

Запорізький національний технічний університет



Радіоелектроніка Інформатика Управління

2(25)'2011

Науковий журнал

Виходить двічі на рік

Видається з березня 1999 року

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

Свідоцтво – серія **КВ № 6904**

Засновник і видавник – Запорізький національний технічний університет

Запоріжжя, ЗНТУ
2011

Постановою президії ВАК України № 1-05/4 від 26.05.2010 р. журнал «Радіоелектроніка, інформатика, управління» (скорочена назва – РІУ), який видається з 1999 року, включений до переліку наукових фахових видань України, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора і кандидата технічних наук та фізико-математичних наук (радіофізика).

Журнал є донором журналу «Telecommunications and Radio Engineering», який видається в США.

Інтернет-сторінка журналу: <http://journal.zntu.edu.ua/ric/index.php?page=index>.

Статті, що публікуються в журналі, реферуються в базах даних та РЖ ВІНІТІ (Росія) і «Джерело» (Україна). Журнал РІУ міститься у міжнародній базі наукових видань Index Copernicus (<http://journals.indexcopernicus.com/index.php>), електронна копія журналу розміщена на сайті Національної бібліотеки України імені В. І. Вернадського НАН України у розділі «Наукова періодика України» за адресою: <http://nbuv.gov.ua/portal/>.

Журнал розповсюджується за Каталогом періодичних видань України (передплатний індекс – 22914).

Науковий журнал друкує оригінальні та оглядові статті науковців ВНЗ і установ України та інших країн відповідно до рубрик:

- | | |
|---|---|
| – радіофізика; | – прогресивні інформаційні технології; |
| – радіоелектроніка та телекомунікації; | – теорія і методи автоматичного управління; |
| – математичне та комп'ютерне моделювання; | – управління у технічних системах. |
| – нейроінформатика та інтелектуальні системи; | |

РЕДАКЦІЙНА КОЛЕГІЯ

Головний редактор – д-р техн. наук Піза Д. М.

Заст. головного редактора – канд. техн. наук Дубровін В. І.

Члени редколегії

д-р техн. наук Андрієнко П. Д.

д-р фіз.-мат. наук Ахметшин А. М.

д-р техн. наук Волков О. В.

д-р фіз.-мат. наук Горбань О. М.

д-р фіз.-мат. наук Горр Г. В.

д-р техн. наук Гостєв В. І.

д-р фіз.-мат. наук Дробахін О. О.

д-р техн. наук Карпуков Л. М.

д-р фіз.-мат. наук Корніч Г. В.

д-р техн. наук Кулік А. С.

д-р фіз.-мат. наук Матюшин В. М.

д-р фіз.-мат. наук Онуфрієнко В. М.

д-р фіз.-мат. наук Погосов В. В.

д-р техн. наук Потапенко Є. М.

д-р техн. наук Толок В. О.

д-р фіз.-мат. наук Чумаченко В. П.

Рекомендовано до видання вченою радою Запорізького національного технічного університету, протокол № 11 від 30.05.2011 р.

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

Журнал зверстаний редакційно-видавничим відділом
Запорізького національного технічного університету.

Адреса редакції: 69063, м. Запоріжжя, вул. Жуковського, 64, ЗНТУ,
редакція журналу «РІУ».

Тел: (061) 769-82-96 – редакційно-видавничий відділ

Факс: (061) 764-46-62

E-mail: rvv@zntu.edu.ua

ЗМІСТ

РАДІОФІЗИКА.....	7
------------------	---

<i>Романенко С. Н., Карпуков Л. М., Борисенко В. А., Львов А. С.</i> ПОВЕРХНОСТНЫЕ ВОЛНЫ НА ГРАНИЦАХ РАЗДЕЛА МЕТАСРЕД. МОДЕЛИРОВАНИЕ ЭФФЕКТОВ СВЕРХРАЗРЕШЕНИЯ.....	7
---	---

РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ.....	14
--	----

<i>Проскурін М. П., Костенко В. Л., Щекотихін О. В., Грушко С. С.</i> ОБГРУНТУВАННЯ НЕОБХІДНОСТІ ПЕРЕХОДУ ДО МІКРОПОТУЖНИХ ТВЕРДОТІЛЬНИХ ІНТЕГРАЛЬНИХ СХЕМ З ОПТОЕЛЕКТРОННИМИ КОМПОНЕНТАМИ ДЛЯ ЦИФРОВИХ АВТОМАТІВ ТА ОБЧИСЛЮВАЛЬНИХ ПРИСТРОЇВ.....	14
---	----

<i>Морцавка С. В., Шама Є. О., Піза Д. М.</i> ВПЛИВ ВИДІВ НОРМУВАННЯ НА ЯКІСТЬ РОЗПІЗНАННЯ РОСЛИННИХ ОБ'ЄКТІВ.....	30
--	----

<i>Сметанин И. Н., Ложковский А. Г., Пиза Д. М., Вербанов О. В.</i> ИССЛЕДОВАНИЕ ПАРАМЕТРОВ ТРАФИКА УЧАСТКА СЕТИ GSM ПРИ ВОЗДЕЙСТВИИ ЛОКАЛЬНЫХ ПЕРЕГРУЗОК.....	23
---	----

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ.....	34
---	----

<i>Агібалов А. П., Поляков М. А.</i> ТРАНСЛЯТОР ПАРАМЕТРОВ МОДЕЛИ КОНЕЧНОГО АВТОМАТА ИЗ СРЕДЫ МАТЛАВ В ПРИЛОЖЕНИЕ ЧЕЛОВЕКО-МАШИННОГО ИНТЕРФЕЙСА.....	34
--	----

<i>Юр Т. В., Харитонов В. Н., Дубровин В. И.</i> МОДЕЛЬ ДИАГНОСТИРОВАНИЯ УЗЛОВ ГАЗОТУРБИННЫХ ДВИГАТЕЛЕЙ С ПРИМЕНЕНИЕМ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ И S-ДИСКРИМИНАНТА.....	60
---	----

<i>Андрущенко Д. М., Варава М. Ю., Неласа Г. В.</i> РОЗПАРАЛЕЛЮВАННЯ ρ - I λ - МЕТОДІВ ПОЛЛАРДА РОЗВ'ЯЗАННЯ ЗАДАЧІ ДИСКРЕТНОГО ЛОГАРИФМУВАННЯ.....	37
--	----

<i>Зайцев С. А., Субботин С. А.</i> ПОСТРОЕНИЕ ДИАГНОСТИЧЕСКИХ МОДЕЛЕЙ НА ОСНОВЕ ПАРАДИГМЫ ОТРИЦАТЕЛЬНОГО ОТБОРА С ИСПОЛЬЗОВАНИЕМ ПРИНЦИПА МАСКИРОВАНИЯ ДЕТЕКТОРОВ.....	65
---	----

<i>Кулик А. С., Лученко О. О., Фирсов С. Н.</i> КОНЦЕПЦИЯ ОБЕСПЕЧЕНИЯ ЖИВУЧЕСТИ СПУТНИКОВЫХ СИСТЕМ УПРАВЛЕНИЯ ОРИЕНТАЦИЕЙ И СТАБИЛИЗАЦИЕЙ.....	41
---	----

<i>Чопоров С. В.</i> ПОСТРОЕНИЕ НЕРАВНОМЕРНЫХ ДИСКРЕТНЫХ СЕТОК ДЛЯ ФУНКЦИОНАЛЬНЫХ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ НА БАЗЕ ТЕОРИИ R-ФУНКЦИЙ.....	70
--	----

<i>Миронова Н. А.</i> ИНТЕГРАЦИЯ МОДИФИКАЦИЙ МЕТОДА АНАЛИЗА ИЕРАРХИИ ДЛЯ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ ГРУППОВЫХ РЕШЕНИЙ.....	47
--	----

<i>Пинчук В. П., Подковалихина Е. А.</i> ОБЪЕКТНО-ОРИЕНТИРОВАННЫЙ ПОДХОД ДЛЯ ПОСТРОЕНИЯ МОДЕЛЕЙ СЛОЖНЫХ СИСТЕМ.....	76
---	----

<i>Кошевой Н. Д., Михайлов А. Г.</i> РАЗРАБОТКА АЛГОРИТМОВ МОДЕЛИРОВАНИЯ РАБОТЫ ДАТЧИКОВ РАСХОДА ТОПЛИВА И ВЗАИМОДЕЙСТВИЯ С ТОПЛИВНОЙ МАГИСТРАЛЬЮ.....	54
--	----

НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ.....82

Асеев Г. Г.
МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА
ДАННЫХ
В ЭЛЕКТРОННЫХ ХРАНИЛИЩАХ: ГЕНЕТИЧЕСКИЕ
АЛГОРИТМЫ.....82

Кротких С. С., Кириченко Л. О.
ИССЛЕДОВАНИЕ ВЫЗВАННЫХ ПОТЕНЦИАЛОВ
В ЭЭГ ЧЕЛОВЕКА С ПОМОЩЬЮ
ДИСКРЕТНОГО ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ.....86

Литвин В. В.
МОДЕЛЮВАННЯ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ
ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ
З ВИКОРИСТАННЯМ ОНТОЛОГІЧНОГО
ПІДХОДУ.....93

Баркалов А. А., Зеленева И. Я., Цололо С. А., Биайрак Х.
УМЕНЬШЕНИЕ ПЛОЩАДИ МАТРИЧНОЙ СХЕМЫ
УСТРОЙСТВА УПРАВЛЕНИЯ С РАЗДЕЛЕНИЕМ
КОДОВ.....101

Пицухина О. А., Клочок А. Ю.
ПОДХОД К ФОРМИРОВАНИЮ ОБРАТНОЙ СВЯЗИ
В ИНТЕЛЛЕКТУАЛЬНЫХ ОБУЧАЮЩИХ СИСТЕМАХ
В СФЕРЕ ВЫСШЕГО ТЕХНИЧЕСКОГО
ОБРАЗОВАНИЯ.....107

Степаненко О. О., Піза Д. М.
ПРОГРАМНИЙ КОМПЛЕКС ДЛЯ АНАЛІЗУ
І ОБРОБКИ ЕХО-ІМПУЛЬСНИХ ЗОБРАЖЕНЬ.....110

Шафроненко А. Ю., Волкова В. В., Бодянский Е. В.
АДАПТИВНАЯ КЛАСТЕРИЗАЦИЯ ДАННЫХ С
ПРОПУЩЕННЫМИ ЗНАЧЕНИЯМИ.....115

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ.....120

Брагина Т. И., Табуницкий Г. В.
АНАЛИЗ ПОДХОДОВ К УПРАВЛЕНИЮ РИСКАМИ
В ПРОГРАММНЫХ ПРОЕКТАХ С ИТЕРАТИВНЫМ
ЖИЗНЕННЫМ ЦИКЛОМ.....120

Дьячук Т. С.
ОЦЕНКА ХАРАКТЕРИСТИК РАСПРЕДЕЛЕННОЙ
СИСТЕМЫ.....124

Гонтарь Н. А., Кудерметов Р. К.
РАЗРАБОТКА ОНТОЛОГИИ СИСТЕМНОГО
ИНЖИНИРИНГА КОСМИЧЕСКИХ СИСТЕМ.....131

Скрупский С. Ю., Луценко Н. В., Скрупская Л. С.
ПАРАМЕТРЫ КОМПРЕССИИ ВИДЕОИНФОРМАЦИИ
В РАСПРЕДЕЛЕННЫХ СИСТЕМАХ.....137

Сметанин Р. И., Тягунова М. Ю.
ПОДХОД К РАСПРЕДЕЛЕНИЮ ПАРАЛЛЕЛЬНЫХ
ВЕТВЕЙ ЗАДАНИЯ В ВЫЧИСЛИТЕЛЬНОЙ
СИСТЕМЕ.....143

УПРАВЛІННЯ У ТЕХНІЧНИХ СИСТЕМАХ.....148

*Александрова Т. Е., Кононенко В. А., Лазаренко А. А.,
Зейн Али*
СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЦИФРОВЫХ
ПД-СТАБИЛИЗАТОРОВ ПОДВИЖНЫХ ОБЪЕКТОВ
С НИЗКОЧАСТОТНЫМИ ФИЛЬТРАМИ БАТТЕРУОРТА
И ЛАНЦОША.....148

Бур'ян С. О., Грищук Т. В.
ЕКСТРЕМАЛЬНА ЕЛЕКТРОМЕХАНІЧНА СИСТЕМА
КЕРУВАННЯ ПАРАЛЕЛЬНО З'ЄДНАНИМИ
НАСОСАМИ ВОДОПОСТАЧАННЯ.....153

Шановні колеги!

Запрошуємо Вас взяти участь у роботі VI міжнародної науково-практичної конференції «Сучасні проблеми і досягнення в галузі радіотехніки, телекомунікацій та інформаційних технологій», яка буде проводитися у Запорізькому національному технічному університеті 19–21 вересня 2012 р.

Мета конференції: аналіз і узагальнення нових теоретичних і практичних результатів у відповідних галузях знань.

Адреса оргкомітету: вул. Жуковського, 64, м. Запоріжжя, Україна, 69063,

ЗНТУ, кафедра радіотехніки та телекомунікацій.

Секретар оргкомітету Колеснікова Євгенія Ісааківна

тел.: (061) 764-32-81, 769-84-31; факс: (061) 764-46-62; e-mail: kolevis@zntu.edu.ua.

CONTENTS

RADIOPHYSICS.....7

*Romanenko S. N., Karpukov L. M., Borisenko V. A.,
Lvov A. C.*

SURFACE WAVES AT THE INTERFACE OF
METAMEDIALS. MODELING OF SUPERRESOLUTION
EFFECTS.....7

RADIO ELECTRONICS AND TELECOMMUNICATIONS.....14

Proskurin N. P., Kostenko V. L., Schekotihin O. V., Grushko S. S.
VALID NECESSITY OF TRANSFER TO MICROPOWER
SOLID STATE INTEGRATED CIRCUITS WITH
OPTRONIC COMPONENTS FOR DIGITAL MACHINES
AND COMPUTING DEVICES.....14

Smetanin I. N., Lozhkovsky A. G., Piza D. M., Verbanov O. V.
RESEARCH OF PARAMETERS OF THE TRAFFIC OF
A SECTION OF NETWORK GSM AT INFLUENCE OF
LOCAL OVERLOADS.....23

Morshchavka S. V., Shama E. O., Piza D. M.
EFFECT OF NORMALIZATION ON THE QUALITY OF
RECOGNITION OF PLANT OBJECTS.....30

MATHEMATICAL AND COMPUTER MODELLING.....34

Agibalov A. P., Polyakov M. A.
TRANSLATOR OF FINITE STATE MACHINE MODEL
PARAMETERS FROM MATLAB ENVIRONMENT INTO
HUMAN-MACHINE INTERFACE APPLICATION.....34

Andrushchenko D. M., Varava M. U., Nelasa G. V.
PARALLELIZATION OF ρ - AND λ - POLLARD'S
METHODS FOR SOLVING THE DISCRETE
LOGARITHM.....37

Kulik A. S., Luchenko O.O., Firsov S. N.
THE CONCEPT OF SURVIVABILITY OF SATELLITE
SYSTEMS, ATTITUDE CONTROL AND
STABILIZATION.....41

Mironova N. O.
INTEGRATION MODIFICATIONS OF THE ANALYTICAL
HIERARCHY PROCESS FOR GROUP DECISION
MAKING SUPPORT SYSTEMS.....47

Koshevoy N. D., Mikhailov A. G.
MODELING ALGORITHMS DEVELOPMENT OF
DEVICE FUNCTION OF THE FUEL CHARGE AND
INTERACTION WITH A FUEL MAIN.....54

Yur T. V., Haritonov V. N., Dubrovin V. I.
DIAGNOSTIC MODEL FOR GAS TURBINE ENGINE
ELEMENTS BASED ON WAVELET ANALYSIS AND S-
DISCRIMINANT.....60

Zaitsev S. A., Subbotin S. A.
THE DIAGNOSIS MODEL BUILDING ON THE BASIS OF
NEGATIVE SELECTION PARADIGM USING THE
PRINCIPLE OF DETECTOR ASKING.....65

Choporov S. V.
GENERATION OF NONUNIFORM HEXAHEDRAL
ELEMENT MESHES FOR FUNCTIONAL MODELS ON
THE BASIS OF R-FUNCTIONS.....70

Pinchuk V. P., Podkoralihina E. A.
OBJECT-ORIENTED APPROACH FOR MODELS
DESIGNING OF THE COMPLEX SYSTEMS.....76

NEUROINFORMATICS AND INTELLIGENT SYSTEMS.....82

Aseyev G. G.
METHODS OF INTELLECTUAL ANALYSIS OF DATA IN
ELECTRONIC DEPOSITORIES:
GENETIC ALGORITHMS.....82

Krotkih S. S., Kirichenko L. O. ,
ANALYSIS OF EVENT-RELATED POTENTIALS
OF EEG SIGNAL USING DISCRETE WAVELET
TRANSFORM.....86

Lytvyn V. V.
INTELLIGENT DECISION SUPPORT SYSTEMS
MODELING USING ONTOLOGICAL
APPROACH.....93

Barkalov A. A., Zelenyova I. J., Tsololo S. A., Biayarek H.
REDUCTION OF TERM MATRIX OF CONTROL
UNIT WITH CODE SHARING.....101

Pishchukhina O. A., Klochok A. Yu.
APPROACH TO THE FEEDBACK FORMING IN
INTELLIGENT LEARNING SYSTEM IN THE SPHERE OF
HIGHER TECHNICAL EDUCATION.....107

Stepanenko O. O., Piza D. M.
PROGRAM COMPLEX FOR ANALYSIS AND
TREATMENT ECHO-PULSE IMAGES.....110

Shafronenko A. Yu. Volkova V. V., Bodyanskiy Yev. V.
ADAPTIVE CLUSTERING WITH MISSING
VALUES.....115

PROGRESSIV INFORMATICS TECHNOLOGIES.....120

Bragina T. I., Tabunshchik G. V.
ANALYSIS RISK MANAGEMENT APPROACHES IN
SOFTWARE PROJECTS WITH AN ITERATIVE
LIFECYCLE.....120

Diachuk T. S.
EVALUATION OF DISTRIBUTED SYSTEM'S
CHARACTERISTICS.....124

Gontar N. A., Kudermetov R. K.
WORKING OUT ONTOLOGY OF SYSTEMS
ENGINEERING OF SPACE SYSTEM.....131

Skrupsky S. Y., Lucenko N. V., Skrupskaya L. S.
THE PARAMETERS OF VIDEO INFORMATION
COMPRESSION PROCESS IN DISTRIBUTED
SYSTEMS.....137

Smetanin R. I., Tyagunova M. Yu.
APPROACH TO THE DISTRIBUTION OF PARALLEL
TASK THREADS IN COMPUTING SYSTEMS.....143

CONTROL IN TECHNICAL SYSTEMS.....148

*Alexandrova T. Ye., Kononenko V. A., Lazarenko A. A.,
Zein Ali*
COMPARATIVE ANALYSIS OF DIGITAL PD-
CTABILIZERS OF MOBILE OBJECTS WITH LOW-
FREQUENCY BUTTERWORT
AND LANCZOS FILTERS.....148

Buryan S., Gryshuk T.
EXTREMAL ELECTROMECHANICAL CONTROL
SYSTEM OF WATER SUPPLY PUMPS CONNECTED IN
PARALLEL.....153

РАДИОФИЗИКА

РАДИОФИЗИКА

RADIOPHYSICS

УДК 537.876.4

Романенко С. Н.¹, Карпуков Л. М.², Борисенко В. А.³, Львов А. С.⁴¹Канд. физ.-мат. наук, доцент Запорожского национального технического университета²Д-р техн. наук, профессор Запорожского национального технического университета³Ст. преподаватель Запорожского национального технического университета⁴Аспирант Запорожского национального технического университета

ПОВЕРХНОСТНЫЕ ВОЛНЫ НА ГРАНИЦАХ РАЗДЕЛА МЕТАСРЕД. МОДЕЛИРОВАНИЕ ЭФФЕКТОВ СВЕРХРАЗРЕШЕНИЯ

Теоретически обосновано появление поверхностных волн (плазмонов), возникающих на границе раздела метасреды и обычного вещества. Представлены результаты численного моделирования эффектов сверхразрешения при использовании плоских линз из метаматериала.

Ключевые слова: метасреда, граница раздела, плазмон, плоская линза, сверхразрешение.

ВВЕДЕНИЕ

В работе [1] Веселаго впервые теоретически исследовал и обосновал возможность создания материалов с отрицательными значениями диэлектрической ϵ и магнитной μ проницаемостей. Такие материалы получили название сред с отрицательной рефракцией (NR-среды «Negative Refraction», метасреды, левосторонние среды, метаматериалы). Интерес к ним обусловлен тем, что такие материалы должны обладать уникальными свойствами, которые отсутствуют у обычных природных веществ. В частности, одно из свойств состоит в том, что при падении волны из обычной среды на границу раздела с левосторонней средой, преломленная в левостороннюю среду волна выйдет под отрицательным углом (с противоположной стороны от нормали в сравнении с обычным преломлением). При этом величина угла преломления определяется известным законом Снеллиуса, который в случае применения к левосторонним средам должен быть модифицирован следующим образом:

$$\frac{\sin \varphi}{\sin \psi} = n_{1,2} = \frac{p_2}{p_1} \sqrt{\frac{\epsilon_2 \mu_2}{\epsilon_1 \mu_1}}.$$

В этом выражении φ – угол падения, ψ – угол преломления, $p_i = +1$ ($i=1, 2$), если среда обычная, и $p_i = -1$, если среда левосторонняя.

Отрицательный угол преломления позволяет осуществить фокусировку поля с использованием плоского слоя метаматериала, что было впервые показано в работе Пендри [2]. Кроме того, на основе свойства фокусировки, в работе [3] представлены результаты эксперимента, в котором был впервые преодолен дифракционный предел, что получило название сверхразрешения.

Другое интересное свойство заключается в том, что на границе раздела метасреды и обычного вещества происходит значительное увеличение амплитуды падающей затухающей волны за счет резонансного возбуждения поверхностных волн (поверхностных плазмонов). При этом длина волны плазмона может быть на несколько порядков короче длины падающей волны, что открывает возможность для развития, в частности, оптики очень коротких поверхностных волн (например, оптических преобразований – отклонение, фокусировка, гауссовы пучки, фотонные кристаллы и т. п.).

Реализованные к настоящему времени структуры метаматериалов представляют собой комбинации металлических проводящих фрагментов различной формы, периодически расположенных в свободном пространстве. Эффективным средством электродинамического моделирования таких структур является интегральное уравнение Поклингтона, обеспечивающее необходимую точность расчетов.

В настоящей работе с использованием аппарата матриц рассеяния теоретически обосновано резонансное увеличение амплитуды затухающих волн при их падении на слой метаматериала. В работе приведены также результаты численного моделирования плоских слоев метаматериалов (плоских линз), обладающих фокусирующими свойствами и позволяющих преодолеть дифракционный предел. Хорошая точность моделирования подтверждена данными эксперимента.

1. УВЕЛИЧЕНИЕ АМПЛИТУДЫ ЗАТУХАЮЩЕЙ ВОЛНЫ В СЛОЕ МЕТАМАТЕРИАЛА

В работе [2] Пендри показал, что дифракционный предел, характерный для обычных линз, может быть преодолен с помощью плоского слоя метаматериала (плоской линзы) с параметрами $\varepsilon = \mu = -1$. При этом, за счет значительного увеличения амплитуды высокочастотных затухающих пространственных гармоник, в процессе их прохождения через слой метаматериала, появляется возможность выявить более мелкие детали изображения и, следовательно, преодолеть дифракционный предел.

Рассмотрим слой метаматериала толщиной d с параметрами $\varepsilon_2, \mu_2 < 0$, расположенный в свободном пространстве (рис. 1). Электрическое поле объекта, излучающего на частоте ω , может быть представлено в виде разложения по пространственным гармоникам в двумерный ряд Фурье

$$\begin{aligned} \vec{E}(x, y, z, t) &= \vec{E}(x, y, z) e^{j\omega t} = \\ &= \sum_n \sum_{k_x, k_y} E_n(k_x, k_y) \cdot e^{-j(k_x x + k_y y + k_z z - \omega t)}, \end{aligned} \quad (1)$$

где

$$k_{zi} = \begin{cases} \sqrt{k_i^2 - (k_x^2 + k_y^2)} = p_i, & \text{если } k_x^2 + k_y^2 = \gamma^2 < k_i^2, \\ -j\sqrt{(k_x^2 + k_y^2) - k_i^2} = -jq_i, & \text{если } k_x^2 + k_y^2 = \gamma^2 > k_i^2, \end{cases} \quad (2)$$

$k_i = \frac{\omega}{c} \cdot \sqrt{\varepsilon_i \cdot \mu_i} = \frac{\omega}{c} \cdot n_i$; $i = 1, 2$ – волновое число среды; n_i – коэффициент преломления среды.

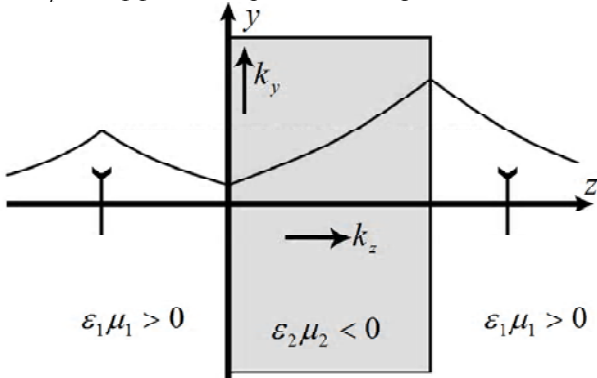


Рис. 1. Прохождение затухающих гармоник через слой метаматериала

Знак (–) перед корнем в (2) выбран из условия затухания волны на бесконечности.

Разложение (1) переводит пространственное (x, y) представление поля объекта в спектральную область (k_x, k_y) , где вся информация об изображении объекта содержится в спектральных компонентах k_x, k_y . При этом малые значения $\gamma = \sqrt{k_x^2 + k_y^2}$ и, следовательно, большие значения $\lambda_{\perp} = 2\pi/\gamma$, соответствуют крупным деталям изображения, и наоборот – большие значения γ (малые $\lambda_{\perp}$) соответствуют мелким деталям изображения. Информация об изображении переносится к слою метаматериала (к линзе) волной, распространяющейся вдоль оси z . Если зависимость этой волны от z имеет вид $e^{-jk_z z}$ ($k_z = k_{z1}$, свободное пространство), то из (2) следует, что при $\gamma < k_1$ (соответственно, $\lambda_{\perp} > \lambda_1$) $k_{z1} = p_1$ – вещественное число и, следовательно, волна распространяется без затухания, перенося информацию об изображении к линзе. Если же $\gamma > k_1$, или $\lambda_{\perp} < \lambda_1$, то $k_{z1} = -jq_1$ – чисто мнимое число и волна будет экспоненциально затухать вдоль оси z , как качественно показано на рис. 1.

Для парциальных ТЕ и ТМ гармоник при нормальном падении на границу раздела сред коэффициенты отражения определяются известными выражениями:

$$\Gamma^{TE} = \frac{k_{z2}/\mu_2 - k_{z1}/\mu_1}{k_{z2}/\mu_2 + k_{z1}/\mu_1}, \quad \Gamma^{TM} = \frac{k_{z2}/\varepsilon_2 - k_{z1}/\varepsilon_1}{k_{z2}/\varepsilon_2 + k_{z1}/\varepsilon_1}.$$

Процесс прохождения затухающих вдоль оси z волн в структуре на рис. 1 описывается суммарными коэффициентами передачи и отражения. В рамках теории длинных линий модель структуры, очевидно, состоит из трех каскадно соединенных элементов: двух скачков параметров, характеризующих границы раздела, и отрезка линии передачи длиной d , моделирующего слой метаматериала, как показано на рис. 2.

Анализ такой каскадной структуры может быть выполнен с использованием аппарата ABCD матриц, однако большей общностью и универсальностью обладает подход на основе матриц рассеяния (S -матриц). В спектральной области границе раздела сред соответствует матрица рассеяния, имеющая вид

$$S_1 = \begin{bmatrix} \Gamma & 1 - \Gamma \\ 1 + \Gamma & -\Gamma \end{bmatrix}, \quad (3)$$

где $\Gamma = \Gamma^{TE}$, либо $\Gamma = \Gamma^{TM}$.

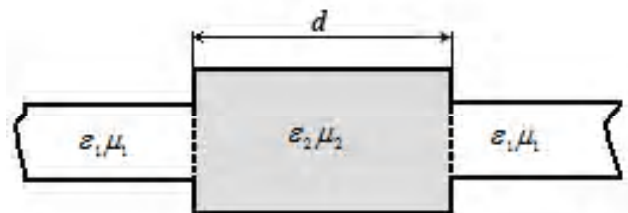


Рис. 2. Модель слоя метаматериала

Матрица рассеяния отрезка линии передачи, длиной d , определяется выражением

$$S_2 = \begin{bmatrix} 0 & e^{-jk_z d} \\ e^{-jk_z d} & 0 \end{bmatrix}, \quad (4)$$

где k_z – постоянная распространения вдоль оси z .

Каждой матрице рассеяния может быть поставлен в соответствие ориентированный граф [4], в результате ориентированный граф всей структуры будет иметь вид, представленный на рис. 3.

Непосредственно по графу структуры находится суммарный коэффициент передачи

$$T_\Sigma = S_{\Sigma 21} = \frac{(1+\Gamma)(1-\Gamma) \cdot e^{-jk_z d}}{1-\Gamma^2 e^{-j2k_z d}} \quad (5)$$

и суммарный коэффициент отражения

$$R_\Sigma = S_{\Sigma 11} = \frac{\Gamma(1-e^{-j2k_z d})}{1-\Gamma^2 e^{-j2k_z d}}. \quad (6)$$

Обозначив для упрощения $K = \frac{k_{z2}/\varepsilon_2}{k_{z1}/\varepsilon_1}$, и рассмат-

ривая лишь ТМ волны с учетом того, что $\varepsilon_2 < 0$, получим

$$\Gamma = \Gamma^{TM} = \frac{k_{z2}/\varepsilon_2 - k_{z1}/\varepsilon_1}{k_{z2}/\varepsilon_2 + k_{z1}/\varepsilon_1} = \frac{k_{z1}/\varepsilon_1 + k_{z2}/\varepsilon_2}{k_{z1}/\varepsilon_1 - k_{z2}/\varepsilon_2} = \frac{1+K}{1-K} \quad (7)$$

При анализе ТЕ волн необходимо использовать Γ^{TE} и учитывать $\mu_2 < 0$. Остальные выкладки аналогичны.

Подставляя (7) в (5) и (6) и учитывая, что $k_{z2} = -jq_2$, находим

$$T_\Sigma = \frac{\left[1 - \left(\frac{1-K}{1+K}\right)^2\right] \cdot e^{+q_2 d}}{1 - \left(\frac{1-K}{1+K}\right)^2 \cdot e^{+2q_2 d}}, \quad (8)$$

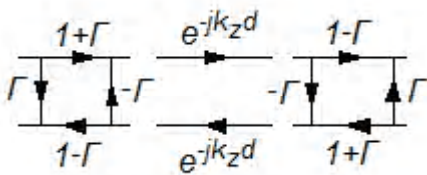


Рис. 3. Ориентированный граф структуры

$$R_\Sigma = \frac{\frac{1-K}{1+K} \cdot (1-e^{+2q_2 d})}{1 - \left(\frac{1-K}{1+K}\right)^2 \cdot e^{+2q_2 d}}. \quad (9)$$

Из (8) видно, что при $K \rightarrow 1$ значение T_Σ растет по экспоненциальному закону и, в зависимости от толщины слоя, может принимать сколь угодно большие значения. Отсюда следует, что если в структуре отсутствуют потери, то амплитуды полей становятся бесконечно большими. В реальных структурах с потерями амплитуды полей остаются конечными.

Более детальный анализ показывает, что в структуре будет наблюдаться резонанс, когда знаменатель в (8) и (9) обращается в нуль. При этом, очевидно, возможны два значения K :

$$K_1 = \frac{1+e^{-q_2 d}}{1-e^{-q_2 d}}, \quad K_2 = \frac{1-e^{-q_2 d}}{1+e^{-q_2 d}}. \quad (10)$$

Наличие двух значений K в (10) соответствует расщеплению резонанса в слое метаматериала и связано с наличием симметричной и антисимметричной поверхностных мод (поверхностных плазмонов) на границах

раздела. При большой толщине слоя $e^{-q_2 d} \ll 1$ расщепление мало и значения $K_1 \approx K_2 \approx 1$. При малой толщине слоя $1+e^{-q_2 d} \gg 1-e^{-q_2 d}$ и расщепление резонансных мод становится большим, при этом $K_1 \gg K_2$.

Из (2), с учетом (10), для толстого слоя метаматериала ($e^{-q_2 d} \ll 1$) можно получить выражение для волнового числа поверхностного плазмона

$$\gamma = \sqrt{\left(\frac{q_1 \cdot \varepsilon_2}{\varepsilon_1}\right)^2 + \left(\frac{\omega}{c}\right)^2 \varepsilon_2 \mu_2}, \quad (11)$$

где q_1 – волновое число падающей затухающей волны; c – скорость света в вакууме.

Из (11) следует, что волновое число плазмона может значительно превышать волновое число падающей волны (при больших $\varepsilon_2/\varepsilon_1$) и, следовательно, длина волны плазмона может быть значительно короче длины падающей волны.

Приведенные выкладки показывают, что при падении затухающей волны на плоский слой метаматериала происходит увеличение амплитуды волны, обусловленное возбуждением на границах раздела резонансных мод, называемых поверхностными плазмонами.

2. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Моделирование эффектов преодоления дифракционного предела (сверхразрешения) проводилось с использованием программы, разработанной на основе интег-

рального уравнения Поклингтона для анализа проволочных структур. На рис. 4 и 5 показаны модель плоской линзы, результаты расчетов в виде 3D-распределения электрического поля в плоскости расположения вибратора-зонда, а также график амплитуды электрического поля вдоль линии перемещения зонда.

Линза составлена из проволочных резонансных элементов в виде спиралей и линейных полуволновых отрезков, возбуждаемых соответственно магнитной и электрической компонентами поля. Представленные расчеты выполнены для размеров элементов структуры, экспериментальное исследование которой проведено в [3].

При моделировании все размеры структуры соответствовали параметрам эксперимента. Каждая спираль в линзе аппроксимировалась 12-ю прямолинейными сегментами одинаковой длины. Как видно, на рисунках 4, 5 явно выделяются максимумы поля, а их положение точно совпадает с расположением источников. Кроме того, результаты расчетов показывают, что разрешающая способность линзы связана с периодом решетки, составляющей структуру линзы. Этот факт отмечается и в ряде работ других авторов. Если оценивать разрешающую способность по уровню половины амплитуды, то для данной линзы расчетная величина разрешения составит $\Delta \approx 10$ мм.

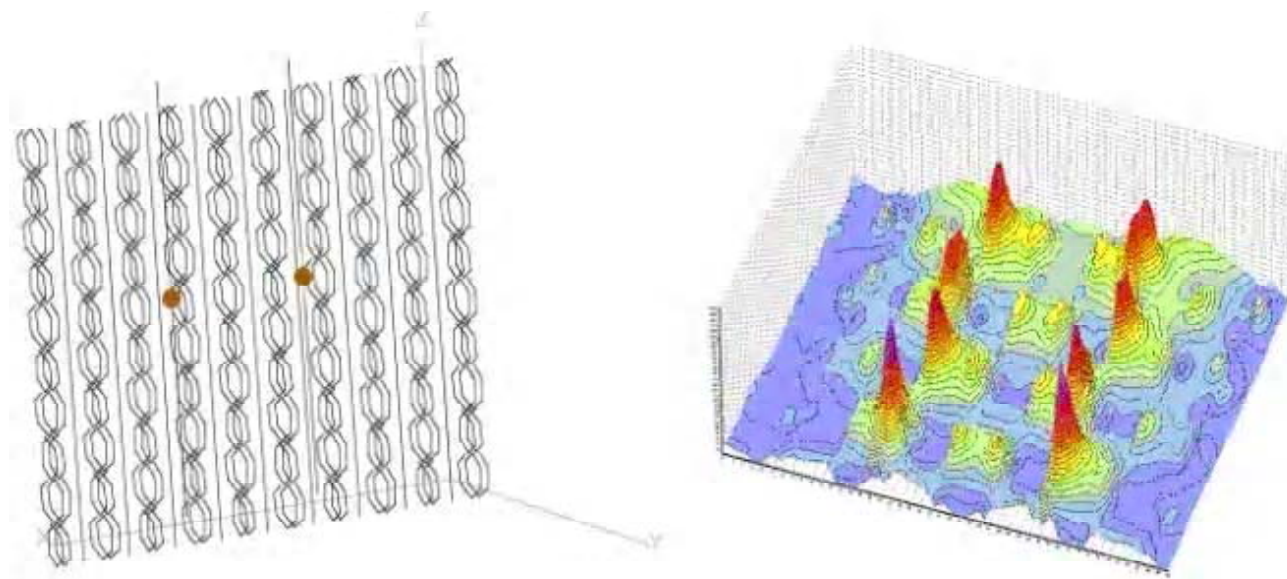


Рис. 4. Модель плоской линзы и 3D-картина электрического поля

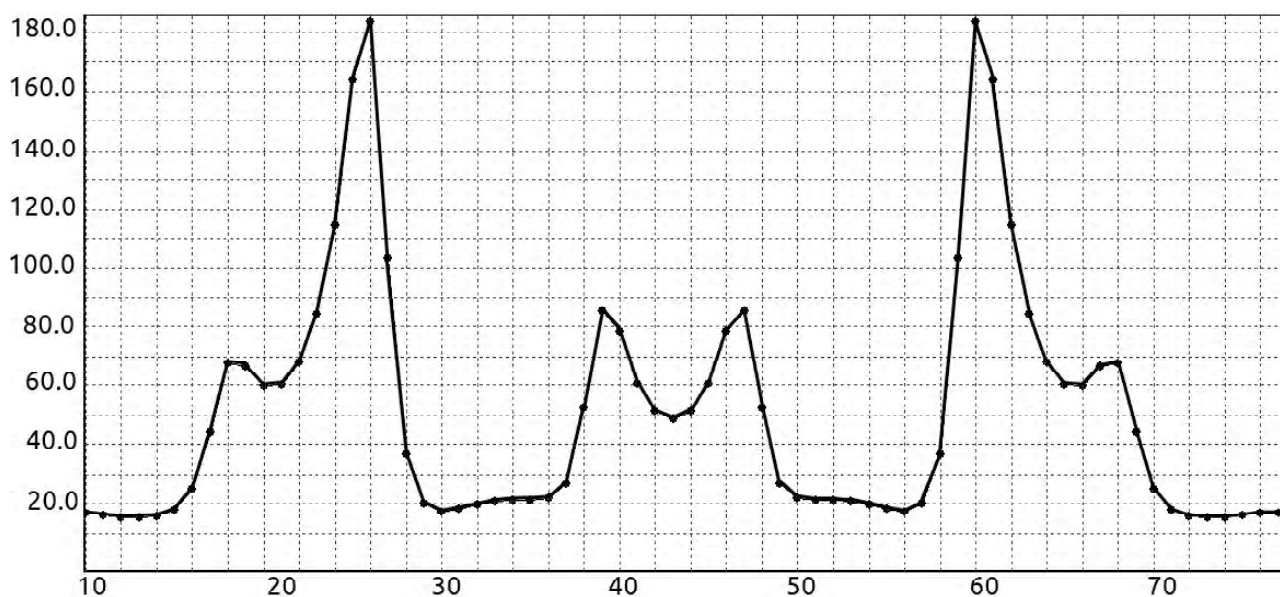


Рис. 5. График распределения z-компоненты электрического поля

Геометрия линзы, схема эксперимента и результаты измерений на резонансной частоте 1664 МГц показаны на рис. 6. В эксперименте расстояние между синфазно возбуждаемыми вибраторами составляло 30 мм, что соответствует примерно $1/6$ длины волны. На рис. 6 видно, что при наличии линзы четко видны два максимума поля, расстояние между которыми точно соответствует положению вибраторов.

При моделировании сближение источников до расстояния в 10 мм дает картину, представленную на рис. 7. Из рисунка следует, что даже для такого предельного сближения при численном моделировании еще можно различить источники. При дальнейшем сближении источников два максимума поля сливаются в один и эффект сверхразрешения исчезает. Следует отметить, что резонансная частота при численном моделировании (1628 МГц) несколько отличалась от резонансной частоты в эксперименте (1664 МГц). По-видимому, это связано с наличием диэлектрика в структуре, а также с несовершенством изготовления макета линзы.

Изменение частоты генератора в обе стороны от резонанса приводит к исчезновению эффекта сверхразрешения, что хорошо видно на рис. 8.

Эффект сверхразрешения можно наблюдать и для структуры, показанной на рис. 9. Здесь линза представляет собой массив чередующихся спиралей разнонаправленной намотки, с попарно ортогональным расположением в пространстве. Результаты моделирования такой структуры приведены на рис. 9 и 10.

Как и в предыдущем случае, на рисунках явно выделяются максимумы поля, расположение которых соответствует положению источников, однако высота максимумов меньше, а ширина – больше и, как следствие, меньше разрешающая способность такой линзы.

ЗАКЛЮЧЕНИЕ

В работе на базе аппарата матриц рассеяния теоретически обоснован эффект возникновения резонансных поверхностных волн (поверхностных плазмонов) при падении затухающей волны на границу раздела метасреды

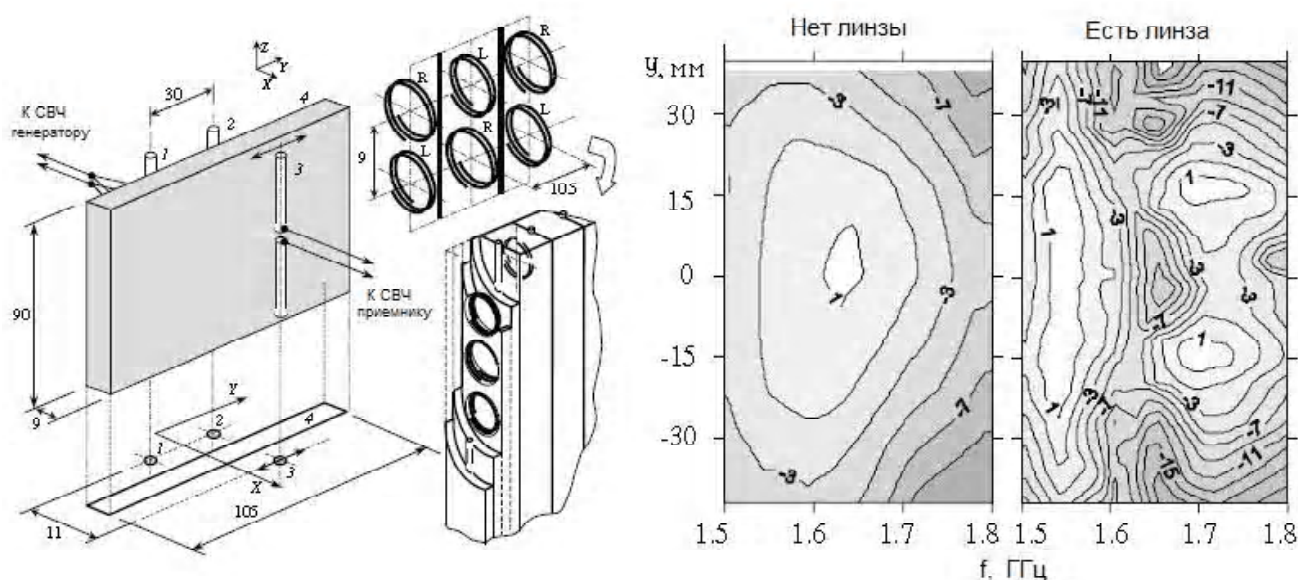


Рис. 6. Геометрия плоской линзы и результаты эксперимента в [3]

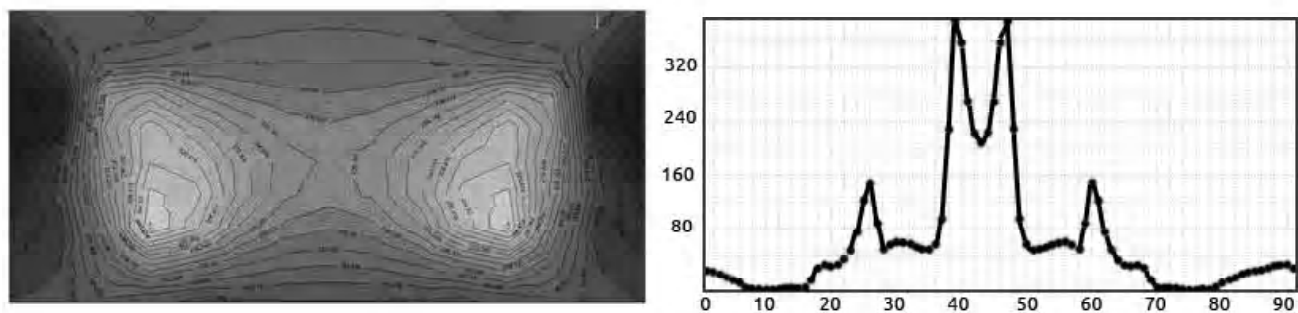


Рис. 7. Линии уровня и график распределения z-компоненты поля при сближении источников

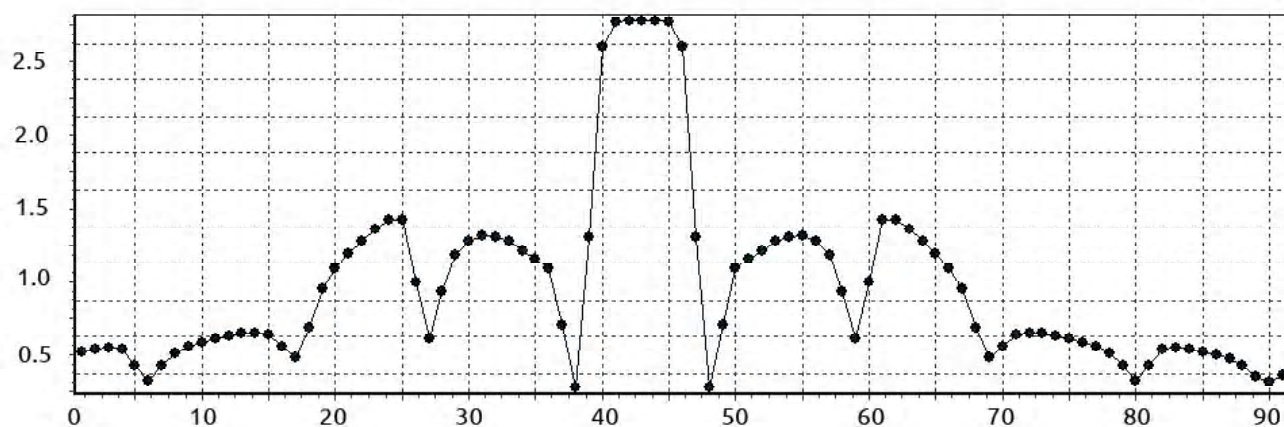


Рис. 8. График z -компоненты поля при расстоянии между источниками 10 мм и $f=1800$ МГц

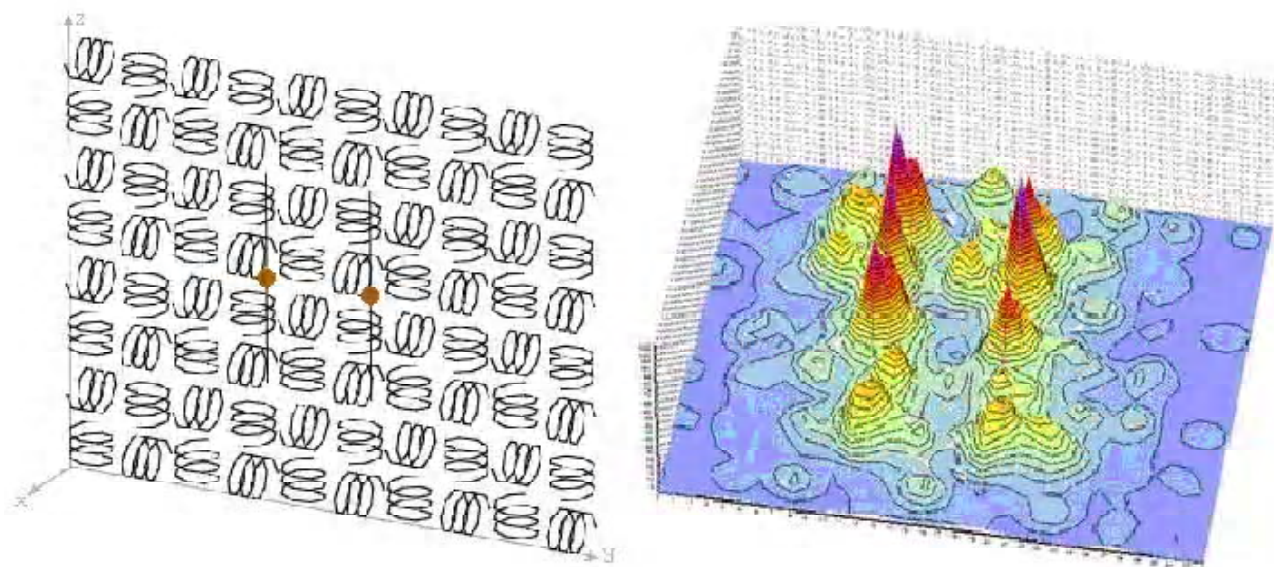


Рис. 9. Линза из разнонаправленных спиралей и 3D-картина электрического поля

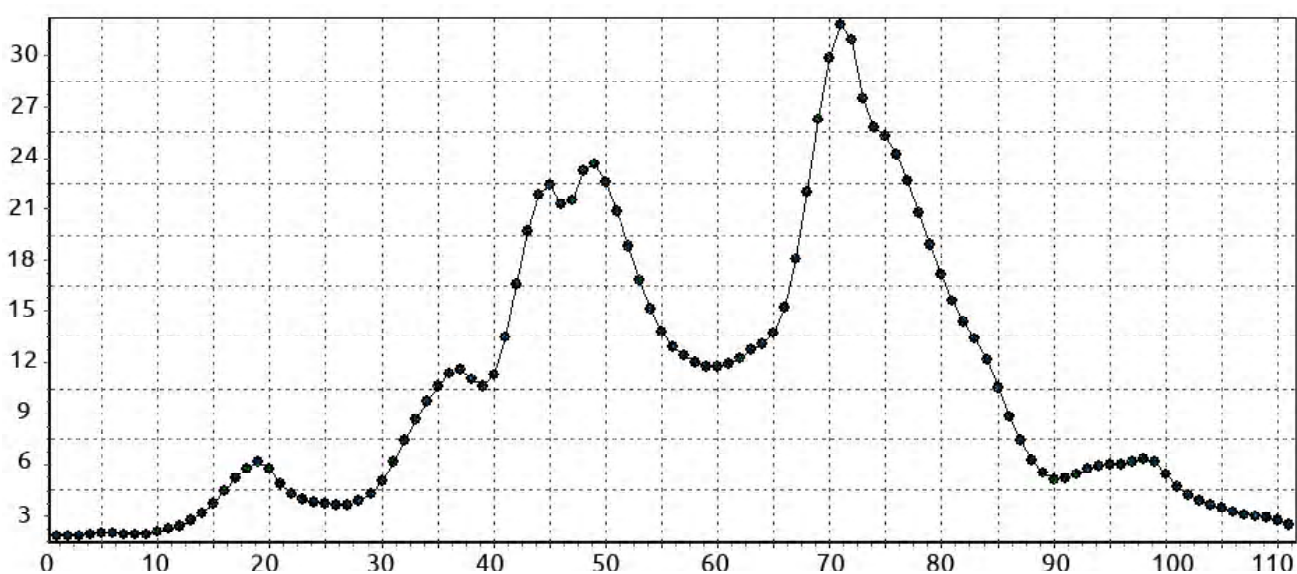


Рис. 10. Распределение z -компоненты электрического поля

и обычного вещества. Получены формулы для коэффициентов прохождения и отражения для плоского слоя метаматериала, расположенного в обычной среде, а также формула для волнового числа плазмона. Показано, что длина волны плазмона может быть значительно короче длины падающей волны.

Также в работе представлены результаты численного моделирования плоской линзы, состоящей из проводящих резонансных элементов в виде спиралей и линейных полуволновых отрезков, а также линзы из массива чередующихся спиралей разнонаправленной намотки, с попарно ортогональным расположением в пространстве. Показано, что такие структуры обладают свойством фокусировки ближнего поля и позволяют преодолеть дифракционный предел. Отмечено, что разрешение линзы определяется периодом решетки, составляющей структуру линзы. Для первой структуры разрешающая способность составляет $\Delta \approx 10$ мм при длине резонансной волны $\lambda = 184$ мм. Для второй структуры разрешение хуже, что обусловлено большим периодом решетки. Результаты моделирования с высокой точностью согласуются с экспериментальными данными, что указывает на адекватность разработанного метода моделирования.

СПИСОК ЛІТЕРАТУРИ

1. Veselago, V. G. The Electrodynamics of Substances with Simultaneously Negative Values of ϵ and μ / V. G. Veselago // Sov. Phys. Usp. – 1968. – vol. 10. – P. 509.
2. Pendry, J. B. Negative Refraction Makes a Perfect Lens / J. B. Pendry // Phys. Rev. Lett. – 2000. – vol. 85. – P. 3966.
3. Lagarkov, F. N. Numerical and experimental investigation of the superresolution in a focusing system based on a plate of «left-handed» material / F. N. Lagarkov, V. N. Kissel // Proc. of the Symp. F, ICMAT. – 2003. – Pp. 157–160.
4. Силаев, М. А. Приложение матриц и графов к анализу СВЧ устройств / М. А. Силаев, С. Ф. Брянцев. – М. : Советское радио, 1970. – 248 с.

Стаття надійшла до редакції 24.03.2011.

Романенко С. М., Карпуков Л. М., Борисенко В. О., Львов О. С.

ПОВЕРХНЕВІ ХВИЛІ НА МЕЖАХ РОЗПОДІЛУ МЕТАСЕРЕДОВИЩ. МОДЕЛЮВАННЯ ЕФЕКТІВ НАДРОЗРІЗНЕННЯ

Теоретично обґрунтовано поява поверхневих хвиль (плазмонів), які виникають на межі розподілу метасередовища та звичайної речовини. Наведені результати чисельного моделювання ефектів надрозрізнення при використанні плоских лінз з метаматеріалу.

Ключові слова: метасередовище, межа розподілу, плазмон, плоска хвиля, надрозрізнення.

Romanenko S. N., Karpukov L. M., Borisenko V. A., Lvov A. C.

SURFACE WAVES AT THE INTERFACE OF METAMEDIALS. MODELING OF SUPERRESOLUTION EFFECTS

Theoretical analysis of the surface waves (plasmons) on the boundary between usual substance and metamaterial and results of numerical modeling of super resolution effects in flat lenses are presented.

Key words: metamedium, interface, plasmon, plane wave, superresolution.

РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ

РАДІОЕЛЕКТРОНИКА ТА ТЕЛЕКОММУНІКАЦИИ

RADIO ELECTRONICS AND TELECOMMUNICATIONS

УДК 004.3

Проскурін М. П.¹, Костенко В. Л.², Щекотихін О. В.¹, Грушко С. С.³¹Канд. техн. наук, доцент Запорізького національного технічного університету²Д-р техн. наук, професор Одеського національного політехнічного університету³Асистент Запорізького національного технічного університету

ОБҐРУНТУВАННЯ НЕОБХІДНОСТІ ПЕРЕХОДУ ДО МІКРОПОТУЖНИХ ТВЕРДОТІЛЬНИХ ІНТЕГРАЛЬНИХ СХЕМ З ОПТОЕЛЕКТРОННИМИ КОМПОНЕНТАМИ ДЛЯ ЦИФРОВИХ АВТОМАТІВ ТА ОБЧИСЛЮВАЛЬНИХ ПРИСТРОЇВ

Розглянуто проблеми передачі даних в металевих напрямних середовищах для систем обробки інформації, керування об'єктами та в цифрових інтегральних схемах. Проведено порівняльний аналіз типів схем електронної логіки та напрямних середовищ. Обґрунтовано можливість використання мікропотужних оптоелектронних елементів для створення оптоелектронної логіки та інтегральних схем на їх основі.

Ключові слова: цифрові інтегральні схеми, оптоелектроніка, оптопара, напрямне середовище, світлодіод, фотоприймач, волоконнооптичний канал зв'язку, цифрові автомати.

ВСТУП

Обробка оптичної інформації набула широкого застосування у техніці, зокрема у приладах і мережах передачі інформаційних потоків, у пристроях управління (ПУ) технологічним і бортовим обладнанням, персональних комп'ютерах (ПК) та цифрових автоматах (ЦА). Це пов'язано з перевагами оптичних і оптоелектронних пристроїв (способів зв'язку на їх основі) над електричними. В основі оптоелектроніки лежать ефекти взаємодії між електромагнітними хвилями (електронейтральними фотонами) і електронами речовин. У пристроях на основі систем випромінювач – фотоприймач (ФП), що з'єднані з волоконнооптичним каналом або кабелем (ВОК), оптичний сигнал від випромінювача без значних втрат проходить великі відстані ($10^2 \dots 10^3$ км) по волоконнооптичних лініях зв'язку (ВОЛЗ). В останні 30 років пристрої,

схеми обробки оптичної цифрової інформації одержали поширення і застосування у різних видах обладнання: системах зв'язку і телекомунікацій, об'єктовому (бортовому, наземному), медицині, військових і цивільних проєктах та ін. [1, 2].

У статті проведено аналіз необхідності переходу до використання оптоелектронних компонентів в інтегральних схемах логіки для цифрових пристроїв і автоматів.

ПОСТАНОВКА ПРОБЛЕМИ В ЗАГАЛЬНОМУ ВИГЛЯДІ

Логічна обробка потоків оптичних цифрових сигналів (ЦС), які передаються по ВОЛЗ, базується на використанні їх відображень у вигляді електричних сигналів (після перетворення типу випромінювання – фотострум:

$L \rightarrow E$ за допомогою фотоприймачів). Для детектування оптичних ЦС використовують фоточутливі пристрої і напівпровідникові (НП) структури: фотодіоди (ФД), фототранзистори (ФТр). Сформовані за допомогою підсилювачів фотоструму у вигляді потоків електричних сигналів, вони обробляються НП цифровими інтегральними схемами (ІС) на базі схем вентилів T^2L , I^2L , E^3L , $T^2LШ$, $МОН$, $МЕН$ та ін. Для одержання вихідних оптичних цифрових сигналів (перетворення типу струм – випромінювання: $E \rightarrow L$) застосовують інші схеми. У них посилені ЦС із виходів ІС модулюють струм випромінювачів на твердотільних НП структурах, до яких відносять інжекційні лазери (ІЛ), лазерні діоди (ЛД) і світлодіоди (СД). Оптичні ЦС передаються на значні відстані по ВОК, що створені на основі оптичнопрозорих діелектричних матеріалів. За допомогою оптоелектронних пристроїв типу оптронів (оптопар), які містять у собі твердотільні випромінювачі і фотоприймачі, забезпечуються численні перетворення типу $L \leftrightarrow E$ [1, 2].

Увага до оптичних і оптоелектронних пристроїв пов'язана з трьома моментами:

1) наявністю обмежень електронних цифрових ІС традиційного типу (табл. 1, рис. 1);

2) перевагами оптичного зв'язку (основні: діапазон частот до 10^{15} Гц, гальванічна розв'язка каскадів перетворення, нечутливість до електромагнітних перешкод, доступність матеріалу (SiO_2) та ін.);

3) зростаючими вимогами до цифрових систем телекомунікацій: збільшення обсягів цифрових потоків (кількісно і якісно) у всіх сферах життя і створення нової інформаційної сфери (знань, відомостей, даних), що доступна більшості бажаючих через Інтернет.

Параметри відомих типів логіки і напрямки розвитку ІС наведені в табл. 1, 2 [3, 4]. Незважаючи на успіхи в освоєнні нано- і субнаносекундного діапазонів, що відповідає частотам $1 \dots 5$ ГГц, розвиток мікро- і наноелектроніки у вигляді цифрових ІС і їх різновидів (схеми логіки $T^2LШ$, I^2L , E^3L , $K-МОН$, $МЕН$ та ін.) характеризується наближенням до кількох видів обмежень: *фізичних* (а),

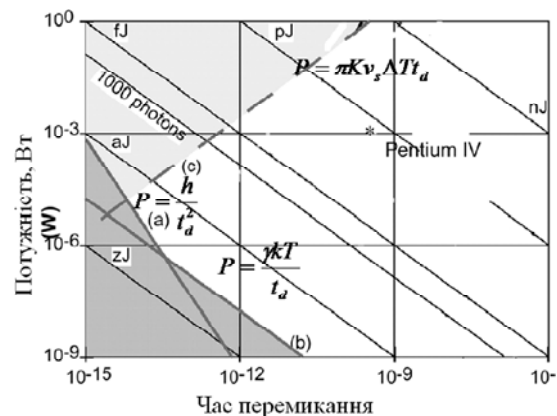


Рис. 1. Графік досягнутих середніх значень параметрів логічних вентилів (біле поле праворуч): потужності ($P_{\text{вик.сер.}}$, Вт), часу ($t_{\text{зт.сер.}}$, нс) і роботи перемикавання ($A_{\text{сер.}}$, пДж) та обмеження властивостей НП матеріалів для Si (сіре поле); * – параметри Pentium IV

конструктивно-технологічних (б), а також *якостей НП матеріалів*, що використовуються (Si, GaAs) (в). Основні з них такі [5, 6]:

1. Використання носія сигналу в ІС у вигляді зарядженої частки (електрона), що обумовлює обов'язкову наявність дводротової лінії (сигнальний – загальний дріт, тобто замкнутість контуру струму) при передаванні сигналу.

2. Властивості напрямних середовищ (НС) для електронів (металеві провідники з параметрами L, R, C , що мають характерний час перемикавання $t_3 \sim RC$ при малому L : між'єднання довжиною $M_{\text{л}} = 10$ мм вносить затримку в поширення сигналу t_3 близько 1 нс).

3. Величина потужності, що споживається, розсіюється в ІС з $N \sim 10^7 \dots 10^8$, $S_{\text{кр.}} \sim 1 \dots 10$ см² (споживання логічного вентиля 0,01...20 мВт при коефіцієнті навантаження $\xi \sim 0,1 \dots 0,3$) становить декілька десятків Вт і більше, що наближається до меж фізичних характеристик НП матеріалів (Si, GaAs) і є обумовленою фундаментальними співвідношеннями.

Таблиця 1. Порівняльні характеристики параметрів логічних вентилів ІС

Тип логіки	Середні значення параметрів логічних елементів (ЛЕ) цифрових ІС			$K_{\text{об.}}$: коефіцієнт об'єднання по входу	$K_{\text{роз.}}$: коефіцієнт розгалуження по виходу
	$P_{\text{сер.}}$, мВт: споживана потужність	$t_{\text{зт.сер.}}$, нс: час перемикавання	$A_{\text{сер.}}$, пДж: робота перемикавання		
$T^2L/T^2LШ$	20/1	5–20/2–10	50–100/10–20	2–5	10
E^3L	20–50	0,5–2	20–50	2–5	10–20
I^2L	0,01–0,1	10–100	0,2–200	1	3–5
МОН	1–10	20–200	50–200	2–5	10–20
КМОН	0,01–0,1	10–50	0,5–5		
ВіКМОН	0,01–0,1	2–10	2–20	2–5	10–100
МЕН	0,1–0,5	0,15–0,5	0,1–0,5	2–5	2–5

Таблиця 2. Основні напрямки розвитку великих ІС

Надвеликі ІС	Мікрозбірки	Надшвидкісні ІС
$N \leq 10^6 - 10^8$	$N_{\Sigma} \geq 10^7 - 10^9$	$N \leq 10^4$
$t_b \geq 50 - 100$ нс		$t_b \geq 0,15 - 2$ нс
$P_b = 0,1 - 0,01$ мВт		$P_b = 0,1 - 20$ мВт

4. Потужність, що витрачається на перезарядження металевої провідної лінії довжиною $M_{\text{лр}}$ сягає 80 % значення потужності, що підводиться до кристала ІС.

Істотно впливають такі характеристики: максимальна кількість зовнішніх виводів K корпусу ІС (визначається розміром площі її кристала $S_{\text{кр}}$); зниження напруги у міжз'єднаннях (досягає значень десятків часток Вольта при значеннях струмів живлення 1...3 А). Це погіршує параметри фронтів імпульсу, знижує завадостійкість ІС, інші її характеристики.

Є питання, пов'язані з електричними зв'язками між кристалами, корпусами ІС, платами через дріотові з'єднання і забезпеченням параметрів передачі ЦС (з урахуванням властивостей з'єднувачів: електричних, частотних, механічних, температурних та ін.):

- у лініях зв'язку, металевих НС всередині і зовні ІС має місце збільшення загасання ЦС від частоти перемикавання (рис. 2) з урахуванням відомої залежності $t_{\text{пер}} \sim k R_{\text{НС}} C_{\text{НС}}$;
- при впливі електромагнітних імпульсів від перемикачів ІС на металевих НС плат виникають електромагнітні перешкоди, амплітуда яких пропорційна довжині лінії $M_{\text{лр}}$;
- низька надійність з'єднувачів, що забезпечують контакт металевих дріотів.

Зазначені фактори можна послабити спеціальними прийомами: зменшенням топологічного розміру, застосуванням знижених значень напруги живлення, струму, оптимізацією трасування і охолодження кристала, застосуванням коаксіальних кабелів, екранування блоків, плат та ін. Але це не вирішує зазначених проблем ІС у цілому, а зі збільшенням рівня інтеграції N , зростанням кількості зв'язків і підвищенням робочих частот ІС вони мають

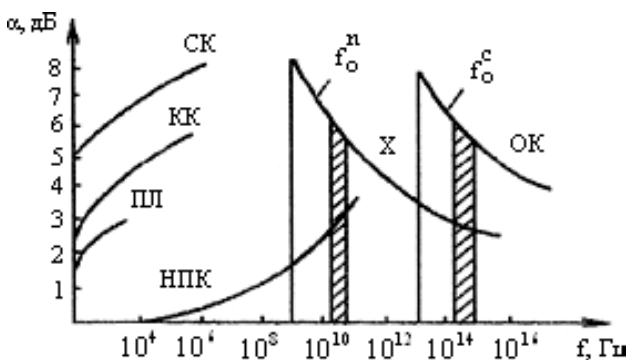


Рис. 2. Частотні залежності загасання напрямних середовищ: КК – коаксіальний кабель, СК – симетричний кабель, ПЛ – повітряна лінія, НПК – надпровідний кабель, ОК – оптичний канал (кабель), X – хвилевід

тенденцію накопичуватися і підсилювати свій негативний вплив.

Наведений аналіз вказує на те, що недоліки технології обробки потоків оптичних ЦС традиційними типами логіки на основі електронних цифрових ІС мають дві складові: використання обраних носіїв сигналів (у вигляді заряджених часток – електронів) і НС у вигляді металевих провідників, доріжок, контактних майданчиків [5, 6].

На погляд авторів, недоліки електронних цифрових ІС можуть бути вирішені через зміну підходів до розробки схем логіки для ІС і ЦА, а саме широким використанням оптоелектронних, оптичних явищ для створення альтернативних схем логічних елементів (ЛЕ) відносно електронних логічних вентилів, що вже існують (табл. 1), і їх впровадження в технології створення «оптичних» і «оптоелектронних» ІС нового покоління.

ПОСТАНОВКА ЦІЛІ ЗАВДАНЬ СТАТТІ ДЛЯ ЇЇ ВИРІШЕННЯ

Рішення вказаних завдань для ІС і ЦА, пристроїв на їх основі можна сформулювати у вигляді цільової функції: що потрібно зробити, щоб, використовуючи вже відомі типи зв'язку, розв'язати поставлені питання. Відповідь є очевидною – необхідна заміна електричного носія (електрона) і металевих НС зв'язком іншої природи, що менш пов'язаний із двома (з трьох) обмеженнями: а) фізичним (обраний вид зв'язку між каскадами в ІС, рис. 2) і б) конструктивно-технологічним (основа перетворення сигналу: тип схем електронної логіки на досягнутому топологічному розмірі, табл. 1), але з урахуванням третього (в) – властивостей НП матеріалів Si, GaAs, що використовуються, – див. рис. 1 (ліве штриховане поле). Це положення має принципове значення, бо з нього виходить, що третє обмеження (в) на даному етапі розвитку технології (рис. 1) ще не є настільки критичними, як а) і б) – тобто має достатньо великий резерв підвищення частоти перемикавання (включно до сірого поля зліва рис. 1).

Можливість обійти означені обмеження ІС відкриває заміна електричного струму (засобу передачі інформації в сучасних ЦА і обчислювальних пристроях – ОП) на сигнали іншої фізичної природи. Переваги оптичного способу зв'язку базуються на використанні електронейтральних фотонів для переносу інформації і властивостях НС, у яких вони поширюються від джерела випромінювання до приймача, та характеру їх взаємодії з електронами твердих речовин [5–7]. Застосування оптичного випромінювання вирішує завдання збору, передачі, перетворення, індикації інформаційних потоків безконтактним (некондуктивним) способом – за допомогою формування інформаційних полів з електрично-непов'язаними ланцюгами джерел і приймачів сигналів (гальванічна розв'язка). Така якість важлива для інформаційних систем і ОП, що перебувають в умовах «забруднення» простору ЕМП (внутрішнього – комутація силових пристроїв, зовнішнього – грозові розряди, радіо-

аційні впливи та ін.) для одержання неспотвореної інформації стану об'єкта, її обробки, видачі сигналів управління. Прилади на основі оптичних і оптоелектронних пристроїв використовуються у зв'язку, системах управління, промисловій автоматизації, техніці [1, 2].

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ ЗА ТЕМАТИКОЮ СТАТТІ

Слід зазначити, що останнім часом з'явилася велика кількість нових розробок удосконалених пристроїв, що використовують оптичні і оптоелектронні прилади (ОЕП), ВОК як робочі елементи для пристроїв виміру і діагностики станів у промисловості, зв'язку, медицині, побуті. У приладах прийому, передачі електричних, оптичних ЦС все більше місце займають оптоелектронні, інтегрально-оптичні компоненти, виконані за планарними технологіями. Оптоелектроніка, інтегральна оптика перетворюються у напрямки, який характеризується багатфункціональністю (при використанні різних фізичних ефектів, їх комбінацій), багатоваріантністю технологічного виконання, способів втілення в прилади і пристрої. На рис. 3, а наведена узагальнена функціональна схема ОЕП [7], а в простішому випадку його можна представити у вигляді оптопари (СД-ФП). На рис. 3, б показана спрощена схема оптоелектронного каналу. Залежно від конкретного застосування, вона подається у вигляді комбінації її частин (ОЕП, канал – ВОК, датчик, модулятор та ін.). Для забезпечення зв'язків у ОП, ЦА і ПК використовують зв'язок по ВОК між кристалами, приладами, компонентами, платами, блоками на основі випромінювачів (ІЛ, ЛД, СД) і ФП на *p-i-n* ФД із робочою частотою f_p до декількох ГГц. Реалізація одного з варіантів міжкомпонентних і міжплатних «вертикальних» зв'язків з ІС на основі СД-ВОП-ФП показана на рис. 4 [8]. У нових розробках матриць регулярних (систематичних) структур типу однорозрядних процесорів, що носять назву «пульсир», також використано подібні оптичні зв'язки між платами для нарощування структури ОП, ЦА за трьома вимірами [9]. Значним досягненням в практичній реалізації оптичних міжкомпонентних зв'язків є створення системної плати для ПК з виключно оптичними зв'язками («горизонтального» типу) російським винахідником [10]. Передавання сиг-

налів оптичним волокном із застосуванням технології WDM (Wavelength Division Multiplexing: мультиплексування по довжині хвилі) дозволяє на 1...3 порядки підвищити продуктивність системи (збільшити пропускну здатність ділянки без прокладки нових кабелів) [1, 2]. Фірмою Bell Labs (корпорація Intel) здійснено передачу 1024 оптичних потоків по одному експериментальному ВОК, створено прототип швидкісного оптичного інтерфейсу вводу/виводу для міжкомпонентних зв'язків кристал – кристал в корпорації Intel [11]. Але найбільш «консервативною» у ОП, ЦА, ПК з використанням потоків оптичних ЦС залишається їх цифрова елементна база обробки інформації і її основа: електронні ІС на відомих типах логіки (табл. 1).

Підсилені і сформовані у вигляді електронних сигналів – відображень оптичних ЦС, вхідні цифрові потоки обробляються НП ІС на базі схем логіки (табл. 1) [3, 4]. З підвищенням продуктивності П (визначається тактовою частотою ОП f_p і розрядністю металевих шин $K = 8...128$ на поверхні кристалу ІС: $\Pi = f_p \times K$), усе більше проявляються обмеження образного інструменту – електронних ІС – при обробці потоків оптичних ЦС. В значній мірі це обумовлює неможливість подолання частотного бар'єра для обраного НС – металевих дротів (доріжок) в ОП і ЦА. Це, зокрема, проявляється в тому, що параметри частоти ядра процесора $f_{яп}$ в ІС сягають значень $\sim 2,5...5$ ГГц, тоді як частота системної шини $f_{сш}$, що зв'язує інші частини (ПК, ОП, ЦА), і має більшу довжину (відповідні значення $R_{НС}, C_{НС}: 1/f_{сш} = t_{пер.} \sim k R_{НС} C_{НС}$) та є приблизно в'ятеро – вдсятеро нижче – $0,5...1$ ГГц. Велика кількість металевих шин потребує використання багатшарових плат, що стає завадою як у технологічному виконанні, так і щодо швидкодії (будь-який дріт має кінцеве значення параметрів R, C, L і, відповідно, кінцеве значення $t_{пер.} \sim k R_{НС} C_{НС}$).

ВИДІЛЕННЯ НЕВИРШЕНИХ РАНІШЕ ЧАСТИН ЗАГАЛЬНОГО ЗАВДАННЯ ТА ПРОПОЗИЦІЇ ЩОДО ЇХ ВИРІШЕННЯ

Огляд досягнень та напрямків досліджень вказує на кілька можливих шляхів вирішення означених проблем

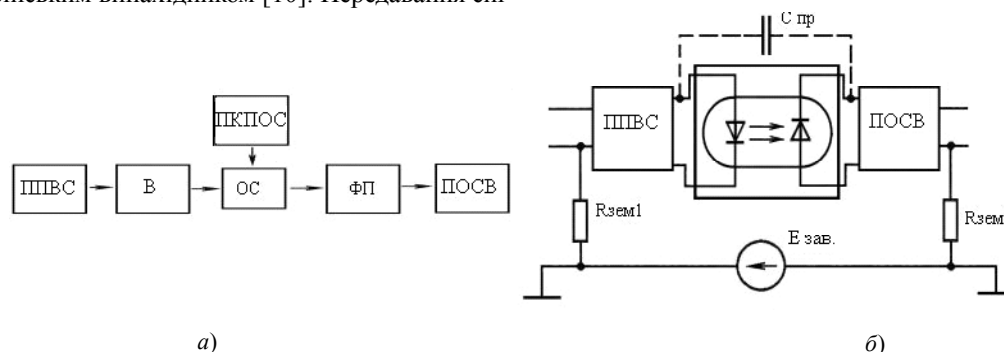


Рис. 3. Узагальнена функціональна схема оптоелектронного пристрою (а); схема оптоелектронного каналу з урахуванням протікання струмів завад крізь паразитні зв'язки (б): ППВС – пристрій перетворення вхідного сигналу; В – випромінювач; ПКПОС – пристрій керування параметрами оптичного середовища; ОС – оптичне середовище; ФП – фотоприймач; ПОСВ – пристрій обробки сигналу на виході; Е зав. – ЕДС наведеної завади

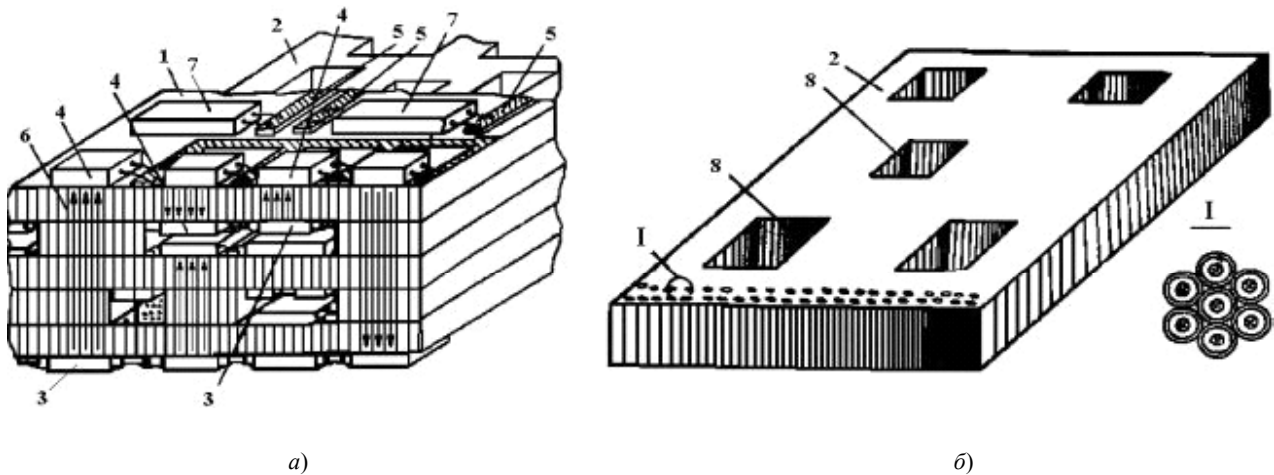


Рис. 4. Реалізація оптичних міжблочних зв'язків на основі пристроїв типу волоконно-оптична пластина (ВОП):

a – конструкція у зборі; *б* – ВОП з ВОК: 1 – плата; 2 – ВОП з ВОК; 3 – блок світловипромінювачів; 4 – блок ФП; 5 – струмопровідні доріжки; 6 – хід оптичних променів; 7 – мікросхема; 8 – вікна для електричних зв'язків; I – збільшений вид зверху на ВОК, що складають ВОП

електронних ІС. Основним напрямком є поступове збільшення використання оптичних компонентів в ІС і організація не тільки входів-виходів у вигляді оптичних зв'язків, але і їх застосування для зв'язку зовні і всередині ІС [1, 2, 5–10]. Тут можливі кілька варіантів розвитку. Перший, що характеризується застосуванням планарних ВОК для більш довгих, ніж в ядрі процесора, зв'язків між каскадами ІС для організації системних шин, вирішується без проведення зміни схем типів логіки (табл. 1). З цим підходом все зрозуміло: концепції звичайних (на макрорівні) сучасних оптичних зв'язків треба втілити на поверхні кристалів ІС за планарними технологіями [2, 5]; між ІС на декількох платах і в блоках ЦА, ОП – наприклад, згідно з технічними рішеннями, що подано в [11].

Другий напрямок, на наш погляд, є більш перспективним і стосується заміни схем логіки, що використовуються для ІС, на альтернативну, наприклад, оптоелектронну або оптичну. Для цього підходу (що передбачає перехід до іншої фізикотопологічної моделі логічного елемента – ЛЕ – і створення, наприклад, робочої схеми оптоелектронного або оптичного інвертора) формування позитивної чи негативної відповіді потребує додаткового аналізу елементів оптопар, їх параметрів. Його мета – дати відповідь на питання про можливість:

- формування елементів оптопар (СД, ФП, ВОК) з розмірами, що досягнуті сучасними технологіями і що можуть функціонувати з параметрами не гіршими за ІС відомих типів логіки;
- вибору базової схеми (або базису) оптичного або оптоелектронного ЛЕ – ОЛЕ;
- досягнення частот перемикання і споживання ОЛЕ на рівні електронних ІС.

Оцінка твердотільних елементів оптопар, їх параметрів і типів (в т.ч. в інтегральному вигляді) достатньо глибоко проведена в [6, 7]. З неї виходить, що швидкодія перемикання СД і ФП може сягати 0,1...1 ГГц, а в особливих режимах перемикання – ще вище. Все це вказує на те,

що частка відомих конструкцій твердотільних елементів оптопар (СД, ФП і ВОК) може бути використана в схемах, наприклад, оптоелектронної логіки для ІС.

До недоліків оптопар СД-ФД відносять мале значення фотоструму, що становить одиниці відсотків від значення струму СД. Для підвищення $K_{\text{фд}}$ в оптопарах СД-ФД використовують вбудовану електронну схему на одному (або декількох) транзисторах, що застосована в оптопарах КОЛ201А (СНД), імпортованих – TLP112, HCPL-0501, HCPL-0700 та ін. [12]. При значенні $K_{\text{фд}} \sim 0,5 \dots 5\%$ для одержання вихідного струму $I_{\text{вих}}$, порядку струму $I_{\text{СД}}$, досить транзистора з коефіцієнтом підсилення $h \geq 21 \dots 210$ (наведено із запасом $\sim 5\%$). До параметрів, що впливають на частотні властивості оптопар, відносять значення ємностей ФП, розмір його області просторового заряду (ОПЗ), щільності струму ($J_{\text{фп}} = I_{\text{фп}}/S_{\text{фп}}$) на робочій частоті f_p фотоперетворення. Наприклад, при модуляції СД інтегральної оптопарі ДВЧ діапазону (0,03...0,3 ГГц) на робочій частоті f_p необхідно, щоб параметри ФП (його час відгуку) відповідали перемикальним параметрам СД. Швидкодіючим ФП є *p-i-n* ФД. На основі таких ФП створені дискретні оптопари в складі СД – *p-i-n* ФД (3ОД120А-1), що відзначаються малими розмірами, високою швидкодією – від десятків нс [12]. Задовільного результату по збільшенню фотоструму досягаються, застосовуючи в оптопарі в якості ФП біполярні ФТр, які сполучають у собі функції ФД і елемента підсилювача-формувача (ПФ) вихідного сигналу. Але за швидкістю вони поступаються оптопарам з окремими *p-i-n* ФД і ВЧ транзисторам приблизно на порядок. На рис. 5, *a-g* наведено відомі схеми і ВАХ оптопар СД-ФТр. Для порівняння на рис. 5, *д* наведено вихідні ВАХ оптопар СД-ФР. Перевагою ВАХ (*г*) перед (*д*) є використання двох режимів: лінійної ділянки і насичення ФТр. В інтегральній оптопарі ДВЧ діапазону частіше застосовують роздільне фотоперетворення (ФП – *p-i-n* ФД з частотою детектування до декількох ГГц) і ПФ: ВЧ транзистор

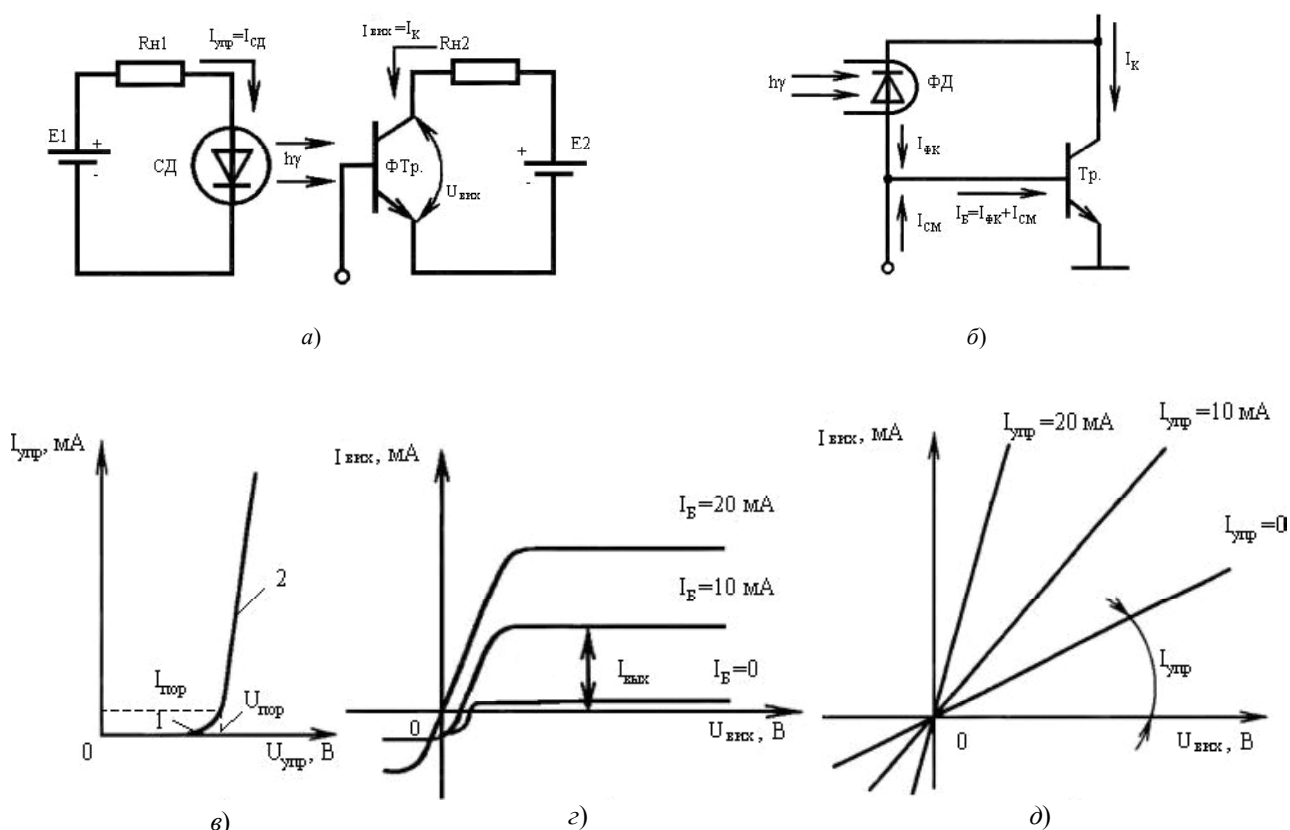


Рис. 5. Електрична схема ОЕП і його ВАХ:

a – схема включення фототранзисторної оптопари (СД-ФТр); *б* – еквівалентна схема *n-p-n* фототранзистора; *в* – вхідна характеристика оптопари (ВАХ СД); *г* – характеристика фототранзисторної оптопари по виходу; *д* – характеристика фоторезисторної оптопари (СД-ФР) по виходу

(Тр) з коефіцієнтом підсилення $h \geq 21 \dots 55$ на частоті f_p [2, 7, 12], що співпадає зі схемою рис. 5, б. Передатні параметри оптопар у режимах малих струмів СД близькі до лінійних, а еквівалентні схеми *n-p-n* ФТр і ФД із *n-p-n* Тр ідентичні [7].

Аналіз параметрів оптопар, елементів ОЕП [6–10], показує, що вони можуть бути використані для створення мікропотужної оптоелектронної елементної бази цифрових пристроїв ДВЧ, УВЧ діапазонів і подолання частини обмежень, що властиві сучасним ІС.

ОСНОВНИЙ МАТЕРІАЛ ДОСЛІДЖЕННЯ З ПОВНИМ ОБҐРУНТУВАННЯМ ОТРИМАНИХ НАУКОВИХ РЕЗУЛЬТАТІВ

Авторами було запропоновано кілька конструкцій структур швидкодіючих елементів мікропотужних оптопар нижньої частини УВЧ діапазону (0,3...3 ГГц) у складі СД-ФП у вигляді малорозмірних їх конструкцій. Інтегральний фотоприймальний пристрій (ІФП) на основі *p-i-n* ФД і ВЧ Тр є схемою з розділним фотоперетворенням (що має максимальну спектральну чутливість при $\lambda \sim 0,8 \dots 0,9$ мкм) і підсиленням фотоструму планарним ВЧ *n-p-n* Тр на Si, спроектованою згідно з вимогами до

ДВЧ приладів [13, 14]. Випромінювач – швидкісний ІЧ СД на сполуках GaAs з поверхневим виходом випромінювання (ПВВ) і довжиною хвилі випромінювання $\lambda \sim 0,8 \dots 0,9$ мкм [15]. Розрахунок структур здійснено: СД (один *p-n* перехід) – на основі аналітичної моделі [16]; ІФП у складі *p-i-n* ФД з ВЧ *n-p-n* Тр (три *p-n* переходи) – методом чисельного моделювання (ПС «Дослідження»), що розроблена за методиками роботи [17].

Частина результатів розрахунків і моделювання параметрів елементів мікропотужної оптопари ДВЧ діапазону (СД – *p-i-n* ФД з ВЧ *n-p-n* транзистором) подана на рис. 6–8.

На рис. 6 подано результати моделювання ІЧ СД ПВВ у вигляді залежностей від напруги ІЧ СД на графіках: щільності оптичної потужності $P_{\text{опт}}$ для двох значень часу життя $t_{\text{нз}}$ основних носіїв заряду (НЗ) (рис. 6, а) і струмів: повного (1) і рекомбінації (2) (при $\lambda = 0,84$ мкм, $I_{\text{СД макс}} = 120$ мкА, $f_{\text{роб}} \sim 0,333$ ГГц) (рис. 6, б). Характерною особливістю наведених даних на рис. 6, а, в є те, що розрахунки проведено для двох значень фізичних параметрів НІ структур: $t_{\text{нз}} = 0,1$ нс (1); $t_{\text{нз}} = 0,2$ нс (2). Значення струму рекомбінації 2 (внутрішній ККД потужності випромінювання) в робочому режимі моделі сягає до 20 % від пов-

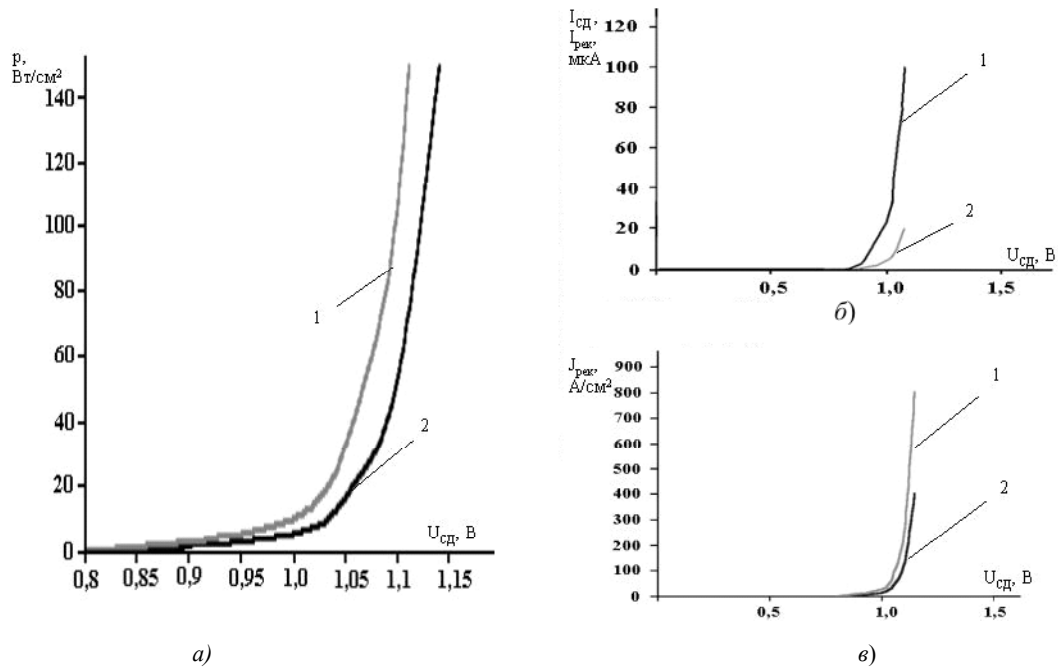
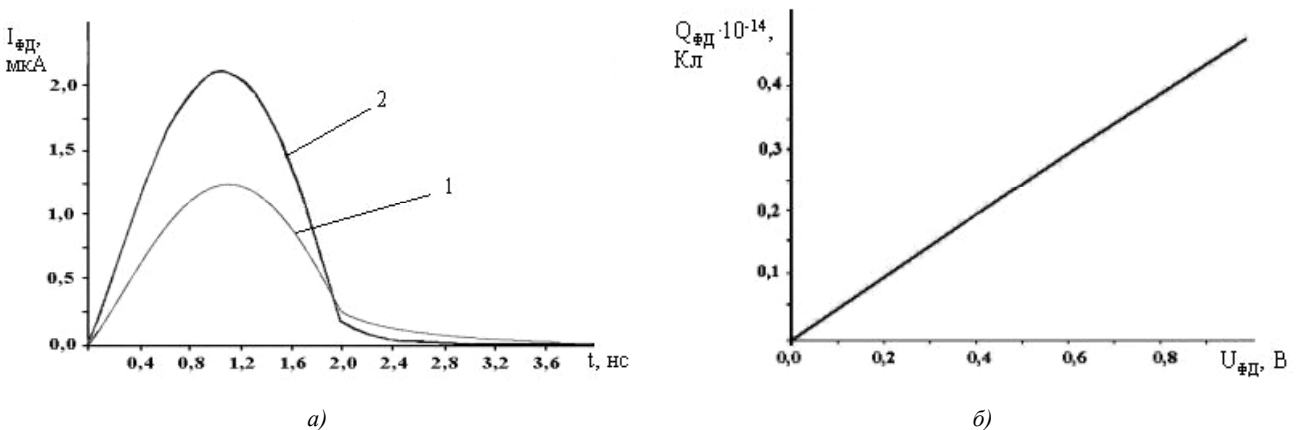


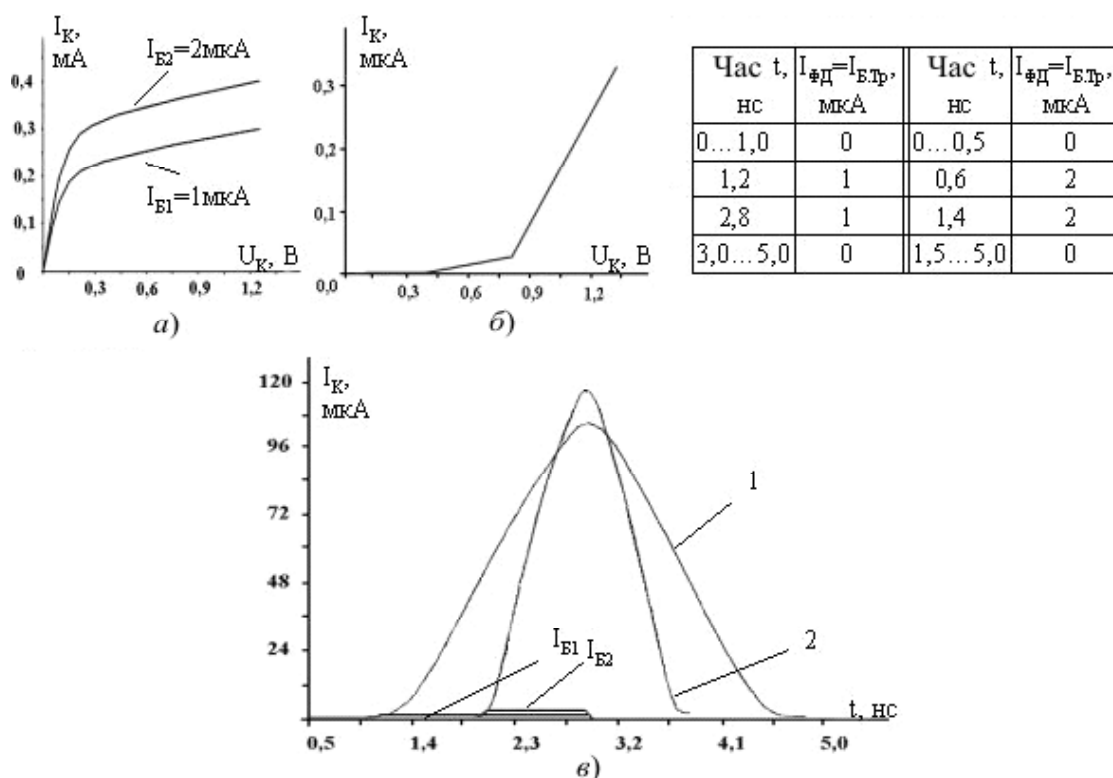
Рис. 6. Частина результатів розрахунків параметрів мікропотужного ІК СД

ного струму інжекції І СД (рис. 6, б). Модуляцію СД запропоновано проводити в режимі «малого сигналу», що забезпечує підвищення частоти перемикавання приблизно на порядок порівняно з режимом «великого сигналу».

На рис. 7 подано результати моделювання ІК $p-i-n$ ФД: фоточутливість при $\lambda = 0,84$ мкм дорівнює $0,45$ А/Вт, гранична частота $f_{\text{гр}} \sim 5$ ГГц, коеф. перетворення до $2 \dots 2,25$ %, приймальне фотовікно $S \sim 25$ мкм². Наведено графіки: значення згенерованого фотоструму ФД (при дії 2 нс синусоїдального оптичного вхідного імпульсу) для двох значень об'ємної швидкості генерації G (рис. 7, а); залежність зміни заряду Q $p-i-n$ ФД від напруги U для розрахунку ємності $C_{\text{ФД}}$, що не перевищує $0,0045$ пФ (рис. 7, б).

На рис. 8 подано результати моделювання підсилювача фотоструму у вигляді ВЧ $n-p-n$ Тр в схемі із загальним емітером: пряма (а) і зворотна (б) ВАХ; значення вихідного струму його колектора (1, 2) для двох трапецієвидних вхідних імпульсів від $p-i-n$ ФД амплітудою $I_{\text{ФД}} = I_{\text{Б1}} = 1$ мкА і $I_{\text{ФД}} = I_{\text{Б2}} = 2$ мкА (в), що задані таблицею. Особливістю їх є те, що з них видно, з якою динамікою перемикавання інтегральний ВЧ $n-p-n$ Тр працює з двома значеннями фотоструму ФД ($I_{\text{ФД}} = I_{\text{Б1}} = 1$ мкА; $I_{\text{ФД}} = I_{\text{Б2}} = 2$ мкА): при значеннях коефіцієнту передачі по струму $h \sim 100$ ($I_{\text{Б1}} = 1$ мкА) і $h \sim 50$ ($I_{\text{Б2}} = 2$ мкА) робоча частота сягає $f_{\text{роб.}} \sim 0,167$ ГГц (1) і $0,333$ ГГц (2) відповідно; розрахункова гранична частота перемикавання (при $h = 1$) $f_{\text{гр}} \sim 2,7$ ГГц.

Рис. 7. Частина результатів розрахунків параметрів ІК $p-i-n$ ФД

Рис. 8. Частина результатів розрахунків параметрів інтегрального ВЧ n - p - n Тр

Результати аналізу, розрахунків і досліджень показали можливість функціонування таких пристроїв, як елементів оптопар при їх відповідних розмірах, в мікропотужних режимах у нижній частині УВЧ діапазону і можливість формування на їх основі ДВЧ, УВЧ оптопар. Параметри часу перемикання $t_{зт.пер.}$ запропонованих елементів оптопар (рис. 6–8) сягають значень електронних типів логіки (табл. 1) при достатньо малих значеннях $P_{сер.}$, мВт; $A_{сер.}$, пДж. Обговорення результатів дослідження проводилось на кількох МНТК [18–21].

ВИСНОВКИ З ПРОВЕДЕНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗРОБОК В ДАНОМУ НАПРЯМІ

На основі проведеного аналізу можна зробити такі висновки.

1. *Динаміка зміни параметрів електронних цифрових ІС* вказує на те, що вони наближаються до певної межі досягнутих параметрів. Вона є різною для відомих типів електронної логіки, пов'язана з мінімальним розміром елемента ІС (топологічний розмір сучасної технології: 1 мкм; ... 100 нм; ... 10 нм), але зв'язана з загальними для ІС факторами, основними з яких є фізичні властивості носіїв інформації, НС і технологічні обмеження.

2. *Фізичні обмеження НП матеріалів і структур*, що входять до складу ІС, ще мають певний запас для підвищення частоти перемикання при заданому значенні $P_{сер.}$ (рис. 1).

3. *Є суттєві переваги ОЕП у складі ОП і ЦА* (для управління об'єктами: передавання, прийом, перетво-

рення інформації в телекомунікаційній техніці та ін.), засновані на використанні фотонів і діелектричних НС і відповідних перевагах оптичного засобу зв'язку.

4. *Аналіз випромінювачів у оптронів, ОЕП* вказує, що прийнятним рішенням для випромінювача у складі оптоелектронних логічних схем є твердотільні пристрої. Серед них виділяються СД – некогерентні випромінювачі на основі GaAs із прийнятним набором характеристик, що відзначаються: малою споживаною потужністю, швидкодією перемикання (10^{-9} с), ККД перетворення електричного струму у випромінювання у межах 10...20 %, значним часом функціонування (10^6 год) і роботою у ВЧ, ДВЧ і УВЧ діапазонах. Є можливість отримати вихід випромінювання як перпендикулярно площині p - n переходу (СД типу ПВВ) так і вздовж неї (більш придатний для планарних технологій отримання ВОК).

5. *Оцінка ФП ДВЧ, УВЧ діапазону і їх параметрів* показує, що структура із роздільним фотоперетворенням і підсиленням типу ІФП (рис. 5, б) забезпечує прийнятну швидкодію на робочій частоті. В пристроях, що існують, схеми ФП сполучаються із входами ІС (на основі Т²ЛШ), не вимагають підвищених напруг живлення (звичайно 1...5 В), мають високу швидкодію – до 10^{-10} с і можливість вибору режиму Тр (ділянки: лінійна або насичення ВАХ) та реалізації адаптивних ФП (через вплив на струм бази Тр).

На основі аналізу можливо зробити припущення, що між етапом розвитку сучасної електронної логіки і майбутньою «оптичною» логікою може перебувати *період застосування оптоелектронної логіки*, наприклад, на

базі використання мало- та мікропотужних режимів елементів оптопар ДВЧ і УВЧ діапазонів у складі оптоелектронних логічних схем (НІ, АБО-НІ, І-НІ, їх комбінацій) для ІС, що можуть сполучати відомі фізичні процеси: оптичні, оптоелектронні і електронні та використовуватимуть їх переваги.

З вищевказаного випливає, що завдання створення мало- і мікропотужних цифрових ІС на елементах оптопар ВЧ, ДВЧ і УВЧ діапазонів (з малим енергоспоживанням для ОП і ЦА) практично назріли. Широке застосування в них елементів оптопар може привести до вирішення частини питань, які вказані в перших частинах статті, а саме: зниження споживаної потужності ІС для ПК, ОП, ЦА (через зниження t_3 при поширенні оптичних ЦС по ВОК) і підвищення широкосмужності оптичних ліній зв'язку і інтерфейсів ІС. Такий підхід відкриє нові можливості розвитку елементної бази ОЕП і використання якостей оптичного середовища для організації міжз'єднань (оптичних інтерфейсів), створення оптоелектронних цифрових ІС для ПК, ОП, ЦА, що можуть конкурувати з відомими типами електронних ІС.

СПИСОК ЛІТЕРАТУРИ

1. Фриман, Р. Волоконно-оптические системы связи / Р. Фриман. – М. : Техносфера, 2003. – 450 с.
2. Волоконные оптические линии связи – ВОЛС: [справочник под ред. С. В. Свечникова, Л. М. Андрушко]. – К. : Техника, 1988. – 239 с.
3. Алексенко, А. Г. Основы микросхемотехники / А. Г. Алексенко. – М. : Юнимедиастиль, 2009. – 448 с.
4. Аванесян, Г. Р. Цифровые интегральные микросхемы : справочник / Г. Р. Аванесян. – М. : Радиотехника, 2008. – 272 с.
5. Свечников, Г. С. Интегральная оптика / Г. С. Свечников. – К. : Наукова думка, 1988. – 166 с.
6. Носов, Ю. Р. Оптоэлектроника / Ю. Р. Носов. – М. : Радио и связь, 1989. – 360 с.
7. Гребнев, А. К. Оптоэлектронные элементы и устройства / А. К. Гребнев, В. Н. Гридин, В. П. Дмитриев. – М. : Радио и связь, 1998. – 336 с.
8. Радиоэлектронный блок с некондуктивными связями : тр. 7-й Междунар. конф. «Волоконная оптика в системах связи, промышленной автоматике, медицине» / В. П. Тесленко, О. В. Щекотихин, Г. А. Полянский, П. А. Ройбул. – Запорожье : ЗНТУ, 2002. – С. 70–72.
9. Шмойлов, В. И. Пульсирующие информационные решетки с матричной коммутацией / В. И. Шмойлов, В. О. Бронзов // Штучний інтелект. – 2008. – № 3. – С. 790–799. – Режим доступу до журн.: www.nbuv.gov.ua/portal/natural/ii/2008_3/JournalAI_2008_3/Razdel8/23_Shmoylev_Bronzov.pdf
10. SciTecLibrary – Новости Науки и Техники. 13.08.2003. Российским специалистом разработана оптоэлектронная системная плата нового поколения, превосходящая даже современные аналоги компании IBM. – Режим доступу до матеріалів: www.sciteclibrary.ru/rus/catalog/pages/5833.htm
11. Корпорация Intel представляет прототип высокоскоростного оптического интерфейса ввода/вывода для межкомпонентных соединений кристалл – кристалл. – Режим доступу до матеріалів: <http://www.intel.com/corporate/europe/emea/rus/country/update/contents/it04041.htm>
12. Юшин, А. М. Оптоэлектронные приборы и их зарубежные аналоги : [справочник]. В 5 т. Т. 1 / А. М. Юшин. – М. : РадиоСофт, 2000. – 512 с.
13. Пат. № 68540А. Україна, МКИ G02F 3/00. Интегральный фотоприёмный пристрой / Белявська О. С., Костенко В. Л., Проскурин М. П. – № 20033076496; заявл. 11.07.2003; опубл. 16.08.2004, Бюл. № 8. – 3 с.
14. Костенко, В. Л. Модели фотоприемников для микроомощных оптопар ВЧ диапазона и оптоэлектронных логических элементов, устройств на их основе / В. Л. Костенко, Е. С. Белявская, Н. П. Проскурин // Нові технології. Науковий вісник Ін-ту економіки і нових технологій. – 2004. – № 3(6). – С. 14–19.
15. Пат. № 12865. Україна, МКИ G02F 3/00. Интегральный светодиод / О. С. Белявська, М. П. Проскурин. – № u 2005 04596; заявл. 17.05.2005; опубл. 15.03.06, Бюл. № 3. – 4 с.
16. Берг, А. Светодиоды / А. Берг, П. Дин ; [пер. с англ. под ред. А. Э. Яновича]. – М. : Мир, 1979. – 688 с.
17. Григоренко, В. П. Моделирование и автоматизация проектирования силовых полупроводниковых приборов / В. П. Григоренко, П. Г. Кузьмин, В. А. Кузьмин, Т. Т. Мнацканов. – М. : Энергоатомиздат, 1988. – 280 с.
18. Костенко, В. Л. Модель интегрального фотоприемника, его параметры и характеристики : труды Междунар. конф. «Сучасний стан та перспективи використання ВОЛЗ. Первинні мережі як транспортна основа телекомунікаційної інфраструктури України» / В. Л. Костенко, Н. П. Проскурин. – Запорожжя : ЗНТУ, 2003. – С. 179–185.
19. Костенко, В. Л. Интегральное фотоприёмное устройство : труды Междунар. конф. «Современные информационные и электронные технологии» / В. Л. Костенко, Н. П. Проскурин, Е. С. Белявская. – Одеса : Од.НПУ, 2003. – С. 193.
20. Proskurin, N. P. The models of photoreceivers optocouples, micropowerful optoelectronic logic elements and devices of HF range : труды 6-й Междунар. конф. «Modern problem of radio Engineering, Telecommunications and Computer Science» / N. P. Proskurin, E. S. Belyavskaya, O. V. Schekotihin. – Львів – Славське : Львівська політехніка, 2004. – С. 549–550.
21. Демиденко, Е. А. Результаты исследований мало- и микроомощных оптоэлектронных логических элементов : труды Междунар. конф. «Сучасні проблеми і досягнення в галузі радіотехніки, телекомунікацій та інформаційних технологій» / Е. А. Демиденко, Н. П. Проскурин. – Запорожжя : ЗНТУ, 2006. – С. 58–60.

Стаття надійшла до редакції 02.09.2010.

Після доробки 03.11.2010.

Проскурин Н. П., Костенко В. Л., Щекотихин О. В., Грушко С. С.

ОБОСНОВАНИЕ НЕОБХОДИМОСТИ ПЕРЕХОДА К МИКРОМОЩНЫМ ТВЕРДОТЕЛЬНЫМ ИНТЕГРАЛЬНЫМ СХЕМАМ С ОПТОЭЛЕКТРОННЫМИ КОМПОНЕНТАМИ ДЛЯ ЦИФРОВЫХ АВТОМАТОВ И ВЫЧИСЛИТЕЛЬНЫХ УСТРОЙСТВ

Рассмотрены проблемы передачи данных в металлических направляющих средах для систем обработки информации, управления объектами в цифровых интегральных схемах. Проведен сравнительный анализ типов схем электронной логики и направляющих сред. Обоснована возможность использования микроомощных оптоэлектронных элементов для создания оптоэлектронной логики и интегральных схем на их основе.

Ключевые слова: цифровые интегральные схемы, оптоэлектроника, оптопара, направляющая среда, светодиод, фотоприемник, волоконнооптический канал связи, цифровые автоматы.

Proskurin N. P., Kostenko V. L., Schekotihin O. V., Grushko S. S.

VALID NECESSITY OF TRANSFER TO MICROPOWER SOLID STATE INTEGRATED CIRCUITS WITH OPTRONIC COMPONENTS FOR DIGITAL MACHINES AND COMPUTING DEVICES

The problems of data transmission in metal directing mediums for the systems of data processing and objects control in digital integrated circuits are considered. Comparative analysis of the types of electronic logic circuits and directing mediums has been carried out. The possibility of using micropower optoelectronic elements for creation of optoelectronic logic and integrated circuits based on it is shown.

Key words: digital integrated circuits, optoelectronics, optocoupler, directing medium, LED, photodetector, fiber-optic communication channel, digital machines.

УДК 621.395

Сметанин И. Н.¹, Ложковский А. Г.², Пиза Д. М.³, Вербанов О. В.⁴

¹ Ст. преподаватель Запорожского национального технического университета

² Д-р техн. наук, профессор, заведующий кафедрой Одесской национальной академии связи им. А. С. Попова

³ Д-р техн. наук, профессор Запорожского национального технического университета

⁴ Ст. преподаватель Одесской национальной академии связи им. А. С. Попова

ИССЛЕДОВАНИЕ ПАРАМЕТРОВ ТРАФИКА УЧАСТКА СЕТИ GSM ПРИ ВОЗДЕЙСТВИИ ЛОКАЛЬНЫХ ПЕРЕГРУЗОК

В данной статье обосновывается необходимость проведения исследований параметров нагрузки для участков сотовой сети связи стандарта GSM, которые подвергаются локальным перегрузкам, приводятся результаты проведенных статистических измерений, и, на основе анализа полученных данных, сделаны выводы об отличии характера потока нагрузки, создаваемого различными видами соединений.

Ключевые слова: сотовая сеть, час наибольшей нагрузки, поток вызовов, длительность разговора, интенсивность нагрузки, число занятий, отказ в обслуживании, локальная перегрузка.

ВВЕДЕНИЕ

С ростом возможностей современных сотовых операторов и количества обслуживаемых ими абонентов, иногда возникают условия, при которых происходит превышение телефонной нагрузки в соте за счет увеличения нагрузки от некоторой части абонентов в этой соте, т. е. возникает локальная перегрузка. Максимальное количество абонентов, собранное в пределах одной соты, можно наблюдать в период проведения массовых мероприятий, таких как День города, концерты и т. д. При этом очень важно качественно обеспечить абонентов сети всеми заявленными услугами. Стабильная и высококачественная работа телекоммуникационной сети зависит от правильного ее проектирования и возможной реконфигурации при ее эксплуатации. Основной исследуемой характеристикой при этом является интенсивность телефонной нагрузки. Причем исходные величины для каждого случая определяются по-своему.

Выбор величины нагрузки при проектировании сети основывается на предполагаемых (прогнозируемых) данных, т. к. точные данные о телефонной нагрузке еще неизвестны. Для внесения изменений при эксплуатации в качестве исходных данных используются уже известные результаты измерений работающей сети. Однако,

и в первом и во втором случаях используются усредненные значения, которые получают при статистических наблюдениях за телефонной нагрузкой. При этом определяемые основные характеристики имеют первичные и вторичные параметры. К первичным можно отнести параметры, которые характеризуют нагрузку, создаваемую потоком вызовов, т. е. среднюю длительность разговора, интенсивность нагрузки и число занятий. В группу вторичных параметров, можно внести количество вызовов абонентов за промежуток времени, длительность паузы между вызовами и время обслуживания вызова. Эта группа параметров может помочь сделать правильный вывод о характеристиках исследуемого потока вызовов и подтвердить либо опровергнуть достоверность предположений о потоковой модели и виде закона распределения случайных величин.

Для этих параметров расчет необходимого объема оборудования, числа линий и каналов может производиться соответствующим методом. Если применить метод, который не будет учитывать реальный поток вызовов, а, следовательно, не отражает их реальную структуру, то точность полученных расчетов будет небольшой, что естественно скажется на качестве обслуживания абонентов и эффективности использования оборудования.

Поэтому возникает необходимость уточнения методики расчета основных характеристик, основанной на модели уже известного потока, с позиции разности между теоретическим и статистическим законами распределения. В результате, непосредственный расчет оборудования основывается на статистических наблюдениях, которые включают в себя измерения первичных параметров, а на измерениях вторичных параметров, основывается оценка модели потока вызовов и выбор либо коррекция соответствующего метода расчета.

Исследования характеристик сетей мобильной связи много раз проводились ранее. Однако они проводились либо для систем стандартов аналоговой транкинговой радиосвязи [1–2], либо исследовалась сеть GSM [2–6], но достаточно давно и не отражает современную реальность, приводящую к локальным перегрузкам на участках сети. Поэтому в условиях возможности локального кратковременного значительного возрастания количества абонентов уточнение методики расчета основных характеристик становится еще более актуальным.

ПРОВЕДЕНИЕ ЭКСПЕРИМЕНТА И ОБРАБОТКА СТАТИСТИЧЕСКИХ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ

Данная статья является результатом проведенной исследовательской работы на сотовой сети оператора мобильной связи стандарта GSM по изучению первичных параметров телефонной нагрузки в современных условиях. При этом исследовался характер всей телефонной нагрузки, обслуживаемой только одним центром мобильной коммутации (MSC). Для более полного отслеживания изменений телефонной нагрузки в сотовой сети, наблюдения проводились в соответствии с Рекомендациями Международного союза электросвязи (ITU-T) в течение 4-х месяцев с учетом концентрации нагрузки [7]. Указанное исследование производилось в г. Запорожье на площади Фестивальной в 2010–2011 годах. Этот район характеризуется как место с высокой плотностью административных, финансовых учреждений, деловых и культурных центров. Также в центральной части площади часто проводятся массовые мероприятия, при проведении которых количество абонентов может превышать пропускную способность соты. Площадь рассматриваемой части равна около $S = 320 \times 275 = 8800 \text{ м}^2$. Вокруг данного участка местности, на различных расстояниях, у оператора мобильной связи стандарта GSM размещено 15 базовых станций (BS), обслуживаемых одним MSC.

Проведение измерений нагрузки осуществлялось в MSC сети путем снятия поступающей учетной информации по сигнализации ОКС-7. Эти данные отражают только ту часть вызовов, которые обслуживаются оператором. Однако, в случаях, когда абонент занят, либо он не отвечает, т.е. когда вызов не завершён тарифицированным разговором абонентов, то эта часть вызовов остается неучтенной. Таким образом, если сравнить

учтенную, т.е. обслуженную телефонную нагрузку, то она окажется меньше реально поступающей.

При возможности локальной перегрузки сети наиболее вероятны случаи получения отказа в обслуживании, когда сеть занята из-за отсутствия свободных каналов, (т.е. заняты все либо зарезервированы для хэндоверных абонентов). В этом случае поступивший вызов вообще останется незафиксированным даже с помощью программного обеспечения системы сигнализации. Для устранения этого недостатка можно предположить, что поступающий поток является простейшим. Тогда необходимо измерить общую длительность промежутков времени, когда заняты все радиоканалы. Теперь, если эту длительность отнести к интервалу времени, в котором проводятся измерения, то таким образом можно рассчитать вероятность потерь по времени P_t^1 . Применительно к простейшему потоку вызовов эта вероятность соответствует вероятности потерь по нагрузке P_n и по вызовам P_v . На основании того, что для простейшего потока потенциальная и поступающая нагрузки также равны, то их величина определяется по полученным значениям фактической обслуженной нагрузки и вероятности потерь из-за отсутствия свободных каналов.

Однако, такой подход правомерен лишь в случае, если поступающий поток абонентских вызовов описывается моделью простейшего потока. Формулы для расчета поступающего потока, который описывается другими моделями, могут быть другими. Так как основные соотношения параметров определены для известных моделей потоков вызовов, то очень важна их точная оценка.

В соответствии с архитектурой сети стандарта GSM MSC выполняет коммутацию между MSC и мобильными либо пользователями внешних сетей, а также управляет работой, связанной с перемещением абонента [8]. Взаимодействие MSC с BS осуществляется через контроллер базовых станций (BSC), по, так называемому, А-интерфейсу. В качестве пользователей внешних сетей к MSC могут быть подключены абоненты стационарной телефонной сети, а также цифровых сетей с интеграцией услуг (ISDN), сети пакетной коммутации (PSDN) и др.²

Из этого следует, что MSC системы стандарта GSM могут обслуживать потоки нагрузки в следующих направлениях связи:

- от пользователей мобильных терминалов к пользователям мобильных терминалов;
- от пользователей мобильных терминалов к пользователям ТФОП;
- от пользователей ТФОП к пользователям мобильных терминалов.

Реконструкция данных по прохождению всех вызовов через MSC в сети определяет множество сценариев устанавливаемых соединений, которые может зафиксировать программа учета нагрузки. Виды соединений, которые были наиболее часто зарегистрированы прог-

¹ Состояние, когда все каналы заняты, отражается в виде «пика» на обобщенной диаграмме нагрузки.

² Состояние, когда все каналы заняты, отражается в виде «пика» на обобщенной диаграмме нагрузки.

раммой, при обслуживании вызовов в MSC, и, являющиеся определяющими при формировании трафика, показаны на рис. 1. Остальные виды связи, (например, переадресация вызова на голосовую почту, звонок в Call-центр обслуживающего оператора сотовой связи) в силу своего крайне редкого использования и, не оказывающие значительного влияния на формирование нагрузки, не рассматривались.

При регистрации данных система сигнализации ОКС-7 позволяет учитывать номера BS, через которые осуществляются вызовы. Такая информация позволяет более

детально исследовать нагрузку на радиоканалы в период локальной перегрузки участка сети.

Все результаты измерений обрабатывались специально разработанной на языке Visual Basic 6 программой и в дальнейшем передавались в табличный редактор Excel пакета MS Office 2007 для построения диаграмм.

В качестве исследуемой BS по результатам полученных данных была выбрана BS-A1, которая работает с перегрузкой во время проведения массовых мероприятий (рис. 2) и исследована ее загруженность в течение 10 последовательных рабочих дней различных периодов года с учетом максимальной концентрации нагрузки³[9].

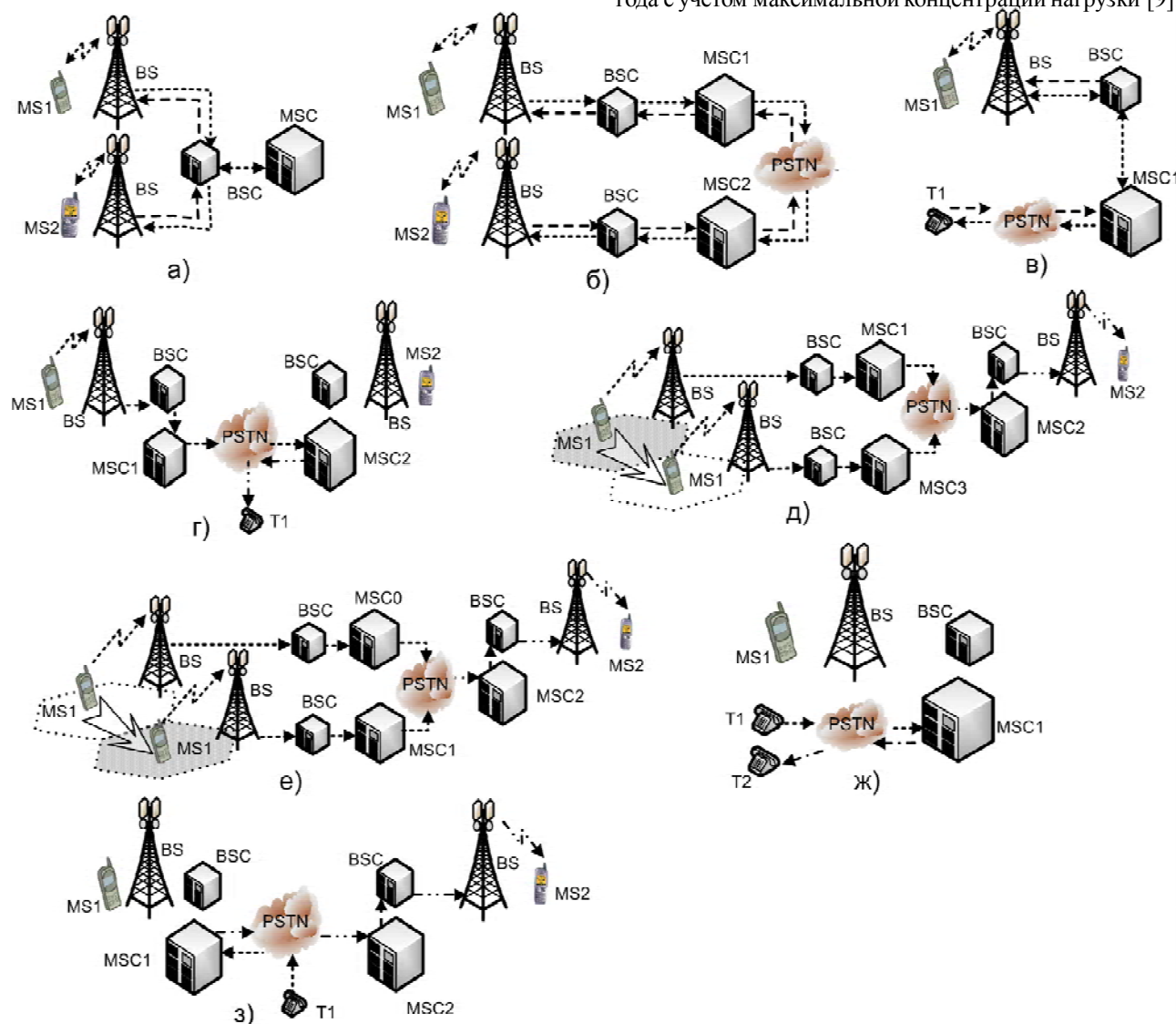


Рис. 1. Основные виды соединений, которые определяют нагрузку в MSC:

а) с MS1 на MS2 либо с MS2 на MS1 (оба абонента находятся в зоне обслуживания одного MSC); б) с MS1 на MS2 либо с MS2 на MS1 (абонент 2 находится в зоне обслуживания другого MSC при соединении MSC через ТфОП); в) с MS1 на терминал Т1 ТфОП либо с терминала Т1 ТфОП на MS1; г) с MS1 на MS2 (абонент 2 установил переадресацию вызова на терминал Т1 ТфОП); д) вызов с MS1 на MS2 с дальнейшим выходным хэндовером (абонент 1 осуществляет вызов абонента 2 из текущей ячейки, а затем перемещается в другую соту и при этом продолжает вести разговор); е) входящий хэндоверный разговор с MS1 на MS2 (когда абонент 1 перемещается в ячейку и продолжает вести разговор, который начал в предыдущей ячейке); ж) с терминала Т1 ТфОП на MS1 (абонент 1 установил переадресацию вызова на другой MS); з) с терминала Т1 ТфОП на MS1 (абонент 1 установил переадресацию вызова на терминал Т2 ТфОП)

³ Наибольшая нагрузка имеет место в феврале, марте, ноябре и декабре – поэтому целесообразно проводить измерения нагрузки именно в эти месяцы [9] гл.3 п.3.2.

На рис. 3 приведенные суточные диаграммы распределения нагрузки по 15-ти минутным интервалам. Полученные результаты исследований сезонной ежедневной загруженности BS показали (рис. 3), что нагрузка для повседневной деятельности в будние дни довольно стабильна и, в зависимости от месяца года, она изменяется незначительно. При этом на декабрь, приходится больше всего разговоров, чем в другие месяцы, а в феврале и марте – нагрузка почти одинакова. Превышение нагрузки в декабре может быть объяснено повышенной активностью абонентов в конце года.

Почасовое суточное распределение нагрузки для рабочих дней, вполне отражает повседневный распорядок жизни. Из обобщенной диаграммы для всех видов связи на рис. 4 видно, как, начиная с 6:00 до 7:30, нагрузка возрастает плавно. Потом до 9:00 нагрузка увеличивается более интенсивно. Далее она снова плавно возрастает до 12:30. С 13:00 до 13:30 наблюдается некоторый спад активности абонентов, что может соответствовать времени обеденного перерыва.

Затем нагрузка нарастает и указывает на явно выраженный час наибольшей нагрузки (ЧНН), который для

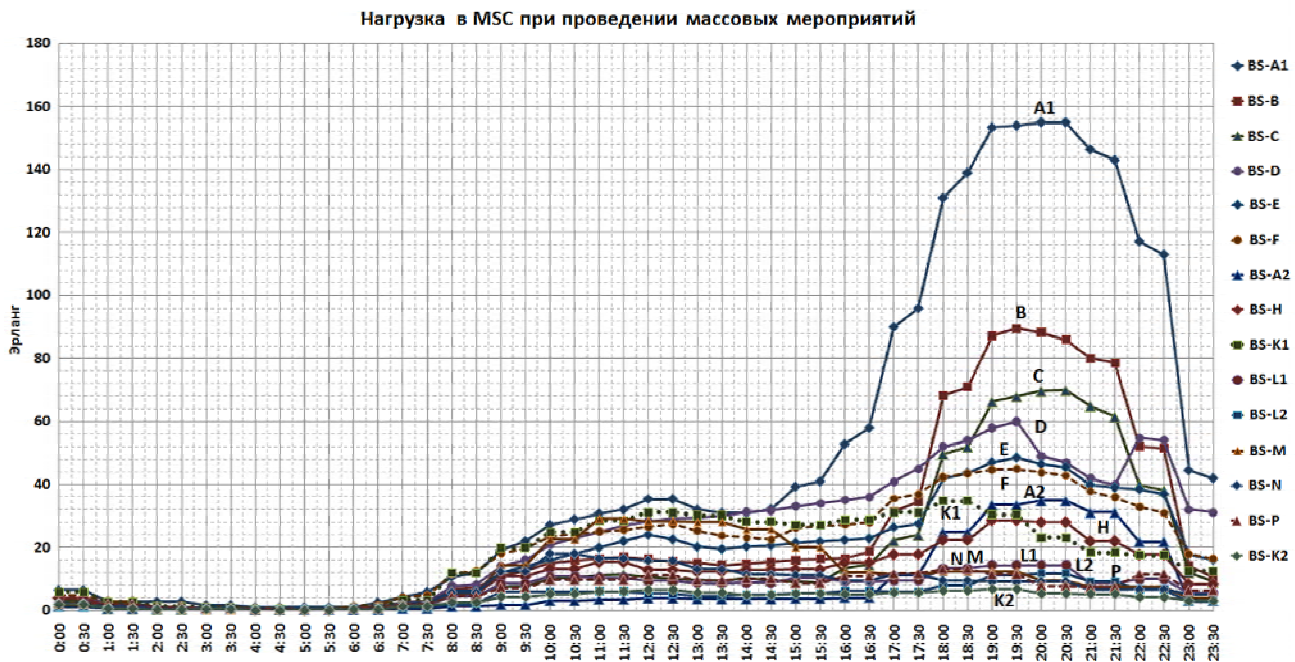


Рис. 2. Распределение нагрузки между BS одного MSC в период проведения массового мероприятия

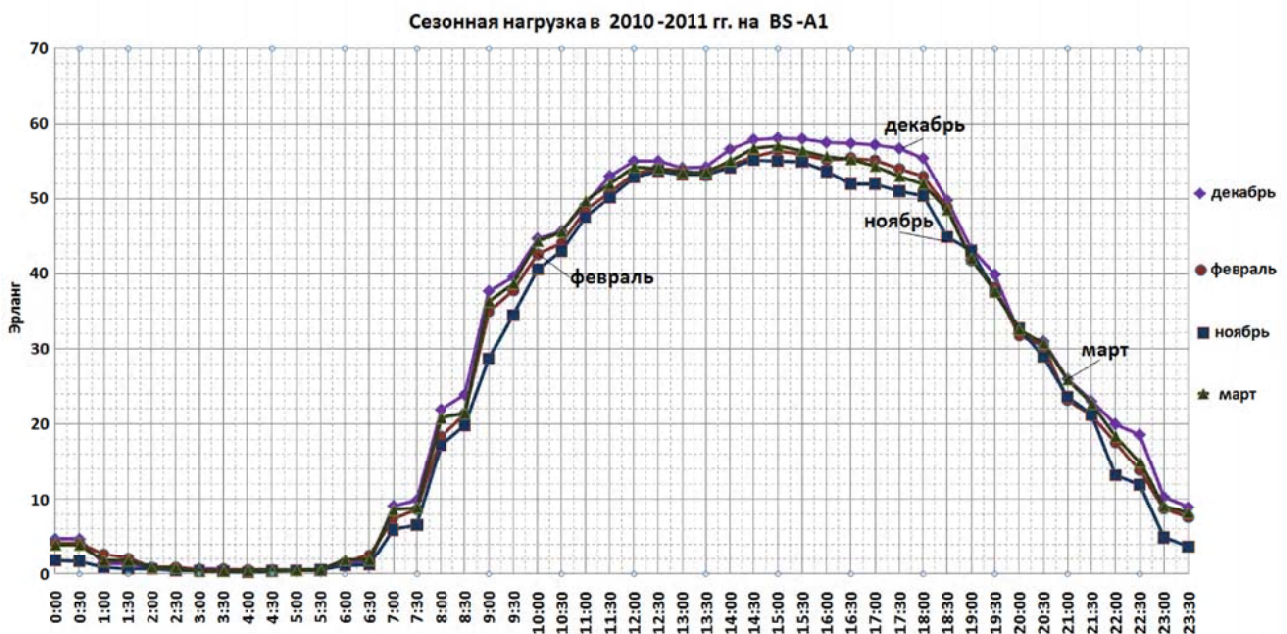


Рис. 3. Суточное распределение ежедневной нагрузки на исследуемой BS по периодам года

всех видов связи приходится на период с 14:30 до 15:30 и его значение достигает величины 56,6 Эрл. (Обобщенные данные распределения нагрузки в ЧНН сведены в табл. 1). В дальнейшем обобщенная нагрузка несколько снижается, а по окончании рабочего дня с 18:00, стабильно убывает.

Анализ результатов исследования нагрузки по направлениям связи показал, что основную загруженность почти в равных долях создают входящие (47,6 % от всей) и исходящие (47,2 %) вызовы. И лишь минимальную часть (1,7 %) от общей нагрузки составляют транзитные для MSC вызовы.

Более детальный анализ суточного распределения ежедневной нагрузки по видам соединений показывает, что нагрузки всех видов связи достаточно равномерно распределены в течение рабочего дня и вечернего времени (рис. 5). Наибольшую долю нагрузки в ЧНН (70 %) образуют соединения, осуществляемые между мобильными терминалами, которые обслуживаются только одним MSC (38,2 %) и несколькими MSC (31,8 %). Так, входящие соединения с MS2 на MS1 в пределах одного MSC (рис. 1, а) в ЧНН с 203 вызовами создают нагрузку в 11,38 Эрл при средней длительности занятия $T = 201,8$ сек., а исходящие с MS1 на MS2 в пределах одного MSC со 191 вызовом создают нагрузку в 11,04 Эрл при средней длительности занятия $T = 208,1$ сек. При этом входящих через несколько MSC соединений на MS1 (рис. 1, б) больше на 1,37 Эрл., чем исходящих от MS1 через ТфОП на MS2. Средняя длительность занятия для перечисленных видов соединений соответствует наличию на территории обслуживаемой BS государственных и деловых организаций, для которых приемлемым можно считать интервал от 120 до 240 секунд [10]⁴.

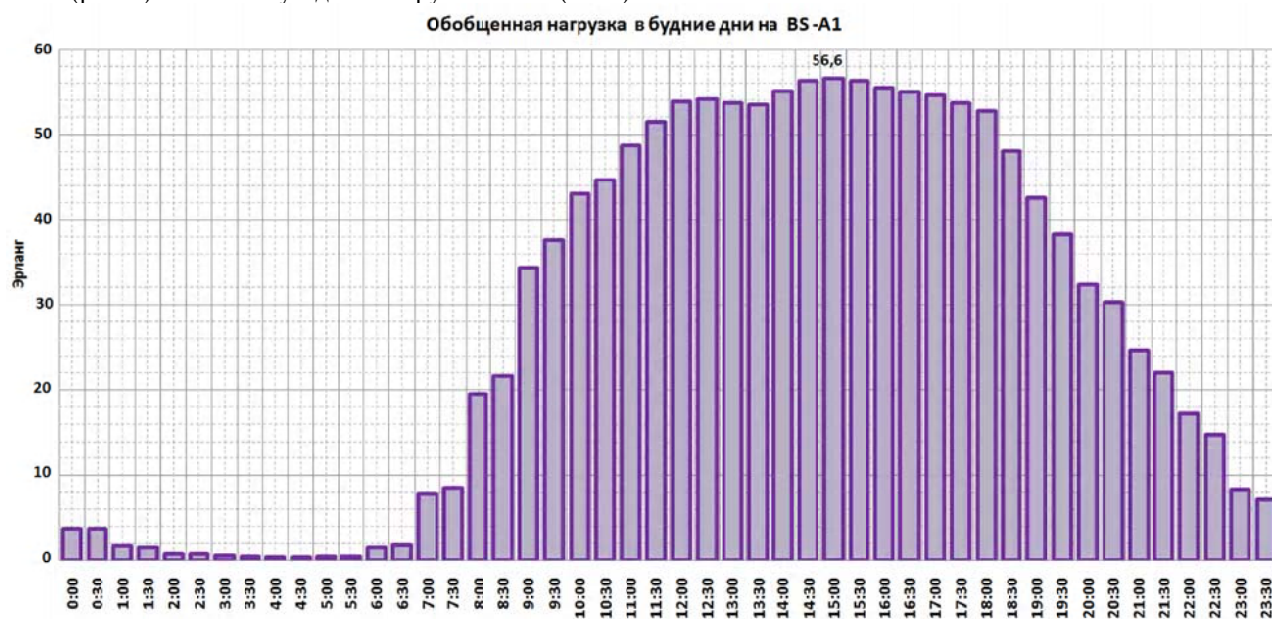


Рис. 4. Обобщенное суточное распределение ежедневной нагрузки BS-A1

Таблица 1. Распределение нагрузки в ЧНН по направлениям и видам связи

Вид нагрузки	исходящая					входящая				транзитная	
Вид соединения	MS1–MS2 один MSC	MS1–MS2 через ТфОП	MS1 на Т1 ТфОП	с MS1 на MS2 переадресация на Т1 ТфОП	хэндовер MS1 на MS2 на другой MSC	MS2–MS1 один MSC	MS2–MS1 через ТфОП	Т1 ТфОП на MS1	хэндовер MS1 на MS2 с другого MSC	с Т1 ТфОП на MS1 переадресация на MS2	Т1 ТфОП на MS1 переадресация Т2 ТфОП
Позиция на рис. 5	а)	б)	в)	г)	д)	е)	ж)	з)	и)	к)	л)
Число вызовов	191	152	59	6	125	203	179	55	112	13	19
Средняя длительность, с	208,1	204,2	100,2	93,6	179	201,8	201	199,64	112,37	108,4	118,7
Нагрузка, Эрл	11,04	8,62	1,64	0,16	6,22	11,38	9,99	3,05	3,48	0,39	0,63
Нагрузка в % к общей	18,8	14,7	2,8	0,3	10,6	19,4	17,0	5,2	6,0	0,7	1,1
Нагрузка по направлениям, Эрл	27,68					27,93				1,02	
Нагрузка по направлениям, в % к общей	47,2					47,6				1,7	

⁴ [10] Приложение А, рисунок А.

Превышение хэндоверных исходящих (рис. 1, *д*) разговоров над входящими (рис. 1, *е*) составляет 1,78 раза. Входящие вызовы с терминала T1 ТфОП на MS1 (рис. 1, *е*) в период ЧНН достигают величины 3,05 Эрл., а к окончанию рабочего дня значительно уменьшается и к 18:30 не превышает 0,91 Эрл., при возросшем до 67 по сравнению с ЧНН количеством вызовов.

Почти 35-ю часть от общей по всем видам связи нагрузки составляют исходящие разговоры с MS1 на терминалы ТфОП (рис. 1, *е*) с показателем 1,64 Эрл. в ЧНН.

Остальные виды соединений не превышают 0,63 Эрл.

Очевидно, что дополнительно требуют отдельных исследований работы BS с перегрузкой, т.к. в эти периоды нагрузка возрастает в 3 раза (достигая во время пика проведения массовых мероприятий 155 Эрл. рис. 2), что имеет очень важное значение для качества обслуживания абонентов.

ОЦЕНКА СТЕПЕНИ ДОСТОВЕРНОСТИ СТАТИСТИЧЕСКИХ ДАННЫХ

В соответствии с Рекомендациями ITU-T высокая точность измерений нагрузки на сети достигается проведением достаточно частых измерений. Поскольку в данном исследовании измерения производились выборочно, то с целью учета реального колебания нагрузки результаты этих измерений подверглись статистической обработке и при этом оценена степень точности полученных данных. Для этого были определены среднее значение ($\bar{x}$) или математическое ожидание (M_x) измеряемой случайной величины (x) [11]

$$M_x \cong \bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad (1)$$

где n – объем выборки.

Также был определен разброс отдельных значений случайной величины относительно среднего значения, оцениваемый величиной среднеквадратического отклонения (δ). Именно по величине δ на основе теоремы Чебышева – Ляпунова можно оценить точность произведенных статистических выборочных наблюдений.

Очевидно, что чем больше проведено измерений, тем ближе к средним значениям основной совокупности, будут располагаться средние значения выборки. Учитывая, что данные измерения были выборочны, то ошибка выборки будет меньше тогда, когда меньшим будет среднеквадратическое отклонение, т. е. в меньшей степени будет колебаться изучаемый параметр.

Так как в данном случае используется среднеквадратическое отклонение не сплошного, а выборочного наблюдения, то приближенно величину средней ошибки выборки можно определить из соотношения [12]:

$$\mu_x \approx \sqrt{\frac{\delta^2}{n}}, \quad (2)$$

где δ – среднеквадратическое отклонение.

В соответствии с теоремой Чебышева-Ляпунова может быть рассчитано значение предельной ошибки выборки (Δ). Эта величина, с определенной долей вероятности, отражает то, что отклонения выборки не превысят заданного значения. Численно определить предельную ошибку выборки можно из соотношения [12]:

$$\Delta_x = z\mu_x = \frac{z\delta}{\sqrt{n}}, \quad (3)$$

здесь z – поправочный коэффициент, который зависит от величины вероятности, с которой гарантируется определенное значение предельной ошибки. Так, например, для вероятности $P(z) = 0,99$ коэффициент $z = 2,6$. С возрастанием z величина вероятности $P(z)$ приближается к единице.

Приведенное соотношение свидетельствует о том, что уменьшение отклонения выборочной средней от основной генеральной средней (т. е. уменьшение предельной ошибки измерений Δ), возможно при уменьшении среднеквадратического отклонения δ либо с увеличением объема произведенной выборки n . Исходя из этого следует: чтобы обеспечить удовлетворительное значение отклонения результатов выборочных измерений от результатов сплошного наблюдения, достаточно меньшего объема выборки n , который уменьшается вместе со среднеквадратическим отклонением. Это стало основой определения предельной ошибки статистических данных.

В результате, для расчета предельной ошибки регистрации статистических данных о нагрузке и количестве вызовов со всех видов и направлений связи с вероятностью $P(z) = 0,99$ оказалась необходимой и достаточной выборка в ЧНН ежедневно четырех 15-ти минутных интервалов в течение 10 рабочих дней, что соответствует 40 замерам. Однако, такой же подход оказывается не приемлемым для расчетов по длительности занятия, т. к. в ЧНН объектом исследования является не средняя длительность занятия на указанных временных интервалах, а каждый отдельный вызов. Поэтому определение предельной ошибки длительности занятия от всех видов вызовов требует большего объема выборки.

Для примера результаты погрешностей расчетов для видов соединений, определяющих основную нагрузку, приведены в табл. 2.

Достоверность полученных результатов подтверждает относительная погрешность расчетов предельной ошибки измерений, которая определена с вероятностью $P(z) = 0,99$ соотношением:

$$\beta = \frac{\Delta_x}{M_x}, \quad (4)$$

и величина которой, ни в одном из направлений и видов связи не превысила 5 %.

Таблица 2. Результаты расчетов погрешности для определяющих видов соединений в ЧНН

Направление	Вид соединения	Параметр	Среднее значение, M_x	Отклонение, δ	Объем выборки, n	Предельная ошибка, Δ	Относительная ошибка, в %
исходящая	MS1– MS2 один MSC	Нагрузка по 15 мин. интервалам, (Эрл.)	11,04	1,19	40	0,49	4,4%
		Число вызовов по 15 мин. интервалам	47,75	5,47	40	2,25	4,7%
		Длительность занятия, (сек)	208,14	240,88	9900	6,29	3,0%
	MS1–MS2 через ТфОП	Нагрузка по 15 мин. интервалам, Эрл	8,62	1,01	40	0,42	4,8%
		Число вызовов по 15 мин. интервалам	38	4,47	40	1,84	4,8%
		Длительность занятия, сек	204,2	238,88	9900	6,24	3,1%
	хэндоверный MS1 на MS2 на другой MSC	Нагрузка по 15 мин. интервалам, Эрл	6,22	0,7	40	0,29	4,6%
		Число вызовов по 15 мин. интервалам	31,25	3,43	40	1,41	4,5%
		Длительность занятия, сек	179	198,96	9900	5,20	2,9%
входящая	MS2– MS1 один MSC	Нагрузка по 15 мин. интервалам, Эрл	11,38	1,28	40	0,53	4,6%
		Число вызовов по 15 мин. интервалам	50,75	5,49	40	2,26	4,4%
		Длительность занятия, сек	201,8	205,87	9900	5,38	2,7%
	MS2–MS1 через ТфОП	Нагрузка по 15 мин. интервалам, Эрл	9,99	1,16	40	0,48	4,8%
		Число вызовов по 15 мин. интервалам	44,75	5,24	40	2,15	4,8%
		Длительность занятия, сек	201	210,62	9900	5,50	2,7%
	Хэндоверный MS1 на MS2 с другого MSC	Нагрузка по 15 мин. интервалам, Эрл	3,48	0,41	40	0,17	4,8%
		Число вызовов по 15 мин. интервалам	28	3,12	40	1,28	4,6%
		Длительность занятия, сек	112,37	127,88	9900	3,34	3,0%

ВЫВОДЫ

Обработанные результаты измерений, после проверки погрешности и всестороннего анализа позволяют сделать определенные выводы:

1. В отличие от более ранних исследований, в которых по типам направлений входящая нагрузка преобладала над исходящей, нынешняя входящая нагрузка к мобильным абонентам (47,2 %) практически такая же, как и исходящая от них (47,6 %), и эти два типа составляют почти всю нагрузку. Доля транзитной нагрузки в общей сравнительно мала (1,7 %).

2. Развитие и улучшение работы сетей сотовой связи стандарта GSM и изменения в тарифной политике операторов мобильной связи, в районах концентрации государственных учреждений, деловых и культурных центров, привели к увеличению длительности ведения разговоров почти до 3,5÷4 минут. При этом для связи между мобильными абонентами длительность входящих (201÷201,8 секунды) и исходящих (204,2÷208,1 секунды) вызовов почти одинакова.

3. Также отразившаяся на тарифной политике и конкурентная борьба за абонентов, привела к значительно уменьшению количества и длительности разговоров между пользователями мобильной и фиксированной связи. Таким образом, появилась значительная разница в длительности ведения разговора (почти 100 секунд) между исходящими вызовами мобильных абонентов к абонентам фиксированной связи (93,6 секунды при количестве вызовов 6) и входящими от абонентов фикси-

рованной связи к абонентам стандарта GSM (199,64 секунды с числом вызовов 55).

СПИСОК ЛИТЕРАТУРЫ

1. Barcelly, F. Idle and Inter-arrival Time Statistics in Public Access Mobile Radio (PAMR) Systems / F. Barcelly, S. Bueno // Proc of the IEEE GLOBECOM 97. –Nov. 98. –Pp. 126–130.
2. Jordan, J. Statistical Modelling of Channel Occupancy in Trunked PAMR Systems / J. Jordan, F. Barcelo // 15th International Teletraffic Congress (ITC15) V. Ramaswami and P.E. Wirth (Editors), Elsevier Science B. V., June 1997. – Pp. 1169–1178.
3. Khedher, H. Traffic Characterization for Mobile Networks / H. Khedher, S. Tabbane, F. Valois // In Vehicular Technology Conference (VTC'2002-Fall 2002), September 2002. – Vancouver, British Columbia, Canada. – Pp. 1485–1489.
4. Boggia, G. Modelling of call dropping in well-established cellular networks In EURASIP Journal on Wireless Communications and Networking. [Электронный ресурс] / G. Boggia, P. Camarda, A. D'Alconzo. – Vol. 2007 Article ID 17826, 11 pages, 2007. doi:10.1155/2007/17826 – Режим доступа к журн. <http://www.hindawi.com/journals/wcn/2007/017826/abs/>
5. Bregni, S. WLC09-1 : An Empirical Study on Statistical Properties of GSM Telephone Call Arrivals / S. Bregni, R. Cioffi, M. Decina // In Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '06). – San Francisco, Calif, USA, November-December 2006. – Pp. 1–5.
6. Bregni, S. An Empirical Study on Time-Correlation of GSM Telephone Traffic / S. Bregni, R. Cioffi, M. Decina // Wireless Communications, IEEE Transactions on. – September 2008. – vol. 7, no. 9. – Pp. 3428–3435.

7. ITU-T : Recommendation No. E.500 (11/98) : Traffic intensity Measurement Principles. – Geneva, 1998. – 16 p.
8. Берлин, А. Н. Цифровые сотовые системы связи / Берлин А. Н. – М. : Эко-Трендз, 2007. – 296 с.
9. Лившиц, Б. С. Теория телетрафика: Учебник для институтов связи / Б. С. Лившиц, А. П. Пшеничников, А. Д. Харкевич. – 2-е изд., перераб. и доп. – М. : Связь, 1979. – 224 с.
10. Системы измерения длительности соединений СИДС подсистем коммутации MSC и IN-платформ сотовой подвижной связи стандарта GSM 900/1800. Методика поверки : МИ 2994-2006 ГСИ – Офиц. изд. – М. : ФГУП «ВНИИФТРИ», 2006. – 18 с. – (Нормативный документ Комитета Российской Федерации по стандартизации, метрологии и сертификации).
11. Гмурман, В. Е. Теория вероятностей и математическая статистика : учеб. пособие для вузов / Гмурман В. Е. – 9-е изд., стер. – М. : Высш. шк., 2003. – 479 с.
12. Максимов, Г. З. Телефонная нагрузка местных сетей связи. учеб. пособие / Г. З. Максимов, А. П. Пшеничников. – М. : Связь, 1969. – 152 с.

Стаття надійшла до редакції 20.05.2011.

Сметанін І. М., Ложковський А. Г., Піза Д. М., Вербанов О. В.

ДОСЛІДЖЕННЯ ПАРАМЕТРІВ ТРАФІКУ ДІЛЯНКИ МЕРЕЖІ GSM ПІД ВПЛИВОМ ЛОКАЛЬНИХ ПЕРЕВАНТАЖЕНЬ

У даній статті обґрунтовується необхідність проведення досліджень параметрів навантаження для ділянок стільникової мережі зв'язку стандарту GSM, які піддаються локальним перевантаженням, наводяться результати проведених статистичних вимірів, і, на основі аналізу отриманих даних, зроблені висновки про відмінність характеру потоку навантаження, який створюється різними видами з'єднань.

Ключові слова: стільникова мережа, година найбільшого навантаження, потік викликів, тривалість розмови, інтенсивність навантаження, число занять, відмова в обслуговуванні, локальне перевантаження.

Smetanin I. N., Lozhkovsky A. G., Pisa D. M., Verbanov O. V.

RESEARCH OF PARAMETERS OF THE TRAFFIC OF A SECTION OF NETWORK GSM AT INFLUENCE OF LOCAL OVERLOADS

In given article necessity of carrying out of researches of parameters of loading for sections of a cellular network of communication of standard GSM which are exposed to local overloads is justified, results of the led statistical measurements, and, on the basis of the analysis of data retrieved are resulted, outputs are drawn on difference of character of a flow of the loading created by various types of connections.

Keywords: a cellular network, busy hour, a flow of calls, a conversation time, intensity of loading, number of occupations, a failure in service, a local overload.

УДК 528.88.042.4

Морщавка С. В.¹, Шама Є. О.², Піза Д. М.³

¹Канд. техн. наук, доцент Запорізького національного технічного університету

²Аспірант Запорізького національного технічного університету

³Д-р техн. наук, професор Запорізького національного технічного університету

ВПЛИВ ВИДІВ НОРМУВАННЯ НА ЯКІСТЬ РОЗПІЗНАННЯ РОСЛИННИХ ОБ'ЄКТІВ

Розглянуто вплив різних видів нормування на якість розпізнання за результатами дистанційного зондування.

Ключові слова: розпізнавання, хеометрика, нормування, спектральні коефіцієнти яскравості.

ВСТУП

При розпізнанні рослинних об'єктів, необхідно проводити аналіз багатоспектральних потоків даних сукупності спектральних коефіцієнтів яскравості, які описуються деяким масивом чисел [1]. На практиці під час експериментів зручно складати матрицю, в якій стовпці відповідають довжинам хвиль, а рядки вибіркам – значенням спектральних коефіцієнтів яскравості (СКЯ), що отримані з виходу аналогово-цифрового перетворювача (АЦП) спектрометра [2]. Такий підхід є загально визнаним в хеометриці. Зазвичай значення x в межах однієї строки змінюються від деякого мінімального значення $x_{\min}$ до деякого максимального значення $x_{\max}$. При переході від однієї вибірки до іншої значення СКЯ можуть значно відрізнятися. Їх абсолютні значення не повинні

впливати на подальшу статистичну обробку цих даних. Оскільки інформація про СКЯ міститься не в абсолютних значеннях відбивних характеристик на певних довжинах хвиль, а в його зміні щодо довжини хвилі, то можна провести нормування одержаних в результаті експериментів характеристик, які в подальшому будуть використовуватись для розпізнання об'єктів [3]. Слід зазначити, що значна кількість методів розпізнавання, в особливості параметричних, є самонормуючими, тобто нечутливи до масштабу, оскільки для обчислень використовується обернена коваріаційна матриця та виконується центрування. Однак, і у цьому випадку, при практичній реалізації, нормування також може позитивно вплинути на якість розпізнавання за рахунок усунення помилок при обчисленнях у процесорах з фіксованою розрядністю.

Приведення до єдиного масштабу забезпечується нормуванням кожної змінної на діапазон розкиду її значень. Для цього функція перетворення $y=f(x)$ повинна мати наступні властивості:

$$y(x_{\min})=K_{\min}; y(x_{\max})=K_{\max}; dy/dx>0. \quad (1)$$

Будь-яка функція з такими властивостями може бути використана для нормування. Наприклад, якщо $x_{\max} \rightarrow \infty$, то можна обрати функцію:

$$y(x) = 1 - \exp\left(1 - \frac{x}{x_{\min}}\right). \quad (2)$$

Тобто, за рахунок вибору відповідної функції можна врахувати різноманітні ефекти спотворення оцінок. Наприклад, схильність набору даних до крайніх оцінок. При цьому, можливо, слід застосовувати для різних наборів даних і різні функції перетворення, що враховують особливості їх поведінки, статусу і тому подібне. Зразкові графіки таких функцій наведені на рис. 1.

Приведення до єдиного масштабу забезпечується нормуванням кожної змінної на діапазон розкиду її значень. У простому варіанті це – лінійне перетворення:

$$\tilde{x}_i = \frac{x_i - x_{i,\min}}{x_{i,\max} - x_{i,\min}}, \quad (3)$$

де $\tilde{x}_i, x_i$ – нормований та вимірний спектральні коефіцієнти яскравості на довжині хвилі λ_i , відповідно; $x_{i,\max}, x_{i,\min}$ – максимальний та мінімальний вимірні спектральні коефіцієнти яскравості, відповідно.

Слід зауважити, що лінійне нормування оптимальне, коли значення змінної щільно заповнюють певний інтервал. Але подібний «прямолінійний» підхід можна застосовувати далеко не завжди. Так, якщо при розпізнанні рослинних об'єктів в даних є відносно рідкісні викиди, що перевищують типовий розкид, то саме ці викиди визначають, згідно з попередньою формулою (3), масштаб нормування. Це приведе до того, що основна маса значень нормованої змінної зосередиться поблизу нуля

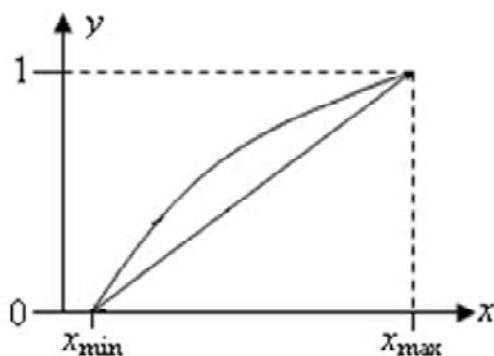


Рис. 1. Графіки функцій нормування

$|\tilde{x}_i| \ll 1$, або, навпаки, поблизу одиниці (крива «ранжована» на рис. 2 має мало значень в середині діапазону). На рис. 2 під назвою «вибірка» наведена для прикладу одна вибірка реальних даних – СКЯ амброзії у логарифмічному масштабі, у відносних одиницях виходу 12 розрядного АЦП спектрометра, в залежності від умовної довжини хвилі – i , що відповідає λ_i . Для рангової статистики цієї вибірки (крива «ранжована» на рис. 2 позначені пунктиром межі $P_0, P_5, P_{10}, P_{20}, P_{30}, P_{50}, P_{70}, P_{80}, P_{90}, P_{100}$, що є процентилями відповідного рівня.

Ще одним фактором є обмеження, що виникають при цифруванні СКЯ за рахунок обмеження розрядної сітки АЦП, що ілюструє крива «вибірка» на рис. 2 в діапазоні від 7 до 16 умовної довжини хвилі.

Для усунення впливу обмежень можна використувати не діапазон від мінімального значення до максимального, а від $N-i$ до $1-N$ процентиля. При цьому N може мати значення від 0 до 0,5 (0,5 відповідає 50-й перцентилі, або медіані). Узагальнення формули (3) матиме вигляд:

$$\tilde{x}_i = (1 - 2 \cdot N) \frac{x_i - x_{P_N}}{x_{P_{1-N}} - x_{P_N}} + N, \quad (4)$$

де x_{P_N} – значення N -процентилі.

Іноді при нормуванні набагато надійніше орієнтуватися не на екстремальні значення, а на статистичні характеристики даних, такі як середнє значення $\bar{x}_i$ і дисперсія σ_i :

$$\tilde{x}_i = \frac{x_i - \bar{x}_i}{\sigma_i}. \quad (5)$$

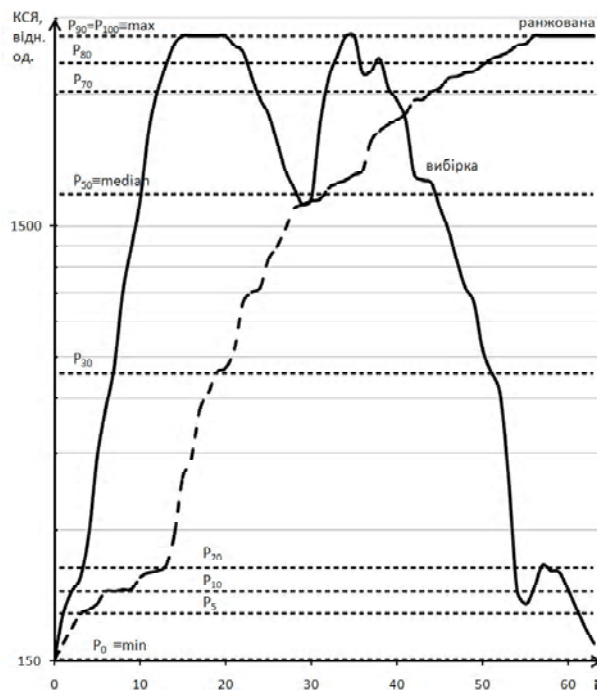


Рис. 2. Використання рангових статистик для нормування

В цьому випадку основна маса даних матиме одиничний масштаб, тобто типові значення усіх змінних будуть порівняними. Перетворення (5) в хеометриці має назву автошкалювання.

Іноді для обмеження діапазону можливих значень вхідних даних використовуються нелінійні гладкі функції, наприклад, – логістичну

$$f(\tilde{x}_i) = (1 + e^{-\tilde{x}_i})^{-1}. \quad (6)$$

Однак до «такого кроку слід вдаватися з обережністю оскільки таке перетворення змінює структуру даних. А у випадку розпізнавання за СКЯ у цьому не має потреби, оскільки максимальні значення вхідних даних і так обмежені розрядністю АЦП спектрометра.

Для порівняння впливу видів нормування на якість розпізнавання рослинних об'єктів, порівнюємо ті види нормування, де масив даних нормується за допомогою лінійного перетворення (4) і за допомогою автошкалювання (5).

УМОВИ І ПРОВЕДЕННЯ ЕКСПЕРИМЕНТУ

Експериментальні дослідження проводилися в два етапи:

- польові дослідження спектрів відбиття від рослин;
- обробка результатів досліджень спектрів.

Польові дослідження проводилися на тестовому полігоні Інституту олійних культур Національної академії аграрних наук, який розташований біля м. Запоріжжя. З отриманих експериментальних даних було відібрано 1000 наборів з 2048 спектральними точками. Критерієм відбору була можливість візуального визначення виду рослин по зображеннях і формі спектральної кривої (відсутність обмежень і надмірної зашумленості). Кількість вибірок різних видів рослин, що були відібрані для подальшого дослідження, та їх назви наведені в табл. 1.

За допомогою пристрою для дослідження спектрів рослин вимірювалися відносні спектральні характеристики в автоматичному режимі з періодом 1 с зі збереженням як поточної спектрограми, так і додаткової інформації (місце розташування і зовнішній вигляд рослини). Характеристики рослин в рядку знімалися послідовно, об'єктів приймача (спектрометричного блоку) переміщався уздовж рядка плавно, без ривків і зупи-

нок. Пристрій для дослідження спектрів складався з: спектрометра Red Tide 650 з волоконно-оптичним кабелем P200-2-UV-VIS і лінзою 74-VIS (фірма Ocean Optics); програмні компоненти, що входять в базовий комплект програмного забезпечення спектрометра: SpectraSuite і OOI_base32; Web-камера Genius VideoCamEye для реєстрації зображення рослин; датчик місцеположення BT-Q920 фірми QSatarz.

ОБРОБКА РЕЗУЛЬТАТІВ ЕКСПЕРИМЕНТУ

Після усереднення 2048 СКЯ, що забезпечуються спектрометром за один замір, до 64, були здобуті 3 масиви чисел: 587×64 – матриця коефіцієнтів відбиття для соняшника, 189×64 – матриця коефіцієнтів відбиття для мишію і 224×64 – матриця коефіцієнтів відбиття для амброзії. Наступним кроком було нормування – кожна з трьох матриць піддавалася нормуванню по лінійному закону нормування згідно формули (4) з різними значеннями N і автошкалювання згідно (5).

Для перевірки якості розпізнавання рослинних об'єктів за експериментальними даними був використаний блок дискримінаційного аналізу General Discriminant Analysis (GDA) програми Statistica 8.0 компанії StatSoft Inc. Він дозволяє ефективно здійснювати статистичний аналіз і обробку експериментальних даних, зокрема визначити статистичні характеристики вибірових сукупностей, підгонку розподілів, здійснювати варіаційний, регресійний і дискримінантний аналіз, будувати гістограми і графіки різної складності. Проведення дискримінаційного аналізу потрібне для того, щоб визначити можливість класифікації рослинних об'єктів за спектральними характеристикам, а також оцінити якість розпізнавання [4]. На даному етапі, для цього аналізу, вибір кількості ознак не брався до уваги, тому, що головною метою було порівняти вплив різних видів нормування на якість розпізнавання при однакових умовах, а не як збільшити якість розпізнавання за рахунок відбору найбільш інформативних ознак.

Були отримані наступні результати. Табл. 2 демонструє показники розпізнавання без проведення нормування взагалі. Табл. 3 дозволяє виявити залежність якості розпізнавання від номеру використаної для нормування процентилі.

Залежність вірогідності вірної класифікації від рівнів (процентилів), що використовуються для нормування, проілюстровано на рис. 3. Для порівняння на рис. 3 наведені вірогідності для випадків використання автошкалювання та розпізнавання без попереднього нормування у вигляді горизонтальних ліній. Згідно з цим рисунком нормування може як покращити якість подальшого розпізнавання, так і погіршити її, але в незначних межах.

Однією з умов застосування методів дискримінантного аналізу є нормальність розподілу вибірок. Методика перевірки на нормальність розподілу описана в міжнародному стандарті ISO 5479-97 [5]. Дозволяється деяке відхилення від нормального розподілу, при якому досить успішно класифікуються вибірки. В цілому вибірки відповідали нормальному розподілу.

Таблиця 1. Видовий склад досліджуваних рослин

Вид рослини	Латинська назва	Клас	Кількість вибірок
соняшник	<i>Heliánthus ánnuus</i>	1	587
амброзія	<i>Ambrósia artemisiifólia</i>	2	224
мишій	<i>Setaria viridis</i>	3	189
всього			1000

Таблиця 2. Класифікаційна матриця без нормування

		Результати класифікації для кожного класу			
		1(соняшник)	2(амброзія)	3(мишій)	Всього
Априорні дані про клас	1(соняшник)	557	26	4	587
	2(амброзія)	53	155	16	224
	3(мишій)	36	9	144	189
	Всього	646	190	164	1000
Вірогідність вірної класифікації		94,89 %	69,20 %	76,19 %	85,6 %

Таблиця 3. Класифікаційна матриця при використанні різних видів лінійного нормування

Рівні для нормування	N	0(min)	0,03	0,05	0,07	0,1	0,2	0,3	авто- шкал.	без норм.
	1-N	1(max)	0,97	0,95	0,93	0,9	0,8	0,7		
Кількість вірно класифікованих вибірок різних класів	2(сон.)	156	156	156	156	154	152	155	154	155
	3 кл.	136	135	147	143	142	143	142	135	144
	1 кл.	556	555	558	560	558	556	554	559	557
Вірогідність вірної класифікації для всіх класів, %		84,8	84,6	86,1	85,9	85,4	85,1	85,1	84,8	85,6

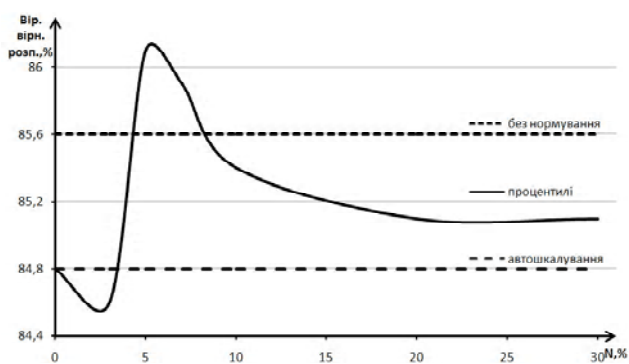


Рис. 3. Залежність вірогідності вірного розпізнавання від рівнів для нормування

ВИСНОВКИ

Перевірка впливу різних видів нормування на якість розпізнавання рослинних об'єктів за результатами дистанційного зондування виявила, що види нормування не значно, але впливають на кінцеву якість розпізнавання. Обробка експериментальних результатів за допомогою дискримінаційного аналізу показала, що для розпізнавання рослин найкраще використовувати лінійне нормування зі значеннями 5 та 95 процентилі замість мінімальних та максимальних, відповідно. При цьому виграш відносно розпізнавання без нормування склав 0,5 %, а відносно автошкалювання 1,3 %. Виграш – незначний, але при вірогідностях вірного розпізнавання близько 90 % кожен додатковий відсоток має значення та досягається значним ускладненням процедури розпізнавання. Крім того, як вже зауважувалося вище, використання розпізнавання без нормування не завжди можливе на практиці. Отримані результати слід віднести до випадків, практичне значення яких не слід узагальнювати без перевірки. На користь цього говорить те, що найбільший вплив нормування створює на розпізнавання лише одного класу – третього, що відповідає мишій – бур'яну з невеликою площею листя, і тому меншим інтегральним коефіцієнтом відбиття. Але результати цієї статті показують, що при вирішенні задач розпізнавання образів

при необхідності нормування початкових даних, слід враховувати їх вплив та обирати їх оптимальні види для досягнення якісного розпізнавання.

СПИСОК ЛІТЕРАТУРИ

1. Пиза, Д. М. Выбор эффективного метода распознавания растений по коэффициентам спектральной яркости / Д. М. Пиза, С. В. Морщавка, Ю. В. Скоробогатов // Радиоэлектроника, информатика, управление. – 1999. – № 1. – С. 85–88.
2. Дистанционное зондирование: количественный подход / [Ш. М. Дейвис, Д. А. Ландгребе, Т. Л. Филлипс и др.]; под ред. Ф. Свейна и Ш. Дейвис. Пер. с англ. М.: Недра. – 1983. – С. 415. (Пер. изд. США, 1978, 396 с.)
3. Морщавка, С. В. Информационные технологии обработки пропашных культур. Радиометрические аспекты / С. В. Морщавка, Д. М. Пиза, Е. А. Белоусов // Радиоэлектроника, информатика, управление. – 2008. – № 2. – С. 45–52.
4. Боровиков, В. П. Популярное введение в программу STATISTICA/ В. П. Боровиков. – М.: Компьютер Пресс, 1998. – 238 с.
5. Статистическое представление данных. Проверка отклонения распределения вероятностей от нормального распределения (ISO 5479-97): ГОСТ Р ИСО 5479-2002. – [Действителен с 2002-01-22] – М.: Госстандарт России, 2002. – 27 с. – (Технический комитет по стандартизации ТК125)

Стаття надійшла до редакції 11.05.2011.

Морщавка С. В., Шама Е. А., Пиза Д. М.

ВЛИЯНИЕ ВИДОВ НОРМИРОВАНИЯ НА КАЧЕСТВО РАСПОЗНАВАНИЯ РАСТИТЕЛЬНЫХ ОБЪЕКТОВ

Рассмотрено влияние разных видов нормирования выборок на качество распознавания по результатам дистанционного зондирования.

Ключевые слова: распознавание, хеометрика, нормирование, спектральные коэффициенты яркости.

Morshchavka S. V., Shama E. O., Piza D. M.

EFFECT OF NORMALIZATION ON THE QUALITY OF RECOGNITION OF PLANT OBJECTS

Influence of different types of samples norm is considered on quality of recognition on results of the remote sensing.

Key words: recognition, xemometrics, normalization, the spectral brightness coefficients.

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL AND COMPUTER MODELLING

УДК 681.5.015

Агибалов А. П.¹, Поляков М. А.²¹Магистрант Запорожского национального технического университета²Канд. техн. наук, доцент Запорожского национального технического университета

ТРАНСЛЯТОР ПАРАМЕТРОВ МОДЕЛИ КОНЕЧНОГО АВТОМАТА ИЗ СРЕДЫ МАТЛАВ В ПРИЛОЖЕНИЕ ЧЕЛОВЕКО-МАШИННОГО ИНТЕРФЕЙСА

Предложена технология и средства автоматического переноса параметров модели конечного автомата из приложения Matlab в приложение человеко-машинного интерфейса. Рассмотрен пример применения технологии к модели электрического аппарата.

Ключевые слова: stateflow, конечный автомат, DDE-диалог, человеко-машинный интерфейс, тег.

ВВЕДЕНИЕ

В настоящее время приложения, созданные с помощью программных средств человеко-машинного интерфейса (HMI, англ. – Human Machine Interface), являются неотъемлемой частью SCADA-систем и контроллерных систем управления техническими объектами и процессами [1, 2]. В среде этих приложений выполняется часть логики пользователя, которая формализуется с помощью конечных автоматов [3, 4].

В ряде случаев, в процессе проектирования этих автоматов, прибегают к их моделированию в среде универсальных пакетов моделирования, например приложения Simulink пакета MATLAB [5]. Библиотека Stateflow этого пакета позволяет графически интерпретировать конечный автомат проектируемой системы управления как событийно-управляемую модель.

В существующей технологии проектирования операция переноса результатов моделирования логики управления пользователя в приложение HMI выполняется вручную, является трудоемкой, требующей высокой квалификации исполнителя.

Целью данной работы является сокращение трудоемкости проектирования приложений HMI путем созда-

ния технологии и средств автоматического переноса параметров модели конечного автомата из приложения Matlab в приложение HMI.

ТЕХНОЛОГИЯ И СРЕДСТВА ПЕРЕНОСА

Предлагаемая технология автоматического переноса параметров модели включает анализ файла модели Simulink и создание матричных эквивалентов элементов модели; инициацию DDE-диалога между приложениями Matlab и HMI; перенос созданных матричных эквивалентов в приложение HMI. Для реализации этой технологии созданы скрипты и функции Matlab, а также VBA-модули, встроенные в проект приложения HMI. Совокупность разработанных средств будем называть транслятором параметров модели конечного автомата из среды MATLAB в приложение человеко-машинного интерфейса, далее – транслятор. Транслятор предназначен для приложений HMI, созданных с помощью программного пакета RSVIEW32 компании Rockwell Automation (США) [2], но предлагаемая технология трансляции может быть использована при разработке трансляторов в другие пакеты HMI.

Программная структура транслятора представлена на рис. 1. Ядро транслятора – скрипт *analysis.m*. В процес-

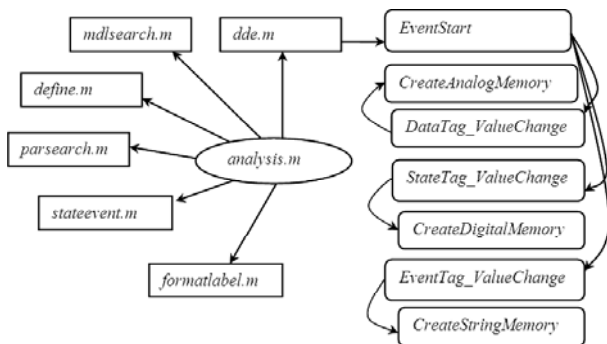


Рис. 1. Программная структура транслятора

се его выполнения вводится имя анализируемой Simulink-модели и имя проекта RSVIEW32, выбранного для генерации параметров модели, вызываются функции и скрипты для осуществления процесса трансляции.

Скрипт *mdlsearch.m* предназначен для поиска координат описания элементов модели. Скрипт *define.m* создан для формирования массивов элементов модели. Скрипт *parsearch.m* служит для поиска вышеназванных параметров в сформированных массивах. Скрипт *stateevent.m* создан для анализа меток состояний. Скрипт *formatlabel.m* отвечает за форматирование анализируемых данных для использования в SCADA-проекте. Скрипт *dde.m* инициализирует DDE-диалог с программой RSVIEW32 и посылает данные в проект; при этом VBA-модуль *EventStart* осуществляет процесс генерации тегов проекта при помощи пар модулей *DataTag_ValueChange*, *CreateAnalogMemory* и *StateTag_ValueChange*, *CreateDigitalMemory*. Скрипт *ddeexcel.m* служит для переноса результатов анализа в таблицу MS Excel для наглядного отображения структуры всей модели Simulink.

В результате работы программы создается база данных тегов SCADA-проекта, наиболее точно отвечающая параметрам элементов Simulink-модели.

ПРИМЕНЕНИЕ ТРАНСЛЯТОРА ДЛЯ ОБРАБОТКИ МОДЕЛИ

Рассмотрим использование транслятора на примере моделирования повторно-кратковременного режима работы электрического аппарата, при котором температура его частей за время нагрузки не достигает установившегося значения, а за время паузы не достигает температуры холодного состояния [6].

Ставится задача трансляции модели электрического аппарата из приложения моделирования в приложение мониторинга RSVIEW32.

Simulink-модель электрического аппарата изображена на рис. 2. Параметрами модели являются время нагрузки tw и паузы tp , постоянные времени нагрева и охлаждения tay и $tay1$ соответственно, температура окружающей среды T и граничная температура нагрева аппарата T_1 . В примере приняты следующие значения констант: $tw = 1$ с, $tp = 2$ с, $tay = 7$ с, $tay1 = 15$ с, $T = 20$ град., $T_1 = 120$ град. Наблюдаемыми величинами являются мгновенная температура нагрева аппарата Th и режим его работы $mode$.

Блок Stateflow Chart реализует функцию конечного автомата. Диаграмма его состояний и переходов изображена на рис. 3.

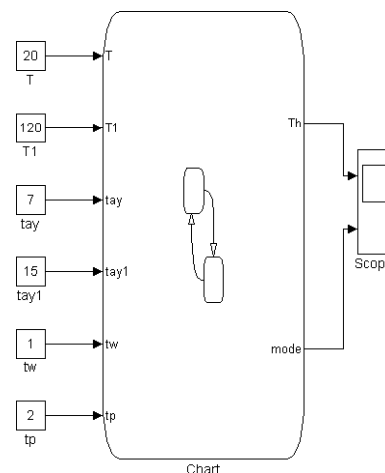


Рис. 2. Simulink-модель электрического аппарата

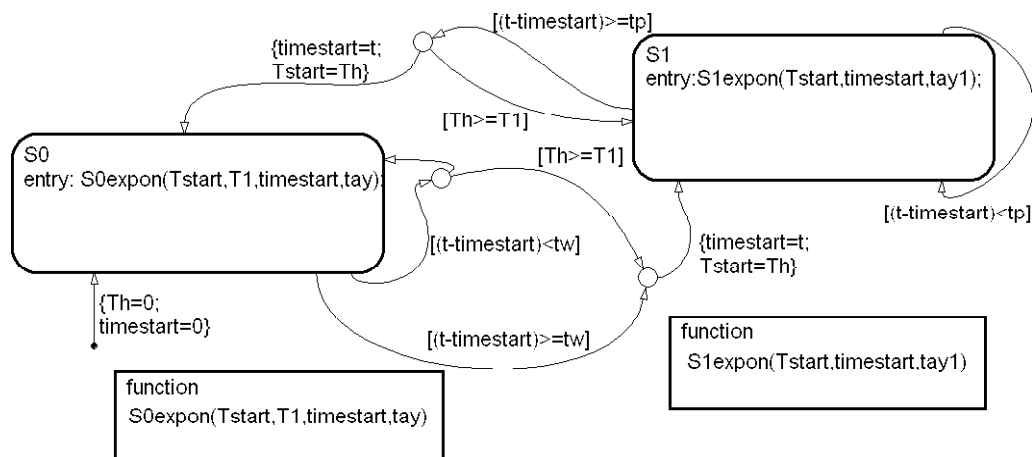


Рис. 3. Диаграмма состояний автомата и переходов блока Stateflow Chart

Автомат находится в одном из двух состояний: *S0* (под нагрузкой) и *S1* (без нагрузки). Каждое состояние характеризуется действием при входе в него (*entry*:), а именно – изменением температуры частей аппарата в зависимости от времени симуляции.

Действие при входе описывается графическими функциями *S0expon* и *S1expon*. Входными переменными функций являются константы модели.

Состояния модели связаны системой переходов, описываемых условиями и действиями, а также соединительными узлами для упрощения диаграммы Stateflow (рис. 3). В процессе симулирования температура аппарата образует экспоненциально нарастающий и спадающий процесс.

Результатами работы транслятора являются:

- 1) матричные эквиваленты элементов модели в рабочей области Matlab (рис. 4);
- 2) база данных тегов в HMI-проекте (рис. 5);
- 3) таблица на листе Excel с параметрами модели автомата (рис. 6).

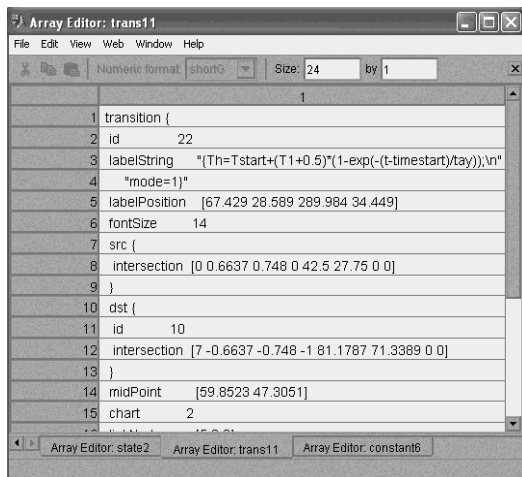


Рис. 4. Массив, отвечающий переходу диаграммы

?	Tag Name	Value	State	Description
1	Y transition1	[(t-timestamp)<tw]	valid	Event description
2	Y transition2	{ttimestamp=t;Tstart=Th}	valid	Event description
3	Y transition3	{ttimestamp=t;Tstart=Th}	valid	Event description
4	Y transition4	[(t-timestamp)<tp]	valid	Event description
5	Y transition5	{Th=0;timestamp=0}	valid	Event description
6	Y tay	7	valid	Input data
7	Y tay1	15	valid	Input data
8	Y T1	21	valid	Input data
9	Y T	20	valid	Input data
10	Y S0	On	valid	State tag

Рис. 5. Монитор тегов RSVIEW32

	A	B	C	D	E	F	G
1	State						
2	label	S0	S1	S0expon_	S1expon_	Tstart	timestamp
3	id		3	4	5	6	
4	entry:	S0expon(Tstart	S1expon(Tstart	timestamp	tay1)		
5	during						
6	exit:						
7	on_event:						
8							
9							
10	id	12	13	14	15	16	17
11	Condition	t-timestamp	<tw		t-timestamp	<tp	t-timestamp
12	Action	timestamp=t;	Tstart=Th	Tstart=Th	timestamp=0		
13	Source	3	8	7	4	SIMULINK	3
14	Destination	9	4	3	4	3	8

Рис. 6. Отчет по трансляции на листе Excel

ВЫВОДЫ

Для создания HMI-проектов с клиент-серверной или распределенной архитектурой используются SCADA-системы. Логика пользователя приложений HMI формализуется в виде конечного автомата и может быть смоделирована в программе Simulink пакета Matlab. Основными элементами модели являются входные и выходные параметры, состояния автомата и переходы между ними.

Разработанный транслятор параметров модели является комплексом программ, выполняемых в среде Matlab и VBA-надстройке RSVIEW32. Трансляция основана на механизме DDE. Трансляции подвергаются:

- 1) параметры состояний: название, id-номер, действия при входе, выходе, во время активности состояния и при совершении установленного события;
- 2) параметры переходов: id-номер, условие совершения, параллельное действие, начальная и конечная точка;
- 3) входные и выходные данные: id-номер, имя переменной, тип данных, тип переменной, величина.

Автоматическое отображение результатов анализа модели на листе Excel упрощает восприятие сложных систем. Разработанный инструмент успешно прошел тестирование на demo-примерах приложения Simulink, в частности, на модели нагревателя, инфракрасной системы наведения ракеты, системы подачи топлива в автомобиле. В то же время предложенная технология имеет ограниченную функциональность, поскольку не генерирует автоматически события проекта RSVIEW32, не создает дисплеи и сигналы тревоги для мониторинга контролируемых величин, и потому имеет перспективу улучшения.

СПИСОК ЛИТЕРАТУРЫ

1. Олсон, Г. Цифровые системы автоматизации и управления / Г. Олсон, Д. Пиани. – С. Пб.: Невский Диалект, 2001. – 557 с.: ил.
2. RSVIEW32: руководство пользователя [Электронный ресурс] / Rockwell Software Inc. – Электрон. дан. – М., [2000]. – Режим доступа: www.eskovostok.ru/_docs/9399-2se32ug-ru.pdf, вільний. – Загл. с экрана.
3. Поляков, М. А. Теоретико-множественная модель интегрированной контроллерной системы управления / М. А. Поляков // Системні технології. – 2009. – № 4. – С. 131–137.
4. Поляков, М. А. Логическое управление объектами электрических систем в среде приложения человеко-машинного интерфейса / М. А. Поляков // Наукові праці Донецького національного технічного університету. – 2009. – № 9(158). – С. 197–201.
5. Дзбни, Д. Simulink 4. Секреты мастерства / Д. Дзбни, Т. Харман. – М.: БИНОМ. Лаборатория знаний, 2003. – 404 с.
6. Электрические и электронные аппараты: учебник для вузов / под ред. Ю. К. Розанова. – 2-е изд., испр. и доп. – М.: Информэлектро, 2001. – 420 с.: ил.

Стаття надійшла до редакції 16.12.2010.
Після доробки 01.02.2011.

Агібалов О. П., Поляков М. О.

ТРАНСЛЯТОР ПАРАМЕТРІВ МОДЕЛІ КІНЦЕВОГО АВТОМАТА ІЗ СЕРЕДОВИЩА MATLAB У ДОДАТОК ЛЮДИНО-МАШИННОГО ІНТЕРФЕЙСУ

Запропоновано технологію та засоби автоматичного переносу параметрів моделі кінцевого автомата з додатку Matlab до додатку людино-машинного інтерфейсу. Розглянуто приклад застосування технології до моделі електричного апарата.

Ключові слова: stateflow, кінцевий автомат, DDE-діалог, людино-машинний інтерфейс, тег.

Agibalov A. P., Polyakov M. A.

TRANSLATOR OF FINITE STATE MACHINE MODEL PARAMETERS FROM MATLAB ENVIRONMENT INTO HUMAN-MACHINE INTERFACE APPLICATION

Technology and means for automatic translation of FSM model parameters from Matlab application to human-machine interface application is proposed. The example of technology application to the electric apparatus model is described.

Key words: stateflow, finite-state machine, DDE-dialog, human-machine interface, tag.

УДК 004.056.53:004.75

Андрущенко Д. М.¹, Варава М. Ю.², Неласа Г. В.³¹Молодий науковий співробітник Запорізького національного технічного університету²Провідний спеціаліст з інформаційних технологій КБ Приватбанк³Канд. техн. наук, доцент Запорізького національного технічного університетуРОЗПАРАЛЕЛЮВАННЯ ρ - І λ - МЕТОДІВ ПОЛЛАРДА РОЗВ'ЯЗАННЯ ЗАДАЧІ ДИСКРЕТНОГО ЛОГАРИФМУВАННЯ

Проведено аналіз ефективності розпаралелювання ρ - і λ -методів Полларда при вирішенні задачі дискретного логарифмування. Наводиться теоретична оцінка часу виконання завдання на паралельній системі. Проведено порівняння результатів практичних і теоретичних розрахунків. Зроблені заміри часу виконання розпаралелених методів.

Ключові слова: методи Полларда, дискретний логарифм, розпаралелювання, складність криптоалгоритма, оцінка складності.

ВСТУП

При створенні систем захисту інформації широко застосовуються асиметричні алгоритми. Надійність таких систем основана на трудомісткості виконання одного з наступних типів зворотних перетворень: розкладання великих чисел на прості множники; обчислення дискретного логарифму; обчислення коренів алгебраїчних рівнянь [1, 2]. Однак розвиток методів прискорених обчислень, в тому числі застосування багатопроцесорних систем паралельних рішень, викликає небезпеку зниження ступеню захисту криптосистем. Тому для оцінки надійності систем захисту та їх вдосконалення необхідно дослідження методів паралельних рішень для проведення названих перетворень.

ПОСТАНОВКА ЗАДАЧІ

Одними з найбільш розповсюджених асиметричних криптосистем є ті, що створені на базі еліптичних кривих $y^2 = x^3 + ax + b \pmod{p}$ над простим полем $GF(p)$. Якщо P є базовою точкою адитивної групи точок еліптичної кривої простого порядку n , точка Q належить заданій групі точок еліптичної кривої, то злом криптографічної системи полягає у розв'язанні рівняння $m \cdot P = Q$ відносно m , де $1 < m < n - 1$ (адитивний аналог задачі дискретного логарифмування).

Надійність систем захисту інформації, злом яких оснований на розв'язанні задачі дискретного логарифму-

вання залежить від величини n . Однак, збільшення n , окрім підвищення надійності, призводить до збільшення часу роботи криптографічних алгоритмів. Тому для побудови ефективних алгоритмів необхідний компромісний варіант, що забезпечує достатню надійність захисту при прийнятному рівні модуля n з точки зору швидкості роботи алгоритму. Поява нових методів прискорення обчислень дискретного логарифму, одним з яких є розпаралелювання, викликає необхідність збільшення параметру n . Тому оцінка ступеня надійності криптографічних систем, дослідження можливості їх злому шляхом розпаралелювання процесу розв'язання задачі дискретного логарифмування є актуальною проблемою.

Відомі різні алгоритми послідовного вирішення цього завдання, в тому числі: великих-малих кроків, Поліга-Хеллмана, ρ -Полларда, λ -Полларда, Адлемана, index-calculus [1–3]. Більшість з них піддаються розпаралелюванню. У даній роботі для перевірки працездатності системи та оцінки ефективності даного підходу були розглянуті ρ -метод і λ -метод Полларда [2].

Метод ρ -Полларда. Ідея методу полягає в побудові послідовності точок Z_i еліптичної кривої

$$Z_i = A_i P + B_i Q, \quad (1)$$

де $1 < A_i, B_i < n - 1$,

яка забезпечує знаходження пари рівних точок Z_i і Z_j :

$$Z_i = Z_j, i \neq j. \quad (2)$$

У цьому випадку одержимо

$$A_i P + B_i Q = A_j P + B_j Q, \quad (3)$$

ЗВІДСИ

$$Q = \frac{A_j - A_i}{B_i - B_j} P. \quad (4)$$

Таким чином,

$$d = \frac{A_j - A_i}{B_i - B_j} (\text{mod } n), \quad B_i \neq B_j. \quad (5)$$

Для реалізації методу оберемо два випадкових числа $A_0, B_0 \in [0, n-1]$, так щоб одночасно A_0 і B_0 не були рівними нулю:

$$Z_0 = A_0 P + B_0 Q. \quad (6)$$

Розіб'ємо відрізок $(0, p)$ на три рівних множини S_1, S_2 й S_3 і обчислимо Z_{i+j} рекурентне за правилом

$$Z_{i+1} = \begin{cases} 2Z_i, & \text{якщо } Z_i^x \in S_1; \\ Z_i + P, & \text{якщо } Z_i^x \in S_2; \\ Z_i + Q, & \text{якщо } Z_i^x \in S_3; \end{cases} \quad (7)$$

де Z_i^x визначає x -координату точки на еліптичній кривій.

Звідси

$$A_{i+1} = \begin{cases} 2A_i, & \text{якщо } Z_i^x \in S_1; \\ A_i + 1, & \text{якщо } Z_i^x \in S_2; \\ A_i, & \text{якщо } Z_i^x \in S_3; \end{cases} \quad (8)$$

$$B_{i+1} = \begin{cases} 2B_i, & \text{якщо } Z_i^x \in S_1; \\ B_i, & \text{якщо } Z_i^x \in S_2; \\ B_i + 1, & \text{якщо } Z_i^x \in S_3. \end{cases} \quad (9)$$

Метод ρ -Полларда проілюстровано на рис. 1.

Метод λ -Полларда. Очевидно, колізію точок можна одержати й іншим шляхом, рухаючись із двох (або більше) різних точок Z_k й Z_l до збігу $Z_{k+s} = Z_{l+t}$. Ця ситуація відображена на рис. 2. Даний метод одержання колізії називається λ -методом Полларда, походження назви якого зрозуміло з рисунка.

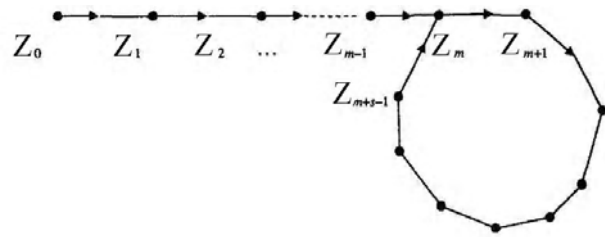


Рис. 1. ρ -метод Полларда

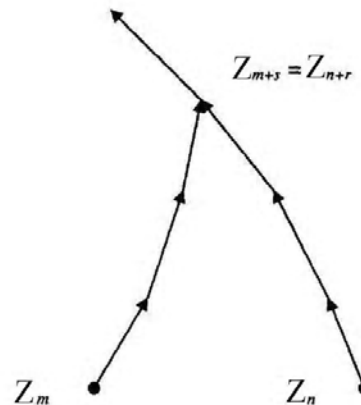


Рис. 2. λ -метод Полларда

Цей метод також може бути основою для розпаралелювання задачі обчислювання дискретного логарифму за допомогою r комп'ютерів, що може в r раз прискорити розв'язання задачі дискретного логарифмування.

РОЗПАРАЛЕЛЮВАННЯ МЕТОДІВ ПОЛЛАРДА

Алгоритм р-Полларда при великих значеннях n стає неефективним. Він може бути поліпшений, якщо всю множину точок еліптичної кривої, включаючи точку нескінченності, розбити на $r > 3$ множин, які не перетинаються. У цьому випадку послідовність Z_i задається у вигляді:

[illegible]

де c_v, d_v – випадкові цілі числа з інтервалу $[0, n-1]$.

При використанні ітераційних формул виду (10) можна зменшити складність (вартість часу) процедури криптоаналізу. Крім того, застосування цього підходу дозволяє ефективно розпаралелювати процес знаходження коефіцієнтів c_v , d_v принаймні на r процесів.

Якщо числа c_ν й d_ν вибрати таким чином, щоб перетинання значень Z_i в областях не було, то й кожну

область можна розбити на необхідне число підобластей, що дозволяє розпаралелювати задачу на необхідну кількість незалежних процесів. Результати роботи кожного із процесів у вигляді трійок чисел $\langle Z_i, c_i, d_i \rangle$ заносяться в загальну базу даних. Центральний процесор здійснює паралельний пошук значень $Z_i = Z_j$, $i \neq j$, і після цього можливе обчислення значення m .

Ілюстрацію розпаралелених ρ -і λ -методів Полларда наведено на рис. 3 та рис. 4.

Однією з головних характеристик паралельних систем є прискорення R паралельної системи, що визначається формулою: $R = T_1 / T_r$, де T_1 – час розв'язання задачі на однопроцесорній системі, а T_r – час розв'язання тієї ж задачі на r -процесорній системі. Нехай $W = W_{\text{ск}} + W_{\text{пр}}$, де W – загальне число операцій, $W_{\text{пр}}$ – число операцій, які можна виконувати паралельно, а $W_{\text{ск}}$ – число скалярних (нерозпаралелених) операцій. Позначимо також через t час виконання однієї операції. Тоді одержуємо оцінку для прискорення R

$$R = \frac{W \cdot t}{(W_{\text{ск}} + \frac{W_{\text{пр}}}{r}) \cdot t} = \frac{1}{a + \frac{1-a}{r}} \xrightarrow{r \rightarrow \infty} \frac{1}{a}, \quad (11)$$

де $a = W_{\text{ск}} / W$.

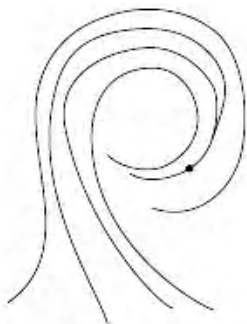


Рис. 3. Розпаралелений ρ -метод Полларда

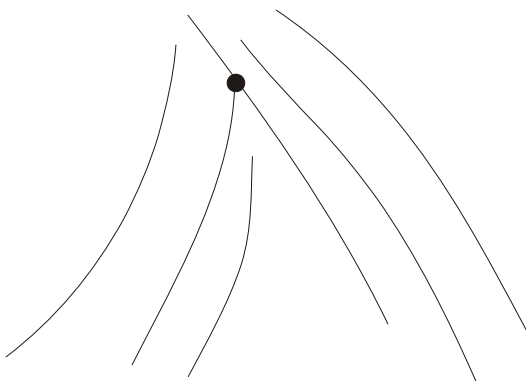


Рис. 4. Розпаралелений λ -метод Полларда

Таким чином, якщо 9/10 програми виконується паралельно, а 1/10 як і раніше послідовно, то більшого, ніж в 10 разів, прискорення одержати в принципі неможливо без втрат якості реалізації паралельної частини коду й кількості використаних процесів. Очевидно, що 10-кратне прискорення досягається тільки в тому випадку, коли час виконання паралельної частини дорівнює нулю.

Розпаралелювання ρ -і λ -методів Полларда дозволить проводити криптоаналіз протоколів не за рахунок підвищення продуктивності окремого комп'ютера, а завдяки масовому використанню паралельно підключених комп'ютерів у мережі.

Складність методу ρ -Полларда оцінюється як [4]:

$$I_{\rho} = \sqrt{\frac{\pi n}{4}}. \quad (12)$$

Однією з переваг ρ -методу Полларда є те, що він допускає розпаралелювання на r незалежних процесів. У цьому випадку складність реалізації кожного із процесів можна оцінити як:

$$I_{\rho_1} = \frac{\sqrt{\frac{\pi n}{4}}}{r} = \sqrt{\frac{\pi n}{4r^2}}. \quad (13)$$

ОСНОВНІ РЕЗУЛЬТАТИ

На рис. 5 зображена залежність часу криптоаналізу L від порядку криптосистеми n й числа розпаралелених процесів r для ρ -методу Полларда. Як видно із графіка, використання паралельного виконання алгоритму, криптоаналіз має сенс, коли число процесів не більше 17–19.

Складність методу λ -Полларда оцінюється як [4]

$$I_{\lambda} = 2\sqrt{n}, \quad (14)$$

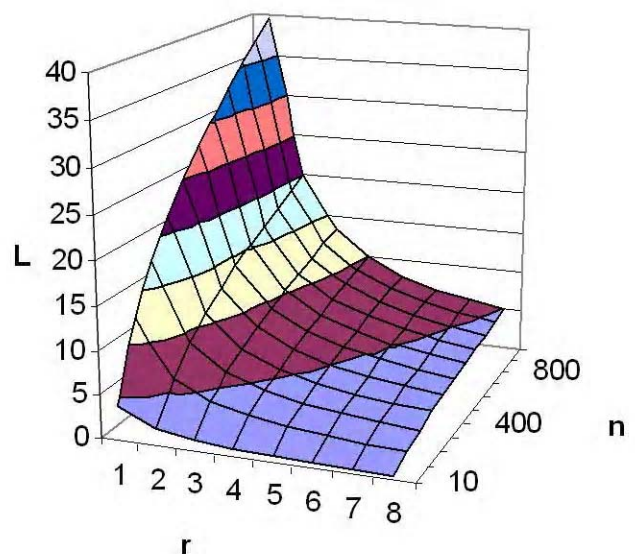


Рис. 5. Складність ρ -методу Полларда

а при розпаралелюванні

$$I_{\lambda_1} = \frac{2\sqrt{n}}{r}. \quad (15)$$

На рис. 6 зображена залежність часу криптоаналізу від порядку криптосистеми й числа розпаралелених процесів для λ -методу Полларда. Як видно із графіка, λ -метод Полларда вимагає більше часу для обчислень, ніж ρ -метод Полларда. Отже, доцільно використовувати ρ -метод Полларда.

Проведемо порівняльний аналіз по складності названих методів. Для цього знайдемо відношення:

$$\frac{I_{\lambda}}{I_{\rho}} = \frac{2\sqrt{n}}{\sqrt{\frac{\pi n}{4}}} = \frac{4}{\sqrt{\pi}} \approx 2,26. \quad (16)$$

З цього відношення випливає, що λ -метод Полларда більше, ніж в 2 рази складніший за метод ρ -Полларда.

ОЦІНКА ЧАСУ ВИКОНАННЯ РОЗПАРАЛЕЛЕНИХ ρ -І λ -МЕТОДІВ ПОЛЛАРДА

У ході виконання дослідження були зроблені заміри часу виконання розпаралелених ρ -і λ -методів Полларда. Програма була запущена на виконання по черзі на 4-х, на 5-х, на 6-х і на 7-х комп'ютерах. Результати тестування наведені в табл. 1 і на рис. 7.

Як видно з таблиці 1 й графіка на рис. 7, час виконання програми методом λ -Полларда в середньому більше, ніж час виконання аналогічної програми методом ρ -Полларда в 2 рази.

Були проведені виміри часу виконання програми методом ρ -Полларда з порядками базових точок $n=2^{64}$, 2^{73} , 2^{84} , 2^{96} на комп'ютерах AMD AtlonXP 1000+. Результати наведені в табл. 2 і на рис. 8. Час розв'язання задачі на одному комп'ютері становить 48 хв для $n=2^{64}$.

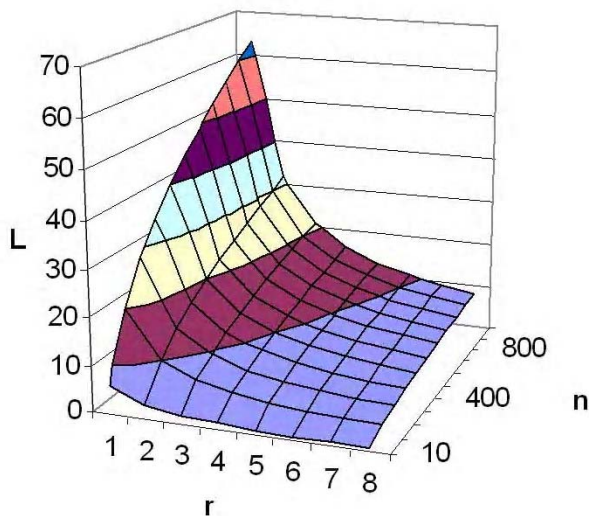


Рис. 6. Складність λ -методу Полларда

Таблиця 1. Час виконання програми, хв.

Кількість процесів, r	Метод	
	ρ -Полларда	λ -Полларда
4	6	10
5	3,5	6,5
6	2	4
7	1,5	2,5

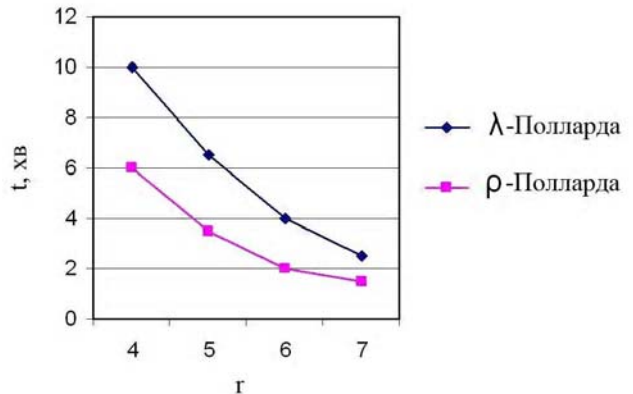


Рис. 7. Час виконання програми

Таблиця 2. Час виконання програми для різних n , хв.

r	$n=2^{64}$	$n=2^{73}$	$n=2^{84}$	$n=2^{96}$
4	22	41	—	—
5	14	26	—	—
6	11	19	38	—
7	10	15	25	37

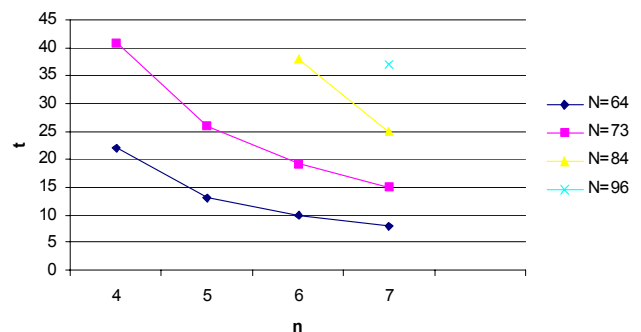


Рис. 8. Час виконання програми, хв.

З табл. 2 і графіка на рис. 8 видно, що програма криптоаналізу дозволяє істотно скоротити час обчислень.

Відомо, що можна прискорити обчислення за λ -методом Полларда, якщо відомий відрізок, на якому знаходиться значення дискретного логарифму [4]. У роботі [5] аналізується інший підхід – можливість використання λ -методу Полларда в тому випадку, коли не відомий відрізок, на якому лежить значення логарифму, шляхом розпаралелювання рішення алгоритму на довільну кількість обчислювальних вузлів r . При цьому весь діапазон, на якому проводиться пошук рішення, ділиться на r рівних частин. Кожен i -й обчислювальний вузол вирі-

шує завдання λ -методом Полларда так, як ніби розв'язання задачі дискретного логарифму лежить на i -му відрізку. Коли розв'язок знайдено одним з обчислювальних вузлів, інші свою роботу зупиняють. Результати обчислень представлені в табл. 3.

Таблиця 3. Залежність середнього прискорення паралельних обчислень від кількості обчислювальних вузлів r

Кількість обчислювальних вузлів r	1	2	3	4
Середнє прискорення обчислень	1	1,06	1,15	1,21

ВИСНОВКИ

Використання паралельних обчислень у криптоаналізі для прискорення методів ρ -і λ -Полларда знижує час виконання розрахунків. Але розглянуті методи розпаралелювання не дозволяють отримати значного прискорення обчислень, і не збільшують небезпеку злому при прийнятних величинах порядку базової точки. Надалі планується модернізація системи для дослідження інших методів розв'язання задачі дискретного логарифмування і порівняльного аналізу їх практичної складності обчислення.

СПИСОК ЛІТЕРАТУРИ

1. Молдовян, Н. А. Введение в криптосистемы с открытым ключом [Текст] / Н. А. Молдовян, А. А. Молдовян // С. Пб. : BHV, 2005. – 288 с.
2. Василенко, О. Н. Теоретико-числовые алгоритмы в криптографии [Текст] / О. Н. Василенко. – М. : МЦНМО, 2003. – 328 с.
3. Андрущенко, Д. М. Практическая оценка стойкости асимметричных криптосистем [Текст] / Д. М. Андрущенко, Г. Л. Козина, Д. М. Пиза // Проблемы информационной безопасности. – 2008. – № 1. – С. 57–62.

4. Сمارт, Н. Криптография [Текст] / Н. Смарт. – М. : Техносфера, 2005. – 528 с.
5. Андрущенко, Д. М. О распараллеливании методов Полларда решения задачи дискретного логарифмирования [Текст] / Д. М. Андрущенко, Г. Л. Козина // VII Всеукраїнська науково-практична конференція студентів, аспірантів та молодих вчених «Теоретичні і прикладні проблеми фізики, математики та інформатики» : збірка тез доповідей учасників. Частина 1. – К., 2009. – С. 10–11.

Стаття надійшла до редакції 18.05.2011.

Андрущенко Д. М., Варава М. Ю., Неласая А. В.

РАСПАРАЛЛЕЛИВАНИЕ ρ - И λ -МЕТОДОВ ПОЛЛАРДА РЕШЕНИЯ ЗАДАЧИ ДИСКРЕТНОГО ЛОГАРИФМИРОВАНИЯ.

Проведен анализ эффективности распараллеливания ρ - и λ -методов Полларда при решении задачи дискретного логарифмирования. Приводится теоретическая оценка времени выполнения задачи на параллельной системе. Проведено сравнение результатов практических и теоретических расчетов. Сделаны замеры времени выполнения распараллеленных методов.

Ключевые слова: методы Полларда, дискретный логарифм, распараллеливание, сложность криптоалгоритма, оценка сложности.

Andrushchenko D. M., Varava M. U., Nelsa G. V.

PARALLELIZATION OF ρ - AND λ - POLLARD'S METHODS FOR SOLVING THE DISCRETE LOGARITHM

The parallelization efficiency of ρ - and λ -methods of Pollard in solving the discrete logarithm is analyzed. The theoretical estimate of the time of the task on a parallel system is given. Comparison of the practical and theoretical calculations carried out. Timing performance threaded methods are made.

Key words: methods of Pollard's, discrete logarithm, paralleling, the complexity of the cryptographic algorithm, estimation of complexity.

УДК:681.5

Кулик А. С.¹, Лученко О. О.², Фирсов С. Н.³

¹Д-р техн. наук, заведующий кафедрой Национального аэрокосмического университета им. М. Е. Жуковского «ХАИ»

²Генеральный директор-Главный конструктор Харtron-Планта,

³Канд. техн. наук, доцент Национального аэрокосмического университета им. М. Е. Жуковского «ХАИ»

КОНЦЕПЦИЯ ОБЕСПЕЧЕНИЯ ЖИВУЧЕСТИ СПУТНИКОВЫХ СИСТЕМ УПРАВЛЕНИЯ ОРИЕНТАЦИЕЙ И СТАБИЛИЗАЦИЕЙ

Сформулированы основные положения обеспечения живучести спутниковых систем ориентации и стабилизации, базирующиеся на принципе самоорганизации посредством глубокого диагностирования аварийного функционального состояния и гибкого восстановления работоспособности объекта.

Ключевые слова: живучесть, диагностирование, нештатная ситуация, самоорганизация.

ВВЕДЕНИЕ

Расширение круга функциональных задач, решаемых современными космическими аппаратами, и увеличе-

ние сроков их активного существования возможно путем обеспечения эффективного и качественного функционирования их бортовых систем как в номинальных

условиях работы, так и при возникновении разнообразных нештатных ситуаций. Обеспечить выполнение целевых функций космического аппарата при возникновении неопределенных нештатных ситуаций возможно путем обеспечения живучести бортовых систем спутника, к которым относится система управления ориентацией и стабилизацией спутника (СУОСС).

В научной литературе традиционно процесс обеспечения живучести СУОСС связывается с повышением надежности работы бортовых систем спутника (безотказностью). Такой подход оправдывает себя, если при рассмотрении СУОСС исключается из рассмотрения сам объект управления, а рассматриваются только информационно-преобразовательные элементы, что само по себе является некорректным, так как объект управления (спутник) представляет собой неотъемлемый элемент контура управления. Понятие живучести значительно шире понятия надежности, так как предусматривает работу СУОСС при наличии нештатных ситуаций, вызванных как внутренними, так и внешними аномальными явлениями. Кроме того, выдвигая предположение о неопределенности отказа и его характеристик, представляющее собой самую «жесткую» гипотезу среди известных, модели и методы, применяемые в теории надежности, становятся не применимы для обеспечения живучести СУОСС, что требует разработки концептуально нового подхода к обеспечению живучести СУОСС.

ПОСТАНОВКА ЗАДАЧИ

Из анализа проблемы обеспечения требуемого времени активного существования спутника на орбите следует, что предпочтительным вариантом для ее решения является подход, связанный с обеспечением живучести СУОСС. СУОСС, обладающие свойством живучести, – это класс адаптивных систем управления с развитыми функциями диагностирования и парирования нештатных ситуаций. Развитая функция диагностирования заключается в глубоком выявлении функциональных причинно-следственных связей возникновения нештатной ситуации с заданной глубиной, другими словами, в поиске устранимых причин отказа СУОСС. В свою очередь развитая функция парирования заключается в гибком управлении избыточными ресурсами, обеспечивающими предотвращение дальнейшего развития нештатной ситуации или такую компенсацию ее последствий, что восстанавливаются измененные или утраченные функциональные свойства системы.

Для решения задач диагностирования и парирования нештатных ситуаций на каждом этапе жизненного цикла СУОСС необходимо формировать множество избыточных ресурсов, обеспечивающих покрытие множества диагностируемых устранимых и неустранимых причин отказов, а также разработать гибкие процедуры диагностирования причин отказа и использования ресурсов для парирования причин отказов, вызвавших нештатную ситуацию.

РАЗРАБОТКА РАЗВИТОЙ ФУНКЦИИ ДИАГНОСТИРОВАНИЯ СУОСС

Способность СУОСС со свойствами живучести самоорганизовываться в нештатных ситуациях базируется на получении достоверной информации о причинах, т. е. на диагнозе. Для получения адекватного диагноза нештатной ситуации требуется разработка «Комплекса взаимосвязанных правил, методов, алгоритмов и средств, необходимых для осуществления диагностирования на всех этапах жизненного цикла объекта» [1], другими словами, диагностического обеспечения. Разработка диагностического обеспечения базируется на формализованном представлении объекта диагностирования, отражающем связь причины нештатной ситуации с ее доступными наблюдениям и измерениям последствиями. Поэтому качество диагностического обеспечения существенно зависит от используемых диагностических моделей. Традиционно в практике разработки и эксплуатации спутниковых систем используют диагностические модели в графической и табличной формах, отражающих только качественную связь между причиной отказа и ее доступным измерению следствием [2–4]. Для количественного отражения причинно-следственных связей предложен новый класс диагностических моделей: диагностические функциональные модели (ДФМ) [5], позволившие формировать принципиально новое диагностическое обеспечение, обладающее следующими характерными особенностями.

Использование ДФМ, отражающих влияние признаков отказа на качество функционирования объекта диагностирования, позволяет описывать нештатные ситуации как в переходных, так и в установившихся режимах функционирования, т. е. в динамических, что позволяет получить существенно большие объемы качественной диагностической информации. ДФМ позволяют отразить изменения качества функционирования, в котором и оценивается техническое состояние объекта диагностирования и выполняемое им функциональное назначение. Применение ДФМ позволяет существенно увеличить глубину диагностирования до устранимой причины отказа с помощью более эффективных избыточных ресурсов. А именно, до конкретного физического вида отказа, что не позволяют сделать известные в теории и используемые на практике диагностические модели. Внедрение диагностического обеспечения на основе ДФМ позволяет найти и сформировать целесообразный с точки зрения современных подходов компромисс (баланс) между эффективностями диагностического и реабилитационного обеспечения спутниковых систем управления ориентацией и стабилизацией, что способствует существенному продлению сроков активного функционирования спутниковых систем.

Диагностическое обеспечение разрабатывается для трехуровневой иерархии (рис. 1). Первый нижний уровень – блоковый. Здесь формируется диагностическое обеспечение (ДО) для блоков датчиков (ДОБД), приводов (ДОБП), вычислителей (ДОБВ) не автономно, а исходя из условий их функционирования в замкнутом контуре.

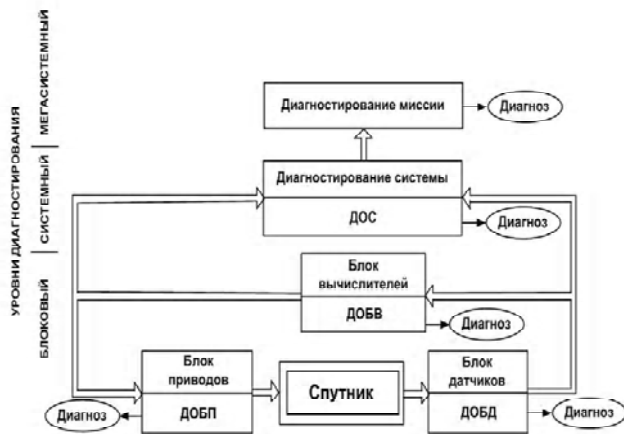


Рис. 1. Функциональная схема иерархического диагностического обеспечения

Второй уровень – системный. Диагностическое обеспечение разрабатывается для всей замкнутой системы управления ориентацией и стабилизацией. Диагностируется выполнение системой ее функций в условиях космического полета. Третий уровень иерархии – мегасистемный. На этом уровне формируется диагностическое обеспечение выполнения космической миссии в целом.

На каждом уровне иерархии при разработке ДО формируются свои специфические для этого уровня подмножества возможных нештатных ситуаций, исходя из целей и задач проекта, опыта разработки, производства и эксплуатации подобных проектов, качества используемых приборов и других факторов, влияющих на качество выполнения космической миссии.

Назначение ДО: оперативная оценка качества функционирования объектов диагностирования в реальном масштабе времени. Качество функционирования объектов диагностирования: блоков, системы в целом, спутника + наземный комплекс сопровождения и управления полетом – определяется требованиями технического задания. На каждом уровне иерархии задача ДО сформировать оперативный диагноз адекватный возникшей нештатной ситуации для последующего эффективного ее парирования с целью обеспечения выполнения поставленных задач каждого этапа жизненного цикла спутниковых систем управления ориентацией и стабилизацией.

Диагностическое обеспечение формируется с использованием трех типов диагностических функциональных моделей: ДФМ для обнаружения отказов, ДФМ для поиска места отказа, ДФМ для определения класса отказа. Эти диагностические модели разрабатываются для решения таких связанных задач диагностирования: обнаружения отказов, поиск места отказа, определение отказа. Последовательное решение таких задач позволяет реализовать принцип последовательного снятия неопределенности. Для каждой задачи формируются множество прямых признаков отказа и множество косвенных. Далее с использованием математических моделей номинального режима функционирования объекта диагностирования строятся математические конструкции,

отражающие связь прямых признаков отказа с косвенными в форме ДФМ. Для ДФМ принципиальным является отражение в полученной структуре имеющихся связей между признаками, а также возможность по доступным измерениям косвенным признакам оценить факт или величину косвенного признака. Для выявления таких свойств ДФМ используются специально разработанные критерии диагностируемости. Структурные свойства ДФМ оцениваются с помощью критериев структурной диагностируемости, а сигнальные свойства ДФМ – с помощью критериев сигнальной диагностируемости. Полная структурная и сигнальная диагностируемость ДФМ гарантирует возможность с помощью этих диагностических моделей получить однозначную информацию для формирования диагноза.

При разработке алгоритмических средств глубокого диагностирования применяются диагностические логические модели (ДЛМ), отражающие бинарную качественную связь между прямыми и косвенными диагностическими признаками объекта. ДЛМ строятся на основании ДФМ с использованием предикатных уравнений по определенным правилам. Применение ДЛМ позволяет ставить и решать задачи минимизации количества косвенных признаков по различным технико-экономическим критериям. Полученные в результате минимизации ДЛМ в табличной форме позволяют перейти к структурам предикатных баз знаний в форме дихотомических деревьев, в узлах которых находятся двужначные предикатные функции, полученные из соответствующих фрагменту дерева ДФМ. Такое дихотомическое дерево с предикатными функциями от дискретных значений косвенных признаков отказов в узлах представляет собой алгоритмическое обеспечение процесса получения полного диагноза функционального состояния объекта диагностирования. Укрупненно получение полного диагноза можно представить с помощью функциональной схемы, отражающей совокупность описанных средств диагностического обеспечения (рис. 2).

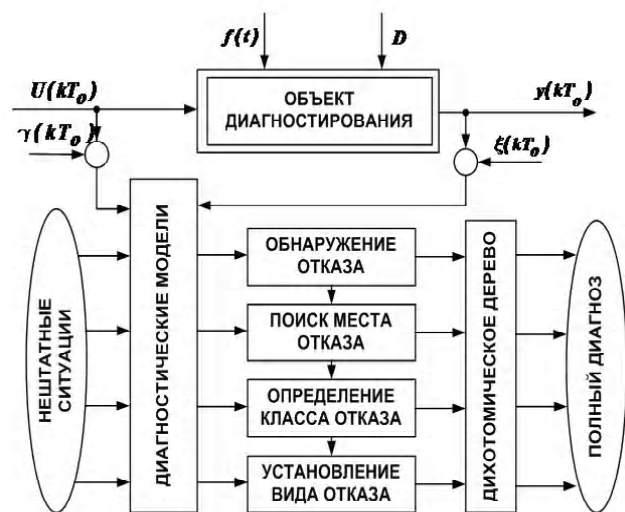


Рис. 2. Функциональная схема средств диагностического обеспечения

РАЗРАБОТКА РАЗВИТОЙ ФУНКЦИИ ПАРИРОВАНИЯ НЕШТАТНЫХ СИТУАЦИЙ СУОСС

Адаптация СУОСС базируется как на получении полного диагноза о функциональном состоянии, вызвавшем нештатную ситуацию, так и на гибком использовании избыточных ресурсов, позволяющих нейтрализовать последствия выявленного вида отказа. Таким образом, для качественного парирования нештатных ситуаций требуется: полный диагноз функционального состояния объекта диагностирования (блока, системы, миссии), отражающий время появления отказа, его места, класса отказа и конкретный физический вид его проявления; наличие избыточных ресурсов, покрывающих множество устранимых видов отказов; процедуры выбора избыточного ресурса, учитывающие предыдущие нештатные ситуации, исходя из текущей ситуации и задач миссии; процедуры эффективного восстановления работоспособности отказавшего объекта диагностирования; средство диагностирования восстановленного объекта диагностирования.

Перспективным подходом к парированию нештатных ситуаций представляется подход с использованием принципа адаптации к особенностям СУОСС, возможным нештатным ситуациям и задачам космической миссии. Особенности такого подхода заключаются в следующем. Парирование нештатных ситуаций производится после обстоятельного, глубокого и качественного диагностирования функционального состояния объектов диагностирования. Другими словами, требуется диагностировать качество выполнения объектом его функционального назначения в режиме выполнения функциональных задач по ориентации и стабилизации спутников. Тестовые режимы диагностирования наиболее распространены в практике диагностирования. В тестовых режимах входные воздействия на объект – это заранее сформированные воздействия, не в полной мере отражающие как характер, так и диагнозы рабочих воздействий. Это обстоятельство приводит к тому, что результаты диагностирования в тестовом режиме, т. е. диагноз, в большей степени отражает способность объекта преобразовывать тестовое воздействие, а не его способность функционировать в замкнутом контуре по выполнению задачи ориентации и стабилизации спутника. Тестовые режимы не позволяют в полной мере учесть влияние связей между блоками, возмущающих воздействий как на связи, так и на блоки, отказы других блоков и подсистем и многое другое, не попадающее в сферу тестового диагностирования и приводящее к частичному, неполному диагнозу.

Вторым необходимым атрибутом качественного парирования нештатных ситуаций в спутниковых системах ориентации и стабилизации является наличие средств восстановления работоспособности. Парирование осуществляется посредством выбора избыточных средств, в наибольшей степени соответствующих сложившейся ситуации в космической миссии, а затем посредством процедур восстановления работоспособности отказав-

шего объекта. Окончательная (завершающая) стадия парирования осуществляется посредством восстановления работоспособности блоков на блоковом уровне, системы в целом на системном уровне, всей миссии на мегами системном уровне. Средства восстановления работоспособности – это различные виды избыточности: функциональной, временной, сигнальной, параметрической, алгоритмической, аппаратной и ряд других. Формирование обоснованного множества избыточных средств – это новая научно-техническая задача. При постановке этой задачи нужно установить связи (установить баланс) между глубиной диагностирования – до устранимого (парируемого) вида отказа – и избыточными ресурсами парирования. Между множеством парируемых видов отказов V и множеством ресурсов парирования R в результате такого исследования должно быть установлено соответствие

$$P: V \rightarrow R. \quad (1)$$

Парирование любого вида отказа из множества возможных – это не тривиальная процедура замены отказавшего фрагмента исправным, а достаточно сложное динамическое восстановление углового положения спутника относительно его центра масс в соответствии с полетным заданием. Каждый вид отказа необходимо рассматривать как специфическое возмущающее воздействие на спутниковую систему ориентации и стабилизации, приводящее к нарушению как требуемых режимов углового состояния спутника, так и технического состояния его системы управления угловым положением. В связи с этим обстоятельством парирование нештатной ситуации, вызванное выявленным в процессе диагностирования конкретным видом отказа, должно производиться в следующей последовательности: вначале выбирается средство для реабилитации ухудшенного технического состояния отказавшего фрагмента, затем с помощью выбранного средства производится восстановление технического состояния, а следовательно, и работоспособности аварийного фрагмента системы. После восстановления работоспособности требуется с помощью соответствующих процедур подстройки восстановить функциональное состояние системы управления до требуемых в данный момент времени характеристик и показателей. Таким образом, ресурсы парирования каждой конкретной нештатной ситуации, вызванной соответствующим конкретным видом отказа, включает: средства реабилитации технического состояния места отказа; процедуры реабилитации; средства компенсации возмущенного углового движения спутника, вызванное видом отказа; процедуры восстановления требуемого, соответствующего полетному заданию функционального состояния спутниковой системы управления угловым положением.

Блоковый уровень включает: блок датчиков, блок приводов и блок вычислителей. На рис. 3 представлена структуризация проблемы гибкого парирования видов отказов в аварийных блоках. Исходным для разрешения про-

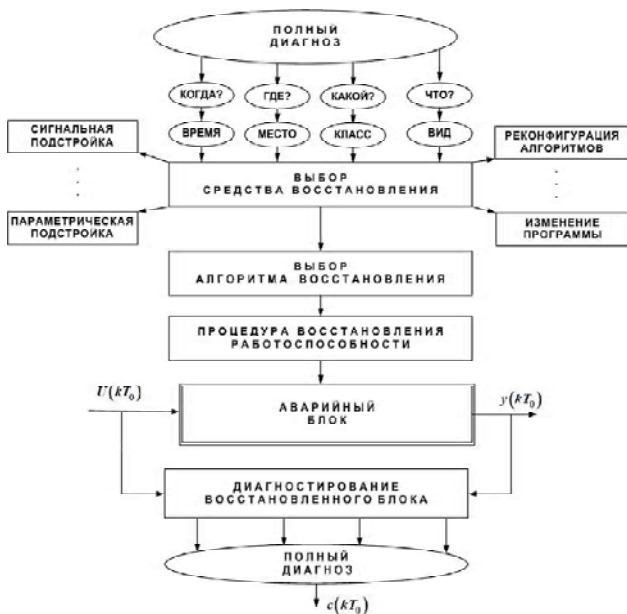


Рис. 3. Структуризация проблемы гибкого парирования видов отказов

блемы есть результат диагностирования функционального состояния блока – полный диагноз. На основании этой информации ставится задача формирования множества ресурсов R парирования множества V – видов отказов конкретного блока. В зависимости от конструкции блока, принципа его действия, множества V подбираются избыточные ресурсы – средства восстановления от различных возможных подстроек до приемлемых средств реконфигурации. Выбор средств восстановления производится на основании предыстории функционирования системы управления, текущего вида отказа, оставшихся избыточных ресурсов и предстоящих задач космической миссии. Выбранное средство восстановления нужно ввести в действие и для этого требуется определенная последовательность действий – алгоритм, который выбирается из множества возможных для текущей ситуации. Затем алгоритм восстановления вводится в действие с помощью соответствующей процедуры, обеспечивающей восстановление работоспособности аварийного блока в пределах (условиях) допустимых полетным заданиям.

После восстановления аварийного блока производится диагностирование его функционального состояния. При этом возможны два исхода. Первый – восстановлена полностью работоспособность блока и его можно использовать в системе управления по назначению. Второй – работоспособность блока не восстановлена и выявлен другой вид отказа, поэтому для парирования этого вида отказа требуется повторить описанный ранее цикл до полного восстановления функциональных свойств аварийного блока. После полного восстановления аварийного блока передается сигнал $c(kT_0)$ на системный уровень, на котором производится парирование системных последствий отказов в блоках с целью последующего выполнения полетного задания.

Отказы в блоках приводят к тому, что спутниковая система управления ориентацией и стабилизацией отклоняется от выполнения своих штатных функций, приводящих к существенному изменению углового положения спутника относительно центра масс по сравнению с заданным программным. Поэтому после выполнения процедур гибкого парирования видов отказов в аварийных блоках системы по сигналам $c(kT_0)$ требуется восстановить программное угловое положение спутника. Причиной нештатных ситуаций на системном уровне являются отказы в межблоковых линиях связи, а также отказы в самой конструкции спутника и существенные внешние возмущающие воздействия. На рис. 4 представлена функциональная схема средств гибкого парирования нештатных ситуаций на системном уровне.

Глубокое диагностирование качества функционирования замкнутой спутниковой системы управления позволяет получить информацию о времени появления отказа, его месте, классе и конкретном физическом проявлении. Анализ возможных системных нештатных ситуаций позволяет сформировать множество видов отказов V системного уровня. Множество избыточных ресурсов для этого уровня R формируется, исходя из необходимости парировать последствия видов отказов множества V . Выбор средств восстановления работоспособности спутниковой системы управления должен производиться с учетом предыдущих системных нештатных ситуаций, оставшихся избыточных ресурсов парирования и предстоящих задач космической миссии.

Выбор средств восстановления работоспособности влечет за собой необходимость выбора алгоритмических средств, т.е. структуры и параметров формализованного действия по парированию текущей нештатной ситуации. После восстановления работоспособности сис-

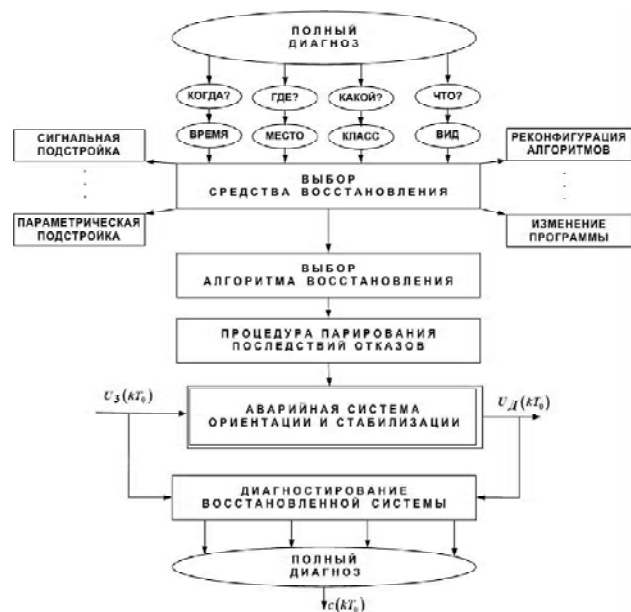


Рис. 4. Функциональная схема средств гибкого парирования системных нештатных ситуаций

темы управления и парирования последствий отказов требуется удостовериться в функциональных свойствах восстановленной системы и качестве выполнения полетного задания. С этой целью производится диагностирование восстановленной спутниковой системы управления ориентацией и стабилизацией. Полный диагноз $c(kT_0)$ позволит принимать решение или о продолжении парирования последствий отказов или о дальнейшем выполнении миссии или о необходимости парирования на мегасистемном уровне.

Отказы на блоковом и системном уровне могут породить такие нештатные ситуации, которые не смогут быть парированы на этих уровнях из-за недостатка избыточных ресурсов, алгоритмических и программных средств, позволяющих парировать возникшую ситуацию и ряда других причин. Мегасистемный уровень парирования нештатных ситуаций – это третий заключительный уровень, на котором обеспечивается выполнения космической миссии в целом. В данном исследовании рассматриваются лишь задачи обеспечения парирования нештатных ситуаций только одной компоненты сложного комплекса, обеспечивающего выполнение космической миссии, а именно, спутниковой системы управления ориентацией и стабилизацией, а не всего комплекса. На рис. 5 представлена совокупность средств, требуемых для парирования нештатных ситуаций на уровне выполнения космической миссии.

После выбора стратегии осуществляется выбор средств ее реализации, исходя из обобщенных критериев всей космической миссии. Использование выбранных средств производится с помощью соответствующей процедуры реализации стратегии через наземный комплекс управления и центр управления полетом по измененному полетному заданию спутника. После оконча-

ния цикла парирования нештатной ситуации в космической миссии производится ее диагностирование с целью оценки текущего состояния посредством полного диагноза. Полный диагноз представляет объективную и обстоятельную информацию о качестве функционирования спутниковой системы управления ориентацией и стабилизацией, необходимую для принятия обоснованного решения о перспективах космической миссии.

Общим, как следует из предыдущего изложения, для каждого уровня парирования является необходимость диагностирования с глубиной до парируемого вида отказа, а также восстановления работоспособности отказавшего объекта. Вместе с тем, появление видов отказов в блоках и системе в целом приводит к накоплению угловых отклонений от полетного задания. Величины этих угловых отклонений существенно зависят не только от периода диагностирования, но и от длительности процедур по восстановлению работоспособности. Накопленные отклонения не позволяют продолжить качественно выполнять полетное задание. Это возмущающие воздействия, дестабилизирующие функционирование спутниковой системы ориентации и стабилизации. Поэтому после процедуры восстановления работоспособности требуется процедура компенсации накопившихся угловых отклонений.

Как следует из приведенных исследований, весь период парирования нештатной ситуации, вызванной каким-либо видом отказа, будет определяться таким выражением

$$T_{\Pi} = \tau_d + \tau_v + \tau_k. \quad (2)$$

Нештатная ситуация – это дестабилизирующее событие для космической миссии, поэтому чем меньше T_{Π} , тем меньше отрицательное влияние такого события на эффективность всего проекта. Вместе с тем, существуют технико-экономические ограничения для каждого проекта, обуславливающие интервальное ограничение для T_{Π} :

$$T_{\Pi \min} \leq T_{\Pi} \leq T_{\Pi \max}, \quad (3)$$

где $T_{\Pi \min}$ – минимально возможный период парирования, а $T_{\Pi \max}$ – максимально возможный период парирования нештатных ситуаций. Очевидно, для того, чтобы вложиться в установленные ограничения для T_{Π} требуется найти рациональное разложение на составляющие τ_d , τ_v , τ_k с учетом их интервального характера. Такое разложение целесообразно свести к оптимизационной задаче.

Итак, предлагаемый подход к разработке реабилитационного обеспечения спутниковых систем управления ориентацией и стабилизацией требует значительных усилий по комплексной формализации процедур восстановления работоспособности блоков, связей и систем, связывающей известные фрагменты аналитического реше-

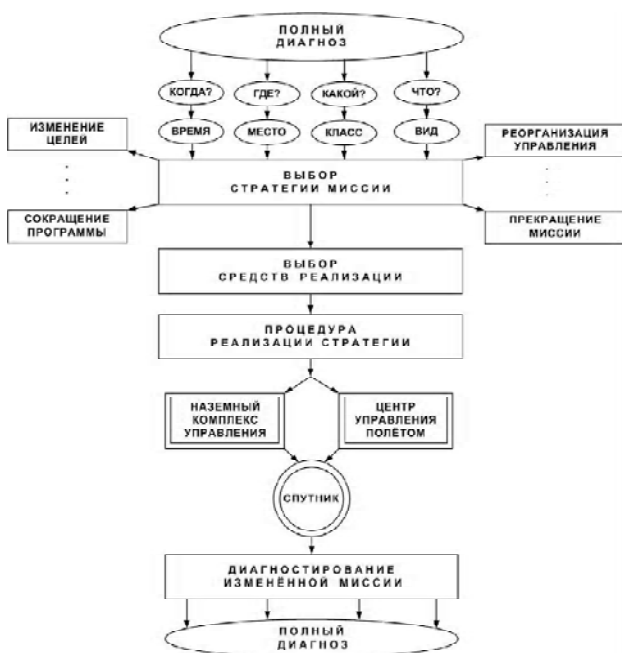


Рис. 5. Функциональная схема средств гибкого парирования отказов миссии

ния отдельных задач в совокупность моделей, методов и средств проектирования. Описанный подход позволяет формализовать разработку реабилитационного обеспечения на платформе гибкого управления избыточными ресурсами.

ВЫВОДЫ

Представленные результаты исследования внедряются в практику создания и эксплуатации живучих спутниковых СУОС на предприятиях, занимающихся разработкой подобных систем, к числу которых относится корпорация «ХАРТРОН». Примерами, демонстрирующими эффективность рассмотренного подхода, являются эксплуатируемые сегодня системы, наделенные свойством живучести на элементном уровне и вновь создаваемые, реализующие системные и мегасистемные уровни живучести космических систем.

СПИСОК ЛИТЕРАТУРЫ

1. *Ахметов, Р. Н.* Методы и модели автономного управления живучестью автоматических космических аппаратов дистанционного зондирования Земли // Р. Н. Ахметов. – Вестник Самарского государственного аэрокосмического университета им. Академика С. П. Королева. – 2008. – №2. – С. 194–210.
2. Космические летательные аппараты. Введение в ракетно-космическую технику: Учеб.пособ. // Ю. Ф. Даниев, А. В. Демченко, В. С. Зевако, А. М. Кулабухов, В. В. Хуторный; под общ. ред. д-ра техн. наук, проф. А. Н. Петренко. – Днепропетровск, 2007. – 521 с.
3. Теоретические основы проектирования информационно-управляющих систем космических аппаратов / В. В. Кульба, Е. А. Микрин, Б. В. Павлов, В. Н. Платонов; под ред.

Е. А. Микрина. Ин-т проблем упр. им. В. А. Трапезникова РАН. – М.: Наука, 2006. – 579 с.

4. Бортовые системы управления космическими аппаратами: Учебные пособия / Бровкин А. Г., Бурдычев Б. Г., Гордийко С. В. и др. Под ред. А. С. Сырова. – М.: Изд-во МАИ-ПРИНТ, 2010. – 340 с.
5. *Лученко О. А.* Определение параметров установки двигателей-маховиков системы ориентации и стабилизации несимметричного малогабаритного космического аппарата / О. А. Лученко, А. Н. Таран, С. Н. Фирсов, В. Н. Постников // Авиационно-космическая техника и технология. – 2011. – №1 (78). – С. 63–68.

Стаття надійшла до редакції 23.05.2011.

Кулік А. С., Лученко О. О., Фірсів С. М.

КОНЦЕПЦІЯ ЗАБЕЗПЕЧЕННЯ ЖИВУЧІСТЮ СУПУТНИКОВИХ СИСТЕМ УПРАВЛІННЯ ОРІЄНТАЦІЄЮ ТА СТАБІЛІЗАЦІЄЮ

Сформульовано основні положення забезпечення живучості супутникових систем орієнтації і стабілізації, що базуються на принципі самоорганізації за допомогою глибокого діагностування аварійного функціонального стану і гнучкого відновлення працездатності об'єкта.

Ключові слова: живучість, діагностування, нештатна ситуація, самоорганізація

Kulik A. S., Luchenko O. O., Firsov S. N.

THE CONCEPT OF SURVIVABILITY OF SATELLITE SYSTEMS, ATTITUDE CONTROL AND STABILIZATION

The fundamentals of position control and stabilization satellite system survivability assurance based on the principle of self-organization through deep diagnosis accident functional conditions and flexible restorative function are presented.

Key words: survivability, diagnosis, worst-case situation, self-organization.

УДК 519.816:004.942

Миронова Н. А.

Асистент Запорізького національного технічного університету

ИНТЕГРАЦИЯ МОДИФИКАЦИЙ МЕТОДА АНАЛИЗА ИЕРАРХИИ ДЛЯ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ ГРУППОВЫХ РЕШЕНИЙ

Проведена систематизация существующих модификаций классического метода анализа иерархий. Выполнена формализация обобщенного подхода анализа иерархий для задач принятия групповых решений.

Ключевые слова: метод анализа иерархий, задача принятия групповых решений, обобщенный подход анализа иерархий, система поддержки принятия групповых решений.

ВВЕДЕНИЕ

В современных условиях большинство важных технических, экономических, социальных, политических решений принимаются в процессе коллективного обсуждения. Групповое принятие решений предпочтительней индивидуального тем, что с увеличением сложности и объема задач один человек не может быть компетентным во всех вопросах и выполнить всей работы по выра-

ботке и реализации решения. Кроме того, отсутствие полной и точной информации, необходимой для принятия решения, приводит к тому, что выбор оптимального исхода происходит именно в результате группового обсуждения руководителями, специалистами, экспертами и консультантами.

Из множества известных методов и подходов к принятию решений наибольший интерес представляют те,

которые дают возможность учитывать многокритериальность и неопределенность, а также позволяют осуществлять выбор решений из множеств альтернатив различного типа при наличии критериев, имеющих разные типы шкал измерения [1]. Одним из таких методов, который учитывает специфику процесса обсуждения проблемной ситуации, является классический метод анализа иерархий (МАИ), предложенный Т. Л. Саати [2].

Целью работы является анализ направлений развития метода анализа иерархий, систематизация существующих модификаций метода, формализация обобщенной модели анализа иерархий для систем поддержки принятия групповых решений.

ПОСТАНОВКА ЗАДАЧИ

Подход, используемый в МАИ [2], предоставляет возможность группе людей взаимодействовать по интересующей их проблеме, модифицировать свои суждения и в результате объединять групповые суждения в соответствии с основным критерием: при проведении попарных сравнений объектов по отношению к некоторой характеристике, или характеристик относительно главной цели. Обратные связи обеспечивают ключ к объединению групповых суждений рациональным образом.

Данный метод позволяет выполнить исследование сложных систем взаимозависимых компонент (ресурсы, желаемые исходы или цели) путем декомпозиции их на отдельные составляющие. Для оценки компонентов систем используется метод последовательного попарного сравнения объектов.

Можно выделить следующие этапы МАИ: построение иерархии, формирование матрицы попарных сравнений (МПС), получение вектора приоритетов, оценка степени согласованности МПС, получение наилучшей альтернативы.

С момента появления МАИ были проведены разнообразные исследования и подробное изучение метода, его преимуществ и потенциальных недостатков, что привело к разработке множества модификаций метода, касающихся, в основном, изменения внутренних процедур (улучшения, усовершенствования определенного этапа) и направленных на исправление присущих классическому методу недостатков.

Для повышения эффективности систем поддержки принятия групповых решений необходимо выполнить:

- анализ развития классического МАИ в соответствии с его этапами;
- формализацию подхода анализа иерархии для задачи группового принятия решения.

АНАЛИЗ РАЗВИТИЯ КЛАССИЧЕСКОГО МЕТОДА АНАЛИЗА ИЕРАРХИЙ

В настоящее время существует множество модификаций МАИ [1–46], которые можно классифицировать относительно следующих критериев:

- по способу измерения предпочтений экспертов (использование различных шкал измерения);
- по типу системы предпочтения экспертов;

- по типу иерархий;
- по методу формирования матриц попарных сравнений;
- по способу оценки и коррекции экспертных суждений;
- по методу синтеза итогового решения;
- в зависимости от структуры метода.

Рассмотрим классификации МАИ относительно вышеприведенных критериев.

1. КЛАССИФИКАЦИЯ МАИ ПО СПОСОБУ ИЗМЕРЕНИЯ ПРЕДПОЧТЕНИЙ ЭКСПЕРТОВ (ИСПОЛЬЗОВАНИЕ РАЗЛИЧНЫХ ШКАЛ ИЗМЕРЕНИЯ)

Один из наиболее существенных недостатков классического МАИ [4] – возможность обработки лишь точечных экспертных оценок, что в большинстве случаев неприемлемо при решении задач технологического предвидения, которые характеризуются наличием концептуальной неопределенности и многофакторных рисков.

Неточность в оценках экспертов и связанные с ней риски можно выразить двумя способами: с помощью точечных оценок и функции распределения вероятностей; с помощью интервальных оценок без распределения вероятностей.

Первичное сравнение элементов с помощью словесной шкалы, затем перевод вербальной оценки с помощью различных шкал измерения предпочтений экспертов (точечные, вероятностное представление точечных оценок и функций распределения, интервальные и нечеткие оценки) приводит к появлению следующих классов метода:

- на основе точечных экспертных оценок (МАИ на основе первого и второго типа иерархий [1], метод анализа сетей [5–6], модификация метода парных сравнений А.А. Павлова [7], МАИ на основе методов парных сравнений относительно стандартов или копированием [1], сокращенная процедура формирования МПС [8–10], модификация метода парных сравнений для случаев с большим количеством оцениваемых параметров [11], способ улучшения согласованности экспертных оценок, заключающийся в изменении структуры иерархии [12], упрощенный вариант МАИ на основе нелинейной свертки критериев: модификация метода собственного вектора [10], идеальный, дистрибутивный и мультипликативный синтез МАИ, метод группового учета бинарных отношений предпочтений альтернатив с дистрибутивным синтезом [13–17], модификация А.П. Ротштейна [18], нечеткий многокритериальный выбор альтернатив: метод наихудшего случая [19]);
- стохастический (стохастический МАИ [20–27]);
- на основе интервальных экспертных оценок (создание модификаций МАИ на основе интервальных оценок [28–30]);
- на основе нечетких экспертных оценок (разработка модификаций МАИ на основе нечетких оценок [31–40], МАС на основе нечетких оценок [41–42]).

Различные способы представления неточности экспертных суждений приводят к необходимости примене-

ния и, следовательно, созданию модифицированных методов получения векторов приоритетов, способов оценки и коррекции экспертных суждений, методов синтеза итогового решения.

Большинство зарубежных и отечественных литературных источников [20–42] посвящено исследованию и разработке модифицированных методов получения вектора приоритетов:

– на основе точечных экспертных оценок: упрощенный вариант МАИ на основе нелинейной свертки критериев: модификация метода собственного вектора [10]; использование энтропии Шеннона для получения вектора приоритетов [41] и др.;

– стохастические [4, 20]: метод имитационного моделирования Монте-Карло, при этом используется лог-нормальное распределение экспертных оценок и метод собственного вектора [21] или взвешенная логит-модель [22–24] и пробит-модель [23], в которых предполагается мультиномиальное распределение весов объектов; регрессионный метод поиска весов и лог-нормальное распределение оценок экспертов [25–27];

– на основе интервальных оценок (интервальные методы получения векторов приоритетов [28–29], интервальный регрессионный анализ [30] и др.);

– на основе нечетких оценок (метод Чанга [37], способы дефазификации [31, 33], нечеткие методы наименьших квадратов: традиционный [38], логарифмический [39] и модифицированный логарифмический [40], метод нечеткого программирования предпочтений и его модификация [36], метод получения вектора приоритетов на основе генетического алгоритма [33] и др.).

Выражение неточности в оценках экспертов в виде нечетких и интервальных чисел приводит к появлению вышеперечисленных классов МАИ, позволяет адекватно представлять неопределенность в суждениях лиц, принимающих решение (ЛПР), и снять ограничения на применение метода.

2. КЛАССИФИКАЦИЯ МАИ ПО ТИПУ СИСТЕМЫ ПРЕДПОЧТЕНИЯ ЭКСПЕРТОВ

Предпочтения могут формироваться одним лицом или коллективом, в зависимости от этого задачи принятия решений с использованием МАИ можно классифицировать на задачи:

- индивидуального принятия решений (классический метод);
- коллективного принятия решений (адаптация метода для задачи принятия групповых решений).

При использовании классического МАИ в процессе принятия групповых решений для нахождения единого группового мнения вычисляют среднегеометрическое оценок участников [1–2]:

$$g_{ij}^A = \sqrt[k]{a_{ij}^1 a_{ij}^2 \dots a_{ij}^k}, \quad (1)$$

где g_{ij}^A – агрегированная групповая оценка, принадлежащая i -й строке и j -му столбцу МПС; k – число МПС,

каждая из которых составлена одним экспертом.

Расчет агрегированной групповой оценки в случае привлечения p ЛПР, у которых различные уровни компетентности, выполняется по формуле:

$$g_{ij}^A = a_{ij}^{\alpha_1} a_{ij}^{\alpha_2} \dots a_{ij}^{\alpha_p}, \quad (2)$$

где $a_{ij}^{\alpha_k}$ – оценка элемента k -м участником с коэффициентом компетентности α_k , $k = \overline{1, p}$, где p – количество участников,

$$\alpha_1 + \alpha_2 + \dots + \alpha_p = 1.$$

Агрегацию суждений участников можно осуществить также с помощью собственных векторов МПС. При этом результаты будут эквивалентны тем, которые получены на уровне элементов матриц, если однородность составленных матриц достаточна и удовлетворяет условию $CR \leq 0,10$ (CR – отношение согласованности).

В пункте 1 приведены модификации МАИ (классы МАИ), которые можно применить только для задач индивидуального принятия решений. Для того, чтобы применить существующие модификации МАИ для задач группового принятия решений необходимо выполнить адаптацию методов на этапе получения обобщенного мнения.

3. КЛАССИФИКАЦИЯ МАИ ПО ТИПУ ИЕРАРХИИ

По характеру связей между критериями и альтернативами выделяют следующие типы иерархий [1–2]:

- первый тип иерархий с одинаковым числом и функциональным составом альтернатив под критериями (каждый критерий, имеющий связь с альтернативами, связан со всеми рассматриваемыми альтернативами);
- второй тип иерархий с различными числом и функциональным составом альтернатив под критериями (каждый критерий, имеющий связь с альтернативами, связан не со всеми рассматриваемыми альтернативами);
- третий тип иерархий с обратной связью.

Для первого типа иерархии разработан алгоритм иерархического синтеза [1], который используется для взвешивания собственных векторов МПС альтернатив весами критериев (элементов), имеющихся в иерархии, а также для вычисления суммы по всем соответствующим взвешенным компонентам собственных векторов нижележащего уровня иерархии. Данная модификация позволяет получить итоговое решение для более детализированной иерархии (например, когда в иерархии присутствуют критерии нижних уровней – подкритерии).

Для второго типа иерархии используется метод определения вектора приоритета альтернатив [1] для случая, когда иерархия имеет один уровень критериев, объединенных фокусом с учетом значимости критериев, и разное количество альтернатив у каждого критерия. Метод предполагает выполнение ряда процедур по структурированию информации и проведению вычислительных операций.

Для третьего типа иерархии разработан метод анализа сетей (МАС) [2, 5–6], который является обобщением МАИ в случае, когда взаимодействием компонентов иерархии и/или их элементов нельзя пренебречь. Многие проблемы принятия решений нельзя представить иерархическими структурами, потому что в них существуют зависимости и взаимодействия между элементами разных уровней иерархии. Часто возникают задачи, в которых важность критериев влияет не только на приоритеты альтернатив (как в иерархиях), но и на приоритеты критериев [5].

МАС основан на построении сетевой модели, описывающей влияние внешних факторов на рассматриваемую систему и взаимное влияние основных элементов системы друг на друга (относительно выбранной цели). Ключевым понятием в рамках данного метода является влияние, т.е. воздействие одного элемента на другой.

4. КЛАССИФИКАЦИЯ МАИ ПО МЕТОДУ ПОПАРНЫХ СРАВНЕНИЙ

Метод попарного сравнения, используемый в классическом МАИ, не всегда может быть эффективно применен в некоторых практических ситуациях [1]:

- оценки и сравнения более девяти объектов (критериев, альтернатив), в таком случае построение однородных МПС становится затруднительным. Это связано с физическими ограничениями интеллекта человека;
- возникновения явления реверса рангов, т.е. изменение порядка ранее прошедших сравнение альтернатив при добавлении новых альтернатив или удалении существующих. Реверс рангов нежелателен при решении ряда прикладных задач, связанных со значительными финансовыми, материальными и социальными затратами на корректировку последствий принимаемых решений или возможностью возникновения конфликтной ситуации между экспертами, готовящими и обосновывающими решения, и ЛПР, несущими ответственность за принятые решения и их последствия;
- сравнения объектов, которые могут поступать эксперту не одновременно, а через определенные промежутки времени; в данной ситуации не представляется возможным попарно сравнить объекты.

Для решения проблемы сравнения и оценки объектов в вышеперечисленных ситуациях наиболее целесообразен *метод сравнения объектов относительно стандартов* [1].

Метод копирования [1] аналогичен методу сравнения объектов относительно стандартов в том плане, что позволяет не нарушать порядок ранее проранжированных альтернатив при добавлении новых, являющихся копиями ранее проранжированных альтернатив. Кроме того, число анализируемых альтернатив при добавлении копий может превышать пороговое значение, равное девяти, установленное для метода попарного сравнения. Данный метод позволяет существенно сократить время экспертов на подготовку исходных данных для анализа

и уменьшить вероятность внесения в них как случайных, так и логических ошибок.

Сокращенная процедура формирования МПС [8–10] обеспечивает устранение следующего недостатка классического МАИ – резкое увеличение количества оценок с увеличением набора элементов; и, как следствие, уменьшение трудоемкости формирования матриц и повышение эффективности их обработки. Процедура позволяет автоматически формировать согласованную МПС («идеальную») на основании оценок эксперта, вынесенных только относительно первого объекта сравнения (в этом случае эксперт выносит $n-1$ суждение). Такую матрицу можно рассматривать как своего рода опорную матрицу при вынесении экспертных суждений.

Модификация метода парных сравнений для случаев с большим количеством оцениваемых параметров [11] основана на использовании сокращенной процедуры формирования МПС [8–10] и разделении результатов экспертиз случайным образом на группы, затем для каждой группы и выборки в целом рассчитывают вектор приоритетов на основании среднеарифметического, медианы и среднегеометрического оценок экспертов.

Модификация метода парных сравнений А. А. Павлова [7] используется в случае большого количества альтернатив (больше 9). Такая задача может возникнуть в двух случаях:

- наилучшая альтернатива не выбирается из набора реально существующих альтернатив, а альтернативы генерируются искусственно для выбора наилучшей, после чего в реализацию этой альтернативы вкладываются существенные ресурсы;
- искусственно генерируются альтернативы; с помощью МАИ находят их результирующие веса, по которым строится аналитическое описание глобальной цели.

5. КЛАССИФИКАЦИЯ МАИ ПО СПОСОБУ ОЦЕНКИ И КОРРЕКЦИИ ЭКСПЕРТНЫХ СУЖДЕНИЙ

Сравнение элементов иерархии осуществляется методом попарных сравнений: эксперты сравнивают альтернативы между собой и дают оценки предпочтения сравниваемых вариантов выбора, которые выражают в так называемой фундаментальной шкале отношений. Результат записывается в виде МПС. На основе полученных МПС формируется вектор приоритетов. Далее выполняется оценка степени согласованности МПС, определяемая таким показателем, как отношение согласованности, которое сравнивается с заданным ЛПР порогом. Если результат сравнения превышает данное число (порог), то это свидетельствует о существенном нарушении логической связности суждений, допущенном экспертом при заполнении матрицы, поэтому исследователю предлагается пересмотреть данные, использованные для построения матрицы, чтобы улучшить согласованность.

Существуют следующие способы оценки и коррекции экспертных суждений:

- 1) *в зависимости от структуры метода:*

– изменение структуры иерархии (способ улучшения согласованности экспертных оценок, заключающийся в изменении структуры иерархии – уменьшение числа элементов на одном уровне иерархии с помощью построения дополнительного, на котором предлагается разместить наиболее близкие между собой элементы рассматриваемого уровня; определение степени близости элементов не поддается формализации и зависит от специфики конкретной задачи [12]);

– пересмотр оценок в МПС (классический метод Саати [2, 43], метод на основе графа [43], метод триплетов [44], улучшенный способ определения несогласованности экспертных суждений [9]);

2) по типу оценок:

– точечные (классический метод Саати [2, 43], метод на основе графа [43], метод триплетов [44], улучшенный способ определения несогласованности экспертных суждений [9]);

– нечеткие (способы оценки согласованности суждений на основе нечетких экспертных оценок [4, 35, 36]);

3) по способу пересмотра:

– единоличный пересмотр значений оценок самим ЛПР [2, 43];

– итеративные (классический метод Саати [2, 43], метод на основе графа [43], метод триплетов [44]).

Существующие способы оценки и коррекции экспертных суждений, в основном, позволяют выполнять коррекцию точечных оценок и не охватывают проверку согласованности нечетких и интервальных матриц попарных сравнений.

6. КЛАССИФИКАЦИЯ МАИ ПО МЕТОДУ СИНТЕЗА ИТОГОВОГО РЕШЕНИЯ

Наиболее простой прием решения многокритериальной задачи заключается в ее сведении к определенной скалярной (однокритериальной) задаче, целевая функция которой представляет собой некоторую комбинацию исходных критериев. При этом используется линейная, мультипликативная и другие формы (свертки) для образования указанной комбинации. Выбор той или иной свертки осуществляется на основе имеющейся информации о решаемой задаче и/или исходя из соображений простоты решения полученной скалярной задачи [45].

Применение МАИ к решению многокритериальной задачи [45] основано на ее скаляризации при помощи линейной свертки критериев. А именно, согласно МАИ наилучшим решением многокритериальной задачи объявляется то, которое доставляет наибольшее возможное значение линейной (аддитивной) свертке критериев.

Применение линейной свертки [10] при решении многокритериальных задач с конечным множеством возможных решений (условие выпуклости множества возможных решений и требование вогнутости критериев), по крайней мере, гипотетически дает возможность обнаружить любое решение, которое может оказаться выбранным, т.е. наилучшим. За пределами указанного

класса применение этой свертки рискованно или же вообще недопустимо, поскольку в результате ее максимизации (при всех возможных значениях неотрицательных коэффициентов) можно пропустить именно то решение, которое в данном случае следует выбрать.

Причина появления реверса рангов в МАИ заключается в использовании линейной свертки альтернатив относительно критериев принятия решений. Более универсальной с точки зрения области применимости является нелинейная свертка [10].

Следовательно, выделим классы МАИ по методу синтеза итогового решения:

1) по способу свертки критериев:

– линейная или аддитивная (идеальный и дистрибутивный синтезы МАИ, метод группового учета бинарных отношений предпочтений альтернатив с дистрибутивным синтезом [13–14]);

– мультипликативная (мультипликативный МАИ [14–17]);

– нелинейная (упрощенный вариант МАИ на основе нелинейной свертки [10]).

2) по типу модели:

– иерархический синтез (получение итогового решения для модификации МАИ с учетом типов иерархий [1]);

– синтез для сетевой модели (метод анализа сетей [5–6]);

– специальные соотношения (нечеткий многокритериальный выбор альтернатив: метод наихудшего случая [18], модифицированный метод А. П. Ротштейна [19]);

3) по типу оценок:

– на основе точечных оценок (МАИ на основе первого и второго типа иерархий [1], МАС [5], МАИ на основе методов парных сравнений относительно стандартов или копированием [1], модификация метода парных сравнений А. А. Павлова [7], сокращенная процедура формирования МПС [8–10], модификация метода парных сравнений для случаев с большим количеством оцениваемых параметров [11], способ улучшения согласованности экспертных оценок, заключающийся в изменении структуры иерархии [12], упрощенный вариант МАИ на основе нелинейной свертки критериев: модификация метода собственного вектора [10], идеальный, дистрибутивный и мультипликативный синтез МАИ, метод группового учета бинарных отношений предпочтений альтернатив с дистрибутивным синтезом [13–17], модификация А. П. Ротштейна [18], нечеткий многокритериальный выбор альтернатив: метод наихудшего случая [19]);

– на основе нечетких оценок (модификации идеального, дистрибутивного и мультипликативного синтеза МАИ, метода группового учета бинарных отношений предпочтений альтернатив с дистрибутивным синтезом [46]).

Существующие классы МАИ по методу синтеза итогового решения позволяют получать наилучшую альтернативу для МАИ на основе точечных оценок. Для нечетких и интервальных МАИ необходимо выполнить разработку методов получения итогового решения на основе нелинейной свертки.

7. КЛАССИФИКАЦИЯ МАИ В ЗАВИСИМОСТИ ОТ СТРУКТУРЫ МЕТОДА

В зависимости от структуры метода можно выделить следующие модификации:

- отсутствует этап формирования МПС (нечеткий многокритериальный выбор альтернатив: метод наихудшего случая [18] не требует трудоемких процедур, связанных с формированием и обработкой МПС; вместо этого используются специальные соотношения, основанные на сравнении с наихудшей альтернативой и наименее важным критерием);

- отсутствует этап получения вектора приоритета (модифицированный метод А.П.Ротштейна [19] не требует процедуры определения собственного вектора и проверки согласованности МПС);

- отсутствует этап оценки степени согласованности МПС (МАС [5–6], модифицированный метод А. П. Ротштейна [19]).

ФОРМАЛИЗАЦИЯ ПОДХОДА АНАЛИЗА ИЕРАРХИИ ДЛЯ ЗАДАЧИ ГРУППОВОГО ПРИНЯТИЯ РЕШЕНИЙ

Исходя из вышеизложенного, задачу принятия группового решения с использованием модификаций МАИ можно представить в виде следующего кортежа:

$$\langle S_0, T, Q | TH(S, A, B, X), \theta, M(TH, \theta), WPO(\theta), \xi(\theta), OA(\theta), G, L, X^* \rangle,$$

где слева от «|» стоят известные переменные, а справа – неизвестные; S_0 – проблемная ситуация описывается количественными или/и качественными характеристиками; описываются условия, связанные с проблемой, причины их возникновения и развития; описание заканчивается кратким формулированием проблемы, которую необходимо решить; T – время, данное для принятия решения; Q – ресурсы, необходимые для принятия решения, не включая ресурсы на реализацию; $TH(S, A, B, X)$ – тип иерархий, представленный в виде графа и матрицы смежности; S – множество признаков или критериев $S = (s_1, s_2, \dots, s_m)$, доопределяющее проблемную ситуацию; A – множество целей $A = (a_1, a_2, \dots, a_k)$; B – множество ограничений $B = (b_1, b_2, \dots, b_n)$; X – множество альтернативных решений $X = (x_1, x_2, \dots, x_n)$, из которых выбирается единственное оптимальное решение X^* ; с целью увеличения степени общности результатов необходимо выбирать любые оцениваемые альтернативные решения как универсальные объекты, характеризующиеся набором количественных или/и качественных показателей; θ – шкала измерения оценок экспертов; $M(TH, \theta)$ – множество МПС, зависящих от типа иерархий, количества объектов в иерархии (цели, критерии, подкритерии, альтернативы), способа формирования МПС; $WPO(\theta)$ – метод обработки МПС (метод получения вектора приоритетов) в зависимости от шкалы измерения оценок эксперта; $\xi(\theta)$ – способ оценки МПС, зависит от типа

оценок, способа пересмотра, структуры метода коррекции суждений экспертов; $OA(\theta)$ – метод синтеза итогового решения, зависит от типа модели, способа свертки критериев и шкалы измерения оценок эксперта; G – единое групповое предпочтение, определяется как функция от индивидуальных предпочтений: $G = f(\theta, X, L)$, где L – решающее правило (способ), позволяющее согласовать индивидуальные предпочтения в единое групповое предпочтение, $G = (g_1, g_2, \dots, g_p)$, p – количество ЛПР в группе.

ВЫВОДЫ

В работе решена актуальная задача исследования направлений развития классического метода анализа иерархий, выполнена систематизация существующих модификаций данного метода, в результате которой были выделены классы МАИ относительно следующих критериев: способ измерения предпочтений экспертов (использование различных шкал измерения); тип системы предпочтения экспертов; тип иерархий; метод формирования матриц попарных сравнений; способ оценки и коррекции экспертных суждений; метод синтеза итогового решения; в зависимости от структуры метода.

Научная новизна работы состоит в том, что впервые предложена формализация обобщенного подхода анализа иерархий, описывающая задачу принятия групповых решений с помощью модификаций МАИ, которая позволяет формировать этапы метода в зависимости от исходных данных.

СПИСОК ЛІТЕРАТУРИ

1. Андрейчиков, А. В. Анализ, синтез, планирование решений в экономике [Текст] / А. В. Андрейчиков, О. Н. Андрейчикова. – М. : Финансы и статистика, 2002. – 368 с.
2. Саати, Т. Л. Принятие решений. Метод анализа иерархий [Текст] / Т. Л. Саати. – М. : Радио и связь, 1993. – 320 с.
3. Миронова, Н. А. Обзор методов аналитической иерархии [Текст] / Н. А. Миронова, В. И. Дубровин // Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта: Материалы международной научной конференции. – Том 1. – Херсон : ХНТУ, 2009. – С. 192–194.
4. Панкратова, Н. Д. Методология обработки нечеткой экспертной информации в задачах предвидения. Ч. 1 [Текст] / Н. Д. Панкратова, Н. И. Недашковская // Проблемы управления и информатики. – 2007. – № 2. – С. 40–55.
5. Андрейчиков, А. В. Принятие решений при зависимостях и обратных связях: Аналитические сети [Текст] / А. В. Андрейчиков, О. Н. Андрейчикова. – М. : Издательство ЛКИ, 2008. – 360 с.
6. Saaty, T. L. Decision making with Dependence and Feedback. The Analytic Network Process. [Текст] / T. L. Saaty; PWS Publications. – Pittsburgh: 2001. – 386 p.
7. Модифікований метод аналізу ієрархій (версія 1, 2) [Текст] / О. А. Павлов, К. І. Ліщук, О. С. Штанькевич, Г. А. Іванова, О. П. Федотов // Вісник Національного технічного університету України «Київський політехнічний інститут». Серія «Інформатика, управління та обчислювальна техніка». – 2009. – № 51. – С. 42–53.

8. Трофимец, В. Я. К вопросу разработки основных вычислительных процедур МАИ : [Электрон. ресурс]. – Режим доступа: <http://zhurnal.ape.relarn.ru/articles/2004/078.pdf>
9. Трофимец, В. Я. К вопросу разработки вспомогательных вычислительных процедур МАИ: [Электрон. ресурс]. – Режим доступа: <http://zhurnal.ape.relarn.ru/articles/2004/079.pdf>
10. Ногин, В. Д. Упрощенный вариант МАИ на основе нелинейной свертки критериев [Текст] / В. Д. Ногин // Журнал вычислительной математики и математической физики. – 2004. – Т. 44. – № 7. – С. 1259–1268.
11. Сайко, В. В. Модификация метода парных сравнений для случаев с большим количеством оцениваемых параметров [Текст] / В. В. Сайко // Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта: Материалы международной научной конференции. – Том 1. – Херсон : ХНТУ, 2009. – С. 210–214.
12. Панкратова, Н. Д. Комплексне оцінювання чутливості рішення на основі методу аналізу ієрархій [Текст] / Н. Д. Панкратова, Н. І. Недашківська // Системні дослідження та інформаційні технології. – 2006. – № 3. – С. 7–25.
13. Недашківська, Н. І. Оцінювання реверсу рангів у методі аналізу ієрархій [Текст] / Н. І. Недашківська // Системні дослідження та інформаційні технології. – 2005. – № 4. – С. 120–130.
14. Triantaphyllou, E. Two new cases o rank reversals when AHP and some of its additive variants are used that do not occur with multiplicative AHP [Текст] / E. Triantaphyllou // Journal of multicriteria decision analysis. – 2001. – № 10. – P. 11–25.
15. Lootsma, F. A. Scale sensitivity in the multiplicative AHP and SMART [Текст] / F. A. Lootsma // Journal of Multi-Criteria Decision Analysis. – 1993. – № 2. – P. 87–110.
16. Lootsma, F. A. The multiplicative AHP, SMART and ELECTRE in a common contex [Текст] / F. A. Lootsma, H. Schuijt // Journal of Multi-Criteria Decision Analysis. – 1997. – vol 6. – № 4. – P. 185–196.
17. Ларичев, О. И. Теория и методы принятия решений [Текст] / О. И. Ларичев. – М. : Логос, 2002. – 392 с.
18. Ротштейн, А. П. Нечеткий многокритериальный выбор альтернатив: метод наихудшего случая [Текст] / А. П. Ротштейн, А. Н. Козачко // Автоматика-2008 : доклады XV международной конференции по автоматическому управлению, 23–26 сентября 2008 г. – Одесса : ОНМА. – С. 488–491.
19. Rotshtein, A. P. Modification of the Saaty Method for the Construction of Fuzzy Set Membership Functions [Текст] / A. P. Rotshtein // The International Conference on Fuzzy Logic and its Applications : Proc. Zichron, Israel, 1997. – P. 125–130.
20. Banuelas, R. Modified analytic hierarchy process to incorporate uncertainty and managerial aspects [Текст] / R. Banuelas, J. Antony // International Journal of Production Research. – 2004. – vol 42. – №18. – P. 3851–3872.
21. Banuelas, R. Six sigma or design for six sigma [Текст] / R. Banuelas, J. Antony // The TQM magazine. – 2004. – 16(4). – P.250–263.
22. Hahn, E. D. Decision making with uncertain judgments: a stochastic formulation of the analytic hierarchy process [Текст] / E. D. Hahn // Decision Sciences. – 2003. – № 34(3). – P. 443–466.
23. Hahn, E. D. Link function selection in stochastic multicriteria decision making models [Текст] / E. D. Hahn // European Journal of Operational Research. – 2006. – № 172(1). – P. 86–100.
24. Phillips-Wren, G. E. A multiple-criteria framework for evaluation of decision support systems [Текст] / G.E. Phillips-Wren, E.D. Hahn, G.A. Forgonne // Omega. – 2004. – № 32(4). – P. 323–332.
25. Laininen, P. Analyzing AHP-matrices by regression [Текст] / P. Laininen, R. P. Hamalainen // European Journal of Operational Research. – 2003. – № 148. – P. 514–524.
26. Crawford, G. A note on the analysis of subjective judgment matrices [Текст] / G. Crawford, C. Williams // Journal of Mathematical Psychology. – 1985. – № 29(4). – P. 387–405.
27. De Jong, P. A statistical approach to Saaty's scaling method for priorities [Текст] / P. De Jong // Journal of Mathematical Psychology. – 1984. – № 28. – P. 467–478.
28. Kazutomi Sugihara Interval priorities in AHP by interval regression analysis [Текст] / Kazutomi Sugihara, Hiroaki Ishii and Hideo Tanaka // European Journal of Operational Research. – 2004. – № 158(3). – P. 745–754.
29. MeiMei Xia Research on the Priority Method of Interval Fuzzy [Текст] / MeiMei Xia, HeFeng Jiang and CuiPing Wei // Applied Mathematical Sciences. – 2009. – № 3(17). – P. 839–843.
30. Lipovetsky, S. Interval estimation of priorities in the AHP [Текст] / Stan Lipovetsky, Asher Tishler // European Journal of Operational Research. – 1999. – № 114(1). – P. 153–164.
31. Ting-Yu Chen Importance-Assessing Method with Fuzzy Number-Valued Fuzzy Measures and Discussions on TFNs And TrFNs [Текст] / Ting-Yu Chen, Tai-Chun Ku // International Journal of Fuzzy Systems. – 2008. – vol.10. – № 2. – P. 104–111.
32. Ахрамейко, А. А. Обобщение метода анализа иерархий Саати для использования нечетко-интервальных экспертных данных: [Электрон. ресурс]. – Режим доступа: http://sedok.narod.ru/_s_files/belorussia_2002.htm.
33. Миронова, Н. А. Метод получения вектора приоритетов из нечетких матриц попарных сравнений [Текст] / Н. А. Миронова, В. И. Дубровин // Искусственный интеллект. – 2009. – № 3. – С. 464–470.
34. Панкратова, Н. Д. Розробка модифікованого методу аналізу ієрархій з урахуванням нечіткої експертної інформації [Текст] / Н. Д. Панкратова, Н. І. Недашківська // Матеріали XIV міжнародної конференції з автоматичного управління (Автоматика-2007). – Ч. 2 – Севастополь : СКУЯС та П. – 2007. – С. 56–58.
35. Панкратова, Н. Д. Методология обработки нечеткой экспертной информации в задачах предвидения. Ч. 2 [Текст] / Н. Д. Панкратова, Н. И. Недашкова // Проблемы управления и информатики. – 2007. – № 3. – С. 49–63.
36. Mikhailov, L. Evaluation of services using a fuzzy Analytic Hierarchy Process [Текст] / L. Mikhailov, P. Tsvetinov // Applied Soft Computing. – 2004. – № 5. – P. 23–33.
37. Chang, D.Y. Applications of the extent analysis method on fuzzy AHP [Текст] / D. Y. Chang // European Journal of Operational Research. – 1996. – vol 95. – №3. – P. 649–655.
38. Ruoning, Xu Fuzzy least-squares priority method in the analytic hierarchy process [Текст] / Ruoning Xu // Fuzzy Sets and Systems. – 2000. – № 3(112). – P. 395–404.
39. Ruoning Xu Fuzzy logarithmic least squares ranking method in analytic hierarchy process [Текст] / Ruoning Xu, Zhai Xiaoyan // Fuzzy Sets and Systems. – 1996. – № 2(77). – P. 175–190.
40. Ying-Ming, Wang A modified fuzzy logarithmic least squares method for fuzzy analytic hierarchy process [Текст] / Ying-Ming Wang, M.S. Taha Elhag, Zhongsheng Hua // Fuzzy Sets and Systems. – 2006. – № 23(157). – P. 3055–3071.

41. Prabjot, K. A fuzzy ANP-based approach for selection ERP vendors [Текст] / K. Prabjot, N.C. Mahanti // International Journal of Soft Computing. – 2008. – № 3(1). – Р. 24–32.
42. Cheng, R. W. A fuzzy ANP-based approach to Evaluate Medical Organizational Performance [Текст] / R. W. Cheng, Che-Wei Chang, Hung-Lung Lin // International and Management Sciences. – 2008. – vol. 19. – № 1. – Р. 53–74.
43. Кораблев, Н. М. Согласование и коррекция экспертных оценок в системах поддержки принятия решений в условиях нечеткой исходной информации [Текст] / Н.М. Кораблев, С.Г. Удовенко, Альзин Фирас // Радиоэлектроника. Информатика. Управление. – 2005. – №2. – С. 116–120.
44. Lamata, M. T. A method for improving the consistency of judgements [Текст] / M. T. Lamata, J.I. Pelaez // International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. – 2002. – Vol 10. – № 6. – Р. 677–686.
45. Ногин, В. Д. Границы применимости распространенных методов скаляризации при решении задач многокритериального выбора [Текст] / В.Д. Ногин // Методы возмущений в гомологической алгебре и динамика систем: межвуз. сб. науч. тр. – Саранск: Изд-во Мордов. ун-та, 2004. – С. 59–68.
46. Миронова, Н. О. Модифіковані методи отримання найкращої альтернативи для методу аналізу ієрархій на основі нечітких експертних оцінок [Текст] / Н.О. Миронова, В.І. Дубровін // Тези доповідей IV міжнародної наукової

конференції «Комп'ютерні науки та інженерія 2010 (CSE-2010)». – Львів : Львівська політехніка, 2010. – С. 246–247.

Стаття надійшла до редакції 12.01.2011.

Після доробки 23.05.2011.

Миронова Н. О.

ІНТЕГРАЦІЯ МОДИФІКАЦІЙ МЕТОДУ АНАЛІЗУ ІЄРАХІЙ ДЛЯ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ ГРУПОВИХ РІШЕНЬ

Проведена систематизація існуючих модифікацій класичного методу аналізу ієрархій. Виконана формалізація узагальненого підходу аналізу ієрархії для задач прийняття групових рішень.

Ключові слова: метод аналізу ієрархій, задача прийняття групових рішень, узагальнений підхід аналізу ієрархій, система підтримки прийняття групових рішень.

Mironova N. O.

INTEGRATION MODIFICATIONS OF THE ANALYTICAL HIERARCHY PROCESS FOR GROUP DECISION MAKING SUPPORT SYSTEMS

The systematization of Analytic Hierarchy Process existent modification is developed. The formalization of Analytic Hierarchy Process generalized approach for group decision making problems is solved.

Key words: Analytic Hierarchy Process, group decision making problem, Analytic Hierarchy Process generalized approach, Group Decision Support Systems.

УДК 681.3

Кошевой Н. Д.¹, Михайлов А. Г.².

¹Д-р. техн. наук, заведующий кафедрой Национального аэрокосмического университета им. Н. Е. Жуковского «Харьковский авиационный институт»

² Канд. техн. наук, доцент Национального аэрокосмического университета им. Н. Е. Жуковского «Харьковский авиационный институт»

РАЗРАБОТКА АЛГОРИТМОВ МОДЕЛИРОВАНИЯ РАБОТЫ ДАТЧИКОВ РАСХОДА ТОПЛИВА И ВЗАИМОДЕЙСТВИЯ С ТОПЛИВНОЙ МАГИСТРАЛЬЮ

В статье рассмотрены принципы построения алгоритмов моделирования работы датчиков расхода топлива. Представлены динамические модели функционирования систем измерения параметров энергоносителей.

Ключевые слова: математические модели, алгоритмы моделирования, датчики, расход топлива.

АКТУАЛЬНОСТЬ ПРОБЛЕМЫ

Существующие системы проектирования измерительных преобразователей (ИП) расхода топлива преимущественно ориентированы на устранение погрешностей приборов, связанных с нелинейностью статических характеристик [1].

В то же время обзор статистических данных показал, что динамические параметры имеют широкую спектральную характеристику и, следовательно, приводят к возникновению динамических погрешностей при совпадении спектров частот проектируемых ИП и топливных магистралей (ТМ) [2].

ПОСТАНОВКА ЗАДАЧИ

При разработке и модернизации системы проектирования ИП расхода топлива возникает необходимость совершенствования методологии проектирования информационно-измерительных систем (ИИС), новых устройств контроля параметров энергоносителей, которые будут применяться в летательных аппаратах (ЛА) или наземных магистральных топливных энергокомплексах (ТЭК). Предлагается выполнить совершенствование в следующих направлениях:

– разработать математические модели функционирования ИИС контроля параметров энергоносителей, которые будут учитывать динамические параметры дат-

чиков и широкую спектральную характеристику ТМ, что позволит минимизировать динамические погрешности проектируемых ИП;

– разработать общие алгоритмы моделирования ИИС контроля параметров энергоносителей ЛА и ТЭК.

ПРЕДЛАГАЕМЫЕ РЕШЕНИЯ

Рассмотрим типовые характеристики (рис. 1) изменения давления p газовой ТМ для различных интервалов времени [2]. Как видно из рисунка, частотный спектр включает в себя как низкочастотную составляющую (помпаж), так и высокочастотную.

Для исследования динамических погрешностей проектируемых ИП и моделирования их работы необходимо сформировать некоторую модель частотных возмущений $M_{FrErr}\{V_{HF}, V_{LF}\}$. Входными переменными данной модели являются коэффициенты влияния высокочастотных и низкочастотных возмущений V_{HF} и V_{LF} соответственно.

С целью минимизации динамических погрешностей в частотной области функционирования датчика необходимо выделить резонансную характеристику проектируемого ИП [3] в области существования двумерного вектора состояния X

$$X = \text{fres}\{m, ku, H, v\}, \quad (1)$$

где m – масса чувствительного элемента; ku – коэффициент упругости пружинного элемента; H – обобщенный геометрический параметр, характеризующий отношение площадей сечения ТМ и чувствительного элемента; v – вязкость рабочего измерительного тела (газа).

Геометрический критерий Cn проектирования ИП будет основан на минимизации дисперсии D ошибок измерений магистрального давления p :

$$\min_D Cn\{l, k, p, W, D\}, \quad (2)$$

где l – набор расстояний между потребителями; k – число потребителей; p – давление на входе потребителя; W – мощность потребителя.

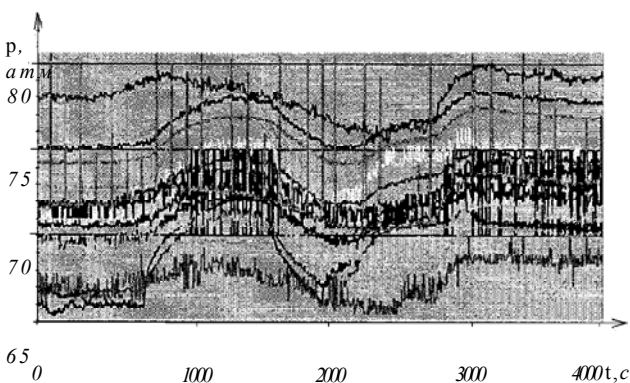


Рис. 1. Типовая характеристика перепадов давления газовой ТМ

В первом приближении будем полагать, что число измерителей и их местоположение совпадает с потребителями энергии (рис. 2).

Тогда общий критерий C проектирования ИП предлагается представить трехкомпонентной системой

$$C = \begin{cases} M_{FrErr}\{V_{HF}, V_{LF}\}, \\ \text{fres}\{m, ku, H, v\}, \\ \min_D Cn\{l, k, p, W, D\}. \end{cases} \quad (3)$$

В качестве модели комбинирования высокочастотных и низкочастотных возмущений предложено использование нелинейного полинома:

$$M_{FrErr} = b_0 + b_1 \cdot \sin(b_2 \cdot \sin(V_{HF}(t)) + V_{LF}(t)), \quad (4)$$

где b_0 – свободный коэффициент, учитывающий постоянное смещение возмущения; b_1 – коэффициент, учитывающий влияние нижних частот; b_2 – коэффициент, учитывающий влияние верхних частот.

На основании выражений (1)–(4) предлагается алгоритмическая реализация частотно-зависимой модели в модуле Simulink системы моделирования Matlab. Работа алгоритма представлена следующими шагами.

1. Моделирование внешних воздействий путем комбинирования высокочастотных и низкочастотных возмущений. Структурная схема данного шага для значений $b_0 = 0$, $b_1 = 5$, $b_2 = 20$ представлена на рис. 3.

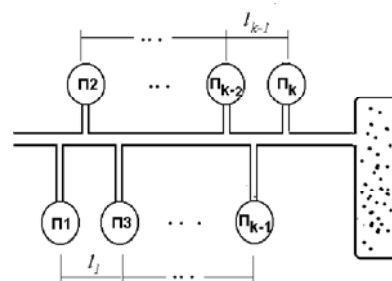


Рис. 2. Структура расположения потребителей и источников ТЭК

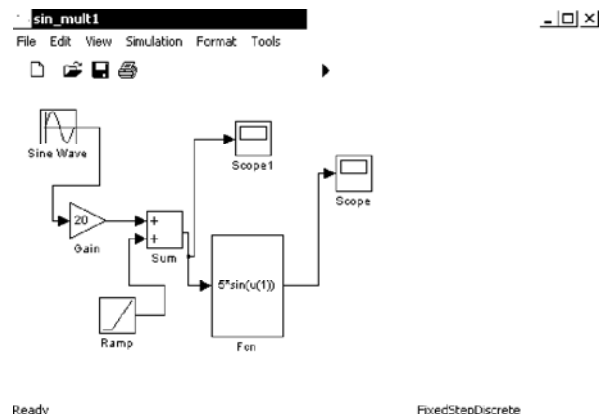


Рис. 3. Схема модели комбинирования высокочастотных и низкочастотных возмущений

2. Реализация частотно-зависимой модели на основе функции моделирования (1), представленной линейным оператором:

$$X' = A \times X, \quad (5)$$

где $A = [0, 1; -1/(ku \cdot m), -H \cdot v / m]$.

Модель комбинирования высокочастотных и низкочастотных возмущений добавлена к частотно-зависимой модели (5) в мультипликативном виде:

$$X' = \{A \times X\} \cdot M_{FrErr}. \quad (6)$$

На рис. 4 представлена структурная схема реализации частотно-зависимой модели в модуле Simulink системы моделирования Matlab.

3. Реализация функции C_n геометрического критерия проектирования на основе выражения (2) выполняется в виде системы k уравнений динамики давления p_k на k -ом участке потребителя:

$$y'_k = -(1/(l_k \cdot W_k))y_k + U, \quad (7)$$

где U – исходное давление топливного резервуара, при этом $U = \sum y_k$.

На рис. 5 представлена структурная схема реализации функции C_n геометрической модели в модуле Simulink для $k=3$, $l_0=0,5$, $l_1=l_2=1$.

4. Оценка результатов моделирования по обобщенному критерию (3). Если минимум дисперсии D достигнут, формируются минимальные параметры проектирования и алгоритм заканчивает свою работу. Иначе переходим к п. 1.

Обобщенная схема моделирования работы датчиков расхода топлива и взаимодействия с топливной магистралью представлена на рис. 6.

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

В соответствии с построенными структурными схемами частотно-зависимых моделей получены следующие результаты моделирования.

На рис. 7 представлен результат моделирования комбинирования высокочастотных и низкочастотных возмущений. На рис. 8 представлен результат реализации частотно-зависимой модели в модуле Simulink системы моделирования Matlab.

Результаты моделирования обобщенной структурной схемы работы датчиков расхода топлива и их взаимодействия с топливной магистралью для значений $v=1,04$, $ku=1,0$, $H=20$ и $v=1,04$, $ku=1,0$, $H=10$ представлены на рис. 9 и рис. 10 соответственно.

Общие результаты проведенных экспериментов приведены на графике (рис. 11). Представлена зависимость дисперсии ошибок измерений при изменении обобщенного конструкционного параметра от $H=2$ до $H=30$,

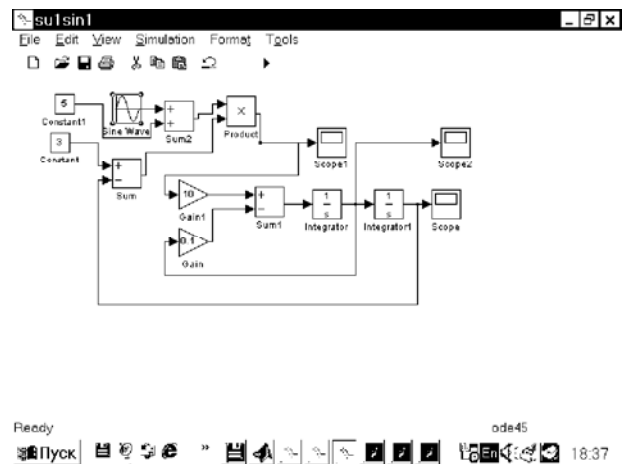


Рис. 4. Структурная схема реализации частотно-зависимой модели в модуле Simulink системы моделирования Matlab

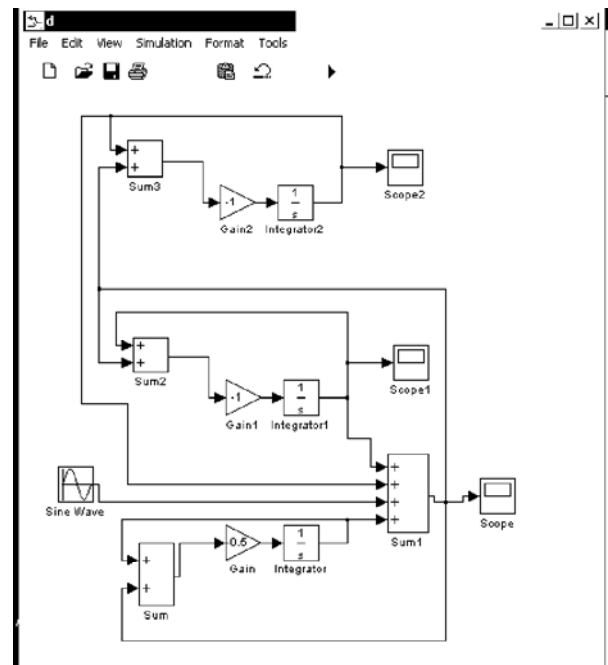


Рис. 5. Структурная схема реализации геометрической модели

при выбранных значениях вязкости рабочего тела – метана $v=1,04$, коэффициента упругости $ku=1,0$, для различных значений l и k . При этом $l_1 < l_2 < l_3$, $k_1 > k_2 > k_3$.

Результаты моделирования позволили выделить области конструктивных параметров проектирования ИП А и С, в которых динамические погрешности измерений максимальны. Определена предпочтительная область В конструктивных параметров проектирования в которой влияние динамических погрешностей измерений минимально.

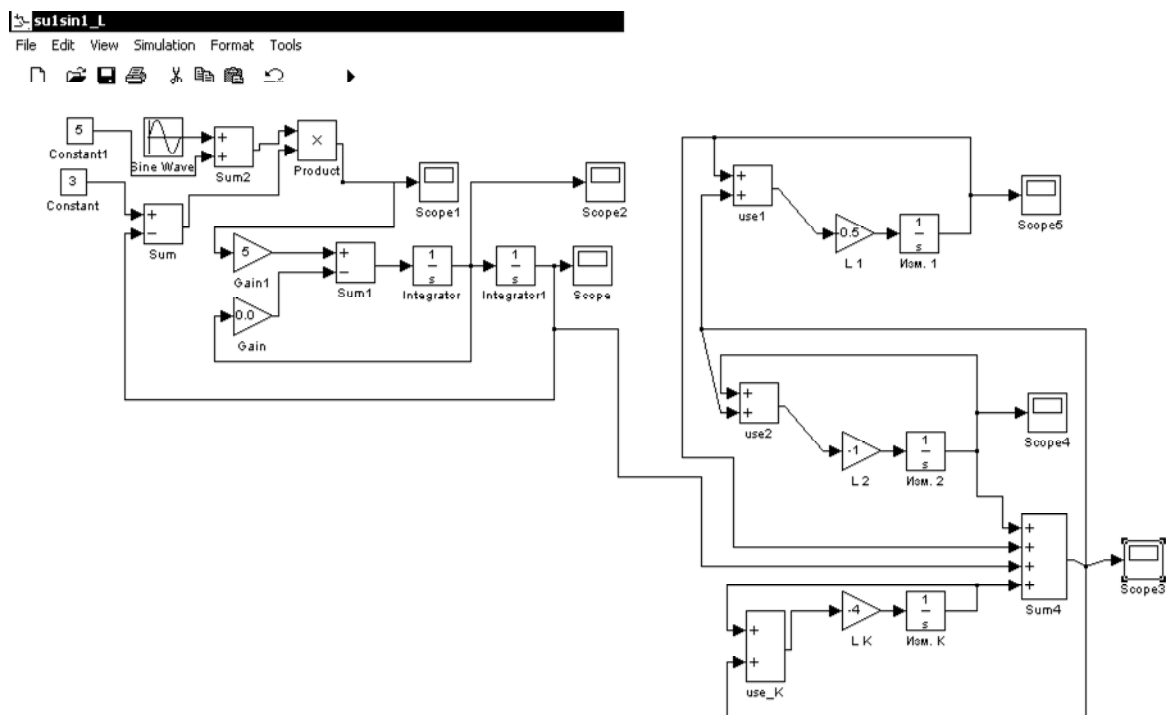


Рис. 6. Обобщенная структурная схема моделирования работы датчиков расхода топлива и их взаимодействия с топливной магистралью

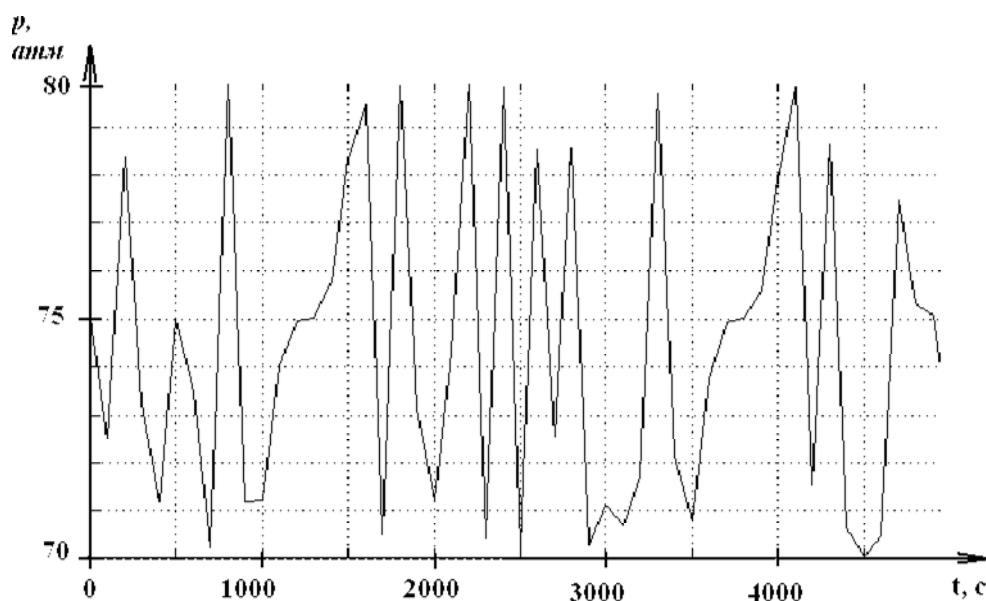


Рис. 7. Результаты моделирования комбинации высокочастотных и низкочастотных возмущений

ВЫВОДЫ

В результате проведенных исследований разработан общий алгоритм моделирования ИИС контроля параметров энергоносителей ЛА и ТЭК, выполнено моделирование общего критерия проектирования ИП. Разработаны математические модели функционирования ИИС контроля параметров энергоносителей, которые учитывают динамические параметры датчиков и широкую

спектральную характеристику ТМ, что позволяет минимизировать динамические погрешности проектируемых ИП. Предложены модель частотных возмущений, резонансная характеристика проектируемого ИП, трехкомпонентный критерий проектирования ИП.

Результаты моделирования позволили сформировать рекомендации при проектировании ИП по выбору обобщенного конструктивного параметра H и коэффициента

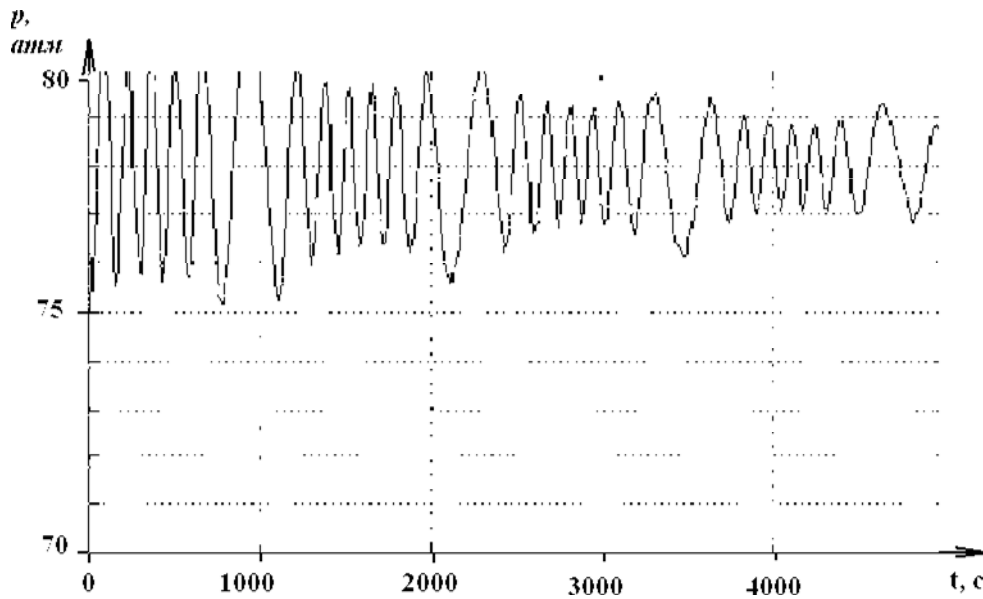


Рис. 8. Результаты моделирования частотно-зависимой модели ТП в модуле Simulink системы моделирования Matlab

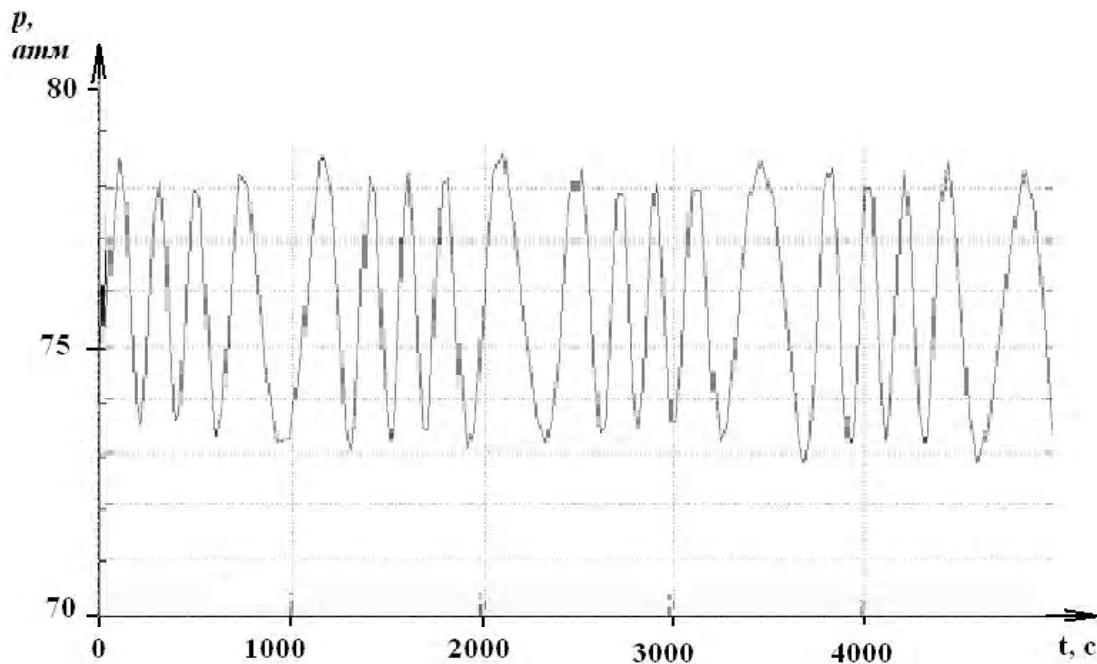


Рис. 9. Результаты моделирования обобщенной структурной схемы работы датчиков расхода топлива и их взаимодействия с топливной магистралью для значений $\nu=1,04$, $k_u=1,0$, $H=20$.

упругости пружинного элемента k_u при заданных значениях массы чувствительного элемента m , вязкости рабочего измерительного тела ν , расстояний между потребителями l и заданном числе потребителей k .

Таким образом, выполнено совершенствование методологии проектирования ИИС и новых устройств контроля параметров энергоносителей, которые могут применяться в ЛА или наземных магистралях ТЭК.

СПИСОК ЛИТЕРАТУРЫ

1. Целищев, Е. Пример проектирования систем контроля в среде AutomatiCS ADT. Фрагменты пилотного проекта / Е. Целищев, М. Савинов, А. Непомнящих // САПР и графика. – 2006. – № 5. – С. 30–37.
2. Свиридов, В. Н. Создание ИВС «Кама» подсистемы «Утечка» / В. Н. Свиридов // Приборы. Системы автоматизации. – 2007. – № 4. – С. 26–29.

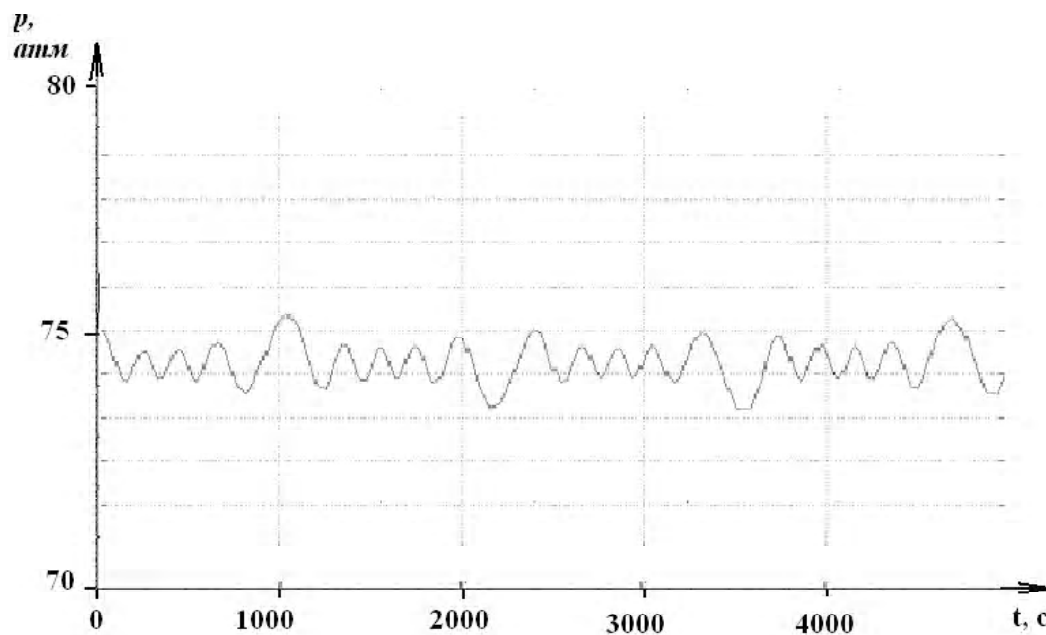


Рис. 10. Результаты моделирования обобщенной структурной схемы работы датчиков расхода топлива и их взаимодействия с топливной магистралью для $v=1,04$, $ku=1,0$, $H=10$

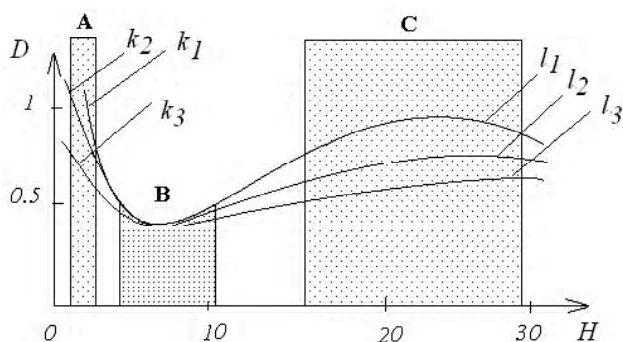


Рис. 11. Общие результаты проведенных экспериментов

3. Михайлов, А. Г. Моделирование структурных схем измерителей расхода топлива / А. Г. Михайлов, А. Р. Сарамольки // *Авиационно-космическая техника и технология*. – 2009. – № 5. – С. 35–39.

Стаття надійшла до редакції 12.05.2011.

Кошовий М. Д., Михайлов А. Г.

РОЗРОБКА АЛГОРИТМІВ МОДЕЛЮВАННЯ РОБОТИ ДАТЧИКІВ ВИТРАТИ ПАЛИВА ТА ВЗАЄМОДІЇ З ПАЛИВНОЮ МАГІСТРАЛЛЮ

В статті розглянуті принципи побудови алгоритмів моделювання роботи датчиків витрати палива. Представлені динамічні моделі функціонування систем вимірювання параметрів енергоносіїв.

Ключові слова: математичні моделі, алгоритми моделювання, датчики, витрата палива.

Koshevoy N. D., Mikhailov A. G.

MODELING ALGORITHMS DEVELOPMENT OF DEVICE FUNCTION OF THE FUEL FLOW AND INTERACTION WITH A FUEL MAIN

In this article the construction principles of device function modeling algorithms of the fuel flow are considered. The functioning dynamic models of parameters measurement systems of fuel are submitted.

Key words: Mathematical models, modeling algorithms, gauges, fuel flow.

Юр Т. В.¹, Харитонов В. Н.², Дубровин В. И.³¹Ассистент Запорожского национального технического университета²Инженер-конструктор ГП ЗМКБ «Прогресс» им. Ивченко, Запорожье³Канд. техн. наук, профессор Запорожского национального технического университета

МОДЕЛЬ ДИАГНОСТИРОВАНИЯ УЗЛОВ ГАЗОТУРБИННЫХ ДВИГАТЕЛЕЙ С ПРИМЕНЕНИЕМ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ И S-ДИСКРИМИНАНТА

В работе предложена модель диагностирования узлов газотурбинного двигателя с применением метода спектрального анализа вибрационных сигналов на основе непрерывного вейвлет-анализа и S-дискриминанта. Разработанная модель может быть использована при проектировании перспективных систем управления силовых установок современных летательных аппаратов для анализа особенностей нестационарных сигналов и диагностирования состояний деталей и узлов ГТД.

Ключевые слова: модель диагностирования, вибросигнал, вейвлет-фильтрация, S-дискриминант, узлы ГТД.

ВВЕДЕНИЕ

Одно из перспективных направлений решения задачи повышения эффективности характеристик силовых установок современных летательных аппаратов касательно адаптивности к условиям эксплуатации и режимам использования самолета, надежности и ресурса лежит в области интегрированных систем автоматического управления и контроля авиационных газотурбинных двигателей (ГТД).

Износ в процессе эксплуатации приводит к изменениям состояния и характеристик узлов двигателя: опор роторных деталей, зубчатых передач, лопаточных машин, камер сгорания, элементов проточной части. В результате значительно изменяются как статические, так и динамические характеристики двигателя, определяющие его свойства как объекта регулирования.

Разработка метода автоматического распознавания технического состояния узлов двигателя, реализуемого в системе автоматического управления (САУ) двигателя, позволит в процессе функционирования двигателя автоматически переходить к режиму управления, наиболее оптимальному в данных условиях.

1. ПОСТАНОВКА ЗАДАЧИ

Одними из эффективных средств технического диагностирования ГТД являются вибродиагностические методы. Вибрационный сигнал несет в себе большое количество информации о техническом состоянии узлов двигателя и может быть достоверным показателем их работоспособности.

В работе [1] авторами предложено усовершенствование статистического метода анализа вибрационных сигналов за счет применения S-дискриминанта [2], что позволило создать метод первичной экспресс-диагностики отклонения технического состояния двигателя от нормального.

Для эффективного анализа вибрационных сигналов ГТД с целью выделения характерных составляющих

и признаков износа узлов авторами усовершенствован метод анализа огибающей высокочастотной случайной вибрации за счет применения банков вейвлетных фильтров и кросс-спектрального анализа [3, 4].

За счет выделения не только гармонических, но и импульсных составляющих, предложенный метод позволяет выделить большее количество характеристических частотных составляющих вибрационных сигналов, что дает более четкие признаки отклонения состояния узлов двигателя от нормального.

Результаты экспериментального исследования эффективности разработанного метода анализа вибрационных сигналов на основе непрерывного вейвлет-преобразования для решения практических задач диагностирования износа подшипников качения приведены в работах [3, 4].

Результаты проведенного исследования показали, что применение вейвлет-фильтрации увеличивает быстродействие и надежность постановки диагноза по сравнению с традиционными статистическими методами и методами, основанными на применении быстрого преобразования Фурье (БПФ).

Кроме того, полученные результаты позволяют сделать вывод об универсальности предложенного метода и возможности его эффективного применения для идентификации различных дефектов двигателя: дефектов подшипников, зубчатых передач, дефектов роторных деталей и т. д.

Целью данной работы является разработка модели диагностирования узлов ГТД, основанной на применении разработанных авторами методов анализа вибрационных сигналов.

2. ВОЗМОЖНОСТИ РЕАЛИЗАЦИИ МЕТОДОВ ДИАГНОСТИРОВАНИЯ УЗЛОВ ГТД

Реализация методов диагностирования рассмотрена на примере двигателя ТВ3-117ВМА-СБМ самолета Ан-140.

Для реализации метода диагностирования узлов ГТД с применением предложенных методов анализа вибрационных сигналов возможны два варианта реализации.

В первом варианте на самолете не предусматривается никаких доработок, а реализация разработанного метода диагностирования осуществляется в наземном технологическом цифровом аппаратно-программном комплексе, который подключается к бортовой аппаратуре контроля вибрации.

У данного варианта существует ограничение. Как показала практическая отработка предложенных методов, зависимость амплитуды характерных частот дефектов подшипников от уровня повреждений элементов подшипников носит нелинейный характер. Имеющее место уменьшение амплитуды при среднем уровне повреждения может быть пропущено при периодическом контроле. В таком случае может быть некорректно определен момент развития более опасных и быстроразвивающихся значительных уровней повреждения элементов конструкции подшипника.

Более перспективным следует считать второй вариант, который предусматривает установку на борту самолета бортового цифрового модуля контроля и диагностики.

Цифровой модуль контроля и диагностики принимает аналоговые сигналы от усилителей заряда штатной бортовой аппаратуры контроля вибрации и производит их обработку. В результате обработки решается ряд задач по распознаванию технического состояния узлов и систем двигателя:

- контроль роторных составляющих;
- контроль переднего редуктора на частотах шестеренчатых и подшипниковых составляющих;
- контроль заднего редуктора и валопровода на частотах шестеренчатых и подшипниковых составляющих;
- контроль коробки приводов на частотах шестеренчатых и подшипниковых составляющих;

- контроль загрязнения масла;
- оценка состояния измерительных трактов «датчик вибрации – усилитель заряда»;
- контроль элементов газозвдушного тракта.

По результатам распознавания выдаются управляющие воздействия (сообщения) в САУ самолета.

Для решения перечисленных задач авторами разработана модель диагностирования узлов ГТД, использующая разработанные методы анализа вибрационных сигналов: статистический метод, улучшенный применением S -дискриминанта, и метод, основанный на применении вейвлетной фильтрации.

3. МОДЕЛЬ ДИАГНОСТИРОВАНИЯ УЗЛОВ ГТД

Разработанная модель диагностирования узлов ГТД основана на существующей модели [5], представляющей систему вибродиагностирования ГТД на основе применения БПФ. Модель была модифицирована с учетом специфики применяемых разработанных методов анализа сигналов. На рис. 1 представлен общий вид разработанной модели.

В модель диагностирования поступает информация двух видов: аналоговая информация от усилителей заряда (УЗ) датчиков вибраций (ДВ) (анализируемый вибрационный сигнал) и цифровые данные по частотам вращения роторов, крутящему моменту на валу винта и др. ($N_{ВВ}$, $N_{СТ}$, $N_{ТК}$, $M_{КР}$). Выходными данными модели являются сообщения, передаваемые в САУ самолета.

В модуле предварительной обработки информации (МПОИ) определяется режим работы двигателя как динамической системы. Если полученные параметры находятся в устойчиво-неизменяемом состоянии в текущем временном диапазоне, то такой режим считается установившимся (УР), в противном случае режим следует считать переходным (ПР).

Вышеописанная характеристика режима работы двигателя важна как на этапе подготовки данных (определе-

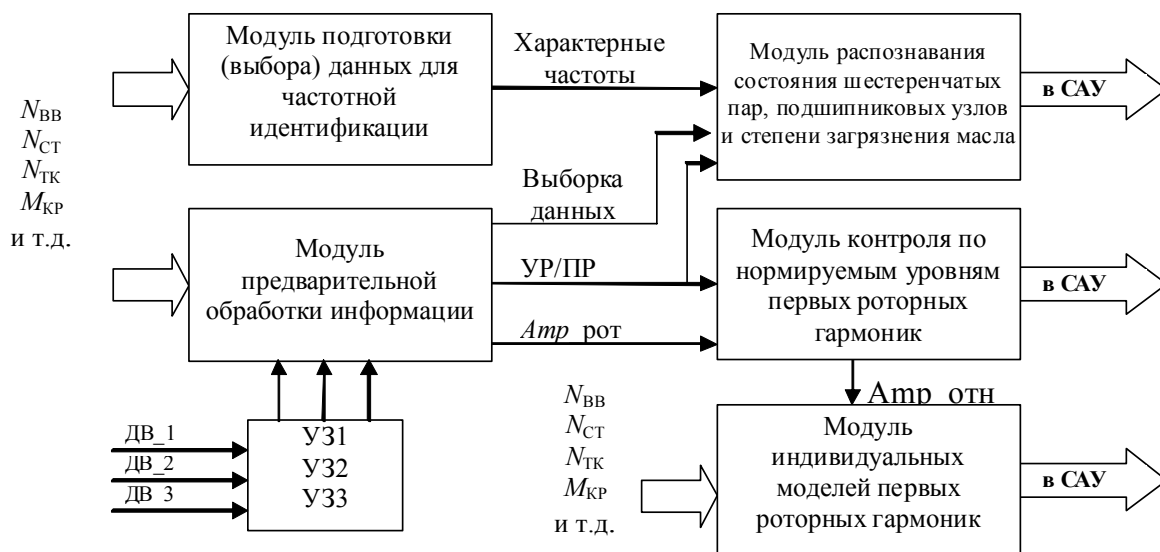


Рис. 1. Модель диагностирования узлов ГТД

ние количества точек в выборке и количества групп выборок данных по вибрациям), так и на этапе контроля и диагностики. В МПОИ происходит оцифровка вибрационного сигнала с отбраковкой сбойных групп данных на основе допускового и перекрестного контроля. После чего данные проходят вейвлет-преобразование с дальнейшим спектральным анализом полученных вейвлет-коэффициентов. Из спектров выделяются составляющие первых роторных гармоник роторов турбокомпрессора, винта и трансмиссии. Амплитудные значения выделенных составляющих спектров передаются в модуль контроля по нормируемым уровням, где происходит их сравнение с тремя типовыми уровнями.

Ограничительные уровни «Опасная вибрация» (ОВ) и «Повышенная вибрация» (ПВ) устанавливаются исходя из комплексных требований прочности конструкции двигателя и его комплектующих агрегатов и их долговечности на протяжении выработки жизненного цикла. Т. е. значения ОВ и ПВ оговариваются параметрами точности. Если текущее значение превысит указанные уровни и это превышение устойчиво будет фиксироваться на протяжении определенного времени, то в САУ самолета будет выдано соответствующее аварийное сообщение.

Информационный уровень «Статистический типовой» представляет зависимость амплитуды той или иной роторной гармоники (статистически максимально вероятной для парка двигателей данного типа) от частоты вращения соответствующего ротора или крутящего момента. Данный уровень используется для вновь устанавливаемого на самолет двигателя на период формирования индивидуальной модели в модуле индивидуальных моделей первых роторных гармоник (МИМПРГ).

Одновременно в модуле контроля по нормируемым уровням первых роторных гармоник производится вычисление текущих значений относительных амплитуд той или иной роторной гармоники по формуле

$$Amp_i \text{ _отн} = 20 \log \frac{Amp_i \text{ _тек}}{Amp_i \text{ _ОВ}}, \text{ дБ},$$

где $Amp_i \text{ _тек}$ – текущее значение амплитуды, $Amp_i \text{ _ОВ}$ – значение амплитуды для уровня опасной вибрации.

Вычисленные значения относительных амплитуд передаются в МИМПРГ.

МИМПРГ отвечает за формирование и хранение в энергонезависимой памяти индивидуальных моделей первых роторных гармоник, а также оценку тренда этих моделей в процессе выработки ресурса.

Индивидуальные модели представляются в виде двухуровневых зависимостей относительных амплитуд от частот вращения соответствующих роторов и крутящего момента. Модель для установившихся режимов представляет собой набор поддиапазонных статистически-вероятностных оценок (максимальных и минимальных, средних значений и границ клиппирования) относительных амплитуд, определенных по первым Z полетам. Модель для переходных режимов представляет постоянное

значение для всех частот вращения роторов и крутящего момента, определенного как максимально-достигнутое значение во всех условиях эксплуатации по первым Z полетам.

Оценка трендов по времени выработки ресурса строится на вычислении S -дискриминанта средних значений моделей для установившихся ресурсов. При условии, если текущее значение S -дискриминанта в каком-либо поддиапазоне превысит заранее установленное значение (из опыта экспериментальных исследований работы метода рекомендуется значение равное семи), то в САУ самолета выдается диагностическое сообщение.

На рис. 2 приведена функционально-логическая модель модуля распознавания состояния шестеренчатых пар, подшипниковых узлов и степени загрязнения масла (МРСШПМ).

Поступающая в МРСШПМ оцифрованная в МПОИ информация «переупаковывается» относительно частоты вращения винта. Данная операция производится для того, чтобы компенсировать влияние низкочастотных процессов в системе «турбокомпрессор-трансмиссия-винт». Отсортированные таким образом данные проходят вейвлет-преобразование в соответствии с методом, представленным в работах [3, 4].

Модуль подготовки данных для частотной идентификации (МПДЧИ) производит расчет характеристических частот узлов двигателя в зависимости от частот вращения роторов.

На основании информации, поступающей из модуля выбора данных для контроля шестерен и подшипников редукторов и коробки приводов, в МРСШПМ формируются задания частотной идентификации спектров вейвлет-коэффициентов и взаимных спектров вейвлет-коэффициентов. Затем производится амплитудное сравнение составляющих, выделенных в соответствии с заданием частотной идентификации.

В соответствии с заданием частотной идентификации спектры вейвлет-коэффициентов и взаимные спектры вейвлет-коэффициентов проходят обработку в специальных согласующих фильтрах. В общем виде формула таких фильтров:

$$Conform Filter (CF) = \sqrt{\sum_{i=1}^N (V_i \bullet w_i)^2},$$

где

$$w_i = \left[\frac{1}{k} \sum_{t=1}^k V_i(t) \right]^{-1}, \quad k < N.$$

Эти фильтры предназначены для оценки «боковых» и комбинационных составляющих относительно основных характерных составляющих. Цель такой обработки – раннее обнаружение изменений в соответствующих узлах. При отсутствии или незначительных изменениях выходное значение фильтра стремится к единице. В противном случае – к нулю. Результаты фильтрации и амплитуды

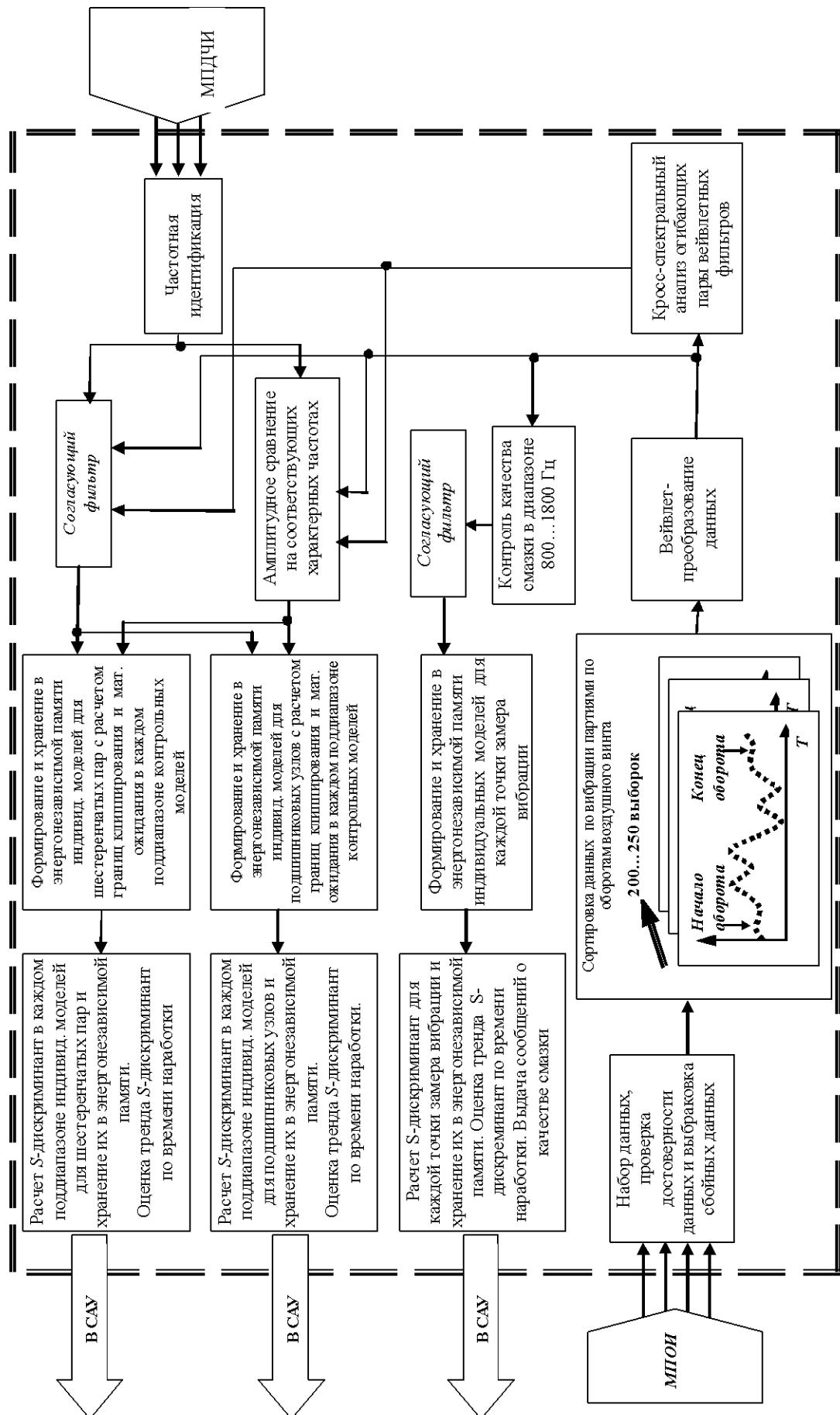


Рис. 2. Модель распознавания состояния шестеренчатых пар, подшипниковых узлов, подлинниковых узлов и степени загрязнения масла

литудных сравнений в дальнейшем проходят через процедуры обобщения в индивидуальных моделях и оценку трендов по аналогии с операциями в МИМПРГ. В МРСШПМ предусмотрена оценка степени загрязнения масла на основе обработки данных в частотном диапазоне 800–1800 Гц.

Предусмотренная в перечне задач по распознаванию технического состояния узлов и систем двигателя оценка состояния измерительных трактов «датчик вибрации – усилитель заряда» основана на сравнительной оценке (по схеме согласующих фильтров) составляющих в частотном диапазоне 1...10 Гц (самые высшие масштабы вейвлет-преобразования) и составляющей зубозацепления ступени перебора (средние масштабы вейвлет-преобразования).

ЗАКЛЮЧЕНИЕ

В данной работе получила дальнейшее развитие модель диагностирования узлов ГТД, которая отличается от существующих применением непрерывного вейвлет-анализа и S-дискриминанта.

Проведенные экспериментальные исследования разработанных методов анализа вибрационных сигналов показали их эффективность при выделении признаков износа узлов двигателя.

При помощи усовершенствованного статистического метода предлагается проводить общую экспресс-диагностику отклонения технического состояния двигателя от нормального и следить за степенью развития дефектов. При помощи метода, основанного на применении вейвлетной фильтрации, проводится спектральный анализ вибрационного сигнала с целью выделения частотных характеристических составляющих, позволяющих судить о месте возникновения дефекта.

Таким образом, модель диагностирования, объединяющая в себе разработанные методы, делает возможным выявление признаков износа узлов ГТД на более ранних этапах развития и позволяет в автоматическом режиме следить за степенью развития дефектов.

Разработанная модель диагностирования может быть использована при проектировании перспективных систем управления силовых установок современных летательных аппаратов для анализа особенностей нестационарных сигналов и диагностирования состояний деталей и узлов ГТД.

СПИСОК ЛИТЕРАТУРЫ

1. Юр, Т. В. Диагностирование технического состояния подшипников качения / Т. В. Юр, В. Н. Харитонов,

В. И. Дубровин // *Авиационно-космическая техника и технология*. – 2009. – № 7 (64). – С. 166–170.

2. Соколова, А. Г. Вибромониторинг машинного оборудования и раннее обнаружение эксплуатационных повреждений / А. Г. Соколова, Ф. Я. Балицкий // *Вестник научно-технического развития*. – 2008. – № 7 (11). – С. 45–50.
3. Юр, Т. В. Метод анализа технического состояния подшипников качения, основанный на использовании вейвлет-математики / Т. В. Юр, В. Н. Харитонов, В. И. Дубровин // *Авиационно-космическая техника и технология*. – 2010. – 10(77). – С. 187–191.
4. Юр, Т. В. Исследование фильтрующих свойств вейвлет-преобразования / Т. В. Юр, В. И. Дубровин, В. Н. Харитонов // *Радиоэлектроника, информатика, управління*. – 2010. – № 2. – С. 157–165.
5. Patent US 7013210 B2. Vibration monitoring system for gas turbine engines / McBrien, Gary M. (Glastonbury, CT), Gottwald, James (Amherst, NH); Goodrich Pump & Engine Control Systems, Inc. (West Hartford, CT, US). – filed on 16-Mar-2004, published on 14-Mar-2006.

Стаття надійшла до редакції 25.05.2011.

Юр Т. В., Харитонов В. Н., Дубровін В. І.

МОДЕЛЬ ДІАГНОСТУВАННЯ ВУЗЛІВ ГАЗОТУРБІННИХ ДВИГУНІВ З ВИКОРИСТАННЯМ ВЕЙВЛЕТ-ПРЕТВОРЕННЯ ТА S-ДИСКРИМІНАНТУ

У роботі запропоновано модель діагностування вузлів газотурбінного двигуна із застосуванням методу спектрального аналізу вібраційних сигналів на основі безперервного вейвлет-аналізу та S-дискримінанту. Розроблена модель може бути використана при проектуванні перспективних систем керування силових установок сучасних літальних апаратів для аналізу особливостей нестационарних сигналів і діагностування станів деталей і вузлів ГТД.

Ключові слова: модель діагностування, вібраційний сигнал, вейвлет-фільтрація, S-дискримінант, вузли ГТД.

Yur T. V., Haritonov V. N., Dubrovin V. I.

DIAGNOSTIC MODEL FOR GAS TURBINE ENGINE ELEMENTS BASED ON WAVELET ANALYSIS AND S-DISCRIMINANT

Diagnostic model for gas turbine elements based on wavelet spectral analysis method of vibration signals and S-discriminant is proposed in the work. The developed model could be used in construction of perspective control systems of aircraft engine units to analyze non-stationary signals and to give a diagnosis of engine units state.

Key words: model of diagnostics, vibrations, wavelet filtering, S-discriminant, gas turbine engine units.

ПОСТРОЕНИЕ ДИАГНОСТИЧЕСКИХ МОДЕЛЕЙ НА ОСНОВЕ ПАРАДИГМЫ ОТРИЦАТЕЛЬНОГО ОТБОРА С ИСПОЛЬЗОВАНИЕМ ПРИНЦИПА МАСКИРОВАНИЯ ДЕТЕКТОРОВ

Получили дальнейшее развитие модель отрицательного отбора с использованием маскированных детекторов и метод ее обучения. Исследовались различные критерии останова для метода обучения модели. Разработан критерий останова, позволяющий выполнить своевременный останов процесса обучения, и таким образом ускорить данный процесс, а также сократить использование вычислительных ресурсов.

Ключевые слова: отрицательный отбор, маскированный детектор, критерий останова.

ВВЕДЕНИЕ

При решении задач технического и медицинского диагностирования возникает необходимость представления диагностической модели в виде продукционных правил вида «если-то». Это обусловлено тем, что продукционные правила отличаются простотой представления знаний и организации логического вывода [1]. На практике зачастую оказывается целесообразным описывать antecedentes правил в бинарном виде, где логическая единица определяет наличие признака у экземпляра, а логический ноль – его отсутствие.

Традиционно используемые для построения диагностических моделей подходы [2, 3] обладают рядом недостатков. Так, например, извлечение правил из искусственных нейронных сетей предполагает наличие дополнительного этапа вербализации [2]. Деревья решений облегчают процедуру генерации правил, однако при решении задач большой размерности получаемые в конечном итоге правила существенно усложняются [3].

Устранение этих недостатков возможно обеспечить при использовании искусственных иммунных систем. В частности, предлагается использовать парадигму отрицательного отбора, благодаря таким ее особенностям, как независимость от представления данных, прозрачность процесса обучения и возможность обучения на экземплярах только одного класса [4]. Однако в существующих реализациях моделей отрицательного отбора значительно снижается точность классификации при увеличении размерности задачи [5].

Для снижения размерности задачи предлагается выполнять отбор наиболее информативных признаков, что позволит повысить точность результатов, увеличит способность модели к обобщению, ускорит процесс обучения и улучшит интерпретируемость результатов.

Цель данной работы заключается в создании метода на основе принципов иммунокомпьютинга, позволяющего строить диагностические модели для случая двух классов и представлять результаты работы в виде про-

дукционных правил, а также осуществлять отбор наиболее информативных признаков.

1. ПОСТАНОВКА ЗАДАЧИ

Рассмотрим задачу построения диагностической модели для случая, когда все экземпляры описываются с помощью бинарных строк длиной l , где каждый разряд соответствует значению входного признака, характеризующего свойства экземпляра. Набор всех возможных l -битовых строк представляет собой пространство признаков U . Известно, что множество U можно представить в виде двух комплементарных подмножеств S и N , описывающих «свои» и «чужие» экземпляры соответственно. Пусть нам дана обучающая выборка S' , состоящая исключительно из «своих» экземпляров. Необходимо построить диагностическую модель, которую в терминах отрицательного отбора опишем набором детекторов D , такую, чтобы на основании этого набора любой экземпляр $x \in U$ можно было однозначно отнести к классу «своих» или «чужих» экземпляров. Также необходимо, чтобы набор детекторов D можно было преобразовать во множество продукционных правил вида «если-то», где antecedentes описывают наличие или отсутствие определенного признака у экземпляра, а консеквент определяет принадлежность экземпляра к множеству S или N . Antecedents правил должны учитывать наиболее информативные признаки.

2. ОТРИЦАТЕЛЬНЫЙ ОТБОР В ИММУНОКОМПЬЮТИНГЕ

Модель отрицательного отбора в искусственных иммунных системах была первой моделью, основанной на поведении биологической иммунной системы.

Основными компонентами модели отрицательного отбора являются детекторы и правила сопоставления. Детекторы способны определить, является ли неизвестный экземпляр экземпляром «чужого» класса, или нет.

Для этого к экземпляру x и детектору d применяется правило сопоставления $match(d, x)$, которое возвращает значение «1», если детектор d активизировался при сопоставлении с экземпляром x , и «0» – в противном случае.

Принцип работы модели заключается в том, что модель содержит набор детекторов, которые толерантны к «своим» экземплярам: $match(d, s) = 0, \forall s \in S$. Поскольку модель отрицательного отбора решает задачу двухклассовой классификации (каждый экземпляр может быть отнесен ко множеству S или N), и $S \cup N = U$, $S \cap N = \emptyset$, то это предполагает, что детекторы будут реагировать на «чужие» экземпляры $n \in N$.

Первый метод обучения такой модели был описан в [6] и состоял из следующих этапов.

1. Этап генерации детекторов. На этом этапе случайным образом формируется набор кандидатов в детекторы C , как правило, превышающий по объему число экземпляров обучающей выборки S' .

2. Цензурирование. На данном этапе каждый кандидат в детекторы $c \in C$ сопоставляется с каждым «своим» экземпляром из обучающей выборки S' .

Если кандидат активизировался (значение правила сопоставления стало равным 1), то он уничтожается, поскольку не может обеспечить толерантность модели к «своим» экземплярам. Все оставшиеся кандидаты формируют набор детекторов D , который может быть использован для обнаружения «чужих» экземпляров.

Несмотря на очевидную ресурсоемкость данного метода, он получил дальнейшее развитие, что заключалось в разработке более оптимальных методов генерации набора детекторов (жадного метода, метода генерации детекторов за линейное время [7]).

3. МЕТОД ОТРИЦАТЕЛЬНОГО ОТБОРА С ПЕРЕСТАНОВКОЙ БИТОВ

Большинство реализаций модели отрицательного отбора обладают существенным недостатком – эффектом «дыр». «Дырами» называют такие экземпляры «чужого» класса, для которых не может быть сформирован ни один детектор, который оставался бы толерантным к «своим» экземплярам, и при этом реагировал на данный «чужой» экземпляр: $a \in N, \forall x \in U, match(u, a), \exists s \in S, match(u, s)$.

Возможное решение данной проблемы описано в [8], при этом модель отрицательного отбора модифицирована таким образом, что каждый детектор представляется не только битовой строкой, а также маской перестановки битов. Эта маска служит для получения дополнительной битовой строки на основе исходной. Маска определяет порядковые номера битов, которые участвуют в перестановке, а также их новое смещение относительно начала строки. Таким образом, с помощью маски формируется вторичное представление детектора (так-

же в виде битовой строки). Например, если детектор представлен исходной битовой строкой $d^1 = \{0, 1, 1, 0\}$, а маска перестановки имеет вид $\{3, 2, 1, 4\}$, то вторичное представление детектора будет иметь вид $d^2 = \{1, 1, 0, 0\}$.

На этапе обучения такие детекторы с множественным представлением подвергаются цензурированию (удаляются) только в случае, если $\forall s \in S, match(d^i, s) = 1, i = 1, \dots, k$, где k – число представлений детектора. Однако при проведении классификации экземпляра x , детектор считается активизированным, если хотя бы одно его представление активизировалось: $match(d^i, x) = 1, i = 1, \dots, k$.

Очевидно, что это увеличит область покрытия детектора, однако приведет к увеличению ошибки первого рода («чужие» экземпляры будут ошибочно относиться к классу «своих»).

Первоначально предполагалось, что маска перестановки битов является общей для всей модели, однако в модели LISYS [9] было предложено использовать индивидуальные маски для каждого детектора в наборе. Применение маски перестановки битов в задаче обнаружения сетевых атак позволило обеспечить точность классификации на 32 % выше, по сравнению с базовым методом.

4. МАСКИРОВАНИЕ

Для уменьшения отрицательного влияния «дыр» на результаты классификации авторами предлагается модификация модели отрицательного отбора. Так же, как и в случае множественного представления детектора, в предлагаемой модели детекторы с маскированием характеризуются увеличенной областью покрытия.

Увеличение области покрытия достигается за счет использования замаскированных битов в строках, представляющих детекторы. Для этого алфавит Ω , на котором определена задача, должен быть дополнен специальным символом Z , обозначающим маскирование: $\Omega = \{0, 1, Z\}$.

Будем считать, что при вычислении правил сопоставления замаскированный бит строки совпадает с любым значением бита другой строки. С использованием замаскированных детекторов становится возможным применение полного соответствия строк в качестве меры подобия двух детекторов, что было затруднительным ранее [10].

Жизненный цикл детектора также предлагается модифицировать, дополнив генерацию и цензурирование итеративным обучением детектора. Это позволит сократить число сгенерированных детекторов.

Извлечение правил из замаскированных детекторов является тривиальной задачей и заключается в следующем. Каждый детектор представляет собой одно продукционное правило, антецедент которого включает в себя все незамаскированные биты детектора (рассматриваемые как наличие или отсутствие соответствующего при-

знака), а консеквент всегда определяет принадлежность к классу «чужих» экземпляров.

5. МЕТОДОБУЧЕНИЯ МОДЕЛИ ОТРИЦАТЕЛЬНОГО ОТБОРА С ИСПОЛЬЗОВАНИЕМ МАСКИРОВАНИЯ ДЕТЕКТОРОВ

Для обучения модели отрицательного отбора с использованием маскирования детекторов предлагается применять описанный ниже метод.

1. Установить номер итерации: $I = 0$. Положить число детекторов, подвергнутых цензурированию: $N_c = 0$. Задать: $D = \emptyset$.

2. Сформировать замаскированный детектор $d = \{d_i\}$, где d_i – значение маски для n -го разряда. Для этого всем битам строки, описывающей детектор d , необходимо установить значение: $\forall n: d_i = Z$. Увеличить номер итерации: $I = I + 1$.

3. Если $\exists s \in S: match(d, s) = 1$, тогда перейти к этапу 3, в противном случае – к этапу 6.

4. Выбрать произвольным образом бит $d_i, i = 1, \dots, l$, $d_i = Z$. Если такого бита не существует, тогда установить $N_c = N_c + 1$ и перейти к этапу 2, в противном случае – перейти к этапу 5.

5. Установить значение i -го бита детектора: $d_i = -s_i$. Перейти к этапу 3.

6. Добавить детектор d в набор детекторов: $D = D \cup \{d\}$. Если выполняется критерий останова (который может быть основан на значениях N_c и I – данный вопрос подробнее рассмотрен в разделе 6), тогда перейти к этапу 7, в противном случае – перейти к этапу 2.

7. Останов.

Важно отметить, что обучение детектора является «жадным» по своей природе. Поэтому для нахождения оптимального набора детекторов рекомендуется во время обучения отбирать «свои» экземпляры в случайном порядке.

Качество полученной диагностической модели будем характеризовать точностью классификации Φ , которая определяется как отношение числа правильно распознанных экземпляров к общему числу экземпляров тестовой выборки.

Производительность работы метода можно повысить, добавив процедуру фильтрации набора детекторов D на этапе 5. На данном этапе необходимо удалить дублирующиеся детекторы, а также детекторы, поглощенные другими детекторами с меньшей строгостью.

Строгость замаскированного детектора определим как:

$$\tau(d) = \sum_{i=1}^l \{1 | d_i \neq Z\}.$$

Детектор d удаляется в результате цензурирования, если $\exists d^*, \forall i = \overline{1, l}, (d_i^* = Z) \vee (d_i^* = d_i) \wedge (\tau(d^*) < \tau(d))$.

При программной реализации данного метода рекомендуется хранить набор детекторов в виде связанного списка, упорядочив его по убыванию строгости детектора.

6. КРИТЕРИИ ОСТАНОВА МЕТОДА ОБУЧЕНИЯ МОДЕЛИ ОТРИЦАТЕЛЬНОГО ОТБОРА

Предложенный метод генерирует вначале детекторы с меньшей строгостью, а затем заполняет получившиеся «дыры» более строгими детекторами. Поскольку заполнение «дыр» весьма несущественно сказывается на точности работы модели, то целесообразно прерывать обучение модели. Для этого необходим критерий останова, который позволит определить полноту покрытия пространства признаков детекторами из набора.

Рассмотрим критерии, на основе которых можно производить останов предложенного выше метода.

Наиболее простым с точки зрения вычислительной сложности является ограничение числа детекторов в популяции. Для этого требуется заранее оценить необходимое число детекторов для обеспечения требуемой точности классификации. Аналогичный подход применялся в [11], что позволило использовать заранее известное фиксированное число детекторов. Однако данный подход не пригоден для останова обучения маскированных детекторов, поскольку детекторы в наборе не являются однородными, а значит, их число не является показателем точности.

На рис. 1 представлен график, отображающий зависимость точности классификации Φ от числа детекторов $N_D = |D|$ в серии вычислительных экспериментов с использованием предложенной модели отрицательного отбора, проведенных в рамках решения задачи диагностирования заболеваний сердца [12].

Из рис. 1 можно видеть, что с ростом числа детекторов N_D увеличивается точность классификации Φ . Таким образом, для обеспечения высокой точности диагностической модели на основе клональной селекции необходимо использовать число детекторов, сопоставимое с объемом обучающей выборки. В то же время для обеспечения обобщающих свойств диагностической

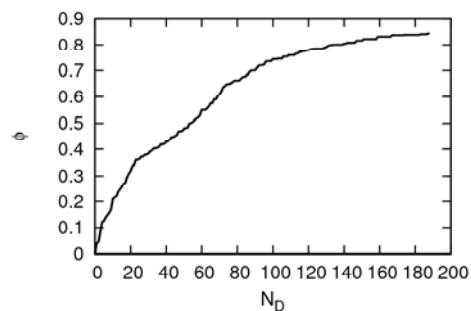


Рис. 1. График зависимости точности классификации Φ от числа детекторов N_D

моделі необхідно, чтобы число детекторов было меньше объема обучающей выборки.

Вместо числа детекторов в наборе предлагается использовать в качестве критерия останова степень покрытия пространства детекторами. Под степенью покрытия пространства заданным набором детекторов будем подразумевать соотношение объема пространства, занятого детекторами из набора D , к объему пространства, занятого «чужими» экземплярами:

$$p = \frac{V_D}{V_N}.$$

При этом считаем, что благодаря этапу цензурирования детекторы будут находиться только в «чужом» пространстве. Очевидно, что одиночный замаскированный детектор d может покрыть до $V(d)$ «чужих» экземпляров, где $V(d) = \prod_{i=1}^l \{2 \mid d_i = Z\}$.

Оценка покрытия группы детекторов не является тривиальной задачей. Так, например, два замаскированных детектора d_1 и d_2 могут покрыть до $V(d_1) + V(d_2) - V(d_1 \wedge d_2)$, где $V(d_1 \wedge d_2)$ – число «чужих» экземпляров, которые покрываются как детектором d_1 , так и детектором d_2 . Для вычисления покрытия большого числа детекторов расчеты еще более усложняются.

Поэтому предлагается использовать приближенную оценку области покрытия детекторов.

В качестве оценки можно использовать число детекторов, которые подверглись цензурированию. Такие детекторы в результате обучения не содержат ни одного замаскированного бита и в то же время реагируют на какой-либо «свой» экземпляр. Логично предположить, что в процессе обучения число таких детекторов будет

возрастать, а значит, его можно использовать в качестве критерия останова метода.

На рис. 2 приведены экспериментально полученные графики зависимости точности работы модели Φ (рис. 2, а) и числа отвергнутых на этапе цензурирования детекторов N_c (рис. 2, б) от числа итераций обучения модели I , используемой при решении задачи диагностирования заболеваний сердца [12].

Из рис. 2, а следует, что с ростом числа итераций I наблюдается увеличение точности классификации Φ , а также числа детекторов, подвергнутых цензурированию N_c , что за счет исключения избыточных детекторов позволяет повысить обобщающие свойства диагностической модели на основе клональной селекции.

Также можно оценить степень покрытия пространства признаков детекторами с помощью метода, предложенного в [13]. Необходимо сгенерировать достаточно большое число случайных экземпляров в пространстве признаков и вычислить соотношение числа экземпляров, распознанных моделью как «чужие» к их общему числу. При таком подходе невозможно оценить абсолютную степень покрытия детекторами, однако можно анализировать ее динамику, и при снижении темпов роста выполнять останова.

Существенным недостатком такого подхода является необходимость создания, хранения и обработки слишком большого набора случайных тестовых данных. В связи с этим целесообразно осуществлять такую проверку не на каждой итерации, а с некоторым шагом.

Приближенную оценку степени покрытия предлагается определять следующим образом.

1. Сгенерировать случайную пробу t . Проба представляет собой замаскированный детектор высокой строгости, т.е. проба описывает несколько расположенных близко точек в пространстве признаков U , при этом число этих точек и их сгруппированность определяется строгостью пробы $\tau(t)$. Предлагается задать строгость 0,2I.

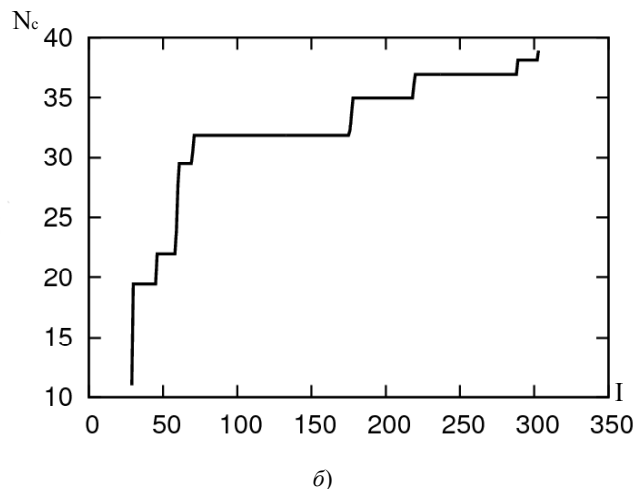
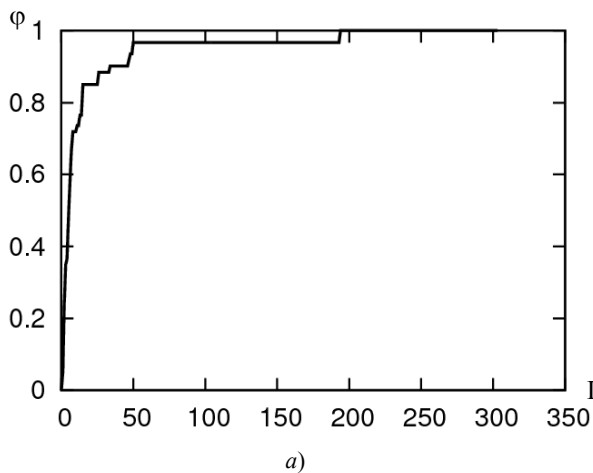


Рис. 2. График зависимости точности классификации Φ (а) и числа отвергнутых на этапе цензурирования детекторов N_c (б) от числа итераций I

2. Положить значение оценки покрытия $e^* = 1$.

3. Вычислить, насколько данная проба t распознается детекторами из набора D . Для этого для каждого детектора $d \in D$ необходимо выполнить подэтапы 3.1–3.4.

3.1. Положить: $e = 0$.

3.2. Если $\exists i, d_i \neq Z \wedge t_i \neq Z \wedge d_i \neq t_i$, то выбрать следующий детектор $d \in D$ и перейти к подэтапу 3.1.

3.3. Для всех $d_i = Z$, положить: $e = 2e$.

3.4. Если $e > e^*$, тогда положить: $e^* = e$. Выбрать следующий детектор $d \in D$ и перейти к подэтапу 3.1.

4. Останов.

В результате работы метода для пробы t значение e^* укажет, сколько точек, содержащихся в пробе, распознаются детекторами из набора D . Проанализировав данные о нескольких таких случайных пробах, можно судить об области покрытия детекторами модели.

Рекомендуется использовать вместо абсолютного значения e^* относительное значение:

$$e' = \frac{e^*}{2^{1-\tau(t)}}.$$

Это позволит использовать пробы различной строгости.

Останов процедуры обучения модели можно производить, если относительный показатель покрытия детекторов модели e' остается неизменным на протяжении нескольких последовательных итераций.

7. ЭКСПЕРИМЕНТЫ И РЕЗУЛЬТАТЫ

Для проведения экспериментов предложенная модель отрицательного отбора и метод ее обучения были реализованы на языке Python.

Предложенная модель отрицательного отбора обучалась и тестировалась на выборке SPECT [12], которая содержит сведения о пациентах с заболеваниями сердца. Каждый экземпляр описан 22 бинарными признаками и может быть отнесен к одному из двух классов (здоров или болен). Точность классификации после обучения составила 95 %. Для обеспечения такой точности потребовалось сформировать 47 детекторов. Малый размер набора детекторов показывает хорошую обобщающую способность предложенной модели.

Также предложенная модель отрицательного отбора применялась для решения задачи диагностирования повторных респираторных заболеваний у детей [14]. Классификация проводилась на основании данных о 23 бинарных признаках, которыми были представлены экземпляры выборки. В результате работы модели на основе 34 детекторов был получен набор продукционных правил, позволяющий определить предрасположенность пациентов к повторным респираторным заболеваниям. Полученная точность работы диагностической модели составила 98 %.

Результаты проведенных экспериментов свидетельствуют о работоспособности разработанного метода и реализующего его программного обеспечения, а также позволяют рекомендовать их для использования при решении практических задач диагностирования.

ВЫВОДЫ

С целью решения актуальной проблемы автоматизации процесса диагностирования объектов, характеризуемых набором бинарных признаков, разработано математическое обеспечение, позволяющее синтезировать диагностические модели для определения класса состояний объектов диагностирования на основе парадигмы отрицательного отбора.

Научная новизна работы заключается в том, что получили дальнейшее развитие модель отрицательного отбора и метод ее обучения, которые модифицированы для использования маскированных детекторов, а также впервые предложен критерий останова метода обучения, что позволяет решать задачи диагностирования с использованием меньшего числа детекторов. Это повышает скорость обучения модели и позволяет снизить требования к вычислительным ресурсам. Также, благодаря маскированию битов в детекторе, учитываются только наиболее информативные признаки, и облегчается получение продукционных правил на основе набора детекторов.

Практическая ценность работы заключается в том, что разработана программная реализация модели отрицательного отбора с использованием маскированных детекторов, которая может использоваться для решения практических задач технического и медицинского диагностирования.

Работа выполнена в рамках госбюджетных тем «Информационные технологии автоматизации распознавания образов и принятия решений для диагностирования в условиях неопределенности на основе гибридных нечеткологических, нейросетевых и мультиагентных методов вычислительного интеллекта» (№ гос. регистрации 0109U007673), «Методы, модели и устройства принятия решений в системах распознавания образов» (№ гос. регистрации 0110U002031).

СПИСОК ЛИТЕРАТУРЫ

1. Субботін, С. О. Подання й обробка знань у системах штучного інтелекту та підтримки прийняття рішень: навч. посібник / С. О. Субботін. – Запоріжжя : ЗНТУ, 2008. – 341 с.
2. Нейроинформатика / [А. Н. Горбань, В. Л. Дунин-Барковский, А. Н. Кирдин и др.]. – Новосибирск : Наука. Сибирское предприятие РАН, 1998. – 296 с.
3. Quinlan, J. R. C4.5: Programs for Machine learning / J.R. Quinlan. – Massachusetts : Morgan Kaufmann Publishers, 1993. – 302 p.
4. Ji, Z. Revisiting negative selection algorithms / Z. Ji, D. Dasgupta // Evolutionary Computation. – 2007. – № 15. – P. 223–251.
5. Stibor, T. On the Use of Hyperspheres in Artificial Immune Systems as Antibody Recognition Regions / T. Stibor, J. Timmis, C. Eckert // Artificial Immune Systems : 5th international conference ICARIS-2006, Oeiras, 4–6 September 2006 : proceedings. – New York : Springer-Verlag, 2006. – P. 215–228.

6. *Forrest, S.* Self-Nonself Discrimination in a Computer / S. Forrest, A. S. Perelson, R. Cherukuri, L. Allen // Research in Security and Privacy : IEEE Symposium, Oakland, 16–18 May 1994 : proceedings. – Los Alamitos: IEEE, 1994. – P. 202–212.
7. *D'haeseleer, P.* An immunological approach to change detection: algorithms, analysis, and implications / P. D'haeseleer, S. Forrest, P. Helman // Computer Security and Privacy : IEEE Symposium, Oakland, 6–8 May 1996 : proceedings. – Los Alamitos: IEEE, 1996. – P. 110–119.
8. *Hofmeyr, S.* Architecture for an artificial immune system / S. Hofmeyr, S. Forrest // Evolutionary Computation. – 2000. – № 8. – P. 443–473.
9. *Balthrop, J.* Revisiting LISYS: parameters and normal behavior / J. Balthrop, S. Forrest, M. Glickman // Evolutionary Computation: congress CEC'02, Honolulu, 12–17 May 2002 : proceedings. – Los Alamitos: IEEE, 2002. – P. 1045–1050.
10. *Balthrop, J.* Coverage and generalization in an artificial immune system / J. Balthrop, F. Esponda, S. Forrest, M. Glickman // Genetic and Evolutionary Computation: conference GECCO-2002, New York, 9–13 July 2002 : proceedings. – San Francisco: Morgan Kaufmann, 2002. – P. 3 10.
11. *Farmer, J.* The immune system, adaptation, and machine learning / J. Farmer, N. Packard, A. Perelson // Physica D : Nonlinear Phenomena. – 1986. – № 2. – P. 187–204.
12. *Frank, A.* SPECT heart data set [Electronic resource] / A. Frank, A. Asuncion. – Irvine : University of California, 2010. – Access mode : <http://archive.ics.uci.edu/ml/datasets/SPECT+Heart>.
13. *Ji, Z.* Negative selection algorithms: from the thymus to V-detector : dissertation ... doctor of philosophy / Z. Ji. – Memphis : The University of Memphis, 2006. – 337 p.
14. *Герасимчук, Т. С.* Использование искусственных иммунных систем для прогнозирования риска развития рекуррентных респираторных инфекций у детей раннего воз-

раста / Т. С. Герасимчук, С. А. Зайцев, С. А. Субботин // Диагностика та лікування інфекційно опосередкованих соматичних захворювань у дітей: міжрегіональна науково-практична конференція, Донецьк, 10–11 лютого 2011 : матеріали. – Донецьк : Норд-прес, 2011. – С. 27–29.

Стаття надійшла до редакції 04.05.2011.

Зайцев С. О., Суботін С. О.

ПОБУДОВА ДІАГНОСТИЧНИХ МОДЕЛЕЙ НА ОСНОВІ ПАРАДИГМИ НЕГАТИВНОГО ВІДБОРУ З ВИКОРИСТАННЯМ ПРИНЦИПУ МАСКУВАННЯ ДЕТЕКТОРІВ

Дістали подальшого розвитку модель негативного відбору з використанням маскованих детекторів та метод її навчання. Досліджувалися різноманітні критерії зупину для методу навчання моделі. Розроблено критерій зупину, що дозволяє виконувати своєчасний зупин процесу навчання, і таким чином прискорити даний процес, а також скоротити використання обчислювальних ресурсів.

Ключові слова: негативна селекція, маскований детектор, критерій зупину.

Zaitsev S. A., Subbotin S. A.

THE DIAGNOSIS MODEL BUILDING ON THE BASIS OF NEGATIVE SELECTION PARADIGM USING THE PRINCIPLE OF DETECTOR MASKING

A new negative selection model using masked detectors with training method has been developed. Various stopping criteria for the training method have been investigated. The stopping criterion is proposed. It allows to terminate training procedure, to speed-up training process and to reduce computational resources.

Key words: negative selection, masked detector, stopping criterion.

УДК [519.674:519.688]:62-11

Чопоров С. В.

Магістр Запорозького національного університета

ПОСТРОЕНИЕ НЕРАВНОМЕРНЫХ ДИСКРЕТНЫХ СЕТОК ДЛЯ ФУНКЦИОНАЛЬНЫХ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ НА БАЗЕ ТЕОРИИ R-ФУНКЦИЙ

В работе рассмотрена проблема математического моделирования геометрических сеток на базе теории R-функций. Предложен подход к построению неравномерных сеток шестигранных элементов.

Ключевые слова: R-функция, сетка, шестигранный элемент, математическая модель.

ПОСТАНОВКА ПРОБЛЕМЫ

Исследование прочности и долговечности проектируемых инженерных конструкций является важной составляющей современной техники. В настоящее время для этих целей широко применяется компьютерное моделирование, позволяющее заменить длительное и дорогостоящее испытание опытного образца изучением соответствующих математических моделей. Следовательно, актуальной является проблема построения математических моделей геометрических объектов, которые

в пространстве занимают определенный объем. Например, сложные инженерные конструкции, сооружения, машины, механизмы и прочее. По сути, речь идет о решении так называемой обратной задачи аналитической геометрии, когда для имеющейся геометрической области строится ее аналитическое описание.

Множество физических объектов реального мира можно определить как сплошные тела, геометрическое моделирование которых можно рассматривать как процесс формализации представления геометрии существу-

ющего или воображаемого объекта. Наиболее общим методом определения множества точек, принадлежащих объекту, является введение предиката, являющегося индикатором принадлежности объекту точки пространства (неявное определение). Простейшей формой такого предиката будет ограничение на знак некоторой действительной функции в виде

$$\Omega = \{(x, y, z) | F(x, y, z) \geq 0\},$$

где неявная функция $F(x, y, z)$ больше нуля внутри области, соответствующей телу Ω , равна нулю на границе и меньше нуля вне тела. Например, для построения математической модели тела в форме эллипсоида с центром в точке (x_0, y_0, z_0) необходимо определить коэффициенты a , b и c в формуле

$$F_{\text{ellipsoid}}(x, y, z) = 1 - \frac{(x - x_0)^2}{a^2} - \frac{(y - y_0)^2}{b^2} - \frac{(z - z_0)^2}{c^2}.$$

Математические модели геометрических объектов нетривиальной формы могут быть определены при помощи более сложных функций конструктивно, используя логические комбинации более простых функций. Такие действия с математическими функциями логически эквивалентны стандартным операциям над множествами. В работах Владимира Логвиновича Рвачева [1–3] разработан достаточно общий подход, получивший название теории R -функций, в котором предложены принадлежащие классу C^m функции для описания теоретико-множественных операций. Наиболее распространенная на практике система R -функций имеет вид (1). Например, тело в форме объединения двух шаров (рис. 1), один из которых радиуса r_0 с центром в начале координат, второй – радиуса r_1 с центром в точке $(0, r_0, 0)$ можно описать формулой (2).

$$\left\{ \begin{array}{l} C \equiv \text{const}, \\ \neg x \equiv -x, \\ x_1 \wedge x_2 \equiv x_1 + x_2 - \sqrt{x_1^2 + x_2^2}, \\ x_1 \vee x_2 \equiv x_1 + x_2 + \sqrt{x_1^2 + x_2^2}. \end{array} \right. \quad (1)$$

$$F_{\text{balls}}(x, y, z) = (r_0^2 - x^2 - y^2 - z^2) \vee [r_1^2 - x^2 - (y - r_0)^2 - z^2]. \quad (2)$$

Поскольку математические модели геометрических объектов в основном используются при численном анализе различных параметров состояния сложных инженерно-технических объектов, на практике требуется построение конечно-элементных дискретных математических моделей. Для этого при использовании R -функций, в первую очередь, требуется разработка методов построения дискретных моделей областей, описанных с помощью неявных функций.

Среди конечных элементов, используемых в пространстве, наиболее распространенными являются элементы тетраэдрической и шестигранной форм. Преимуществом первого типа элементов является большая топологическая гибкость, второго – возможность уменьшения размерности систем линейных алгебраических уравнений за счет использования меньшего количества элементов (необходимо использовать пять тетраэдров для представления объема элемента, представленного одним шестигранником, рис. 2).

При этом функции формы линейного тетраэдрического конечного элемента имеют вид

$$N_T(x, y, z) = \alpha_0 + \alpha_1 x + \alpha_2 y + \alpha_3 z, \quad \alpha_i \in R, i = \overline{0, 3}, \quad (3)$$

а функции формы шестигранного конечного элемента –

$$N_H(x, y, z) = \alpha_0 + \alpha_1 x + \alpha_2 y + \alpha_3 z + \alpha_4 xy + \alpha_5 xz + \alpha_6 yz + \alpha_7 xyz, \quad \alpha_i \in R, i = \overline{0, 7}. \quad (4)$$

Наличие нелинейных членов в соотношении (4) приводит к тому, что градиенты шестигранного конечного элемента (в отличие тетраэдрического) не постоянны и изменяются вдоль одной из координатных плоскостей. Таким образом, применение шестигранных конечных элементов является более предпочтительным в вычислительном плане.

На практике в основном требуется построение нерегулярных сеток, где количество конечных элементов скачкообразно в местах, где расположены так называемые сингулярности (особенности) объекта (острые углы, отверстия, трещины и т. п.). С другой стороны, построение первоначального приближенного разбиения требует последующей его оптимизации.

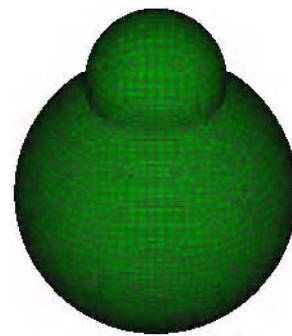


Рис. 1. Тело в форме объединения двух шаров: $r_0 = 0,8$, $r_1 = 0,4$

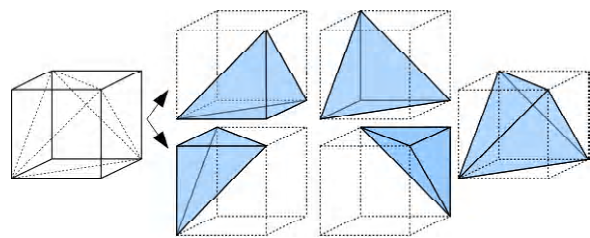


Рис. 2. Представление шестигранника пятью тетраэдрами

Таким образом, разработка методов построения неравномерных дискретных моделей трехмерных объектов, заданных при помощи R -функций, на базе шестигранных конечных элементов является сложной и актуальной задачей.

АНАЛИЗ ПУБЛИКАЦИЙ ПО ТЕМЕ ИССЛЕДОВАНИЯ

Методы построения нерегулярной сетки шестигранных конечных элементов можно разделить на прямые и не прямые. Прямые методы используют непосредственное построение шестигранных ячеек внутри заданной границы, в то время как не прямые базируются на преобразовании некоторого исходного дискретного представления области.

Среди прямых методов можно выделить фронтальные методы (наиболее распространенной тут идеей является исчерпание области пласт за пластом, формируя таким образом и дискретизацию границы, и дискретизацию внутренней части области [4]) и методы, использующие блочную декомпозицию [5–7] или суперпозицию [8–10].

Методы декомпозиции и блочной декомпозиции основаны на идее (полуавтоматической или автоматической) декомпозиции тела на кубоподобные блоки, для каждого из которых дискретизация может быть получена отображением на дискретизацию единичного куба, с последующим объединением дискретизаций частей. Для автоматизации процесса декомпозиции на блоки используются срединные поверхности. Достоинством данной группы методов является их высокая скорость и качество получаемой сетки. Основным недостатком – затруднительность их применения для дискретизации сложных нестандартных областей.

Основная идея методов суперпозиции заключается в использовании начальной сетки, которая может быть более или менее просто создана для части пространства, содержащей объект, и последующей адаптации сетки к границе области. Ключевым шагом данной группы методов является генерация сетки высокого качества около границы объекта. Для адаптации начальной сетки могут быть использованы техники изоморфизма [8, 9] и проекции. При всей прагматичности такого подхода результирующие алгоритмы данного типа методов являются весьма трудоемкими для реализации [11].

Непрямые методы можно разделить на фронтальные, в основе которых идея послышной конвертации начальной сетки тетраэдров [12, 13], методы, использующие двойственное представление (STC-представление [14]) и методы поэлементной конвертации (например, на основе шаблонов). Преимуществом данной группы методов является отсутствие вырожденных элементов. Недостатком – необходимость построения качественной начальной дискретизации, что само по себе является трудоемкой задачей.

МЕТОД СУПЕРПОЗИЦИИ ДЛЯ ФУНКЦИОНАЛЬНОГО ПРЕДСТАВЛЕНИЯ НА БАЗЕ ТЕОРИИ R -ФУНКЦИЙ

Пусть тело Ω представлено неявной функцией на базе теории R -функций. Использование методов теории R -функций при неявном представлении тела Ω требует учета особенностей такого представления при разработке алгоритмов получения дискретного представления. Неявное представление дает правило для проверки принадлежности точки Ω , однако, не дает правило генерации системы точек и топологии элементов, образующих Ω . Это делает более предпочтительным использование в качестве алгоритмической базы метода суперпозиции совместно с сеточным методом и техникой изоморфизма [8–10] на шаге адаптации пограничного слоя элементов. Тогда общий алгоритм получения дискретного представления тела заданного неявной функцией можно представить схемой, приведенной на рис. 3.

На первом шаге алгоритма для формирования начальной сетки достаточно покрыть область равномерной сеткой с шагом h и определить значение R -функций в узлах сетки. Затем последовательно отсеивать элементы, в узлах которых есть внешние точки или которые не соответствуют критерию близости к границе области (на расстоянии $\Delta = 0,5h$ есть внешние точки).

Граница полученной таким образом начальной сетки является многогранником, каждая грань которого является четырехугольником. Следовательно, ее можно рассматривать как неструктурированную сетку четырехугольных элементов, для которой может быть получена изоморфная (однозначно соответствующая) сетка на границе тела путем установления для каждого узла соответствующего граничного, а для каждого ребра границы начальной сетки – ребро на границе. Таким образом,

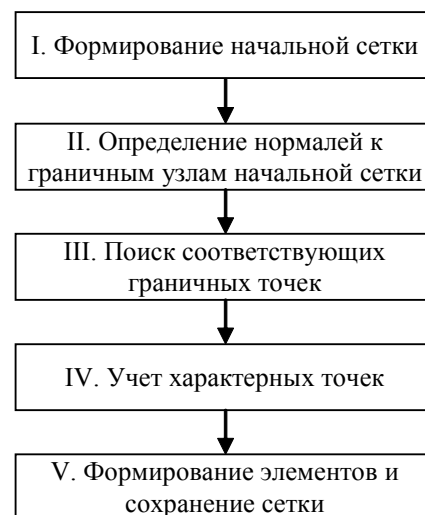


Рис. 3. Общая схема построения дискретной модели

каждому четырехугольнику будет соответствовать четырехугольная грань на границе тела, которые совместно определяют шестигранный элемент.

На втором шаге к каждому граничному узлу начальной сетки определяется нормаль как среднее арифметическое нормалей смежных в узле граней (рис. 4). Полученные нормали определяют направления для поиска соответствующих граничных точек на третьем шаге.

На четвертом шаге происходит учет характерных точек (геометрических особенностей, концентраторов напряжений и т. д.), так как потеря таковых может приводить к значительной утере точности при использовании полученного дискретного представления численными методами. При этом характерные точки могут быть учтены как при помощи техники перемещения необходимого граничного узла, так и при помощи коррекции направления нормали соответствующего узла начальной сетки и последующей репроэкции.

На последнем шаге алгоритма при помощи техники изоморфизма формируются элементы, соединяя узлы начальной сетки с соответствующими узлами на границе (рис. 5).

Следует отметить, что в результате применения описанного подхода генерируются элементы приблизительно одинакового размера, что позволяет применять для последующего улучшения качества сеток методы на основе локального сглаживания Лапласа.

УНИВЕРСАЛЬНЫЙ ШАБЛОН ДЛЯ ПОСТРОЕНИЯ НЕРАВНОМЕРНЫХ СЕТОК

Наиболее распространенным подходом к получению неравномерной начальной сетки для методов на основе суперпозиции является использование множества шаб-

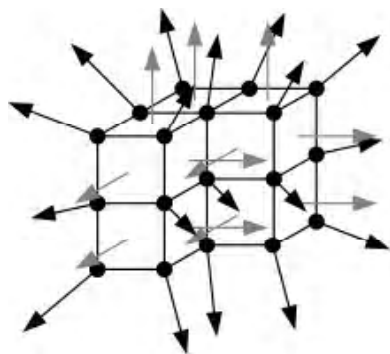


Рис. 4. Определение нормалей к узлам начальной сетки

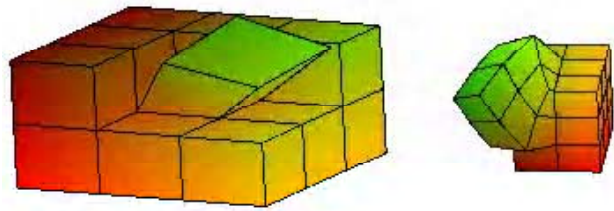


Рис. 5. Построение граничного слоя шестигранных элементов

лонов сгущения с таблицами соответствия шаблонов [8–10]. Идея алгоритма заключается в рекурсивном разбиении области, в которую помещен исходный объект, на 27 элементов до тех пор, пока в узлах сетки удовлетворяет некоторая управляющая функция (пример такого сгущения на плоскости приведен на рис. 6), с последующим восстановлением топологии сетки. При этом в одном узле не должны быть смежными части, у которых разность уровней рекурсии больше единицы.

При построении неравномерной адаптивной сетки для трехмерной области с учетом вращения получается 22 различных конфигурации (256 без учета вращения, рис. 7), покрывающих все возможные ситуации взаимного расположения частей с разными уровнями рекурсии. Такое количество необходимых конфигураций делает трудным их построение, верификацию и программирование.

В работах [8, 9] предложено использовать упрощенное множество шаблонов, которое изображено на рис. 8.

В работе [10] разработано альтернативное множество шаблонов (рис. 9), которое позволяет генерировать неравномерные сетки с меньшим использованием шаблона на 8 (разбиения куба на 27 частей).

Однако и первое, и второе множество шаблонов в ряде случаев приводит к необходимости использования шаблона, соответствующего конфигурации 8 (разбиению на 27 элементов), что приводит к избыточному сгущению сетки. Для решения этой задачи, в таких случаях, можно использовать универсальный шаблон (рис. 10), конфигурации граничных элементов которого будут соответствовать одному из шаблонов множества 2.

Для корректного определения координат внутренних узлов может быть использовано изопараметрическое преобразование. Такой подход может быть использован также для сгущения построенных дискретных моделей.

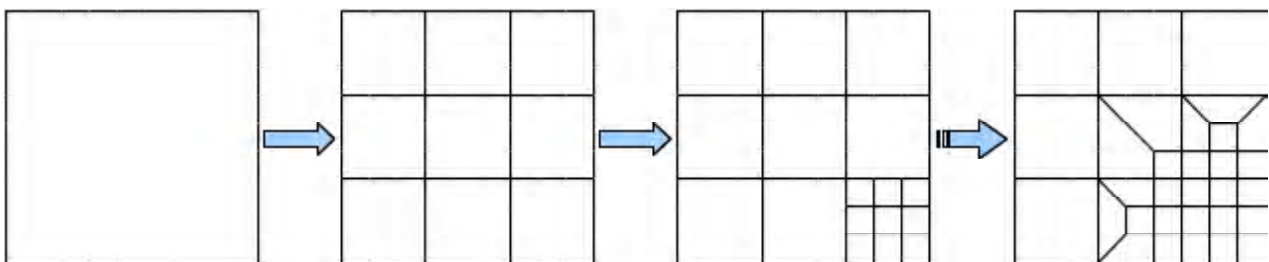


Рис. 6. Идея сгущения сетки

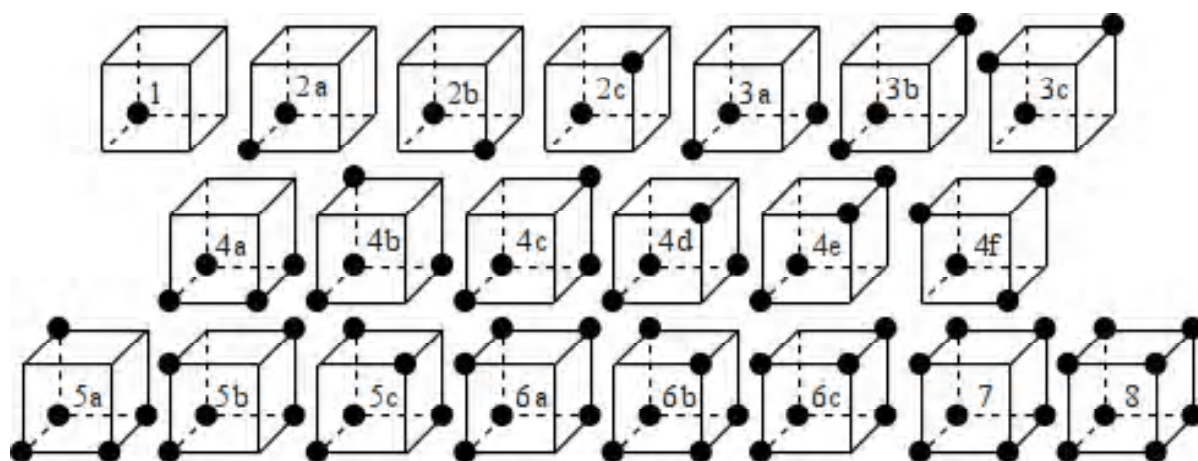


Рис. 7. Возможные конфигурации в трехмерном пространстве

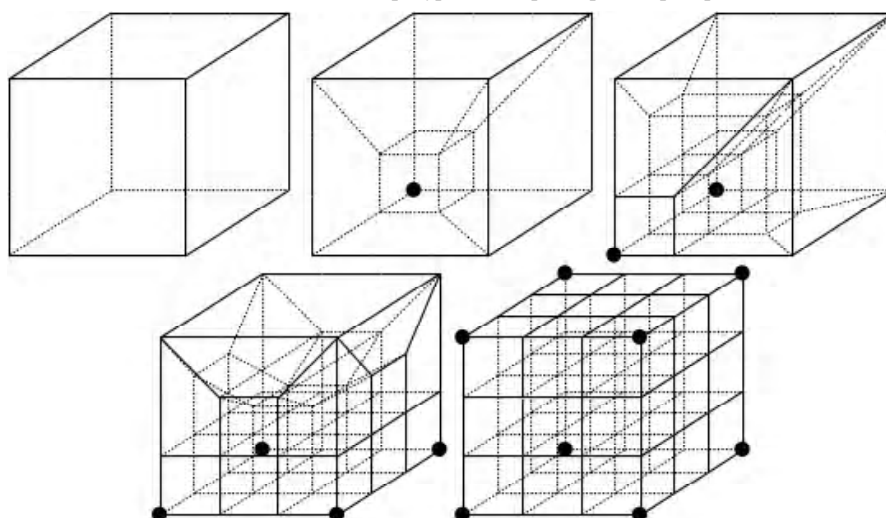


Рис. 8. Множество шаблонов 1

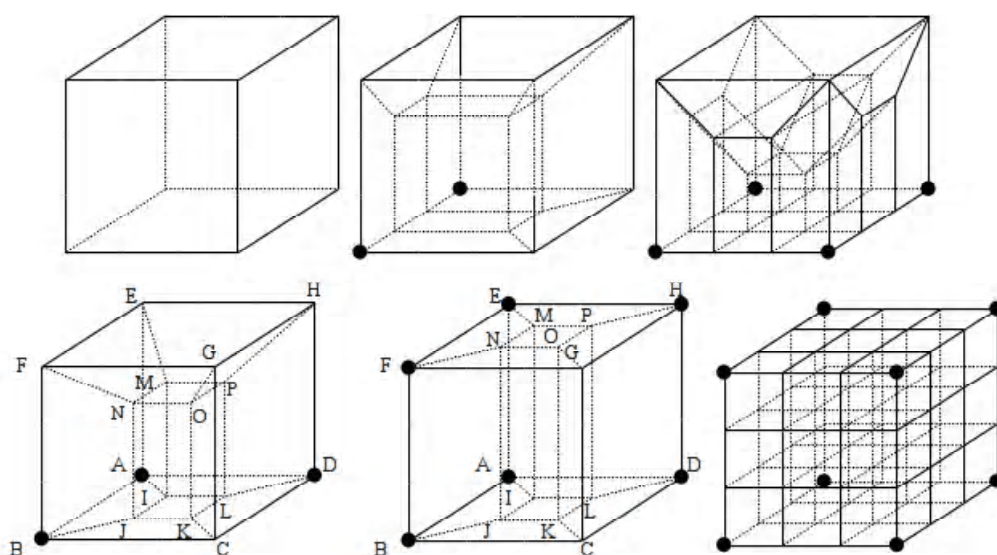


Рис. 9. Множество шаблонов 2

ВЫВОДЫ

Таким образом, в работе разработан подход к получению дискретных моделей, соответствующих неявным аналитическим моделям, построенным на базе теории

R -функций. Предложен подход к построению неравномерных дискретных моделей на базе разработанного универсального шаблона. Результаты работы реализованы в виде подсистемы САПР, которая может быть ис-

пользована при моделюванні і проектуванні різних технічних деталей і об'єктів (некоторые примеры приведены на рис. 11).

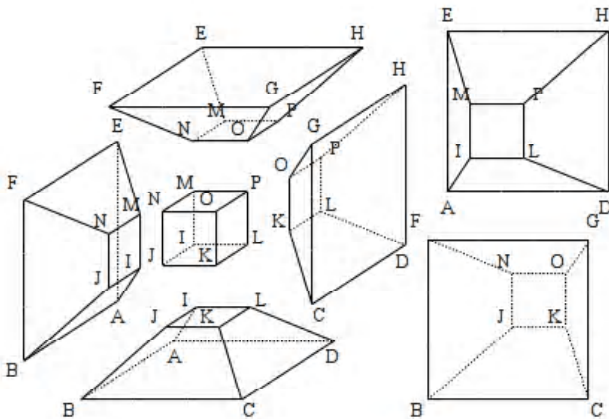


Рис. 10. Универсальный шаблон

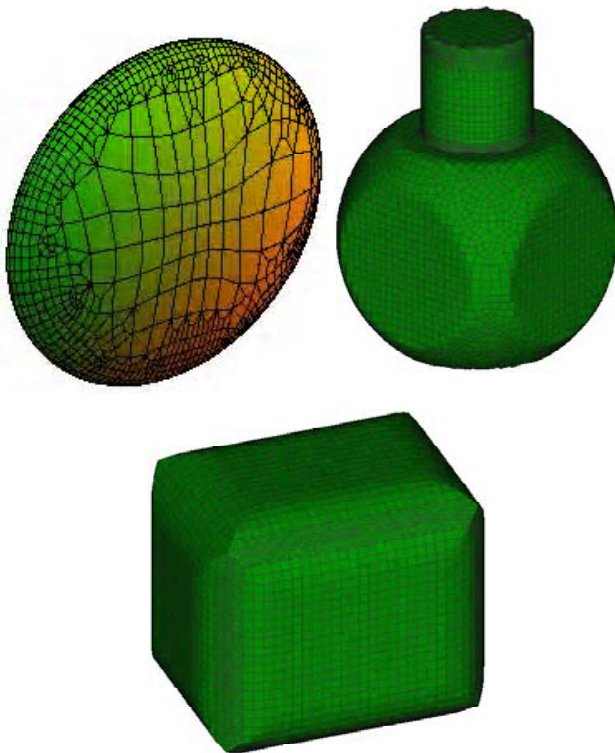


Рис. 11. Примеры неравномерных сеток

СПИСОК ЛІТЕРАТУРИ

1. Рвачев, В. Л. Теория R -функций и некоторые ее приложения / В. Л. Рвачев. – К. : Наукова думка, 1982. – 552 с.
2. Рвачев, В. Л. Новые подходы к построению уравнений трехмерных локусов с помощью R -функций / В. Л. Рвачев, А. В. Толок, Р. А. Уваров, Т. И. Шейко // Вісник Запорізького державного університету. – 2000. – № 2. – С. 119–130.
3. Рвачев, В. Л. Введение в теорию R -функций / В. Л. Рвачев, Т. И. Шейко // Проблемы машиностроения. – 2001. – Т. 4, № 1–2. – С. 46–58.
4. Staten, M. L. Unconstrained Paving & Plastering: A New Idea for All Hexahedral Mesh Generation / M. L. Staten,

- S. J. Owen, T. D. Blacker // Proceedings, 14th International Meshing Roundtable. – Sandia National Laboratories: Springer-Verlag, 2005. – P. 399–416.
5. Armstrong, C. G. Medials for Meshing and More / C. G. Armstrong, D. J. Robinson, R. M. McKeag, T. S. Li, S. J. Bridgett, R. J. Donaghy, C. A. McGleenan // Proceedings, 4th International Meshing Roundtable. – Sandia National Laboratories, 1995. – P. 277–288.
6. Holmes, D. J. Generalized Method of Decomposition Solid Geometry into Hexahedron Finite Elements / D. J. Holmes // Proceedings, 4th International Meshing Roundtable. – Sandia National Laboratories, 1995. – P. 141–152.
7. Liu, S.-S. Automatic Hexahedral Mesh Generation by Recursive Convex and Swept Volume Decomposition / S.-S. Liu, R. Gadh // Proceedings, 6th International Meshing Roundtable. – Sandia National Laboratories, 1997. – P. 217–231.
8. Schneiders, R. A Grid-based Algorithm for the Generation of Hexahedral Element Meshes / R. Schneiders // Engineering with Computers. – 1996. – № 12. – P. 168–177.
9. Schneiders, R. Octree-based Generation of Hexahedral Element Meshes / R. Schneiders, R. Schindler, F. Weiler // 5th Annual International Meshing Roundtable. – 1996. – P. 205–216.
10. Ito, Y. Octree-based reasonable-quality hexahedral mesh generation using a new set of refinement templates / Yasushi Ito, Alan M. Shih, Bharat K. Soni // International Journal For Numerical Methods in Engineering. – 2009. – № 77 (13). – P. 1809–1833.
11. Thompson, J. F. Hand book of grid generation / Joe F. Thompson, Bharat K. Soni, Nigel P. Weatheril. – New York: CRC Press, 1999. – 1200 p.
12. Owen, S. J. H-Morph: An Indirect Approach to Advancing Front Hex Meshing / S. J. Owen, S. Saigal // International Journal for Numerical Methods in Engineering. – 2000. – № 1 (33). – P. 289–312.
13. Owen, S. J. Hex-domain mesh generation using 3D constrained triangulation / S. J. Owen // Computer Aided Design. – 2001. – № 3(33). – P. 211–220.
14. Murdoch, P. The spatial twist continuum: A connectivity based method for representing all-hexahedral finite element meshes / P. Murdoch, S. E. Benzley, T. D. Blacker, S. A. Mitchel // Finite Elements in Analysis and Design. – 1997. – № 28. – P. 137–149.

Стаття надійшла до редакції 11.05.2011.

Choporov S. V.

GENERATION OF NONUNIFORM HEXAHEDRAL ELEMENT MESHES FOR FUNCTIONAL MODELS ON THE BASIS OF R -FUNCTIONS

The problem of mathematical modeling of geometrical objects on the basis of R -functions is described in the article. Author propose an approach for generation of nonuniform hexahedral element meshes.

Key words: R -function, mesh, hexahedral element, mathematical model.

Чопоров С. В.

ПОБУДОВА НЕРІВНОМІРНИХ ДИСКРЕТНИХ СІТОК ДЛЯ ФУНКЦІОНАЛЬНИХ МАТЕМАТИЧНИХ МОДЕЛЕЙ НА БАЗІ ТЕОРІЇ R -ФУНКЦІЙ

В роботі розглянута проблема математичного моделювання геометричних об'єктів на базі теорії R -функцій. Запропоновано підхід до побудови нерівномірних сіток шестигранних елементів.

Ключові слова: R -функція, сітка, шестигранний елемент, математична модель.

ОБЪЕКТНО-ОРИЕНТИРОВАННЫЙ ПОДХОД ДЛЯ ПОСТРОЕНИЯ МОДЕЛЕЙ СЛОЖНЫХ СИСТЕМ

Предложена методика построения динамической модели системы с локализованными свойствами на основе объектно-ориентированного подхода. Рассмотрен пример системы автоматического регулирования температуры рабочего тела с пропорционально-интегрально-дифференциальным законом регулирования, приведена методика построения классов и функций, составляющих программную реализацию модели, обсуждаются возможности предлагаемой методики.

Ключевые слова: компьютерное моделирование систем, моделирование динамики системы, объектно-ориентированный подход, система терморегулирования.

ВВЕДЕНИЕ

Параллельно с развитием аппаратных и программных средств вычислительной техники интенсивно эволюционируют и общие подходы моделирования сложных систем различной природы. Однако имеющиеся в настоящее время методы построения динамических моделей технических систем все же не удовлетворяют в полной мере растущим запросам разработчиков систем различной этимологии [1, 2]. Остается открытым вопрос о формализации синтеза компьютерной модели сложной системы из заданного множества субмоделей. Определенным потенциалом в создании более продвинутых методов моделирования систем обладает подход, который называют объектно-ориентированным.

МЕТОДИКА ПОСТРОЕНИЯ ДИНАМИЧЕСКОЙ МОДЕЛИ СИСТЕМЫ

Ниже рассмотрена методика построения динамической модели технической системы на основе объектно-ориентированного подхода и ее программной реализации. Хорошим примером (не слишком громоздким и, в то же время, достаточно характерным) системы как объекта моделирования, является система автоматического регулирования температурой некоторого рабочего тела, структура которой показана на рис. 1. Модель системы в нашем примере состоит из следующих элементов:

- 1) рабочее тело;
- 2) нагреватель;
- 3) регулятор;

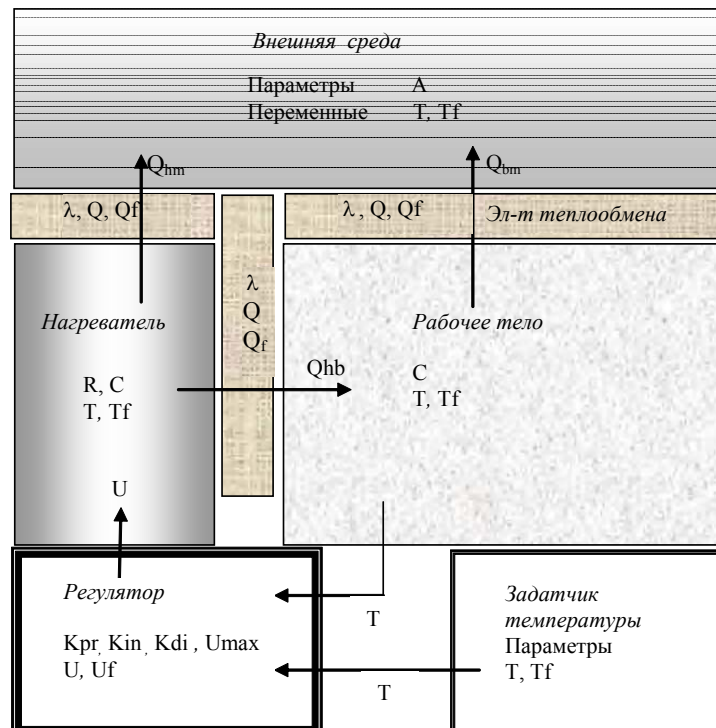


Рис. 1. Схема взаимодействия объектов в системе терморегулирования

- 4) элемент теплообмена «нагреватель – внешняя среда»;
- 5) элемент теплообмена «рабочее тело – внешняя среда»;
- 6) элемент теплообмена «нагреватель – рабочее тело»;
- 7) датчик температуры;
- 8) внешняя среда.

Принимается следующая методология моделирования.

Элементы системы типизированы, т. е. каждый из элементов системы принадлежит определенному виду. Модель элемента системы данного вида представлена соответствующим классом. Состояние элемента системы определяется внутренними процессами, отображаемыми в субмодели, и воздействиями на него других элементов системы. Состояние элемента системы описывается набором переменных его состояния.

Каждый из элементов системы получает информацию о состоянии тех элементов системы, которые оказывают на него влияние. Обобщенная информация о наличии такого влияния для каждого элемента системы представлена матрицей межэлементных воздействий M , которая для рассматриваемой системы имеет следующий вид:

	1	2	3	4	5	6	7	8
1	0	0	0	0	1	1	0	0
2	0	0	1	1	0	1	0	0
3	1	0	0	0	0	0	1	0
4	0	1	0	0	0	0	0	1
5	1	0	0	0	0	0	0	1
6	1	1	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0

Номера строк и номера столбцов в этой матрице соответствуют номерам элементов системы, перечисленным ранее. Единичное значение элемента матрицы M_{ik} означает, что элемент с номером i испытывает воздействие со стороны k -того элемента системы. Матрица M используется для построения интерфейсов функций, описывающих поведение элементов системы.

Моделирование поведения системы выполняется на заданном отрезке времени $[0, t_m]$. Временной шаг моделирования может быть переменным. При этом выполнение каждого шага включает в себя 3 этапа.

1. Каждый элемент системы получает информацию о состояниях воздействующих на него в данный момент времени элементов.

2. Каждый из элементов системы вычисляет свое состояние в следующий момент времени с учетом величины текущего шага моделирования (не переходя при этом в следующее состояние). Вычисления выполняются на основе соответствующей субмодели элемента системы.

3. Каждый из элементов системы реализует (переводит себя в) следующее состояние.

В соответствии с идеологией объектно-ориентированного моделирования (ООМ) все элементы модели системы типизированы, т. е. относятся к определенному классу и каждый элемент модели системы представляется объектом соответствующего класса. При этом струк-

тура, свойства и поведение объекта данного класса, порождаемого моделирующей программой, однозначно определяется описанием этого класса. Класс определяет информационную структуру элемента модели системы и содержит набор функций (методов), определяющих эволюцию его состояния. При этом структура межэлементных взаимодействий, определяемая матрицей M , представлена в соответствующих классах в виде списков аргументов функций – членов класса, осуществляющих выполнение второго этапа очередного шага моделирования системы.

СТРУКТУРА И ОПИСАНИЕ КЛАССОВ ЭЛЕМЕНТОВ СИСТЕМЫ

Обобщенную структуру класса, описывающего произвольный элемент системы, можно представить следующим образом:

```
class <имя типа элемента>
{
    <список параметров модели элемента системы
    с указанием их типов>
    <описание переменных состояния текущего мо-
    мента времени>
    <описание переменных состояния следующего
    момента времени>
    void next(список аргументов – состояний воз-
    действующих элементов);
    void step(); // реализация
    следующего состояния системы
    <конструкторы и другие стандартные элементы
    класса>
};
```

Приведем описание каждого из классов рассматриваемой модели. Описание конструкторов и других стандартных элементов классов опустим для краткости.

Класс Body (рабочее тело)

```
class Body
{
public:
    double C;
    double T, Tf;
    void next(Inter<Heater, Body>&,
    Inter<Body, Medium>&);
    void step() { T=Tf; }
    Body() { C=0.0; T=0.0; Tf=0.0; }
    Body(double c, double x) { C=c; T=x; }
};

void Body::next(Inter<Heater, Body>& hb,
Inter<Body, Medium>& bm)
{
    Tf= T+(h/C)*(hb.Q-bm.Q);
}
```

Модель элемента системы типа Body имеет один параметр: теплоемкость рабочего тела C , по одной переменной состояния для текущего и следующего моментов времени T и Tf (температура рабочего тела). Функция next, вычисляющая следующее состояние системы (т. е. значение Tf), через список своих аргументов получает информацию о состояниях элементов теплообмена «нагреватель – рабочее тело» и «рабочее тело – внешняя среда». Приведенный алгоритм для функции next

класа Body соответствует рабочей формуле первого порядка интегрирования дифференциального уравнения теплового баланса рабочего тела:

$$\frac{dT}{dt} = \frac{1}{C} (Q_{hb} - Q_{bm}), \quad (1)$$

где Q_{hb} – тепловой поток от нагревателя к рабочему телу, Q_{bm} – тепловой поток от рабочего тела во внешнюю среду (в приведенном описании класса эти величины имеют имена $hb.Q$ и $bm.Q$).

Класс Heater (нагреватель)

```
class Heater
{ public:
double R, C;
double T, Tf;
void next(Regul&, Inter<Heater,Medium>&,
Inter<Heater,Body>&);
void step() { T=Tf; }
Heater(double r, double c, double To) { R=r;
C=c; T=To; }
};
void Heater::next(Regul& reg,
Inter<Heater,Medium>& hm, Inter<Heater,Body>& hb)
{ double P=sqr(reg.U)/R;
Tf=T+(h/C)*(Phm.Q-hb.Q);
}
```

Модель нагревателя имеет два параметра: R – электрическое сопротивление нагревателя, C – его теплоемкость. Параметром состояния является температура (переменные T и Tf). Функция `next` получает информацию о состоянии регулятора, а также элементов теплообмена «нагреватель – внешняя среда» и «нагреватель – рабочее тело». Алгоритм функции `next` включает в себя вычисление тепловой мощности нагревателя (входного теплового потока):

$$P = \frac{U^2}{R}, \quad (2)$$

и температуры для следующего момента времени. Для вычисления последней использована рабочая формула для уравнения теплового баланса нагревателя такого же вида, как и в случае рабочего тела.

Класс Regul (регулятор)

```
class Regul
{public:
double Umax, Kpr, Kin, Kdi;
double Uf, U;
void next(Body&, Targ&);
void step() { U=Uf; }
Regul(double Um, double kpr, double kin, double kdi)
{Umax=Um; Kpr=kpr; Kin=kin; Kdi=kdi; U=0; }
};
void Regul::next(Body& body, Targ& targ)
{static int M; M++;
```

```
static double S=0.0;
static Body body0;
static Targ targ0;
if (M==1) { body0=body; targ0=targ; }
double Dx=targ.T-body.T, Dxo=targ0.T-body0.T,
D=(Dx-Dxo)/h;
S+=h*(Dx+Dxo)/2;
Uf=Kpr*Dx+Kin*S+Kdi*D;
if (Uf<0) Uf=0.0;
if (Uf>Umax) Uf=Umax;
body0=body; targ0=targ;
}
```

Субмодель регулятора соответствует универсальному пропорционально-интегрально-дифференциальному закону регулирования. Параметрами модели регулятора являются величины: K_{pr} – коэффициент пропорционального регулирования, K_{in} – коэффициент интегрального регулирования, K_{di} – коэффициент дифференциального регулирования, U_{max} – максимальное напряжение на выходе регулятора. Параметром состояния является напряжение на выходе регулятора (переменные U , Uf).

Функция `next` построена на основе уравнения пропорционально-интегрально-дифференциального регулирования. Напряжение U , выдаваемое регулятором, определяется значением и поведением во времени температуры рабочего тела T и температуры задатчика Z :

$$U = K_{pr}(Z-T) + K_{in} \int_0^t (Z-T)dt + K_{di} \frac{d}{dt}(Z-T), \quad (3)$$

Класс Inter (элемент теплообмена)

Основным назначением субмодели вида «элемент теплообмена» является описание механизма передачи тепла от одного элемента системы к другому. Хотя такая субмодель может учитывать и другие сущности. В нашем примере элементы теплообмена являются безинерционными и, таким образом, представляют механизм теплопередачи в «чистом виде».

При построении данного класса используются возможности параметризованных классов (шаблонов). Класс `Inter` является шаблоном с параметрами `type1` и `type2`, с помощью которых передаются типы воздействующих объектов – элементов системы.

```
template <class type1, class type2>
class Inter
{ public:
double Ld;
double Q, Qf;
Inter(double ld) { Ld=ld; Q=0; }
void next(type1, type2);
void step() { Q=Qf; }
};
template <class type1, class type2>
void Inter<type1, type2>::next(type1 A, type2 B)
{ Qf=Ld*(A.T-B.T);
}
```


Объект класса Inter содержит один параметр (L_d – коэффициент теплопроводности соответствующего материала) и один параметр состояния (Q , Q_f – передаваемые тепловые потоки). Функция `next` вычисляет тепловой поток в соответствии с уравнением:

$$Q = \lambda(T_a - T_b), \quad (4)$$

где T_a , T_b – температуры источника и приемника тепла, λ – коэффициент теплообмена.

Класс Targ (задатчик температуры)

Задатчик температуры является целеустанавливающим элементом системы. Он представляет закон изменения температуры со временем или просто заданную температуру. Состояние задатчика в нашем примере не зависит от состояний других элементов системы.

```
class Targ
{public:
double T, Tf;
Targ() { T=Tf=Tmin; }
Targ(double A) { T=A; }
void next();
void step() { T=Tf; }
};
void Targ::next()
{const float A=300;
if (t<tm/2) Tf=t/tm*A; else Tf=A*(1.0-t/tm);
}
```

Параметров данная субмодель в нашем примере не имеет. Вместо этого, параметры задатчика могут быть инкапсулированы в функцию – член класса `next`. В нашем примере мы имеем один такой параметр. Параметром состояния задатчика является выдаваемая им температура (переменные T , T_f).

Функция `next` содержит принятый закон изменения температуры: линейный рост (первая фаза) и линейное снижение (вторая фаза) температуры рабочего тела:

$$T = \begin{cases} T_f = \frac{t}{t_m} A, & \text{если } t < \frac{1}{2} t_m, \\ T_f = A(1 - \frac{t}{t_m}), & \text{если } t \geq \frac{1}{2} t_m. \end{cases} \quad (5)$$

Параметры A – максимальная температура процесса, t_m – общее время моделирования системы удобно представить как общие параметры моделирования, присутствующие в программе в виде глобальных переменных.

Класс Medium (внешняя среда)

Данный класс представляет субмодель внешней среды.

```
class Medium
{public:
double A1, A2;
double T, Tf;
Medium(double a1, double a2) { A1=a1; A2=a2; T=A1; }
void next();
void step() { T=Tf; }
};
```

```
void Medium::next()
{if (t<tm/2) Tf=A1; else Tf=A2;
}
```

Субмодель имеет два параметра ($A1$ и $A2$) и одну переменную состояния – температура внешней среды (T , T_f).

Функция `next` содержит информацию об изменении состояния внешней среды: ее температура в момент времени $t = t_m/2$ скачкообразно изменяется от значения $A1$ до $A2$.

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Результаты моделирования приведены на рис. 2 и 3. На рисунках приведены следующие кривые: температура задатчика, напряжение на выходе регулятора, температура нагревателя, температура рабочего тела, все величины как функции времени. Результаты моделирования на рис. 2 соответствуют следующим условиям: температура внешней среды постоянна, задатчик формирует сигнал в форме П-импульса. Параметры модели:

1. Задатчик: $T_{min}=10$, $T_{max}=100$.
2. Внешняя среда: $T_{medi}=10$.
3. Элементы теплообмена: $L_{bm}=10$, $L_{hm}=0.1$, $L_{hb}=15$.
4. Рабочее тело: $C_{body}=20$.
5. Нагреватель: $R_{heat}=100$, $C_{heat}=2$.
6. Регулятор: $U_{max}=400$, $K_{pr}=20$, $K_{in}=3$, $K_{di}=0.1$.

Рис. 3 соответствует таким условиям: температура внешней среды постоянна, задатчик формирует сигнал в форме треугольного импульса. Параметры модели:

1. Задатчик: $T_{min}=10$, $T_{max}=100$.
2. Внешняя среда: $T_{medi}=0$.
3. Элементы теплообмена: $L_{bm}=10$, $L_{hm}=0.1$, $L_{hb}=15$.
4. Рабочее тело: $C_{body}=20$.
5. Нагреватель: $R_{heat}=100$, $C_{heat}=2$.
6. Регулятор: $U_{max}=400$, $K_{pr}=20$, $K_{in}=12$, $K_{di}=0$.

ВЫВОДЫ

На примере конкретной системы показана результативность объектно-ориентированного подхода для построения компьютерной модели, позволяющей отслеживать фазовую траекторию системы. Показаны особенности построения программной реализации модели на основе объектно-ориентированного подхода. Основными преимуществами такого подхода являются следующие.

1. Модульность, позволяющая с минимальными затратами вносить функциональные и структурные изменения в модель анализируемой системы.
2. Возможность формализации процесса построения программной реализации динамической модели системы с локализованными свойствами различной этимологии и природы.
3. Потенциальная возможность построения моделей систем высокого уровня сложности, включающих большое число элементов и связей.

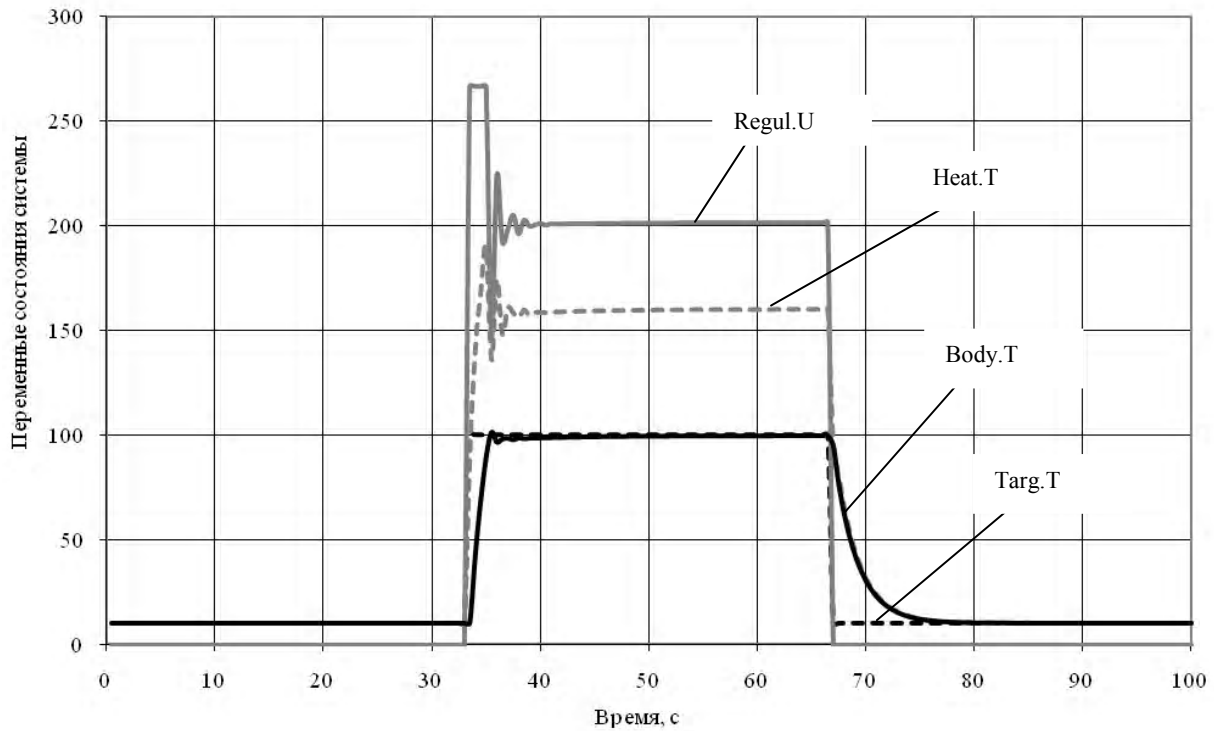


Рис. 2. Результат моделирования при ступенчатом изменении задаваемой температуры

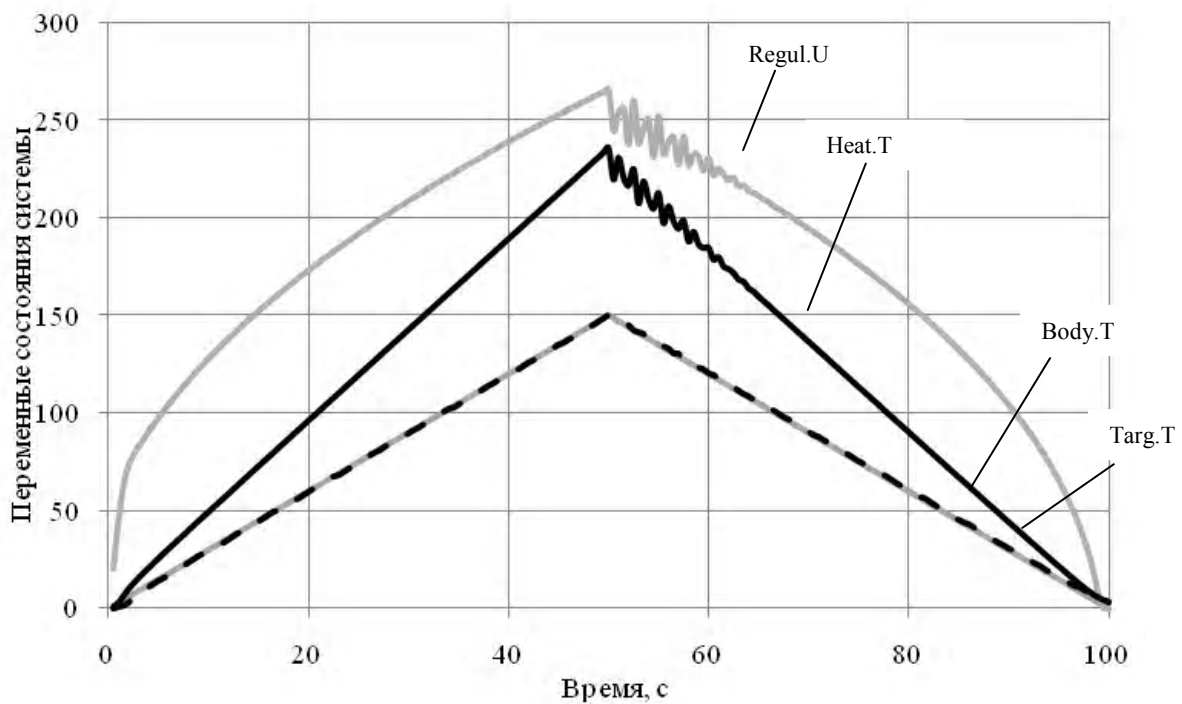


Рис. 3. Результат моделирования для функции задатчика треугольной формы

СПИСОК ЛИТЕРАТУРЫ

1. Колесов, Ю. Б. Объектно-ориентированное моделирование сложных динамических систем / Ю. Б. Колесов. – СПб. : Изд-во С.Пб. ГПУ, 2004. – 239 с.

2. Колесов, Ю. Б. Моделирование систем. Динамические и гибридные системы / Ю. Б. Колесов, Ю. Б. Сениченков. – С. Пб. : БХВ-Петербург, 2006. – 224 с.

Стаття надійшла до редакції 15.03.2011.

Пінчук В. П., Подковаліхіна О. О.

ОБ'ЄКТНО-ОРІЄНТОВАНИЙ ПІДХІД ДЛЯ ПОБУДОВИ МОДЕЛЕЙ СКЛАДНИХ СИСТЕМ

Запропоновано методика побудови динамічної моделі системи з локалізованими властивостями на основі об'єктно-орієнтованого підходу. Розглянуто приклад системи автоматичного регулювання температури робочого тіла з пропорційно-інтегрально-диференціальним законом регулювання, наведено методику побудови класів і функцій, що складають програмну реалізацію моделі, обговорюються можливості запропонованої методики.

Ключові слова: комп'ютерне моделювання систем, моделювання динаміки системи, об'єктно-орієнтований підхід, система терморегулювання.

Pinchuk V. P., Podkovichina E. A.

OBJECT-ORIENTED APPROACH FOR MODELS DESIGNING OF THE COMPLEX SYSTEMS

A method for constructing a dynamic model of the system with localized features based on object-oriented approach was proposed. An example of a system for automatic temperature control of the working body with a proportional-integral-differential law of regulation was considered.

A technique for constructing classes and their functions that comprise the software implementation of the model was led, the opportunities of the proposed method were discussed.

Key words: computer simulation of the systems, modeling the dynamics of the system, object-oriented approach, the thermal control system.

НЕЙРОИНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

НЕЙРОИНФОРМАТИКА И ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

NEUROINFORMATICS AND INTELLIGENT SYSTEMS

УДК 519.168:004.658

Асеев Г. Г.

*Д-р техн. наук, профессор, заведующий кафедрой Харьковской
государственной академии культуры*

МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В ЭЛЕКТРОННЫХ ХРАНИЛИЩАХ: ГЕНЕТИЧЕСКИЕ АЛГОРИТМЫ

Представлен один из возможных методов интеллектуального анализа данных в электронных хранилищах большого объема – генетические алгоритмы и их модификация.

Ключевые слова: интеллектуальный анализ, электронное хранилище, нейронная сеть, генетические алгоритмы.

ВВЕДЕНИЕ

В настоящее время в электронных хранилищах данных (ХД) корпоративных информационных систем хранятся терабайты различной текстовой и числовой информации. Для обнаружения, извлечения и интеллектуального анализа этих данных используются методы Knowledge Discovery in Databases и Data mining [1]. В [1] были описаны некоторые рекомендации, следуя которым можно подготовить качественные данные в нужном объеме для анализа: первичные источники данных, хранение данных, подготовка исходного набора данных, предобработка и очистка исходных данных [2], трансформация, нормализация, выдвижение гипотез и построение модели Data Mining [3]. Данная работа продолжает цикл статей, посвященных методам интеллектуального анализа данных в электронных хранилищах большого объема, в частности модификации генетических алгоритмов.

ГЕНЕТИЧЕСКИЕ АЛГОРИТМЫ – МАТЕМАТИЧЕСКИЙ АППАРАТ

Такие свойства генетических алгоритмов, как адаптивность, робастность, возможность распараллеливания вычислений и отыскание глобального экстремума принятой функции приспособленности, обеспечили их эф-

фективное использование для решения различных задач в пространствах высокой размерности в ХД. Примером подобной задачи может служить обучение нейросети, то есть подбора таких значений весов, при которых достигается минимальная ошибка.

Из биологии мы знаем, что любой организм может быть представлен своим *фенотипом*, который фактически определяет, чем является объект в реальном мире, и *генотипом*, который содержит всю информацию об объекте на уровне хромосомного набора. При этом каждый ген, то есть элемент информации генотипа, имеет свое отражение в фенотипе. Разработчик генетических алгоритмов выступает в данном случае как «создатель», который должен правильно установить законы эволюции, чтобы достичь желаемой цели как можно быстрее. Впервые эти нестандартные идеи были применены к решению оптимизационных задач в середине 70-х годов [4]. Примерно через десять лет появились первые теоретические обоснования этого подхода [5, 6]. В дальнейшем генетические алгоритмы доказали свою конкурентоспособность при решении многих *NP*-трудных задач [7] и особенно в практических приложениях, где математические модели имеют сложную структуру и применение стандартных методов типа ветвей и границ, динамического или линейного программирования крайне затруднено.

В наиболее часто встречающейся разновидности генетического алгоритма для представления генотипа объекта применяются битовые строки. При этом каждому атрибуту объекта в фенотипе соответствует один ген в генотипе объекта. Ген представляет собой битовую строку, чаще всего фиксированной длины, которая представляет собой значение этого признака.

Генетический алгоритм работает с представленными в конечном алфавите строками S конечной длины l , которые используются для кодировки исходного множества альтернатив W . Строки представляют собой упорядоченные наборы из l элементов: $S = (s_1, s_2, \dots, s_l)$, каждый из которых может быть задан в своем собственном алфавите V_i , $i = \overline{1, L}$, где алфавит V_i является множеством из r_i символов: $V_i = \{v_{ij}, j = \overline{1, r_i}\}$. Для решения конкретной задачи требуется однозначно отобразить конечное множество альтернатив W на множество строк подходящей длины (очевидно, что длина строк зависит от алфавитов, используемых для их задания).

Для работы алгоритма необходимо на множестве строк $U^m(V_1, V_2, \dots, V_m)$ задать неотрицательную функцию $F(S)$, определяющую показатель качества, «ценность» строки $SO U^m(V_1, V_2, \dots, V_m)$. Алгоритм производит поиск строки, для которой

$$F^*(S) = \arg \max_{S \in U^m(V_1, V_2, \dots, V_m)} F(S).$$

Если на множестве W задана целевая функция $f(w)$, то функцию $F(S)$ на множестве строк $U^m(V_1, V_2, \dots, V_m)$ можем определить следующим образом: $F(S) = f(w)$, если элемент w при отображении исходного множества W на множество строк был сопоставлен строке S .

Генетический алгоритм за один шаг производит обработку некоторой популяции строк. Популяция $G(t)$ на шаге t представляет собой конечный набор строк:

$$G(t) = (S_1^t, S_2^t, \dots, S_N^t), S_k^t \in U^m(V_1, V_2, \dots, V_m), k = \overline{1, N},$$

где N – размер популяции, причем строки в популяции могут повторяться.

Анализ работы алгоритма удобно производить, используя аппарат схем. Схемой в генетическом алгоритме называют описание некоторого подмножества строк. Схема $H = (h_1, h_2, \dots, h_m)$ может рассматриваться как строка, алфавиты для элементов которой дополнены специальным символом «#»:

$$H \in U^m(V_1^H, V_2^H, \dots, V_m^H), V_i^H = V_i \cup \{ \# \}.$$

Если в некоторой позиции r схемы H присутствует символ «#», то такая позиция называется свободной, а сам символ «#» интерпретируется как произвольный символ из алфавита V_r . Позиция q схемы H называется

фиксированной, если в этой позиции присутствует один из символов алфавита V_q . Схема H , в которой определены фиксированные и свободные позиции, описывает подмножество $U_H \subseteq U^m(V_1, V_2, \dots, V_m)$, содержащее такие строки, у которых элементы, соответствующие фиксированным позициям схемы, совпадают с символами схемы, а элементы, соответствующие свободным позициям схемы, являются произвольно заданными в соответствующих алфавитах:

$$U_H = \left\{ S \mid S \in U^m(V_1, V_2, \dots, V_m) \wedge (\forall i (i \in I_{[1, m]} \wedge h_i \neq \#) \rightarrow (s_i = h_i)) \right\},$$

где $I_{[1, m]}$ – множество целых чисел отрезка $[1, m]$.

Например, для множества строк $U_{(V_1, V_2, V_3, V_4, V_5)}^5$, где $U_i = \{0, 1\}$, $V_i = \overline{1, 5}$, схема $H_1 = \langle 1 \# \# \# 0 \rangle$ задает такое множество строк, у которых первым элементом является символ «1», пятым – «0», а остальные – либо «0», либо «1». Строки «10010», «11110» являются примерами строк, принадлежащих множеству U_{H_1} .

Часть популяции $G(t) = (S_1^t, S_2^t, \dots, S_N^t)$, строки которой удовлетворяют схеме H , обозначают $G_H(t) = (S_1^{H, t}, S_2^{H, t}, \dots, S_{n(H, t)}^{H, t})$, где $n(H, t)$ – число строк схемы H в популяции $G(t)$, и называют подпопуляцией, соответствующей схеме H .

ПРОЦЕДУРА ОПТИМИЗАЦИИ

В общем случае процедура оптимизации на основе обычного последовательного комплекс-метода выглядит следующим образом: требуется отыскать минимум некоторой функции, как правило, многоэкстремальной:

$$E(x) \rightarrow \min_{x \in R^n},$$

достаточно общего вида, при этом, о характере этой функции не делается практически никаких априорных предположений.

Будем использовать функции приспособленности следующего вида:

$$E(x) = \sum_{cl} E(cl) = \sum_{cl} \sum_d w_{cl, d} = \sum_{cl} \sum_d sf_{cl, d} \times \times ids_{cl, d} = \sum_{cl} \sum_d sf_{cl, d} \times \log \frac{N}{ds_{cl}},$$

где $E(cl)$ – функция приспособленности для хромосомы cl ; $w_{cl, d}$ – нормализованные данные о хромосоме c для документа d ; $sf_{cl, d}$ – частота встречаемости терма (или набора термов), представленного хромосомой cl ; $ids_{cl, d}$ – инверсная частота встречаемости терма (или

набора термов), представленного хромосомой cl ; ds_{cl} – число документов, содержащих комбинации хромосомы cl ; N – общее число документов в ХД.

Все данные должны быть представлены в двоичном коде (1 – если терм (или набор термов) содержится в документе, 0 – в противном случае). В ХД используются хромосомы, максимальная длина которых составляет до нескольких сотен генов, причем некоторые из них могут быть пустыми. Согласно этому утверждению, число повторяющихся термов в решении может варьироваться от 2 до нескольких сотен. Поскольку пространство решений очень большое, предлагается использовать мутацию, фиксированную между 50 и 70 процентами, и в конечном итоге каждая хромосома будет подвержена мутации для новой популяции.

Работа алгоритма начинается с формирования начального комплекса

$$x_i(0) = (x_{i1}(0), x_{i2}(0), \dots, x_{ij}(0), \dots, x_{in}(0))^T, \quad i = 1, 2, \dots, N \geq n+1,$$

представляющего собой «облако» (популяцию) точек (векторов), достаточно произвольно расположенных в n -мерном пространстве факторов. Среди множества этих точек находится «наихудшая» $x_N(0)$, в которой значение функции $E(x_N(0))$ максимально, после чего эта точка отражается через центр тяжести всех остальных вершин-точек, формируя новый комплекс $x_N(1)$, $i = 1, 2, \dots, N$. Такое отражение вместе с растяжением и сжатием обеспечивают движение комплекса к экстремуму функции $E(x)$, при этом, благодаря достаточно случайному распределению точек «облака», поиск имеет глобальный характер.

С формальной точки зрения, рассмотрим процесс оптимизации на k -й итерации поиска, когда сформирован комплекс $x_i(k)$, $i = 1, 2, \dots, N$. Среди множества точек $x_i(k)$ находится «наихудшая», такая, что

$$E(x_N(k)) = \max_i \{E(x_1(k)), \dots, E(x_N(k))\},$$

после чего определяется центр тяжести «облака» без наихудшей точки:

$$x_C(k) = \frac{1}{N-1} \left(\sum_{i=1}^N x_i(k) - x_N(k) \right).$$

Далее $x_N(k)$ отражается через центр тяжести $x_C(k)$, формируя новую вершину комплекса $x_R(k)$, которая теоретически расположена ближе к экстремуму, чем $x_N(k)$ и $x_C(k)$, т. е.

$$E(x_R(k)) < E(x_C(k)) < E(x_N(k)).$$

Операция отражения формально имеет следующий вид:

$$x_R(k) = x_C(k) + \eta_R(x_C(k) - x_N(k)) =$$

$$= \frac{1}{N-1} x_1(k) + \dots + \frac{1}{N-1} x_{N-1}(k) + \frac{\eta_R}{N-1} x_1(k) + \dots \\ \dots + \frac{\eta_R}{N-1} x_{N-1}(k) - \eta_R x_N(k) = X(k)R,$$

где η_R – параметр шага отражения, часто полагаемый равным единице, $X(k) = (x_1(k), x_2(k), \dots, x_{N-1}(k))$ – $(n \times N)$ – матрица координат вершин комплекса,

$$R = \left(-\eta_R, \frac{1+\eta_R}{N-1}, \dots, \frac{1+\eta_R}{N-1} \right)^T - (N \times 1) - \text{вектор}.$$

В случае, если отраженная вершина $x_R(k)$ окажется «наилучшей» среди всех остальных точек комплекса, т. е.:

$$E(x_R(k)) < E(x_C(k)) < E(x_N(k)), \quad i = 1, 2, \dots, N-1,$$

производится операция растяжения комплекса в направлении от центра тяжести $x_C(k)$ до $x_R(k)$ согласно выражению

$$x_E(k) = x_C(k) + \eta_E(x_R(k) - x_C(k)) = X(k)E,$$

где η_E – параметр шага растяжения, часто полагаемый равным двум:

$$E = \left(-\eta_E \eta_R, \frac{1-\eta_E(1-\eta_R)}{N-1}, \dots, \frac{1-\eta_E(1-\eta_R)}{N-1} \right)^T.$$

Если же $x_R(k)$ окажется наихудшей среди всех $x_i(k)$, комплекс сжимается согласно соотношению

$$x_S(k) = x_C(k) + \eta_S(x_R(k) - x_C(k)) = X(k)S,$$

где η_S – параметр шага сжатия, обычно полагаемый равным 0,5:

$$S = \left(-\eta_S \eta_R, \frac{1-\eta_S(1-\eta_R)}{N-1}, \dots, \frac{1-\eta_S(1-\eta_R)}{N-1} \right)^T.$$

При $\eta_S = 1$, $\eta_E = 2$, $\eta_R = 0,5$ приходим к простым выражениям:

$$R = \left(-1, \frac{2}{N-1}, \dots, \frac{2}{N-1} \right)^T, \quad E = \left(-2, \frac{1}{N-1}, \dots, \frac{1}{N-1} \right)^T,$$

$$S = \left(-0,5, \frac{1}{N-1}, \dots, \frac{1}{N-1} \right)^T.$$

Таким образом, в процессе своего движения к экстремуму оптимизируемой функции комплекс на каждой итерации теряет одну наихудшую вершину и приобретает одну новую точку так, что на $(k+1)$ -й итерации новый комплекс также имеет N точек-вершин.

В генетических алгоритмах в результате селекции из популяции одновременно исключаются несколько особей с наихудшими (максимальными) значениями функции приспособленности. В связи с этим представляется целесообразным ввести алгоритм комплекс-метода с отражением, растяжением и сжатием сразу нескольких вершин [8, 9].

Итак, пусть на k -й итерации процесса оптимизации имеется комплекс $x_i(k)$, $i = 1, 2, \dots, N$ с $P < N$ наихудшими вершинами $x_{H_p}(k)$, $p = 1, 2, \dots, P$. Тогда координаты центра тяжести комплекса без вершин $x_{H_p}(k)$ задаются выражением

$$x_C(k) = \frac{1}{N-P} \left(\sum_{i=1}^N x_i(k) - \sum_{p=1}^P x_{H_p}(k) \right),$$

а процедура отражения описывается системой уравнений

$$\begin{cases} x_{R_1}(k) = x_C(k) + \eta_R(x_C(k) - x_{H_1}(k)), \\ \vdots \\ x_{R_P}(k) = x_C(k) + \eta_R(x_C(k) - x_{H_P}(k)). \end{cases}$$

В случае, если среди отраженных вершин оказывается $Q \leq P$ наилучших, комплекс растягивается в их направлении согласно уравнениям

$$\begin{cases} x_{E_1}(k) = x_C(k) + \eta_E(x_{R_1}(k) - x_C(k)), \\ \vdots \\ x_{E_Q}(k) = x_C(k) + \eta_E(x_{R_Q}(k) - x_C(k)). \end{cases}$$

Если, далее, среди отражаемых вершин окажется $U \leq P$ наихудших, комплекс сжимается в их направлении согласно уравнениям

$$\begin{cases} x_{S_1}(k) = x_C(k) + \eta_S(x_{R_1}(k) - x_C(k)), \\ \vdots \\ x_{S_U}(k) = x_C(k) + \eta_S(x_{R_U}(k) - x_C(k)). \end{cases}$$

Таким образом, комплекс-метод приобретает черты генетического алгоритма, у которого в результате селекции на каждой итерации из популяции удаляется несколько наихудших особей.

Объединяя введенную модификацию комплекс-метода с холландовской генетической процедурой, приходим к алгоритму, реализующему идею искусственного отбора, состоящую в данном случае в том, что из популяции не только удаляются наихудшие особи, но и одновременно создаются их «антиподы», обладающие улучшенными свойствами.

Работа такого алгоритма образована последовательностью следующих шагов:

- создание начальной популяции, образованной $P(0)$ особями хромосомами – вершинами комплекса;
- операция кроссовера с увеличением популяции $P_{CR}(0) > P(0)$;
- операция мутации $P_M(0) > P_{CR}(0)$;
- операция инверсии $P_I(0) > P_M(0)$;
- первая селекция (определение наихудших особей) без сокращения популяции $P_{SEL1}(0) = P_I(0)$;
- операция отражения с удалением P наихудших особей $P_R(0) < P_{SEL1}(0)$;
- операция растяжения без увеличения популяции $P_E(0) = P_R(0)$;
- операция сжатия без увеличения популяции $P_I(0) = P_E(0)$;
- вторая селекция с удалением $P_W(0)$ наихудших особей $P_{SEL2}(0) = P_I(0) = P(1)$ и формирование популяции $P(1)$ для следующей итерации алгоритма.

ВЫВОДЫ

Описанный в разделе математический аппарат холландовских генетических алгоритмов имеет ряд недостатков. В частности, они характеризуются низкой скоростью сходимости, не позволяющей им отыскивать решение за приемлемое время. Также генетические алгоритмы являются чувствительными к выбору параметров алгоритма, например, размера популяции, вероятностей кроссовера и мутации и т. п.

Эти и некоторые другие особенности генетических алгоритмов послужили толчком к созданию их различных модификаций. В некоторых модификациях, например, предлагается использовать, кроме классических генетических операторов кроссовера, мутации и инверсии дополнительные операторы. Например, такие как операторы объединения (fusion) и разделения (fission). Операция объединения заключается в том, что два аллеля соединяются в один. Операция разделения предполагает замену одного аллеля другим случайным аллелем. В результате происходит разделение кластеров [10].

В основе рассматриваемого алгоритма лежит синтез обычного эволюционного генетического подхода с идеями адаптивной оптимизации и, прежде всего, последовательного комплекс-метода отыскания экстремума функций многих переменных. При этом в каждый момент времени текущая популяция отождествляется с «облаком» – комплексом точек в пространстве переменных-факторов, а кроме традиционных генетических операторов мутации, кроссовера и инверсии дополнительно вводятся операторы комплекс-поиска, такие как отражение, растяжение и сжатие. Работа предложенного алгоритма протестирована на выборке Reuters-21578 [8, 9]. Было установлено, что предложенный алгоритм работает быстрее и дает более точные результаты (в среднем 8–10 %) по сравнению со стандартными генетическими алгоритмами.

СПИСОК ЛІТЕРАТУРИ

1. Асеев, Г. Г. Проблема обнаружения нового знания в хранилищах данных методами Knowledge Discovery in Databases / Г. Г. Асеев // Вестник НТУ «ХПИ». – 2006. – № 19. – С. 62–70.
2. Асеев, Г. Г. Методы интеллектуальной предобработки данных в электронных хранилищах / Г. Г. Асеев // Радіоелектроніка, інформатика, управління. – 2010. – № 2(23). – С. 106–111.
3. Асеев, Г. Г. Методы интеллектуального анализа данных в электронных хранилищах / Г. Г. Асеев // Бионика интеллекта : науч.-техн. журнал. – 2008. – № 1(70). – С. 28–33.
4. Растрин, Л. А. Случайный поиск – специфика, этапы истории и предрасудки / Л. А. Растрин // Вопросы кибернетики. – Вып. 33. – 1988. – С. 3–12.
5. Holland, J. H. Adaptation in natural and artificial systems / John H. Holland. – Ann Arbor : University of Michigan Press, 1985. – 305 p.
6. Rechenberg, I. Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der Biologischen Information / Rechenberg I. – Freiburg : Fromman, 1983. – P. 135–143.
7. Goldberg, D. E. Genetic algorithms in search, optimization, and machine learning / David E. Goldberg. – [USA] : Addison-Wesley, 1989. – 752 p.
8. Волкова, В. В. Возможностная фаззи-кластеризация текстовых массивов в реальном времени на основе самообучающейся нейронной сети / В. В. Волкова, Б. В. Колчигин // Факультетская научно-практическая молодежная школа-семинар студентов, аспирантов и молодых ученых «Информационные интеллектуальные системы» : тезисы докл. – Харьков : ХНУРЭ, 2008. – С. 30–33.
9. Волкова, В. В. Комбинированное обучение самоорганизующихся карт с нечетким выводом / В. В. Волкова, Е. В. Махиборода // Факультетская научно-практическая молодежная школа-семинар студентов, аспирантов и молодых ученых «Информационные интеллектуальные системы» : тезисы докл. – Харьков : ХНУРЭ, 2008. – С. 30–33.
10. Рассел, С. Искусственный интеллект. Современный подход / С. Рассел, П. Норвиг. – М. : Вильямс, 2006. – 1408 с.

Стаття надійшла до редакції 21.01.2011.

Асеев Г. Г.

МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ В ЕЛЕКТРОННИХ СХОВИЩАХ: ГЕНЕТИЧНІ АЛГОРИТМИ

Представлено один з можливих методів інтелектуального аналізу даних в електронних сховищах великого об'єму – генетичні алгоритми і їх модифікація.

Ключові слова: інтелектуальний аналіз, електронне сховище, нейронна мережа, генетичні алгоритми.

Aseyev G. G.

METHODS OF INTELLECTUAL ANALYSIS OF DATA IN ELECTRONIC DEPOSITORIES: GENETIC ALGORITHMS

One of the possible methods of data intellectual analysis in high-volume electronic depositories is presented – genetic algorithms and their modification.

Key words: intellectual analysis, electronic depository, neuron network, genetic algorithms.

УДК 004:519.2

Кротких С. С.¹, Кириченко Л. О.²

¹Аспирант Харьковського національного університету радіоелектроніки

²Канд. техн. наук, доцент Харьковського національного університету радіоелектроніки

ИССЛЕДОВАНИЕ ВЫЗВАННЫХ ПОТЕНЦИАЛОВ В ЭЭГ ЧЕЛОВЕКА С ПОМОЩЬЮ ДИСКРЕТНОГО ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ

В работе с помощью дискретного вейвлет-преобразования проведен анализ изменения частотной структуры электроэнцефалограммы с вызванными потенциалами после воздействия стимула. Реализован метод определения времени реакции на стимул, основанный на оценивании вейвлет-энтропии и относительной вейвлет-энтропии сегментов энцефалограммы.

Ключевые слова: электроэнцефалограмма, вызванные потенциалы, дискретное вейвлет-преобразование, спектр вейвлет-энергии, вейвлет-энтропия, относительная вейвлет-энтропия.

ВВЕДЕНИЕ

Головной мозг человека обладает электрическим полем, которое характеризуется электрическими скачками от нескольких миллиампер до нескольких сотен миллиампер. Каждый нейрон мозга генерирует изменение электрического потенциала, этот процесс можно измерить. Для анализа функций головного мозга широко используется электрический энцефалограф, который поз-

воляет измерить изменение электрического потенциала в отдельных участках головного мозга человека. Электроэнцефалограмму (ЭЭГ) можно рассматривать как фиксацию последовательности умственных задач, выполняемых мозгом субъекта. Различные умственные задачи или условия, в которых находится человек, имеют свои уникальные отражения в результирующем ЭЭГ сигнале. Таким образом, наличие тех или иных изменений в

сигнале может свидетельствовать об изменении условий или внутреннего физиологического состояния человека [1, 2].

ЭЭГ измеряет электрический потенциал с помощью нескольких электродов, расположенных в определенных точках на поверхности головы субъекта. Количество электродов, в зависимости от модели энцефалографа, может быть от 8 до 64. Если субъект жив, электрическая мозговая активность присутствует постоянно. За исключением воздействий мускульной активности и моргания глаз, сигнал распределен на частотах от 0,5 Гц до 40 Гц. Этот промежуток принято разделять на частотные диапазоны, называемые ритмами. Под понятием «ритм» на ЭЭГ подразумевается определенная полоса частот, соответствующая некоторому состоянию мозга. К ритмам ЭЭГ взрослого бодрствующего человека относятся альфа-, бета-, гамма-, дельта-, тета- ритмы (табл. 1). Присутствие каждого из диапазонов зависит от условий и состояния субъекта [2].

Одним из методов исследования активности головного мозга является использование вызванных потенциалов, связанных с событиями (ВП), которые представляют реакцию мозга на внешние события (стимулы). В англоязычной литературе распространен термин ERP (event-related potentials). В зависимости от стимула существуют зрительные, слуховые, соматосенсорные ВП, возможно получение ВП при раздражении любого периферического нерва.

Традиционная методика регистрации ВП заключается в том, что в течение определенного времени после подачи стимула производится вычисление амплитуд ЭЭГ через временные интервалы, зависящие от частоты квантования. Полученные данные суммируются, в результате чего амплитуда стабильно возникающих после стимула потенциалов неуклонно возрастает, а амплитуда ритмов спонтанной ЭЭГ в той же степени уменьшается. Для получения истинных амплитуд ВП амплитуда в каждой точке делится на число стимулов.

По скорости реакции на стимул (латентности) ВП делятся на компоненты, вызывающие быструю реакцию в стволе головного мозга (реакция возникает в течении 0–12 мс с момента стимуляции), компоненты средней задержки (12–50 мс после стимуляции) и компоненты с долгой задержкой или когнитивные компоненты ВП (реакция возникает начиная от 50 мс после стимуляции и обычно до 500 мс). ВП с долгой задержкой связаны с когнитивными процессами, такими как обращение к памяти, эмоции или выполнение задач на внимание. Изменение амплитуды и латентности определенных компонентов позволяет

диагностировать различные заболевания в ранней стадии развития, например, поражения проводящих путей или нервных центров, предынсультные состояния, психические заболевания, эпилепсию и др. [3–5].

ПОСТАНОВКА ЗАДАЧИ

Основным инструментом анализа частотных характеристик сигнала является преобразование Фурье. Показателем меры спектральной сложности сигнала является спектральная энтропия. Вся энергия упорядоченного сигнала, такого как синусоида, сконцентрирована в соответствующих гармониках, что говорит о низкой энтропии сигнала. С другой стороны, зашумленный сигнал содержит широкий диапазон частот, следовательно, обладает высокой энтропией. Однако применение спектрального анализа к исследованию ЭЭГ весьма ограничено, поскольку участки ВП являются весьма кратковременными и сильно нестационарными. Этот недостаток может быть частично решен использованием оконного преобразования Фурье. Но при узком окне частотное разрешение будет слишком мало, а при широком – временная локализация становится неточной. Это ограничение становится критическим, когда сигнал содержит кратковременные изменения частот, такие как ВП. Преодолеть это ограничение можно, используя частотно-временное представление сигнала, как это предусматривается вейвлет-преобразованием, не требующим допущений о стационарности. Вейвлет-анализ вызванных потенциалов в ЭЭГ человека и использование вейвлет-энтропии как меры упорядоченности участков ЭЭГ-сигнала было рассмотрено в работах [5–9].

Целью представленной работы является анализ изменения частотной структуры ЭЭГ после воздействия стимула с помощью дискретного вейвлет-преобразования и определение времени реакции на внешние воздействия, основанное на применении вейвлет-энтропии.

КРАТКИЕ СВЕДЕНИЯ ИЗ ВЕЙВЛЕТ-АНАЛИЗА

Рассмотрим основные методы исследования стохастических временных рядов, основанные на применении вейвлет-анализа [10–13]. Основная идея кратномасштабного вейвлет-анализа заключается в том, что разложение ряда производится по ортогональному базису, образованному сдвигами и кратномасштабными копиями вейвлетной функции. Базисные функции $\psi(t)$ называются вейвлетами, если они удовлетворяют ряду условий, в частности, определены на пространстве комплекснозначных функций с ограниченной энергией, колеблются вокруг оси абсцисс, быстро сходятся к нулю и имеют нулевой момент первого порядка.

Произвольная функция $s(t)$ может быть разложена по вейвлет-базису $\psi_{m,k}(t) = a^{\frac{m}{2}} \psi(a^m t + k)$ с параметрами масштаба a и сдвига k :

$$s(t) = \sum_{m,k=-\infty}^{\infty} S_{m,k} \psi_{m,k}(t), \quad (1)$$

Таблица 1. Основные ритмы ЭЭГ человека

Ритмы	Частотный диапазон
Гамма (μ)	40–70 Гц
Бета (β)	14–30 Гц
Альфа (α)	8–13 Гц
Тета (θ)	4–7 Гц
Дельта (δ)	0,5–3 Гц

где коэффициенты вейвлет-спектра $S_{m,k}$ определяются скалярным произведением

$$S_{m,k} = \langle s(t), \psi_{m,k}(t) \rangle = \int_{-\infty}^{\infty} s(t) \psi_{m,k}(t) dt. \quad (2)$$

В дискретном вейвлет-преобразовании (ДВП) параметры масштаба и сдвига задаются обычно в виде степенных функций: $a = 2^{-m}$, $k = k_0 \cdot 2^{-m}$, $m, k_0 \in \mathbb{Z}$. Дискретные вейвлеты используются, как правило, в паре со связанными с ними дискретными скейлинг-функциями. Скейлинг-функции имеют с вейвлетами общую область задания и определенное соотношение между значениями. При заданных материнском вейвлете ψ и соответствующей скейлинг-функции ϕ аппроксимирующие коэффициенты $\text{apr}(j, k)$ и детализирующие коэффициенты $\text{det}(j, k)$ ДВП для процесса $X(t)$ определяются следующим образом:

$$\begin{aligned} \text{apr}(j, k) &= \int_{-\infty}^{\infty} X(t) \phi_{j,k}(t) dt, \\ \text{det}(j, k) &= \int_{-\infty}^{\infty} X(t) \psi_{j,k}(t) dt, \end{aligned} \quad (3)$$

где j – параметр масштаба, k – параметр сдвига ($j, k \in \mathbb{Z}$) и

$$\begin{aligned} \phi_{j,k}(t) &= 2^{-j/2} \phi(2^{-j}t - k), \\ \psi_{j,k}(t) &= 2^{-j/2} \psi(2^{-j}t - k). \end{aligned} \quad (4)$$

Кратномасштабный анализ, выполняемый с помощью ДВП, заключается в разбиении исследуемого временного ряда на две составляющие: аппроксимирующую и детализирующую, с последующим аналогичным дроблением аппроксимирующей до заданного уровня декомпозиции сигнала. Временной ряд $X(t)$ представляется в виде суммы аппроксимирующей компоненты $\text{approx}_N(t)$ и детализирующих компонент $\text{detail}_j(t)$:

$$\begin{aligned} X(t) &= \text{approx}_N(t) + \sum_{j=1}^N \text{detail}_j(t) = \\ &= \sum_{k=1}^{N_a} \text{apr}(N, k) \phi_{j,k}(t) + \sum_{j=1}^N \sum_{k=1}^{N_j} \text{det}(j, k) \psi_{j,k}(t), \end{aligned} \quad (5)$$

где N – выбранный максимальный уровень разложения ряда, N_j – количество детализирующих коэффициентов на уровне j , N_a – количество аппроксимирующих коэффициентов на уровне N .

Выбор типа вейвлет-функции и количества уровней разложения является важным вопросом при выполнении

ДВП. Обычно вейвлет-функция подбирается в зависимости от временных и частотных характеристик каждого анализируемого сигнала. Максимальный уровень разложения зависит от того, какие частотные диапазоны необходимо исследовать.

Величина энергии на заданном уровне вейвлет-разложения j с количеством детализирующих вейвлет-коэффициентов N_j определяется как

$$E_j = \frac{1}{N_j} \sum_{k=1}^{N_j} \text{det}^2(j, k). \quad (6)$$

Набор величин E_j для каждого уровня разложения составляет спектр вейвлет-энергии ряда. Полная вейвлет-энергия спектра представляет собой сумму энергий каждого уровня:

$$E_{\text{tot}} = \sum_{j=1}^N E_j, \quad (7)$$

где N – максимальный уровень разложения. Относительная вейвлет-энергия показывает распределение вейвлет-энергии по уровням разложения:

$$p_j = \frac{E_j}{E_{\text{tot}}}. \quad (8)$$

Вейвлет-энтропия WE является количественной мерой упорядоченности сигнала и определяется по формуле:

$$WE = - \sum_{j=1}^N p_j \ln(p_j). \quad (9)$$

Относительная вейвлет-энтропия RWE является мерой подобия распределений вейвлет-энергий $\{p_j\}$ и $\{g_j\}$ для двух сигналов или двух участков одного сигнала и вычисляется как

$$\text{RWE}(p|q) = \sum_{j=1}^N p_j \ln\left(\frac{p_j}{q_j}\right). \quad (10)$$

Значение относительной вейвлет-энтропии строго положительно и близко к нулю, если распределения энергий двух сигналов близки. Чем больше значение RWE, тем больше расхождение между распределениями и, соответственно, между сигналами.

АНАЛИЗ СТРУКТУРЫ ВРЕМЕННЫХ РЯДОВ С ПОМОЩЬЮ ВЕЙВЛЕТ-ЭНТРОПИИ

Покажем на модельных примерах, что адаптивное оценивание вейвлет-энтропии позволяет отследить появление сегментов с различным характером поведения. Для проведения численных исследований в качестве основных

модельных временных рядов были выбраны гармонические сигналы и реализации процессов авторегрессии.

На рис. 1 вверху слева представлен модельный сигнал, до середины реализации являющийся суммой двух синусов частотой 2 и 7 Гц, а после – синусоидой 13 Гц. Справа представлены спектры вейвлет-энергии на разных временных интервалах, полученные разложением сигнала на 7 уровней. Гармоники сигналов попадают в соответствующие диапазоны частот. Значения вейвлет-энтропий для приведенных спектров равны 1.33 и 1.21. Очевидно, что сигнал с одной гармонической составляющей является более упорядоченным, чем с двумя. Для определения структурных различий между первой и второй половинами сигнала, сигнал был разбит на участки длиной 1 с. Для каждого из участков по формуле (9) была рассчитана вейвлет-энтропия. Изменение величины вейвлет-энтропии во времени показано на рис. 1 внизу слева. Очевидно, что значение вейвлет-энтропии достигает наибольших значений в местах склеивания двух сигналов разной частоты.

На рис. 2 (вверху) представлена модельная реализация процессов авторегрессии, содержащая 5000 отсчетов. Первая часть реализации соответствует авторегрессии 1-го порядка с параметром $\phi = 0,3$. Вторая часть является авторегрессией 2-го порядка с коэффициентами $\phi_1 = 0,3$ и $\phi_2 = 0,4$. Для определения структурных различий между первой и второй половинами временного ряда реализация была разбита на интервалы, содержащие 500 отсчетов каждый. Для каждого из участков была рассчитана относительная вейвлет-энтропия. В средней части рис. 2 показано изменение величины относительной энтропии, вычисленной согласно формуле (10), где в качестве базового спектра $\{p_j\}$ выбран спектр авторегрессии 1-го порядка. В нижней части рис. 2 представлены значения величины относительной энтропии, где в качестве базового спектра выбран спектр вейвлет-энергии авторегрессии 2-го порядка.

Несмотря на значительную близость корреляционной структуры рядов, значения относительной вейвлет-энт-

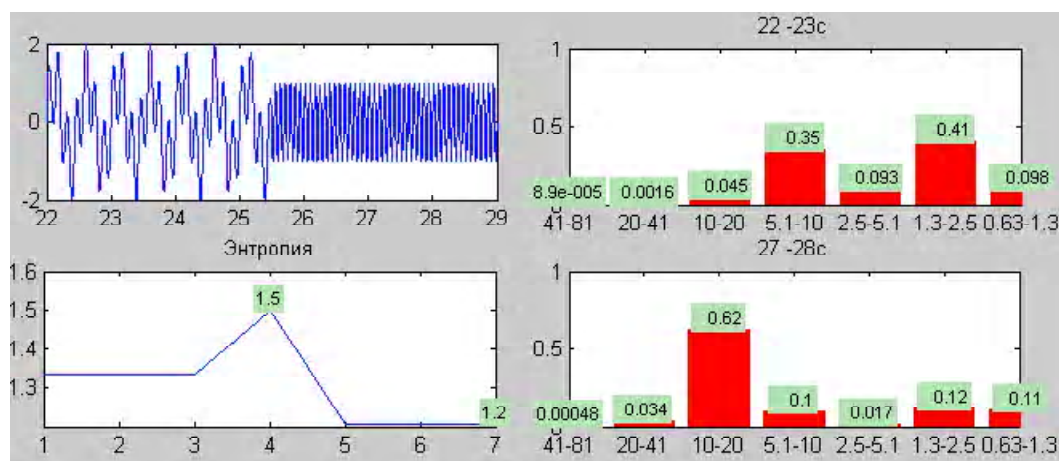


Рис. 1. Гармонический сигнал, спектр вейвлет-энергии и изменение вейвлет-энтропии

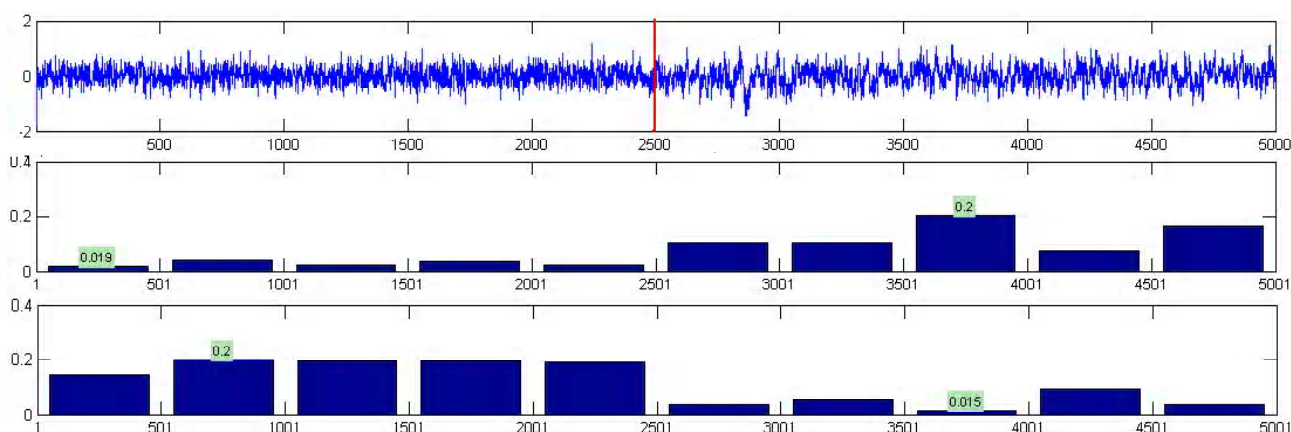


Рис. 2. Временной ряд авторегрессий 1-го и 2-го порядка (вверху); относительная вейвлет-энтропия, базовый спектр авторегрессии 1-го порядка (в середине); относительная вейвлет-энтропия, базовый спектр авторегрессии 2-го порядка (внизу)

ропии для одной половины ряда близки к нулю, а для второй на порядок больше. Это говорит о наличии частотной близости между участками сигнала и выбранным базовым вейвлет-спектром. Значения относительной вейвлет-энтропии изменяются при изменении частотной структуры временного ряда. Оба модельных примера показывают, что адаптивное оценивание вейвлет-энтропии позволяет успешно обнаруживать участки упорядоченности и изменения частотных компонент в различных сигналах.

ПРИМЕНЕНИЕ ОТНОСИТЕЛЬНОЙ ВЕЙВЛЕТ-ЭНТРОПИИ ДЛЯ АНАЛИЗА ВЫЗВАННЫХ ПОТЕНЦИАЛОВ В ЭЭГ-СИГНАЛЕ ЧЕЛОВЕКА

Исходные данные для исследований в данной работе были получены на специализированном сайте [14]. Данные содержали ЭЭГ десяти здоровых людей, записанные с помощью электроэнцефалографа системы BCI2000, который включает 64 электрода. Для каждого из испытуемых было записано по два ЭЭГ-сигнала: фоновая ЭЭГ (субъект сидит в удобном кресле с открытыми глазами в расслабленном состоянии) и ЭЭГ субъекта, выполняющего задачу на внимание. Во втором случае человек сидит в удобном кресле, напротив него расположен экран, на котором с частотой 8 секунд появляются маркеры с правой или левой стороны. При появлении маркера справа, субъект должен сжать правую кисть, слева – левую. Сигнал регистрируется с максимальной частотой 160 Гц. Длительность каждой записи ЭЭГ составляет 4 минуты.

Для исследования частотной структуры ЭЭГ было произведено многоуровневое вейвлет-разложение исходных сигналов. Количество уровней разложения было выбрано согласно основным исследуемым ритмам мозга: 40–80 Гц (гамма-ритм), 20–40 Гц (бета-ритм), 10–20 Гц (альфа-ритм), 5–10 Гц (тета-ритм), 2,5–5 Гц (дельта-ритм). Вейвлет-спектр не повторяет точно частотные ритмы че-

ловека, так как в соответствии с алгоритмом ДВП на последующих уровнях можно получить частотные диапазоны только в два раза короче, чем на предыдущем уровне. В качестве базовой вейвлет-функции был использован вейвлет семейства Добеши (Daubechies) степени 5.

На рис. 3 (вверху) показан график ЭЭГ человека, релаксирующего с закрытыми глазами. На соответствующем спектре вейвлет-энергии просматривается доминирование альфа-ритма, который активизируется при расслаблении с закрытыми глазами. В нижней части рис. 3 представлен график ЭЭГ человека, выполняющего сжатие и разжимание кисти руки. В этом случае на спектре вейвлет-энергии заметно преобладание низких частот в диапазоне 0,6–1,3 Гц.

При анализе ВП время реакции на стимул можно определять двумя способами [7]. В первом случае исследуется динамика вейвлет-энтропии WE. Время реакции t_{WE} – это время после стимула, за которое вейвлет-энтропия показывает наименьшее значение с момента подачи стимула. Данное время необходимо на перестройку частоты в период после стимула. Во втором случае анализируется динамика относительной вейвлет-энтропии, которая рассчитывается по формуле (10) с фоновым ЭЭГ-сигналом, когда субъект находится в расслабленном состоянии. Время реакции t_{RWE} – это время после стимула, за которое относительная энтропия достигает максимума. В момент t_{RWE} участок ЭЭГ с вызванными потенциалами и фоновая ЭЭГ показывают максимальную степень расхождения.

Для расчета вейвлет-спектра фоновых сигналов соответствующая запись ЭЭГ была разбита на равные непересекающиеся интервалы по 500 мс, для каждого в соответствии с формулами (6)–(8) были получены и усреднены значения вейвлет-энергии. Для ЭЭГ, содержащей стимулы, были проделаны следующие операции: выбраны участки длиной 4 секунды, начиная за 2 секунды до подачи сти-

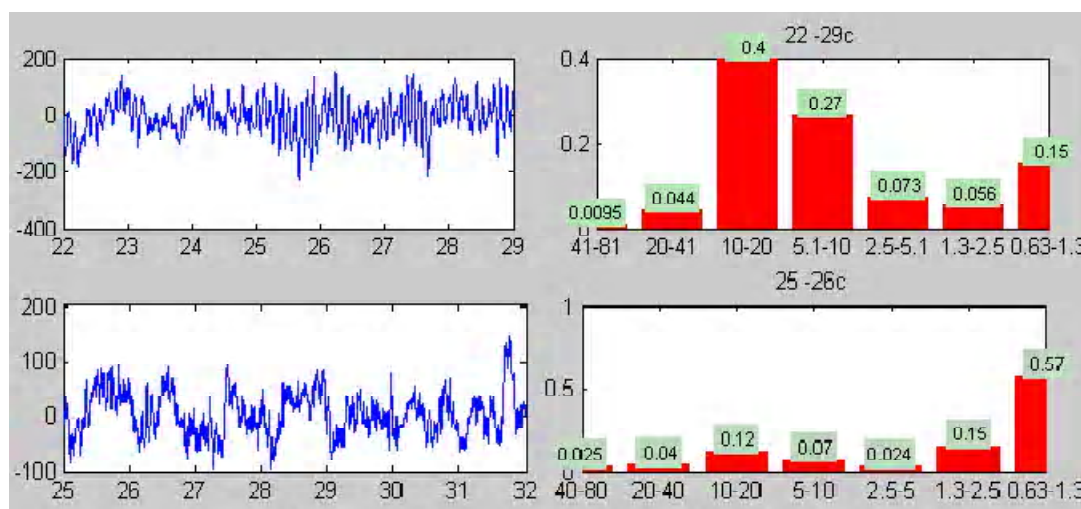


Рис. 3. ЭЭГ и спектр вейвлет-энергии: субъект в расслабленном состоянии (вверху), субъект выполняет сжатие и разжимание кисти (внизу)

мула; участки разбиты на непересекающиеся сегменты длиной 500 мс; для каждого участка получен спектр вейвлет-энергии и рассчитаны значения вейвлет-энтропии и относительной вейвлет энтропии к фоновой ЭЭГ.

Анализ распределений вейвлет-энергии (рис. 4) показывает, что после подачи стимула (вертикальная линия отмечает момент стимула) происходит смещение энергии в диапазон более низких частот: из альфа-ритма в тета-ритм. Время после стимула, в течение которого преобладает тета-ритм, варьируется от двух до четырех секунд.

Анализ распределения вейвлет-энергии в первые 500 мс после стимула показал, что распределение частот зависит от выбора электрода, с которого производится запись. Во фронтальной и центральной частях в период после стимула преобладает частота 5–10 Гц, а в затылочной – частота 10–20 Гц (рис. 5).

На рис. 6 представлены диаграммы изменения вейвлет энтропии (вверху) и относительной вейвлет-энтропии (внизу) для двух разных участков энцефалограммы со стимулом. В результате воздействия стимула значение энтропии начинает убывать, пока не достигает минимума в момент t_{WE} . Это связано с концентрацией энергии в области определенной частоты, соответствующей данному воздействию. Относительная вейвлет-энтропия через время t_{RWE} после стимула демонстрирует всплеск, как следствие максимального расхождения между распределением энергии сигнала со стимулом и фонового сигнала. Проведенные исследования показали, что для разных случаев время реакции на внешнее воздействие варьировалось от 300 до 600 мс, причем разница между величинами t_{WE} и t_{RWE} достигала 200 мс.

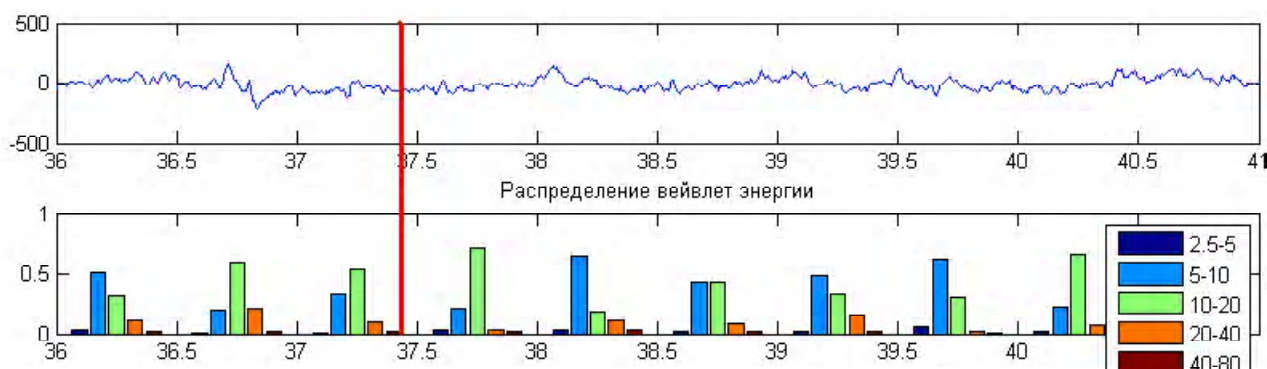


Рис. 4. Спектры вейвлет-энергии ЭЭГ-сигнала до и после стимула

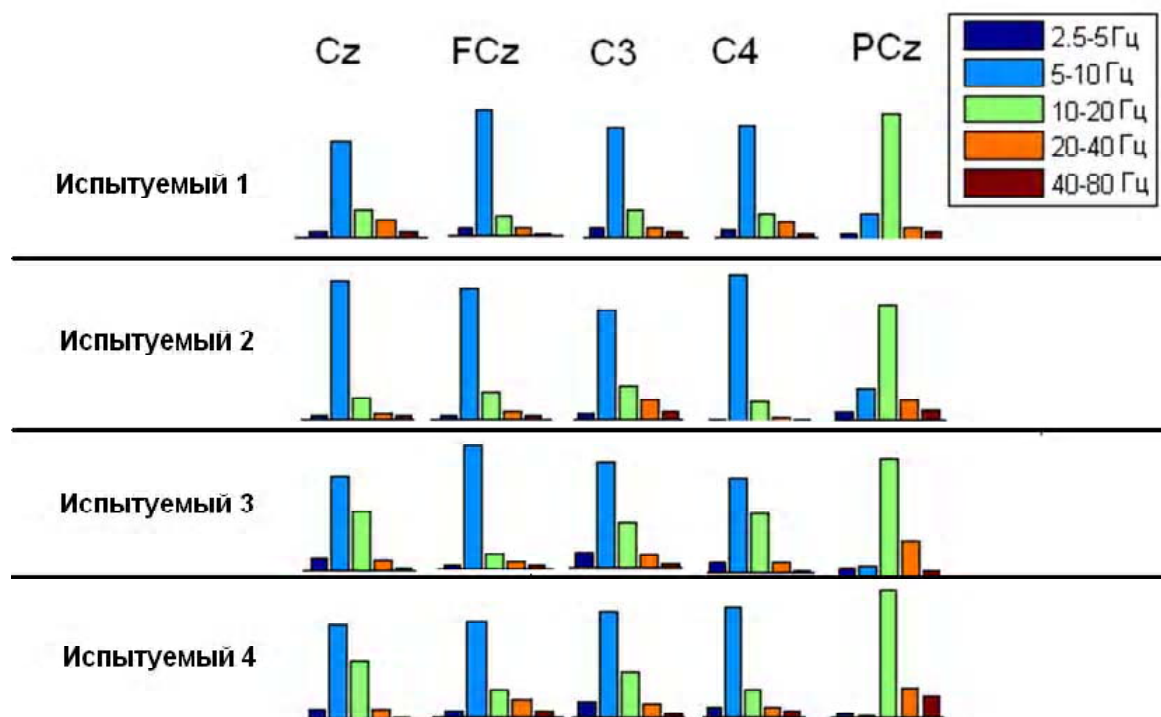


Рис. 5. Распределения вейвлет-энергии в первые 500 мс после стимула для 5 основных отведений ЭЭГ

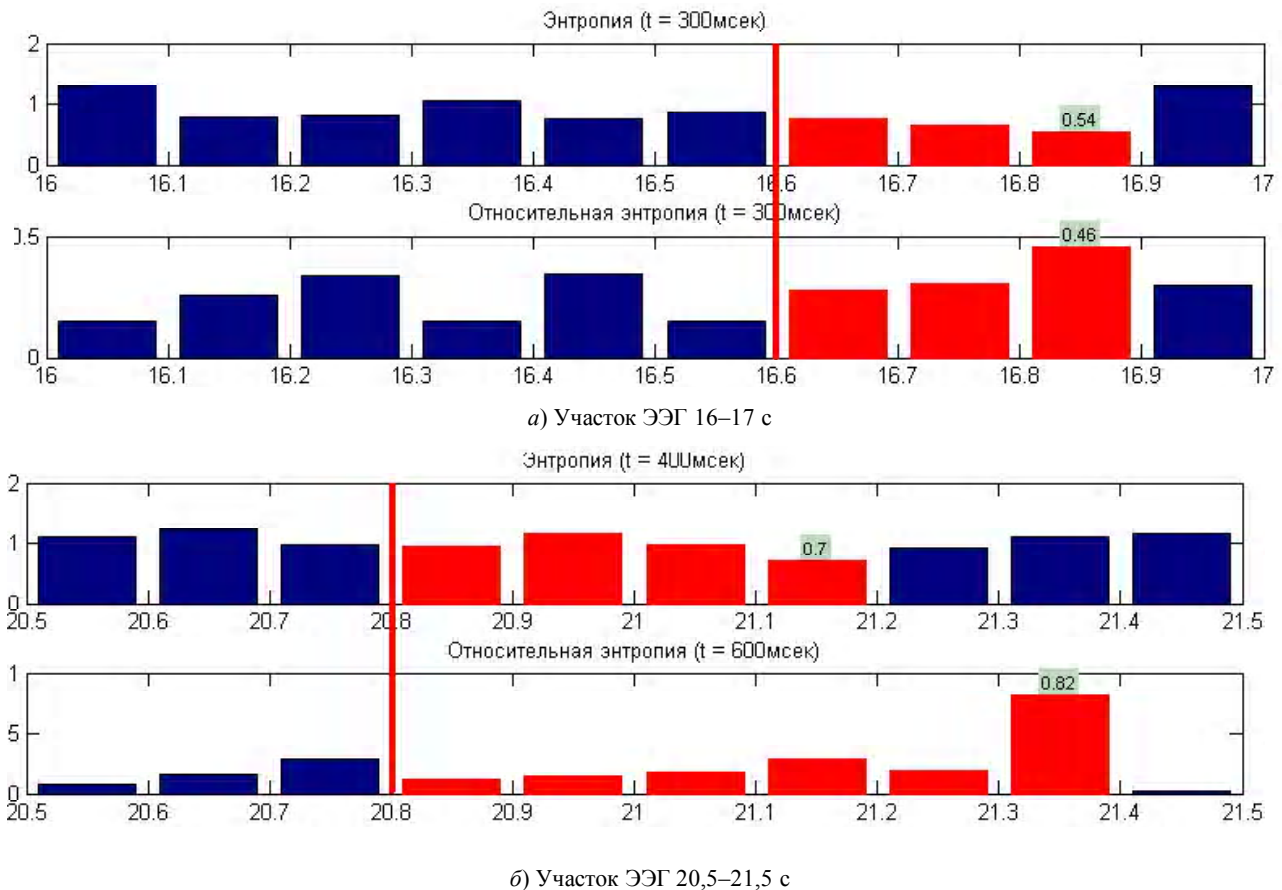


Рис. 6. Изменение вейвлет-энтропии после стимула (вверху), изменение относительной энтропии после стимула (внизу); вертикальная линия отмечает моменты стимулов

ВЫВОДЫ

В данной работе продемонстрировано, что ДВП является эффективным инструментом для анализа ЭЭГ и обнаружения отличительных особенностей колебаний связанных с ВП до и после стимула. Исследования ЭЭГ показали, что в состоянии покоя с открытыми глазами в ЭЭГ человека преобладает альфа ритм, однако при подаче визуального стимула происходит кратковременная синхронизация, характеризующаяся снижением вейвлет-энтропии, затем наблюдается период длиной 300–600 мс с преобладанием тета-ритма. Адаптивное оценивание вейвлет-энтропии и относительной энтропии на основе фоновое ЭЭГ-сигнала позволяет определить время реакции на внешний возбудитель. Результаты работы могут быть полезны при диагностике неврологических заболеваний, где необходима точная оценка времени реакции на стимул и характер колебаний ЭЭГ после стимула.

СПИСОК ЛИТЕРАТУРЫ

1. Рангайян, Р. М. Анализ биомедицинских сигналов: практический подход: пер. с англ. / Р. М. Рангайян. – М. : Физматлит, 2007. – 440 с.
2. Зенков, Л. Р. Функциональная диагностика нервных болезней: руководство для врачей / Л. Р. Зенков, М.А. Ронкин. – М. : Медицина, 1991. – 640 с.
3. Рутман, Э. М. Вызванные потенциалы в психологии и психофизиологии / Э.М. Рутман. – М.: Наука, 1979. – 216 с.
4. Опыт применения вызванных потенциалов в клинической практике / под ред. В. В. Гнездицкого, А. М. Шамшиновой. – М. : АОЗТ «Антидор», 2001. – 480 с.
5. Basar, E. EEG Brain Dynamics. Relation between EEG and Brain Evoked Potentials / E.Basar. – Amsterdam: Elsevier, 1980. – 434 p.
6. Basar, E. Event-related oscillations are 'real brain responses' / wavelet-analysis and new strategies / E. Basar, M. Schürmann, T. Demiralp, C. Basar-Eroglu, A. Ademoglu // International Journal of Psychophysiology. – 2001. – V. 39. – P. 91–127.
7. Rosso, O. A. Wavelet entropy: a new tool for analysis of short duration brain electrical signals / O. A. Rosso // Journal of Neuroscience Methods. – 2001. – V. 105. – P. 65–75.
8. Yordanova, J. Wavelet entropy analysis of event-related potentials indicates modality-independent theta dominance / J. Yordanova, V. Kolev, O. A. Rosso, M. Schürmann, O.W. Sakowitz, M. Цзгюрен, E. Basar // Journal of Neuroscience Methods. – 2002. – V. 117. – P. 99–109.
9. Giannakakis, G. A. Entropy Differentiations of Event Related Potentials in Dyslexia / G. A. Giannakakis, N. N. Tsiaparas, M. S. Xenikou, Ch. Papageorgiou, K. S. Nikita // proceeding 8th IEEE International Conference on BioInformatics and Bioengineering, Athens, Greece, 2008.
10. Малла, С. Вэйвлеты в обработке сигналов: пер. с англ. / С. Малла. – М. : Мир, – 2005. – 671 с.

11. Смоленцев, Н. К. Основы теории вейвлетов. Вейвлеты в MATLAB / Н. К. Смоленцев. – М.: ДМК Пресс, 2005. – 304 с.
12. Дьяконов, В. П. Вейвлеты. От теории к практике / В. П. Дьяконов – М.: СОЛОН-Пресс, 2002. – 440 с.
13. Zunino, L. Wavelet Entropy of Stochastic Processes / L. Zunino, D. G. Perez, M. Garavaglia, O. A. Rosso // Physica A. – 2007. – V. 379, N. 2. – P. 503–512.
14. PhysioNet: the research resource for complex physiologic signals [Электронный ресурс] – Режим доступа: www.physionet.org

Стаття надійшла до редакції 19.05.2011.

Кротких С. С., Кіріченко Л. О.

ДОСЛІДЖЕННЯ ВИКЛИКАНИХ ПОТЕНЦІАЛІВ У ЕЕГ ЛЮДИНИ ЗА ДОПОМОГОЮ ДИСКРЕТНОГО ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ

В роботі за допомогою дискретного вейвлет-перетворення проведено аналіз змінювання частотної структури елект-

роенцефалограми з викликаними потенціалами після впливу стимулу. Було реалізовано метод визначення часу реакції на стимул, який заснований на оцінюванні вейвлет-ентропії та відносної вейвлет-ентропії сегментів енцефалограми.

Ключові слова: Електроенцефалограма, викликані потенціали, дискретне вейвлет-перетворення, спектр вейвлет-енергії, вейвлет-ентропія, відносна вейвлет-ентропія.

Krotkih S. S., Kirichenko L. O.

ANALYSIS OF EVENT-RELATED POTENTIALS OF EEG SIGNAL USING DISCRETE WAVELET TRANSFORM

In this work we use discrete wavelet transform for analyzes the frequency structure of EEG signal with evoked potentials after effect of stimulus. The method for determining the response time to a stimulus, based on the evaluation of the wavelet entropy and relative wavelet entropy EEG, has been implemented.

Key words: electroencephalogram, evoked potentials, discrete wavelet transform, wavelet energy spectrum, wavelet entropy, relative wavelet entropy.

УДК 004.853+004.832

Литвин В. В.

Канд. техн. наук, доцент Національного університету «Львівська політехніка»

МОДЕЛЮВАННЯ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ З ВИКОРИСТАННЯМ ОНТОЛОГІЧНОГО ПІДХОДУ

Досліджено функціонування інтелектуальних систем підтримки прийняття рішень, ядром баз знань яких є онтології. Розглядаються інтелектуальні агенти чотирьох типів. Для кожного типу розроблено метрику, яку використовують для визначення релевантності пропонованих системою рішень.

Ключові слова: інтелектуальна система підтримки прийняття рішень, інтелектуальний агент.

АНАЛІЗ ДОСЛІДЖЕНЬ ТА ПОСТАНОВКА ПРОБЛЕМИ В ЗАГАЛЬНОМУ ВИДІ

Технологія інтелектуальних систем підтримки прийняття рішень (ІСППР) є одним із найрозвинутіших напрямків штучного інтелекту. Дослідження у цій області полягають у розробці автоматизованих інформаційних систем, які застосовуються у тих областях діяльності людини, які вимагають логічного міркування, певної майстерності та досвіду.

Сучасний рівень розвитку ІСППР відбувається у двох напрямках розроблення інтелектуальних агентів (ІА) [1]:

– ІА, засновані на прецедентах (англійською – Case-Based Reasoning, або CBR);

– ІА планування діяльності (пошук у просторі станів).

Вибір ІА залежить від задачі. Метод виведення за прецедентами ефективний, коли основним джерелом знань про задачу є досвід, а не теорія; рішення не є унікальними для конкретної ситуації, а можуть бути використані

в інших випадках; метою розв’язування задачі є отримати не гарантований вірний розв’язок, а кращий з можливих. Виведення, засноване на прецедентах, є методом побудови ІСППР, які приймають рішення щодо даної проблеми або ситуації за наслідками пошуку аналогій, що зберігаються в базі прецедентів [2]. Такий прецедент називають релевантним. З математичної точки зору серед елементів множини прецедентів $Pr = \{Pr_1, Pr_2, \dots, Pr_N\}$ релевантним Pr_k є прецедент для якого відстань до поточної ситуації S є найменшою, тобто

$$Pr_k = \arg \min_i d(Pr_i, S).$$

ІА планування діяльності повинен досягнути цільового стану. Насамперед він повинен побудувати план досягнення цього стану із всіма можливими альтернативами. Процес планування ґрунтується на декомпозиції. Задача планування ZP містить 3 складові [3]: множину

станів St , множину дій A , множину цільових станів $Goal$ (станів мети); тобто

$$ZP = \langle St, A, Goal \rangle.$$

Отже, для планування діяльності ІА повинен вміти оцінювати стани та дії.

Як бачимо для обох класів ІСППР необхідна метрика. У першому випадку для оцінювання релевантності прецедентів, у другому випадку – для оцінювання релевантності станів. Від способу визначення цієї метрики напряму залежить ефективність функціонування ІА. На наш погляд такий спосіб повинен базуватись на чіткому і аргументованому стандарті баз знань. У галузі інженерії знань таким стандартом стали онтології [4]. Тому нами запропоновано для побудови метрики використовувати онтології.

Під моделлю онтології O розуміють трійку вигляду:

$$O = \langle C, R, F \rangle,$$

де C – поняття; R – відношення між поняттями; F – інтерпретація понять та відношень (аксіоми). Аксіоми встановлюють семантичні обмеження для системи понять та відношень.

На сьогодні розрізняють три типи онтологій: предметно-орієнтовані (Domain-oriented), орієнтовані на прикладну задачу (Task-oriented) та загальні онтології (Top-level).

Мета роботи – побудувати моделі функціонування інтелектуальних систем підтримки прийняття рішень, які базуються на онтологіях в залежності від класу задач.

Постановка задачі. Побудувати метрики для оцінювання релевантності прецедентів в задачах пошуку аналогій та оцінювання релевантності станів для задач пошуку цільового стану. Ці метрики повинні базуватись на онтологіях. Апробувати отримані моделі функціонування ІСППР шляхом розроблення прикладних ІСППР.

КЛАСИФІКАЦІЯ ПРИКЛАДНИХ ЗАДАЧ З ТОЧКИ ЗОРУ ЇХ МЕТРИЗАЦІЇ

Проаналізувавши клас задач, для яких розробляють ІСППР, можна зробити висновок, що всі задачі можна поділити на два підкласи. Існує клас задач для яких суттєве значення понять (властивостей). Сюди відносяться задачі діагностики захворювань, розпізнавання образів, класифікація явищ на основі збору даних, тощо. Такі задачі назвемо ознаковими. Для іншого класу задач не є суттєвим значення понять, а скоріше їх семантика або частотність зустрічання термінів в тексті і т. д. Сюди можна віднести кластеризацію інформаційних ресурсів, класифікацію текстів згідно УДК, інтелектуальні пошукові системи, реферування та анотування текстових документів. Такий клас задач назвемо семантичними задачами. В результаті отримаємо поділ ІСППР за двома вимірами так, як це зображено на рис. 1. У кожній чверті перераховано окремі задачі, які попадають у відповідний клас.



Рис. 1. Класи задач, для розв'язування яких використовують ІСППР

Для ефективного функціонування ІА необхідно побудувати метрику, на основі якої визначати релевантність станів чи прецедентів. На наш погляд побудова такої метрики напряму залежить від класу задач: семантичні вони чи ознакові.

Отже, загалом, на нашу думку, виділяється чотири різних класи задач, які розв'язують ІСППР. Зріз за напрямками потребує дві різні функціональні моделі (пошук релевантних прецедентів та планування діяльності), зріз за типом задачі – використання двох різних метрик для розв'язування цих задач та оцінки якості отриманих розв'язків. Далі розглянемо всі ці класи задач, однак, насамперед введемо поняття адаптивної онтології.

Ефективність адаптації онтології бази знань до особливостей предметної області визначають закладені в її структуру елементи та механізми її адаптації шляхом самонавчання під час експлуатації. Одним з підходів до реалізації таких механізмів є автоматичне зважування понять бази знань (БЗ) та семантичних зв'язків між ними під час самонавчання. Цю роль беруть на себе коефіцієнти важливості понять та зв'язків [5]. Коефіцієнт важливості поняття (зв'язку) – це чисельна міра, котра характеризує значущість певного поняття (зв'язку) у конкретній предметній області і динамічно змінюється за певними правилами у процесі експлуатації системи. Отже, ми розширимо поняття онтології, ввівши в її формальний опис коефіцієнти важливості понять та відношень. Тому таку онтологію ми будемо визначати як п'ятірку:

$$O = \langle C, R, F, W, L \rangle,$$

де W – важливість понять C ; L – важливість відношень R .

Визначену, таким чином, онтологію будемо називати адаптивною, тобто такою, що адаптується до ПО за рахунок модифікації понять та коефіцієнтів важливості цих понять і зв'язків між ними [6]. Така онтологія однозначно представляється у вигляді зваженого концептуального графа (КГ) [7]. Тому метрику ми будемо будувати, використовуючи зважені КГ.

МЕТРИЗАЦІЯ ІНТЕЛЕКТУАЛЬНИХ АГЕНТІВ, БАЗОВАНИХ НА ПРЕЦЕДЕНТАХ

Вочевидь, що в залежності від прецеденту ваги понять різні. Тобто насправді W – вектор вимірності кількості прецедентів $W=(W_1, W_2, \dots, W_N)$. Надалі будемо розглядати лише один прецедент, тобто нижній індекс у вазі понять опускатимемо.

Побудуємо метрику для семантичних задач, які використовують прецеденти. Нами пропонується визначати відстань між прецедентом і ситуацією, як суму відстаней між найважливішими поняттями прецедента та поточного випадку. Таких важливих понять може бути одне, два; однак якщо їх є більше-рівне трьох, то нами пропонується вибирати перші три. Ця кількість визначена на основі опитувань експертів різних ПО і вважається ними оптимальною. У такому випадку ми маємо 3 центри ваг i -го прецедента pr_i^1, pr_i^2, pr_i^3 і 3 центри ваг поточної ситуації s^1, s^2, s^3 . Тоді існує 9 різних відстаней $d(pr_i^j, s^k)$, $j=1,2,3; k=1,2,3$. Вибираємо 3 найменші з них та їх сумуємо. Отримана таким чином сума й буде відстанню між прецедентом та поточною ситуацією. Найважливішим є поняття, яке є центром ваг КГ. Центром ваг КГ є поняття, середня відстань від якого до всіх інших понять є найменшою. Вочевидь, що визначена таким чином відстань залежатиме від того як ми визначимо відстань між двома суміжними вершинами КГ. Для цього пропонується визначати відстані між вершинами, що з'єднані зв'язком як

$$d_{ij} = \frac{Q}{L_{ij}(W_i + W_j)}, \quad (1)$$

де W_i та W_j – коефіцієнти важливості вершин C_i та C_j відповідно; L_{ij} – коефіцієнт важливості зв'язку між вершинами; Q – константа, яка залежить від конкретної онтології. Приймемо, що $L_{ii} = \infty$, тоді $d_{ii} = 0$.

Далі знаходимо центри ваг концептуального графа.

Це перші три вершини для яких середня відстань $\bar{d}_i$ є найменшою:

$$\bar{d}_i^* = \min_i \bar{d}_i. \quad (2)$$

Середня відстань $\bar{d}_i$ для вершини C_i обчислюється згідно формули:

$$\bar{d}_i = \frac{\sum_{j=1, j \neq i}^n d_{ij}^*}{n-1}, \quad (3)$$

де n – кількість вершин графа; d_{ij}^* – найкоротший шлях між вершинами C_i та C_j , який обчислюється за допомогою відомих алгоритмів, наприклад Форда, Дейкстри, Флойда-Уоршалла [8].

Далі згідно концептуального графа, що задає онтологію прецедента шукаємо відстань від даного прецедента до поточної ситуації. Якщо поняття поточної ситуації не входять в концептуальний граф, то онтологію даного прецедента доповнюємо онтологією всього ІА до якого входить цей прецедент. Якщо ж необхідне поняття далі не входить в онтологію ІА, то його відсутність зумовлює ріст відстані до безмежності, що означає не близькість прецедента із поточною ситуацією.

Зазначимо, що запропонована таким чином відстань задовольняє трьом аксіомам метрики [5].

Ефективність розробленої відстані покажемо на прикладі аналізу анотацій наукових статей. Розглянемо три анотації статей.

1. *The work is carried out in a direction of information technologies development focused on the natural language information processing. On position of author offered model the problem of putting a language material in order with the help of the uniform standard, it is considered rather significant for the given class of technologies. It is one of the central problem on way of development of the effective technologies for language information processing. (III, № 4, 2004, с. 613)*

Відповідний концептуальний граф першої анотації поданий на рис. 2.

2. *The article presents the basic concepts on researching and solving the task of automatic knowledge retrieval from text documents. Industrial system that solves the stated above task is described as well as one of its main application. (III, №3, 2004, с. 668).*

Відповідний концептуальний граф матиме вигляд, як наведено на рис. 3.

3. *The paper is dedicated to the problem of automated text consistency analysis. It is proposed to implement text consistency via text logic analysis with the attraction of the knowledge of application domain, natural language and normative base. (III, № 3, 2004, с. 660).*

Концептуальний граф третьої анотації наведений на рис. 4.

Значення ваг понять та зв'язків взято із тестової онтології «напряму комп'ютерних наук» для прецеденту «штучний інтелект». У випадках, коли у графі немає зв'язку, вагу такого зв'язку вважали рівною 5, що є середнім значенням ваги зв'язків. Користуючись формулою (1), в якій Q приймається рівним 100, отримуємо зважені графи, які наведено на рис. 5.

Використовуючи алгоритм Дейкстри та роблячи відповідні обчислення за формулами (2) та (3), отримуємо, що центрами ваг відповідних графів є:

$$\bar{d}^1 = \{1\} = \{\text{'technologies'}\}, \quad \bar{d}^2 = \{4\} = \{\text{'system'}\},$$

$$\bar{d}^3 = \{5\} = \{\text{'knowledge'}\}.$$

1-й і 3-й тексти легко зв'язуються за допомогою вершини 'natural_language' (5-а в 1-му тексті і 7-а в 3-му тексті). Тоді $d_{5,11}^1 = 0,42$; $d_{5,7}^3 = 0,23$. А відстань між

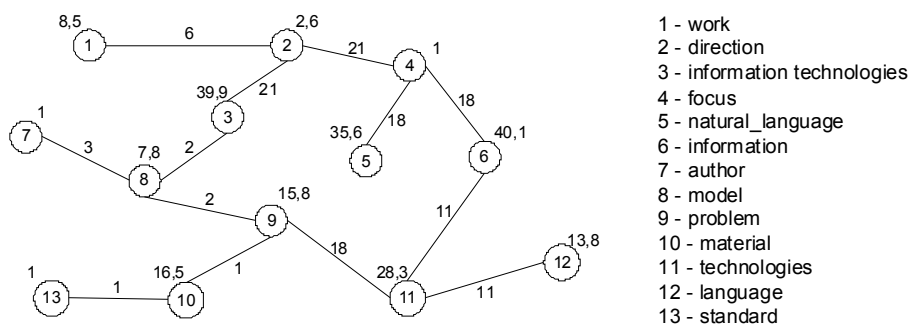


Рис. 2. Концептуальний граф 1-ї анотації

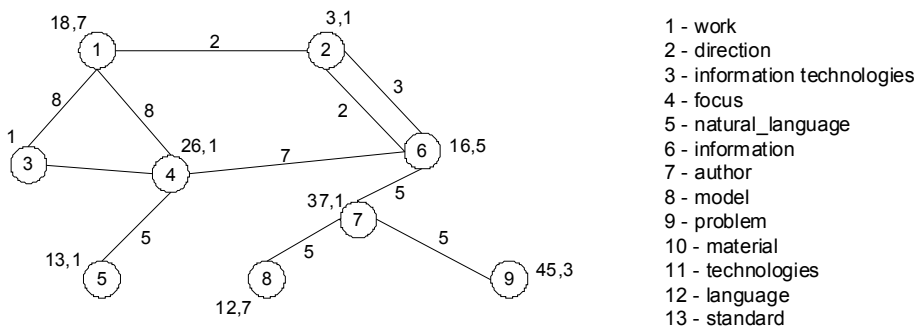


Рис. 3. Концептуальний граф 2-ї анотації

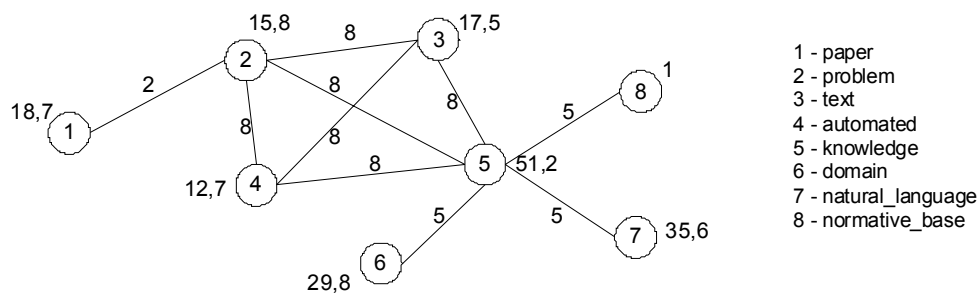


Рис. 4. Концептуальний граф 3-ї анотації

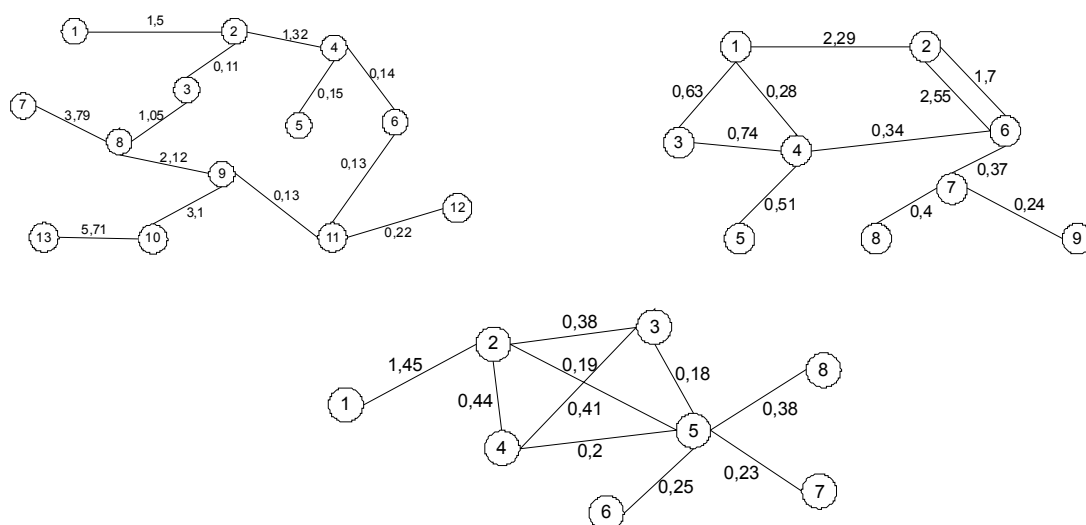


Рис. 5. Зважені графи для трьох анотацій

1-м і 3-м текстом тоді рівна 0,65.

$$\bar{d}^{13} = 0,42 + 0,23 = 0,65.$$

2-й і 3-й тексти легко зв'язуються за допомогою вершини 'knowledge' (7-а в 2-му тексті і 5-а в 3-му тексті). Тоді, $d_{4,7}^2 = 0,71$, що є відстанню між 2-м та 3-м текстом, оскільки 'knowledge' – центр ваг 3-го графа. Тобто $\bar{d}^{23} = 0,71$.

Для обчислення відстані між 1-м і 2-м текстом скористаємося 3-м текстом, оскільки в перших двох немає спільних вершин. Тоді

$$\bar{d}^{12} = \bar{d}^{13} + \bar{d}^{23} = 0,65 + 0,71 = 1,36.$$

Отже, найближчими за змістом є 1-й і 3-й текст, найдалшими є 1-й і 2-й текст.

Побудуємо метрику для ознакових задач, які використовують прецеденти.

Нехай множина прецедентів $Pr = \{Pr_1, Pr_2, \dots, Pr_N\}$ описується характеристиками (властивостями) $X = \{x_1, x_2, \dots, x_M\}$. D_i – домен властивості x_i , w_{i_l} – коефіцієнт важливості властивості x_i прецедента Pr_i . Значення властивості x_i позначимо $z_i = z(x_i)$. Отже

$$Pr_i \leftrightarrow X_i = \{x_{i_1} = z_{i_1}, x_{i_2} = z_{i_2}, \dots, x_{i_k} = z_{i_k}\}, \text{ де } z_{i_j} \in D_{i_j}.$$

Позначимо I_i – множина індексів властивостей прецедента Pr_i . Тоді відстань між прецедентом Pr_i та поточною ситуацією S визначається як:

$$d_i = \sum_{i_l \in I_i} \varphi(z_{i_l}, z_{i_l}^S), \quad (4)$$

де z_{i_l} – значення властивості x_{i_l} прецедента Pr_i ; $z_{i_l}^S$ – значення властивості x_{i_l} поточної ситуації S ; $\bar{I}_i$ – множина індексів важливих властивостей прецедента Pr_i ; $\bar{I}_i = \bar{I}_{i_1} \cup \bar{I}_{i_2} \cup \dots \cup \bar{I}_{i_{N_i}}$; N_i – кількість властивостей, необхідних розглянути щоб прийняти рішення стосовно прецедента Pr_i . Тобто

$$\bar{I}_{i1} = \left\{ i_{s1} \left| i_{s1} = \arg \max_{i_l \in I_i} w_{i_l} \right. \right\},$$

$$\bar{I}_{i2} = \left\{ i_{s2} \left| i_{s2} = \arg \max_{i_l \in I_i / i_{s1}} w_{i_l} \right. \right\},$$

$$\bar{I}_{i3} = \left\{ i_{s3} \left| i_{s3} = \arg \max_{i_l \in I_i / i_{s1} / i_{s2}} w_{i_l} \right. \right\}, \dots$$

Розглянемо функцію $\varphi(\xi, \eta)$. Очевидно, що ξ – може бути діапазоном, тобто нечіткою підмножиною $\xi \subseteq D$, де D – універсальна множина; числовим значенням або нечисловим значенням. В залежності від цього $\varphi(\xi, \eta)$ визначається по своєму, а саме:

$$\varphi(\xi, \eta) = \begin{cases} 1 - \mu_\xi(\eta), & \xi - \text{нечітка множина,} \\ \lambda \cdot |\xi - \eta|, & \xi, \eta - \text{числові значення,} \\ 1 - \mu(\xi, \eta), & \xi, \eta - \text{нечислові значення,} \end{cases} \quad (5)$$

де $\mu_\xi(\eta)$ – коефіцієнт впевненості того, що η належить нечіткій підмножині ξ ; λ – числова величина, яка залежить від ПО, щоб добуток $\lambda \cdot |\xi - \eta| \in [0, 1]$; $\mu(\xi, \eta) \in [0, 1]$ – нечітка величина подібності значень ξ та η . Наприклад $\mu(\xi, \eta) = 1$, якщо $\xi = \eta$, $\mu(\xi, \eta) = 0,9$, якщо $\xi \approx \eta$, $\mu(\xi, \eta) = 0$, якщо $\xi \neq \eta$.

Розглянемо приклад функціонування ІА в області медицини, а саме захворювань опорно-рухового апарату. За основу візьмемо результати досліджень, що одержали у Львівському Національному медичному університеті ім. Данила Галицького Е. Х. Заремба й О. О. Зімба. З деякими з них можна ознайомитися в [9]. Прецедентами є можливі захворювання. Для прикладу розглянемо три хвороби: $Pr = \{\text{'Ревматизм'}, \text{'Артрит'}, \text{'Подагра'}\}$.

Провівши онтологічний інжиніринг ПО, ми одержали множину властивостей, які необхідно дослідити та їх важливість в залежності від прецеденту. Наведемо деякі з них: $X = \{\text{'Ступінь недиференційованої дисплазії сполучної тканини'}, \text{'Концентрація ендотеліну-1 у плазмі крові'}, \text{'Температура'}, \text{'Біль у суглобі'}\}$. Розглянемо детальніше ці ознаки. Домен першої є нечислові значення, другої і третьої – відрізки, четвертої – бінарне значення. Так $D_{\text{С-н-д-с-т}} = \{\text{слабке, середнє, сильне}\}$, $D_{\text{К-е-п-к}} = [0; 3,5]$, $D_{\text{Температура}} = [36,42]$, $D_{\text{Біль у суглобі}} = \{\text{Так(1), Немає(0)}\}$. Значення цих ознак для ревматизму: $\text{Температура}_{\text{Ревматизм}} = \{(36|0), (37|0,6), (38|1), (39|1), (40|0,7)\}$ – нечітка множина, $\text{Біль у суглобі}_{\text{Ревматизм}} = \{\text{Так(1)}\}$.

Важливість цих властивостей (коефіцієнти W) ми одержали методами статистичного аналізу (було досліджено 110 пацієнтів). Так для захворювання на ревматизм важливими симптомами є: наявність лади ($w=0,7$), астенічна статура ($w=0,65$), тонка шкіра ($w=0,45$), сколіоз ($w=0,35$). Перший симптом являє собою бінарну величину, що приймає значення з множини $\{\text{Так(1), Немає(0)}\}$, дві наступні – нечіткі величини, що приймають значення з відрізка $[0; 1]$ (зрозуміло, що 0 – відсутність симптому, 1 – повна впевненість у його наявності) і четвертий симптом теж нечітка змінна, тільки її носій – відрізок $[0;3]$. Для ревматизму приймаємо такі значення симптомів: $z_{\text{ладу}} = 1$, $z_{\text{ас}} = 1$, $z_{\text{тк}} = 1$, $z_{\text{сколіоз}} \in [1,5; 3]$.

Нехай при обстеженні пацієнта Y ми одержали такі значення властивостей $y_{\text{ладу}}, y_{\text{ас}}, y_{\text{тк}}, y_{\text{сколіоз}}$. Тоді відстань до прецедента 'Ревматизм' вимірюється як:

$$d_{\text{ревматизм}} = 0,7 \cdot |1 - y_{\text{ладу}}| + 0,65 \cdot |1 - y_{\text{ac}}| + 0,45 \cdot |1 - y_{\text{mk}}| + 0,35 \cdot |1 - \mu_{[1,5;3]}(y_{\text{сколіоз}})|.$$

Аналогічним чином можна виміряти відстань до інших захворювань ('Артрит', 'Подагра'). Пацієнт хворий тим захворюванням, відстань до якого є найменшою.

МЕТРИЗАЦІЯ ІНТЕЛЕКТУАЛЬНИХ АГЕНТІВ ПЛАНУВАННЯ ДІЯЛЬНОСТІ

Розглянемо спочатку ознакові задачі. Нехай $v(St(i))$ – оцінка стану $St(i)$. a_{ij}^k – перехід із стану $St(i)$ в стан $St(j)$, використовуючи альтернативу α_k . Наприклад, для зняття захисного покриття з поверхні трубопроводу можна використати три альтернативи: *механічний, хімічний або термічний спосіб зняття захисного покриття*. $v(a_{ij}^k)$ – оцінка дії a_{ij}^k . Стан мети *Goal* визначається необхідністю деякої підмножини ознак X досягнути певних значень $z(x, Goal) \forall x \in X$.

Будь-який стан $St(i)$ задається своєю множиною ознак Y_i , які набувають значень $z(y, St(i)) \forall y \in Y_i$.

Для оцінки стану $St(i)$ необхідно здійснити відображення Ψ множини ознак та їх значень стану $St(i)$ в множину ознак та значень стану *Goal* за рахунок онтології, тобто

$$\Psi: Y_i \xrightarrow{0} X. \quad (6)$$

Тоді оцінка стану $v(St(i))$ обчислюється

$$v(St(i)) = d(St(i), Goal) = \sum_{x \in X} w_x \varphi(z(x, St(i)), z(x, Goal)), \quad (7)$$

де w_x – важливість ознаки x в адаптивній онтології, функція φ така ж як у формулі (5).

Вочевидь, що чим оцінка стану менша, тим стан кращий.

У наших дослідженнях для вибору дій ІА ми спиратимемося на раціональність агента як прагнення мінімізувати витрати ресурсів для досягнення кінцевого стану.

Тому вважаємо, що кожна дія a_{ij}^k однозначно визначається витратами ресурсів g_{ij}^k (ціна переходу зі стану в стан), де $k = 1, 2, \dots, n_i$. n_i – кількість альтернатив α_k для здійснення переходу a_{ij} . Кожна з альтернатив характеризується витратами ресурсів та терміном експлуатування. Інформація про альтернативи та витрати ресурсів повинна зберігатися в онтології. Інформація про значення ознак та вигоди від переходу в стан (терміни експлуатування тощо) зберігається в базі даних. Вочевидь, що можуть появлятися нові альтернативи, тому інтелектуальному агенту необхідно постійно відстежувати наукові видання з метою їх пошуку та заносити в онтологію.

Оцінка дії прямопропорційна витраті ресурсів, тобто:

$$v(a_{ij}^k) = E \cdot g_{ij}^k, \quad (8)$$

де E – скалярна величина, яка зводить вимір оцінки дії до одного виміру з оцінкою станів.

Загалом рішення стосовно вибору дії на основі альтернативи здійснюємо згідно формули:

$$0(a_{ij}^k) = \omega v(a_{ij}^k) + (1 - \omega) v(St(j)), \quad (9)$$

де $\omega \in [0, 1]$ – частка альтернативи, яку ІА віддає в процесі прийняття рішення, інша частка належить стану в який він перейде. Перехід здійснюємо у той стан для якого $0(a_{ij}^k)$ найменше.

Після оцінки дій та станів, задача вибору шляху зводиться до задачі асинхронного динамічного програмування [10]. Використовуючи методи придатні для розв'язування таких задач, знаходимо розв'язок у вигляді шляху переходу з початкового у кінцевий стан.

Для семантичних задач про стан мети *Goal* наперед щось важко сказати. Наприклад, для задачі реферування текстових документів станом мети є кінцевий реферат, однак ми лише можемо собі уявляти як він приблизно має виглядати. Оцінка стану в такій задачі співпадає з оцінкою важливості семантичної одиниці (слово, лексема, речення), залежно від задачі. Нами запропоновано для таких задач будувати метрику на основі зважування міри TF-IDF онтологією ПО [11]. Тобто

$$v(St) = (TF - IDF) \cdot W. \quad (10)$$

Така оцінка містить істотні переваги у порівнянні з іншими оцінками, оскільки у ній одночасно враховується як частотний аналіз зустрічання термінів у тексті (TF-IDF), так і специфіка предметної області до якої належить тематика цього тексту. Новий стан для задач реферування полягає в додаванні у квазіреферат нових речень. Детальніше ця задача нами розглянута у [11].

МЕТОДИ ЗАДАННЯ ВАГ ВАЖЛИВОСТІ ПОНЯТЬ ТА ЗВ'ЯЗКІВ

Розглянемо методи задання початкових коефіцієнтів важливості понять та зв'язків та їх модифікацію в процесі функціонування інтелектуального агента для розв'язування ознакових та семантичних задач.

Методи задання початкових ваг (коефіцієнтів важливості) понять та зв'язків:

- за рахунок експертних оцінок;
- присвоєння випадковим чином;
- за рахунок аналізу (статистичного, інтелектуального) інформаційних джерел, які описують ПО в якій функціонує ІА.

Окрім того, ці ваги можуть мати обмеження на величину, наприклад їх значення знаходяться у відрізку $[0, 1]$ (ймовірнісні методи), або без обмежень на величину (нагромаджувальні методи). Оскільки онтологія формує собою таксономію понять, то використовуючи мову

об'єктно-орієнтованого підходу, кожне поняття являє собою клас.

Для семантичних задач нами запропоновано такий метод обчислення ваг класів:

1. Повна вага W_j^i класу онтології дорівнює сумі власної ваги $W0_j^i$, ваги підкласів Ws_j^i та ваги суміжних класів Wn_j^i (класів, зв'язаних з даним класом не IS-A зв'язком):

$$W_j^i = W0_j^i + Ws_j^i + Wn_j^i, \quad (11)$$

де $Ws_j^i = \sum_k Wc_k^{i+1} \cdot L_{j,k}$ – вага k підкласів j -го класу i -го рівня, причому для кореневого класу рівень $i = 0$; $Wc_k^{i+1} = W0_k^{i+1} + Ws_k^{i+1}$ – вага класу C_k^{i+1} ; $L_{j,k}$ – вага зв'язку між класами C_j^i та C_k^{i+1} .

Перерахунок окремих компонент повної ваги класу відображено на схемі (рис. 6).

2. У момент внесення на $i+1$ -й рівень нового підкласу йому присвоюється власна вага $W0_j^{i+1}$, рівна половині власної ваги класу вищого i -го рівня:

$$W0_j^{i+1} = \frac{1}{2} W0_j^i. \quad (12)$$

Вага класу Wc_j^i та усіх батьківських класів аж до кореневого збільшується на величину ваги новоствореного підкласу:

$$Wc_j^m = Wc_j^m + W0_j^{i+1}, \quad \forall m \leq i. \quad (13)$$

3. Під час встановлення зв'язку між поняттями k_1 та k_2 між відповідними вершинами графа онтології з'являється ребро, а до ваги суміжних класів Wn_1 додається вага Wc_2 і навпаки – до Wn_2 додається вага нового суміжного до нього класу Wc_1 , так що:

$$Wn_j = \sum_k Wc_k \cdot L_{j,k}. \quad (14)$$

Повторне встановлення зв'язків приводить до появи кратних ребер у графі.

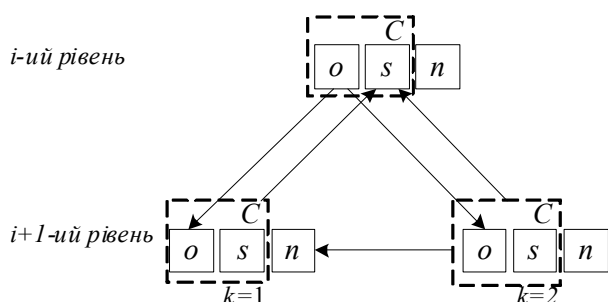


Рис. 6. Схема перерахунку окремих компонент повної ваги класу

4. Кратність ребер відображає частоту зустрічання V пари семантично пов'язаних понять $L_{i+1} = V \cdot L_i$. Кратні ребра після перерахунку не збільшують валентність вершини.

5. Вага екземпляра БЗ дорівнює повній вазі його класу.

Таким чином, визначена модель онтології БЗ дає змогу розраховувати вагові коефіцієнти своїх компонентів у процесі їх додавання, вилучення і використання під час експлуатації системи, завдяки чому реалізує механізм адаптації до заданої користувачем ПО [12].

Зауважимо, що насправді коефіцієнти важливості понять та зв'язків між ними є векторними величинами розмірності кількості прецедентів або різних тематик визначених в конкретній ПО. Так, якщо ми розглядаємо задачу класифікації текстових документів, то в межах онтології має описуватись кілька різних тем. Тому коефіцієнти важливості понять залежать від тематики. Нехай онтологія O описує m тем ПО – $Th_1, Th_2, \dots, Th_m$, тоді коефіцієнти ваг понять є вектор, компонентами якого є відповідні значення в залежності від теми, тобто $W = (W_1, W_2, \dots, W_m)$. Для процесу автоматизованого реферування наперед треба вибрати тему, до якої належатиме текстовий документ, що опрацьовується, щоб система використовувала правильні ваги [13].

Розглянемо визначення ваг понять для ознакових задач на основі інтелектуального аналізу даних, а саме на основі побудови дерева рішень (ДР). Як відомо ознакові задачі дозволяють для пошуку релевантних прецедентів будувати ДР [14]. Однак, ДР не є панацеєю, оскільки згадувані ознаки, що лежать на відповідній гілці, що задає прецедент, не гарантують врахування повної множини ознак, які необхідно врахувати для знаходження релевантного прецедента. Нами пропонується використовувати ДР для визначення ваг базових термінів, які задають деякий прецедент, а потім на основі онтології ПО розвинути отримані ваги на всю онтологію для відповідного прецедента. Тоді для пошуку релевантного прецедента слід використовувати значення тих n понять, які для відповідного прецедента мають найбільші ваги.

Розглянемо гілку дерева рішень. Вершини (ознаки) цієї гілки знаходяться на k рівнях. Вочевидь, що чим вищий рівень, тим значуща ознака, яка на цьому рівні знаходиться. Ця евристична думка має бути відображена в значеннях ваг цих ознак. Крім того, пропонується ці ваги нормувати, тобто щоб їх сума для кожного прецедента (гілки) була рівна 1.

Розглянемо два способи визначення ваг базових ознак, які задовольняють вищі описані два припущення.

1 спосіб. Арифметичні ваги. Визначаються як відношення різниці $(k+1)$ рівня дерева та рівня, на якому знаходиться ознака до суми всіх рівнів гілки, тобто базуються на сумі арифметичної прогресії:

$$w_i = \frac{k+1-i}{\sum_{j=1}^k j} = \frac{k+1-i}{\frac{(1+k)k}{2}}.$$

2 спосіб. Геометричні ваги. Базуються на сумі геометричної прогресії:

$$w_i = \frac{2^{k-i}}{2^k - 1}.$$

Отримані на основі ДР ваги назвемо вагами базових ознак прецедента і позначимо таку множину ваг W_B . Тепер необхідно їх розвинути на всю онтологію ПО, використовуючи таксономію понять онтології, відношення між поняттями та їх інтерпретацію. Математично (формально) цей процес запишемо у вигляді:

$$W_B \xrightarrow{0} W. \quad (15)$$

Розмноження ваг на всю онтологію залежить від визначення (аксіоматизації) класів, їх ієрархії (вертикальний зв'язок) та горизонтальних зв'язків. У роботі [15] типам зв'язків присвоєні такі коефіцієнти важливості: для зв'язку типу спеціалізації («IS-A») – $\sigma = 0,9$; для узагальнення («kind-of») – $\gamma = 0,4$; для причинного зв'язку («caused-by») – $\rho_{CBY} = 0,3$; для характеризуючого зв'язку («characterized-by») – $\rho_{WRT} = 0,2$. Саме такі значення ми використовували під час розроблення вище описаних прикладних ІСППР.

Отже, загалом процес функціонування ІСППР на основі онтологій складається із кроків, наведених на рис. 7. Для його зображення використано діаграму діяльності (UML Activity) [16].

Для розроблення ІСППР було обрано такі програмні засоби (ПЗ):

– Для побудови онтології – редактор онтологій Protégé ;

– Для запису правил бази знань – SWRL, який входить в Protégé ;

– База даних – СКБД MySQL;

– Модуль керування процесом розв'язування задачі – мова програмування PHP.

В залежності від типу задачі в базі даних зберігаються: перелік прецедентів та відповідних до них рішень, ваги важливості понять адаптивної онтології, типи відношень та їх ваги, значення ознак, історія значень ознак (для задач планування).

ВИСНОВКИ

Розроблено математичні моделі функціонування інтелектуальних систем підтримки прийняття рішень в залежності від класу задачі. Всі ці моделі використовують метрику для знаходження релевантних прецедентів або визначення релевантності станів. Для побудови таких метрик використовуються онтології. З цією метою у загальноприйнятий трьохелементний кортеж, який задає онтологію (множина понять, відношень та їх інтерпретація), авторами додано дві скалярні величини (важливість понять та відношень), які використовуються для обчислення необхідних відстаней. Розглянуто способи завдання початкових коефіцієнтів важливості понять та зв'язків, зокрема на основі інтелектуального аналізу даних.

Розглянуті чотири класи задач. Визначено перелік задач, які входять до цих класів. Наведено приклади функціонування окремих інтелектуальних систем, які базуються на розроблених моделях.

СПИСОК ЛІТЕРАТУРИ

1. Каменнова, М. С. Корпоративные информационные системы: технологии и решения / М.С.Каменнова // Системы Управления Базами Данных. – 1995. – № 3. – С. 88–99.

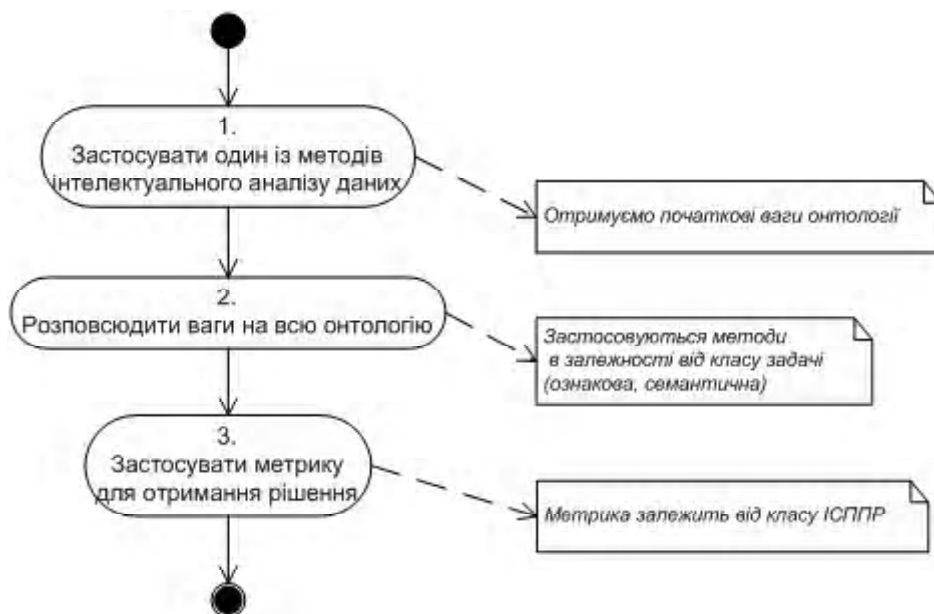


Рис. 7. Діаграма діяльності ІСППР на основі онтологій

2. *Funk, P.* Advances in Case-Based Reasoning / P. Funk, P. A. Gonzalez-Calero // 7th European Conference, ECCBR 2004. – Madrid, Spain. – P. 375–380.
3. *Рассел, С.* Искусственный интеллект / С. Рассел, П. Норвиг. – М. : СПб. ; К. : Вильямс, 2006. – 1408 с.
4. *Gruber, T. R.* A translation approach to portable ontologies / T.R.Gruber // Knowledge Acquisition. – 1993. – N 5 (2). – P. 199–220.
5. Інтелектуальні системи, базовані на онтологіях // Д. Г. Досин, В. В. Литвин, Ю. В. Нікольський, В. В. Пасічник. – Львів : Цивілізація, 2009. – 414 с.
6. *Литвин, В. В.* Мультиагентні системи підтримки прийняття рішень, що базуються на прецедентах та використовують адаптивні онтології / В. В. Литвин // Радіоелектроніка, інформатика, управління. – 2009. – № 2 (21). – С. 120–126.
7. *Sowa, J.* Conceptual graphs for a database interface / J.Sowa // IBM Journal of Research and Development. – Vol. 20. – № 4. – 1976. – P. 336–357.
8. *Свами, М.* Графы, сети и алгоритмы / М. Свами, К. Тхула-сираман. – М. : Наука, 1984. – 408 с.
9. *Литвин, В. В.* Проектування інтелектуальних агентів прийняття рішень в просторі ознак з використанням онтологічного підходу / В. В. Литвин, Р. Р. Даревич, Д. Г. Досин, Н. В. Шкутяк // Штучний інтелект. – 2010. – № 3. – С. 254–262.
10. *Скобелев, В. Г.* Локальные алгоритмы на графах / В. Г.Скобелев. – Донецк : ИПММ НАН Украины, 2003. – 217 с.
11. *Крайовський, В. Я.* Основні підходи до розроблення програмного комплексу автоматичного реферування текстових документів / В. Я. Крайовський В. В. Литвин, Н. Б. Шаховська // Інститут проблем моделювання в енергетиці. – 2009. – Випуск 51. – С. 178–186.
12. *Даревич, Р. Р.* Оцінка подібності текстових документів на основі визначення інформаційної ваги елементів бази знань / Р. Р. Даревич, Д. Г. Досин, В. В. Литвин, З. Т. Назарчук // Штучний інтелект. – 2006. – № 3. – С. 500–509.
13. *Литвин, В. В.* Метод автоматизованого реферування текстових документів з використанням онтологій / В. В. Литвин, В. А. Гайдін, О. Ю. Пшеничний // Складні системи і процеси. – Запоріжжя. – 2009. – № 1. – С. 81–87.
14. *Цветков, А. М.* Разработка алгоритмов индуктивного вывода с использованием деревьев решений / А. М. Цветков // Кибернетика и системный анализ. – № 1. – 1993. – С. 174–178.
15. *Bulskov, H.* On Querying Ontologies and Databases / H. Bulskov, R. Knappe, R. Andreasen // FQAS. – 2004. – P. 191–202.
16. *Фаулер, М.* UML в кратком изложении / М. Фаулер, К. Скотт. – М. : Мир, 1999. – 340 с.

Стаття надійшла до редакції 02.12.2010.

Литвин В. В.

МОДЕЛИРОВАНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ С ИСПОЛЬЗОВАНИЕМ ОНТОЛОГИЧЕСКОГО ПОДХОДА

Исследовано функционирование интеллектуальных систем поддержки принятия решений, ядром баз знаний которых являются онтологии. Рассматриваются интеллектуальные агенты четырех типов. Для каждого типа разработана метрика, которая используется для определения релевантности предлагаемых системой решений.

Ключевые слова: интеллектуальная система поддержки принятия решений, интеллектуальный агент.

Lytvyn V. V.

INTELLIGENT DECISION SUPPORT SYSTEMS MODELING USING ONTOLOGICAL APPROACH

The functioning of intelligent decision support systems with knowledge base based on ontology is investigated. Four types of intelligent agents are considered. Metrics for each of them is developed and used to determine the relevance of solutions proposed by the system.

Keywords: Intelligent Decision Support System, Intelligent Agent.

УДК 004.3

Баркалов А. А.¹, Зеленева И. Я.², Цололо С. А.³, Биайрак Х.⁴

¹ Д-р техн. наук, профессор Донецкого национального технического университета

^{2,3} Канд. техн. наук, доцент Донецкого национального технического университета

⁴ Аспирант Донецкого национального технического университета

УМЕНЬШЕНИЕ ПЛОЩАДИ МАТРИЧНОЙ СХЕМЫ УСТРОЙСТВА УПРАВЛЕНИЯ С РАЗДЕЛЕНИЕМ КОДОВ

В статье предложена модель композиционного микропрограммного устройства управления с разделением кодов, ориентированная на реализацию логической схемы устройства в базе заказных матриц. В модели используется представление адреса вершины алгоритма управления в виде конкатенации кодов ОЛЦ и кода компоненты ОЛЦ. Этот подход позволяет уменьшить число входов и выходов схемы формирования функций возбуждения.

Ключевые слова: композиционное устройство управления, матричная схема, операторная линейная цепь (олц), разделение кодов.

ВВЕДЕНИЕ

При реализации схемы устройства управления (УУ) необходимо учитывать особенности элементного бази-

са и алгоритмы управления [1]. Если алгоритм управления представлен линейной граф-схемой алгоритма (ГСА), то для его реализации целесообразно использо-

вать модель композиционного микропрограммного устройства управления (КМУУ) [2]. В настоящей работе рассматривается задача реализации схемы КМУУ в базе заказных матричных схем [3], которые широко используются при массовом производстве средств вычислительной техники и автоматики. В этом случае возникает задача уменьшения площади кристалла, занимаемого схемой КМУУ [4; 5]. Решение этой задачи позволяет уменьшить потребляемую мощность и повысить выпуск годных кристаллов. При этом рассматривается случай выполнения условий, позволяющих использовать модель КМУУ с разделением кодов [2].

Целью исследования является уменьшение аппаратных затрат в схеме КМУУ с разделением кодов за счет использования нескольких источников классов псевдоэквивалентных ОЛЦ.

Задачей исследования является разработка метода, позволяющего уменьшить площадь, занимаемую схемой адресации микрокоманд в КМУУ с разделением кодов.

РЕАЛИЗАЦИЯ КМУУ С РАЗДЕЛЕНИЕМ КОДОВ НА ЗАКАЗНЫХ МАТРИЦАХ

Пусть ГСА Γ представлена множествами вершин V и дуг E , соединяющих эти вершины. При этом $V = \{b_0, b_E\} \cup E_1 \cup E_2$, где b_0 – начальная вершина ГСА, b_E – конечная вершина ГСА, E_1 – множество операторных вершин, где $|E_1| = M$, E_2 – множество условных вершин. В вершинах $b_q \in E_1$ записаны наборы микроопераций $Y(b_q) \subseteq Y$, где $Y = \{y_1, \dots, y_N\}$ – множество микроопераций. В вершинах $b_q \in E_2$ записаны элементы множества логических условий $X = \{x_1, \dots, x_L\}$. Пусть ГСА является линейной, то есть включает более 75 % операторных вершин [3].

Сформируем множество ОЛЦ $C = \{\alpha_1, \dots, \alpha_G\}$ ГСА Γ , где каждая из ОЛЦ является последовательностью операторных вершин и каждой паре ее соседних компонент b_i, b_j соответствует дуга $\langle b_i, b_j \rangle \in E$. Каждая ОЛЦ имеет только один выход O_g и произвольное число входов I_g ($g = 1, \dots, G$). Формальные определения ОЛЦ, их входов и выходов можно найти в [3]. Каждая вершина

$b_q \in E_1$ соответствует микрокоманде MI_q , хранимой в управляющей памяти (УП) КМУУ по адресу $A(b_q)$. Для адресации микрокоманд достаточно

$$R = \lceil \log_2 M \rceil \quad (1)$$

бит, представленных переменными $T_r \in T$, где $|T| = R$.

Пусть ОЛЦ $\alpha_g \in C$ включает F_g компонент и пусть $Q = \max(F_1, \dots, F_G)$. В этом случае для кодирования компонент достаточно

$$R_C = \lceil \log_2 Q \rceil \quad (2)$$

переменных, образующих множество τ , где $|\tau| = R_C$. Для кодирования ОЛЦ $\alpha_g \in C$ достаточно

$$R_C = \lceil \log_2 G \rceil \quad (3)$$

переменных, образующих множество T , где $|T| = R_0$.

Пусть $K(\alpha_g)$, $K(b_q)$ соответственно код ОЛЦ $\alpha_g \in C$ и код компоненты некоторой ОЛЦ. Тогда адрес микрокоманды, соответствующей вершине $b_q \in E_1$, может быть представлен в виде конкатенации

$$A(b_q) = K(\alpha_g) * K(b_q). \quad (4)$$

В выражении (4) b_q является компонентой ОЛЦ $\alpha_g \in C$, а операция конкатенации обозначается знаком $*$. Выполним адресацию компонент так, чтобы для каждой ОЛЦ $\alpha_g \in C$ их коды возрастали в естественном порядке. При этом первая компонента любой ОЛЦ имеет код 0, вторая – 1, и так далее до Q .

В этом случае для реализации схемы УУ может быть использована модель КМУУ с разделением кодов (рис. 1).

Назовем эту модель КМУУ U_1 . В КМУУ U_1 матрицы M_1 и M_2 образуют блок адресации микрокоманд (БАМ), а матрицы M_3 и M_4 – управляющую память (УП). Блок БАМ реализует системы функций возбуждения триггеров счетчика СТ и регистра RG:

$$\Phi = \Phi(X, \tau), \quad (5)$$

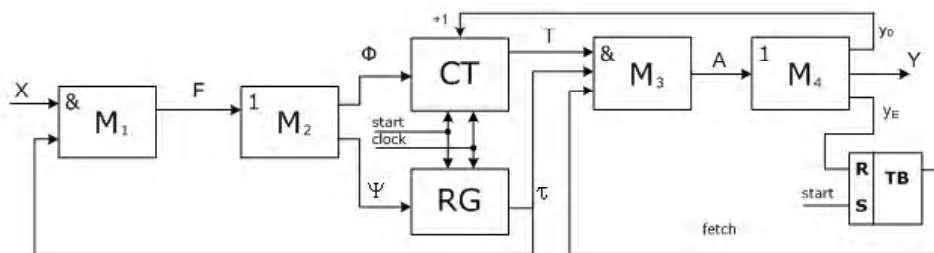


Рис. 1. Матричная реализация КМУУ с разделением кодов

$$\Psi = \Psi(X, \tau). \quad (6)$$

При этом матрица M_1 реализует систему термов $F = \{F_1, \dots, F_H\}$, входящих в функции (5)–(6). Матрица M_2 реализует функции (5)–(6), как дизъюнкции термов $F_h \in F$:

$$\left. \begin{aligned} \Phi_r &= \bigvee_{h=1}^H C_{rh} F_h (r = \overline{1, R_0}); \\ \Psi_r &= \bigvee_{h=1}^H C_{rh} F_h (r = \overline{1, R_C}). \end{aligned} \right\} \quad (7)$$

В функциях (7) $C_{rh} \in \{0, 1\}$ и $C_{rh} = 1$, если и только если терм F_h входит в функцию Φ_r или Ψ_r .

Матрица M_3 реализует систему функций A , соответствующих адресам микрокоманд (4):

$$A(b_q) = \left(\bigwedge_{r=1}^{R_0} \tau_r^{l_{gr}} \right) \cdot \left(\bigwedge_{r=1}^{R_C} T_r^{l_{qr}} \right). \quad (8)$$

В формуле (8) $l_{gr} \in \{0, 1\}$ – значение r -го разряда кода $K(\alpha_g)$, где $\alpha_g \in C$ и b_q входит в эту ОЛЦ; $l_{qr} \in \{0, 1\}$ – значение r -го разряда кода $K(b_q)$; $\tau_r^0 = \overline{\tau_r}$, $\tau_r^1 = \tau_r (r = \overline{1, R_0})$, $\tau_r^0 = \overline{\tau_r}$, $\tau_r^1 = \tau_r (r = \overline{1, R_C})$. Отметим, что некоторые разряды этих кодов могут быть несущественными, однако этот случай в данном случае не рассматривается. Для формирования терма $A_m \in A(m = \overline{1, M})$ соответствующий терм $A(b_q)$ умножается на переменную Fetch. При Fetch=1 разрешается выборка микрокоманд из УП.

Матрица M_4 реализует систему функций

$$y_n = \bigvee_{m=1}^M C_{nm} A_m, \quad (9)$$

где $n \in \{0, 1, \dots, N, E\}$, а $C_{nm} = 1$, если и только если функция y_n формируется в микрокоманде с индексом m . Переменная y_0 служит для увеличения содержимого СТ на единицу, что позволяет адресовать компонент одной ОЛЦ без использования блока БАМ. Переменная y_E формируется при достижении конечной вершины ГСА Γ и служит признаком окончания алгоритма. Если $y_E = 1$, то триггер TF обнуляется, переменная Fetch=1 и выборка микрокоманд прекращается.

Как видно из рис. 1, сигнал Start служит для установки начальных значений (Fetch=1, CT=RG=0). Сигнал Clock используется для переключения элементов памяти (СТ и RG).

Очевидно, КМУУ U_1 является автоматом Мура, состояния которого представлены функциями $\tau_r \in \tau$. Не-

достатком этого устройства является значительное число термов $F_h \in F$, реализуемых матрицей M_1 . Для уменьшения этого параметра можно ввести преобразователь кодов (ПК) [6]. Этот подход позволяет уменьшить число термов в системе (5)–(6) до некоторой величины H_0 , определяемой характеристиками эквивалентного автомата Мили. Однако ПК потребляет некоторые ресурсы кристалла. В настоящей статье предлагается подход, позволяющий гарантированно уменьшить параметр H до H_0 и уменьшить площадь, занимаемую схемой ПК.

ОСНОВНАЯ ИДЕЯ ПРЕДЛАГАЕМОГО МЕТОДА

Напомним, что ОЛЦ $\alpha_i, \alpha_j \in C$ являются псевдоэквивалентными (ПОЛЦ), если их выходы связаны с входом одной и той же вершиной ГСА Γ [2]. Пусть $\Pi_C = \{B_1, \dots, B_I\}$ разбиение множества ОЛЦ C_1 на классы ПОЛЦ. При этом ОЛЦ $\alpha_g \in C_1$, если ее выход не связан с вершиной b_E . Выполним кодирование ОЛЦ $\alpha_g \in C$ так, чтобы максимально возможное число классов $B_i \in \Pi_C$ входило в один обобщенный интервал R_0 -мерного булева пространства.

Теперь множество Π_C может быть представлено в виде $\Pi_C^1 \cup \Pi_C^2$. Пусть $B_i \in \Pi_C^1$, если класс $B_i \in \Pi_C$ представляется одним обобщенным интервалом R_0 -мерного булева пространства. В противном случае $B_i \in \Pi_C^2$. Очевидно, $\Pi_C^1 \cap \Pi_C^2 = \emptyset$ и $\Pi_C^1 \cup \Pi_C^2 = \Pi_C$. Закодируем классы $B_i \in \Pi_C^2$ двоичными кодами $K(B_i)$ разрядности

$$R_2 = \lceil \log_2 G_2 \rceil, \quad (10)$$

где $G_2 = |\Pi_C^2| + 1$. Используем для кодирования классов $B_i \in \Pi_C^2$ переменные $z_r \in Z$, где $|Z| = R_2$.

Исходная ГСА Γ служит для нахождения системы обобщенных формул переходов [2]. Разделим эту систему S на две подсистемы: $S = S_1 \cup S_2$. Пусть подсистема S_1 задает переходы для классов $B_i \in \Pi_C^1$, а подсистема S_2 – для классов $B_i \in \Pi_C^2$. В этом случае для реализации схемы КМУУ на заказных матрицах предлагается модель U_2 (рис. 2).

В КМУУ U_2 блок БАМ представлен матрицами M_1^1 , M_1^2 и M_2 . Матрица M_1^1 реализует термы $F_h \in F^1$, входящие в подсистему формул перехода S_1 . Термы $F_h \in F^1$ задаются формулами:

$$F_h = \left(\bigwedge_{r=1}^{R_0} \tau_r^{l_{gr}} \right) \cdot X_h (h = \overline{1, H_1}). \quad (11)$$

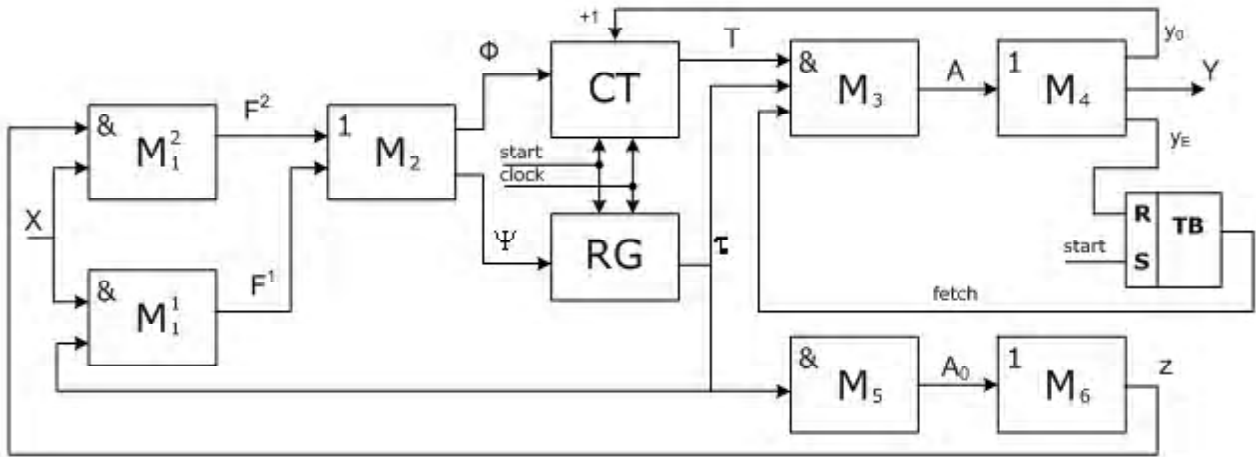


Рис. 2. Структурная схема КМУУ

В формуле (11) конъюнкция X_h соответствует части формул переходов, содержащей логические условия $x_l \in X$. Отметим, что $l_{gr} \in \{0, 1, *\}$ и τ_r^* . Это связано с тем, что классы $B_i \in \Pi_C^1$ представляются обобщенными интервалами, то есть некоторые кодирующие переменные имеют неопределенные значения.

Матрица M_1^2 реализует термы $F_h \in F^2$, входящие в подсистему формул S_2 . Термы $F_h \in F^2$ задаются формулами:

$$F_h = \left(\bigwedge_{r=1}^{R_2} z_r^{l_{ir}} \right) \cdot X_h \left(h = \overline{1, H_2} \right). \quad (12)$$

В формуле (12) $l_{ir} \in \{0, 1, *\}$ – значение r -го разряда кода $K(B_i)$ класса $B_i \in \Pi_C^2$. При этом неопределенное значение $*$ может появиться, если число возможных кодов больше числа классов $B_i \in \Pi_C^1$. Очевидно, $z_r^0 = \overline{z_r}$, $z_r^1 = z_r$ и $z_r^* = 1$ ($r = \overline{1, R_2}$). Кроме того, выполняется равенство $H_0 = H_1 + H_2$.

Матрица M_2 реализует функции (7), однако теперь они зависят от термов $F_h \in F^1 \cup F^2$. Матрицы M_3 и M_4 реализуют управляющую память; они полностью идентичны соответствующим матрицам КМУУ U_1 . Это же справедливо и для триггера TF. Матрицы M_5 и M_6 реализуют преобразователь кодов ПК. Матрица M_5 реализует термы $A_i \in A_0$, входящие в систему функций

$$Z = Z(\tau). \quad (13)$$

Термы A_i определяются выражением

$$A_i = \bigwedge_{r=1}^{R_0} \tau_r^{l_{ir}} \left(i = \overline{1, M_0} \right), \quad (14)$$

где $l_{ir} \in \{0, 1, *\}$ – значение r -го разряда кода класса $B_i \in \Pi_C^2$, $\tau_r^0 = \overline{\tau_r}$, $\tau_r^1 = \tau_r$ и $\tau_r^* = 1$ ($r = \overline{1, R_0}$). Параметр M_0 определяют в ходе синтеза КМУУ U_2 ; он зависит от результат кодирования ОЛЦ $\alpha_g \in C$. Матрица M_6 формирует функции (13), как некоторые дизъюнкции от термов (14):

$$z_r = \bigvee_{i=1}^{M_0} C_{ri} A_i \left(r = \overline{1, R_0} \right). \quad (15)$$

В формуле (15) булева переменная $C_{ri} = 1$, если и только если функция z_r зависит от терма A_i .

Принцип функционирования КМУУ U_2 очевиден. Разница между КМУУ U_1 и U_2 заключается в следующем:

1. Используются два источника кодов классов псевдоэквивалентных ОЛЦ. Один источник – регистр RG, второй – блок преобразователя кодов ОЛЦ в коды классов ПОЛЦ.

2. Число термов в схеме БАМ гарантировано равняется этому параметру эквивалентного автомата Мили. При этом площадь, занимаемая блоком ПК, уменьшается по сравнению с подходом, основанным на преобразовании кодов всех ОЛЦ $\alpha_g \in C_1$.

Отметим, что КМУУ с разделением кодов реализует-ся только при выполнении условия

$$R_0 + R_C = R. \quad (16)$$

МЕТОД СИНТЕЗА И РАЗВИТИЕ ОСНОВНОЙ ИДЕИ

В настоящей работе предлагается метод синтеза КМУУ U_2 , включающий следующие этапы:

1. Формирование множества ОЛЦ C и C_1 по исходной граф-схеме алгоритма Г.

2. Определение параметров R , R_0 и R_C . Если условие (16) выполняется, то процесс синтеза продолжается.

3. Формирование разбиения Π_C множества ОЛЦ C_1 на классы псевдоэквивалентных ОЛЦ.

4. Кодирование ОЛЦ $\alpha_g \in C$ и их компонент.

5. Формирование содержимого управляющей памяти.

6. Формирование множеств Π_C^1 и Π_C^2 . Кодирование классов $B_i \in \Pi_C^2$.

7. Формирование системы обобщенных формул переходов S и разбиение ее на подсистемы S_1 и S_2 .

8. Формирование таблицы переходов для классов $B_i \in \Pi_C^1$ по системе S_1 . Построение системы функций $\Phi = \Phi(\tau, X)$ и системы $\Psi = \Psi(\tau, X)$.

9. Формирование таблицы переходов для классов $B_i \in \Pi_C^2$ по системе S_2 . Построение системы функций $\Phi = \Phi(Z, X)$ и системы $\Psi = \Psi(Z, X)$.

10. Построение таблицы преобразователя кодов и формирование системы функций $Z = Z(\tau)$.

11. Реализация схемы КМУУ на заказных матрицах по полученным системам функций.

Детализируем некоторые этапы синтеза, связанные с построением таблиц переходов. Пусть система S_1 включает формулу перехода:

$$B_3 \rightarrow x_1 x_2 b_3 \vee x_1 x_2 b_8 \vee x_1 b_{12}. \quad (17)$$

Пусть $K(B_3) = 011$, $A(B_3) = 10010$, $A(B_8) = 11001$ и $A(B_{12}) = 11100$. Очевидно, $R_0 = 2$, $T = \{T_1, T_2\}$, $R_C = 3$, $\tau = \{\tau_1, \tau_2, \tau_3\}$. Таблица переходов для классов $B_i \in \Pi_C^1$ имеет следующие столбцы: B_i – класс ПОЛЦ; $K(B_i)$ – код класса $B_i \in \Pi_C^1$; $A(b_q)$ – адрес входа некоторой ОЛЦ $\alpha_g \in C$; X_h – входной сигнал, определяющий переход из входов ОЛЦ $\alpha_g \in B_i$ в вершину b_q ; Ψ_h – функция возбуждения триггеров регистра RG, принимающие единичное значение на переходе $\langle B_i, b_q \rangle$; Φ_h – функция возбуждения триггеров счетчика СТ, принимающие единичное значение на переходе $\langle B_i, b_q \rangle$; $h = \overline{1}, H_1$ – номер перехода. Для формулы (17) этот фрагмент таблицы имеет следующий вид (табл. 1).

Как следует из табл. 1, $\Psi = \{D_1, D_2, D_3\}$ и $\Phi = \{D_4, D_5\}$. Из табл. 1 имеем, например, следующие функции:

$$\begin{aligned} D_2 &= \overline{\tau_1} \tau_2 \tau_3 x_1 \overline{x_2} \vee \overline{\tau_1} \tau_2 \tau_3 x_1; \\ D_4 &= \overline{\tau_1} \tau_2 \tau_3 x_1 x_2. \end{aligned} \quad (18)$$

Термы системы (18) реализуются на матрице M_1^1 , а сами функции – на матрице M_2 .

Пусть система S_2 включает формулу перехода

$$B_5 \rightarrow x_2 b_{19} \vee x_3 b_{26}. \quad (19)$$

Пусть $K(B_5) = 01$, $A(B_{19}) = 00101$, $A(B_{26}) = 01100$, то есть $R_2 = 2$, $Z = \{z_1, z_2\}$. Таблица переходов для системы S_2 имеет такой же вид, как и таблица переходов для системы S_1 (табл. 2).

Из табл. 2 можно, например, получить функции

$$D_2 = \overline{z_1} z_2 \overline{x_3}; \quad D_5 = \overline{z_1} z_2 x_3. \quad (20)$$

Термы системы (20) реализуются на матрице M_1^2 , а сами функции – на матрице M_2 .

Пусть $\Pi_C^2 = \{B_2, B_5, B_6\}$, $K(B_2) = 01$, $K(B_5) = 10$ и $K(B_6) = 11$. При этом код 00 является признаком того, что переходы происходят из классов $B_i \in \Pi_C^1$. Этому факту соответствует переменная y_{C1} ; в данном случае $y_{C1} = \overline{z_1} z_2$. Пусть $B_2 = \{\alpha_2, \alpha_3\}$, $B_5 = \{\alpha_4, \alpha_5\}$ и $B_6 = \{\alpha_6, \alpha_7\}$, а ОЛЦ имеют коды, соответствующие двоичным эквивалентам их индексов. Таблица преобразователя кодов имеет следующие столбцы: α_g , $K(\alpha_g)$, B_i , $K(B_i)$, Z_i , i . Здесь столбец Z_i содержит перемен-

Таблица 1. Фрагмент таблицы переходов для формулы (17)

B_i	$K(B_i)$	$A(b_q)$	X_h	Ψ_h	Φ_h	h
B_3	011	10010	$x_1 x_2$	D_1	D_4	1
		11001	$x_1 \overline{x_2}$	$D_1 D_2$	D_5	2
		11100	$\overline{x_1}$	$D_1 D_2 D_3$	–	3

Таблица 2. Фрагмент таблицы переходов для формулы (19)

B_i	$K(B_i)$	$A(b_q)$	X_h	Ψ_h	Φ_h	h
B_5	01	00101	x_3	D_3	D_5	1
		01199	$\overline{x_3}$	$D_2 D_3$	–	2

ные z_r , равные единице в коде $K(B_i)$. Для нашего примера таблица ПК представлена в табл. 3.

Из табл. 3 с учетом минимизации имеем: $z_1 = \tau_1$, $z_2 = \tau_2$. При этом матрица M_5 в схеме КМУУ отсутствует. Однако, это частный случай. В общем случае система термов, полученных из таблицы ПК, реализуется на матрице M_5 , а функции $z_r \in Z$ – на матрице M_6 .

Отметим, что термы $F_h \in F^1$ могут зависеть только от логических условий $x_l \in X^1$, где $X^1 \subset X$. Аналогично, термы $F_h \in F^2$ могут зависеть только от логических условий $x_l \in X^2$, где $X^2 \subset X$. В лучшем случае имеем равенство $X^1 \cap X^2 = \emptyset$. Это позволяет уточнить часть схемы КМУУ U_2 , реализующей функции Ψ и Φ (рис. 3).

Отметим, что наличие переменной y_{c1} не вносит задержку во время такта КМУУ U_2 по сравнению с КМУУ U_1 . Это связано с тем, что значение переменной y_{c1} формируется до прихода правильных значений логических условий. Итак, площади матриц M_1^1 , M_1^2 и M_2 могут быть найдены следующим образом:

$$\begin{aligned} S(M_1^1) &= 2(L_1 + R_C + 1) \cdot H_1; \\ S(M_1^2) &= 2(L_2 + R_0) \cdot (H_2 + 1); \\ S(M_2) &= H_0(R_0 + R_C). \end{aligned} \quad (21)$$

Таблица 3. Таблица ПК КМУУ U_2

α_g	$K(\alpha_g)$	B_i	$K(B_i)$	Z_i	i
α_2	010	B_2	01	z_2	1
α_3	011	B_2	01	z_2	2
α_4	100	B_5	10	z_1	3
α_5	101	B_5	10	z_1	4
α_6	110	B_6	11	$z_1 z_2$	5
α_7	111	B_6	11	$z_1 z_2$	6

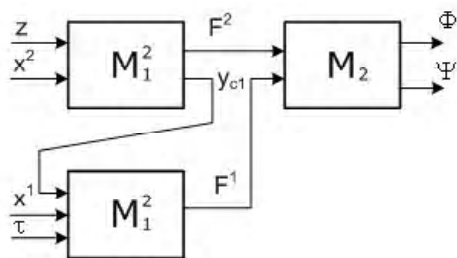


Рис. 3. Уточненная часть схемы КМУУ U_2

В системе (21) имеем $L_1 = |X^1|$, $L_2 = |X^2|$, $|F^1| = H_1$,

$|F^2| = H_2$, $H_1 + H_2 = H_0$.

ЗАКЛЮЧЕНИЕ

В работе предложен метод, позволяющий уменьшить сложность матричной реализации КМУУ с разделением кодов. Этот метод основан на использовании двух источников кодов классов псевдоэквивалентных ОЛЦ. При этом матрица термов схемы адресации КМУУ разбивается на две части.

Такой подход гарантирует уменьшение числа термов в схеме адресации до величины, определяемой эквивалентным автоматом Мили. Выбор между КМУУ U_1 и U_2 может быть выполнен аналитическим путем, то есть без реализации схемы устройства. Отметим, что предложенный метод применим только для линейных граф-схем алгоритма при выполнении условия (16). Если условие (16) не выполняется, то это приводит к резкому увеличению площадей матриц M_3 и M_4 по сравнению с КМУУ с общей памятью [2].

Научная новизна статьи заключается в усовершенствовании метода синтеза КМУУ с разделением кодов. Внесенные изменения позволяют использовать несколько источников кодов псевдоэквивалентных ОЛЦ, что приводит к уменьшению аппаратных затрат в схеме адресации.

Практическая значимость результатов работы заключается в уменьшении числа базовых элементов, которые необходимы для реализации схемы КМУУ на кристалле. Это позволяет удешевить реализацию сложных цифровых систем, в состав которых входят устройства управления, реализованный по модели КМУУ.

СПИСОК ЛИТЕРАТУРЫ

1. Baranov Samary. Logic Synthesis for Control Automata. / Samary Baranov. – Kluwer Academic Publishers, 1994. – 312 p.
2. Barkalov, A. Logic Synthesis for Compositional Microprogram Control Units. / A. Barkalov, L. Titarenko. – Berlin : Springer, 2008. – 273 p.
3. Smith, Michael. Application-Specific Integrated Circuits / Michael Smith. – Boston : Addison-Wesley, 1997. – 1040 p.
4. Баркалов, А. А. Синтез микропрограммных автоматов на заказных и программируемых СБИС. / А. А. Баркалов, Л. А. Титаренко. – Донецк : УНИТЕХ, 2009. – 336 с.
5. Bayrek Hathot Approach to Realization of Compositional Microprogramming Control Unit with Code Converter on Custom-Made Matrix. / A. Barkalov, I. Zelenyova, A. Miroshkin, H. Bayrek // Наукові праці Донецького національного технічного університету. Серія «Інформатика, кібернетика та обчислювальна техніка». Випуск 11(164). – Донецьк : ДонНТУ, 2010. – С. 71–74.
6. Barkalov, A. Basic principles of logic design. / A. Barkalov, L. Titarenko. – Zielona Gora : University of Zielona Gora Press, 2010. – 295 p.

Стаття надійшла до редакції 03.05.2011.

Баркалов О. О., Зеленьова І. Я., Цололо С. О., Біайрак Х.

ЗМЕНШЕННЯ ПЛОЩІ МАТРИЧНОЇ СХЕМИ ПРИБОРУ КЕРУВАННЯ З РОЗДІЛЕННЯМ КОДІВ

У статті запропонована модель композиційного мікропрограминого пристрою керування з розділенням кодів, яка орієнтована на реалізацію схеми пристрою у базисі замовних матриць. Запропонована модель використовує представлення адреси вершини алгоритму керування у вигляді конкатенації кодів ОЛЛ та коду компоненти ОЛЛ. Такий підхід зменшує число входів і виходів схеми формування функцій збудження.

Ключові слова: композиційний пристрій керування, матрична схема, операційний лінійний ланцюг (олл), розділення кодів.

Barkalov A. A., Zelenyova I. J., Tsololo S. A., Biayarek H.

REDUCTION OF TERM MATRIX OF CONTROL UNIT WITH CODE SHARING

The structures of compositional microprogram control unit with code sharing are proposed. Structures allow reducing the complexity of the matrix realization in the device's circuit. The proposed method is based on using of node's address representation as a concatenation of OLC codes and code of OLC components. The proposed method allow reducing the complexity of the matrix realization in the device's circuit.

Key words: compositional control unit, matrix circuit, operator linear chain (olc), code sharing.

УДК 378.14:004.421

Пищукина О. А.¹, Ключок А. Ю.²¹Канд. техн. наук, доцент Национального аэрокосмического университета им. М. Е. Жуковского «ХАИ»²Ассистент Национального аэрокосмического университета им. М. Е. Жуковского «ХАИ»

ПОДХОД К ФОРМИРОВАНИЮ ОБРАТНОЙ СВЯЗИ В ИНТЕЛЛЕКТУАЛЬНЫХ ОБУЧАЮЩИХ СИСТЕМАХ В СФЕРЕ ВЫСШЕГО ТЕХНИЧЕСКОГО ОБРАЗОВАНИЯ

Предложен подход к формированию обратной связи процесса обучения с использованием интеллектуальных компьютерных систем, особенностью которого является поэтапный контроль процесса обучения, определение места возникновения ошибок, в отдельных случаях – распознавание причин их возникновения и формирование рекомендаций по их устранению.

Ключевые слова: интеллектуальные компьютерные обучающие системы, высшее техническое образование, обратная связь.

ВВЕДЕНИЕ

Сокращение часов аудиторной нагрузки и увеличение объемов самостоятельной работы студентов представляет собой одну из особенностей процесса обучения в современном вузе, ориентированном на западные стандарты образования. Основную сложность для студентов технических специальностей вызывает усвоение постоянно растущего объема информации и расширения круга задач, умение решать которые необходимо для обеспечения компетентности и конкурентоспособности выпускников технических вузов на современном рынке труда. Вместе с тем, следует учитывать ментальные отличия украинских студентов, связанные со слабой мотивацией и (или) неумением работать самостоятельно, что способствует недостаточному усвоению учебного материала, и как следствие, снижению общего образовательного уровня. Необходимо отметить также недостаточную подготовку учащихся средних школ по базовым техническим дисциплинам для эффективного обучения в высших учебных заведениях.

Эффективным путем разрешения данного противоречия, связанного с необходимостью активного обучения студентов, ограничением энергетических ресурсов преподавателя, фиксированным количеством часов практических и лабораторных занятий согласно разработанным учебным и рабочим программам вузов, является использование компьютерных средств для организации самостоятельной работы студентов, а именно – интеллектуальных обучающих систем [1, 2].

Характерной чертой обучения студентов технических специальностей является необходимость формирования у них конкретных навыков и умений, направленных на практическое решение примеров и задач, требующих конечного ответа, чаще всего в виде числа или математического выражения. Эта особенность накладывает специфические требования к структуре и наполнению интеллектуальной обучающей системы, связанные с многократным повторением и решением предложенных задач, и требует включения в ее состав следующих итерационных блоков [3, 4]: 1) информационного блока, содержащего конкретные рекомендации по способам

решения задач; 2) блока примеров, где студент частично участвует в процессе их решения; 3) блока контроля, предлагающего выбор уровня сложности тестовых примеров и задач, и реализующего оценивание студента в результате их решения; 4) блока оценивания и рекомендаций, реализующего обратную связь учебного процесса.

ПОСТАНОВКА ПРОБЛЕМЫ

В результате проведенного анализа существующих разработок компьютерных обучающих программ в сфере технического образования было выявлено, что именно формирование обратной связи представляет собой наибольшую трудность при разработке компьютерных систем обучения, т. к. в эту связь необходимо включить большое количество критериев оценивания знаний и диагностирования причин возникновения ошибки (чаще всего слабоформализуемых) [5, 6]. Вместе с тем предложенные критерии не всегда в полной мере отражают сущность и функции обратной связи, реализуемой преподавателем, а также в изученных разработках не учтены особенности и специфика обратной связи при изучении технических дисциплин (большое количество математических конструкций, сложность расчетов и т. д.). Таким образом, задача формирования обратной связи интеллектуальной обучающей системы, включающей в себя функции оценивания и рекомендации для исправления ошибок, и являющейся адекватной процессу обучения технической специальности в высшем учебном заведении представляет собой актуальную научно-практическую задачу.

РЕШЕНИЕ ПРОБЛЕМЫ

При изучении сущности обратной связи учебного процесса были выявлены ее основные составляющие, реализуемые преподавателем: функция оценивания, функция поиска места ошибки в случае неверного ответа студента и рекомендательная функция повторного изучения материала по результатам определяемого места ошибки. Две последние функции в интеллектуальной системе обучения были объединены в блок диагностирования ошибок (рис. 1).

Блок диагностирования выполняет обнаружение ошибки при решении практических задач по разветвляющемуся алгоритму: обнаружение факта ошибки при несовпадении результата решения с эталонным заданным значением, далее – обнаружение места ошибки на каком-либо этапе решения задачи, определение класса ошибки, т. е. ее принадлежность к лекционной теме или типу решаемых задач, затем – диагностирование «пробела» в знаниях и рекомендации по улучшению полученного результата, например, – изучить теоретический материал (с указанием тем или глав), повторить алгоритм решения предложенных задач, перейти к решению задач более простого уровня и т. д. (рис. 2.)

Для оценивания уровня знаний в интеллектуальной обучающей системе предлагается учитывать не только знание алгоритма решения и способность применить его к предлагаемой задаче, но также рассматривать все множество допущенных ошибок и степень их влияния на результат, с целью устранения субъективности закладываемых параметров оценки.

Множество ошибок, допускаемых студентами в процессе решения задач, предложено декомпозировать на следующие подмножества (признак классификации – уровень ошибки): комплексные или системные ошибки – подмножество $\{A\}$, грубые ошибки – подмножество $\{B\}$, исправимые ошибки – подмножество $\{C\}$. Каждое из этих подмножеств содержит определенные виды ошибок, выявленные в процессе педагогической деятельности [3], и каждому виду ошибки присвоен свой вес в зависимости от влияния этой ошибки на исход решения; значения весов сведены в матрицу P размерностью $(n \times 1)$.

Например, системные ошибки, связанные с непониманием метода или неверного применения способа решения имеют наибольший вес (1–2 балла), грубые ошибки, связанные, например, с неправильной подстановкой коэффициентов имеют вес 0,5–1 балла, исправимые ошибки, такие как ошибки в расчетах или ошибки из-за невнимательности имеют минимальный вес 0,1–0,2 балла. Для того, чтобы предоставить возможность обучаемому повысить оценку и самостоятельно устранить допущенные ошибки введены параметры, отражающие способ исправления ошибки, которым также присваи-

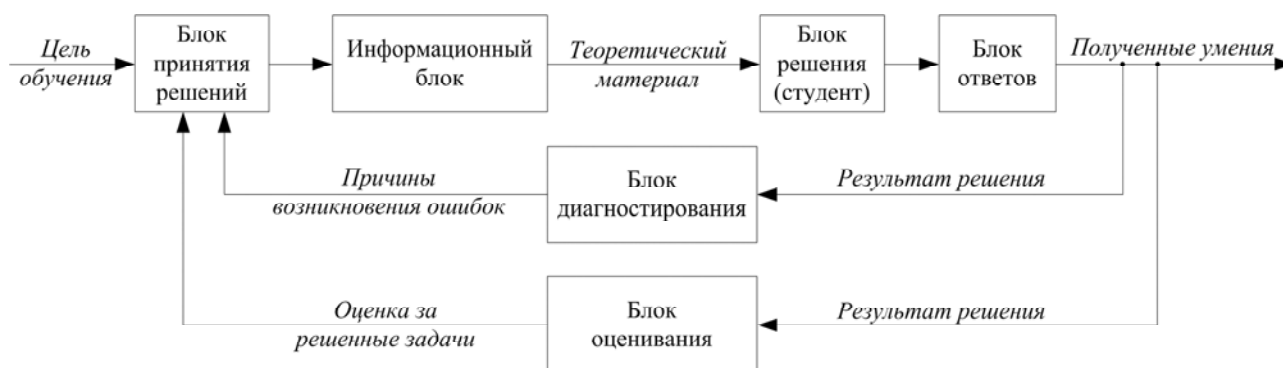


Рис. 1. Схема реализации прямых и обратных связей в интеллектуальной обучающей системе

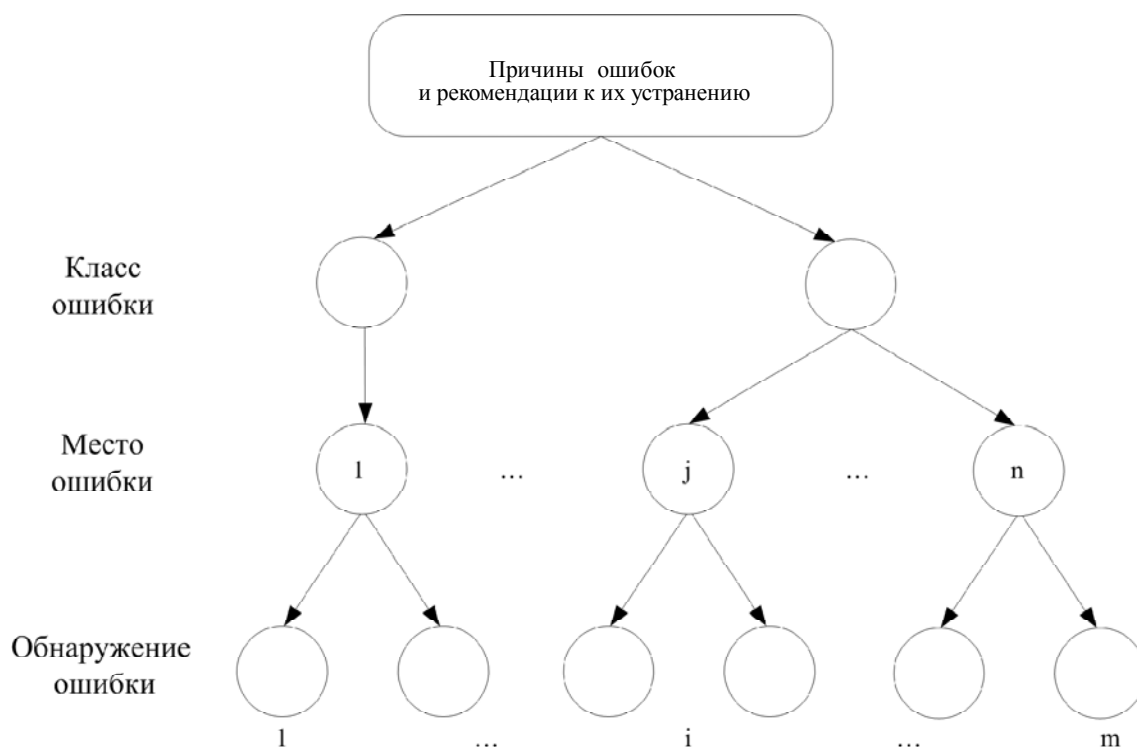


Рис. 2. Схема диагностирования возникновения ошибок

вается вес, имеющий отрицательное значение и уменьшающий баллы матрицы P ; значение весов исправления ошибки сведено в матрицу весов исправления ошибок Q размерностью $(n \times n)$. Среди способов исправления выделены 3 основные группы: самостоятельное исправление (после информации о том, что она допущена), устранение ошибки после рекомендации с указанием ее места, устранение ошибки интеллектуальной системой после нескольких неудачных попыток обучаемого. Для формализации оценивания предлагается использовать критерий максимизации разности между ожидаемым результатом (высший балл по соответствующей шкале) S и полученными штрафными баллами за допущенные ошибки R , а также показатель ошибок, определяющий степень влияния допущенных ошибок на результат решения H (сумма элементов матрицы H представляет собой штрафные баллы R):

$$K = (S - R) \rightarrow \max; H = P \times Q. \quad (1)$$

Предложенный подход к контролю и оцениванию уровня знаний, а также диагностированию и формированию множества рекомендаций при использовании компьютерных средств обучения находит применение на кафедре систем управления летательными аппаратами Национального аэрокосмического университета им. Н. Е. Жуковского «ХАИ» в рамках разработки компьютерных обучающих программ по основным читаемым курсам [7, 8, 9].

ЗАКЛЮЧЕНИЕ

Предлагаемый подход позволяет реализовать не только прямые связи процесса обучения в заданной предметной области с использованием компьютерных средств, а также его наиболее значимую часть – обратную связь, с помощью которой осуществляется поэтапный контроль процесса обучения, определяются места возникновения ошибок, в отдельных случаях распознаются причины их возникновения и формируются рекомендации по их устранению и улучшению результатов обучения.

Особенностью предложенного подхода является возможность учитывать не только количество ошибок при оценивании сформированных умений, а также степень их влияния на результат и качество помощи интеллектуальной системы, оказанной при исправлении допущенных ошибок, что более полно и точно отражает обратную связь учебного процесса, связанную с оцениванием обучающихся. Вместе с тем, реализация достаточно сложной процедуры диагностирования причин возникновения ошибок требует систематической доработки сформированных компьютерных модулей с целью повышения эффективности использования интеллектуальной обучающей системы.

СПИСОК ЛИТЕРАТУРЫ

1. Информационно-аналитические модели управления технически высшими учебными заведениями [текст] / А. Н. Гуржий, В. С. Кривцов, А. С. Кулик и др. ; Нац. аэрокосмич. ун.-т им. Н. Е. Жуковского «Харьковский авиационный институт». – Х. : ХАИ, 2004. – 386 с.

2. Башмаков, А. И. Разработка компьютерных учебников и обучающих систем / А. И. Башмаков, И. А. Башмаков. – М. : Информационно-издательский дом «Филинъ», 2003. – 616 с.
3. Пищукина, О. А. Информационная технология разработки компьютерных обучающих программ / О. А. Пищукина, Д. В. Бирюкова, О. В. Клименко // Радиоелектронні і комп'ютерні системи. – 2006. – №2 (14). – С. 57–62.
4. Пищукина, О. А. Интеллектуальные обучающие системы в сфере технического образования / Пищукина О. А. // Інформаційні комп'ютерні технології в машинобудуванні. – 2010. – С.118.
5. Пищукина, О. А. Особенности разработки компьютерных обучающих программ / О. А. Пищукина, Н. В. Нечипорук, Д. В. Бирюкова, О. В. Клименко // Актуальні проблеми сучасних наук: теорія та практика. – 2006» – Том 9. – Дніпропетровськ : Наука і освіта. – 2006. – С. 84–86.
6. Пищукина, О. А. Формирование подхода к контролю и оценке знаний и умений при разработке компьютерных обучающих программ / О. А. Пищукина, О. В. Клименко, Д. В. Бирюкова // Системи підтримки прийняття рішень. Теорія і практика – К. : ІПММС НАНУ. – 2007. – С.134–136.
7. Пищукина, О. О. Комп'ютерна програма «Навчальна програма розв'язання диференційних рівнянь операторним методом» / О. О. Пищукина, Д. В. Бирюкова, О. В. Клименко // Свід. про реєстрацію авторського права на твір № 17725. – зареєстр. в Держ. департ. інтелектуальної власності Мін. освіти і науки України; реєстр. 28.08.2006 р.
8. Пищукина, О. О. Комп'ютерна програма «Навчальна програма розв'язання диференційних рівнянь методом Ейлера» / О. О. Пищукина, Д. В. Бирюкова, О. В. Клименко // Свід. про реєстрацію авторського права на твір № 17651. – зареєстр. в Держ. департ. інтелектуальної власності Мін. освіти і науки України; реєстр. 15.08.2006 р.
9. Дергачев, К. Ю. Формирование комплекса интеллектуальных обучающих программ при решении навигационных задач / К. Ю. Дергачев, О. А. Пищукина, А. Ю. Клочок // Людина і космос. – Днепропетровск, 2011. – С. 211.

Стаття надійшла до редакції 26.05.2011.

Пищукина О. О., Клочок А. Ю.

ПІДХІД ДО ФОРМУВАННЯ ЗВОРОТНЬОГО ЗВ'ЯЗКУ В ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ, ЩО НАВЧАЮТЬ, У СФЕРІ ВИЩОЇ ТЕХНІЧНОЇ ОСВІТИ

Запропоновано підхід до формування зворотнього зв'язку процесу навчання з використанням інтелектуальних комп'ютерних систем, особливістю якого є поетапний контроль процесу навчання, визначення місця виникнення помилок, в окремих випадках – розпізнавання причин їх виникнення і формування рекомендацій з їх усунення.

Ключові слова: інтелектуальні комп'ютерні системи, що навчають, вища технічна освіта, зворотний зв'язок.

Pishchukhina O. A., Klochok A. Yu.

APPROACH TO THE FEEDBACK FORMING IN INTELLIGENT LEARNING SYSTEM IN THE SPHERE OF HIGHER TECHNICAL EDUCATION

The approach to the feedback of learning process forming while using intelligent computer system is offered. It's feature is step-by-step control of learning process, determining of mistakes origin point, in other case – causes of its occurrence and forming recommendations of its eliminating.

Key words: intelligent computer learning system, higher technical education, feedback.

УДК 004.932.4

Степаненко О. О.¹, Піза Д. М.²

¹Канд. техн. наук, доцент Запорізького національного технічного університету

²Д-р техн. наук, професор Запорізького національного технічного університету

ПРОГРАМНИЙ КОМПЛЕКС ДЛЯ АНАЛІЗУ Й ОБРОБКИ ЕХО-ІМПУЛЬСНИХ ЗОБРАЖЕНЬ

На основі аналізу методів обробки ехо-імпульсних зображень було виявлено недоліки існуючих підходів і запропоновано два нових методи. Розроблені методи і програмна система, що базується на них, дозволяють підвищити ефективність візуального аналізу медичних та сейсмічних ультразвукових зображень.

Ключові слова: математична модель, метод лінійного передбачення, суперпозиція імпульсного сигналу, спектр, ехо-імпульсне зображення, модуль математичних обчислень.

ВСТУП

За останні 40 років ультразвук став важливою діагностичною методикою. Його потенціал у відображенні медичної діагностики було визнано у 1930-х і 1940-х роках, коли Теодор і Фридріх Дуссікі спробували використовувати

ультразвук для того, щоби діагностувати пухлини мозку. Однак, лише у 1970-х, робота цих та інших піонерів досліджень ультразвуку реально принесла свої плоди.

Разом з технологічними вдосконаленнями, ультразвук прогресував від великої громіздкої машини, яка відтво-

рювала неоптимальні зображення, до переносного, зручного для використання і складного пристрою. Така еволюція потребувала тісного поєднання фізики, фізіології, медицини, техніки і керування [1].

Однак, навіть на сучасному етапі розвитку ультразвуку зображення, які отримуються за його допомогою, завжди неідеальні і містять шуми різного виду, які вносять істотні спотворення та ускладнюють процес прийняття рішень, заснований на їх візуальному аналізі. Це пояснюється як самою природою ультразвуку, так і недосконалістю апаратури [2]. Тому актуальною є задача розробки методів та алгоритмів покращення якості ультразвукових та ехо-імпульсних зображень та їх програмна реалізація, з метою покращення їх якості та спрощення їх візуального аналізу.

ПОСТАНОВКА ЗАДАЧІ

Якщо розглядати ультразвукове зображення як набір суперпозицій невідомих імпульсних сигналів, що відповідає методам отримання цих зображень, то задача їх обробки полягає в аналізі цих суперпозицій.

Існуючі методи аналізу висувають до сигналів суперпозиції обмеження, які не можуть бути виконані в рамках задач ультразвукової діагностики. Зокрема, метод інверсної фільтрації вимагає апіорного знання форми елементарного імпульсу, а метод кепстрального аналізу припускає, що всі імпульси мають однакову форму, причому висувається обмеження на їх число [3, 4].

Реалії практичних задач (особливо в сейсмології і медичній ультразвуковій діагностиці) змушують вести аналіз сигнальних суперпозицій, що включають сотні і тисячі елементарних імпульсів. Причому через ефекти дифракції, реверберації, розсіювання і загасання форми цих імпульсів не тільки невідомі, але й різні, що не дає змоги використати жоден із відомих методів для аналізу таких суперпозицій [5].

Отже, актуальною є розробка і програмна реалізація таких методів аналізу суперпозицій сигналів, які б не висували вимог до форми і кількості сигналів у суперпозиції.

ОСОБЛИВОСТІ РЕАЛІЗАЦІЇ РОЗРОБЛЕНИХ МЕТОДІВ ТА АЛГОРИТМІВ

У результаті вивчення існуючих методів аналізу суперпозицій імпульсних сигналів і виявлення їх недоліків було створено два нових методи обробки ехо-імпульсних зображень.

Запропоновані методи обробки ультразвукових зображень, базовані на застосуванні адаптованих до даної галузі методів лінійного передбачення, які переважно використовуються в задачах обробки голосу, та сигнального підпростору для аналізу суперпозицій імпульсних сигналів.

Метод аналізу суперпозицій імпульсних сигналів, заснований на ідеї методу параметричного спектрального аналізу, дозволяє підвищити його розподільну здатність у часовій області і достовірність виявлення меж слабо-контрастних ділянок шаруватих структур без використання будь-якої апіорної інформації щодо форми елемен-

тарних імпульсів. Метод є стійким до впливу вимірювальних шумів.

Запропонований алгоритм візуалізації результатів аналізу ехо-імпульсних зображень у рамках методу параметричного спектрального аналізу відноситься до області методів комп'ютерного бачення, оскільки візуалізуються полюси, а не реальні фізичні амплітуди, які відповідають моделі лінійного передбачення Фур'є-спектральних характеристик реєстрованих часових імпульсних суперпозицій. У межах імпульсної ультразвукової діагностики основна проблема використання методів параметричного спектрального аналізу пов'язана з використанням системи ЧАРП (часового автоматичного регулювання підсилення), що призводить до підйому амплітудно-часових характеристик, оскільки яскравість ультразвукових ехо-імпульсних зображень має бути позитивною.

Отже для реалізації методів спектрального аналізу другого порядку аналізована послідовність має бути знакозмінною. Для цих цілей доцільно використовувати перетворення Гільберта кожної аналізованої сигнальної послідовності, що забезпечує результат, близький до результату чисельного диференціювання, але гарантує стійкість до вимірювальних шумів (у порівнянні з методом чисельного диференціювання).

Додаткову обробку яскравості синтезованих зображень пропонується виконувати методами еквалізації гістограм або евристичного зрізування.

Модель суперпозиції сигналів, що перекриваються, за умови, що залежність еквівалентного шумового спектрального компонента близька до білого шуму, доцільно подати у вигляді

$$\hat{S}(f_k) = \sum_{m=1}^{M=2N} p(m)S(f_{k-m}), k = 1, 2, \dots, K, \quad (1)$$

де $p(m)$ – коефіцієнти лінійного прогнозування; M – порядок фільтра лінійного прогнозування; K – число частот, на яких вимірюється спектральна характеристика $S(f)$ (воно, як правило, збігається із числом часових відліків сигнальної послідовності $s(t)$ [6].

Задача підвищення роздільної здатності візуального аналізу суперпозицій сигналів, що перекриваються, вирішується на основі використання моделі синтезованого «відфільтрованого» ехо-імпульсного зображення:

$$\hat{s}(t) = \frac{\sigma^2}{\left[1 - \sum_{i=1}^M p_i e^{-j2\pi\Delta f t}\right]^2}, \quad (2)$$

де σ^2 розглядається як дисперсія помилки лінійного прогнозування для заданого порядку моделі M ; $\Delta f = f_k - f_{k-1}$ – величина, що характеризує дискретність спектральної характеристики $S(f)$, яка виникає через кінцеве значення часового інтервалу реєстрації суперпозиції $s(t)$.

Структура запропонованого алгоритму є наступною:

1. Перехід з часової області первинних вимірювань у спектральну область на основі використання алгоритму прямого перетворення Фур'є (спектральний аналіз I-го порядку).

2. Розрахунок моделі лінійного прогнозування для спектральної області. Порядок моделі M лінійного прогнозування визначає значуще число імпульсів, що відображаються і які будуть продетектовані на наступній стадії.

3. Зворотний перехід зі спектральної області в часову область на основі використання алгоритму нелінійного параметричного спектрального аналізу (спектральний аналіз II-го порядку) на основі синтезованої моделі (2) [6].

Схема алгоритму наведена на рис. 1.

Тобто, на вході системи присутня вхідна амплітудно-імпульсна послідовність (суперпозиція зашумлених сигналів, що перекриваються), на виході – відфільтрована «синтезована» часово-залежна послідовність, що вимірюється в відношеннях амплітуд.

Оскільки основним об'єктом досліджень були ультразвукові зображення, то вихідна інформація надавалася як зображення відповідного формату, що зчитувались у комп'ютер у вигляді чисельної матриці. При цьому описаний алгоритм застосовувався для кожного стовпця аналізованого зображення, тому якщо число стовпців дорівнювало 400, то алгоритм застосовувався 400 разів з подальшою візуалізацією результатів уже як для єдиного синтезованого зображення.

Метод обробки ультразвукових зображень, заснований на методі сигнального підпростору, направлений на нейтралізацію впливу багатократних перевідбиттів, є стійким до впливу вимірювальних і структурних шумів і не вимагає апріорного знання моделі аналізованої структури.

Він базується на використанні властивостей власних векторів кореляційної матриці C , спектральної залежності $S(f)$ сигнальної суперпозиції, що аналізується, і розкладанні ортогонального простору, утвореного власними векторами кореляційної матриці, на два ортогональних підпростори: сигнальний і шумовий. З математичної точки зору розв'язання задачі ґрунтується на пошуку такої нової системи координат, у якій інформативні гармоніки спектральної залежності $S(f)$ були б ортогональними в межах заданого частотного інтервалу F (обумовленого тривалістю інтервалу вимірів і частотою дискретизації сигнальної суперпозиції).

Матрицю C можна подати у вигляді суми двох кореляційних матриць, а перехід у часову область здійснюється

на основі моделі (відображаються полюси знаменника моделі (3)):

$$\hat{s}(t) = 1/s_i^T C_q s_i; \quad S_i^T = \begin{bmatrix} e^{-j2\pi f_1 t} & \dots & e^{-j2\pi f_N t} \end{bmatrix};$$

$$f \in F; \quad t_i = \text{var}, \quad (3)$$

де інтервал $\Delta t = t_i - t_{i-1}$ може бути обраний як завгодно малим [7].

ОСОБЛИВОСТІ РЕАЛІЗАЦІЇ ПРОГРАМНИХ КОМПОНЕНТІВ

Особливістю запропонованої системи є той факт, що методи, які вона реалізує, із мінімальними змінами можуть бути застосовані до різних практичних галузей (найбільш очевидними є галузі сейсмічної розвідки вуглеводнів і ультразвукової медичної діагностики).

Таким чином, одними з найкритичніших вимог до системи є її гнучкість і здатність масштабування. Виходячи саме з цих міркувань і була розроблена її структура.

Потреба у гнучкості системи диктувала необхідність використання для її розробки модульного підходу, причому кількість модулів не повинна була бути сталою величиною, до того ж, вони могли бути виконані з використанням різних технологій і мов програмування. Це дозволяло організувати введення до системи нових модулів (які, наприклад, могли містити модифікації методів або їх нові реалізації) без будь-яких ускладнень. Поєднання модульного підходу з об'єктно-орієнтованим підходом, окрім цього, також вирішує і проблему масштабованості системи [8].

Враховуючи те, що програмне забезпечення розроблялось для комп'ютерів на основі операційних систем сімейства Microsoft Windows, у якості платформи розробки ядра системи було вирішено обрати платформу .NET, а в якості основної мови програмування – C#.

Значним недоліком використання платформи .NET є низька швидкодія генерованого байт-коду, тому для розробки критичних для продуктивності модулів доцільно використовувати інші мови програмування. В якості такої додаткової мови програмування використовувалась мова Fortran. Це пов'язано з її націленістю на вирішення математичних задач, а також наявністю швидких компіляторів. Для отримання відповідних модулів системи використовувалась реалізація компілятора Intel Fortran Compiler 10, як одного із найкращих компіляторів Fortran у плані швидкодії кінцевих програм і бібліотек [9].

Вищесказане висувало також вимоги до середовища, у якому велася розробка системи. Вибір середовища, адекватного вирішуваній задачі, було одним з вирішальних факторів успішної розробки. Очевидним рішенням було обрання середовища Microsoft Visual Studio 2005.

На рис. 2 наведена функціональна схема системи обробки ультразвукових зображень, розроблена з урахуванням наведених вище міркувань щодо вимог до її структури.

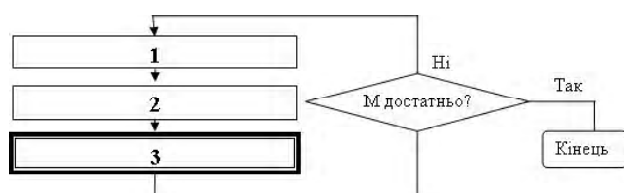


Рис. 1. Схема нового алгоритму підвищення роздільної здатності

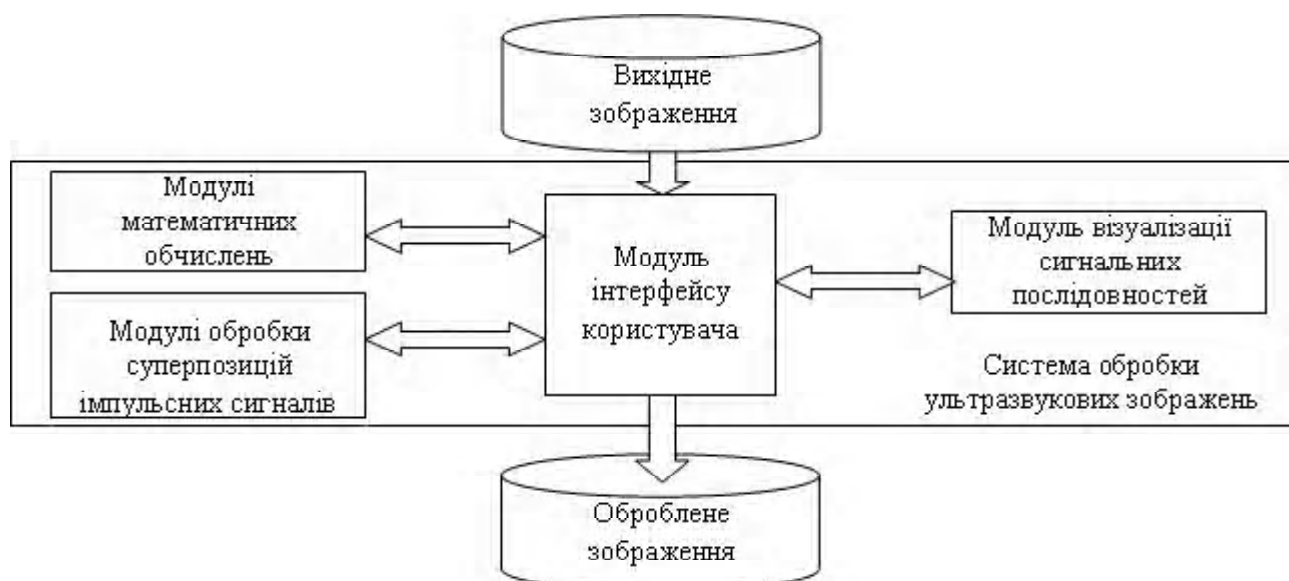


Рис. 2. Функціональна схема програми

Із рисунку видно, що система може розглядатися як сукупність чотирьох взаємопов'язаних логічних частин: модуля інтерфейсу користувача, модуля візуалізації сигнальних послідовностей, модулів математичних обчислень, модулів обробки суперпозицій імпульсних сигналів. Модуль інтерфейсу користувача є програмою, яка функціонує як сполучна ланка між іншими модулями системи. Він надає користувачеві змогу використовувати сервіси, що їх надає система, за допомогою екранних форм. При розробці модуля використовувалась мова програмування C#.

Модуль візуалізації сигнальних послідовностей являє собою елемент керування, який подає послідовність відліків дискретного сигналу у вигляді відповідного графіку сигналу або спектру. Він заснований на ієрархії класів Windows Forms і реалізований на мові програмування C#.

Модулі математичних обчислень – це поєднання розроблених за допомогою мов C# і C динамічних бібліотек, які надають можливість використання таких засобів, як комплексні числа, матриці комплексних чисел, швидке перетворення Фур'є.

Модулі обробки суперпозицій імпульсних сигналів – дві групи модулів, які реалізують адаптивний метод лінійного передбачення (бібліотека, реалізована на C#), і метод сигнального підпростору (поєднання бібліотек, реалізованих за допомогою C# і Fortran).

АЛГОРИТМ ФУНКЦІОНУВАННЯ ПРОГРАМИ

Першою дією програми після запуску є сканування директорії, в якій мають знаходитися модулі розширення системи (саме в такому вигляді було реалізовано методи обробки суперпозицій імпульсних сигналів). Знайдені коректні модулі додаються у список, який далі буде використовуватися користувачем при виборі методу обробки. Після цього користувачеві надається екранна форма, за допомогою якої буде здійснюватися взаємодія користувача з системою.

Вхідними даними для програми слугують ехо-імпульсні зображення в одному із розповсюджених графічних форматів (BMP, GIF, JPEG, PNG або TIFF) і параметри обраного користувачем методу.

У загальному вигляді алгоритм функціонування програми наведено на рис. 3.



Рис. 3. Алгоритм роботи програми

Після запуску програми користувачеві надається вікно, аналогічне наведеному на рис. 4. Як видно з рисунку, головне вікно програми можна розділити на декілька областей: – *заголовок вікна* використовується стандартним чином – для переміщення вікна, його згортання, розгортання і закриття (тобто, виходу з програми); *головне меню* містить: меню роботи з файлами («File»), меню налаштування графіків («Plots»), меню допомоги («Help»); *область візуалізації зображень* містить: вхідне зображення («Input image»), графік суперпозиції сигналів стовпця вхідного зображення («Input signal»), графік амплітудного спектра суперпозиції сигналів стовпця вхідного зображення («Input spectrum»), оброблене зображення («Output image»), графік суперпозиції сигналів стовпця обробленого зображення («Output signal»), графік амплітудного спектра суперпозиції сигналів стовпця обробленого зображення («Output spectrum»), перемикач номера стовпця вхідного і обробленого зображення, який розглядається на графіках («View column»), час обробки останнього зображення («Time elapsed»).

Область інформації про вхідне зображення («Image info») містить шлях до відкритого зображення («File name») і його геометричні розміри («Image size»). Тут також знаходиться перемикач перетворення Гільберта («Apply Hilbert transform») – активувати його рекомендується при обробці ультразвукових медичних зображень; *область налаштувань методів обробки зображень* («Processing options») містить список наявних методів («Processing») і дозволяє визначати їх параметри (кнопка «Options...»), а також – список методів обробки отриманого зображення («Postprocessing») і кнопку активації обробки («Process»); *рядок стану* містить відомості про поточний стан програми: якщо вона знаходиться у черговому стані – напис «Ready», якщо вона знаходиться у стані обробки зображення – напис «Processing», смугу поточного прогресу і кількість оброблених стовпців зображення.

ВИСНОВКИ

В роботі розглянуто проблему обробки ехо-імпульсних зображень з метою підвищення їх якості при візуалізації. Вона полягає в аналізі суперпозицій невідомих імпульсних сигналів, які складають ці зображення. Існуючі підходи до аналізу, а саме, метод інверсної фільтрації та метод кепстрального аналізу, не можуть бути використані в даній області через ряд обмежень, які вони накладають на вхідні суперпозиції. Тому було запропоновано методи, вільні від обмеження існуючих підходів, і засновані на використанні методів лінійного передбачення і сигнального підпростору у спектральній області.

За розробленими методами реалізовано програмне забезпечення для обробки ультразвукових зображень. Програму реалізовано на платформі .NET за допомогою мов C# і Fortran з використанням вільних бібліотек

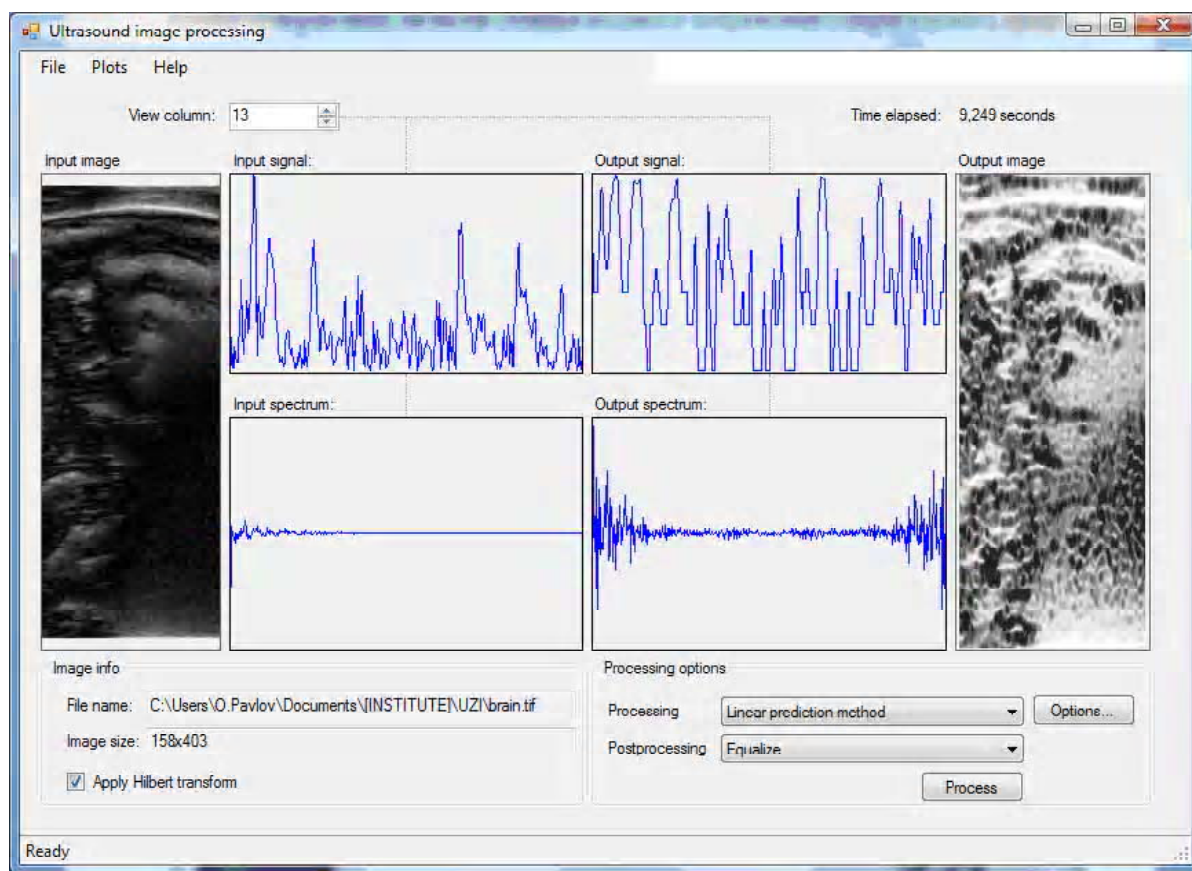


Рис. 4. Головне вікно програми

швидкого перетворення Фур'є fft та пошуку власних значень і векторів матриць EISPACK. Програма може успішно експлуатуватися на комп'ютерах під керуванням операційних систем Microsoft Windows 2000/XP/2003/Vista. Мінімальними апаратними вимогами є: процесор класу Intel Pentium III тактовою частотою 500 МГц; 96 МБ оперативної пам'яті; 200 МБ вільного місця на жорсткому диску; монітор SVGA і відеоадаптер з підтримкою розподільної здатності 1024x768.

Галузь застосування програми – підвищення якості зображень, отриманих за допомогою пристроїв ультразвукових досліджень або сейсмограм.

СПИСОК ЛІТЕРАТУРИ

1. *Newman, P.* The history of ultrasound / P. Newman, G. Rozycki // *Surgical Clinics of North America*. – 1998. – № 2 (78). – Р. 179–195.
2. *Митьков, В. В.* Практическое руководство по ультразвуковой диагностике. Общая ультразвуковая диагностика / В. В. Митьков. – М.: Видар, 2003. – 720 с.
3. *Иванов, В. К.* Теория линейных некорректных задач и ее приложения / В. К. Иванов, В. В. Васин, В. П. Танана. – М.: Наука, 1978. – 200 с.
4. *Чайделс, Д. Д.* Кепстр и его применение при обработке данных / Д. Д. Чайделс, Д. П. Скиннер, Р. Ч. Кемерайт // *ТИИЭР*. – 1977. – № 10. – С. 5–23.
5. *Степаненко, А. А.* Повышение чувствительности ультразвуковой диагностики на основе метода параметрического спектрального анализа второго порядка / А. А. Степаненко, А. М. Ахметшин // *Клиническая информатика и телемедицина*. – 2005. – Т. 2, № 1. – С. 98–100.
6. *Степаненко, А. А.* Разложение суперпозиций неизвестных импульсных сигналов методом адаптивного спектрального анализа второго порядка / А. А. Степаненко, А. М. Ахметшин // *Искусственный интеллект*. – 2005. – № 3. – С. 610–618.

УДК 519.7:004.8

7. *Степаненко, А. А.* Анализ эхо-импульсных изображений слоистых структур: метод сигнального подпространства / А. А. Степаненко, А. М. Ахметшин // *Радиоэлектроника, информатика, управление*. – 2004. – № 1. – С. 5–9.
8. *O'Docherty, M.* Object-Oriented Analysis and Design / M. O'Docherty. – Wiley & Sons, 2005. – 560 p.
9. *Рыжиков, Ю.* Современный Фортран / Ю. Рыжиков. – Минск: Корона Принт, 2007. – 288 с.

Стаття надійшла до редакції 14.04.2011.

Степаненко А. А., Пиза Д. М.

ПРОГРАММНЫЙ КОМПЛЕКС ДЛЯ АНАЛИЗА И ОБРАБОТКИ ЭХО-ИМПУЛЬСНЫХ ИЗОБРАЖЕНИЙ

На основе анализа методов обработки эхо-импульсных изображений были выявлены недостатки существующих подходов и предложены два новых метода. Разработанные методы и базирующаяся на них программная система позволяют повысить эффективность визуального анализа медицинских и сейсмических ультразвуковых изображений.

Ключевые слова: математическая модель, метод линейного предсказания, суперпозиция импульсного сигнала, спектр, эхо-импульсное изображение, модуль математических вычислений.

Stepanenko O. O., Piza D. M.

PROGRAM COMPLEX FOR ANALYSIS AND TREATMENT ECHO-PULSE IMAGES

On the basis of analysis of methods of processing of echo-pulse images the lacks of existent approaches were reduced and two new methods are offered. Worked out methods and programmatic system which is based on them, allow to promote efficiency of visual analysis medical and seismic ultrasonic images.

Key words: mathematical model, method of linear prediction, superposition of impulsive signal, spectrum, echo-pulse images, module of mathematical calculations.

Шафроненко А. Ю.¹, Волкова В. В.², Бодянский Е. В.³

¹Сталкер-исследователь Харьковского национального университета радиоелектроніки

²Старший преподаватель Харьковского национального университета радиоелектроніки

³Д-р техн. наук, профессор Харьковского национального университета радиоелектроніки

АДАПТИВНАЯ КЛАСТЕРИЗАЦИЯ ДАННЫХ С ПРОПУЩЕННЫМИ ЗНАЧЕНИЯМИ

Предложена адаптивная нейросетевая система, позволяющая решать задачу кластеризации данных с пропущенными значениями в on-line режиме с постоянной коррекцией восстанавливаемых элементов таблицы и центроидов кластеров. Введенная нейросистема характеризуется высоким быстродействием и простотой численной реализации.

Ключевые слова: адаптивная нейросетевая система, восстановление пропусков, кластеризация.

ВВЕДЕНИЕ

Задача кластеризации данных является важным элементом общей проблемы Data Mining, а для ее решения

сегодня существует множество подходов и алгоритмов: от сугубо интуитивных и эвристических до строго математических [1–9]. Вместе с тем, во многих задачах Data

Mining, включая, естественно, кластеризацию, исходные таблицы данных «объект-свойство» могут содержать пустые клетки (пропуски), информация в которых по тем или иным причинам отсутствует. Задаче восстановления таких пропущенных наблюдений уделялось достаточное внимание [6, 10–12], при том наиболее эффективными в данной ситуации оказались подходы, основанные на математическом аппарате мягких вычислений (вычислительного интеллекта) [13, 14], и, прежде всего, искусственных нейронных сетей [12, 15–17]. Вместе с тем, известные подходы к восстановлению пропусков и традиционные алгоритмы кластеризации работоспособны лишь в случаях, когда исходная таблица задана априорно и число ее строк или столбцов не изменится в процессе обработки. В то же время существует достаточно широкий класс задач, когда данные поступают на обработку последовательно в on-line режиме, при этом заранее неизвестно, какой из обрабатываемых векторов-образов может содержать пропуски. При этом процессы восстановления данных и их кластеризации должны протекать одновременно в реальном времени.

ПОСТАНОВКА ЗАДАЧИ

Пусть задана исходная $(N \times n)$ таблица ($N \geq 1, n \geq 1$) «объект-свойство».

Таблица 1, содержащая информацию об N объектах, каждый из которых описывается $(1 \times n)$ – вектором-строкой признаков $\underline{x}_i = (x_{i1}, \dots, x_{ip}, \dots, x_{ij}, \dots, x_{in})$, при этом предполагается, что N_G строк могут иметь по одному пробелу, а $N_F = N - N_G$ заполнены полностью. В процессе обработки таблицы необходимо заполнить пропуски и сформировать m кластеров. При этом требуется, чтобы восстановленные элементы были бы в определенном смысле «наиболее правдоподобны» или «близки» к априори неизвестным закономерностям, содержащимся в таблице, которые могут неизвестным образом изменяться в процессе обработки.

АДАПТИВНЫЙ АЛГОРИТМ ЗАПОЛНЕНИЯ ПРОПУСКОВ

Для начала представим таблицу 1 в виде $(N \times n)$ – матрицы X , в которой отсутствует один элемент x_{kj} или в более общем случае отсутствуют N_G элементов. Предполагается [10], что между столбцами $\vec{x}_j = (x_{1j}, \dots, x_{ij}, \dots, x_{kj}, \dots, x_{Nj})^T$ существует линейная корреляция, с учетом которой и про-

изводится восстановление пропуска с помощью уравнения линейной регрессии

$$\hat{x}_{kj} = w_{j0} + w_{j1}x_{k1} + w_{j2}x_{k2} + \dots + w_{j,j-1}x_{k,j-1} + \dots + w_{jn}x_{kn} = \underline{X}_{kj} w_j, \quad (1)$$

где $w_j = (w_{j0}, w_{j1}, \dots, w_{jn})^T$ – $(n+1)$ – вектор-столбец параметров, подлежащих определению; $\underline{X}_{kj} = (1, x_{k1}, \dots, x_{k,j-1}, x_{k,j+1}, \dots, x_{kn})$ – $(1 \times n)$ вектор-строка признаков k -го объекта без kj -го элемента и с единицей на первой позиции.

Вектор неизвестных параметров w_j может быть найден с помощью стандартного метода наименьших квадратов, для чего из матрицы X следует исключить k -ю строку, j -й столбец, добавить слева столбец, образованный единицами, и на основе полученной $((N-1) \times n)$ – матрицы X_j рассчитать оценки параметров

$$w_j = (X_j^T X_j)^+ X_j^T \vec{X}_j, \quad (2)$$

где $\vec{X}_j = (x_{1j}, \dots, x_{ij}, \dots, x_{k-1,j}, x_{k+1,j}, \dots, x_{Nj})^T$.

Если же пробелы имеются в N_G строках и в разных столбцах, из матрицы X исключаются все эти строки и на основании усеченной $(N_F \times n)$ матрицы n раз вычисляются векторы параметров w_j (2) для всех $j = 1, 2, \dots, n$. Далее с помощью уравнения (1) заполняются все пробелы полученными оценками $\hat{x}_{kj}$.

Если же данные поступают на обработку последовательно строка за строкой, вместо оценки (2) может быть использован рекуррентный метод наименьших квадратов, который, однако, предлагает неизменность во времени всех связей, существующих в матрице данных. Экспоненциально же взвешенный рекуррентный метод наименьших квадратов может быть неустойчив при малых значениях параметра сглаживания.

В связи с этим для on-line анализа данных представляется целесообразным использование адаптивных алгоритмов обучения, имеющих как фильтрующие, так и следящие (для нестационарных ситуаций) свойства [18]. Если в процессе обработки массива данных было проанализировано и восстановлено N строк, то с приходом $(N+1)$ -го наблюдения оценки w_j уточняются с помощью адаптивной рекуррентной процедуры

$$\begin{cases} w_j(N+1) = w_j(N) + r_j^{-1}(N+1)(x_{N+1,j} - \underline{X}_{N+1,j} w_j(N)) \underline{X}_{N+1,j}^T, \\ r_j(N+1) = \alpha r_j(N) + \|\underline{X}_{N+1,j}\|^2, \end{cases} \quad (3)$$

где $0 \leq \alpha \leq 1$ – параметр сглаживания, задающий компромисс между процессами фильтрации наблюдений и слежением за изменяющимися характеристиками данных.

Таблица 1.

	1	...	p	...	j	...	n
1	x_{11}	...	x_{1p}	...	x_{1j}	...	x_{1n}
...	...	...	...	...	...	...	...
i	x_{i1}	...	x_{ip}	...	x_{ij}	...	x_{in}
...	...	...	...	...	...	...	...
k	x_{k1}	...	x_{kp}	...	x_{ki}	...	x_{kn}
...	...	...	...	...	...	...	...
N	x_{N1}	...	x_{Np}	...	x_{Nj}	...	x_{Nn}
...	...	...	...	...	...	...	...

Обработку информации в режиме последовательного поступления данных удобно организовать с помощью нейросетевой системы, основными элементами которой являются n -параллельно работающих адаптивных линейных ассоциаторов (ALA) [19], настраиваемых с помощью алгоритма (3). При этом, если в k -й момент времени на вход системы поступает вектор наблюдений $\vec{x}_k = (x_{k1}, x_{k2}, \dots, x_{kin})$ полный или с пропусками, на выходе появляется он же, если $\vec{x}_k$ полный, или его оценка, если в $\hat{x}_k$ содержались пробелы.

АДАПТИВНЫЙ АЛГОРИТМ КЛАСТЕРИЗАЦИИ

Для решения задачи кластеризации в on-line режиме целесообразно воспользоваться самоорганизующейся картой Т. Кохонена [20], имеющей простую архитектуру с прямой передачей информации и кроме нулевого рецепторного слоя содержащей единственный слой нейронов, чаще всего тех же адаптивных линейных ассоциаторов. Каждый нейрон связан с каждым рецептором нулевого слоя прямыми связями и со всеми остальными нейронами поперечными внутрислойными (латеральными) связями. Именно латеральные связи обеспечивают возбуждение одних нейронов и торможение других.

Благодаря такой организации сети, каждый нейрон ALA получает всю информацию об анализируемом векторе-образе и генерирует на своем выходе соответствующий отклик, после чего между нейронами возникает конкуренция, в результате которой определяется единственный нейрон-победитель с максимальным выходным сигналом. Этот сигнал по латеральным связям обеспечивает возбуждение ближайших «соседей» победителя и подавление реакции далеко отстоящих нейронов. Таким образом в процессе конкурентного самообучения формируются группы нейронов, каждый из которых максимальным образом реагирует на образы «своих» кластеров-подобластей входного пространства сигналов.

Рассмотрим простейшую карту Кохонена, имеющую 1D топологию, n рецепторов и m (по числу возможных кластеров) нейронов, каждый из которых характеризуется собственным вектором синаптических весов $w_q^K, q = 1, 2, \dots, m$, определяющих центроид соответствующего кластера.

Каждый нейрон сети в k -й момент времени получает на входы n -мерный входной вектор

$$\tilde{x}_k = \frac{\hat{x}_k^T}{\|\hat{x}_k\|} \quad (4)$$

и генерирует на своем выходе сигнал

$$y_{kl} = w_l^{KT}(k-1)\tilde{x}_k, \quad (5)$$

зависящий от вектора синаптических весов $w_l^K(k-1)$, настроенных по данным, полученным до момента k , на определенную область (кластер) входного пространства. Близкие в смысле используемой метрики входные векторы $\tilde{x}_k$ и $\tilde{x}_i$ могут возбуждать либо один и тот же нейрон w_q^K , либо два нейрона соседа w_q^K и w_{q+1}^K или w_q^K и $w_{q-1}^K, q = 1, 2, \dots, m$.

В основе алгоритма самоорганизации карты Кохонена лежат принципы конкурентного самообучения [19], при этом как и любая другая процедура настройки нейронной сети, работа алгоритма начинается с инициализации синаптических весов сети, которые обычно выбираются с помощью генератора случайных чисел, при этом для каждого из нейронов должно выполняться условие

$$\|w_l^K(0)\| = 1, l = 1, 2, \dots, m. \quad (6)$$

Процедура самоорганизации реализуется в три основных этапа: конкуренции, кооперации и синаптической адаптации и начинается с анализа образа $\tilde{x}_k$, поступающего с рецепторного слоя на все нейроны сети. Для каждого из нейронов вычисляется расстояние

$$D(\tilde{x}_k, w_l^K(k-1)) = \|\tilde{x}_k - w_l^K(k-1)\|, \quad (7)$$

при этом, если входные векторы предварительно нормированы с помощью выражения (4) так, что

$$\|\tilde{x}_k\| = 1, \quad (8)$$

а в качестве расстояния (7) используется евклидова метрика, то мерой расстояния между $\tilde{x}_k$ и $w_l^K(k-1)$ может служить

$$\begin{aligned} D(\tilde{x}_k, w_l^K(k-1)) &= 2(1 - \tilde{x}_k^T w_l^K(k-1)) = \\ &= 2(1 - \cos(\tilde{x}_k, w_l^K(k-1))), \end{aligned} \quad (9)$$

а мерой близости – скалярное произведение

$$-1 \leq \tilde{x}_k^T w_l^K(k-1) = \cos(\tilde{x}_k, w_l^K(k-1)) \leq 1. \quad (10)$$

Далее определяется нейрон-победитель, «ближайший» ко входному образу $\hat{x}_k$ такой, что

$$D(\tilde{x}_k, w_l^K(k-1)) = \min_l D(\tilde{x}_k, w_l^K(k-1)), \quad (11)$$

после чего, временно опуская этап кооперации, можно подстроить синаптические веса сети с помощью модифицированного правила самообучения

$$w_q^K(k) = \begin{cases} \frac{w_q^K(k-1) + \eta(k)(\tilde{x}_k - w_q^K(k-1))}{\|w_q^K(k-1) + \eta(k)(\tilde{x}_k - w_q^K(k-1))\|}, & \text{если } q\text{-й нейрон победил,} \\ w_q^K(k-1) & \text{в противном случае.} \end{cases} \quad (12)$$

Несложно видеть, что процедура (12) реализует принцип «победитель получает все», при этом вектор синаптических весов $w_q^K(k-1)$ нейрона-победителя «подтягивается» ко входному вектору $\tilde{x}_k$ на расстояние, определяемое шагом настройки $\eta(k)$.

Регулирование величины шага обычно производится, исходя из эмпирических соображений, при этом весьма удобной является процедура [21]

$$\eta(k) = r^{-1}(k), r(k) = \alpha r(k-1) + \|\tilde{x}_k\|^2 = \alpha r(k-1) + 1. \quad (13)$$

Несложно заметить структурную схожесть алгоритмов обучения (3) и (12), (13).

Одной из особенностей карты Кохонена является наличие этапа кооперации в процессе самоорганизации, когда нейрон-победитель определяет локальную область топологического соседства, в которой возбуждается не только он сам, но и его ближайшее окружение, при этом более близкие к победителю нейроны возбуждаются сильнее чем удаленные. Эта топологическая область определяется функцией соседства Φ_{ql} , зависящей от расстояния $D(w_q^K(k-1), w_l^K(k-1))$ между победителем $w_q^K(k-1)$ и любым из нейронов сети $w_l^K(k-1)$. Как правило, Φ_{ql} – это ядерная функция симметричная относительно максимума в точке $D(w_q^K(k-1), w_q^K(k-1)) = 0$ и принимающая в ней единичное значение, монотонно убывающая с ростом расстояния и стремящаяся к нулю при $D(w_q^K(k-1), w_l^K(k-1)) = 2$. При этом наиболее часто в качестве функции соседства используются гауссиан, конус, функции Епанечникова, «мексиканская шляпа» и т. п.

Использование функции соседства приводит к модифицированному правилу обучения

$$\begin{cases} w_l^K(k) = \frac{w_l^K(k-1) + r^{-1}(k)\Phi_{ql}(\tilde{x}_k - w_l^K(k-1))}{\|w_l^K(k-1) + r^{-1}(k)\Phi_{ql}(\tilde{x}_k - w_l^K(k-1))\|} \forall l, \\ r(k) = \alpha r(k-1) + 1, \end{cases} \quad (14)$$

реализующему принцип «победитель получает больше». При $\Phi_{ql} = \delta_{ql} = 1$ при $q=l$ и 0 в остальных случаях, вновь приходим к алгоритму (12), обеспечивающему на каждом такте настройку единственного нейрона-победителя $w_q^K(k-1)$.

В принципе, для обучения самоорганизующейся карты можно вообще обойтись без нахождения победителя, а в качестве функции соседства использовать некоторую конструкцию, зависящую от выходных сигналов нейронов (5) так, что

$$\begin{cases} w_l^K(k) = \frac{w_l^K(k-1) + r^{-1}(k)\varphi(y_{kl})(\tilde{x}_k - w_l^K(k-1))}{\|w_l^K(k-1) + r^{-1}(k)\varphi(y_{kl})(\tilde{x}_k - w_l^K(k-1))\|} \forall l, \\ r(k) = \alpha r(k-1) + 1. \end{cases} \quad (15)$$

Поскольку

$$y_{kl} = w_l^{KT}(k-1)\tilde{x}_k = \cos(\tilde{x}_k, w_l^K(k-1)), \quad (16)$$

то наиболее простой такой конструкцией является

$$\varphi(y_{kl}) = \frac{1 + y_{kl}}{2}, \quad (17)$$

удовлетворяющая всем условиям, предъявляемым к ядерным функциям соседства. Тогда окончательно алгоритм обучения карты Кохонена для нахождения центроидов m кластеров имеет вид

$$\begin{cases} w_l^K(k) = \frac{w_l^K(k-1) + r^{-1}(k)\frac{1 + y_{kl}}{2}(\tilde{x}_k - w_l^K(k-1))}{\|w_l^K(k-1) + r^{-1}(k)\frac{1 + y_{kl}}{2}(\tilde{x}_k - w_l^K(k-1))\|} \forall l, \\ r(k) = \alpha r(k-1) + 1. \end{cases} \quad (18)$$

Нейросетевая архитектура, реализующая процесс кластеризации данных с пропусками, приведена на рис. 1.

Как видно, введенная нейронная сеть образована однотипными нейронами – адаптивными линейными ассоциаторами, настраиваемыми с помощью достаточно простых и очевидных алгоритмов обучения.

ЗАКЛЮЧЕНИЕ

Рассмотрена задача кластеризации данных, содержащихся в таблицах «объект-свойство» и имеющих пропуски, в режиме последовательного поступления этих данных на обработку. Предложена адаптивная нейросетевая система, позволяющая решать эту задачу в on-line режиме с постоянной коррекцией восстанавливаемых элементов таблицы и центроидов кластеров. Введенная нейросистема образована набором адаптивных линейных ассоциаторов, характеризуется высоким быстродействием и простотой численной реализации.

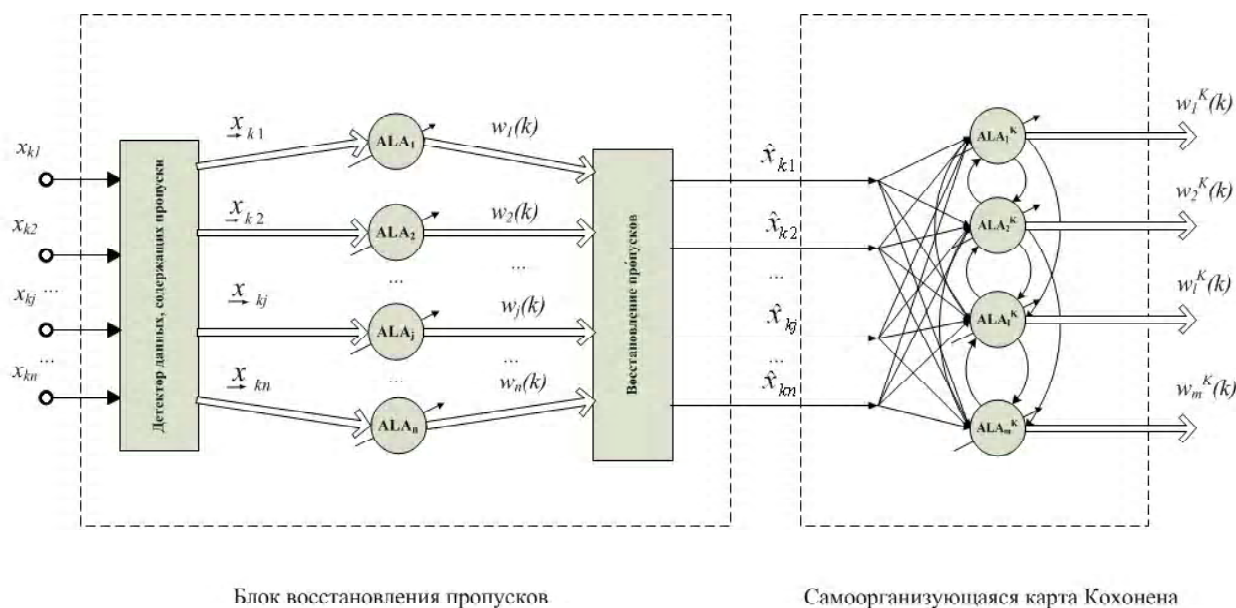


Рис. 1. Искусственная сеть для адаптивной кластеризации данных с пропущенными значениями

СПИСОК ЛІТЕРАТУРИ

1. Jain, A. Data clustering: A review / A. Jain, M. Murty, P. Flynn // ACM Computing Surveys. – 1999. – №3(31). – P. 264–323.
2. Дюк, В. Data Mining / В. Дюк, А. Самойленко. – С. Пб. : Питер, 2001. – С. 368.
3. Friedman, J. The Elements of Statistical Learning. Data Mining, Inference and Prediction / J. Friedman, T. Hastie, R. Tibshirani. – Berlin : Springer, 2003. – P. 552.
4. Xu, R. Survey of clustering algorithms / R. Xu, D. II. Wunsch // IEEE Trans on Neural Networks. – 2005. – №3 (16). – P. 645–678.
5. Han, J. Data Mining: Concepts and Techniques / J. Han, M. Kamber. – Amsterdam : Morgan Kaufman Publ, 2006. – P. 754.
6. Gan, G. Data Clustering: Theory, Algorithms and Applications / G. Gan, Ch. Ma, J. Wu. – Philadelphia : SIAM, 2007. – P. 466.
7. Abonyi, J. Cluster Analysis for Data Mining and System Identification / J. Abonyi, B. Feil. – Basel : Birkhaeuser, 2007. – P. 303.
8. Olson, D. Z. Advanced Data Mining Techniques / D. Z. Olson, D. Dursun. – Berlin : Springer, 2008. – P. 180.
9. Xu, R. II. Clustering / R. Xu, D. C. Wunsch. – Hoboken, N. J. : John Wiley & Sons, Inc., 2009. – P. 341.
10. Загоруйко, Н. Г. Эмпирические предсказания. – Новосибирск : Наука, 1979. – С. 120.
11. Gorban, A. Principal Manifolds for Data Visualization and Dimension Reduction. Lecture Notes in Computational Science and Engineering, Vol. 58 / A. Gorban, B. Kegl, B. Wunsch, A. (Eds.) Zinovyev. – Berlin-Heidelberg – New York : Springer, 2007. – P. 330.
12. Marwala, T. Computational Intelligence for Missing Data Imputation, Estimation, and Management / Marwala T. // Knowledge Optimization Techniques. – Hershey – New York : Information Science Reference, 2009. – P. 303.
13. Maimon, O. Soft computing for Knowledge Discovery and Data Mining / O. Maimon, L. Rokach. – New York : Springer-Verlag, 2007. – P. 448.
14. Rutkowski, L. Computational Intelligence Methods and Techniques / L. Rutkowski. – Berlin-Heidelberg : Springer-Verlag, 2008. – P. 514.
15. Bishop, C. M. Neural Networks for Pattern Recognition. – Oxford : Clarendon Press, 1995. – P. 482.
16. Gorban, A. N. Neural network modeling of data with gaps / A. N. Gorban, A. A. Rossiev, D. C. Wunsch II // Радіоелектроніка, інформатика, управління. – 2000. – №1 (3). – С. 47–55.
17. Tkacz, M. Artificial neural networks in incomplete data sets processing / In: Eds. M. A. Klopotek, S. T. Wierzchon, K. Trojanowski // Intelligent Information Processing and Web Mining. – Berlin-Heidelberg : Springer – Verlag, 2005. – P. 577–583.
18. Bodyanskiy, Ye. An adaptive learning algorithm for a neuro-fuzzy network / Ye. Bodyanskiy, V. Kolodyazhniy, A. Stephan // Ed. B. Reusch «Computational Intelligence. Theory and Applications». – Berlin-Heidelberg – New York : Springer, 2001. – P. 68–75.
19. Haykin, S. Neural Networks. A Comprehensive Foundation / Haykin, S. – Upper Saddle River, N. J.: Prentice Hall, 1999 – P. 842.
20. Kohonen, T. Self-Organizing Maps / T. Kohonen. – Berlin: Springer-Verlag, 1995. – P. 362.
21. Бодянский, Е. В. Искусственные нейронные сети: архитектуры, обучение, применения / Е. В. Бодянский, О. Г. Руденко. – Х. : ТЕЛЕТЕХ, 2004. – С. 372.

Стаття надійшла до редакції 06.05.2011.

Шафроненко А. Ю., Волкова В. В., Бодянский С. В.

АДАПТИВНА КЛАСТЕРИЗАЦІЯ ДАНИХ З ПРОПУЩЕНИМИ ЗНАЧЕННЯМИ.

Запропонована адаптивна нейромережева система, що дозволяє вирішувати задачу кластеризації даних з пропущеними значеннями в реальному часі з постійним коригуванням відновлених елементів таблиці і центрів кластерів. Впроваджена нейронна система характеризується високою швидкістю і простою числової реалізації.

Ключові слова: адаптивна нейромережева система, відновлення пропусків, кластеризація.

Shafronenko A.Yu. Volkova V. V., Bodyanskiy Yev. V.

ADAPTIVE CLUSTERING WITH MISSING VALUES

In this paper the adaptive neural network system that solves the clustering problem of data with gaps is proposed. This system allows to process the data in the on-line mode with a constant correction of recoverable table's elements and centers of clusters. A proposed neural system has high speed and simple numerical realization.

Key words: adaptive neural network system, the restoration of gaps, clustering.

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

ПРОГРЕССИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

PROGRESSIV INFORMATICS TECHNOLOGIES

УДК 004.052

Брагина Т. И.¹, Табунщик Г. В.²¹Аспирант Запорожского национального технического университета²Канд. техн. наук, доцент Запорожского национального технического университета

АНАЛИЗ ПОДХОДОВ К УПРАВЛЕНИЮ РИСКАМИ В ПРОГРАММНЫХ ПРОЕКТАХ С ИТЕРАТИВНЫМ ЖИЗНЕННЫМ ЦИКЛОМ

Выполнен обзор основных рисков программных проектов с итерационным жизненным циклом. Предложены рекомендации по управлению выявленными рисками в зависимости от выбранной модели разработки программных проектов. Выделены наиболее критичные риски и проведен их качественный анализ.

Ключевые слова: управление рисками, программные проекты, реакция на риск.

1. ВВЕДЕНИЕ

Разработка информационных систем связана с определенной совокупностью рисков – неопределенных событий или условий, наступление которых отрицательно или положительно сказывается на целях проекта. Каждый проект задумывается и разрабатывается на основании ряда гипотез, сценариев и допущений – факторов, которые для целей планирования считаются верными, реальными или определенными без привлечения доказательств. Неопределенность в рамках программного проекта (ПП) следует обязательно рассматривать в качестве потенциального источника возникновения рисков проекта, так как решение проектных задач управления в условиях неопределенности порождает риски выхода из бюджета и расписания, что ставит под угрозу достижение целей ПП. Следовательно, процессы управления рисками являются наиболее важной компонентой процессов принятия решений и управления проектами.

2. ПОСТАНОВКА ЗАДАЧИ

Процессы управления рисками проекта включают в себя [1]:

- планирование управления рисками;
- идентификацию рисков;

- качественный анализ рисков;
- количественный анализ рисков;
- планирование реагирования на риски;
- мониторинг и управление рисками.

Эти процессы взаимодействуют как друг с другом, так и с процессами из других областей знаний. В зависимости от потребностей проекта в каждом процессе могут принимать участие один или несколько человек или групп. Каждый процесс имеет место, по крайней мере, один раз в ходе каждого проекта, а если проект разделен на фазы – то в одной или нескольких фазах проекта.

Исходными данными для выявления рисков является имеющаяся информация как об общих, так и о специфических для данного проекта рисках, связанная с бизнесом, технологиями, организациями и внешними условиями. Основным фактором, влияющим на возможные риски, является принятая в компании модель программного обеспечения (ПО), т. е. структура, определяющая последовательность выполнения и взаимосвязи процессов, действий и задач на протяжении жизненного цикла (ЖЦ). Существуют также дополнительные факторы, которым должно быть уделено внимание: имеющийся у команды опыт, применяемые внутри организации подходы к рискам, выраженные в виде правил и инструк-

ций, а также вся информация о проекте, известная на данный момент. После сбора необходимых данных проектная группа должна работать над заблаговременным выявлением рисков и создавать стратегии и планы по управлению ими, должны быть разработаны планы по решению проблем в случае их возникновения [2].

Для различных моделей ПО можно выделить основные, наиболее часто встречающиеся риски, характерные для каждой из них. Выделим наиболее распространенные модели, ориентированные на итеративный процесс разработки – Rational Unified Process (RUP), Microsoft Solutions Framework (MSF) и Agile (eXtreme Programming, Crystal, Feature Driven Development) [3].

Предвосхищение потенциальных проблем и заблаговременная подготовка четко составленных планов по борьбе с ними сокращает временные затраты в критических ситуациях и ограничивает негативный эффект, создаваемый рисками. Поэтому руководитель проекта, определив модель разработки ПО, которую будет использовать для проекта, должен осознавать сильные и слабые стороны данной модели, а также то, что управление рисками будет зависеть от характерных черт выбранной модели [4, 5].

Целью работы является разработка рекомендаций для управления рисками при использовании различных моделей разработки ПО. Для этого необходимо идентифицировать характерные риски разработки ПО, определить наиболее эффективные методы управления данными рисками, учитывая различия между моделями, выделить наиболее критичный риск для каждой модели.

3. ИДЕНТИФИКАЦИЯ РИСКОВ ПРОГРАММНЫХ ПРОЕКТОВ С ИТЕРАТИВНЫМ ЖИЗНЕННЫМ ЦИКЛОМ

Анализ литературы [6–10] позволяет выделить основные риски проектов по разработке ПО:

- внутренние нарушения календарного планирования;
- изменение требований;
- текучесть кадров;
- нарушение спецификаций;
- низкая производительность;
- недостаточное внимание к проекту со стороны руководства компании;
- отсутствие мотивации персонала компании.

Нарушения календарного планирования

Данный риск появляется из-за определенных нарушений (или полного несоответствия) процесса планирования бюджета времени и средств. Это можно рассматривать как ошибку собственно календарного планирования в противовес ошибкам осуществления проекта. Ошибка календарного планирования – не только реальный риск, но и самый крупный из пяти главных рисков по степени влияния на расхождение плана проекта и реального исполнения.

Ошибки календарного планирования можно рассматривать как тенденцию неверно судить о размерах про-

дукта, который предстоит создать. В случае, когда календарное планирование строится без учета размера продукта, весьма вероятен перерасход времени на 50–80 %.

Статистика по IT-проектам показывает, что многие компании не выполняют предварительной оценки размера проекта, выбирая вместо этого календарное планирование от конца к началу или принимая желаемое за действительное. Те компании, которые прилагают серьезные усилия для определения размера продукта, могут сократить влияние ошибок календарного планирования до 15 %.

Изменение требований

Программное обеспечение, которое разрабатывается командой, всегда предназначено для того, чтобы соответствовать области деятельности клиента. Однозначно эта область не останется статичной за время создания программного обеспечения. Она будет изменяться со скоростью, диктуемой рынком и собственными темпами технологического развития. Разница между тем, что клиенты хотят в начале и в конце периода разработки возникает из-за изменений, которые произошли в данной области бизнеса за это время.

С точки зрения проекта, это изменение всегда является увеличением требований. Даже удаление того, что уже создано – своего рода увеличение, поскольку требует дополнительной работы.

Текучесть кадров

Возможность текучести кадров обычно не рассматривается в процессе оценки проекта, в то время как персонал представляет собой основной ресурс фирмы по разработке программных проектов. Текучесть кадров может привести к упущению прибыли, потере квалифицированных рабочих, снижению качества продукции и другим проблемам.

Нарушение спецификаций

Реализация риска нарушения спецификаций чаще всего является фатальной для проекта.

Нарушение спецификаций относится к краху процесса переговоров по установлению требований, которые идут в начале любого проекта. Когда существует глубокий конфликт, не позволяющий прийти к согласию, то часто результат маскируют. В итоге приходится столкнуться с отложенными проблемами, и конфликт разгорается вновь. В худшем случае это происходит на очень поздней стадии проекта, когда истрачены отведенные на проект ресурсы. Проект очень уязвим в этой стадии, и отказ любой из сторон поддерживать его может привести к быстрому прекращению разработки – процентный диапазон прекращения проектов из-за нарушения спецификаций составляет 10–15 % [6].

По отношению к данному риску может использоваться только политика ликвидации риска. Для рассмотрения проблемы неоднозначности, используемой для сокрытия разногласий, определим прекращение прений по поводу спецификаций как то, что все стороны подписа-

лись под входными и выходными граничными условиями проекта и определениями данных, вплоть до данных элементарного уровня из всех входящих и исходящих потоков данных или функций для создания данных. Хотя соглашение по потокам данных может быть только частью требуемого согласования, но это — ключевая часть. Поскольку описания данных менее склонны к неоднозначности, чем описания функций, заключаем, что отказ от претензий по входящим и исходящим потокам данных является хорошим показателем согласия. Когда такое согласие достигнуто, риск прекращения следует исключить из рассмотрения.

Низкая производительность

Различия между командами проектов в производительности в целом несколько сглажены и всегда меньше, чем индивидуальные различия. Более того, некоторые различия индивидуальной производительности возникают из-за четырех главных рисков, описанных выше. После устранения воздействия других рисков и распространения индивидуальных различий на команды можно определить, что этот фактор, как правило, сбалансирован: одинакова вероятность как позитивных, так и негативных изменений производительности. Сбалансированный риск (низкая или высокая производительность) лишь вносит шум в процесс. Он расширяет диапазон неопределенности без сдвига среднего ожидаемого показателя.

Существует множество причин (сопровождение действующих систем, повышение квалификации, участие в подготовке технико-коммерческих предложений, участие в презентациях, административная работа, отпуска, праздники, больничные), по которым участники проекта не смогут работать по проекту 100 % своего времени. При принятии риска рекомендуется планировать, что разработчики, которые назначены в проект на 100 %, будут реально работать над задачами проекта в среднем от 24 до 32 часов в неделю вместо 40 часов [11].

Недостаточное внимание к проекту со стороны руководства компании

Часть неудач по созданию ПО объясняется фрагментарностью подхода к процессу управления. Например, концентрируя свои усилия на технической стороне реализации, управляющий проектом уделяет недостаточное внимание проблемам управления человеческими ресурсами. Или, наоборот, техническая сторона контролируется не на достаточном уровне либо допускаются несоответствия в графике разработки.

Отсутствие мотивации персонала компании

Мотивация сотрудников занимает одно из центральных мест в управлении созданием ПП. Руководитель организации должен сам выбирать каким образом стимулировать каждого работника для достижения целей организации. Если этот выбор сделан удачно, то руководитель получает возможность координировать усилия многих людей и сообща реализовывать потенциальные возмож-

ности группы разработчиков. В ином случае, отсутствие мотивации персонала приведет к срыву сроков, несоответствию необходимому качеству и надежности ПП, в крайнем случае прекращению разработки проекта.

4. ПЛАНИРОВАНИЕ РЕАКЦИИ НА РИСК ДЛЯ ПРОЕКТОВ С ИТЕРАЦИОННЫМ ПРОЦЕССОМ РАЗРАБОТКИ

По отношению к выявленным рискам возможны следующие действия [6]:

- ликвидация риска (например, за счет устранения причины);
- уменьшение риска (например, за счет использования дополнительных защитных средств);
- принятие риска (и выработка плана действия в соответствующих условиях);
- переадресация риска (например, путем заключения страхового соглашения).

Реакции на возникновение различных рисков и, соответственно, методы управления рисками будут варьироваться в зависимости от выбранной модели ПО.

В табл. 1 рассмотрены методы управления рисками для наиболее распространенных моделей ПО с итерационным циклом разработки – RUP, MSF и Agile [4], учитывающие особенности каждого типа проекта.

В рамках MSF управление рисками рассматривается как процесс их выявления, анализа и эффективной превентивной работы над ними. Эффективный процесс выявления и управления рисками помогает достичь разумных компромиссов между опасностями и открывающимися возможностями. В случае управления проектом с большим количеством данных целесообразно использовать MSF, т.к. систематизация и структуризация информации в форме базы данных позволит упростить работу с ними. Но в то же время изменение требований потребует пересмотра структуры и формата базы данных, что повлечет большие трудовые и временные затраты.

RUP акцентирует внимание на планировании ЖЦ и отдельных итераций, управлении рисками, наблюдаемости хода проекта и метриках проекта. В RUP управление рисками означает определение и оценку рисков, принятие линии поведения, направленной на устранение, снижение вероятности риска, а также выбор действий на случай реализации риска. RUP опирается на планирование, что является сильной стороной модели, но подвергает проект повышенной угрозе возникновения ошибок календарного планирования.

Особенностью Agile является ожидание изменений, то есть в гибком процессе проектная команда не пытается зафиксировать требования в начале проекта и затем следовать жестко определенному плану. Изменения могут быть сделаны на сколь угодно позднем этапе проекта, что повышает возможность возникновения рисков ситуаций. Для решения этих проблем в Agile чаще всего используют итеративную разработку с возможно более короткой продолжительностью итерации, опираясь на тщательный подбор разработчиков. Для работы по ме-

Таблица 1. Возможная реакция на риск в зависимости от типа программного проекта

Тип проекта Тип риска	RUP	MSF	Agile
Нарушения календарного планирования, срыв сроков	Перерасчет сроков выполнения этапов работ	Планирование временного резерва	Выделение сверхурочных трудочасов, повышенное внимание к предварительному планированию
	Ликвидация или уменьшение данного риска возможно путем оценки длительности определенных работ или всего проекта несколькими методами: – опрос экспертов – проведение интервью соответствующих квалифицированных специалистов; – мозговой штурм – проведение общего собрания, на котором специалисты по очереди высказывают свои мнения о необходимых временных затратах на определенные типы работы. Споры и замечания не допускаются, все мнения записываются, группируются и оцениваются по типам и характеристикам работ; – сбор данных – проведение сбора данных по нескольким проектам относительно масштабов недооценки размера проекта; – имитационное моделирование – моделирование и анализ неопределенности в оценках основных показателей проекта (денежные и временные затраты).		
Изменение требований	учет увеличения трудоемкости и временных затрат в случае возможного роста требований, например, на 50 % (принятие риска)	переоценка проекта каждый раз, когда требования добавляются/изменяются (ликвидация риска)	подписание контракта с компенсацией затрат (перерасчет риска заказчику)
Текущая кадров	определение причин, по которым люди покидают компанию. Для изучения мотивов ухода большое значение имеет сбор и анализ информации о них. В первую очередь это сведения об общем числе уволившихся, долях сотрудников различных возрастных категорий, о работниках с низкой и высокой квалификацией, а также об их стаже работы и образовании	определение общих потерь времени как число месяцев, которое потребуется новому сотруднику на достижение того же уровня производительности, который был у замененного им работника	– учет в оценках трудоемкости издержек на обучение сотрудников; – уменьшение потерь от текучести кадров за счет привлечения на начальном этапе избыточного числа участников
Нарушение спецификаций	Политика ликвидации риска – подписание договора между заказчиком и компанией с описанием входных и выходных условий, оценка рисков независимыми экспертами		
Низкая производительность	улучшение организации производства и труда – повышение норм труда и расширения зон обслуживания; уменьшение числа сотрудников, не выполняющих нормы; упрощение структуры управления; повышение уровня специализации	структурные изменения – изменения удельных весов отдельных сотрудников в зависимости от трудоемкости выполняемой работы	повышение технического уровня производства – внедрение новых видов оборудования и программного обеспечения
Недостаточное внимание к проекту со стороны руководства компании	Проведение промежуточных отчетных собраний с приглашением заказчика	Промежуточная оценка прогресса разработки проекта руководством компании с составлением отчетности перед заказчиком	Составление договора между заказчиком и компанией с поэтапной сдачей проекта
Отсутствие мотивации персонала компании	Проведение тренингов, корпоративных собраний, внедрение методики team building	Премии персоналу в зависимости от успешности выполнения проектов	Распространение акций фирмы среди персонала, выделение разработчикам % от прибыли с проекта

тодологии Agile команда разработчиков должна состоять из высококвалифицированных специалистов, уход любого из которых наносит большой ущерб разработке проекта.

ВЫВОДЫ

Проведен качественный анализ рисков в зависимости от выбранной модели ПП и разработаны рекомендации по реагированию на выявленный риск.

Для RUP наиболее критичным оказался риск нарушения календарного планирования, для MSF – изменение требований, для Agile – текучесть кадров. Предложенные реакции позволят снизить негативное влияние рисков на конечные цели проектов и позволят выполнить проект в установленные сроки с заданным бюджетом.

СПИСОК ЛИТЕРАТУРЫ

1. Microsoft Solutions Framework. Дисциплина управления рисками MSF, вер. 1.1: [Электрон. ресурс]. – Режим доступа: <http://www.microsoft.com/rus/msf>.
2. PMBOK Руководство к Своду знаний по управлению проектами, 3-е изд. / PMBOK, Project Management Institute. – PMI, 2004. – 411 с.
3. Груздо, И. В. Повышение качества программного проекта за счет управления рисками / И. В. Груздо // Моделирование в экономике, организация производства та управління проектами. – 2009. – 1. – С. 141–146.
4. ДеМарко, Т. Вальсируя с медведями Управление рисками в проектах по разработке программного обеспечения / Том ДеМарко, Тимоти Листер, p.m. Office, 2005. – 190 с.
5. Брагина, Т. И. Сравнительный анализ итеративных моделей разработки программного обеспечения / Т. И. Брагина, Г. В. Табунщик // Радиоэлектроника, информатика, управління. – 2010. – № 2. – С. 130–139.
6. Bragina, T. Comparative Analysis of Software Development Models for Electro-technical Systems / T. Bragina, G. Tabunshchik // Proc. of Int. Conf. on Modern Problem of Radio Engineering, Telecommunications and Computer Science TCSET'2010, February 19–23, 2010, Lviv–Slavsko, Ukraine. – P. 347.
7. Брагина, Т. И. Классификация моделей итеративной разработки программного обеспечения / Т. И. Брагина, Г. В. Табунщик // Системный анализ. Информатика. Управление : материалы Всеукраинской научно-практической конференции, САИУ–2010, Март 04–05, 2010, Запорожье, Украина. – Запорожье : КПУ, 2010. – С. 23–25.
8. Галатенко, В. А. Управление рисками: обзор употребительных подходов / Галатенко В.А. // Информационный бюллетень. – №11(162). – 2006. – С. 1–15
9. Липаев, В. В. Анализ и сокращение рисков проектов программных средств / В.В. Липаев // Jet Info. – 2005. – №1. – С. 1–36.
10. Астахов, А. Как управлять рисками информационной безопасности? / А. Астахов // CISA. – ноябрь 2006. – С.121 – 124.
11. Петренко, С. Методики и технологии управления информационными рисками / С. Петренко, С. Симонов // IT Manager. – 2003. – №3. – С. 57–61.

Стаття надійшла до редакції 09.11.2010.

Після доробки 11.03.2011.

Брагіна Т. І., Табунщик Г. В.

АНАЛІЗ ПІДХОДІВ ДО КЕРУВАННЯ РИЗИКАМИ В ПРОГРАМНИХ ПРОЕКТАХ З ІТЕРАЦІЙНИМ ЖИТТЄВИМ ЦИКЛОМ

Виконано огляд головних ризиків програмних проектів з ітераційним життєвим циклом. Запропоновані рекомендації з управління виявленими ризиками в залежності від обраної моделі розробки програмних проектів. Виділено найбільш критичні ризики та проведено їх якісний аналіз.

Ключові слова: керування ризиками, програмні проекти, реакція на ризик.

Bragina T. I., Tabunshchik G.V.

ANALYSIS RISK MANAGEMENT APPROACHES IN SOFTWARE PROJECTS WITH AN ITERATIVE LIFECYCLE

In the article there is performed a review of the main risks in the software projects with an iterative lifecycle. The authors developed guidelines for the risks identified management, depended on the software development projects model. The most critical risks are identified and the qualitative analysis was made by the authors.

Key words: the risk management, the software projects, reaction on the risk.

УДК 004.75

Дьячук Т. С.

Асистент Запорозького національного технічного університета

ОЦЕНКА ХАРАКТЕРИСТИК РАСПРЕДЕЛЕННОЙ СИСТЕМЫ

Выполнен обобщенный анализ распределенной системы с точки зрения систем массового обслуживания. Определены характеристики, которые могут быть исследованы с помощью систем массового обслуживания.

Ключевые слова: система массового обслуживания, Grid, кластер, ресурс, задача.

ВВЕДЕНИЕ

Оптимизация ресурсов является одной из важных задач в современном мире. Существует множество видов

ресурсов: материальные, информационные, финансовые, интеллектуальные, трудовые, энергетические, времени и другие. Эффективное использование ресурсов

[1] в сфере высокопроизводительных вычислений не является исключением, так как вычислительная нагрузка на информационные системы постоянно растет. Получить максимальный результат при минимальных затратах стремятся как владельцы ресурсов, так и пользователи этих ресурсов. Для облегчения проектирования, контроля, планирования использования и анализа эффективности распределенной системы стремятся свести все измерители ресурсов к одному – например, денежному. При этом возникают трудности из-за отсутствия методов оценки некоторых видов ресурсов в денежном выражении. Поэтому при оценке эффективности использования ресурсов применяются несколько показателей, в совокупности отражающих уровень потребления ресурсов. В данной работе рассматривается распределенная система с помощью математического аппарата систем массового обслуживания. Данная модель является упрощенной в связи со сложностью реальных систем, но в целом позволяет получить характеристики, которые можно использовать для анализа свойств реальных распределенных систем.

ПОСТАНОВКА ЗАДАЧИ И ЦЕЛЬ РАБОТЫ

В настоящее время широко используется идея объединения ресурсов различных организаций в одну распределенную вычислительную систему в рамках Grid-проектов. В связи с чем возникает актуальная задача оптимального развития Grid, которая заключается в повышении производительности системы с одной стороны и минимизации затрат на сопровождение инфраструктуры с другой. В Grid-системе существуют два уровня планирования [2]: глобальный – распределение задач между кластерами (объединение машин в рамках одной организации) и локальный – распределение внутри кластера.

Пользовательские задачи распределяются между кластерами, учитывая информацию о задачах и текущем состоянии кластеров. Считаем, что одна задача выполняется на одном кластере, то есть не разбивается по нескольким. Подбирается кластер, с подходящими по характеристикам машинами и позволяющий решить задачу за минимальное время.

Анализируется запрос на выполнение задачи и рассылается всем кластерам, которые могут его выполнить. Каждый кластер на основании информации, содержащейся в запросе, своей текущей загрузки и уже составленного расписания возвращает время, до которого задача может быть выполнена. Принимается решение на основе этой информации и выбирается кластер, который удовлетворяет предельному времени выполнения задачи и имеет наименьшую стоимость (или удовлетворяет предельной стоимости и имеет минимальное время выполнения). Затем с выбранным кластером заключается соглашение и посылается задача на выполнение. Иерархическая система распределения задач имеет два уровня:

– распределение задач по кластерам (обслуживается брокерами);

– распределение многопроцессорных задач на кластере (обслуживается локальными планировщиками кластеров).

Миграция задач с одного кластера на другой также влияет на эффективность распределения. Задача может быть возвращена брокеру (с некоторым штрафом), если кластер получит больше выгодных задач от других брокеров или в случае сбоя. Брокер перераспределяет возвращенную задачу на другой кластер.

Выбор ресурсов и распределение вычислительной нагрузки осуществляется таким образом, чтобы удовлетворить все требования пользователей и оптимизировать общее время выполнения и стоимость используемых ресурсов.

В данной работе предложена упрощенная модель оценки характеристик кластера рабочих станций.

МОДЕЛЬ ЗАГРУЗКИ КЛАСТЕРА

Рассмотрим процесс решения задач на кластере с точки зрения систем массового обслуживания [3]. Загрузка и получение результатов на каждой рабочей станции представляют собой случайные события.

Будем считать, что вероятность получения результата на рабочей станции в малом промежутке времени Δt пропорциональна этому промежутку, то есть

$$\Delta p(t) = p(t + \Delta t) - p(t) = [1 - p(t)]\lambda \Delta t,$$

где $p(t)$ – вероятность того, что за время t получено решение задачи; λ – коэффициент пропорциональности.

Выражение в квадратных скобках соответствует вероятности отсутствия решения в момент времени t .

При $\Delta t \rightarrow 0$ получим дифференциальное уравнение

$$dp(t) = [1 - p(t)]\lambda dt,$$

откуда

$$\frac{dp(t)}{1 - p(t)} = \lambda dt.$$

После интегрирования

$$\int_0^t \frac{dp(t)}{1 - p(t)} = \int_0^t \lambda dt,$$

получим

$$\ln[1 - p(t)] = -\lambda t \quad \text{или} \quad p(t) = 1 - e^{-\lambda t}.$$

Плотность распределения времени решения задачи соответственно равна

$$r(t) = \frac{dp(t)}{dt} = \lambda e^{-\lambda t},$$

то есть имеет экспоненциальный закон распределения.

В простейшем случае все рабочие станции имеют одинаковые характеристики, и плотности распределения

решения будут соответственно одинаковы. Величина λ отражает обобщенные характеристики рабочей станции и представляет собой интенсивность решения задач на станции.

Рассмотрим поток задач, поступающих на кластер. Будем считать вероятность поступления задачи в бесконечно малом промежутке времени Δt также пропорциональна величине этого промежутка, то есть

$$\Delta s(t) = s(t + \Delta t) - s(t) = [1 - s(t)]\mu\Delta t,$$

где $s(t)$ – вероятность того, что за время t получено решение задачи; μ – коэффициент пропорциональности.

Выражение в квадратных скобках соответствует вероятности отсутствия задач в момент времени t .

При $\Delta t \rightarrow 0$ получим дифференциальное уравнение

$$ds(t) = [1 - s(t)]\mu dt,$$

откуда

$$\frac{ds(t)}{1 - s(t)} = \mu dt.$$

После интегрирования

$$\int_0^t \frac{ds(t)}{1 - s(t)} = \int_0^t \mu dt,$$

получим

$$\ln[1 - s(t)] = -\mu t \text{ или } s(t) = 1 - e^{-\mu t}.$$

Плотность распределения времени поступления задачи соответственно равна

$$v(t) = \frac{ds(t)}{dt} = \mu e^{-\mu t},$$

то есть также имеет экспоненциальный закон распределения.

На рис. 1 приведен граф состояний системы.

Рассмотрим состояния рабочих станций кластера:

S_0 – состояние, когда все рабочие станции загружены решением задач;

S_1 – состояние, при котором на одной из рабочих станций завершено решение;

.....

S_i – состояние, когда решение завершено на i рабочих станциях;

.....

S_n – состояние завершения решений на всех рабочих станциях.

Обозначим через $p_0(t), p_1(t), \dots, p_i(t), \dots, p_n(t)$ вероятности нахождения системы в соответствующих состояниях. Тогда можно записать следующие дифференциальные уравнения

$$\frac{dp_0(t)}{dt} = -\lambda_{0,1}p_0(t) + \mu_{1,0}p_1(t);$$

$$\frac{dp_1(t)}{dt} = \lambda_{0,1}p_0(t) - (\mu_{1,0} + \lambda_{1,2})p_1(t) + \mu_{2,1}p_2(t);$$

.....

$$\frac{dp_i(t)}{dt} = \lambda_{i-1,i}p_{i-1}(t) - (\mu_{i,i-1} + \lambda_{i,i+1})p_i(t) + \mu_{i+1,i}p_{i+1}(t);$$

.....

$$\frac{dp_n(t)}{dt} = \lambda_{n-1,n}p_{n-1}(t) - \mu_{n,n-1}p_n(t);$$

$$\sum_{i=0}^n p_i = 1.$$

В граничном стационарном режиме при $t \rightarrow \infty$ будем иметь систему алгебраических уравнений

$$0 = -\lambda_{0,1}p_0(t) + \mu_{1,0}p_1(t);$$

$$0 = \lambda_{0,1}p_0(t) - (\mu_{1,0} + \lambda_{1,2})p_1(t) + \mu_{2,1}p_2(t);$$

.....

$$0 = \lambda_{i-1,i}p_{i-1}(t) - (\mu_{i,i-1} + \lambda_{i,i+1})p_i(t) + \mu_{i+1,i}p_{i+1}(t);$$

.....

$$0 = \lambda_{n-1,n}p_{n-1}(t) - \mu_{n,n-1}p_n(t);$$

$$\sum_{i=0}^n p_i = 1.$$

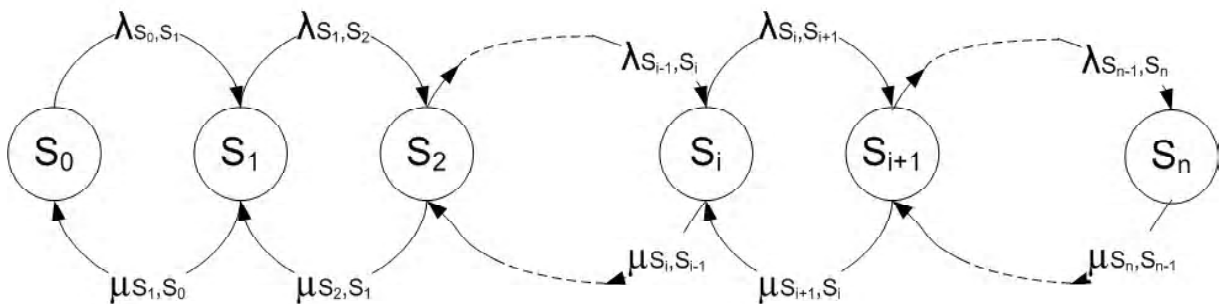


Рис. 1. Граф состояний системы

Для кластера, состоящего из n рабочих станций значения $\lambda_{i-1,i}$ и $\mu_{i,i-1}$ соответственно равны

$$\lambda_{i,i+1} = (n-i)\lambda; \quad \mu_{i+1,i} = (i+1)\mu, \quad i = 0, \dots, n-1.$$

Ожидаемое количество свободных рабочих станций кластера $S_{\text{св}}$ и средняя загрузка рабочих станций $S_{\text{заг}}$ равны соответственно

$$S_{\text{св}} = 1p_1 + 2p_2 + \dots + np_n;$$

$$S_{\text{заг}} = np_0 + (n-1)p_1 + \dots + p_{n-1}.$$

Справедливо также следующее утверждение $S_{\text{заг}} = n - S_{\text{св}}$, так как

$$S_{\text{св}} + S_{\text{заг}} = 1p_1 + 2p_2 + \dots + (n-1)p_{n-1} + np_n +$$

где

$$M = \begin{bmatrix} -n\lambda & \mu & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 \\ n\lambda & -(n-1)\lambda - \mu & 2\mu & \dots & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & (n-1)\lambda & -(n-2)\lambda - 2\mu & \dots & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -(n-i+1)\lambda - i\mu & (i+1)\mu & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & (n-i+1)\lambda & -(n-i)\lambda - (i+1)\mu & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & (n-i)\lambda & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 2\lambda & -\lambda - (n-1)\mu & n\mu \\ 1 & 1 & 1 & 1 & 1 & 1 & \dots & 1 & 1 & 1 \end{bmatrix}; P = \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_{i-1} \\ p_i \\ p_{i+1} \\ \vdots \\ p_{n-1} \\ p_n \end{bmatrix}; I = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \quad (1)$$

Определим значение вероятности p_i с помощью определителей. Учитывая особенности матрицы, отметим следующее свойство определителей: определитель треугольной матрицы равен произведению диагональных элементов. Кроме того, покажем, что произведению диагональных элементов матрицы будет иметь определитель вида:

$$\Delta = \begin{vmatrix} a_{11} & b_{12} & 0 & \vdots & 0 & 0 & 0 & 0 & 0 & \vdots & 0 & 0 & 0 \\ 0 & a_{22} & b_{23} & \vdots & 0 & 0 & 0 & 0 & 0 & \vdots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \vdots & \dots & \dots & \dots \\ 0 & 0 & 0 & \vdots & a_{i-2,i-2} & b_{i-2,i-1} & 0 & 0 & 0 & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \vdots & 0 & a_{i-1,i-1} & b_{i-1,i} & 0 & 0 & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \vdots & 0 & 0 & a_{ii} & 0 & 0 & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \vdots & 0 & 0 & 0 & a_{i+1,i+1} & 0 & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \vdots & 0 & 0 & 0 & b_{i+2,i+1} & a_{i+2,i} & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \vdots & 0 & 0 & 0 & 0 & b_{i+3,i} & \vdots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \vdots & 0 & 0 & 0 & 0 & 0 & \vdots & a_{n-2,n-2} & 0 & 0 \\ 0 & 0 & 0 & \vdots & 0 & 0 & 0 & 0 & 0 & \vdots & b_{n-1,n-2} & a_{n-1,n-1} & 0 \\ 0 & 0 & 0 & \vdots & 0 & 0 & 0 & 0 & 0 & \vdots & 0 & b_{n,n-1} & a_{nn} \end{vmatrix}. \quad (2)$$

Нетрудно видеть, что последовательное разложение определителей по элементам столбцов, начиная с последнего столбца до столбца $(i+1)$, которые представляют собой диагональные элементы, приведет к треугольной матрице, значение которой будет также произведение диагональных элементов. В результате получим:

$$\Delta = \prod_{i=1}^n a_{ii}.$$

Найдем определитель Δ матрицы M системы. Для этого преобразуем матрицу складыванием предыдущей строки с каждой последующей за исключением последней. В результате получим

$$\det M = \Delta = \begin{vmatrix} -n\lambda & \mu & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & -(n-1)\lambda & 2\mu & \dots & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & -(n-2)\lambda & \dots & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -(n-i+1)\lambda & (i+1)\mu & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & -(n-i)\lambda & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & -\lambda & n\mu \\ 1 & 1 & 1 & 1 & 1 & 1 & \dots & 1 & 1 & 1 \end{vmatrix}.$$

Выполним разложение определителя по элементам нижней строки. Алгебраические дополнения первого и последнего элементов этой строки представляют собой треугольные матрицы, алгебраические дополнения остальных элементов представляют определители вида (2).

Алгебраическое дополнение элемента $(n+1)$ строки и $(n+1)$ столбца представляет собой треугольную матрицу $n \times n$, диагональ которой содержит все отрицательные элементы. Значение определителя будет иметь знак положительный, если n имеет четное значение, и отрицательный, если n – нечетное. С учетом значений элементов диагонали будем иметь $(-1)^n n! \lambda^n$.

Рассмотрим алгебраические дополнения элементов $(n+1)$ строки и i -го столбца. Знак алгебраического дополнения для элементов $(n+1)$ строки и i -го столбца будет определяться степенью $(-1)^{n+1+i}$. С другой стороны, определитель будет иметь вид матрицы (2), причем значения элементов диагонали столбцов от 1 до $(i-1)$ будет отрицательным, а значения элементов диагонали столбцов от i до n – положительным. Следовательно, знак оп-

ределителя минора определится степенью $(-1)^{i-1}$. Отсюда знак алгебраического дополнения будет определяться произведением $(-1)^{n+1+i} \cdot (-1)^{i-1} = (-1)^{n+2i}$, то есть если n имеет четное значение, то положительный и отрицательный, если n – нечетное.

С учетом значений элементов диагонали будем иметь $(-1)^n \frac{n!}{i!} \lambda^{n-i} \frac{n!}{(n-i)!} \mu^i = (-1)^n n! C_n^i \lambda^{n-i} \mu^i$, где C_n^i – число сочетаний из n по i . Отсюда имеем

$$\Delta = (-1)^n n! \sum_{i=0}^n C_n^i \lambda^{n-i} \mu^i = (-1)^n n! (\lambda + \mu)^n,$$

где $\sum_{i=0}^n C_n^i \lambda^{n-i} \mu^i = (\lambda + \mu)^n$ согласно биному Ньютона.

При нахождении определителей Δ_i , у которых столбец i заменен на вектор свободных членов, который представлен 1 в $(n+1)$ строке и нулями для остальных строк, его значение и знак будет таким же, как и для алгебраического дополнения элементов $(n+1)$ строки и i столбца.

$$\Delta_i = \begin{vmatrix} -n\lambda & \mu & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 \\ n\lambda & -(n-1)\lambda - \mu & 2\mu & \dots & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & (n-1)\lambda & -(n-2)\lambda - 2\mu & \dots & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -(n-i+1)\lambda - i\mu & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & (n-i+1)\lambda & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 2\lambda & -\lambda - (n-1)\mu & n\mu \\ 1 & 1 & 1 & 1 & 1 & 1 & \dots & 1 & 1 & 1 \end{vmatrix}.$$

Складывая предыдущую строку с каждой последующей, за исключением последней, получим

$$\Delta_i = \begin{vmatrix} -n\lambda & \mu & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -(n-1)\lambda & 2\mu & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & -(n-2)\lambda & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -(n-i+1)\lambda & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & -\lambda & n\mu \\ 1 & 1 & 1 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 \end{vmatrix}.$$

После разложения по элементам столбца i получим

$$\Delta_i = (-1)^n n! C_n^i \lambda^{n-i} \mu^i.$$

Отсюда

$$p_i = \frac{\Delta_i}{\Delta} = C_n^i \frac{\lambda^{n-i} \mu^i}{(\lambda + \mu)^n}.$$

Оценим ожидаемое количество свободных станций $S_{св}$ для загрузки новых заданий:

$$S_{св} = \sum_{i=1}^n i C_n^i \frac{\lambda^{n-i} \mu^i}{(\lambda + \mu)^n}.$$

Так как $C_n^i = \frac{n}{i} C_{n-1}^{i-1}$, то можно записать

$$S_{св} = \sum_{i=1}^n i \frac{n}{i} C_{n-1}^{i-1} \frac{\lambda^{n-i} \mu^i}{(\lambda + \mu)^n} = \frac{n\mu}{(\lambda + \mu)} \sum_{i=1}^n C_{n-1}^{i-1} \frac{\lambda^{n-i} \mu^{i-1}}{(\lambda + \mu)^{n-1}}.$$

Выражение под знаком суммы представляет собой сумму вероятностей всех состояний системы для количества станций, равном $(n-1)$, то есть равно единице. В результате получим

$$S_{св} = \frac{n\mu}{(\lambda + \mu)}.$$

Количество загруженных станций $S_{заг}$ равно

$$S_{заг} = n - S_{св} = n - \frac{n\mu}{\lambda + \mu} = \frac{n\lambda}{\lambda + \mu}.$$

Готовность системы для принятия новых задач определяется вероятностью

$$G = 1 - \frac{\lambda^n}{(\lambda + \mu)^n}.$$

Вероятность отказа в обслуживании:

$$P_{отк} = \frac{\lambda^n}{(\lambda + \mu)^n}.$$

Для любой системы массового обслуживания, в которой заявка может обслуживаться одним каналом, среднее число заявок, обслуживаемых в единицу времени, определяется как произведение числа занятых каналов на плотность потока обслуживания. В нашем случае, количество задач K , решаемых в единицу времени, определится произведением $S_{заг}$ на величину λ . Таким образом, получим

$$K = S_{заг} \lambda = \frac{n\lambda^2}{\lambda + \mu}.$$

Рассмотрим случай разбиения каждой задачи на подзадачи, которые могут выполняться параллельно на различных рабочих станциях. Пусть каждая задача представлена в виде m подзадач, имеющих примерно одинаковые характеристики. Среднее время решения подзадачи будет в m раз меньше и, следовательно, интенсивность ее решения $\lambda_{n/3}$ будет в m раз больше, то есть

$$\lambda_{n/3} = m\lambda.$$

Поток $\mu_{n/3}$ поступающих подзадач также увеличится в m :

$$\mu_{n/3} = m\mu.$$

Количество подзадач, решаемых в единицу времени определится выражением

$$K_{n/3} = \frac{n\lambda_{n/3}^2}{\lambda_{n/3} + \mu_{n/3}} = \frac{n(m\lambda)^2}{m\lambda + m\mu} = \frac{nm\lambda^2}{\lambda + \mu}.$$

Если в кластере одновременно решается M задач, имеющих одинаковые приоритеты, то среднее количество подзадач, относящихся к одной задаче, решаемых в

единицу времени, равно

$$k = \frac{K_{n/3}}{M} = \frac{nm\lambda^2}{(\lambda + \mu)M}.$$

Среднее время решения задачи равно

$$\bar{t} = \frac{m}{k} = \frac{m(\lambda + \mu)M}{nm\lambda^2} = \frac{(\lambda + \mu)M}{n\lambda^2}.$$

Если поступление на решение новых подзадач рассматривать как следствие получения очередного решения, то $\lambda = \mu$. Таким образом, поток запросов на кластер меняется во времени в зависимости от интенсивности решения задач. Чем быстрее решаются задачи, тем сильнее увеличивается поток задач на глобальном уровне. Если есть простаивающие станции, то поток задач увеличивается. Среднее время на решение задачи определится выражением

$$\bar{t} = \frac{2M\lambda}{n\lambda^2} = \frac{2M}{n\lambda}.$$

Отсюда вытекает очевидный факт, что ожидаемое время на решение задачи прямо пропорционально количеству задач в системе и обратно пропорционально количеству рабочих станций в кластере.

Пусть каждая задача представлена различным количеством подзадач m_i , тогда общее количество подзадач равно

$$m_{\Sigma} = \sum_{i=1}^M m_i.$$

Будем считать, что подзадачи имеют одинаковые характеристики с точки зрения времени их решения, то есть интенсивность решения одинакова:

$$\forall (i = 1, M) \left(\frac{\lambda_i}{m_i} = \lambda_{n/3} = \text{const} \right),$$

где λ_i – интенсивность решения задачи i , величина обратная ожидаемому времени решения.

В этом случае среднее количество подзадач, решаемых в единицу времени для каждой задачи и определится стратегией планирования. Для случая, когда планирование осуществляется между подзадачами независимо от того, к каким задачам они относятся, среднее количество подзадач, решаемых в единицу времени, распределится пропорционально количеству подзадач в каждой задаче:

$$k_i = \frac{K_{n/3} m_i}{m_{\Sigma}},$$

а ожидаемое время решения задачи равно

$$\bar{t}_i = \frac{m_i}{k_i} = \frac{m_i m_{\Sigma}}{K_{n/3} m_i} = \frac{m_{\Sigma}}{K_{n/3}}.$$

Для данного случая время решения каждой задачи одинаковое.

Для кластера ожидаемое количество загруженных рабочих станций в этом случае будет иметь вид

$$S_{\text{заг}} = \frac{n\lambda_{n/3}}{\lambda_{n/3} + \mu_{n/3}},$$

а среднее количество подзадач, решаемых в единицу времени равно

$$K_{n/3} = S_{\text{заг}} \lambda_{n/3} = \frac{n\lambda_{n/3}^2}{\lambda_{n/3} + \mu_{n/3}}.$$

Если планирование осуществляется между задачами независимо от их объема (одни и те же ресурсы выделяются каждой задаче), то среднее время решения i -той задачи равно

$$\bar{t}_i = \frac{m_i}{K_{n/3}} = \frac{m_i (\lambda_{n/3} + \mu_{n/3})}{n\lambda_{n/3}^2}.$$

Из полученного выражения следует, что более короткие задачи будут решаться быстрее.

Оценим стоимость выполнения i -ой задачи

$$\text{Cost}_i = \bar{t}_i \cdot g,$$

где g – стоимость использования ресурсов кластера в единицу времени.

ВЫВОДЫ

В работе предложена упрощенная модель оценки характеристик кластера рабочих станций. Данная модель отражает основные закономерности работы многопроцессорной системы. Результаты исследований могут быть применены администраторами при настройке оборудования кластерных систем для повышения эффективности их использования. Система массового обслуживания аналогичного типа приведена в [3], только в нашем случае применена другая трактовка и расширены исследуемые характеристики системы.

Часто кластеры строятся из однородных элементов, поэтому принято допущение, что все рабочие станции кластера имеют приблизительно одинаковые характеристики. Если рассматривать случай, когда кластер строится на основе компьютеров с разными характеристиками, то количество состояний резко возрастает, и системы уравнений становятся очень громоздкими, что не позволяет получить закономерности в общем случае.

В будущей работе планируется исследовать поступление задач с разными характеристиками, а также рассмотреть возможность назначения приоритетов. Дальнейшее развитие связано также с объединением кластеров на глобальном уровне для изучения их взаимодействия.

СПИСОК ЛІТЕРАТУРИ

1. Дьячук, Т. С. Дослідження ефективності роботи алгоритму планування ресурсів для розподіленої системи / Т. С. Дьячук, Р. К. Кудерметов, А. М. Літава, Р. І. Сметанін // Вісник КДПУ імені Михайла Остроградського. – Кременчук : КДУ, 2010. – Вип. 1/2010 (60) частина 1. – С. 47–50.
2. Эвристики распределения задач для брокера ресурсов Grid [Электронный ресурс] / А. И. Аветисян, С. С. Гайсарян, Д. А. Грушин, Н. Н. Кузюрин, А. В. Шокуров // Труды Института системного программирования РАН – 2004. – Режим доступа: <http://citforum.univ.kiev.ua/nets/digest/grid/index.shtml> – Загл. с экрана.
3. Клейнрок, Л. Теория массового обслуживания. Пер. с англ. / Клейнрок, Л. ; пер. И. И. Грушко; ред. В.И. Нейман. – М. : Машиностроение, 1979. – 432 с.

Стаття надійшла до редакції 17.03.2011.

УДК 004.9:004.82

Дьячук Т. С.

ОЦІНКА ХАРАКТЕРИСТИК РОЗПОДІЛЕНОЇ СИСТЕМИ

Виконано узагальнений аналіз розподіленої системи з точки зору систем масового обслуговування. Визначено характеристики, які можуть бути досліджені за допомогою систем масового обслуговування.

Ключові слова: система масового обслуговування, Grid, кластер, ресурс, завдання.

Diachuk T. S.

EVALUATION OF DISTRIBUTED SYSTEM'S CHARACTERISTICS

The distributed system in terms of queueing systems is analyzed. The characteristics that can be studied using queueing systems are presented.

Key words: queueing system, Grid, cluster, resource, task.

Гонтарь Н. А.¹, Кудерметов Р. К.²

¹Ассистент Запорожского национального технического университета

²Канд. техн. наук, профессор Запорожского национального технического университета

РАЗРАБОТКА ОНТОЛОГИИ СИСТЕМНОГО ИНЖИНИРИНГА КОСМИЧЕСКИХ СИСТЕМ

В статье обоснована необходимость создания онтологии системного инжиниринга (СИ). Онтология СИ вместе с переходом к модели-ориентированному СИ повысит эффективность разработки сложных систем. Предложен прототип онтологической модели, основанный на дескриптивной логике и формализованный с помощью языка OWL DL.

Ключевые слова: системный инжиниринг, онтология, MBSE, дескриптивная логика, OWL.

ВВЕДЕНИЕ

Согласно определению Международного Совета по Системной Инженерии (International Council on Systems Engineering, INCOSE) системный инжиниринг (СИ) – это междисциплинарный подход и средства для создания успешных систем [1, 2]. На протяжении всего жизненного цикла (ЖЦ) систем (концептуальное проектирование, разработка, изготовление, испытания, эксплуатация и утилизация) СИ фокусирует внимание разработчиков на глубоком анализе и отслеживании потребностей пользователей создаваемой системы и функциональных требованиях к ней. СИ интегрирует дисциплины и коллективы разработчиков в единое распределенное виртуальное пространство проекта, в котором согласованы процессы разработки и изготовления системы, взаимодействия разработчиков, применяемые ими понятия, правила, методы, инструменты и т. д.

СИ – итерационный процесс и каждое новое, принятое в ходе ЖЦ системы, решение отражается на конечном результате разработки. Поэтому важным является создание на основе современных информационных технологий благоприятной среды, отражающей текущие

процессы СИ, множество моделей системы, возможные варианты решений и ошибки.

В одной из последних инициатив INCOSE – «Systems Engineering Vision 2020» декларируются принципы модели-ориентированного подхода к СИ, который получил название MBSE (Model-based Systems Engineering) [3, 4]. MBSE – это формализованное применение моделирования для поддержки системных требований, разработки, анализа, верификации и валидации систем на фазе концептуального проектирования и далее при разработке и последующих фазах ЖЦ системы. Основными задачами MBSE являются:

- обеспечить согласованное взаимодействие между заказчиками и разработчиками систем, а также между коллективами разработчиков;
- уменьшить риски и неопределенности и контролировать их в течение всего ЖЦ системы;
- поддерживать процессы контроля границ системы, ее агрегации, декомпозиции, интеграции, сертификации и прогнозирования функциональных возможностей;
- обеспечить согласованность множества системных аспектов, видений и моделей системы;

- обеспечить обзоримость и возможность отслеживания состояния разработки больших и сложных систем;
- за счет ясного, определенного представления системы в виде моделей, устранять неоднозначности при участии разработчиков с разными культурами и языками и использовании ими различных средств проектирования;
- обеспечить необходимое понимание вклада всех составных частей системы в ее миссию.

В качестве языка моделирования для СИ и MBSE перспективным считается язык SysML, построенный на базе языка UML и принятый как индустриальный стандарт консорциумом Object Management Group (OMG) в 2007 г. [5]. Язык SysML включает не только все возможности языка UML моделировать программные системы, но и дополняет его диаграммами, необходимыми для моделирования любых технических систем, в частности диаграммами требований, блоков (рис. 1) и параметрической диаграммой. Однако следует отметить, что SysML включает много избыточных и практически неиспользуемых диаграмм, а также, из-за неоднозначности некоторых конструкций и наличия в спецификации неформальных описаний, язык не всегда позволяет точно выразить семантические особенности системы.

ПОСТАНОВКА ЗАДАЧИ

В настоящее время интенсивно создаются онтологии различных предметных областей (ПрО) и исследуются возможности их использования в MBSE. В научных и промышленных сообществах системных инженеров сформировалось видение развития MBSE и интеграции его с онтологическим подходом, которое предполагает последовательное развитие онтологий от прототипов до стандартов (рис. 2) [6].

Для успешного перехода от документо-ориентированного СИ к модели-ориентированному СИ (т. е. к MBSE) и эффективного его использования в различных ПрО необходимо создание множества моделей самого MBSE и охватываемых им процессов. В частности, для однозначного понимания командами разработчиков этапов, процессов и задач MBSE, правильной обработки его артефактов приложениями необходимо создание машинно-интерпретированных представлений всех составляющих СИ и MBSE. Решению этой задачи будет способствовать создание онтологий СИ и MBSE. В данной работе ставится цель обосновать возможность создания таких онтологий.

ОНТОЛОГИЯ СИСТЕМНОГО ИНЖИНИРИНГА

Онтология формальная спецификация разделяемой концептуальной модели. Здесь под «концептуальной» моделью подразумевается абстрактная модель, описывающая систему понятий ПрО. «Разделяемая» модель означает согласованное ее понимание определенным сообществом (группой людей). «Спецификацией» является описание системы понятий в явном виде. «Формальность» модели подразумевает машиночитаемое представление данных [7].

Создание онтологии СИ позволит получить в распоряжение системных инженеров такую модельную стратегию СИ, которая объединит совокупность понятий, правил, отношений, данных и информации СИ, независимых от отдельных стандартов и терминов и не скрытых в программном коде. Если в MBSE SysML обеспечивает синтаксическую совместимость моделей процессов и систем, то онтология – их семантическую интероперабельность.

Онтология СИ не будет нести в себе новых знаний, но она формализует наработанные знания и явные заключе-

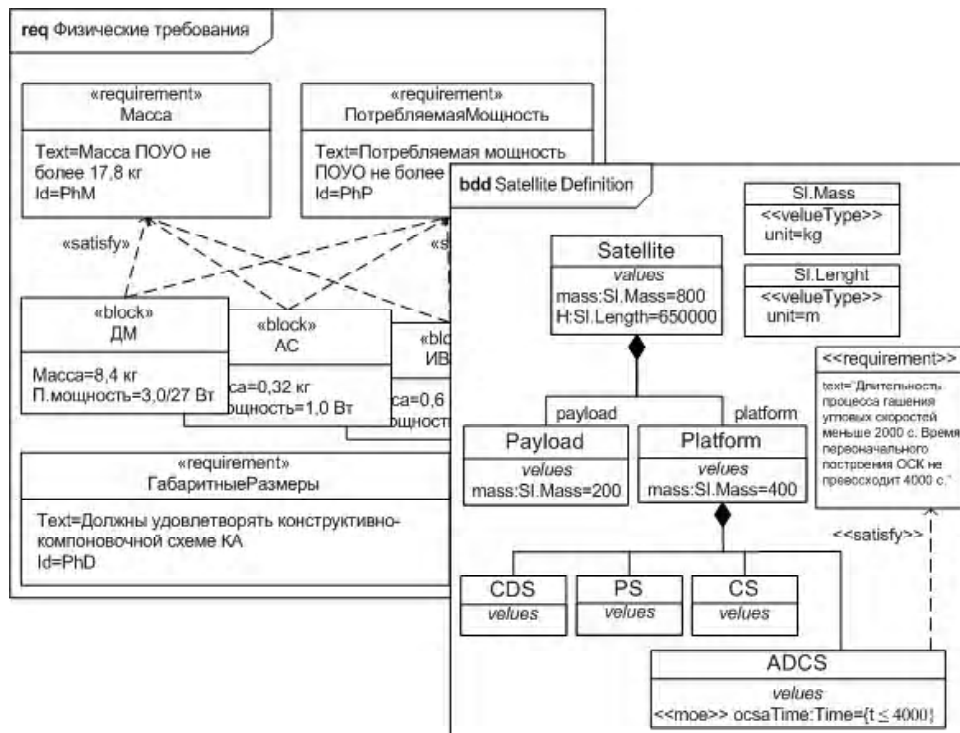


Рис. 1. Диаграммы требований и блоков для модели КС на языке SysML

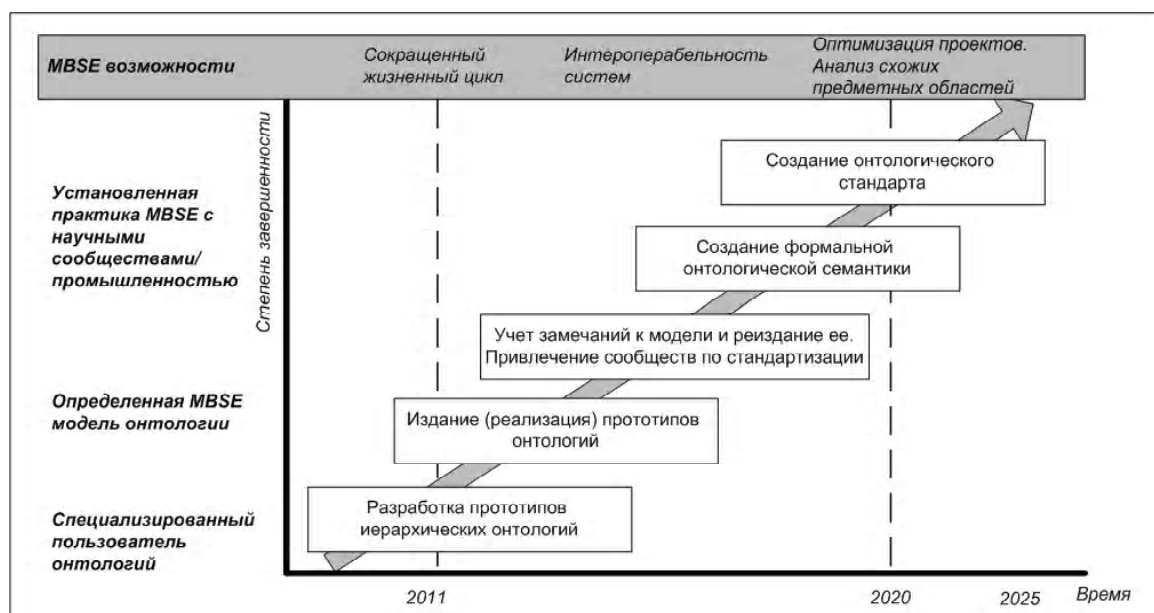


Рис. 2. Место онтологии в развитии СИ

ния и обеспечит их согласованность. Эти знания будут элементарными для восприятия системными инженерами и интероперабельными в распределенных компьютерных системах. Возможно, первые экземпляры онтологии СИ будут избыточными, громоздкими и противоречивыми в силу большого множества дисциплин самого СИ и дисциплин, которыми он управляет на протяжении ЖЦ системы.

В настоящее время имеются огромный международный опыт и знания в области создания космических систем (КС) и, поэтому становится вопрос об эффективном хранении и обработке этих знаний. Онтологическое представление СИ КС позволит специфицировать структуру и семантику терминов двух предметных областей, выразить различные формы сложных ограничений целостности и будет поддерживать работу с распределенными ресурсами, формализованную базу единой терминологии для разработчиков, автоматическую обработку запросов, интеграцию ее или в нее других специфицированных формализованных прикладных ПрО. На рис. 3 показаны возможные области использования онтологий и в частности онтологии СИ КС.

Определение характеристик КС обычно идет от функционального описания к физическому. Физическое описание формализуется в дереве продукта, которое представляет собой структуру, исходящую из упорядоченного и исчерпывающе разбитого на составные части конечного продукта, представленного в виде последовательности уровней, на которых расположены его составляющие [8, 9]. Эти составляющие конечного продукта меняют свою сущность по мере разбиения от верхнего уровня до нижних уровней декомпозиции. Как правило, они становятся менее сложными функционально, более компактными и самодостаточными в физическом смысле, включая при этом меньшее число используемых технологий. Поэтому онтология СИ КС должна отражать не только процессы СИ, но и техническую и структурную декомпозиции КС.

Разработку принципов СИ и MBSE для КС осуществляют организации NASA, ECSS, INCOSE, OMG и др. Стандарты, спецификации и рекомендации этих организаций должны стать источником информации для формирования понятий онтологии СИ КС. В настоящее время работы по инженерии знаний и онтологии в области КС проводятся NASA в рамках проекта NEXIOM (NASA Exploration Information Ontology Model) [10]. Проект NEXIOM разрабатывается с целью обеспечения:

- возможности отслеживать все изменения системы и ее элементов на протяжении всего ЖЦ (трассируемость);
- возможности проводить полнофункциональный анализ системы;
- поддержки интерпретации и доступа к информации, как со стороны разработчиков, так и программных агентов.

Данный проект должен регулировать техническую, экономическую и организационную части процесса разработки систем, вплоть до стандартизации документооборота и рабочего времени инженеров. NEXIOM нацелен на создание модели, которая будет содержать данные и знания о существующих стандартах хранения, организации и обработки данных, методах и средствах моделирования, опыте предыдущих успешных разрабо-

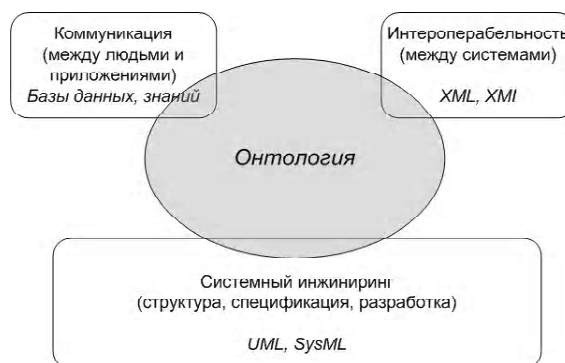


Рис. 3. Область применения онтологии СИ КС

ток систем. Предполагается, что стандарты и спецификации NEXIOM будут развиваться во времени с учетом потребностей разработчиков. Одной из целей рабочей группы проекта NEXIOM является поиск экономически эффективных способов обмена информацией между NASA и промышленностью и ожидается, что онтологии будут играть в этом ключевую роль.

Эффективность создания онтологий возрастает за счет того, что онтологии ПрО могут разрабатываться независимо, а затем проходить процессы согласования и адаптации между собой. Поэтому, учитывая, что проект NEXIOM направлен на организацию и формализацию знаний всего процесса создания КС, целесообразно создавать онтологии отдельных дисциплин этого процесса, в частности СИ КС, как дисциплины и процесса, интегрирующего множество других дисциплин и команд разработчиков на протяжении всего ЖЦ систем.

В качестве инструмента для создания онтологии СИ КС предполагается использовать программную среду проектирования онтологий *Protégé*, которая включает редактор онтологий, позволяющий проектировать онтологию, разворачивая таксономическую структуру абстрактных или конкретных классов и слотов. Структура онтологии создается аналогично иерархической структуре каталога. На основе сформированной онтологии *Protégé* может генерировать формы получения знаний для введения экземпляров классов и подклассов. *Protégé* позволяет вводить в онтологию классы методом заполнения форм, составленных из элементов мета-описания.

Одним из важных шагов при проектировании онтологии является распределение классов в таксономию. Смысловые подклассы одного и того же класса в онтологии должны находиться на одном уровне. Подклассов должно быть не меньше двух и не больше двенадцати, иначе, по мнению разработчиков *Protégé*, можно выделить еще один подкласс.

Основные концепты (понятия) СИ КС будут представлять собой объекты онтологии, реализованные как классы с характерными атрибутами. Такой способ представления информации позволяет работать как с СИ в целом, так и с ее структурными элементами в отдельности. Мы используем способ нисходящей разработки, который начинается с определения самых общих понятий ПрО СИ КС с последующей их структуризацией. Наиболее фундаментальные понятия СИ КС будут классами, которые находятся в верхней части таксономического дерева, корнем которого является класс «*THING*», охватывающий все множество концептов онтологии.

На рис. 4 показан простейший пример, демонстрирующий возможность создания онтологии СИ КС с помощью *Protégé*. В качестве источников данных для этого примера использовались основные понятия, отношения между ними и атрибуты, изложенные в стандартах ECSS, NASA, INCOSE и публикациях, посвященных разработке КС.

Не существует единственного «правильного» способа или методологии разработки онтологий, каждый аналитик составляет план и этапы онтологического анализа,

опираясь на свой опыт и существующие проекты онтологий. Тем не менее, можно выделить некоторые фундаментальные правила разработки онтологии [11]:

- необходимо рассматривать множество альтернативных способов моделирования и моделей ПрО. Лучшее решение почти всегда зависит от предполагаемого приложения и ожидаемых расширений;

- разработка онтологии – это обязательно итеративный процесс;

- понятия в онтологии должны быть близки к объектам (физическим или логическим) и отношениям в интересующей нас ПрО. Скорее всего, это существительные (объекты) или глаголы (отношения) в предложениях, которые описывают данную ПрО.

ФОРМАЛИЗАЦИЯ ОНТОЛОГИИ СИ С ПОМОЩЬЮ ДЕСКРИПТИВНОЙ ЛОГИКИ

Для выражения фундаментальных понятий при формализации данных о сложных системах и их концептуальном моделировании целесообразно использовать дескриптивные логики (ДЛ). ДЛ состоят из двух уровней: интенционального и экстенционального. Интенциональный уровень определяет множество понятий и отношений, описывающих концептуальную структуру ПрО (классы, свойства, отношения, аксиомы). Экстенциональный уровень содержит множество экземпляров элементов, описанных в интенциональном уровне [12]. С помощью ДЛ определяются основные элементы онтологии, которые позволяют отделить знания о ПрО от оперативных знаний и опыта.

Существует несколько диалектов дескриптивной логики, которые отличаются выразительными возможностями. Для описания онтологии СИ КС выбран диалект AL, который считается базовым для ДЛ и содержит необходимые и достаточные конструкции для описания созданного безэкземплярного прототипа онтологии, в котором имеются: T – концепт «*THING*»; $\perp$ – концепт «*NOTHING*»; $\cap$ – логическая связка (конъюнкция); $\exists$ – квантор существования; $\forall$ – квантор ограничение значения; $\neg$ – логическая связка (отрицание).

Концепты онтологии определяются индуктивно следующим образом:

- все имена концептов, T и $\perp$ являются AL-концептами;

- если A является AL-концептом, то и $\neg A$ является AL-концептом;

- если $A0$ и $A1$ являются AL-концептами, а r – имя роли, то $A0 \cap A1$, $\exists r.A0$, $\exists r.A1$, $\forall r.A0$, $\forall r.A1$ являются AL-концептами.

Онтологическая модель описывается унарными и бинарными предикатами. Унарные предикаты представляют собой фундаментальные понятия и соответствуют классам, которые находятся в корне таксономического дерева. Например, для онтологии СИ КС определены унарные предикаты: «*SYSTEMS – ENGINEERING*» – концепт абстрактных классов онтологии, «*REQUIREMENT* –

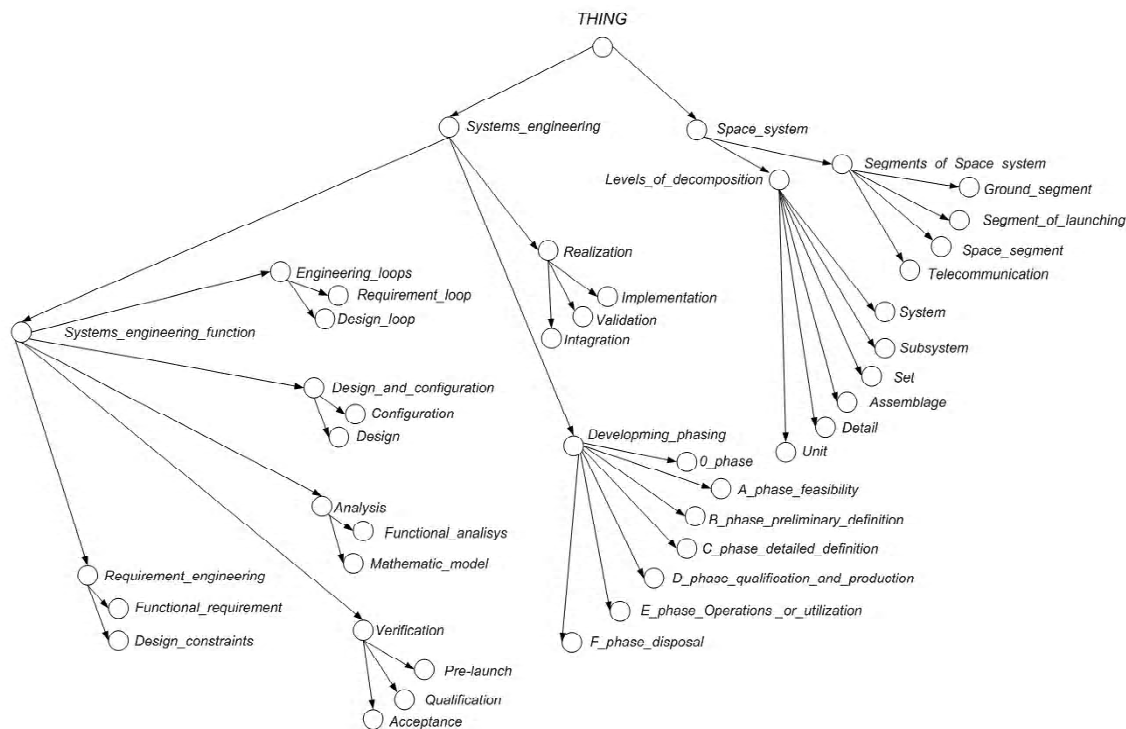


Рис. 4 . Таксономия классов онтологии СИКК

ENGINEERING» выделенное понятие Про СИ, определенное как класс онтологии.

Бинарные предикаты – это роли. «*hasInstance*» бинарный предикат, определяющий понятие того, что с концептом, стоящим справа от роли, имеет место существование такого отношения «*hasInstance.A*».

С помощью предикатов в онтологиях строятся отношения и аксиомы. Например, следующая формула показывает отношение «*существование концепта, определенного как экземпляр класса FUNCTIONAL – REQUIREMENT*»:

$$\exists \text{hasInstance.FUNCTIONAL – REQUIREMENT} \quad (1)$$

В онтологиях всегда predeterminedены два класса «THING» и «NOTHING». «THING» это множество всех концептов онтологии. «NOTHING» это пустое множество концептов. Таким образом, если мы введем концепт «SYSTEMS – ENGINEERING», то он будет экземпляром класса:

$$\text{SYSTEMS – ENGINEERING} \exists \text{hasInstance.T} \quad (2)$$

Для части онтологии, представляющей онтологию дисциплины «инженерия требований», СИ КС в качестве концептов или унарных предикатов и прием и, которые будут выделенными понятиями СИ и определенными как классы онтологии.

В самом общем случае, терминологические аксиомы первого рода (включений) (3) и второго рода (равенств) (4) имеют вид:

$$A0 \sqsubseteq A1, r0 \sqsubseteq r1; \quad (3)$$

$$A0 \equiv A1, r0 \equiv r1. \quad (4)$$

Для онтологии СИ КС с помощью аксиом первого рода определим таксономию концептов онтологии. Например, класс «REQUIREMENT – ENGINEERING» является подклассом «SYSTEMS – ENGINEERING», а класс, в свою очередь, является надклассом класса:

$$\begin{aligned} \text{«REQUIREMENT – ENGINEERING»} &\sqsubseteq \\ &\sqsubseteq \text{«SYSTEMS – ENGINEERING»}, \end{aligned} \quad (5)$$

$$\begin{aligned} \text{«FUNCTIONAL – REQUIREMENT»} &\sqsubseteq \\ &\sqsubseteq \text{«REQUIREMENT – ENGINEERING»}. \end{aligned} \quad (6)$$

Таким образом, ДЛ позволяет описать модель онтологии системного инжиниринга с точки зрения математики, создать устойчивые непротиворечивые формулировки термов онтологии. ДЛ обладает мощным аппаратом формализации понятий и при дальнейшей работе с онтологией позволяет рассматривать все элементы как с точки зрения их семантики, так и со стороны логических законов.

Несмотря на существующее множество технологий и моделей хранения, интерпретации и обработки данных, онтология усиливает возможности концептуального моделирования. Работа с онтологической моделью реализует абстрактное, единое, семантическое моделирование Про (интенциональный уровень), независящее от спецификации этой модели (экстенциональный уровень) и служащее в качестве средства создания эталонного инструмента взаимодействия разработчиков между собой и между программными приложениями.

Следует учитывать, что создание онтологии требует экспертных знаний в исследуемой Про и значительных

затрат времени. Онтология содержит формализованные данные, но не отображает методы, функции и процессы ПрО, поэтому онтология СИ будет являться лишь универсальным источником данных о ПрО СИ.

OWL-ОНТОЛОГИЯ СИ

Для формального машинного представления онтологий Консорциум W3C предложил язык описания онтологий OWL (Ontology Web Language), который имеет различные по выразительности диалекты, спроектированные для использования отдельными сообществами разработчиков и пользователей [13].

Диалект OWL DL наиболее популярен и используется при необходимости представления знаний о ПрО в максимально выразительном формализованном виде без потери полноты вычислений (все заключения гарантировано будут вычисляемыми) и разрешимости (все вычисления гарантированно завершатся за определенное время). Формальной основой онтологий на языке OWL DL является ДЛ, т.е. логика предикатов, предназначенная для представления терминологического знания о ПрО. OWL DL включает все языковые конструкции OWL с ограничениями вроде разделения типа (класс не может быть частным свойством, а свойство не может быть индивидом или классом).

Каждый индивид в OWL представлении является членом класса `owl:Thing`. Таким образом, каждый определенный класс автоматически является подклассом `owl:Thing`. Специфичные для данной ПрО корневые классы (понятия) определяются простым объявлением именованного класса. OWL также определяет пустой класс `owl:Thing`.

Упомянутая выше программная среда проектирования онтологий *Protégé* позволяет разрабатывать OWL-онтологии, а также автоматически их генерировать из ранее созданных таксономий классов. Ниже приведены фрагменты сгенерированных в *Protégé* OWL-описаний концептов, определений классов и аксиом онтологии СИКС. Так, введенные

нами выше классы «*SYSTEMS – ENGINEERING*» и «*REQUIREMENT – ENGINEERING*», определяются следующим образом:

```
<owl:Class rdf:ID="Systems_engineering"/>
<owl:Class rdf:ID="Requirement_engineering"/>
```

Определение общего класса онтологии «*SYSTEMS – ENGINEERING*» и принадлежность его классу «*THING*» (1) выражается так:

```
<rdfs:domain>
<owl:Class>
<owl:unionOf rdf:parseType="Collection">
<rdf:Description rdf:about=
"http://www.w3.org/2002/07/owl#Thing"/>
<owl:Class rdf:about="#Systems_engineering"/>
</owl:unionOf>
</owl:Class>
</rdfs:domain>
```

Аксиомы включения (5) и (6) онтологии СИКС, записываются в виде:

```
<owl:Class rdf:about="#Functional_requirement">
<rdfs:subClassOf
rdf:resource="#Requirement_engineering"/>
</owl:Class>
<owl:Class rdf:about="#Requirement_engineering">
<rdfs:subClassOf
rdf:resource="#Systems_engineering"/>
</owl:Class>
```

OWL-онтологию (документ, описанный на языке OWL) СИКС при необходимости можно преобразовать в XML-документ и использовать его для обмена в распределенной компьютерной системе поддержки СИ. С другой стороны, OWL-документ можно конвертировать в XMI-формат для обмена моделями и, затем, в модель на языке SysML (UML). Следовательно, в зависимости от целей конкретных процессов СИ и на различных этапах ЖЦ системы можно использовать языки OWL и SysML (рис. 5)

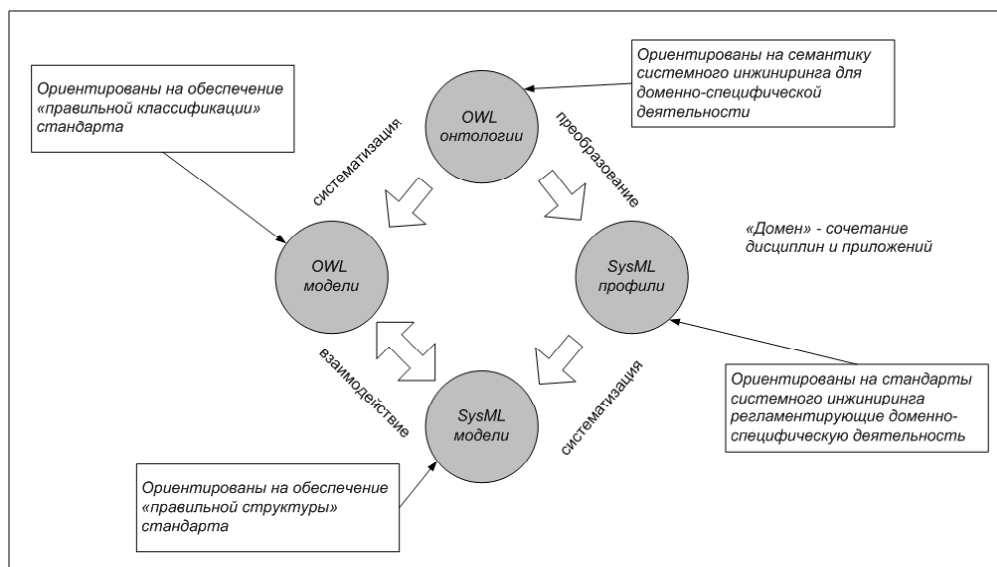


Рис. 5. Взаимодействие между OWL и SysML моделями

ВЫВОДЫ

Таким образом, в работе обоснована необходимость создания онтологии СИ КС, возможность ее формализации с помощью диалекта AL дескриптивной логики и представления ее на языке OWL DL. На примере разработанного прототипа продемонстрированы основные принципы, которые будут положены в основу дальнейшей разработки онтологии СИ КС. В качестве инструментального средства проектирования онтологии выбран пакет *Protégé*. Онтология СИ КС будет интегрирована в процесс MBSE, что позволит повысить эффективность использования MBSE для разработки сложных систем.

СПИСОК ЛИТЕРАТУРЫ

1. Systems Engineering Handbook [Text] / INCOSE. – TP-2003-016-02, Version 2a, 1 June 2004. – USA: INCOSE, 2004. – 300 p.
2. ГОСТ Р ИСО/МЭК 15288 – 2005. Системная инженерия. Процессы жизненного цикла систем [Текст]. – Введен 2005-12-29. – М.: Изд-во стандартов, 2006. – 57 с.
3. Systems Engineering Vision 2020 [Электронный ресурс] / INCOSE-TP-2004-004-02, September, 2007. – Режим доступа: http://www.incose.org/ProductsPubs/pdf/SEVision2020_20071003_v2_03.pdf
4. Kossiakoff A. Systems engineering: principles and practice [Text] / Alexander Kossiakoff et al. – 2nd ed. (Wiley series in systems engineering and management; 67) – WILEY, 2011. – 559 p.
5. OMG Systems Modeling Language (OMG SysML™) [Электронный ресурс]. – Режим доступа: www.sysml.org/docs/specs/OMGSysML-v1.2-10-06-02.pdf
6. Ontology in MBSE / INCOSE // International Council on Systems Engineering IW2011 MBSE Workshop, January 29, 2011. – USA, 2011. – 11 p.
7. Gruber, Thomas R. A translation approach to portable ontology specifications [Text] / Thomas R. Gruber // Appeared in Knowledge Acquisition. – 1993. – №5(2). – P. 199–220.
8. ECSS-E-ST-10C. Space engineering. System engineering general requirements [Text]. – 6 March 2009. – ECSS Secretariat SA-ESTEC, Requirements & Standards Division. Noordwijk, Netherlands, 2009. – 100 p.

9. Systems Engineering Handbook [Text] / Aeronautics and Space Administration, Headquarters Washington DC 20546, SP-2007-6105 Rev1. – Washington, 2007. – 360 p.
10. NASA Exploration Information Ontology Model (NEXIOM) primer and vision [Text] / National Aeronautics and Space Administration, Headquarters Washington DC 20546. – Washington, 2005. – 18 p.
11. Noy N. Ontology Development 101: A Guide to Creating Your First Ontology [Электронный ресурс] / Natalya F. Noy, D. L. McGuinness // Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001. – Режим доступа: http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html.
12. Baader F. The Description Logic Handbook [Text] / F. Baader et al. – Cambridge University Press, 2003. – 574 p.
13. OWL Web Ontology Language Guide / W3C [Электронный ресурс]. – Режим доступа: <http://www.w3.org/TR/owl-guide/>

Стаття надійшла до редакції 23.05.2011.

Гонтар Н. А., Кудерметов Р. К.

РОЗРОБКА ОНТОЛОГІЇ СИСТЕМНОГО ІНЖИНІРИНГУ КОСМІЧНИХ СИСТЕМ

У статті обґрунтована можливість створення онтологій системного інжинірингу (СІ). Онтологія СІ разом з переходом до моделі-орієнтованого СІ підвищить ефективність розробки складних систем. Запропоновано прототип онтологічної моделі, створений за допомогою дескриптивної логіки і формалізований за допомогою мови OWL DL.

Ключові слова: системний інжиніринг, онтологія, MBSE, дескриптивна логіка, OWL.

Gontar N. A., Kudermetov R. K.

WORKING OUT ONTOLOGY OF SYSTEMS ENGINEERING OF SPACE SYSTEM

This article grounded the need to create ontology of systems engineering (SE). SE ontology with the transition to model-based approach SE will increase the efficiency of development of complex systems. The prototype of the ontology model based on description logic and formalized through language OWL DL is proposed.

Key words: systems engineering, ontology, MBSE, description logic, OWL.

УДК 004.627: 004.272.26

Скрупский С. Ю.¹, Луценко Н. В.², Скрупская Л. С.³

^{1, 3} Ассистент Запорожского национального технического университета

² Старший преподаватель Запорожского национального технического университета

ПАРАМЕТРЫ КОМПРЕССИИ ВИДЕОИНФОРМАЦИИ В РАСПРЕДЕЛЕННЫХ СИСТЕМАХ

Проанализированы основные параметры компрессии видеоинформации в распределенных системах. Экспериментально исследовано влияние этих параметров на коэффициент сжатия, уровень искажения и время сжатия видеоинформации в распределенной системе.

Ключевые слова: видеоинформация, распределенная система, компрессия, битрейт, уровень искажения.

ВВЕДЕНИЕ

В настоящее время видеоинформацию хранят в архивах в сжатом цифровом виде, поскольку в несжатом

виде она занимает значительные объемы памяти [1]. Проблема хранения видеоинформации особенно актуальна с распространением стандарта телевидения высо-

кой четкости (HDTV) [2], так как видеоинформация, удовлетворяющая такому стандарту, занимает существенно большие объемы памяти по сравнению с видеоинформацией, удовлетворяющей стандарту SDTV [3, 4]. Современные методы сжатия видеоинформации обладают высокой вычислительной сложностью [5], поэтому для достижения приемлемого для пользователя времени сжатия компрессию видеоинформации выполняют при помощи параллельных и распределенных компьютерных систем [6–8].

ПОСТАНОВКА ЗАДАЧИ

При сжатии видеоинформации в распределенной системе преследуется цель получения качественного результата с хорошим коэффициентом сжатия за относительно приемлемое для пользователя время. Достижение этой цели возможно как путем повышения производительности применяемого в распределенной системе оборудования, так и путем варьирования параметров кодера, осуществляющего компрессию видеоинформации.

Рассмотрим видеопоследовательности, разделяемые на части при помощи метода [9] и сжимаемые на узлах распределенной системы. Специфика, таких видеопоследовательностей заключается в том, что после разделения они ограничены одной сценой и их смежные кадры характеризуются значительной корреляционной зависимостью. Следовательно, наиболее трудоемкая операция оценки движения блоков в алгоритмах сжатия может быть выполнена на относительно небольшой ширине области поиска векторов движения без значительного ущерба качеству сжатия. В данной работе ставится задача экспериментально проверить эту гипотезу, а так же исследовать влияние параметров сжатия видеоинформации, разделяемой на части при помощи метода [9] и сжимаемой на узлах распределенной системы, на показатели результата компрессии.

Суть рассматриваемого метода разделения видеоинформации на части в распределенных системах заключается в вычислении коэффициентов корреляции между всеми смежными кадрами видеопоследовательности. Затем для коэффициента корреляции текущей пары кадров вычисляется значение функции чувствительности к смене сцен анализатора видеопоследовательности. Если значение функции чувствительности превышает значение модуля коэффициента корреляции, то видеопоследовательность разделяется на две части. Первая часть содержит кадры исходной видеопоследовательности до определенного с помощью функции чувствительности места слабой межкадровой корреляции, а вторая видеопоследовательность – с кадрами после места слабой межкадровой корреляции. Если же значение функции чувствительности меньше значения модуля коэффициента корреляции, тогда вычисляется значение функции чувствительности для следующей пары смежных кадров, которое сравнивается с соответствующим значением модуля коэффициента корреляции.

АНАЛИЗ ПАРАМЕТРОВ КОМПРЕССИИ ВИДЕОИНФОРМАЦИИ

Рассмотрим основные параметры компрессии видеоинформации:

- битрейт – количество бит, формируемых в единицу времени для кодирования видеопоследовательности;
- ширина области поиска векторов движения операции оценки движения (ОД) блоков кадра, которую обозначим как Ω ;
- количество восстановленных кадров операции ОД блоков кадра (*ReFrames*).

Для видеопоследовательностей, удовлетворяющих стандарту HDTV, битрейт 0,5–2 Мбит/с условно считается малым, 2–6 Мбит/с – средним, а 6 Мбит/с и более на данный момент принято считать большим [10]. Чем больше битрейт закодированной видеопоследовательности, тем меньше ее искажение относительно исходной, но тем больший объем памяти она занимает.

В соответствии с операцией ОД блоки кадры разбиваются на блоки, далее в рассматриваемом кадре для каждого блока, обозначаемого C_n , выполняется поиск в области Ω восстановленного кадра наиболее «похожего» блока, обозначаемого C_k , смещенного на вектор движения [11]. Ширина области поиска влияет как на качество результата, так и на время сжатия видеопоследовательности. Типичными значениями ширины области поиска ОД блоков кадра полным перебором являются 16, 32, 64 и 128 пикселей [10].

Для более точного поиска векторов движения в современных стандартах сжатия видеоинформации применяют уточнение векторов движения до половины или даже до четверти пиксела. Для этого путем интерполяции формируют три дополнительных базовых блока, смещенных относительно найденного блока C_k на половину пиксела вправо, на половину пиксела вниз и на половину пиксела вправо-вниз. Затем из полученных четырех векторов-кандидатов выбирается лучший по какому-либо критерию вектор. Аналогично выполняется уточнение до четверти пиксела. Чем точнее поиск в операции ОД, тем ниже уровень искажения результата, но тем больше времени потребуется кодеру для сжатия видеоинформации.

Для оценки движения блоков текущего кадра могут быть использованы не один, а несколько предыдущих или последующих в порядке воспроизведения восстановленных кадров. Тогда для совместного поиска векторов движения используются эти восстановленные кадры, что позволяет улучшить качество результата сжатия, но при этом увеличивается объем вычислений в алгоритме сжатия. Типичными значениями количества восстановленных кадров операции ОД являются 1, 3, 5 и 7 кадров [10].

К основным показателям результата сжатия относятся следующие [11]:

- коэффициент сжатия;
- уровень искажения;
- время сжатия видеопоследовательности.

Коэффициент сжатия характеризует эффективность сжатия – отношение размера выходного файла, содержащего сжатую видеопоследовательность, к размеру входного файла, содержащего исходную видеопоследовательность. Чем эффективнее метод сжатия, тем меньше коэффициент сжатия.

Объективной характеристикой качества результата сжатия является уровень искажения восстановленной видеопоследовательности относительно исходной. Наиболее распространенной метрикой уровня искажения является Peak signal-to-noise ratio – PSNR, которая измеряется в дБ и вычисляется по формуле:

$$PSNR(x, y) = 10 \cdot \log_{10} \frac{255^2 \cdot W \cdot H}{\sum_{i=0}^W \sum_{j=0}^H (x_{ij} - y_{ij})^2}, \quad (1)$$

где W и H – ширина и высота кадра видеопоследовательности, соответственно; x – восстановленная видеопоследовательность; y – исходная видеопоследовательность; i, j – индексы пиксела в кадре. Чем больше значение PSNR, тем ниже уровень искажения и качественнее сжатая видеопоследовательность.

Время сжатия видеопоследовательности – это время, затрачиваемое кодером для осуществления алгоритма сжатия видеопоследовательности.

ОРГАНИЗАЦИЯ ВЫЧИСЛИТЕЛЬНЫХ ЭКСПЕРИМЕНТОВ

В распределенную компьютерную систему компрессии видеоинформации на узел-анализатор поступает видеоинформация для сжатия. Анализатор принимает ее и выполняет разделение видеопоследовательностей на части предложенным в [9] методом.

Получаемые в результате разделения части видеопоследовательностей по мере появления становятся в очередь на сжатие вычислительными узлами. Если какой-либо вычислительный узел свободен, то он изымает видеопоследовательность из очереди и сжимает ее с заданными параметрами при помощи программы, реализующей алгоритмы стандарта MPEG-4/H264 [12]. В качестве такой программы использован видеокomppressor FFmpeg [13]. Сжатая часть видеоинформации поступает в хранилище. Если все ресурсы заняты, то видеопоследовательность ожидает в очереди пока не освободится хотя бы один процессор. Такая организация вычислительного процесса позволяет избежать накладных расходов на синхронизацию процессоров, а так же сжимать части видеопоследовательностей по мере их поступления от узла-анализатора.

Для проведения экспериментов использован предоставленный для исследования методов сжатия видеоинформации кластер Института проблем моделирования в энергетике (ИПМЭ) им. Г. Е. Пухова г. Киев: 4 узла следующей конфигурации:

- процессор Intel Xeon 5405;
- оперативная память 4×2 Гб DDR-2 на каждый вычислительный узел;

– коммуникационная среда InfiniBand 20Гб/с.

На узлах кластера, задействованных в экспериментах, помимо FFmpeg установлено промежуточное ПО (middleware): пакет ARC [14] с планировщиком Torque [15] на операционной системе CentOS 5.2.

В экспериментах использованы тестовые видеопоследовательности, удовлетворяющие стандарту HDTV, подробнее описанные в работе [9] и приведенные в таблицах данной статьи. Они были получены конкатенацией файлов, содержащих общепринятые для тестирования методов сжатия видеопоследовательности [16, 17] с разрешением кадра 1920×1080 пикселей. Конкатенация обеспечивала дополнительные смены сцен. После разделения тестовых видеопоследовательностей на части анализатором при помощи упомянутого выше метода, получаемые части видеопоследовательностей ограничены одной сценой, а их смежные кадры характеризуются значительной корреляционной зависимостью.

Во всех экспериментах применено дискретное косинусное преобразование (DCT) для устранения интракадровой избыточности и контекстно-адаптивное двоичное арифметическое кодирование (CABAC) для статистического кодирования без потерь [11, 18]. Од блоков выполнялась полным перебором векторов движения с четвертьпиксельной точностью. Параметры DCT и CABAC являются общепринятыми в сжатии видеоинформации, поэтому в данной работе подробно на них останавливаться не будем.

Путем варьирования параметров кодера, осуществляющего компрессию видеоинформации в распределенной системе, экспериментально было исследовано влияние каждого из проанализированных параметров компрессии на показатели результата сжатия.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Далее приводятся усредненные значения результатов многократных экспериментов на четырех узлах кластера при сжатии описанных выше тестовых видеопоследовательностей. Такие видеопоследовательности однотипны, т. е. ограничены одной сценой, их смежные кадры характеризуются значительной корреляционной зависимостью, следовательно, результаты экспериментов для данной выборки видеопоследовательностей позволяют сформировать выводы в целом относительно влияния параметров сжатия видеоинформации, разделяемой на части при помощи метода [9] и сжимаемой на узлах распределенной системы, на показатели результата компрессии. В следующих таблицах величина «среднее значение PSNR» вычислена по формуле (1) и усреднена по всем кадрам видеопоследовательности.

Коэффициенты сжатия, средние значения PSNR и время сжатия, полученные в результате экспериментов при битрейте 0,5 Мбит/с и пяти восстановленных кадрах ($ReFrames=5$) для ОД блоков, приведены в табл. 1–3.

Таким образом, при битрейте 0,5 Мбит/с расширение области поиска векторов движения с 16 до 128 пикселей ведет к незначительному улучшению качества

результата (до 1,36 %) и увеличению коэффициента сжатия (до 7,14 %). Время сжатия существенно возрастает (в 9,2 раза) за счет увеличения количества сравниваемых векторов движения при расширении области поиска. При этом для описания векторов движения требуется больший объем памяти, поэтому коэффициент сжатия возрастает.

Результаты аналогичных экспериментов, но при битрейте 4 Мбит/с и пяти восстановленных кадрах для ОД блоков приведены в табл. 4–6.

Таблица 1. Коэффициенты сжатия при битрейте 0,5 Мбит/с и $ReFrames=5$

Видеопоследовательность	Ω , пикселей			
	16	32	64	128
Blue sky Pedestrian area part1	$8,6 \cdot 10^{-4}$		$8,7 \cdot 10^{-4}$	
Blue sky Pedestrian area part2	10^{-3}			
Riverbed Station part1	$2,4 \cdot 10^{-3}$		$2,5 \cdot 10^{-3}$	
Riverbed Station part2	$7,4 \cdot 10^{-4}$		$7,5 \cdot 10^{-4}$	
Rush hour Tractor part1	$8,1 \cdot 10^{-4}$		$8,2 \cdot 10^{-4}$	
Rush hour Tractor part2	$1,4 \cdot 10^{-3}$		$1,5 \cdot 10^{-3}$	

Таблица 2. Средние значения PSNR (дБ) при битрейте 0,5 Мбит/с и $ReFrames=5$

Видеопоследовательность	Ω , пикселей			
	16	32	64	128
Blue sky Pedestrian area part1	32,45	32,47	32,55	32,56
Blue sky Pedestrian area part2	32,58	32,65	32,7	32,76
Riverbed Station part1	27,98	28,16	28,28	28,36
Riverbed Station part2	34,9	34,96		34,97
Rush hour Tractor part1	35,02	35,08	35,14	
Rush hour Tractor part2	29,54	29,61	29,68	29,72

Таблица 3. Время сжатия (с) при битрейте 0,5 Мбит/с и $ReFrames=5$

Видеопоследовательность	Ω , пикселей			
	16	32	64	128
Blue sky Pedestrian area part1	18,07	28,19	57,74	147,82
Blue sky Pedestrian area part2	27,09	42,43	87,38	222,43
Riverbed Station part1	33,71	55,51	115,54	289,67
Riverbed Station part2	27,75	43,85	89,82	236
Rush hour Tractor part1	17,72	26,54	52,58	136,77
Rush hour Tractor part2	34,37	56,66	121	317,02

Таблица 4. Коэффициенты сжатия при битрейте 4 Мбит/с и $ReFrames=5$

Видеопоследовательность	Ω , пикселей
	16, 32, 64, 128
Blue sky Pedestrian area part1	$6 \cdot 10^{-3}$
Blue sky Pedestrian area part2	
Riverbed Station part1	$8,2 \cdot 10^{-3}$
Riverbed Station part2	$5,9 \cdot 10^{-3}$
Rush hour Tractor part1	$6,3 \cdot 10^{-3}$
Rush hour Tractor part2	$6,5 \cdot 10^{-3}$

Таким образом, при битрейте 4 Мбит/с расширение области поиска векторов движения ведет к незначительному улучшению качества результата (до 0,44 %). Качество сжатия улучшается менее заметно, чем при битрейте 0,5 Мбит/с, поскольку кодер может формировать больше бит за секунду для описания векторов движения, следовательно, расширять область поиска не обязательно. При этом коэффициент сжатия не изменяется, а время сжатия существенно увеличивается (в 11,3 раз).

Подобные эксперименты были проведены так же при битрейте 8 Мбит/с и пяти восстановленных кадрах для ОД блоков. Результаты приведены в табл. 7–9.

Таблица 5. Средние значения PSNR (дБ) при битрейте 4 Мбит/с и $ReFrames=5$

Видеопоследовательность	Ω , пикселей			
	16	32	64	128
Blue sky Pedestrian area part1	41,08	41,1	41,1	41,2
Blue sky Pedestrian area part2	40,9			40,91
Riverbed Station part1	31,55	31,6	31,65	31,69
Riverbed Station part2	41,88		41,89	
Rush hour Tractor part1	41,92	41,93		
Rush hour Tractor part2	36,09	36,11	36,14	36,16

Таблица 6. Время сжатия (с) при битрейте 4 Мбит/с и $ReFrames=5$

Видеопоследовательность	Ω , пикселей			
	16	32	64	128
Blue sky Pedestrian area part1	51,24	95,3	204,82	504
Blue sky Pedestrian area part2	70,54	129,5	283,4	728
Riverbed Station part1	70,1	129,11	287,27	746
Riverbed Station part2	61,79	112,07	247,7	698,16
Rush hour Tractor part1	57,37	104,06	230,58	588
Rush hour Tractor part2	63,53	104,51	264,1	727,1

Таблица 7. Коэффициенты сжатия при битрейте 8 Мбит/с и $ReFrames=5$

Видеопоследовательность	Ω , пикселей
	16, 32, 64, 128
Blue sky Pedestrian area part1	$12,2 \cdot 10^{-3}$
Blue sky Pedestrian area part2	$12,1 \cdot 10^{-3}$
Riverbed Station part1	$15,4 \cdot 10^{-3}$
Riverbed Station part2	$12,4 \cdot 10^{-3}$
Rush hour Tractor part1	$12,5 \cdot 10^{-3}$
Rush hour Tractor part2	$12,7 \cdot 10^{-3}$

Таблица 8. Средние значения PSNR (дБ) при битрейте 8 Мбит/с и $ReFrames=5$

Видеопоследовательность	Ω , пикселей			
	16	32	64	128
Blue sky Pedestrian area part1	42,7	42,72	42,73	42,74
Blue sky Pedestrian area part2	42,46			
Riverbed Station part1	33,97		34	
Riverbed Station part2	42,48			
Rush hour Tractor part1	42,81			
Rush hour Tractor part2	38,35	38,36	38,37	38,38

Таким образом, при битрейте 8 Мбит/с расширение области поиска векторов движения ведет к еще менее существенному улучшению качества результата (до 0,08 %). При этом коэффициент сжатия не изменяется, а время сжатия существенно увеличивается (примерно в 12,5 раз). Следовательно, расширять область поиска векторов движения при битрейте 8 Мбит/с и более не целесообразно.

Для исследования влияния количества восстановленных кадров операции ОД блоков на показатели результата сжатия выполнены аналогичные эксперименты, в которых при фиксированном битрейте 4 Мбит/с изменялось количество восстановленных кадров ОД. Результаты экспериментов приведены в табл. 10–15.

Таким образом, из таблиц 4–6 и 10–15 видно, что увеличение количества восстановленных кадров с одного до семи для ОД блоков при фиксированном битрейте 4 Мбит/с незначительно повышает качество сжатия (до 0,7 %) и не влияет на коэффициент сжатия, однако время сжатия увеличивается существенно (в 3 раза).

Таблица 9. Время сжатия (с) при битрейте 8 Мбит/с и $ReFrames=5$

Видеопоследовательность	Ω , пикселей			
	16	32	64	128
Blue sky Pedestrian area part1	81,2	155,44	324,5	778
Blue sky Pedestrian area part2	94,06	177,11	393,18	1005
Riverbed Station part1	98,24	188,12	435	1155
Riverbed Station part2	85,52	160,73	358,8	929
Rush hour Tractor part1	80,22	155,1	347	870
Rush hour Tractor part2	81,08	151,38	362,3	1017

Таблица 10. Коэффициенты сжатия при битрейте 4 Мбит/с и $ReFrames=1$

Видеопоследовательность	Ω , пикселей			
	16	32	64	128
Blue sky Pedestrian area part1	$6 \cdot 10^{-3}$			
Blue sky Pedestrian area part2				
Riverbed Station part1	$8,2 \cdot 10^{-3}$	$8,3 \cdot 10^{-3}$		
Riverbed Station part2	$5,9 \cdot 10^{-3}$			
Rush hour Tractor part1	$6,3 \cdot 10^{-3}$			
Rush hour Tractor part2	$6,5 \cdot 10^{-3}$			

Таблица 11. Средние значения PSNR (дБ) при битрейте 4 Мбит/с и $ReFrames=1$

Видеопоследовательность	Ω , пикселей			
	16	32	64	128
Blue sky Pedestrian area part1	41,07			
Blue sky Pedestrian area part2	40,81			
Riverbed Station part1	31,41	31,42	31,47	31,51
Riverbed Station part2	41,86	41,86	41,86	41,86
Rush hour Tractor part1	41,85			
Rush hour Tractor part2	36	36,02	36,04	36,06

Таблица 12. Время сжатия (с) при битрейте 4 Мбит/с и $ReFrames=1$

Видеопоследовательность	Ω , пикселей			
	16	32	64	128
Blue sky Pedestrian area part1	33,21	57,21	116,17	279,6
Blue sky Pedestrian area part2	43,12	73,35	152,43	381,6
Riverbed Station part1	36,32	60,7	129,45	328,53
Riverbed Station part2	43,18	73,35	155,48	393,75
Rush hour Tractor part1	35,66	59,28	126	315
Rush hour Tractor part2	40,57	67,91	147,42	389

Таблица 13. Коэффициенты сжатия при битрейте 4 Мбит/с и $ReFrames=7$

Видеопоследовательность	Ω , пикселей
	16, 32, 64, 128
Blue sky Pedestrian area part1	$6 \cdot 10^{-3}$
Blue sky Pedestrian area part2	$6 \cdot 10^{-3}$
Riverbed Station part1	$8,2 \cdot 10^{-3}$
Riverbed Station part2	$5,9 \cdot 10^{-3}$
Rush hour Tractor part1	$6,3 \cdot 10^{-3}$
Rush hour Tractor part2	$6,5 \cdot 10^{-3}$

Таблица 14. Средние значения PSNR (дБ) при битрейте 4 Мбит/с и $ReFrames=7$

Видеопоследовательность	Ω, пикселей			
	16	32	64	128
Blue sky Pedestrian area part1	41,09		41,1	
Blue sky Pedestrian area part2	40,92			
Riverbed Station part1	31,59	31,64	31,69	31,73
Riverbed Station part2	41,88			
Rush hour Tractor part1	41,93	41,94		
Rush hour Tractor part2	36,1	36,13	36,15	36,16

Таблица 15. Время сжатия (с) при битрейте 4 Мбит/с и $ReFrames=7$

Видеопоследовательность	Ω , пикселей			
	16	32	64	128
Blue sky Pedestrian area part1	69,22	125,44	273,33	685
Blue sky Pedestrian area part2	92,81	175	388,6	1007
Riverbed Station part1	89,98	167,74	374,25	982
Riverbed Station part2	76,26	142,82	322,05	850
Rush hour Tractor part1	71,59	135,36	305,58	788
Rush hour Tractor part2	80,84	150,89	356,38	993,2

ВЫВОДЫ

Результаты экспериментов позволяют сделать выводы о том, что на видеопоследовательности, получаемые после разделения видеоинформации на части в распределенной системе предложенным в [9] методом, параметры компрессии оказывают следующее влияние:

– повышение битрейта с 0,5 Мбит/с до 8 Мбит/с ведет к существенному улучшению качества (до 31,59 %), но, в то же время, к значительному ухудшению коэффициента сжатия (в 16,75 раз) и увеличению времени сжатия (в 6,6 раз);

– увеличение количества восстановленных кадров с одного до семи для ОД блоков при фиксированном битрейте 4 Мбит/с незначительно повышает качество сжатия (до 0,7 %) и не влияет на коэффициент сжатия, при этом время сжатия увеличивается существенно (в 3 раза);

– при фиксированном битрейте расширение области поиска векторов движения для ОД блоков существенно увеличивает время сжатия (в 9,2 раз при битрейте 0,5 Мбит/с, в 11,3 раз при битрейте 4 Мбит/с и в 12,5 раз при битрейте 8 Мбит/с). При этом, чем больше битрейт, тем меньше выигрыш в качестве от расширения области поиска ОД блоков. Коэффициент сжатия при битрейте 0,5 Мбит/с увеличивается от расширения области поиска ОД блоков, при битрейте 4 Мбит/с и более – не изменяется.

В дальнейшем представляется перспективным исследовать влияние динамического изменения ширины области поиска векторов движения для каждого блока кадра видеопоследовательности на производительность компрессии видеоинформации в распределенных системах.

СПИСОК ЛІТЕРАТУРИ

1. Lossless Video Compression for Archives: Motion JPEG2k and Other Options / Media Matters Whitepapers. – Режим доступа: \www/ URL: <http://www.media-matters.net/docs/WhitePapers/WPMJ2k.pdf> – 1.02.2006 г. – Загл. с экрана.
2. Parameter values for the HDTV standards for production and international programme exchange : ITU-R Recommendation BT.709, 2008. – 32 p.
3. Understanding and Using High-Definition Video / Adobe Systems. – Access mode: \www/ URL: <http://www.adobe.com/products/premiere/pdfs/hdprimer.pdf> – 1.11.2010 г. – Title from screen.
4. H.261: Video codec for audiovisual services : ITU-T Recommendation H.261, 1993. – 29 p.
5. Performance analysis and architecture evaluation of MPEG-4 video codec system : IEEE International Symposium on Circuits and Systems / H.-C. Chang, L.-G. Chen, M.-Y. Hsu, Y.-C. Chang. – Geneva, 2000. – P. 449–452.
6. Performance Evaluation of Parallel MPEG-4 Video Coding Algorithms on Clusters of Workstations : PARELEC '04 Proceedings of the international conference on Parallel Computing in Electrical Engineering / A. Rodriguez, A. Gonzalez, M. P. Malumbres. – Washington DC, 2004. – P. 354–357.
7. Parallel Scalability of H.264 : Proceedings of the first Workshop on Programmability Issues for Multi-Core Computers / C. H. Meenderinck, A. Azevedo, M. Alvarez, B. H. Juurlink, A. Ramirez. – Goteborg, Sweden, 2008.
8. ON-DEMAND HD VIDEO USING JINI BASED GRID : ICME 2008 / S. Kent, P. Broadbent, N. Warren, S. Gulliver, 2008. – P. 1045–1048.
9. Скрупский, С. Ю. Повышение эффективности сжатия видеоинформации в распределенных компьютерных системах / С. Ю. Скрупский // Электронное моделирование. – 2011. – №6 (33). – С. 57–72.
10. H.264/MPEG-4 AVC – Access mode: \www/ URL: http://en.wikipedia.org/wiki/H.264/MPEG-4_AVC – 28.09.2010 г. – Title from screen.
11. Скрупский, С. Ю. Методы компрессии видеоинформации / С. Ю. Скрупский // Наукові праці Донецького національного технічного університету. Серія: «Обчислювальна техніка та автоматизація». – 2011, №21 (183). – С. 122–130.
12. Advanced video coding for generic audiovisual services : ITU-T Recommendation H.264 and ISO/IEC 14496-10 (AVC), 2003. – 90 p.
13. FFmpeg. – Режим доступа: \www/ URL: <http://www.ffmpeg.org/> – 1.10.2010 г. – Загл. с экрана.
14. Ellert, M. Advanced Resource Connector middleware for lightweight computational Grids / M. Ellert, M. Grønager, A. Konstantinov, B. Köny, J. Lindemann, I. Livenson, J. L. Nielsen, M. Niinimäki, O. Smirnova, A. Wäänänen // Future Generation Computer Systems. – 2007, Vol. 23, Issue 2 – P. 219–240.
15. Система пакетной обработки заданий torque. Руководство пользователя. – М. : Т-Платформы, 2008. – 31 с.
16. YUV Video Sequences / Arizona State University. – Режим доступа: \www/ URL: <http://trace.eas.asu.edu/yuv/> – 1.11.2010 г. – Загл. с экрана.
17. Xiph.org Test Media. – Режим доступа: \www/ URL: <http://media.xiph.org/video/derf/> – 1.11.2010 г. – Загл. с экрана.
18. Парфенова, А. О. Порівняльний аналіз основних стандартів відео для передачі по 4-G мережам / А. О. Парфенова, А. Ю. Макаренко, С. Б. Могильний // Вісник Національного технічного університету України «КПІ». – 2010, №40. – С. 171–176.

Стаття надійшла до редакції 10.05.2011.

Скрупський С. Ю., Луценко Н. В., Скрупська Л. С.

ПАРАМЕТРИ КОМПРЕСІЇ ВІДЕОІНФОРМАЦІЇ В РОЗПОДІЛЕНИХ СИСТЕМАХ

Проаналізовані основні параметри компресії відеоінформації в розподілених системах. Експериментально досліджено вплив цих параметрів на коефіцієнт стискування, рівень викривлення та час стискування відеоінформації в розподіленій системі.

Ключові слова: відеоінформація, розподілена система, компресія, бітрейт, рівень викривлення.

Skrupsky S. Y., Lucenko N. V., Skrupskaya L. S.

THE PARAMETERS OF VIDEO INFORMATION COMPRESSION PROCESS IN DISTRIBUTED SYSTEMS

The paper deals with the analysis of the main parameters of video information compression process in distributed systems. The influence of these parameters on compression coefficient, distortion level and compression time of video information in distributed system is experimentally investigated.

Key words: video information, distributed system, compression, bit rate, distortion level.

ПОДХОД К РАСПРЕДЕЛЕНИЮ ПАРАЛЛЕЛЬНЫХ ВЕТВЕЙ ЗАДАНИЯ В ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЕ

Рассмотрена проблема распределения заданий в вычислительной системе для случая, когда длительность выполнения параллельных ветвей задания, которые необходимо распределить, различна. Для повышения эффективности распределения разработан метод. Показано, что предложенный метод является эффективным, поскольку, используя его, время выполнения заданий заметно сокращается.

Ключевые слова: вычислительная система, grid, параллельные вычисления, ветви задания, распределение заданий, время выполнения задания, время работы системы, время ожидания.

ВВЕДЕНИЕ

Сегодня направление параллельных вычислений в GRID имеет большой интерес со стороны научных исследований, поскольку увеличение скорости вычислений является весьма актуальным для всех, чья деятельность связана с большим объемом вычислительных работ [1]. Использование параллельных вычислений позволяет решать задачи, требующие больших временных затрат, за меньшее время [2, 3]. Однако при параллельном вычислении бывают случаи, когда некоторые ветви задания вычисляются раньше и ожидают окончания работы остальных. При этом вычислительные узлы простаивают, что сказывается на общем времени выполнения заданий, а, следовательно, и на загрузке всей системы. Таким образом, сокращение времени простоя позволит увеличить скорость вычислений.

В системах, в которых приоритет заданий играет важную роль, диспетчеризация нагрузки происходит поочередно для каждого задания и большее внимание уделяется распределению ветвей параллельного задания между компьютерами. В таком случае зачастую количество ветвей любого из заданий меньше количества компьютеров.

ПОСТАНОВКА ЗАДАЧИ

Рассмотрим вычислительную систему, в которую в определенный момент времени поступают параллельные задания, т. е. состоящие из параллельно выполняемых ветвей. Для сокращения общего времени выполнения всех поступающих для решения заданий и при этом минимизации нагрузки на вычислительные ресурсы следует определить, какая ветвь выполняемого параллельного задания будет выполняться на каком компьютере. Таким образом, необходимо решить проблему эффективного распределения параллельных ветвей задания между вычислительными узлами системы.

Пусть каждый из компьютеров имеет различную производительность (у некоторых компьютеров производительность может совпадать, но не у всех), длительность

выполнения каждой из ветвей одного задания также различна, в результате чего время выполнения задания в целом определяется наибольшей длительностью выполнения одной из ветвей этого задания. Также следует отметить, что количество ветвей не превышает количества компьютеров.

Пусть данная вычислительная система состоит из n компьютеров, а очередь заданий состоит из q заданий (каждое задание содержит l_i ветвей, где $i = \overline{1, q}$). Тогда математическая модель этой задачи может быть записана следующим образом.

Мощность эталонного компьютера – P^e .

Мощность k -го компьютера – P_k , где $k = \overline{1, n}$.

Тогда коэффициент мощности k -го компьютера –

$$c_k = \frac{P^e}{P_k}, \quad (1)$$

т. е. чем больше мощность k -го компьютера, тем меньше его коэффициент мощности.

Во время распределения i -го задания свободно только a_i компьютеров. Учитывая (1), соответственно, коэффициент мощности m -го компьютера –

$$c_m = \frac{P^e}{P_m}, \quad (2)$$

где $m = \overline{1, a_i}$.

Тогда время выполнения j -ой ветви i -го задания на m -м компьютере t_{ijm} :

$$t_{ijm} = t_{ij}^e \times c_m, \quad (3)$$

где $j = \overline{1, l_i}$, t_{ij}^e – время выполнения j -ой ветви i -го задания на эталонном компьютере.

Тогда время выполнения i -го задания в данной системе T_i :

$$T_i = \max\{t_{ijm}\}, \quad (4)$$

Время ожидания окончания выполнения всего задания m -м компьютером, закончившим выполнение j -ой ветви i -го задания, обозначим как Δt_{ijm} :

$$\Delta t_{ijm} = T_i - t_{ijm}. \quad (5)$$

Суммарное время ожидания окончания всех ветвей i -го задания Δt_i :

$$\Delta t_i = \sum_{j=1}^{l_i} \Delta t_{ijm}. \quad (6)$$

Таким образом, для сокращения общего времени выполнения всех поступающих для решения заданий необходимо, чтобы

$$\begin{cases} T_i \rightarrow \min \\ \Delta t_i \rightarrow \min \end{cases} \quad (7)$$

Введем понятие времени ожидания k -м компьютером начала выполнения $(i+r)$ -го задания после окончания выполнения i -го задания, где $r = \overline{1, (q-1)}$. Обозначим это время как Δt_{ki} .

Тогда время работы системы, потраченное на выполнение всех заданий, рассчитывается как максимальное время работы одного из n компьютеров этой системы и равняется сумме длительностей ветвей заданий, выполненных на конкретном k -м компьютере плюс сумма всех Δt_{ki} для этого же компьютера.

А суммарное время ожидания всей системы Δt :

$$\Delta t = \sum_{k=1}^n \sum_{i=1}^q \Delta t_{ki} + \sum_{i=1}^q \Delta t_i.$$

Таким образом, необходимо, чтобы, учитывая (7), время работы системы стремилось к минимуму и $\Delta t \rightarrow \min$.

СУЩЕСТВУЮЩИЕ РЕШЕНИЯ

Для сравнения рассмотрим простейший случай, когда в каждый момент времени распределяется одно задание и самая длительная ветвь задания распределяется на наиболее мощный компьютер, а далее распределение происходит в порядке возрастания длительностей ветвей заданий и убывания мощностей компьютеров.

Согласно такому подходу, порядковые номера компьютеров и ветвей совпадают, т.е. $m = j$.

Такой подход обеспечивает выполнение первого из условий (7), но не в полной мере обеспечивает второе.

Для выполнения обоих условий в данной работе предлагается метод, в котором распределяются ветви заданий таким образом, чтобы были заняты также и менее мощные компьютеры. При этом вновь рассчитанное время выполнения j -й ветви i -го задания, t_{ijm} , не превышало уже рассчитанное время выполнения i -го задания, T_i . Таким образом, получаем, что выполняются условия (5) и для распределения $(i+1)$ -го задания, выполняемого одновременно с i -м на других компьютерах, остаются вычислительные узлы с большей средней мощностью. За счет этого общее время выполнения Q заданий уменьшится.

РАЗРАБОТАННЫЙ МЕТОД

Рассмотрим работу разработанного метода на примере. На рис. 1 изображено задание, содержащее 6 ветвей вычислений, которое выполняется в системе, состоящей из 6 компьютеров с одинаковой мощностью, — здесь все компьютеры можно считать эталонными. Длительности каждой ветви задания соответственно равны: 100, 80, 70, 60, 30, 20 условных единиц времени.

На рис. 1 система выглядит в виде 6 строк, каждая из которых представляет собой компьютер. Первая ячейка указывает номера компьютера; вторая — коэффициент мощности компьютера (здесь он равен 1.00, т. к. любой из компьютеров можно считать эталонным); третья — непосредственно графически представленную распределенную ветвь задания. Распределенная ветвь задания представлена в виде двух частей (или одной части для самой длительной ветви): собственно время выполнения с номером ветви задания и время ожидания окончания выполнения всех ветвей задания (заштрихованная область). При этом первое число номера ветви задания, обозначает номер задания, второе — номер ветви. К примеру, номер 1.2 указывает, что это распределена вторая ветвь первого задания.

Под изображением системы на рис. 1 представлены краткие статистические данные о работе системы с распределенными заданиями: время работы T и суммарное время ожидания системы Δt .

1	1.00	1.1	
2	1.00	1.2	
3	1.00	1.3	
4	1.00	1.4	
5	1.00	1.5	
6	1.00	1.6	
Статистика:			
Время работы:			100
Суммарное время ожидания:			240

Рис. 1. Задание, выполняемое в системе, состоящей из компьютеров одинаковой мощности

Время выполнения такого задания в данной системе:

$T = \max\{100, 80, 70, 60, 30, 20\} = 100$, а суммарное время ожидания окончания всех ветвей задания:

$$\Delta t = (100 - 100) + (100 - 80) + (100 - 70) + (100 - 60) + (100 - 30) + (100 - 20) = 240.$$

Рассмотрим более сложный случай. Используем вычислительную систему из 12 компьютеров с коэффициентами мощности соответственно: 0,5; 0,75; 0,8; 1; 1,2; 1,25; 1,4; 1,5; 2; 2,1; 2,1; 2,5. Запустим для выполнения в ней два задания, таких же, как и рассмотренное выше, и проведем расчеты времени работы и суммарного времени ожидания.

Согласно рассмотренному простейшему подходу, время выполнения каждой ветви первого задания рассчитывается следующим образом:

$$t_{111} = t_{11}^e \times c_1 = 100 \times 0,5 = 50,$$

$$t_{122} = t_{12}^e \times c_2 = 80 \times 0,75 = 60,$$

$$t_{133} = t_{13}^e \times c_3 = 70 \times 0,8 = 56,$$

$$t_{144} = t_{14}^e \times c_4 = 60 \times 1 = 60,$$

$$t_{155} = t_{15}^e \times c_5 = 30 \times 1,2 = 36,$$

$$t_{166} = t_{16}^e \times c_6 = 20 \times 1,25 = 25.$$

Тогда время выполнения первого задания в данной системе:

$T_1 = \max\{50, 60, 56, 60, 36, 25\} = 60$, а суммарное время ожидания окончания всех ветвей первого задания:

$$\Delta t_1 = (60 - 50) + (60 - 60) + (60 - 56) + (60 - 60) + (60 - 36) + (60 - 25) = 73.$$

Для второго задания время выполнения каждой ветви имеем:

$$t_{211} = t_{21}^e \times c_7 = 100 \times 1,4 = 140,$$

$$t_{222} = t_{22}^e \times c_8 = 80 \times 1,5 = 120,$$

$$t_{233} = t_{23}^e \times c_9 = 70 \times 2 = 140,$$

$$t_{244} = t_{24}^e \times c_{10} = 60 \times 2,1 = 126,$$

$$t_{255} = t_{25}^e \times c_{11} = 30 \times 2,1 = 63,$$

$$t_{266} = t_{26}^e \times c_{12} = 20 \times 2,5 = 50.$$

А время выполнения второго задания в данной системе и суммарное время ожидания окончания всех ветвей второго задания соответственно:

$$T_2 = \max\{140, 120, 140, 126, 63, 50\} = 140,$$

$$\Delta t_1 = (140 - 140) + (140 - 120) + (140 - 140) + (140 - 126) + (140 - 63) + (140 - 50) = 201.$$

В таком случае время работы системы равно 140 условных единиц времени, а суммарное время ожидания системы – 274.

Согласно разработанному методу, время выполнения первого задания в данной системе рассчитывается аналогично и так же равняется $T_1 = 60$. Время выполнения каждой j -ой ветви первого задания на каждом m -м компьютере t_{1jm} представлено в табл. 1.

Из этой таблицы выбираются значения, которые максимально приближены к T_1 , но не более его. Эти значения в таблице выделены серым цветом. Выбор производится по очереди для каждого столбца, начиная с первого. При этом никакие два значения не должны принадлежать одному и тому же столбцу или строке.

Тогда суммарное время ожидания окончания всех ветвей первого задания:

$$\Delta t_1 = (60 - 50) + (60 - 60) + (60 - 56) + (60 - 60) + (60 - 60) + (60 - 50) = 24.$$

Время выполнения каждой j -й ветви второго задания на каждом m -м компьютере t_{2jm} и выбранные значения представлены в табл. 2.

Таблица 1. Время выполнения ветвей первого задания

Номер ветви Номер компьютера	1	2	3	4	5	6
1	50	40	35	30	15	10
2	75	60	52,5	45	22,5	15
3	80	64	56	48	24	16
4	100	80	70	60	30	20
5	120	96	84	72	36	24
6	125	100	87,5	75	37,5	25
7	140	112	98	84	42	28
8	150	120	105	90	45	30
9	200	160	140	120	60	40
10	210	168	147	126	63	42
11	210	168	147	126	63	42
12	250	200	175	150	75	50

Таблица 2. Время выполнения ветвей второго задания

Номер ветви Номер компьютера	1	2	3	4	5	6
5	120	96	84	72	36	24
6	125	100	87,5	75	37,5	25
7	140	112	98	84	42	28
8	150	120	105	90	45	30
10	210	168	147	126	63	42
11	210	168	147	126	63	42

Тогда время выполнения второго задания в данной системе:

$T_2 = \max\{120, 100, 98, 90, 63, 42\} = 120$, а суммарное время ожидания окончания всех ветвей второго задания:

$$\Delta t_2 = (120 - 120) + (120 - 120) + (120 - 98) + (120 - 75) + (120 - 63) + (120 - 42) = 202.$$

РЕЗУЛЬТАТЫ РАБОТЫ

В результате распределения заданий в данной системе согласно разработанному методу, время работы сис-

темы равно 120 условных единиц времени, а суммарное время ожидания системы – 226.

На рис. 2, 3 очевидны преимущества предлагаемого метода:

– распределив первое задание по компьютерам системы, не смотря на то, что время выполнения одинаково, время ожидания окончания первого задания гораздо меньше, а средняя мощность компьютеров, оставшихся в распоряжении второго задания, выше;

– распределив второе задание, имеем меньшее общее время выполнения и ожидания, что и требовалось получить.

1	0.50	1.1	
2	0.75	1.2	
3	0.80	1.3	
4	1.00	1.4	
5	1.20	1.5	
6	1.25	1.6	
7	1.40		
8	1.50		
9	2.00		
10	2.10		
11	2.10		
12	2.50		
Статистика:			
Время работы:		60	
Суммарное время ожидания:		73	

1	0.50	1.1	
2	0.75	1.2	
3	0.80	1.3	
4	1.00	1.4	
5	1.20	1.5	
6	1.25	1.6	
7	1.40	2.1	
8	1.50	2.2	
9	2.00	2.3	
10	2.10	2.4	
11	2.10	2.5	
12	2.50	2.6	
Статистика:			
Время работы:		140	
Суммарное время ожидания:		274	

Рис. 2. Пример распределения заданий с использованием существующего метода

1	0.50	1.1	
2	0.75	1.2	
3	0.80	1.3	
4	1.00	1.4	
5	1.20		
6	1.25		
7	1.40		
8	1.50		
9	2.00	1.5	
10	2.10		
11	2.10		
12	2.50	1.6	
Статистика:			
Время работы:		60	
Суммарное время ожидания:		24	

1	0.50	1.1	
2	0.75	1.2	
3	0.80	1.3	
4	1.00	1.4	
5	1.20	2.1	
6	1.25	2.4	
7	1.40	2.3	
8	1.50	2.2	
9	2.00	1.5	
10	2.10	2.6	
11	2.10	2.5	
12	2.50	1.6	
Статистика:			
Время работы:		120	
Суммарное время ожидания:		226	

Рис. 3. Пример распределения заданий с использованием предлагаемого метода

ВЫВОДЫ

Таким образом, в данной статье был рассмотрен подход к распределению ветвей параллельного задания между компьютерами гетерогенной вычислительной среды. Был продемонстрирован пример оптимизации распределения, используя рассмотренный метод, эффективно решающий поставленную задачу.

СПИСОК ЛИТЕРАТУРЫ

1. Linux. Кластер. Практическое руководство по параллельным вычислениям [Электронный ресурс] / авт. курса Ю. Сбитнев – Электрон. текстовые дан. – [Волгоград?], [199–?] – Режим доступа: <http://cluster.linux-ekb.info>, свободный. – Загл. с экрана.
2. Шпаковский Г. И. Реализация параллельных вычислений: кластеры, многоядерные процессоры, грид, квантовые компьютеры / Г. И. Шпаковский. – Минск : БГУ.– 2010. – 155 с.
3. Технология запуска параллельных задач в различных распределенных средах / [В. М. Волохов, Д. А. Варламов, Н. Ф. Сурков, А. В. Пивушков, А. В. Волохов] // Распределенные вычисления и Грид-технологии в науке и образовании : труды 4-й междунар. конф. (Дубна, 28 июня – 3 июля, 2010 г.). – Дубна : ОИЯИ. – С. 329–334.

Стаття надійшла до редакції 20.05.2011.

Сметанін Р. І., Тягунова М. Ю.

ПІДХІД ДО РОЗПОДІЛУ ПАРАЛЕЛЬНИХ ГІЛОК ЗАВДАННЯ В ОБЧИСЛЮВАЛЬНІЙ СИСТЕМІ

Розглянуто проблему розподілу завдань в обчислювальній системі для випадку, коли тривалість виконання паралельних гілок завдання, які необхідно розподілити, різна. Розроблено метод для підвищення ефективності розподілу. Показано, що запропонований метод є ефективним, оскільки, використовуючи його, час виконання завдань помітно скорочується.

Ключові слова: обчислювальна система, grid, паралельні обчислення, гілки завдання, розподілення завдань, час виконання завдання, час роботи системи, час очікування.

Smetanin R. I., Tyagunova M. Yu.

APPROACH TO THE DISTRIBUTION OF PARALLEL TASK THREADS IN COMPUTING SYSTEMS

The problem of the tasks distribution in the computer system with different execution time of parallel threads, which must be distributed, is considered. Method to improve distribution efficiency is developed. Method is effective, because it much reduces execution time of task.

Keywords: computing system, grid, parallel computations, task threads, tasks distribution, task execution time, system work time, waiting time.

УПРАВЛІННЯ У ТЕХНІЧНИХ СИСТЕМАХ

УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ СИСТЕМАХ

CONTROL IN TECHNICAL SYSTEMS

УДК 621.11-32

Александрова Т. Е.¹, Кононенко В. А.², Лазаренко А. А.³, Зейн Али³¹Канд. техн. наук, доцент Национального технического университета «ХПИ»²Канд. техн. наук, старший научный сотрудник Национального технического университета «ХПИ»³Аспирант Национального технического университета «ХПИ»

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЦИФРОВЫХ ПД-СТАБИЛИЗАТОРОВ ПОДВИЖНЫХ ОБЪЕКТОВ С НИЗКОЧАСТОТНЫМИ ФИЛЬТРАМИ БАТТЕРУОРТА И ЛАНЦОША

Рассматривается проблема построения цифровых ПД-стабилизаторов подвижных объектов с зашумленными высокочастотными помехами входными сигналами. Проведен сравнительный анализ стабилизаторов с фильтрами Баттеруорта и Ланцоша различных порядков.

Ключевые слова: цифровой ПД-стабилизатор, низкочастотный фильтр, дифференцирующий фильтр.

ПОСТАНОВКА ПРОБЛЕМЫ

В качестве чувствительных элементов систем стабилизации подвижных объектов, в частности, объектов военного назначения, обычно используется гироскопические платформы с нелинейной системой коррекции [1, 2]. В работе [3] показано, что оси таких гироплатформ совершают высокочастотные колебания, приводящие к зашумленности измеряемых параметров движения высокочастотными помехами. При разработке цифровых стабилизаторов подвижных объектов широкое распространение получили низкочастотные цифровые фильтры, подавляющие высокочастотные помехи. Среди большого разнообразия цифровых фильтров наиболее распространенными и эффективными являются низкочастотные фильтры Баттеруорта и Ланцоша [4]. В данной статье рассматривается проблема рационального использования цифровых фильтров различного порядка при построении ПД-стабилизаторов подвижных объектов.

АНАЛИЗ ПУБЛИКАЦИЙ ПО ТЕМЕ ИССЛЕДОВАНИЙ

В работах [4, 5] приведены АЧХ и ФЧХ фильтров Баттеруорта и Ланцоша различного порядка. Показано, что фильтры Баттеруорта, имея практически идеальную АЧХ, в смысле подавления высокочастотных помех, вносят значительное фазовое запаздывание при прохождении дискретного сигнала. Фильтры же Ланцоша, наоборот, вносят значительное фазовое опережение, возрастающее с повышением порядка фильтра и с улучшением его фильтрующих свойств. В работах [6, 7] получены аналитические соотношения для АЧХ и ФЧХ фильтров Баттеруорта и Ланцоша различных порядков, позволяющие произвести оценку динамических свойств цифровых стабилизаторов с данными фильтрами.

ЦЕЛЬ РАБОТЫ

Целью настоящей статьи является сравнительный анализ динамических характеристик ПД – стабилизаторов подвижных объектов при использовании цифровых фильтров различного порядка.

ОСНОВНАЯ ЧАСТЬ

Рассмотрим цифровой ПД – стабилизатор, структурная схема которого приведена на рис. 1.

Для рассмотренного стабилизатора можно записать

$$Z\{U[nT]\} = \{k_\phi W_\phi(z) + k_\phi W_\lambda(z)\} Z\{\phi[nT]\}, \quad (1)$$

где $Z\{\cdot\}$ – символ Z-преобразования. Тогда дискретная передаточная функция ПД-стабилизатора, выполненного по параллельной схеме, равна

$$W_{\text{ПД}}(z) = \frac{Z\{U[nT]\}}{Z\{\phi[nT]\}} = k_\phi W_\phi(z) + k_\phi W_\lambda(z). \quad (2)$$

Передаточные функции фильтров Баттеруорта и Ланцоша второго порядка записываются в виде [4]

$$W_\phi(z) = \frac{a_0 + a_1 z^{-1} + a_2 z^{-2}}{b_0 + b_1 z^{-1} + b_2 z^{-2}}; \quad (3)$$

$$W_\lambda(z) = c_0 + c_1 z^{-1} - c_1 z^{-3} - c_0 z^{-4}. \quad (4)$$

Подставляя (3) и (4) в формулу (2), получим передаточную функцию ПД-стабилизатора, представленного на рис. 1.

$$W_{\text{ПД}}(z) = \frac{k_\phi(d_0 + d_1 z^{-1} + d_2 z^{-2}) + k_\phi(f_0 + f_1 z^{-1} + f_2 z^{-2} + f_3 z^{-3}) - \frac{-f_4 z^{-4} - f_5 z^{-5} - f_6 z^{-6}}{1 + l_1 z^{-1} + l_2 z^{-2}}}{1 + l_1 z^{-1} + l_2 z^{-2}}, \quad (5)$$

где соответствующие постоянные связаны с параметрами фильтров (3) и (4) соотношениями

$$d_0 = \frac{a_0}{b_0}; \quad d_1 = \frac{a_1}{b_0}; \quad f_0 = c_0; \quad f_1 = \frac{b_0 c_1 + b_1 c_0}{b_0};$$

$$f_2 = \frac{b_1 c_1 + b_2 c_0}{b_0}; \quad f_3 = \frac{c_1(b_2 + b_0)}{b_0}; \quad f_4 = \frac{b_0 c_0 + b_1 c_1}{b_0};$$

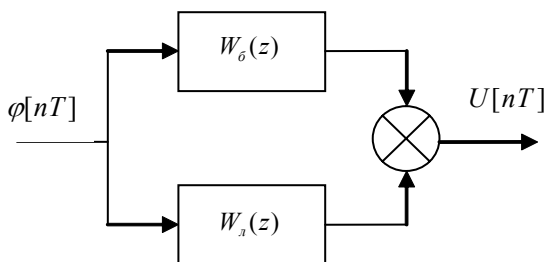


Рис. 1. ПД-стабилизатор, выполненный по параллельной схеме:

$W_\phi(z)$ – фильтр Баттеруорта; $W_\lambda(z)$ – фильтр Ланцоша

$$f_5 = \frac{b_1 c_0 + b_2 c_1}{b_0}; \quad f_6 = \frac{b_2 c_0}{b_0};$$

$$l_1 = \frac{b_1}{b_0}; \quad l_2 = \frac{b_2}{b_0}.$$

К обеим частям соотношения (5) применим операцию обратного Z-преобразования, в результате чего получаем разностное уравнение рассматриваемого ПД-стабилизатора

$$U[nT] = k_\phi \{d_0 \phi[nT] + d_1 \phi[(n-1)T] + d_2 \phi[(n-2)T]\} + k_\phi \{f_0 \phi[nT] + f_1 \phi[(n-1)T] + f_2 \phi[(n-2)T] + f_3 \phi[(n-3)T] - f_4 \phi[(n-4)T] - f_5 \phi[(n-5)T] - f_6 \phi[(n-6)T]\} - l_1 U[(n-1)T] - l_2 U[(n-2)T]. \quad (6)$$

В качестве примера рассмотрим фильтры (3) и (4) с параметрами: $a_0=0,08073$; $a_1=0,16147$; $b_0=1,48256$; $b_1=-1,83854$; $b_2=0,677890$; $c_0=5$; $c_1=2,5$. Используя соотношение (6), построим АЧХ и ФЧХ рассматриваемого ПД-стабилизатора, приведенного на рис. 2, при различных значениях варьируемых параметров k_ϕ , k_ϕ . Для этого на входе линейного цифрового стабилизатора (6) сформируем синусоидальную решетчатую функцию

$$\phi[nT] = A \sin \omega n T. \quad (7)$$

При этом на выходе стабилизатора, в соответствии с алгоритмом (6), имеет место синусоидальная решетчатая функция

$$U[nT] = B(\omega T) \sin[\omega(nT) + \varphi(nT)]. \quad (8)$$

Для каждого текущего значения ωT вычисляется значение АЧХ

$$M(\omega T) = \frac{B(\omega T)}{A},$$

и измеряется фазовый сдвиг $\varphi(\omega T)$ выходного сигнала (8) относительно входного (7).

Анализ АЧХ и ФЧХ позволяет сделать вывод, что стабилизатор (6) плохо подавляет высокочастотные помехи и вносит в систему при низких частотах значительное фазовое опережение. Объясняется это тем, что в схеме, представленной на рис. 1, на фильтр Ланцоша подается зашумленный высокочастотными помехами входной сигнал, а подавление фильтром Ланцоша высоких частот не всегда эффективно.

В связи с изложенным, предлагается использование последовательно-параллельной схемы ПД-стабилизатора, представленной на рис. 3. В этой схеме на фильтр Ланцоша подается уже отфильтрованный низкочастотным фильтром Баттеруорта сигнал $\tilde{\varphi}(nT)$.

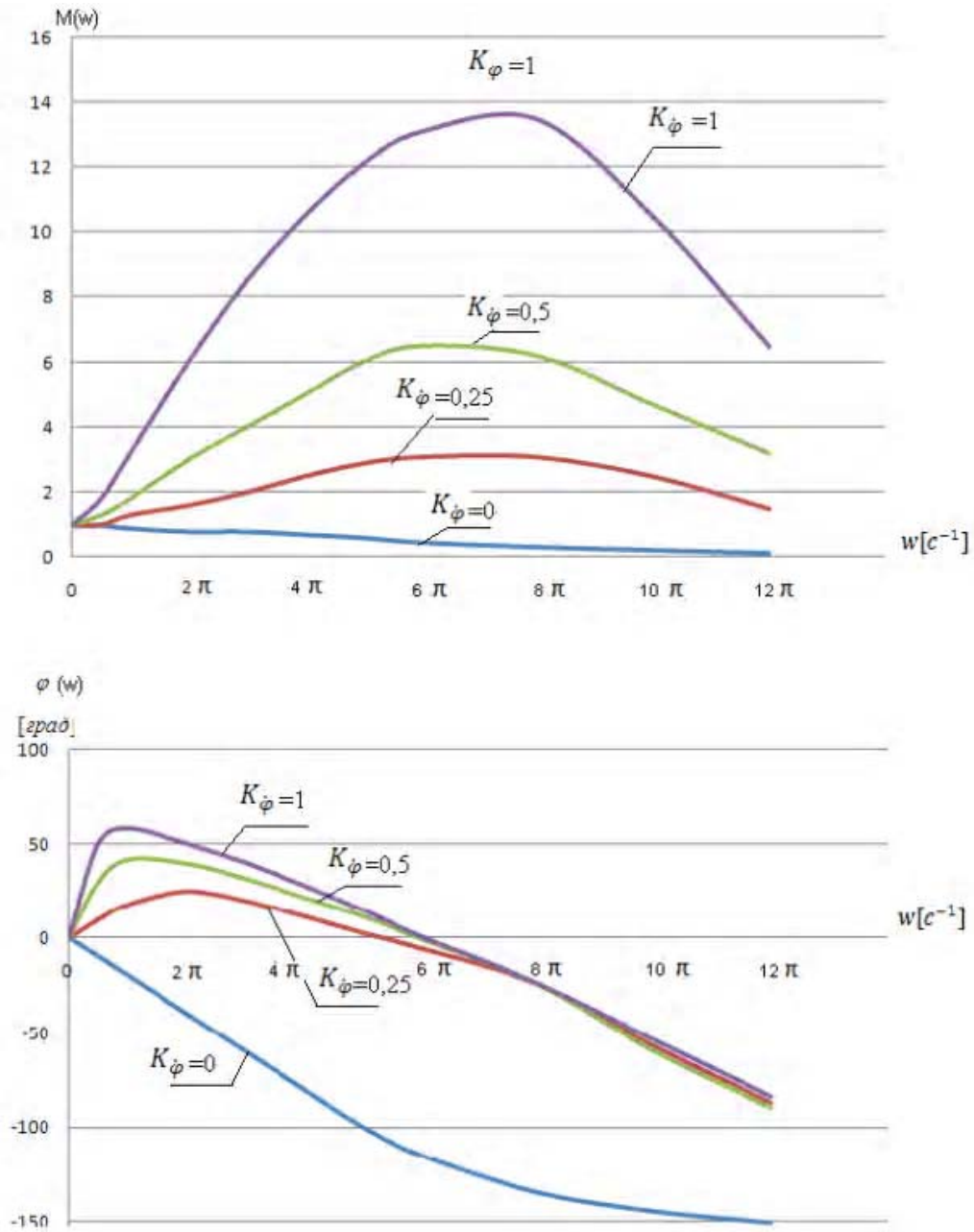


Рис. 2. АЧХ и ФЧХ ПД-стабилизатора, выполненного по параллельной схеме

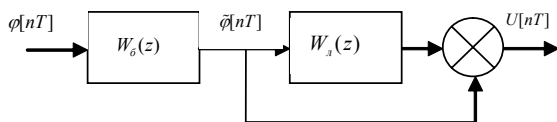


Рис. 3. ПД-стабилизатор, выполненный по последовательно-параллельной схеме

$$W_{\text{ПД}}(z) = \frac{k_\phi \{d_0 + d_1 z^{-1} + d_2 z^{-2} + k_\phi (f_0 + f_1 z^{-1} + f_2 z^{-2} - f_2 z^{-4} - f_1 z^{-5} - f_0 z^{-6})\}}{1 + l_1 z^{-1} + l_2 z^{-2}}, \quad (9)$$

где постоянные f_0 , f_1 и f_2 равны:

$$f_0 = \frac{a_0 c_0}{b_0}; \quad f_1 = \frac{a_0 c_1 + a_1 c_0}{b_0}; \quad f_2 = \frac{a_0 c_2 + a_1 c_1 + a_2 c_0}{b_0},$$

В этом случае передаточная функция стабилизатора при использовании фильтров второго порядка записывается в виде

а соответствующее передаточной функции (9) разностное уравнение ПД-стабилизатора имеет вид

$$\begin{aligned}
 U[nT] = & k_{\varphi} \{ d_0 \varphi[nT] + d_1 \varphi[(n-1)T] + \\
 & + d_0 \varphi[(n-2)T] + k_{\dot{\varphi}} (f_0 \varphi[nT] + \\
 & + f_1 \varphi[(n-1)T] + f_2 \varphi[(n-2)T] - \\
 & - f_2 \varphi[(n-4)T] - f_1 \varphi[(n-5)T] - \\
 & - f_0 \varphi[(n-6)T]) \} - l_1 U[(n-1)T] - \\
 & - l_2 U[(n-2)T].
 \end{aligned} \quad (10)$$

На рис. 4 приведены АЧХ и ФЧХ стабилизатора (10). Подавление высоко-частотных помех стабилизатором

(10) значительно эффективнее, чем стабилизатором (6), а фазовое опережение, вносимое фильтром Ланцоша, значительно ниже, чем при использовании стабилизатора (6). Кроме того, выбором значения коэффициента $k_{\dot{\varphi}}$ при определенной частоте фазовый сдвиг может быть установлен равным нулю. Обычно это целесообразно при собственной частоте колебаний подвижного объекта. Так, для объектов бронетехники собственная частота колебаний поддресоренной массы обычно составляет 1,2 – 1,5 Гц.

Анализ цифрового ПД-стабилизатора, выполненного по последовательно-параллельной схеме, показывает, что его высокие динамические свойства обусловлены тем, что на вход дифференцирующего фильтра Ланцоша подается выходной сигнал фильтра Баттеруорта,

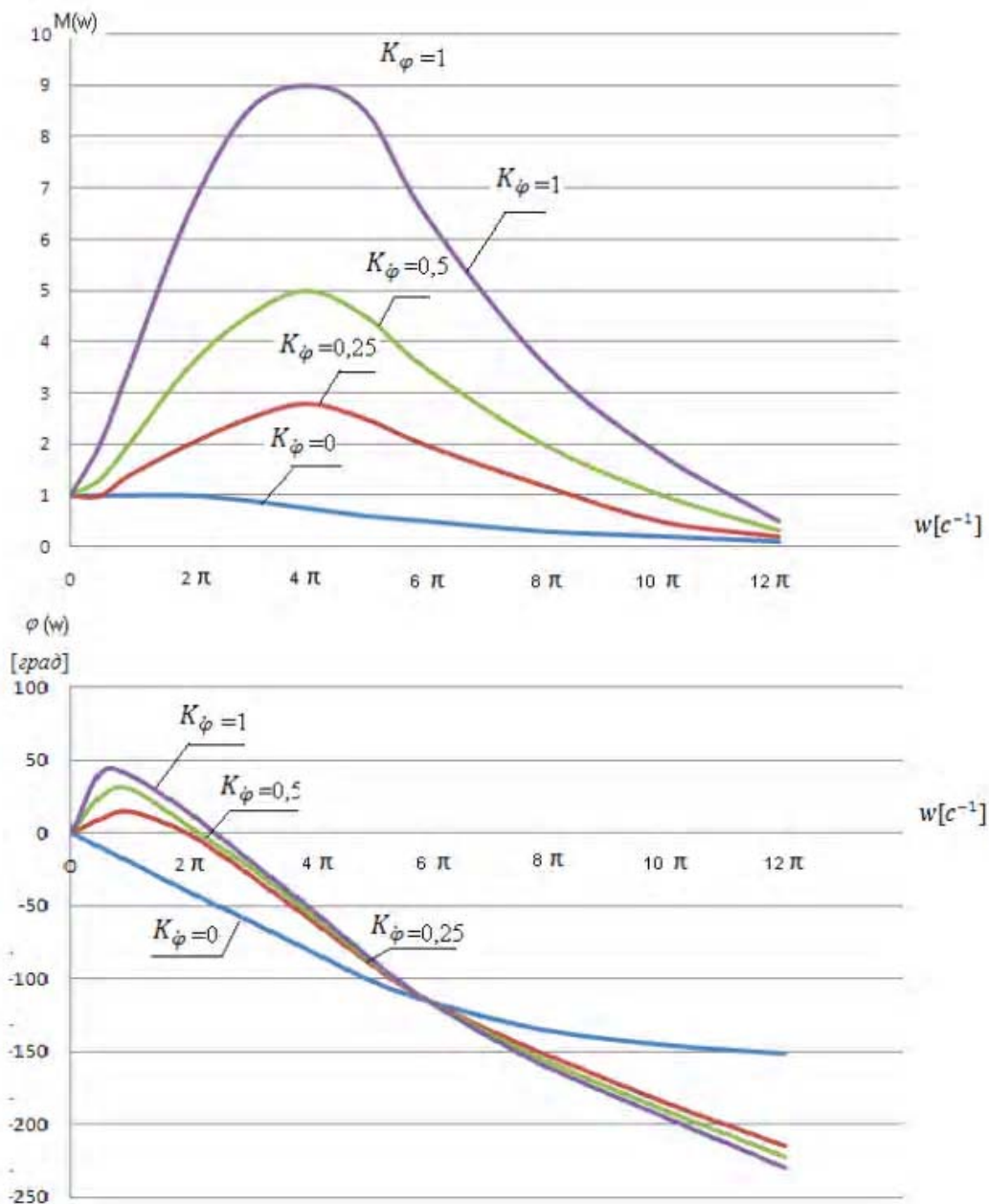


Рис. 4. АЧХ и ФЧХ ПД-стабилизатора выполненного по последовательно-параллельной схеме

который практически не содержит высокочастотных помех. Рассмотрим ПД-стабилизатор, представленный на рис. 3, с фильтром Баттеруорта третьего порядка с передаточной функцией

$$W_6(z) = \frac{a_0(1+3z^{-1}+3z^{-2}+z^{-3})}{1+d_{10}z^{-1}+d_{20}z^{-2}+b_{30}z^{-3}}, \quad (11)$$

и фильтром Ланцоша первого порядка с передаточной функцией

$$W_6(z) = c_2(1-z^{-2}). \quad (12)$$

В результате получаем передаточную функцию ПД-стабилизатора в виде

$$W_{\text{ПД}}(z) = \frac{k_{\phi}a_0\{1+3z^{-1}+3z^{-2}+z^{-3} + k_{\phi}c_2(1+3z^{-1}+2z^{-2}-2z^{-3} - 3z^{-4}-z^{-5})\}}{1+d_{10}z^{-1}+d_{20}z^{-2}+d_{30}z^{-3}}, \quad (13)$$

Соответствующий передаточной функции (13) алгоритм стабилизации записывается в виде

$$\begin{aligned} U[nT] = & k_{\phi}a_{10}\{\phi[nT] + 3\phi[(n-1)T] + \\ & + 3\phi[(n-2)T] + \phi[(n-3)T] + \\ & + k_{\phi}c_2(\phi[nT] + 3\phi[(n-1)T] + \\ & + 2\phi[(n-2)T] - 2\phi[(n-3)T] - \\ & - 3\phi[(n-4)T] - \phi[(n-5)T])\} - \\ & - d_{10}U[(n-1)T] - d_{20}U[(n-2)T] - \\ & - d_{30}U[(n-3)T]. \end{aligned} \quad (14)$$

Пусть параметры фильтров (11) и (12) составляют $a_{10}=0,012045$; $d_{10}=-1,93847$; $d_{20}=1,37229$; $d_{30}=-0,33745$; $c_2=12,5$.

Расчет АЧХ и ФЧХ ПД-стабилизатора (14) позволяет сделать вывод о том, что динамические характеристики стабилизаторов (10) и (14) практически совпадают. Некоторое ухудшение фильтрующих свойств фильтра Ланцоша первого порядка по сравнению с фильтром второго порядка компенсируется использованием фильтра Баттеруорта третьего порядка с практически идеальной АЧХ, который весьма эффективно фильтрует высокочастотные помехи.

Таким образом, цифровые ПД-стабилизаторы подвижных объектов целесообразно формировать по последовательно-параллельной схеме в соответствии с алгоритмами стабилизации (10) и (14).

СПИСОК ЛІТЕРАТУРИ

1. Корнеев, В. В. Основы автоматики и танковые автоматические системы / В. В. Корнеев, Л. П. Кузьмин, К. И. Павличук, М. И. Кузнецов. – М. : Академия БТВ, 1976. – 545 с.
2. Аблесімов, О. К. Автоматичне керування рухомими об'єктами і технологічними процесами. т. 3. Автоматичне керування озброєнням танків / О. К. Аблесімов, Є. Є. Александров, І. Є. Александрова. – Харків : НТУ «ХПІ», 2008. – 444 с.
3. Александров, Е. Е. Оценка точности стабилизации поля зрения прицела танковой пушки / Е. Е. Александров, Т. Е. Александрова, А. А. Лазаренко // Артиллерийское и стрелковое вооружение. – 2011. – № 3. – с. 40–44.
4. Хемминг, Р. В. Цифровые фильтры / Р. В. Хемминг. – М. : Недра, 1984. – 221 с.
5. Ивашко, А. В. Методы и алгоритмы цифровой обработки сигналов / А. В. Ивашко. – Харьков: НТУ «ХПИ», 2005. – 240 с.
6. Александров, Е. Е. Синтез цифровых нерекурсивных фильтров для информационно-управляющих систем / Е. Е. Александров, Т. Е. Александрова, В. А. Кононенко / Техническая электродинамика. Специальный выпуск. Силовая электроника и энергоэффективность. – 2011 – т. 1. – с. 163–168.
7. Александров, Е. Е. Сравнительный анализ цифровых дифференцирующих фильтров / Е. Е. Александров, Т. Е. Александрова, В. А. Кононенко // В кн. : Автоматика / Automatics – 2011. XVIII Міжнародна конференція з автоматичного управління. Матеріали конференції. – Львів : Видавництво Львівської політехніки, 2011. – С. 303–304.

Стаття надійшла до редакцій 25.05.2011.

Александрова Т. Є., Кононенко В. О., Лазаренко А. О., Зейн Алі

ПОРІВНЯЛЬНИЙ АНАЛІЗ ЦИФРОВИХ ПД-СТАБІЛІЗАТОРІВ РУХОМИХ ОБ'ЄКТІВ З НИЗЬКОЧАСТОТНИМИ ФІЛЬТРАМИ БАТТЕРУОРТА І ЛАНЦОША

Розглядається проблема побудови цифрових ПД-стабілізаторів рухомих об'єктів із спотвореними завадами вхідними сигналами. Проведено порівняльний аналіз стабілізаторів з фільтрами Баттеруорта і Ланцоша різних порядків.

Ключові слова: цифровий ПД-стабілізатор, низькочастотний фільтр, диференціюючий фільтр.

Alexandrova T. Ye., Kononenko V. A., Lazarenko A. A., Zein Ali

COMPARATIVE ANALYSIS OF DIGITAL PD-STABILIZERS OF MOBILE OBJECTS WITH LOW-FREQUENCY BUTTERWORT AND LANCZOS FILTERS

The problem of constructing a digital PD-stabilizers of mobile objects with noise input signals is discussed. The comparative analysis of the stabilizers with Butterworth and Lanczos filters of the various orders is considered.

Key words: digital PD-stabilizer, low-pass filter, differentiating filter.

ЕКСТРЕМАЛЬНА ЕЛЕКТРОМЕХАНІЧНА СИСТЕМА КЕРУВАННЯ ПАРАЛЕЛЬНО З'ЄДНАНИМИ НАСОСАМИ ВОДОПОСТАЧАННЯ

У статті йдеться про застосування екстремального енергоефективного керування для двох паралельно з'єднаних насосів. Один з насосів укомплектований частотно-керованим асинхронним електроприводом. Представлено опис алгоритму екстремального керування, його математична модель та структурна схема системи.

Ключові слова. Насос водопостачання, електромеханічна система, екстремальне керування, енергоефективність, регульований електропривод, паралельне з'єднання насосів.

ВСТУП

Системи водопостачання та насосні установки відносяться до числа найбільш енергоспоживаючих технологічних об'єктів. У зв'язку з постійним подорожчанням та вичерпуванням енергетичних та водних ресурсів питанням енергоефективного керування насосних установок почали приділяти значної уваги. Саме тому, все більш актуальним стає питання розробки та впровадження енергоощадливих та енергоефективних технологій у даній сфері. На практиці широке розповсюдження знайшли системи стабілізації напору та програмне керування насосами на основі графіків добового водоспоживання [1]. Більша енергоефективність може бути досягнута при максимально можливому зменшенні швидкості, яке допустиме для конкретного режиму водоспоживання, та забезпеченні продуктивності насоса, яка відповідає потребам споживачів. Так звані «інтерактивні» алгоритми керування забезпечують автоматичний пошук мінімально-достатньої швидкості насосу, яка задовольняє потреби споживачів [2, 3]. Розглянуті вище способи забезпечують певний рівень енергозбереження, але оскільки робоча точка не завжди опиняється на лінії максимального ККД при використанні вищерозглянутих алгоритмів керування, енергоефективність системи може знижуватися. Ця задача розв'язується за допомогою екстремальної системи керування ККД насосної установки [4]. Також використовуються двоканальні екстремальні системи автоматичного керування насосними установками [5], що дозволяють не тільки підвищити ККД насосу, але й зменшити втрати у привідному двигуні. Але у багатьох випадках системи водопостачання не обмежуються одним насосом. Застосовуються послідовне, паралельне та змішане з'єднання для реалізації певних режимів роботи. Застосування екстремального керування для послідовного з'єднання насосів, один з яких керований за швидкістю, дозволяє підвищити ККД керованого насосу [4]. Цей алгоритм може бути застосований також для паралельного з'єднання агрегатів з метою підвищення енергоефективності системи в цілому.

МЕТА РОБОТИ

Метою роботи є підвищення енергоефективності електромеханічної системи автоматичного керування паралельно з'єднаними насосами водопостачання шляхом розробки екстремального алгоритму керування ККД керованого та некерованого насосів. Працездатність системи перевіряється шляхом математичного моделювання.

ЕКСТРЕМАЛЬНИЙ АЛГОРИТМ КЕРУВАННЯ ККД КЕРОВАНОГО ЗА ШВИДКІСТЮ НАСОСА ДЛЯ ПАРАЛЕЛЬНО З'ЄДНАНИХ АГРЕГАТІВ

Розглянемо роботу екстремального алгоритму для паралельного з'єднання агрегатів, враховуючи, що один з насосів керований за швидкістю, інший – некерований. Нехай робоча точка керованого насосу A_1 знаходиться зліва від лінії максимального ККД на характеристиці 1, як показано на рис. 1. Характеристика некерованого насосу має вигляд 2, сумарна – 3. У деякий момент часу екстремальний енергоефективний контролер зменшує швидкість обертання регульованого насосу на фіксоване значення $\Delta\omega$. Його напірна характеристика позначена 5, а сумарна характеристика обох насосів – 7. Новими робочими точками відповідно стануть B , B_1 і B_2 . З метою стабілізації продуктивності до значення Q_A споживачі змушені зменшити гідравлічний опір мережі до значення, при якому характеристика мережі прийме вигляд 6. Робочі точки перейдуть у положення C , C_1 та C_2 (для некерованого насосу Н2). На наступному кроці алгоритму контролер знову зменшить оберти першого насосу на фіксоване значення $\Delta\omega_1$ (характеристика 9, сумарна – 8) і перемістить робочі точки в D , D_1 та D_2 (для насосу Н2). Це викличе відповідну реакцію споживачів до стабілізації витрат води.

Такий процес відбуватиметься доти, поки робоча точка насосу Н1 не виявиться праворуч кривої максимального ККД (точка G_1). Після цього контролер фіксовано збільшить оберти цього насосу, що призведе до збільшення продуктивності та необхідності споживачам прикри-

Рівняння для вихідної заданої частоти на основі структурної схеми матиме вигляд

$$\Delta f = \frac{KT_0}{z-1} \text{sign}(\omega_1 - \omega_1^*). \quad (1)$$

Перетворимо рівняння (1) для запису у різниці

$$\Delta f((n+1)T_0) - \Delta f(nT_0) = KT_0 \text{sign}(\omega_1(nT_0) - \omega_1^*(nT_0)). \quad (2)$$

Враховуючи, що бажана швидкість ω_1^* визначається на основі величини бажаної продуктивності при максимальному ККД, рівняння (2) переписується у вигляді

$$\Delta f((n+1)T_0) = \Delta f(nT_0) + KT_0 \text{sign}(\omega_1(nT_0) - f(Q_1(nT_0))). \quad (3)$$

ОСОБЛИВОСТІ ЗАСТОСУВАННЯ ЕКСТРЕМАЛЬНОГО АЛГОРИТМУ КЕРУВАННЯ ККД КЕРОВАНОГО ЗА ШВИДКІСТЮ НАСОСА ДЛЯ ПАРАЛЕЛЬНО З'ЄДНАНИХ АГРЕГАТИВ З ЖОРСТКИМИ НАПІРНИМИ ХАРАКТЕРИСТИКАМИ

У випадку, якщо напірні характеристики насосів мають велику жорсткість, траєкторія руху робочої точки може не перетнути лінію максимального ККД насосу при застосуванні екстремального алгоритму через достатньо велику зміну величини продуктивності навіть при невеликій зміні частоти обертання, як показано на рис. 3.

Як видно з рисунку, за рахунок майже лінійної ділянки сумарних напірних характеристик, робоча точка регульованого насосу буде рухатися по траєкторії $A_1 - B_1 - C_1 - D_1 - E_1$, у той час як робоча точка на виході системи буде рухатися по траєкторії $A - B^* - C^* - D^* - E^*$. Точками на траєкторії регульованого насосу, яким відповідають точки B^* , C^* , D^* та E^* можна знехтувати, враховуючи велику жорсткість напірних характеристик. Траєкторія руху робочої точки нерегульованого насосу не зміниться від зображеної на рис. 1. Подальший такий

рух без обмеження за швидкістю може призвести до того, що робоча точка регульованого насосу перейде на вісь H , тобто його продуктивність стане нульовою. Вихідна продуктивність буде дорівнювати продуктивності нерегульованого насосу.

Якщо робоча точка буде знаходитися справа від лінії максимального ККД, вона також не опиниться на ній через роботу алгоритму на збільшення швидкості у цій зоні.

У такому випадку екстремальний алгоритм необхідно скоректувати. Розташування робочої точки зліва від лінії максимального ККД не призведе до підвищення ефективності, тому розглянемо випадок, коли остання розташована справа. Якщо змінити напрям роботи алгоритму у цій зоні, то при зменшенні швидкості робоча точка може опинитися на лінії максимального ККД. Рівняння регулятора з урахуванням зміни напрямку роботи алгоритму матиме вигляд

$$\Delta f((n+1)T_0) = \Delta f(nT_0) + KT_0 \text{sign}(f(Q_1(nT_0)) - \omega_1(nT_0)). \quad (4)$$

Принцип роботи регулятора зображений на рис. 4. Як видно з рис. 4, траєкторія робочих точок $A_1 - B_1 - C_1 - D_1$ регульованого насосу перетинає лінію максимального ККД за умови, що рух починається зліва і споживачі реагують на зміну продуктивності. Після перетину лінії максимального ККД ця траєкторія буде визначатися послідовними переходами з точки D_1 в C_1 і навпаки, які забезпечують коливання в зоні максимального ККД регульованого насосу. Незначними переходами, що відбуваються при стабілізації споживачами продуктивності нехтуємо, при умові, що напірні характеристики мають досить високу жорсткість.

Таке розташування вихідної робочої точки найчастіше застосовується у системах, що містять паралельне з'єднання насосів, для збільшення вихідної продуктивності. Тому екстремальний алгоритм у такій формі найкраще підходить для паралельного з'єднання насосів з жорсткими напірними характеристиками.

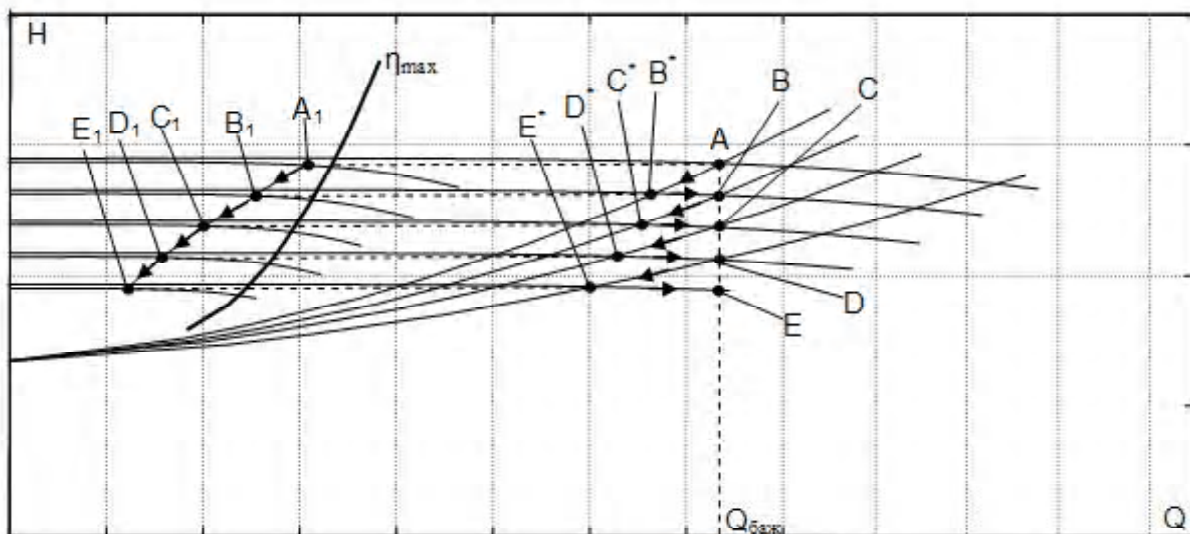


Рис. 3. Траєкторія руху робочої точки регульованого насосу з жорсткими напірними характеристиками

ЕКСТРЕМАЛЬНИЙ АЛГОРИТМ КЕРУВАННЯ ККД НЕКЕРОВАНОГО ЗА ШВИДКІСТЮ НАСОСА ДЛЯ ПАРАЛЕЛЬНО З'ЄДНАНИХ АГРЕГАТІВ

При паралельному з'єднанні насосних агрегатів можливий варіант реалізації екстремального керування ККД для некерованого за швидкістю насосом за рахунок зміни швидкості керованого.

Структурна схема екстремального регулятора представлена на рис. 5.

На структурній схемі введені наступні позначення: Q_2 – фактична виміряна продуктивність нерегульованого насосу; Q_2^* – продуктивність нерегульованого насоса при якій забезпечується максимальний ККД (для окремого насоса – постійна величина, та як його швидкість незмінна, а на кожній напірній характеристиці існує тільки одне значення продуктивності, при якій забезпечується максимальний ККД); T_0 – період квантування екстремального контролера; Δf^* – необхідна величина зменшення завдання частоти регульованого насосу; K – параметр, що задає амплітуду зміни завдання частоти.

Рівняння для вихідної заданої частоти на основі структурної схеми матиме вигляд

$$\Delta f = \frac{KT_0}{z-1} \text{sign}(Q_2 - Q_2^*). \quad (5)$$

Перетворимо рівняння (5) для запису у різниціях

$$\Delta f((n+1)T_0) = \Delta f(nT_0) + KT_0 \text{sign}(Q_2(nT_0) - Q_2^*(nT_0)). \quad (6)$$

Алгоритм для такого варіанту ілюструється на рис. 6.

Нехай робоча точка A_2 знаходиться зліва від лінії максимального ККД нерегульованого насосу Н2. У момент часу T_0 екстремальний енергоефективний контролер зменшує швидкість обертання регульованого насоса на фіксоване значення $\Delta\omega$. Його напірна характеристика позначена 5, а сумарна характеристика обох насосів – 7. Робоча точка нерегульованого насоса переходить у B_2 .

З метою стабілізації продуктивності до значення Q_4 споживачі змушені зменшити гідравлічний опір мережі до значення, при якому характеристика мережі прийме вигляд 6. Робочі точки перейдуть у положення С, C_1 та C_2 (для нерегульованого насосу Н2). На наступному кроці

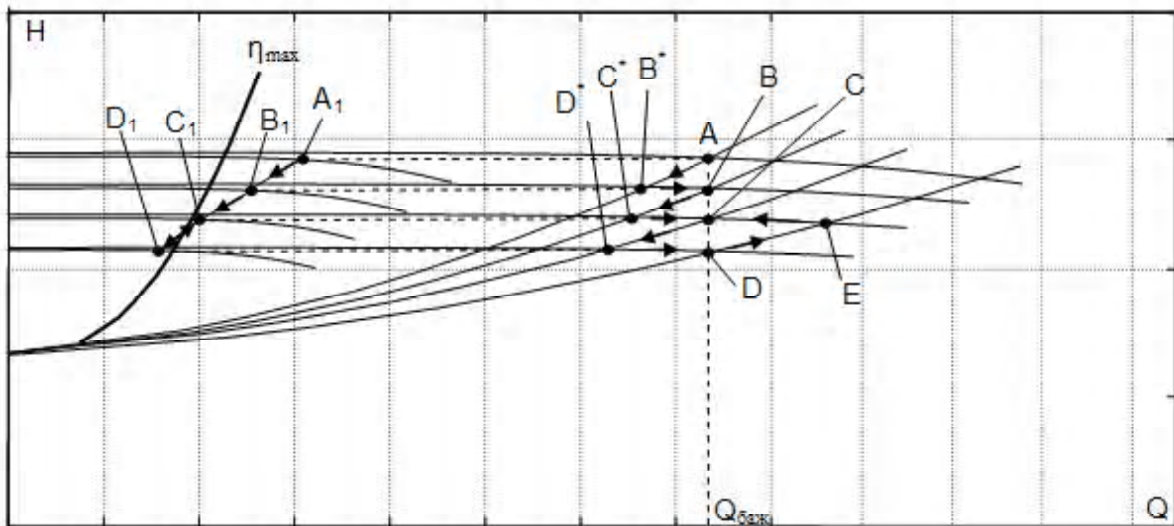


Рис. 4. Траєкторія руху робочої точки регульованого насоса з жорсткими напірними характеристиками для другої зони

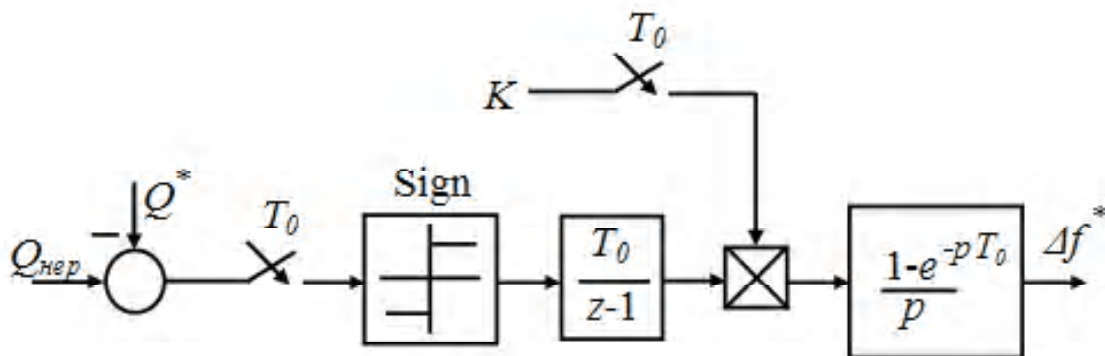


Рис. 5. Структурна схема екстремального енергоефективного регулятора ККД для нерегульованого за швидкістю насоса

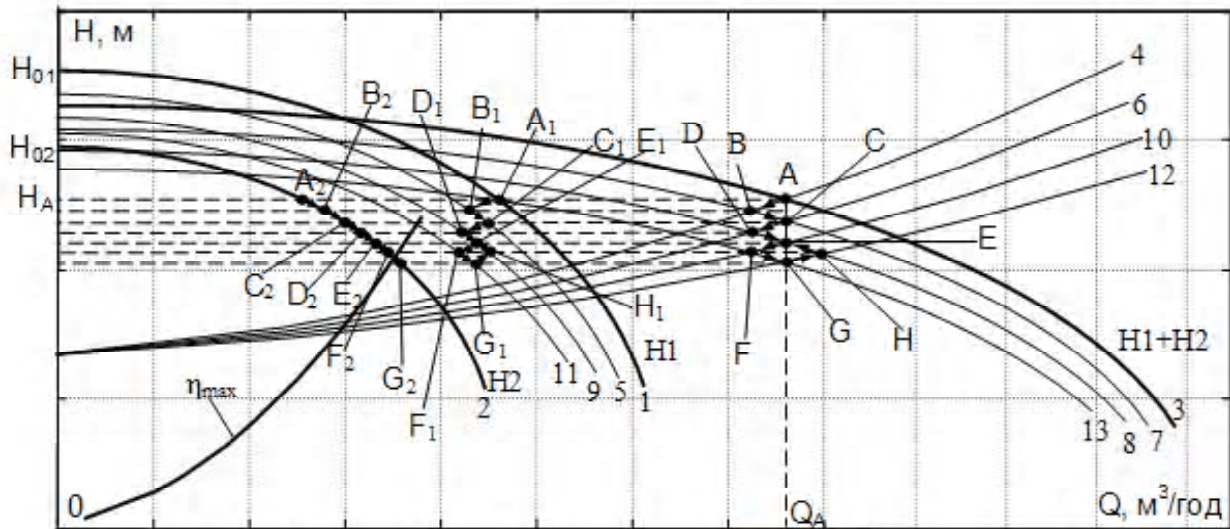


Рис. 6. Робота екстремального алгоритму при паралельному з'єднанні насосних агрегатів для нерегульованого за швидкістю насосу

алгоритму $2T_0$ контролер знову зменшить оберти першого насоса на фіксоване значення $\Delta\omega_1$ (характеристика 9, сумарна – 8) і перемістить робочі точки в D , D_1 та D_2 (для насоса H2). Це викличе відповідну реакцію споживачів до стабілізації витрат води.

Такий процес відбуватиметься доти, поки робоча точка насоса H2 не виявиться праворуч кривої максимального ККД (точка G_2), тобто Q_2 стане більше Q_2^* . Після цього контролер фіксовано збільшить оберти цього насоса, що призведе до збільшення продуктивності та необхідності споживачам прикривати крани. У результаті робота першого насоса характеризуватиметься циклічною послідовністю наступних робочих точок $E_1 - F_1 - G_1 - H_1 - E_1$. Для нерегульованого насоса послідовність буде наступною $E_2 - F_2 - G_2 - F_2 - E_2$. У результаті роботи алгоритму ККД некерованого насоса буде коливатися в околі значення, що визначається продуктивністю в точці G .

Якщо споживачі не реагуватимуть на зміну продуктивності, то за рахунок руху по напірній характеристиці вправо, робоча точка також опиниться за лінією максимального ККД і далі буде рухатися в її околі за рахунок зміни швидкості обертання керованого насоса.

Якщо вихідна робоча точка опиниться у другій зоні, то алгоритм буде працювати аналогічно на збільшення швидкості керованого насоса.

МАТЕМАТИЧНА МОДЕЛЬ ЕКСТРЕМАЛЬНОЇ СИСТЕМИ АВТОМАТИЧНОГО КЕРУВАННЯ ПАРАЛЕЛЬНО З'ЄДНАНИМИ НАСОСАМИ

Математична модель двох паралельно з'єднаних насосів описується наступними рівняннями [6]:

$$H = H_{01n}\omega_1^2 / \omega_{n1}^2 - a_{n1}Q_1^2 - \chi_1 dQ_1 / dt, \quad (7)$$

$$H = H_{02n}\omega_2^2 / \omega_{n2}^2 - a_{n2}Q_2^2 - \chi_2 dQ_2 / dt, \quad (8)$$

$$H = H_{cm} + aQ^2 - \chi dQ / dt, \quad (9)$$

$$Q = Q_1 + Q_2, \quad (10)$$

$$M_{C1} = \rho g Q_1 H / \eta_1 \omega_1, \quad (11)$$

$$M_{C2} = \rho g Q_2 H / \eta_2 \omega_2, \quad (12)$$

де Q_1 , Q_2 – продуктивності першого та другого насосів відповідно; Q – сумарна продуктивність першого та другого насоса; H_{01n} та H_{02n} – номінальні напори при нульових подачах першого та другого насосів при номінальних швидкостях відповідно; ω_1 , ω_2 – швидкості обертання першого та другого насосів відповідно; ω_{n1} , ω_{n2} – номінальні швидкості обертання першого та другого насосів відповідно; χ_1 , χ_2 – сталі часу інтегрування першого та другого насосів відповідно; χ – загальна стала часу інтегрування; H_{cm} – геодезична висота підйому води; a_{n1} , a_{n2} – номінальні гідравлічні опори першого та другого насосів відповідно; a – гідравлічний опір мережі; M_{C1} , M_{C2} – моменти навантаження на валах двигунів першого та другого насосів відповідно; ρ – густина води; g – прискорення вільного падіння; η_1 , η_2 – ККД першого та другого насосів відповідно; H – напір першого та другого насоса; t – час.

Наявність зворотних клапанів на виході кожного з насосів враховується наступним чином: якщо продуктивність насоса стає від'ємною, то вона приймається рівною нулю.

ККД насосів є нелінійними функціями, які залежать від положень робочих точок на напірних характеристиках $\eta_i = f(Q_i, H_i)$. В роботі ці залежності апроксимуються двошаровими нейронними мережами типу «feed-forward backpropagation» з 10, 1 нейронами та функціями активації tansig, purelin у відповідних шарах. Для навчання нейронних мереж використана база даних 4000 робочих точок з каталогу фірми виробника [7].

Крива максимального ККД керованого за обертами насоса також апроксимована аналогічною нейронною мережею, навченою на основі даних 250 робочих точок.

Привідні асинхронні двигуни насосів описуються нелінійними двофазними моделями в нерухомій системі координат статора а–b [8]. Перетворювач частоти реалізує квадратичний закон керування $u/f^2 = \text{const}$ [4].

Структурна схема екстремальної системи керування показана на рис. 7.

ДОСЛІДЖЕННЯ ЕКСТРЕМАЛЬНОЇ СИСТЕМИ КЕРУВАННЯ

Для дослідження вибрано два однакові насоси фірми Vogel Pumpen CNX-100-65-400 [7] з наступними максимальними даними: потужність 156,1 кВт, напір 240 м, продуктивність 164,9 м³/год, ККД 62,9 %, частота обертання 2900 об/хв. Номінальні дані приводних асинхронних двигунів наступні: потужність 160 кВт, синхронна частота обертання 3000 об/хв, ККД 96 %, лінійна напруга 380 В, коефіцієнт потужності 0,9, ковзання 0,019, активний опір статора 0,0117 Ом, індуктивність статора 0,0126 Гн, активний опір ротора 0,0094 Ом, індуктивність ротора

0,0127 Гн, взаємна індуктивність статора та ротора 0,0124 Гн. При моделюванні прийнято $H_{cm} = 150$ м, $\chi = 0,5$ с/(м²/год), $\chi_1 = \chi_2 = 0,05$ с/(м²/год), приведений момент інерції на валах двигунів $J = 1,4$ кг м². Розрахункові параметри насосів на основі напірних характеристик є наступними: $H_{01} = H_{02} = 235$ м, $a_{n1} = a_{n2} = 9018$ м/(м³/с)². Напірні характеристики насосів досить жорсткі, тому під час моделювання використані алгоритми (4) та (6).

Результати досліджень екстремальної системи керування ККД керованого насосу представлені на рис. 8, а ККД некерованого насосу – на рис. 9.

При дослідженні екстремальної системи керування ККД керованого насосу вихідна робоча точка приймалась такою, щоб робоча точка керованого насоса знаходилась праворуч від кривої максимального ККД. Поведінка споживачів моделювалася як система стабілізації продуктивності з ПІ-регулятором. Результати досліджень показали, що швидкість керованого насосу ω_1 зменшується, доки робоча точка не опиниться зліва від лінії максимального ККД, при цьому задана швидкість ω_1^*

починає зменшуватися а швидкість двигуна ω_1 збільшуватись. У результаті ККД керованого насосу η_1 буде коливатися навколо максимального значення, а ККД некерованого насосу η_2 буде коливатися навколо деякого значення, яке залежить від початкового розташування робочої точки. Траєкторія руху робочої точки відповідає алгоритму (4).

Підключення регулятора ККД некерованого насосу показано пунктирними лініями на рис. 7. При дослідженні

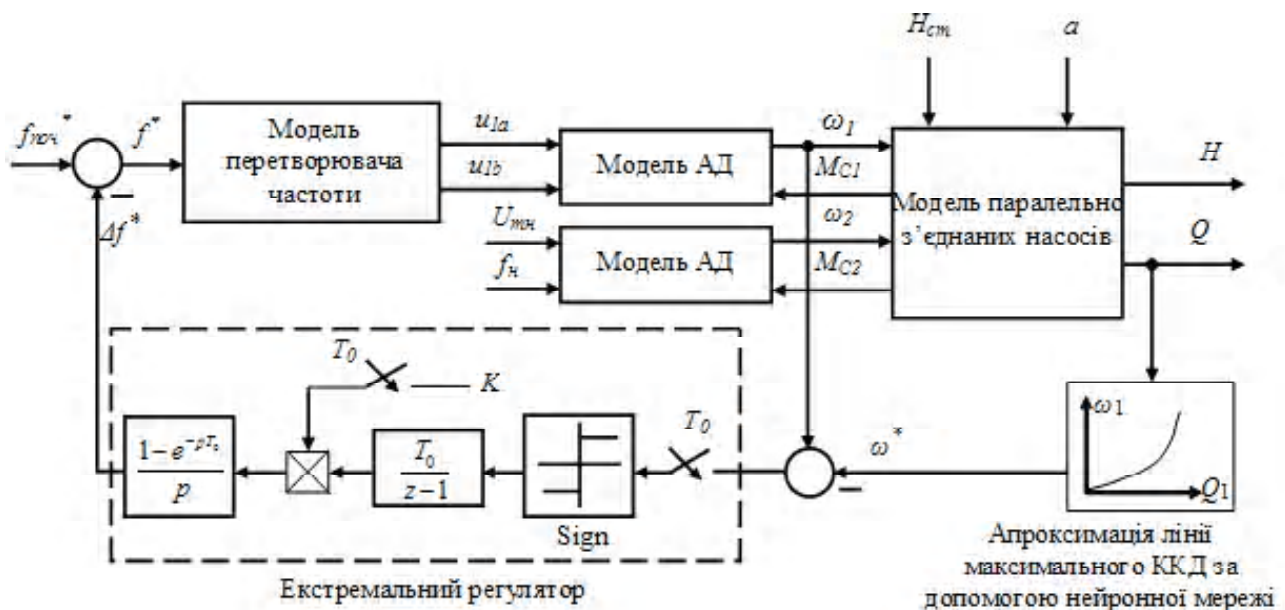


Рис. 7. Структурна схема екстремальної системи керування

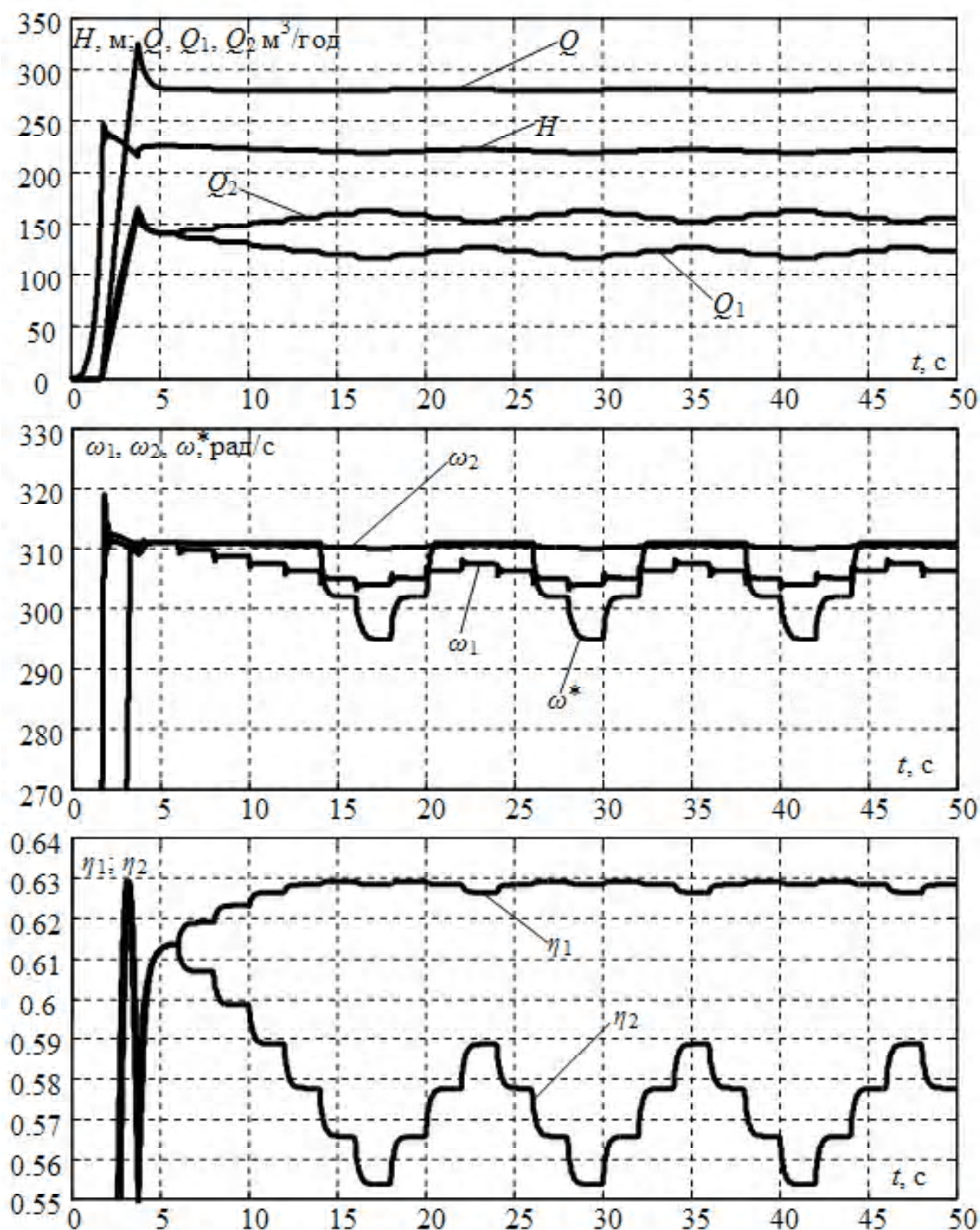


Рис. 8. Дослідження екстремальної системи керування ККД керованого насосу

системи керування вихідна робоча точка насосу знаходилася ліворуч від кривої максимального ККД. Результати досліджень показали, що швидкість керованого насосу ω_1 зменшується, доки продуктивність некерованого насосу Q_2 не стане більше за задану продуктивність Q_2^* .

У результаті ККД некерованого насосу η_2 буде коливатися навколо максимального значення, а ККД керованого насосу η_1 буде коливатися навколо деякого значення, яке залежить від початкового розташування робочої точки.

При застосуванні пропонованого алгоритму у гідравлічних системах, що наповнюють великі резервуари, ви-

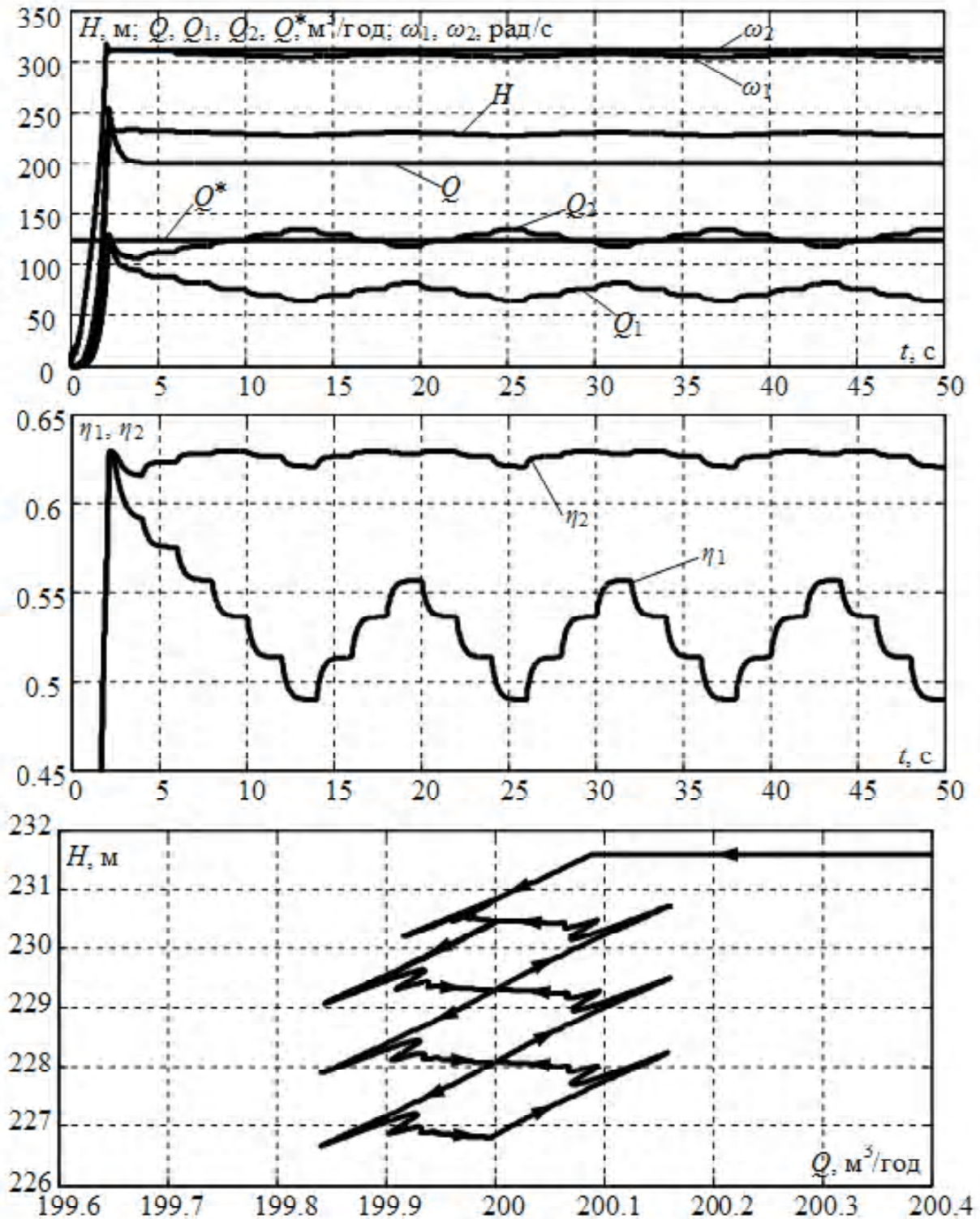


Рис. 9. Дослідження екстремальної системи керування ККД некерованого насосу

моги, щоб споживачі стабілізували продуктивність, є необов'язковими. Дослідження такої системи показані на рис. 10.

Як видно з рис. 10, швидкість регульованого насосу ω_1 поступово зменшується доки робоча точка нерегульованого насосу не опиниться справа від лінії максимального ККД. На 20-й секунді відбувається плавне змен-

шення гідравлічного опору мережі на 10 % від поточного значення, при цьому алгоритм не втрачає свої властивості. ККД некерованого насосу η_2 залишається коливатися в околі максимального значення, а ККД керованого η_1 збільшується. На 50-й секунді відбувається збільшення гідравлічного опору мережі на 20 %. Анало-

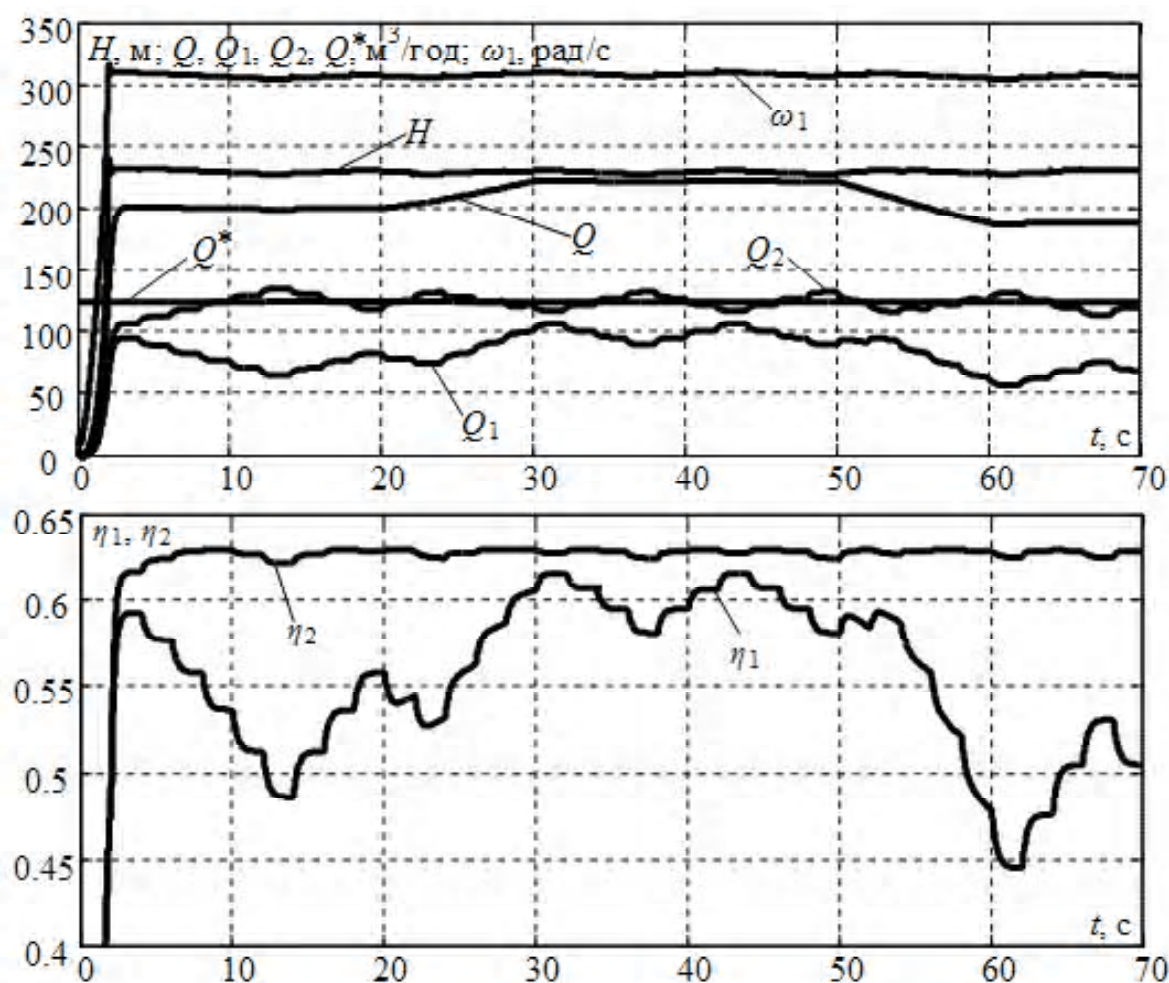


Рис. 10. Дослідження екстремальної системи керування ККД некерованого насосу за відсутності реакції споживачів

гічно попередньому випадку алгоритм залишається працездатним.

ВИСНОВКИ

При паралельному з'єднанні двох насосів, один з яких оснащений керованим асинхронним електроприводом, можливе застосування екстремального енергоефективного керування. При цьому можливий варіант керування ККД керованого насосу та ККД некерованого за рахунок зміни швидкості керованого насосу. Розташування вихідних робочих точок обмежується величиною жорсткості напірних характеристик насосів. Застосування другого варіанту керування може не передбачати реакцію споживачів, а сам алгоритм у цьому випадку нечутливий до зміни гідравлічного опору мережі під час роботи насосів. Діапазон допустимих продуктивностей установки залежить від геодезичної висоти підйому води та мінімально можливого гідравлічного опору мережі.

СПИСОК ЛІТЕРАТУРИ

1. *Kiselychnyk, O.* Overview of energy efficient control solutions for water supply systems / O. Kiselychnyk, M. Bodson, H. Werner // Transactions of Kremenichuk State Polytechnic Univ. Kremenichuk: KSPU. – N3/2009 (56), part 1. – 2009. – P. 40–45.
2. *Попович, М. Г.* Експериментальні дослідження роботи інтерактивного енергозберігаючого контролера на гібридній моделі насосної установки / М. Г. Попович, О. І. Киселичник, С. О. Бур'ян, О. Ф. Соколовський // Вісник Кременчуцького державного політехнічного університету імені Михайла Остроградського. – Кременчук : КДПУ. – Випуск 3/2007 (44), Частина 1. – С. 72–75.
3. *Pechenik, M.* Experimental research of interactive energy saving controller of water supply pump based on flow rate measurement / M. Pechenik, O. Kiselychnyk, S. Buryan // Вісник Національного технічного університету «Харківський політехнічний інститут». Проблеми автоматизованого електроприводу. Теорія і практика. – Харків : НТУ «ХПІ». – 2010. – № 28. – С. 272–274.
4. *Popovich, M.* Extremal electromechanical control system of water supply pumps connected in series / Popovich M., Kiselychnyk O., Buryan S. // Transactions of Kremenichuk State Polytechnic Univ. Kremenichuk : KSPU. – N 3/2010 (62), part 2. – 2010. – P. 37–41.
5. *Бур'ян, С. О.* Двоканальна екстремальна електромеханічна система автоматичного керування насосною установкою / С. О. Бур'ян, Т. В. Гришук // Вісник Національного технічного університету «Харківський політехнічний інститут».

- Проблеми автоматизованого електроприводу. Теорія і практика. – Х. : НТУ «ХПІ», – 2010. – № 28. – С. 176–179.
6. Попович, М. Г. Проблеми теорії автоматизації багатоагрегатних насосних установок на основі принципу пасивності / М. Г. Попович, О. І. Кіселичник // Технічна електродинаміка. Проблеми сучасної електротехніки. – Частина 5. – Київ. – 2006. – С. 54–59.
 7. *ITT Industries*. Pump Selection Program / *ITT Industries* // Vogel Select CD. – Jan. – 2009.
 8. Marino, R. Exponentially convergent rotor resistance estimation for induction motors / R. Marino, S. Peresada, P. Tomei // *IEEE Trans. on Industrial Electronics*. – 1995. – Vol. 42, No. 5. – P. 508–515.

Стаття надійшла до редакції 15.02.2011.

Бурян С. А., Гришук Т. В.

ЭКСТРЕМАЛЬНАЯ ЭЛЕКТРОМЕХАНИЧЕСКАЯ СИСТЕМА УПРАВЛЕНИЯ ПАРАЛЛЕЛЬНО СОЕДИНЕННЫМИ НАСОСАМИ ВОДОСНАБЖЕНИЯ

В статье рассматривается применение экстремального энергоэффективного управления для двух параллельно соединен-

ных насосов. Один из насосов комплектуется частотно-управляемым асинхронным электроприводом. Представлены описание алгоритма экстремального управления, его математическая модель и структурная схема системы.

Ключевые слова. Насос водоснабжения, электромеханическая система, экстремальное управление, энергоэффективность, регулируемый электропривод, параллельное соединение насосов.

Buryan S., Gryshuk T.

EXTREMAL ELECTROMECHANICAL CONTROL SYSTEM OF WATER SUPPLY PUMPS CONNECTED IN PARALLEL

The paper concerns the application of the extremal energy efficient control for two pumps connected in parallel. One of the pumps is with the controlled induction motor drive. The extremal control algorithm description, its mathematical model and the system block diagram are presented.

Key words: Water supply pump, electromechanical system, extremal control, efficiency, controlled drive, parallel connection of pumps.

ДО ВІДОМА АВТОРІВ

Журнал «Радіоелектроніка, інформатика, управління» (PIU) призначений для публікації найбільш значимих наукових і практичних результатів досліджень учених вищих навчальних закладів і наукових організацій.

Журнал включений у перелік наукових видань України, у яких можуть публікуватися результати дисертаційних робіт на здобуття вчених ступенів доктора і кандидата технічних наук і фізико-математичних наук (радіофізика).

Статті, що опубліковано в журналі, реферуються в реферативних журналах і базах даних ВІНГІ (Росія) і «Джерело» (Україна). Журнал міститься у міжнародній базі наукових видань Index Copernicus (<http://journals.indexcopernicus.com/index.php>). Інтернет-сторінка журналу:

<http://journal.zntu.edu.ua/ric/index.php?page=index>.

Журнал видається два рази в рік і розповсюджується за Каталогом періодичних видань України (передплатний індекс – 22914).

Для розгляду питання про публікацію статті в редакцію журналу необхідно вислати поштою або представити особисто наступне:

- 1) рукопис (роздруковку) статті, підписаний на останній сторінці всіма авторами, в двох екземплярах;
- 2) відомості про авторів;
- 3) оригінал експертного висновку про можливість відкритого опублікування статті;
- 4) супровідний лист-клопотання з організації, де була виконана робота (або лист автора);
- 5) рецензію від фахівця в даній області з вченим ступенем доктора наук. Підпис рецензента обов'язково мусить бути завірений.
- 6) диск з наступними файлами:
 - електронна версія статті, повністю ідентична роздруковці;
 - відомості про авторів;
 - рисунки у графічному форматі .tif.

Файли з матеріалами статті можна вислати електронною поштою або передати особисто на оптичному диску або USB-накопичувачі.

Вимоги до оформлення статті. Приймаються статті, набрані в редакторі Microsoft Word.

Параметри сторінки:

- розмір паперу – A4 (210x297);
- орієнтація – книжкова;
- шрифт – Times New Roman, розмір – 12 pt;
- міжрядковий інтервал – полуторний;
- верхнє поле – 20 мм, нижнє – 20 мм, ліве – 25 мм, праве – 15 мм.

Сторінки рукопису повинні бути пронумеровані. Не допускаються розбіжності рукопису з електронною версією статті. Текст рукопису не повинен мати рукописних виправлень та позначок.

Послідовність розміщення матеріалу статті:

- 1) індекс УДК;
- 2) прізвища й ініціали авторів, назва статті, анотація й ключові слова українською мовою (для громадян України);
- 3) прізвища й ініціали авторів, назва статті, анотація й ключові слова російською мовою;
- 4) прізвища й ініціали авторів, назва статті, анотація й ключові слова англійською мовою;
- 5) текст статті;
- 6) список літератури.

Текст статті. Приймаються статті російською, українською та англійською мовами. Розмір статті до 0,5 авторського аркуша. У статті слід уникати зайвої деталізації, проміжних формул і висновків; не слід наводити відомі факти, повторювати зміст таблиць і ілюстрацій у тексті. Стаття не повинна мати граматичних або інших помилок, а також повинна відповідати тематиці журналу й вимогам ВАК щодо фахових видань.

Структура тексту статті мусить містити такі необхідні елементи: постановка проблеми в загальному виді і її зв'язок з важливими науковими або практичними завданнями; аналіз останніх досліджень і публікацій, у яких розпочато розв'язання даної проблеми, і на які опирається автор; виділення нерозв'язаних раніше частин загальної проблеми, яким присвячується стаття; формулювання цілей статті (постановка завдання); виклад основного матеріалу дослідження з повним обґрунтуванням отриманих наукових результатів, висновки по даному дослідженню й перспективи подальших досліджень у даному напрямку. Матеріал публікації мусить бути розбитий на підрозділи не більше двох рівнів.

Рисунки розміщуються в тексті й додатково додаються в окремих файлах (формат .tif з роздільною здатністю 150–300 dpi, чорно-білі або у градаціях сірого). Розмір рисунків не повинен перевищувати ширини сторінки (17 см) або ширини колонки (8 см). Написи на рисунках бажано виконувати шрифтом Times New Roman, розмір 10. Рисунки нумерують і підписують униз.

Формули виконуються за допомогою вбудованого в Word редактора Microsoft Equation. Формули нумерують у круглих дужках праворуч. Формули великого розміру записуються в кілька рядків.

Нумерація рисунків, формул і таблиць наскрізна однакова.

Обсяг **анотації** не повинен перевищувати 40 слів.

Ключові слова наводяться в називному відмінку у кількості до десяти слів.

Список літератури наприкінці статті подається мовою оригіналу і складається в порядку згадування посилань у тексті й відповідно до діючого стандарту на бібліографічний опис. Посилання на літературу в тексті нумеруються поспільно й позначаються цифрою у квадратних дужках.

У відомостях про авторів необхідно навести:

- 1) прізвище, ім'я, по батькові (повністю);
- 2) учений ступінь;
- 3) посаду;
- 4) місце роботи;
- 5) електронну адресу;
- 6) робочий, домашній, мобільний телефони.

Статті, які не відповідають зазначеним вимогам, не приймаються до розгляду.

Всі статті проходять закриті рецензування і в разі потреби можуть бути повернуті автору на доробку. Редакція залишає за собою право на літературну редакцію тексту статті без повідомлення автору.

Рукописи й диски не вертаються, коректура та відбитки статей авторам не надсилаються.

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

Тел.: (061) 7-698-2-96 – редакційно-видавничий відділ
(061) 7-644-6-62 – головний редактор

Факс: (061) 7-642-1-41

E-mail: rvv@zntu.edu.ua

Наукове видання

**Радіоелектроніка,
інформатика,
управління**

№ 2/2011

Науковий журнал

Головний редактор – д-р техн. наук Піза Д. М.

Заст. головного редактора – канд. техн. наук Дубровін В. І.

Комп'ютерний моделювання та верстання

Зуб С. В.

Оригінал-макет підготовлено у редакційно-видавничому відділі ЗНТУ

Підписано до друку 06.12.2011. Формат 60×84/8.

Папір офс.Різогр. друк Ум. друк. арк. 19.

Тираж 300 прим. Зам. № 1789.

69063, м. Запоріжжя, ЗНТУ, друкарня, вул. Жуковського, 64