критерия.

$$\begin{aligned} \frac{d}{dt} \bar{J}(t_f, t) &= 2(\bar{x}(t_f, t) - x^*)^T F \frac{d}{dt} \bar{x}(t_f, t) = \\ &= 2(\bar{x}(t_f, t) - x^*)^T F W(t_f, t, x(t)) B(t) u(t). \end{aligned}$$

Последний шаг перед получением выражения для искомого управления - формирование дифференциального уравнения для желаемого поведения $\bar{J}(t_f, t)$. Поскольку конечное значение эквивалентного критерия совпадает со значением исходного критерия, необходимо выбрать такое дифференциальное уравнение, чтобы при любых начальных условиях его решение стремилось к заданному достижимому уровню J^* . Данный момент - наиболее неопределенный в рассматриваемом методе. Во-первых, часто неизвестно J^* (согласно (1) и (7) его теоретическое минимальное значение - нуль, однако оно может быть недостижимо). Во-вторых, неизвестно поведение $\bar{J}(t_f, t)$, приводящее к J^* . Тем не менее, во многих случаях можно задаться уравнением

$$\frac{d}{dt} \bar{J}(t_f, t) = \frac{1}{T} (J^* - \bar{J}(t_f, t)), \quad (8)$$

решение которого, как известно, имеет вид экспоненты с постоянной времени T , приближающееся к J^* при $t \rightarrow \infty$. При этом уже при $t^* > 3T$ $\bar{J}(t_f, t^*) \approx J^*$. Незвестное J^* можно подбирать, начиная от его нулевого значения. Критерием достижимости заданного J^* является совпадение траекторий фактического и желаемого $\bar{J}(t_f, t)$, определяемых соответственно (7) и (8).

Приравнивая два последних выражения и выполняя псевдообращение, получим

$$\left. \begin{aligned} u(t) &= A^T(AA^T)q, \\ A &= (\bar{x}(t_f, t) - x^*)^T F W(t_f, t, x(t)) B(t), \\ q &= \frac{1}{2T} (J^* - (\bar{x}(t_f, t) - x^*)^T F (\bar{x}(t_f, t) - x^*)). \end{aligned} \right\} \quad (9)$$

Как видно из приведенной схемы метода, его можно отнести к группе методов обратной задачи динамики [3]. Его особенность - использование для построения управления эквивалентной модели конечного состояния. Данная схема успешно использовалась для построения терминального управления в ряде прикладных задач [4], [5], [6]. Специфика модели (2), требующая модификации приведенной схемы, заключается в трудностях непосредственного вычисления частной производной $\partial f / \partial v^c$, входящей в (5), из-за наличия связи между v^b и v^c .

РЕШЕНИЕ ЗАДАЧИ

Вычисляя частную производную вектора правых частей дифференциальных уравнений (1) по вектору независимых управлений с учетом "двойной" зависимости f от v^c - непосредственно и неявно через v^b , получим

$$\begin{aligned} \frac{\partial}{\partial v^c} f(v, x, v^c, v^b) &= \frac{\partial}{\partial v^c} f(v, x, v^c, v^b) + \\ &+ \frac{\partial}{\partial v^b} f(v, x, v^c, v^b) \cdot \frac{\partial}{\partial v^c} v^b. \end{aligned} \quad (10)$$

Для устранения необходимости непосредственного дифференцирования обратной зависимости $v^b = \Phi^{-1}$ найдем частную производную уравнения связи по v^c .

$$\begin{aligned} \frac{\partial}{\partial v^c} \Phi(x, v^c, v^b) &= \frac{\partial}{\partial v^c} \Phi(x, v^c, v^b) + \\ &+ \frac{\partial}{\partial v^b} \Phi(x, v^c, v^b) \cdot \frac{\partial}{\partial v^c} v^b = 0. \end{aligned}$$

Предполагая невырожденность якобиана связей $\partial \Phi / \partial v^b$, что должно иметь место при корректно заданных связях, получаем

$$\frac{\partial}{\partial v^c} v^b = - \left[\frac{\partial}{\partial v^b} \Phi(x, v^c, v^b) \right]^{-1} \cdot \frac{\partial}{\partial v^c} \Phi(x, v^c, v^b). \quad (11)$$

Подставляя (11) в (10), получаем искомую $\partial f / \partial v^c$.

$$\begin{aligned} \frac{\partial}{\partial v^c} f(v, x, v^c, v^b) &= \frac{\partial}{\partial v^c} f(v, x, v^c, v^b) - \\ &- \frac{\partial}{\partial v^b} f(v, x, v^c, v^b) \cdot \left[\frac{\partial}{\partial v^b} \Phi(x, v^c, v^b) \right]^{-1} \cdot \frac{\partial}{\partial v^c} \Phi(x, v^c, v^b). \end{aligned} \quad (12)$$

Входящие в (12) и в (5) частные производные правых частей дифференциальных уравнений и уравнений связи существуют в силу указанных выше предположения о

свойствах функций f и Φ .

Пример. Рассмотрим пример из гидромеханики.

Имеется трубопроводная сеть, граф которой изображен на рисунке 1.

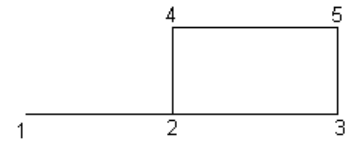


Рисунок 1

Предполагается, что геометрические уровни вершин 1, 2, 3 $z_1 = z_2 = z_3 = 0$, уровни вершин 4, 5 $z_4 = z_5 = 5$ м. В вершинах 1 и 3 имеются напорные насосы. Трубы 2-4 и 3-5 - цилиндрические емкости высотой 5 м, - предназначены для длительного хранения питьевой воды. Емкость 2-4 имеет диаметр 3 м, емкость 3-5 - диаметр 5 м. Остальные трубы диаметром 1 м и длиной 10 м каждая предназначены для подвода воды (труба 1-2) и соединения емкостей (трубы 2-3 и 4-5). В вершине 5 предполагается свободный выход воды через трубу, которая в рассматриваемую систему не входит. Целью управления является прокачка воды через емкости с помощью насосов таким образом, чтобы вода в каждой емкости обновлялась каждые два часа.

Модель данной системы без учета процессов перемешивания строится на основе уравнений Бернулли [7] для неустановившегося напорного движения несжимаемой жидкости в предположении абсолютной жесткости стенок труб. Система уравнений Бернулли для рассматриваемой задачи имеет вид

$$h_i + z_i = h_j + z_j + \mu_{ij} \frac{U_{ij}^2}{2g} + \alpha_0 \frac{L_{ij}}{g} \cdot \frac{dU_{ij}}{dt},$$

$$(i, j) \in \{(1, 2), (2, 4), (4, 5)\},$$

$$h_2 + z_2 = h_3 + z_3 - \Delta h_3 + \mu_{23} \frac{U_{23}^2}{2g} + \alpha_0 \frac{L_{23}}{g} \cdot \frac{dU_{23}}{dt}, \quad (13)$$

$$h_3 + z_3 + \Delta h_3 = h_5 + z_5 + \mu_{35} \frac{U_{35}^2}{2g} + \alpha_0 \frac{L_{35}}{g} \cdot \frac{dU_{35}}{dt}.$$

Здесь h_i, h_j - пьезометрические напоры в i -й и j -й вершинах сети, U_{ij} - средняя скорость потока в трубе $i-j$, L_{ij} - длина трубы $i-j$, $g = 9,807 \text{ м} \cdot \text{с}^{-2}$ - ускорение свободного падения, $\alpha_0 \in [1,03; 1,05]$ - коэффициент коррекции количества движения. Параметр μ_{ij} определяется выражением [8]

$$\mu_{ij} = \alpha + \lambda_{ij} \frac{L_{ij}}{D_{ij}} + \zeta_{ij}, \quad \lambda_{ij} = \left(2,035 \cdot \lg \frac{D_{ij}}{2k_{ij}} - 1,74 \right)^{-2}, \quad (14)$$

где $\alpha \in [1, 05; 2, 00]$ - коэффициент коррекции скорости, учитывающий степень турбулентности потока (большие значения соответствуют меньшей турбулентности), D_{ij} , ζ_{ij} - диаметр и коэффициент местного сопротивления i - j -й трубы. Коэффициент сопротивления λ_{ij} (коэффициент Дарси) рассчитывается по приведенной выше формуле Кармана-Никурадзе для шероховатых труб, в которой k_{ij} - средняя высота бугорков шероховатости. Член Δh_3 отражает наличие насоса в месте соединения трубы (2,3) и емкости (3,5). Так как насос, нагнетая воду в емкость, одновременно отсасывает ее из трубы (2,3) и наоборот, указанный член входит в уравнения для (2,3) и (3,5) с разными знаками. Пьезометрический напор h_5 при свободном выходе потока из вершины 5 нулевой.

Наряду с (13), (14), в динамическую модель входят уравнения сохранения потока в виде системы уравнений для расходов во внутренних вершинах сети.

$$Q_{12} = Q_{23} + Q_{24}, \quad Q_{24} = Q_{45}, \quad Q_{23} = Q_{35}. \quad (15)$$

Для вершины 5 уравнение сохранения не записываем, так как из нее имеется свободный выход потока.

Расходы Q_{ij} связаны со средними скоростями U_{ij} соотношением

$$Q_{ij} = \sigma_{ij} \cdot U_{ij}, \quad \sigma_{ij} = \pi D_{ij}^2 / 4,$$

где σ_{ij} - площадь поперечного сечения i - j -й трубы.

Последняя группа уравнений модели - уравнения для уровней воды в емкостях

$$\frac{dS_{24}}{dt} = U_{24}, \quad \frac{dS_{35}}{dt} = U_{35}.$$

Указанные уровни отображают не уровни заполнения емкостей (они на протяжении всего времени работы системы предполагаются заполненными), а местоположения выделенных частиц воды, по которым можно судить о степени обновления воды в емкостях.

Целью прокачки является достижение уровнями S_{24} и S_{35} значения $S_z = 5$ м за заданное время 7200с.

Соответствующий критерий терминальной задачи в этом случае имеет вид:

$$J = [S_{24}(t_f) - S_z]^2 + [S_{35}(t_f) - S_z]^2 \rightarrow J^*, \quad (16)$$

где минимально возможное значение J^* нулевое.

Переходя в (14), (15) от скоростей к уровням, перепишем модель в следующем эквивалентном виде:

$$\left. \begin{aligned} \frac{dQ_{ij}}{dt} &= -\beta_{ij} Q_{ij}^2 + \gamma_{ij} (h_i + z_i - h_j - z_j), \\ (i, j) &\in \{(1, 2), (2, 4), (4, 5)\}, \\ \frac{dQ_{23}}{dt} &= -\beta_{23} Q_{23}^2 + \gamma_{23} (h_2 + z_2 - h_3 - z_3 + \Delta h_3), \\ \frac{dQ_{35}}{dt} &= -\beta_{35} Q_{35}^2 + \gamma_{35} (h_3 + z_3 - h_5 - z_5 + \Delta h_3), \\ \frac{dS_{24}}{dt} &= \frac{Q_{24}}{\sigma_{24}}, \quad \frac{dS_{35}}{dt} = \frac{Q_{35}}{\sigma_{35}}, \quad \frac{dh_1}{dt} = u_1(t), \quad \frac{d\Delta h_3}{dt} = u_2(t), \\ Q_{12} &= Q_{23} + Q_{24}, \quad Q_{24} = Q_{45}, \quad Q_{23} = Q_{35}, \\ \beta_{ij} &= \frac{\mu_{ij}}{2\alpha_0 L_{ij} \sigma_{ij}}, \quad \gamma_{ij} = \frac{\sigma_{ij} \cdot g}{\alpha_0 L_{ij}}, \\ (i, j) &\in P = \{(1, 2), (2, 3), (2, 4), (3, 5), (4, 5)\}. \end{aligned} \right\} \quad (17)$$

Управляемые посредством $u_1(t)$, $u_2(t)$ напоры h_1 , $\Delta h_3 \in h^c$ здесь предполагаются независимыми, от которых через уравнения расходов зависят другие напоры. Для выявления указанных зависимостей продифференцируем уравнения расходов из (17) и подставим в них соответствующие выражения для производных от расходов. В результате получим следующую систему линейных уравнений относительно зависимых напоров $h^b = colon(h_2, h_3, h_4)$:

$$A(\gamma) \cdot h^b = b(Q, \gamma, h^c),$$

где

$$A(\gamma) = \begin{pmatrix} -\gamma_{12} - \gamma_{23} - \gamma_{24} & \gamma_{23} & \gamma_{24} \\ \gamma_{24} & 0 & -\gamma_{24} - \gamma_{45} \\ \gamma_{23} & -\gamma_{23} - \gamma_{35} & 0 \end{pmatrix},$$

$$b(Q, \gamma, h^c) = \begin{pmatrix} -\beta_{12} Q_{12}^2 - \beta_{23} Q_{23}^2 - \beta_{24} Q_{24}^2 - \gamma_{12} h_1 - \gamma_{23} \Delta h_3 - \\ -\gamma_{12} (z_1 - z_2) + \gamma_{23} (z_2 - z_3) + \gamma_{24} (z_2 - z_4) \\ \beta_{24} Q_{24}^2 - \beta_{45} Q_{45}^2 - \gamma_{24} (z_2 - z_4) + \gamma_{45} (z_4 - z_{45}) - \\ -\gamma_{45} h_5 \\ \beta_{23} Q_{23}^2 - \beta_{35} Q_{35}^2 + (-\gamma_{23} + \gamma_{35}) \Delta h_3 - \\ -\gamma_{23} (z_2 - z_3) + \gamma_{35} (z_3 - z_5) - \gamma_{35} h_5 \end{pmatrix}.$$

Система уравнений (5) в данном случае имеет вид

$$\frac{d}{d\vartheta} W(\vartheta, t, x(t)) =$$

$$\begin{bmatrix} \frac{\partial}{\partial Q} f_Q(\vartheta, x, h^c, h^b) & 0_{5 \times 2} & \frac{\partial}{\partial h^c} f_Q(\vartheta, x, h^c, h^b) \\ \frac{\partial}{\partial Q} f_s(\vartheta, x, h^c, h^b) & 0_{2 \times 2} & 0_{2 \times 2} \\ 0_{2 \times 5} & 0_{2 \times 2} & 0_{2 \times 2} \end{bmatrix} \times \begin{matrix} x = \bar{x}(\vartheta, t) \\ h^c = \bar{h}^c(\vartheta, t) \end{matrix}$$

$$x = colon(Q, S, h^c),$$

$$\frac{d}{d\vartheta} \bar{Q}_{ij}(\vartheta, t) = -\beta_{ij} \bar{Q}_{ij}^2 + \gamma_{ij} (\bar{h}_i + z_i - \bar{h}_j - z_j),$$

$$(i, j) \in \{(1, 2), (2, 4), (4, 5)\},$$

$$\frac{d}{d\vartheta} \bar{Q}_{23}(\vartheta, t) = -\beta_{23} \bar{Q}_{23}^2 + \gamma_{23} (\bar{h}_2 + z_2 - \bar{h}_3 - z_3 + \Delta h_3),$$

$$\frac{d}{d\vartheta} \bar{Q}_{35}(\vartheta, t) = -\beta_{35} \bar{Q}_{35}^2 + \gamma_{35} (\bar{h}_3 + z_3 - \bar{h}_5 - z_5 + \Delta h_3),$$

$$\bar{h}_5 = h_5,$$

$$\frac{d\bar{S}_{24}}{dt} = \frac{\bar{Q}_{24}}{\sigma_{24}}, \quad \frac{d\bar{S}_{35}}{dt} = \frac{\bar{Q}_{35}}{\sigma_{35}}, \quad \frac{d\bar{h}_1}{dt} = 0, \quad \frac{d\Delta h_3}{dt} = 0,$$

$$\bar{h}^6 = A(\gamma)^{-1} b(Q, g, \bar{h}^c),$$

$$\frac{d}{dQ} f_Q(\vartheta, x, h^c, h) = \text{diag}\{-2\beta_{ij} Q_{ij}\}, \quad (i, j) \in P,$$

$$\frac{d}{dQ} f_s(\vartheta, x, h^c, h) = \text{diag}\{1/\sigma_{ij}\}, \quad (i, j) \in \{(2, 4), (3, 5)\},$$

$$\frac{d}{dh^c} f_Q(\vartheta, x, h^c, h) = \frac{df_Q}{dh^c} + \frac{df_Q}{dh^6} \cdot A^{-1}(\gamma) \cdot \frac{d}{dh^c} b(Q, \gamma, \bar{h}^c),$$

$$\frac{df_Q}{dh^c} = \begin{pmatrix} \gamma_{12} & 0 \\ 0 & \gamma_{23} \\ 0 & 0 \\ 0 & \gamma_{35} \\ 0 & 0 \end{pmatrix}, \quad \frac{df_Q}{dh^6} = \begin{pmatrix} -\gamma_{12} & 0 & 0 \\ \gamma_{23} & -\gamma_{23} & 0 \\ \gamma_{24} & 0 & -\gamma_{24} \\ 0 & \gamma_{35} & 0 \\ 0 & 0 & \gamma_{45} \end{pmatrix},$$

$$\frac{d}{dh^c} b(Q, \gamma, \bar{h}^c) = \begin{pmatrix} -\gamma_{12} & -\gamma_{23} \\ 0 & 0 \\ 0 & \gamma_{35} - \gamma_{23} \end{pmatrix},$$

$$\vartheta \in [t, t_f], \quad W(t, t, x(t)) = I, \quad \bar{Q}_{ij}(t, t) = Q_{ij}(t), \quad (i, j) \in P,$$

$$\bar{S}_{ij}(t, t) = S_{ij}(t), \quad (i, j) \in \{(2, 4), (3, 5)\},$$

$$\bar{h}_1(t, t) = h_1(t), \quad \Delta \bar{h}_3(t, t) = \Delta h_3(t).$$

Управление (9) конкретизируется в виде:

$$\left. \begin{aligned} u(t) &= A^T(AA^T)q, \quad A = \begin{pmatrix} A_1 & A_2 \end{pmatrix}, \\ A_1 &= (\bar{S}_{24}(t_f, t) - Sz) \cdot W_{68}(t_f, t, x(t)) + \\ &\quad + (\bar{S}_{35}(t_f, t) - Sz) \cdot W_{78}(t_f, t, x(t)), \\ A_2 &= (\bar{S}_{24}(t_f, t) - Sz) \cdot W_{69}(t_f, t, x(t)) + \\ &\quad + (\bar{S}_{35}(t_f, t) - Sz) \cdot W_{79}(t_f, t, x(t)), \\ q &= \frac{1}{2T} (J^* - (\bar{S}_{24}(t_f, t) - Sz)^2 - (\bar{S}_{35}(t_f, t) - Sz)^2), \\ u_1(t) &= \frac{A_1 \cdot q}{A_1^2 + A_2^2}, \quad u_2(t) = \frac{A_2 \cdot q}{A_1^2 + A_2^2}. \end{aligned} \right\} (18)$$

Предлагаемый алгоритм исследовался с помощью компьютерного моделирования для следующих параметров: $T = 400$ с, $k_{ij} = 0,0001$ м, $\alpha_0 = 1,04$, $\alpha = 2$,

$J^* = 0$. Шаг численного интегрирования дифференциальных уравнений методом Рунге-Кутты четвертого порядка полагался равным 50с, конечное время $t_f = 7200$ с. Начальные условия выбирались из предположения, что в начальном состоянии вода в трубах и емкостях неподвижна. Этому соответствуют нулевые начальные значения для всех расходов, уровней и управляющего напора Δh_3 . Начальное же значение управляющего напора $h_1 = 5$ м, так как этот напор противостоит напору водяных столбов в емкостях высотой 5м. Результаты моделирования представлены на рисунке 2, где изображены зависимости изменения уровней, управляющих напоров, а также желаемая и фактическая зависимости $\bar{J}(t_f, t)$. Последние изображены на третьем графике в виде практически совпадающих кривых.

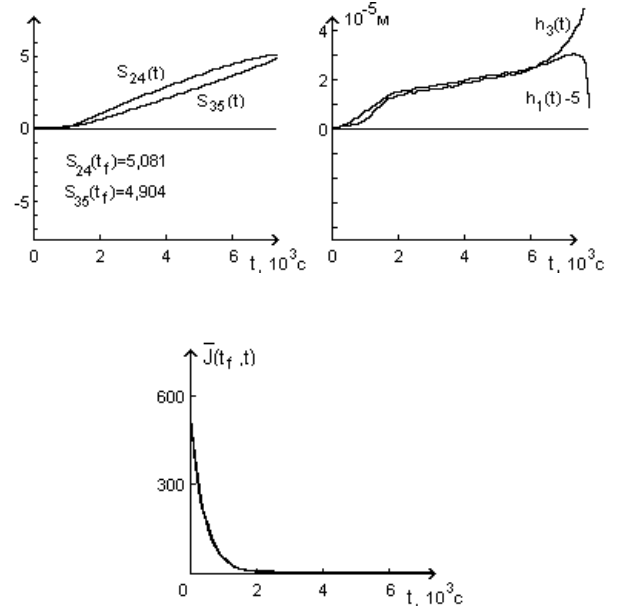


Рисунок 2

Желаемое поведение $\bar{J}(t_f, t)$, согласно (8), определяется выражением

$$\bar{J}(t_f, t) = (\bar{J}(t_f, t_0) - J^*) \cdot \exp\left[-\frac{t-t_0}{T}\right].$$

Фактическое поведение определяется, согласно (7), (16) и (18), выражением

$$\bar{J}(t_f, t) = J^* - 2 \cdot T \cdot q.$$

Совпадение указанных желаемой и фактической зависимостей свидетельствует о достижимости в данном

классе управлений выбранных значений Sz, J^* .

Для сравнения предпринимались попытки подобрать постоянные значения управляющих напоров, обеспечивающих прокачку обеих емкостей за заданное время. Наилучший результат, достигнутый при этом, получен при $h_1 = 0,00004$, $\Delta h_3 = 0$ и равен $S_{24}(t_f) = 14,047$; $S_{24}(t_f) = 5,070$. Уровень 14,047 означает, что за заданное время емкость (2,4) прокачивается почти 3 раза, в то время как другая емкость (3,5) - только один раз. Кроме того, напор первого насоса (т.е. его энергетические затраты) при этом почти на порядок больше суммарных энергетических затрат двух насосов при предлагаемом алгоритме управления. Время расчета управления на компьютере 486DX66 - около четырех минут.

Следует отметить, что алгоритм (17) малочувствителен к погрешностям его параметров. Так, искажение параметров β_{ij}, γ_{ij} в алгоритме расчета управления (18) на одну треть по отношению к тем же параметрам в (17) приводит к изменению конечных значений управляемых уровней примерно на десять процентов.

ВЫВОД

Приведенные в данной статье, а также в работах [3],[4],[5] результаты свидетельствуют о практической ценности метода конечных состояний для управления

многомерными нелинейными системами достаточно широкого класса. Метод может применяться и в случае ограничений на величины управляющих сигналов [4]. Расчет управлений при этом производится по (9) и корректируется до максимальных или минимальных допустимых значений при выходе на ограничение. Время управления в этом случае может увеличиться, однако работоспособность алгоритма сохранится.

ПЕРЕЧЕНЬ ССЫЛОК

1. Шушляпин Е.А. Синтез линейных и нелинейных систем управления конечным положением на основе моделей конечного состояния // Проблемы управления и информатики - 1997. - №3. - С.10-16.
2. Алексеев В.М. Об одной оценке возмущений обыкновенных дифференциальных уравнений // Вестн. Московск. ун-та.- Сер.1. Математика, механика - 1961.-№2.-С.28-36.
3. Батенко А.П. Системы терминального управления. - М.: Радио и связь, 1984.
4. Шушляпин Е.А., Канов Л.Н. Синтез управления переходными процессами в нелинейных системах электропривода// Радиоэлектроника, информатика, управление. - 1999. - №1. - С.136-139.
5. Канов Л.Н., Шушляпин Е.А. Терминальное управление процессами коммутации двухобмоточных электромагнитов постоянного тока // Оптимизация производственных процессов: сб. науч. трудов. - Севастоп. гос. техн. ун-т. - 1999. - Вып.1. - С.102-107.
6. Канов Л.Н., Шушляпин Е.А. Терминальное управление динамикой синхронного электродвигателя // Оптимизация производственных процессов: сб. науч. трудов. - Севастоп. гос. техн. ун-т. - 1999. - Вып.2. - С.96-102.
7. Чугаев Р.Р. Гидравлика. - Л.: Энергия, 1975. - 600с.: ил.
8. Павленко В.Г. Основы механики жидкости. - Л.: Судостроение, 1988. - 240с.:ил.

Надійшла 29.02.2000
Після доробки 21.03.2000

АВТОРЫ НОМЕРА

D. C. Wunch II,

Distinguished Professor of Computer Engineering, University of Missouri-Rolla

Архипов Александр Евгеньевич,

доктор технических наук, профессор Национального технического университета Украины "Киевский политехнический институт"

Архипова София Анатольевна,

преподаватель Национального технического университета Украины "Киевский политехнический институт"

Базылев Богдан Юрьевич,

аспирант Запорожской государственной инженерной академии

Беликов Сергей Борисович,

доктор технических наук, ректор Запорожского государственного технического университета

Бодянский Евгений Владимирович,

доктор технических наук, профессор Харьковского государственного технического университета радиоэлектроники

Бутвин Борис Леонидович,

кандидат технических наук, ведущий научный сотрудник Центрального научно-исследовательского института Вооруженных сил Украины

Васильева Евгения Владимировна,

ассистент Запорожского государственного технического университета

Вершина Александр Иванович,

кандидат технических наук, доцент Запорожского государственного технического университета

Гапоненко Николай Прокофьевич,

кандидат технических наук, доцент Запорожского государственного технического университета

Горбань Александр Николаевич,

доктор физико-математических наук, заме-

ститель директора по науке Института вычислительного моделирования Сибирского отделения Российской академии наук, заведующий кафедрой Красноярского государственного технического университета

Горбань Александр Николаевич,

доктор физико-математических наук, проректор Запорожского института государственного и муниципального управления

Горбачев Модест Николаевич,

кандидат технических наук, старший научный сотрудник Института общей энергетики НАН Украины (г.Киев)

Гостев Владимир Иванович,

доктор технических наук, начальник кафедры Киевского военного института управления и связи

Деев Сергей Георгиевич,

ассистент Запорожского государственного технического университета

Диянкова Елена Владимировна,

сотрудник Всероссийского НИИ технической физики Российского федерального ядерного центра

Диянкова Светлана Анатольевна,

сотрудник Всероссийского НИИ технической физики Российского федерального ядерного центра

Долгов Алексей Юрьевич,

старший преподаватель Приднестровского государственного университета, г. Тирасполь

Дорогов Александр Юрьевич,

кандидат технических наук, докторант Санкт-Петербургского государственного электротехнического университета

Дорошенко Юлия Николаевна,

аспирантка Запорожского государственного технического университета

-
- Дубровин Валерий Иванович,*
кандидат технических наук, докторант
Запорожского государственного техниче-
ского университета
- Зиновьев Андрей Юрьевич,*
аспирант Института вычислительного моде-
лирования Сибирского отделения Россий-
ской академии наук
- Карпуков Леонид Матвеевич,*
кандидат технических наук, докторант За-
порожского государственного технического
университета
- Касьян Константин Николаевич,*
кандидат технических наук, доцент Запо-
рожского государственного технического
университета
- Касьян Николай Николаевич,*
старший преподаватель Запорожского госу-
дарственного технического университета
- Квичанский Андрей Валерьевич,*
сотрудник Всероссийского НИИ техниче-
ской физики Российского федерального
ядерного центра
- Коноваленко Наталья Алексеевна,*
студентка Запорожского государственного
технического университета
- Корельский Дмитрий Викторович,*
аспирант Запорожского государственного
технического университета
- Кофанов Юрий Николаевич,*
доктор технических наук, профессор Мос-
ковского государственного института элект-
роники и математики (технического уни-
верситета)
- Кравчина Виталий Викторович,*
начальник лаборатории ОКБ "Элмис",
г.Запорожье
- Крищук Владимир Николаевич,*
кандидат технических наук, заведующий
кафедрой Запорожского государственного
технического университета
- Кузьмина Людмила Владимировна,*
ассистент Запорожского государственного
технического университета
- Куренный Эдуард Григорьевич,*
доктор технических наук, профессор До-
нецкого государственного технического
университета
- Левыкина Виктория Ивановна,*
ассистент Запорожского государственного
технического университета
- Литвинов Евгений Николаевич,*
кандидат технических наук, доцент Нацио-
нального технического университета Украи-
ны "Киевский политехнический институт"
- Лихоузов Алексей Сергеевич,*
аспирант Национального технического уни-
верситета Украины "Киевский политехниче-
ский институт"
- Лихоузова Татьяна Анатольевна,*
аспирантка Национального технического
университета Украины "Киевский политех-
нический институт"
- Лозинский Андрей Орестович,*
кандидат технических наук, доцент Госу-
дарственного университета "Львівська полі-
техніка"
- Лопин Вячеслав Николаевич,*
кандидат технических наук, доцент Кур-
ского государственного медицинского уни-
верситета
- Месюра Владимир Иванович,*
кандидат технических наук, доцент Вин-
ницкого государственного технического
университета
- Мухамадиев Рим Фанавиевич,*
сотрудник Всероссийского НИИ техниче-
ской физики Российского федерального
ядерного центра
- Мухамадиева Татьяна Анатольевна,*
сотрудник Всероссийского НИИ техниче-
ской физики Российского федерального
ядерного центра
- Онищенко Сергей Анатольевич,*
студент Запорожского государственного
технического университета
- Онуфриенко Владимир Михайлович,*
кандидат физико-математических наук, де-

-
- кан Запорожского государственного технического университета
- Пархоменко Анжелика Владимировна,*
кандидат технических наук, доцент Запорожского государственного технического университета
- Питенко Александр Андреевич,*
аспирант Института вычислительного моделирования Сибирского отделения Российской академии наук
- Побийтеч Олег Александрович,*
адъюнкт Киевского военного института управления и связи
- Поздняков Андрей Анатолиевич,*
студент Запорожского государственного технического университета
- Попов Сергей Витальевич,*
аспирант Харьковского государственного технического университета радиотехники
- Поставец Лилия Владимировна,*
студентка Запорожского государственного технического университета
- Потапенко Евгений Михайлович,*
доктор технических наук, профессор Запорожского государственного технического университета
- Росицев Алексей Анатольевич,*
аспирант Института вычислительного моделирования Сибирского отделения Российской академии наук
- Саенко Елена Романовна,*
заместитель декана Запорожского государственного университета
- Самолчев Петр Александрович,*
аспирант Запорожского государственного технического университета
- Сарафанов Альберт Викторович,*
кандидат технических наук, докторант Московского государственного института электроники и математики (технического университета)
- Сенашова Мария Юрьевна,*
кандидат физико-математических наук, научный сотрудник Института вычислительного моделирования Сибирского отделения Российской академии наук
- Сердюк Сергей Никитович,*
кандидат технических наук, доцент Запорожского государственного технического университета
- Слюсарова Татьяна Ивановна,*
аспирантка Запорожского государственного технического университета
- Сокирук Виталий Виталиевич,*
студент магистратуры Винницкого государственного технического университета
- Субботин Сергей Александрович,*
инженер-программист Запорожского государственного технического университета
- Табуницкий Галина Владимировна,*
аспирантка Запорожского государственного технического университета
- Терехов Сергей Александрович,*
сотрудник Всероссийского НИИ технической физики Российского федерального ядерного центра
- Федорова Нелли Николаевна,*
сотрудник Всероссийского НИИ технической физики Российского федерального ядерного центра
- Флора Валентин Данилович,*
кандидат технических наук, доцент Запорожского государственного технического университета
- Хорев Александр Павлович,*
директор Симферопольского почтамта
- Цыганков Владимир Дмитриевич,*
кандидат технических наук, начальник отдела ЦНИТИ, г.Москва
- Чепкасов Юрий Иванович,*
кандидат технических наук, доцент Донецкого государственного технического университета

Черникова Лидия Вячеславовна,
аспирантка Донецкого государственного
технического университета

Шамровский Александр Дмитриевич,
кандидат физико-математических наук,
профессор Запорожской государственной
инженерной академии

Шило Галина Николаевна,
аспирантка Запорожского государственного

технического университета

Шушлятин Евгений Андреевич,
кандидат технических наук, докторант
Севастопольского государственного техни-
ческого университета

Яремчук Евгений Викторович,
аспирант Винницкого государственного тех-
нического университета