p-ISSN 1607-3274
e-ISSN 2313-688X

Радіоелектроніка
Інформатика
Управління

Радиоэлектроника
Информатика
Управление

Radio Electronics
Computer Science
Control

2016/1

ISSN 1607-3274
9 771607 327005 61>

Запорізький національний технічний університет

Радіоелектроніка, інформатика, управління

Науковий журнал

Виходить чотири рази на рік
№ 1(36) 2016
Заснований у січні 1999 року.
Засновник і видавець – Запорізький національний технічний університет.
ISSN 1607-3274 (друкований), ISSN 2313-688X (електронний).
Запоріжжя
ЗНТУ
2016

Запорожский национальный технический университет

Радиоэлектроника, информатика, управление

Научный журнал

Выходит четыре раза в год
№ 1(36) 2016
Основан в январе 1999 года.
Основатель и издатель – Запорожский национальный технический университет.
ISSN 1607-3274 (печатный), ISSN 2313-688X (электронный).
Запорожье
ЗНТУ
2016

Zaporizhzhya National Technical University

Radio Electronics, Computer Science, Control

The scientific journal

Published four times per year
№ 1(36) 2016
Founded in January 1999.
Founder and publisher – Zaporizhzhya National Technical University.
ISSN 1607-3274 (print), ISSN 2313-688X (on-line).
Zaporizhzhya
ZNTU
2016

Науковий журнал «Радіоелектроніка, інформатика, управління» (скорочена назва – РІУ) видається Запорізьким національним технічним університетом (ЗНТУ) з 1999 р. періодичністю чотири номери на рік.

Зареєстрований Державним комітетом інформаційної політики, телебачення та радіомовлення 29.01.2003 р. Свідоцтво про державну реєстрацію друкованого засобу масової інформації серія КВ №6904.

ISSN 1607-3274 (друкований), **ISSN** 2313-688X (електронний).

Наказом Міністерства освіти і науки України № 1328 від 21.12.2015 р. «Про затвердження рішень Атестаційної колегії Міністерства щодо діяльності спеціалізованих вчених рад від 15 грудня 2015 року» **журнал включений до переліку наукових фахових видань України,** в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора і кандидата фізико-математичних та технічних наук.

В журналі безкоштовно публікуються наукові статті англійською, російською та українською мовами.

Правила оформлення статей подано на сайті: http://ric.zntu.edu.ua/information/authors.

Журнал забезпечує **безкоштовний відкритий он-лайн доступ** до повнотекстових публікацій.

Журнал дозволяє авторам мати авторські права і зберігати права на видання без обмежень. Журнал дозволяє користувачам читати, завантажувати, копіювати, поширювати, друкувати, шукати або посилатися на повні тексти своїх статей. Журнал дозволяє повторне використання його вмісту у відповідності з CC ліцензією CC-BY.

Опублікованими статтям присвоюється унікальний ідентифікатор цифрового об'єкта DOI.

Журнал реферується та індексується у провідних міжнародних та національних реферативних журналах і наукометричних базах даних, а також розміщується у цифрових архівах та бібліотеках з безкоштовним доступом у режимі on-line (у т. ч. DOAJ, DOI, CrossRef, EBSCO, eLibrary.ru / РИНЦ, Google Scholar, Index Copernicus, INSPEC, ISSN, Ulrich's Periodicals Directory, WorldCat, ВІНІТІ, Джерело), повний перелік яких подано на сайті: http://ric.zntu.edu.ua/about/editorialPolicies#custom-0.

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

Тематика журналу містить: радіофізику, мікро-, нано- і радіоелектроніку, апаратне і програмне забезпечення комп'ютерної техніки, комп'ютерні мережі і телекомунікації, теорію алгоритмів і програмування, оптимізацію і дослідження операцій, міжмашинну і людино-машинну взаємодію, математичне і комп'ютерне моделювання, обробку даних і сигналів, управління в технічних системах, штучний інтелект, включаючи системи, засновані на знаннях, і експертні системи, інтелектуальний аналіз даних, розпізнавання образів, штучні нейронні і нейро-нечіткі мережі, нечітку логіку, колективний інтелект і мультиагентні системи, гібридні системи.

Усі статті, пропоновані до публікації, одержують **об'єктивний розгляд**, що оцінюється за суттю без урахування раси, статі, віросповідання, етнічного походження, громадянства або політичної філософії автора(ів).

Усі статті проходять двоступінчасте закрите (анонімне для автора) **резензування** штатними редакторами і незалежними рецензентами – провідними вченими за профілем журналу.

Рекомендовано до видання вченою радою ЗНТУ, протокол № 10 від 21.03.2016 р.

Журнал зверстаний редакційно-видавничим відділом ЗНТУ.

Веб-сайт журналу: http://ric.zntu.edu.ua.

Адреса редакції: Редакція журналу «РІУ», Запорізький національний технічний університет, вул. Жуковського, 64, м. Запоріжжя, 69063, Україна.

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

Факс: (061) 764-46-62

E-mail: rvv@zntu.edu.ua

Научный журнал «Радиоэлектроника, информатика, управление» (сокращенное название – РИУ) издается Запорожским национальным техническим университетом (ЗНТУ) с 1999 г. периодичностью четыре номера в год.

Зарегистрирован Государственным комитетом информационной политики, телевидения и радиовещания 29.01.2003 г. (Свидетельство о государственной регистрации печатного средства массовой информации серия КВ №6904).

ISSN 1607-3274 (печатный), **ISSN** 2313-688X (электронный).

Приказом Министерства образования и науки Украины № 1328 от 21.12.2015 г. «Об утверждении решений Аттестационной коллегии Министерства относительно деятельности специализированных ученых советов от 15 декабря 2015 года» **журнал включен в перечень научных профессиональных изданий Украины,** в которых могут публиковаться результаты диссертационных работ на соискание ученых степеней доктора и кандидата физико-математических и технических наук.

В журнале бесплатно публикуются научные статьи на английском, русском и украинском языках.

Правила оформления статей представлены на сайте: http://ric.zntu.edu.ua/information/authors.

Журнал обеспечивает **бесплатный открытый он-лайн доступ** к полнотекстовым публикациям. Журнал разрешает авторам иметь авторские права и сохранять права на издание без ограничений. Журнал разрешает пользователям читать, загружать, копировать, распространять, печатать, искать или ссылаться на полные тексты своих статей. Журнал разрешает повторное использование его содержания в соответствии с CC лицензией CC-BY.

Опубликованным статьям присваивается уникальный идентификатор цифрового объекта DOI.

Журнал реферируется и индексируется в ведущих международных и национальных реферативных журналах и наукометрических базах данных, а также размещается в цифровых архивах и библиотеках с бесплатным доступом on-line (в т .ч. DOAJ, DOI, CrossRef, EBSCO, eLibrary.ru / РИНЦ, Google Scholar, Index Copernicus, INSPEC, ISSN, Ulrich's Periodicals Directory, WorldCat, ВИНИТИ, Джэрэло), полный перечень которых представлен на сайте: http://ric.zntu.edu.ua/about/editorialPolicies#custom-0.

Журнал распространяется по Каталогу периодических изданий Украины (подписной индекс – 22914).

Тематика журнала включает: радиофизику, микро-, нано- и радиоэлектронику, аппаратное и программное обеспечение компьютерной техники, компьютерные сети и телекоммуникации, теорию алгоритмов и программирования, оптимизацию и исследование операций, межмашинное и человеко-машинное взаимодействие, математическое и компьютерное моделирование, обработку данных и сигналов, управление в технических системах, искусственный интеллект, включая системы, основанные на знаниях, и экспертные системы, интеллектуальный анализ данных, распознавание образов, искусственные нейронные и нейро-нечеткие сети, нечеткую логику, коллективный интеллект и мультиагентные системы, гибридные системы.

Все статьи, предлагаемые к публикации, получают **объективное рассмотрение**, которое оценивается по существу без учета расы, пола, вероисповедания, этнического происхождения, гражданства или политической философии автора(ов).

Все статьи проходят двухступенчатое закрытое (анонимное для автора) **рецензирование** штатными редакторами и независимыми рецензентами – ведущими учеными по профилю журнала.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Рекомендовано к изданию ученым советом ЗНТУ, протокол № 10 от 21.03.2016.

Журнал сверстан редакционно-издательским отделом ЗНТУ.

Веб-сайт журнала: http://ric.zntu.edu.ua.

Адрес редакции: Редакция журнала «РИУ», Запорожский национальный технический университет, ул. Жуковского, 64, г. Запорожье, 69063, Украина.

Тел.: +38-061-769-82-96 – редакционно-издательский отдел

Факс: (061) 764-46-62

E-mail: rvv@zntu.edu.ua

The scientific journal «Radio Electronics, Computer Science, Control» is published by the Zaporizhzhya National Technical University (ZNTU). since 1999 with periodicity four numbers per year.

The journal is registered by the State Committee for information policy, television and radio broadcasting of Ukraine in 29.01.2003. The journal has a State Registration Certificate of printed mass media (series KB №6904).

ISSN 1607-3274 (print), **ISSN** 2313-688X (on-line).

By the Order of the Ministry of Education and Science of Ukraine from 21.12.2015 № 1328 "On approval of the decision of the Certifying Collegium of the Ministry on the activities of the specialized scientific councils dated 15 December 2015" **journal is included in the list of scientific specialized periodicals of Ukraine,** where the results of dissertations for Doctor of Science and Doctor of Philosophy in Mathematics and Technical Sciences may be published.

The journal publishes scientific articles in English, Russian, and Ukrainian free of charge.

The **article formatting rules** are presented on the site: http://ric.zntu.edu.ua/information/authors.

The journal provides policy of **on-line open (free of charge) access** for full-text publications. The journal allow the authors to hold the copyright without restrictions and to retain publishing rights without restrictions. The journal allow readers to read, download, copy, distribute, print, search, or link to the full texts of its articles. The journal allow reuse and remixing of its content, in accordance with a CC license CC-BY.

Published articles have a unique digital object identifier (DOI).

The journal is abstracted and indexed in leading international and national abstractig journals and scientometric databases, and also placed to the digital archives and libraries with a free on-line access (including DOAJ, DOI, CrossRef, EBSCO, eLibrary.ru / РИНЦ, Google Scholar, Index Copernicus, INSPEC, ISSN, Ulrich's Periodicals Directory, WorldCat, VINITI (All-Russian Institute of scientific and technical information), Djerelo), full list of which is presented on the site: http://ric.zntu.edu.ua/about/editorialPolicies#custom-0.

The journal is distributed:by the Catalogue of Ukrainian periodicals (the catalog number is 22914).

The journal scope: radio physics, micro-, nano- and radio electronics, computer hardware and software, computer networks and telecommunications, algorithm and programming theory, optimization and operations research, machine-machine and man-machine interfacing, mathematical modeling and computer simulation, data and signal processing, control in technical systems, artificial intelligence, including knowledge-based and expert systems, data mining, pattern recognition, artificial neural and neuro-fuzzy networks, fuzzy logics, swarm intelligence and multiagent systems, hybrid systems.

All articles proposed for publication receive an **objective review** that evaluates substantially without regard to race, sex, religion, ethnic origin, nationality, or political philosophy of the author(s).

All articles undergo a two-stage **blind peer review** by the editorial staff and independent reviewers – the leading scientists on the profile of the journal.

EDITORIAL BOARD

Recommended for publication by the Academic Council of ZNTU, protocol № 10 dated 21.03.2016.

The journal is imposed by the editorial-publishing department of ZNTU.

The journal web-site is http://ric.zntu.edu.ua.

The address of the editorial office: Editorial office of the journal «Radio Electronics, Computer Science, Control», Zaporizhzhia National Technical University, Zhukovskiy street, 64, Zaporizhzhya, 69063, Ukraine.

Tel.: +38-061-769-82-96 – the editorial-publishing department.

Fax: +38-061-764-46-62

E-mail: rvv@zntu.edu.ua

ЗМІСТ

CONTENTS

РАДІОФІЗИКА

РАДИОФИЗИКА

RADIOPHYSICS

УДК 004.02:621.391.8:537.86

Borulko V. F.[1], Vovk S. M.[2]

[1]PhD, Senior Scientist, Senior Scientist of department of Applied and Computer Radio Physics, Dnipropetrovs'k National University, Dnipropetrovs'k, Ukraine

[2]PhD, Associate professor, Associate professor of department of Automated Systems of Information Processing, Dnipropetrovs'k National University, Dnipropetrovs'k, Ukraine

MINIMUM-DURATION FILTERING

Myriad filtering and meridian filtering are known as robust methods of signal processing. The theory of these methods is based on the generalized Cauchy distribution and maximum-likelihood criterion. Based on the "Principle of Minimum Duration", we present an alternative approach to justify and generalize the myriad and meridian filtering methods. The proposed approach shows that the myriad and meridian filtering methods are special cases of the minimum-duration filtering methods derived from a concept of "signal quasi-duration". Mathematically, this concept is implemented through the concept of a functional (i.e., a function of a function) by using the proposed set of cost functions. On this foundation, a "superfamily" of quasi-duration functional is built, and a general class of minimum-duration filtering methods which depends on the three free-adjustable parameters is introduced. The numerical simulations are performed to compare the proposed and conventional methods for the problem of filtering a constant signal which is distorted by a mixture of Cauchy, Laplacian and Gaussian noise.

Keywords: myriad filtering, meridian filtering, duration.

NOMENCLATURE

PMD Principle of Minimum Duration;
ML maximum-likelihood;
RMSE Root Mean Square Error.
A unknown signal amplitude, or filter output;
b constant in the equalizing procedure;
D functional of strict duration, or "strict duration";
$D(A)$ one-variable objective function corresponded to D;
$D^{(\alpha,\beta,...)}$ functional of quasi-duration, or "quasi-duration" ;
$D^{(\alpha,\beta,...)}(A)$ one-variable objective function corresponded to $D^{(\alpha,\beta,...)}$;
$D_S^{(\alpha,\beta,q)}$ quasi-duration based on the superset of cost functions;
$D_M^{(\alpha,q)}$ quasi-duration based on the generalized Meshalkin cost function;
$f(t)$ shape of the observed signal;
$g(t)$ observed signal;
g_i ith sample of the observed signal;
N number of signal samples, or filter window size;
p tail constant of the generalized Cauchy distribution;
$p(z)$ probability density function;
q free-adjustable parameter called "smoothing degree";
$s(t)$ arbitrary signal;
T time interval;
t time argument;
x argument of cost function;
y intermediate variable in the equalizing procedure;
y_k kth approximation to y;
α – free-adjustable parameter associated with the scale parameter;
β – free-adjustable parameter associated with the shape of data distribution;
σ – scale parameter, or standard deviation of noise;
$\chi(x)$ – ideal cost function;
$\psi(x)$ – arbitrary cost function;
$\psi^{(\alpha,\beta,...)}(x)$ – "quasi-duration" cost function;
$\psi_R^{(\beta)}(x)$ – "root cost function";
$\psi_R^{(\alpha,\beta,q)}(x)$ – "root cost function with smoothing", or "q-smoothed root cost function";
$\psi_{\log}^{(\alpha,q)}(x)$ – "q-smoothed logarithmic cost function";
$\psi_{med}^{(\alpha,q)}(x)$ – "q-smoothed median cost function";
$\psi_D^{(\alpha,\beta,q)}(x)$ – "generalized Demidenko cost function";
$\psi_M^{(\alpha,q)}(x)$ – "generalized Meshalkin cost function";
$\psi_S^{(\alpha,\beta,q)}(x)$ – member of the "superset of cost function";

INTRODUCTION

The principle of maximum likelihood is a mathematical foundation of many filtering methods. To use this principle, it is necessary to specify the joint probability density

DOI 10.15588/1607-3274-2016-1-1

function which should be maximized over all desired parameters. In the case of independent and identically distributed data samples, the mathematical expression for the joint probability density function is considerably simplified with reducing the computational complexity and filter structure.

Based on the maximum-likelihood principle, the myriad filtering [1–2] and the meridian filtering [3–4] have been introduced as robust methods [5–6] of signal processing in impulsive environments. These methods are based on the assumptions that the signal samples are independently Cauchy-distributed and meridian-distributed, respectively. In spite of the common features of these methods, each of them presents an individual class of nonlinear filtering methods with the one free-tunable parameter associated with the scale parameter of noise distribution. Later, a more general class has been proposed [7], where the filtering methods depend on the two free-adjustable parameters associated with the scale parameter and the tail-constant of the generalized Cauchy distribution.

In this paper, we present a larger class of filtering methods. This class depends on the three free parameters that need to be adjusted, where the first two parameters coincide with the two parameters mentioned above, and the third parameter is associated with the shape of data distribution. In contrast to [7], this class is based on the "Principle of Minimum Duration" (PMD) [8–10]. Therefore, it will be referred to as the class of "minimum-duration filtering."

At the beginning of this paper, we describe the problem statement based on the PMD. In general, the PMD states a non-energy criterion, when the signal processing should produce a signal with a minimum duration. In this paper, we restrict the study to the approximation problem with the one amplitude parameter and show that the minimum-duration filtering is derived from the concept of "signal quasi-duration" by the PMD. Mathematically, this concept is based on the concept of a functional (i.e., a function of a function), where the cost function of the function, which describes the signal, is used. We construct a new set of cost functions and build a "superfamily" of quasi-duration. On this basis, for the discrete case we introduce a general class of the minimum-duration filtering methods. At the end of this paper, the performance of the minimum-duration methods is numerically compared to that of the conventional ones for the problem of filtering a constant signal which is distorted by a mixture of Cauchy, Laplacian and Gaussian noise.

1 PROBLEM STATEMENT

The original problem statement is to build the filtering methods by using the PMD. The mathematical problem statement requires a formalization of the concept of "signal duration". In this paper, we use the two concepts, namely "strict duration" and "quasi-duration" [8].

The strict duration, D, of any signal, $s(t)$, is defined as a measure of the nonzero signal values. Mathematically, D is the functional

$$D \equiv D[s(t)] = \int_{-\infty}^{\infty} \chi[s(t)]dt\,, \qquad (1)$$

where $\chi(x) = \begin{cases} 1, & if\ x \neq 0 \\ 0, & if\ x = 0 \end{cases}$ plays the role of an ideal cost function which we call the "Titchmarsh cost function" (p. 319 in [11]). Since D is the functional, it will be also referred to as the "functional of strict duration." Despite an obvious physical interpretation, the strict duration cannot be constructively applied to the problem formulation [8].

The concept of quasi-duration enables to formulate the problems constructively, although it has a less obvious physical interpretation. The quasi-duration, $D^{(\alpha,\beta,...)}$, is the functional

$$D^{(\alpha,\beta,...)} \equiv D^{(\alpha,\beta,...)}[s(t)] = \int_{-\infty}^{\infty} \psi^{(\alpha,\beta,...)}[s(t)]dt, \quad (2)$$

where $\alpha,\beta,...$ are the real free-adjustable parameters, $\psi^{(\alpha,\beta,...)}(x)$ is the continuous cost function with the property: $\psi^{(\alpha,\beta,...)}(x) \to \chi(x)$ as parameters $\alpha,\beta,...$ go to their boundary values. This property provides the limiting process from $D^{(\alpha,\beta,...)}$ to D, that makes sense in the noiseless case. Since $D^{(\alpha,\beta,...)}$ is the functional, it will be also referred to as the "quasi-duration functional."

Phenomenologically, the PMD is formulated as: "After processing, the signal duration should be minimal." An applicability domain of the PMD depends on what concept is used. If the concept of the strict duration is used, the PMD can be applied to time-limited and noise-free signal. However, if the concept of the quasi-duration is used, the applicability domain of PMD is significantly extended, since it becomes possible to process both the noisy signal and the almost time-limited signal (e.g., the Gaussian pulse). On the other hand, there are two situations when using and implementing the PMD. In the first one, the signal is considered as a solution, which is either a sum of the known and unknown signal components or an unknown signal of a given class. In this situation, the PMD is implemented either by variation of the unknown signal component or by variation of the signal itself. In the second one, the signal is considered as an error (or residual) signal. Provided that the approximation problem is formulated for a finite number of unknown parameters, the PMD is implemented by variation of these parameters [10].

Further the approximation problem with the one unknown linear parameter which is the signal amplitude is considered. Let $g(t)$ be the signal observed in the time interval T, and $f(t)$ be the shape of this signal. Let A be the unknown signal amplitude. Then the error signal is

$$s(t) = g(t) - Af(t); \quad t \in T\,. \qquad (3)$$

Applying the PMD to (1) with (3) leads to the following problem

$$\arg\min_A D(A) = \arg\min_A \int_{-\infty}^{\infty} \chi[g(t) - Af(t)]dt\,, \qquad (4)$$

where D becomes a one-variable objective function, $D(A)$, to be minimized in A. The advantage of (4) is that the

complete destruction of a large part of $g(t)$ will not change the solution. The disadvantage of (4) is that the negligible noise makes it impossible to solve this problem.

Applying the PMD to (2) with (3) yields

$$\arg\min_{A} D^{(\alpha,\beta,\ldots)}(A) = \arg\min_{A} \int_{-\infty}^{\infty} \psi^{(\alpha,\beta,\ldots)}[g(t) - Af(t)]dt\,, \quad (5)$$

where $D^{(\alpha,\beta,\ldots)}$ becomes the one-variable objective function, $D^{(\alpha,\beta,\ldots)}(A)$, to be minimized in A. The problem (5) defines the minimum-duration estimator of A based on the concept of signal quasi-duration. Provided that $f(t) = \text{const}$, from (5) the mathematical statement of the filtering problem for the constant signal is readily obtained. The advantage of (5) is that by selecting a cost function, which should give the Titchmarsh cost function in the limit, one can eliminate the drawback of (4) and naturally satisfy the two additional and important requirements, namely: 1) robust behavior for large values of the argument when the contribution of the small number of large values is limited; 2) smooth behavior for small values of the argument to perform the optimum processing of a large number of small values. The first one determines the filter behavior with respect to the impulses (outliers), and the second one determines the filter behavior with respect to the noise (inliers). Therefore, the mathematical problem is reduced to finding such a cost function, which satisfies the requirements mentioned above and generalizes the some known filtering methods (in particular, the average, median, myriad, and meridian filtering). The way to solve this problem is to introduce a finite set of free-adjustable parameters which control the cost function behavior, and, consequently, which control the filtering result in a given noise environment.

2 REVIEW OF THE LITERATURE

The cost functions technique is the theoretical basis to build the methods of the optimal data processing and, in particular, of the optimal filtering. For example, using a quadratic cost function $\psi(x) = x^2$ leads to the method of average filtering, which is the optimal in the case of Gaussian noise. The use of an absolute cost function $\psi(x) = |x|$ leads to the median filtering, which is the optimal method in the case of Laplacian noise. An important feature of these methods is that they do not require any adjustment to the noise parameters. However, this is not the case for the other methods. For example, the methods of myriad filtering and meridian filtering are obtained by using the cost functions $\psi(x) = \log(x^2 + \alpha^2)$ and $\psi(x) = \log(|x| + \alpha);\ \alpha > 0$, respectively, which provide the optimal filtering in the cases of Cauchy noise and "Meridian" noise [1–4]. Unlike previous methods, these methods depend on the noise scale parameter σ and require the optimal tuning of the free parameter as $\alpha = \sigma$. Similarly, the use of the cost function $\psi(x) = \log(|x|^p + \alpha^p);\ \alpha > 0;\ 0 < p \le 2$ results in the construction of the filtering method in the case when the noise has a generalized Cauchy distribution. Here, there are two adjustable parameters, α and p [7]. Unfortunately, the cost functions mentioned above do not provide the limiting process to $\chi(x)$. Therefore, they can not be used as a function $\psi^{(\alpha,\beta,\ldots)}(x)$. However, they can be derived as special (limiting) cases for the given noise environment.

The limiting process from the quasi-duration functional to the functional of strict duration is the theoretical foundation of the proposed approach. This requirement noticeably limits the set of cost functions that can be used in (5). In [8, 10, 13] the following sets were considered: 1) one-parameter set of the "root cost functions"; 2) two two-parameter sets of the root cost functions with the quadratic and absolute smoothing; 3) two three-parameter sets of the "q-smoothed root cost functions" and "generalized Demidenko cost functions".

The use of the one-parameter set of the "root cost functions" leads to the following. Let $\psi^{(\alpha,\beta,\ldots)}(x)$ be the cost function

$$\psi_R^{(\beta)}(x) = |x/\Lambda|^{\beta};\quad 0 < \beta < 1\,, \quad (6)$$

where Λ is the constant with the physical dimension of x; further $\Lambda = 1$. Since $\lim_{\beta\to+0} \psi_R^{(\beta)}(x) = \chi(x)$, the function (6) may be used in (2). In addition, the following properties hold: (*i*) $\psi_R^{(\beta)}(x)$ is neither convex nor concave for all real x; (*ii*) $\psi_R^{(\beta)}(x)$ has even symmetry, where $\psi_R^{(\beta)}(0) = 0$ and $\psi_R^{(\beta)}(\pm 1) = 1$; (*iii*) $\lim_{x\to-0}[d\psi_R^{(\beta)}(x)/dx] = -\infty$ and $\lim_{x\to+0}[d\psi_R^{(\beta)}(x)/dx] = \infty$. Due to its typical behavior, the function $\psi_R^{(\beta)}(x)$ was named as the "root cost function" (even when β is an irrational number) [10]. Let the quasi-duration functional based on (6) be denoted as $D^{(\beta)}$. Then substituting (6) into (5) yields [10]

$$\arg\min_{A} D^{(\beta)}(A) = \arg\min_{A} \int_{T} |g(t) - Af(t)|^{\beta}\, dt;\ 0 < \beta < 1\,, \quad (7)$$

where $D^{(\beta)}(A)$ is the objective function to be minimized in A. The advantage of (7) is that the solution of (7) makes sense in the cases when there is the complete destruction of the large part of data and when the data are distorted by noise. The disadvantages of (7) are the following. First, the noise appearance can lead to the bias of estimator. Second, the noise nature is left out of account. The latter means that the cost function which depends on more free parameters and which takes into account the noise nature can be more efficient than (6).

The two-parameter sets of the root cost functions with the quadratic and absolute smoothing are defined by the similar equations that by introducing a smoothing degree, q, have been summarized as [10]

$$\psi_R^{(\alpha,\beta,q)}(x) = k_R^{(\alpha,\beta,q)}[(1 + |x|^q/\alpha^q)^{\beta/q} - 1]\,, \quad (8)$$

where $\alpha > 0;\ 0 < \beta \le 1;\ 0 < q < \infty;\ \beta < q$; $k_R^{(\alpha,\beta,q)} = 1/[(1 + |x_1|^q/\alpha^q)^{\beta/q} - 1]$ and $\psi_R^{(\alpha,\beta,q)}(x_1) = 1$. This function represents a three-parameter set of the

"q-smoothed root cost functions". Passing to the limit in (8) as $\beta \to +0$ yields the "q-smoothed logarithmic cost function"

$$\psi_{\log}^{(\alpha,q)}(x) = k_{\log}^{(\alpha,q)} \ln(1+|x|^q/\alpha^q), \qquad (9)$$

where $\alpha > 0$; $0 < q < \infty$; $k_{\log}^{(\alpha,q)} = 1/\ln(1+|x_1|^q/\alpha^q)$. This function is the generalization of the myriad and meridian cost functions, which are for $q = 2$ and $q = 1$, respectively. It is easy to see that in the case of generalized Cauchy distribution the "smoothing degree" parameter q coincides with the tail constant p of this distribution.

Contrariwise, passing to the limit in (8) as $\beta \to 1$ yields

$$\psi_{med}^{(\alpha,q)}(x) = k_{med}^{(\alpha,q)}[(1+|x|^q/\alpha^q)^{1/q} - 1]; \quad 0 < q < \infty, \qquad (10)$$

where $k_{med}^{(\alpha,q)} = [(1+|x_1|^q/\alpha^q)^{1/q} - 1]$; $x_1 \neq 0$ is the normalization point, where $\psi_{med}^{(\alpha,q)}(x_1) = 1$. Since the $\psi_{med}^{(\alpha,q)}(x)$ approaches the absolute cost function for $|x|^q/\alpha^q >> 1$, further it will be referred to as the "q-smoothed median cost function" (despite the fact that for $q > 1$ and for $q < 1$ the behavior of (10) is different near the zero value of x). If $q = 2$, the special case of (10) is a pseudo-Huber cost function [12]. But if $q = 1$, from (10) one obtains the absolute cost function.

The analysis of (8) shows that under fixed q the limiting process from $\psi_R^{(\alpha,\beta,q)}(x)$ to $\chi(x)$ is performed by α and β (when their boundary values are zero) just in the one direction (first by α and then by β). This drawback has been eliminated in [13] by generalizing the cost functions proposed by E. Z. Demidenko [14] and L. D. Meshalkin [15].

The function $\psi_D(x) = \dfrac{x^2}{x^2+\alpha^2}$ was suggested by E. Z. Demidenko for the regression problem [14]. By introducing b and q parameters, it can be generalized to [13]

$$\psi_D^{(\alpha,\beta,q)}(x) = k_D^{(\alpha,\beta,q)}\left[1 - \left(1 + \frac{|x|^q}{\alpha^q}\right)^{-\beta/q}\right], \qquad (11)$$

where $\alpha > 0$, $0 < \beta < \infty$, $0 < q < \infty$, and $k_D^{(\alpha,\beta,q)} = \dfrac{(1+|\alpha/x_1|^q)^{\beta/q}}{(1+|\alpha/x_1|^q)^{\beta/q} - (|\alpha/x_1|^q)^{\beta/q}}$ is defined by $\psi_D^{(\alpha,\beta,q)}(x_1) = 1$. It is easy to check, that the Titchmarsh cost function $\chi(x)$ is the limit of (11) as $\beta \to \infty$, and the q-smoothed logarithmic cost function $\psi_{\log}^{(\alpha,q)}(x)$ is the limit of (11) as $\beta \to +0$.

The function $\psi_M(x) = 1 - \exp(-x^2/\alpha^2)$ was suggested by L. D. Meshalkin for robust estimation [15], [16]. It can be generalized to [13]

$$\psi_M^{(\alpha,q)}(x) = k_M^{(\alpha,q)}\left[1 - \exp\left(-\frac{|x|^q}{\alpha^q}\right)\right], \qquad (12)$$

where $\alpha > 0$, $0 < q < \infty$ and $k_M^{(\alpha,q)} = \left[1 - \exp(-|x_1|^q/\alpha^q)\right]^{-1}$ is defined by $\psi_M^{(\alpha,q)}(x_1) = 1$. It is obvious that as $\alpha \to 0$ the function $\psi_M^{(\alpha,q)}(x)$ tends to $\chi(x)$. In addition, for fixed $\alpha > 0$ it approaches the $\chi(x)$ as $q \to 0$, and it has a form of the "rectangular hole" as $q \to \infty$ [13].

3 MATERIALS AND METHODS

It is seen that (11) is a continuation of (8) on negative values of β. Hence, there exists a continuous (by α, β and q) set of cost functions which is defined by the common member

$$\psi_S^{(\alpha,\beta,q)}(x) = k_S^{(\alpha,\beta,q)}\left[\left(1 + \frac{|x|^q}{\alpha^q}\right)^{\beta/q} - 1\right], \qquad (13)$$

where $k_S^{(\alpha,\beta,q)} = 1/[(1+|x_1|^q/\alpha^q)^{\beta/q} - 1]$, $-\infty < \beta \le 1$, $\alpha > 0$, $0 < q < \infty$, $\beta < q$. This set incorporates the following cost functions: 1) q-smoothed median cost functions for $\beta = 1$ (in particular, the pseudo-Huber cost function); 2) q-smoothed root cost functions for $0 < \beta < 1$; 3) q-smoothed logarithmic cost functions for $\beta = 0$ (in particular, the meridian and myriad cost functions); 4) generalized Demidenko cost functions for $-\infty < \beta < 0$; 5) Titchmarsh cost function for $\beta \to -\infty$. Since this set is sufficiently representative, it will be referred to as a "superset" of cost functions. Below, the Meshalkin cost function is also included in this superset.

With regard to the small values of x, the contribution of α is different for each cost function derived from (13). To eliminate this shortcoming, we have proposed to equalize the second-order derivative of the cost functions at zero [13]. Further, the method for producing a modified superset of cost functions is presented for $q = 2$. The similar technique may be obtained for any $0 < q < \infty$.

Let α_{myr}^2 be the fixed value of α^2 for the myriad cost function. Let the equality: $d^2\psi_S^{(\alpha,\beta,2)}(x)/dx^2 = d^2\psi_{\log}^{(\alpha,2)}(x)/dx^2$ be hold at $x = 0$. Then

$$\frac{(\beta/2)}{y\left[\left(1 + \dfrac{|x_1|^2}{y}\right)^{\beta/2} - 1\right]} = \frac{1}{b}, \qquad (14)$$

where $y = \alpha_S^2$ denotes α^2 for the function $\psi_S^{(\alpha,\beta,2)}(x)$ and $b = \alpha_{myr}^2 \ln(1+|x_1|^2/\alpha_{myr}^2)$. Equation (14) states the equalizing problem, where y is to be determined, and can be solved by using the Newton's method for finding roots

$$y_{k+1} = y_k - \frac{u(y_k)}{u'(y_k)}, \qquad (15)$$

where y_k is the kth approximation to y, and where $u(y_k)$ and $u'(y_k)$ are the reduced function and its first derivative at y_k, respectively. The convergence of (15) is ensured by the following approach. For $-\infty < \beta < 0$, the function $u(y) = y\left[\left(1+\frac{|x_1|^2}{y}\right)^{\beta/2} - 1\right] - \frac{\beta}{2}b = 0$ should be used with its first derivative $u'(y) = \left(1+\frac{|x_1|^2}{y}\right)^{\beta/2}\left[1-\frac{(\beta/2)|x_1|^2}{y+|x_1|^2}\right] - 1$ and with the initial guess $y_0 = \alpha_{myr}^2$. For $0 < \beta < 1$, the function $u(y) = \left[\left(1+y|x_1|^2\right)^{\beta/2} - 1\right]/y - \frac{\beta}{2}b = 0$ should be used with its first derivative $u'(y) = \frac{(\beta/2)(1+|x_1|^2 y)^{\beta/2-1}|x_1|^2 y - (1+|x_1|^2 y)^{\beta/2} + 1}{y^2}$ and with the initial guess $y_0 = 1/\alpha_{myr}^2$. The equalizing procedure based on (15) typically converges to machine precision within 3–5 iterations.

Fig. 1 represents the superset and the modified superset of cost functions for $q = 2$ and $x_1 = 1$. Fig. 1a shows the superset without the equalizing procedure, when all cost functions have the same value $\alpha^2 = 0.01$. In this figure, there are depicted the pseudo-Huber (curve 1), q-smoothed root with $\beta = 1/2$ (curve 2), myriad (curve 3) and generalized Demidenko (the curves 4–7 are for $\beta = -1; -2; -5; -100$ in (13), respectively) cost functions. It is seen that the sequence of these cost functions tends to $\chi(x)$ as $\beta \to -\infty$. Fig. 1b shows the modified superset of cost functions, which is produced by the equalizing procedure for $\alpha_{myr}^2 = 0.01$. Here, the curve 7 is similar to the graph of the Meshalkin cost function, coinciding visually with it. Thus, in the limit as $\beta \to -\infty$, the Meshalkin cost function is obtained. This fact can readily be proved, since the finiteness of the second-order derivative at zero is provided on condition that $\alpha_S^2 = const * \beta$. Substituting this value into (13) with $q = 2$ and passing to the limit in (13) as $\beta \to -\infty$ yield the Meshalkin cost function.

Based on (13), the quasi-duration functional can be expressed as

$$D_S^{(\alpha,\beta,q)} \equiv D_S^{(\alpha,\beta,q)}[s(t)] = \int_{-\infty}^{\infty} \psi_S^{(\alpha,\beta,q)}[s(t)]dt = k_S^{(\alpha,\beta,q)} \times$$

$$\times \int_{-\infty}^{\infty}\left[\left(1+\frac{|s(t)|^q}{\alpha^q}\right)^{\beta/q} - 1\right]dt, \qquad (16)$$

where $\alpha > 0$, $-\infty < \beta \le 1$, $0 < q < \infty$, and $\beta < q$. It has the following special cases: 1) q-smoothed median functional

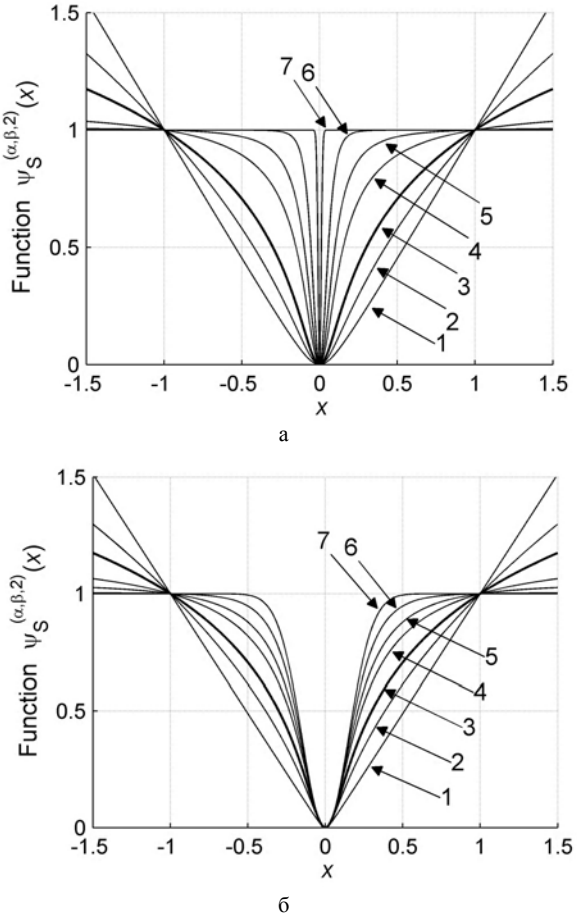


Figure 1 – Superset of cost functions (a) without and (b) with the equalizing procedure

(when $\beta = 1$); 2) pseudo-Huber functional (when $\beta = 1$ and $q = 2$); 3) q-smoothed root functional (when $0 < \beta < 1$); 4) q-smoothed logarithmic functional (when $\beta \to 0$); 5) myriad functional (when $\beta \to 0$ and $q = 2$); 6) meridian functional (when $\beta \to 0$ and $q = 1$); 7) generalized Demidenko functional (when $-\infty < \beta < 0$); 8) Demidenko functional (when $\beta = -2$ and $q = 2$). The generalized Meshalkin functional defined by

$$D_M^{(\alpha,q)}[s(t)] = k_M^{(\alpha,q)} \int_{-\infty}^{\infty}\left[1 - \exp\left(-\frac{|s(t)|^q}{\alpha^q}\right)\right]dt \quad (17)$$

is also derived from (16) in the limit as $\beta \to -\infty$ after the equalizing procedure. Thus, (16) determines a large family of functionals, which will be referred to as the "superfamily of quasi-duration". On this basis, the minimum-duration estimate is defined as a solution of the appropriate optimization problem related to minimizing the quasi-duration (16). Assuming $f(t) = const$, below we write the general class of the minimum-duration filtering methods for the discrete case with the following notations: g_i is the sample of observed signal, N is the number of signal samples in filter window, and A denotes the unknown amplitude value, which is the filter output. This class is given by the problem

$$\arg\min_{A}\left\{k_S^{(\alpha,\beta,q)}\sum_{i=1}^{N}\left[\left(1+\frac{|g_i-A|^q}{\alpha^q}\right)^{\beta/q}-1\right]\right\}, \quad (18)$$

where $\alpha>0$, $-\infty<\beta\le 1$, $0<q<\infty$, and $\beta<q$. Note, that $k_S^{(\alpha,\beta,q)}>0$ for $0<\beta\le 1$ and $k_S^{(\alpha,\beta,q)}<0$ for $-\infty<\beta<0$. The filtering consists in finding the optimal value of A. The special cases of (18) can be written more simply without the nonessential constants.

The computational complexity of the optimization problem (18) can be reduced [17], when instead of the optimal value, $A\in R$, where R is the set of real numbers, the quasi-optimal value, $A\in\{g_i\,|\,i=1,...,N\}\subset R$, is computed as the filter output. It is seen that the quasi-optimal value of A makes, at least, one of the terms in the sum (18) equal to zero.

4 EXPERIMENTS AND RESULTS

We have compared the potential capability of the minimum-duration methods in the problem of filtering the constant signal distorted by additive noise. Further, we have selected the q-smoothed root filtering ((18) with $q=2$ and $\beta=1/2$), Demidenko filtering ((18) with $q=2$ and $\beta=-2$), and Meshalkin filtering ((18) with $q=2$ and $\beta\to\infty$). These methods have been compared with the methods of average, median, pseudo-Huber, myriad and maximum-likelihood (ML) filtering. The three cases, when the constant signal with amplitude $A=1$ was additively distorted by the Cauchy noise (case 1), by the mixture of Cauchy and Laplacian noise (case 2), and by the mixture of Cauchy and Gaussian noise (case 3), have been examined with the same value of the scale parameter for each type of noise. The mixture of Cauchy and Laplacian noise was formed with a priori probabilities of 1/2 for each. The mixture of Cauchy and Gaussian noise was formed with a priori probabilities of 2/3 and 1/3, respectively. Thus, in these cases, the probability density function, $p(z);z\in R$, of noisy signal is defined by $p(z)=p_C(z)=\frac{1}{\pi}\frac{1}{(z-A)^2+\sigma^2}$ (in the case 1), $p(z)=\frac{1}{2}p_C(z)+\frac{1}{2}p_L(z)$; $p_L(z)=\frac{1}{2\lambda}\exp\left(-\frac{|z-A|}{\lambda}\right)$; $\lambda=\frac{\sigma}{\sqrt{2}}$ (in the case 2) and $p(z)=\frac{2}{3}p_C(z)+\frac{1}{3}p_G(z)$; $p_G(z)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-(z-A)^2/2\sigma^2\right]$ (in the case 3).

For the calculations, the fragment of noisy constant signal was numerically simulated on a set of N=121 discrete samples, where N is the length of filter window. The filtering was to estimate the signal amplitude by finding the location of the global minimum of the corresponding optimization problem with the relative accuracy 0,1%. Further, the estimator's Root Mean Square Error (RMSE) averaged arithmetically over 10000 noise realizations have been calculated as a function of the free-adjustable parameter α that associated with the scale parameter of a given noise distribution. In these calculations, the modified superset of cost functions was used for $q=2$.

Fig. 2 represents the calculated RMSE/σ ratio vs the α/σ ratio, where $\sigma=0{,}1$. The solid curves with heavy dots depict the calculated values for the (1) pseudo-Huber filtering, (2) q-smoothed root filtering with q=2 and $\beta=1/2$, (3) myriad filtering, (4) Demidenko filtering, and (5) Meshalkin filtering; the dotted curve (6) with heavy dots depicts the calculated values for ML filtering; the dotted horizontal line depicts the calculated value for the median filtering.

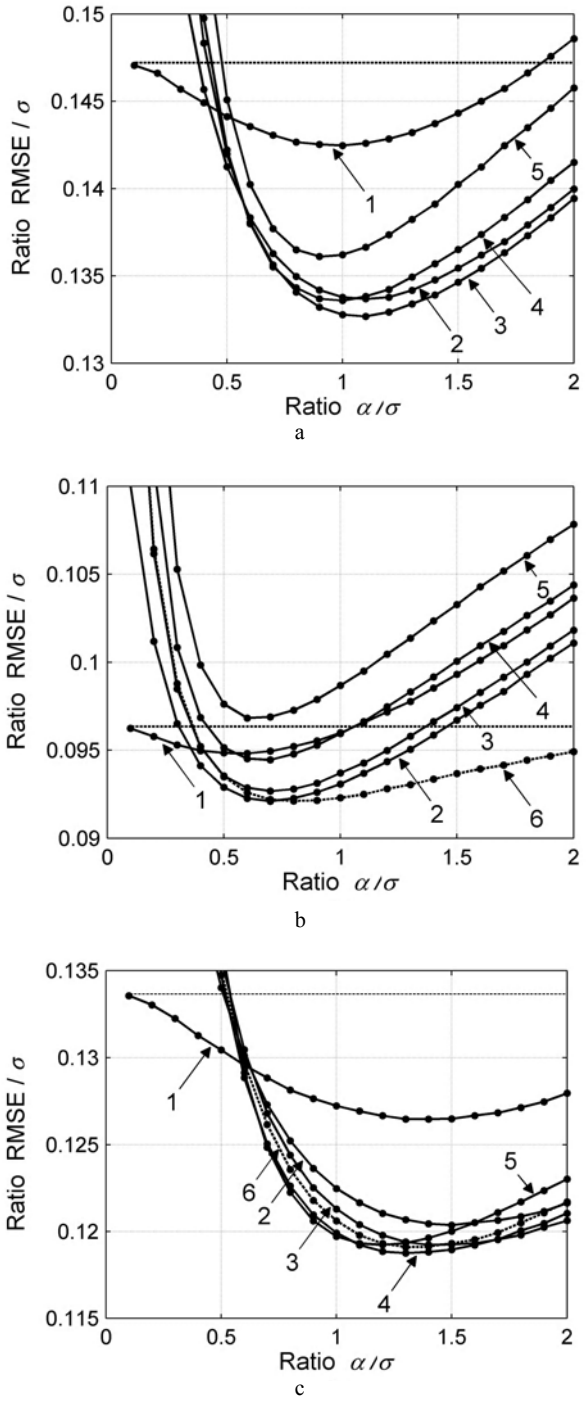


Figure 2 – RMSE/σ ratio vs α/σ ratio for the constant signal distorted by the (a) Cauchy noise, (b) mixture of Cauchy and Laplacian noise, and (c) mixture of Cauchy and Gaussian noise

5 DISCUSSION

As shown in Fig. 2a, in the case 1 the myriad filtering, which is also the ML filtering, provides the best result at $\alpha/\sigma = 1{,}1$ (theoretically, $\alpha/\sigma = 1$). Moreover, it provides the best results within the range $0{,}6 \le \alpha/\sigma \le 2$. However, if $\alpha/\sigma \le 0{,}4$ and α goes to zero, the performance of the myriad filtering (as well as the other minimum-duration filtering with the exception of the pseudo-Huber filtering) dramatically deteriorates; whereas that of the median filtering remains steady throughout. Note also, if the α/σ ratio becomes larger than 2, the RMSE/σ ratio becomes the larger for all these methods (except for the median filtering) and tends to the value which is obtained by the average filtering. In the case 2, Fig. 2b shows that the q-smoothed root filtering (with $q = 2$ and $\beta = 1/2$) provides the best result within the range $0.4 \le \alpha/\sigma \le 0.7$ and achieves the performance of the ML filtering at $\alpha/\sigma = 0.7$. However, for $\alpha/\sigma \ge 0.8$ the ML filtering, which tends to the median filtering as α goes to infinity, is the best. In the case 3, Fig. 2c shows the advantage of the Demidenko filtering, although this advantage is not significant here. This figure shows that within the range $0.6 \le \alpha/\sigma \le 2$ the Demidenko filtering is slightly better than the ML filtering. This can be explained by the fact that the ML estimator has no optimum properties for finite samples. In addition, for all these three cases, the selected minimum-duration filtering methods, used with $q = 2$ (which was close to the optimal value of q) and optimal α, were 5–8 times better than the average filtering. However, the smaller N was, the smaller the advantage.

Thus, these numerical simulations show the following. For the problem of filtering the noisy constant signal, the potential of the minimum-duration filtering exceeds the potential of the median and average filtering. As would be expected, the myriad filtering is the best for the Cauchy noise, the q-smoothed root filtering (with $q = 2$ and $\beta = 1/2$) is the best for the given mixture of Cauchy and Laplacian noise, and the Demidenko filtering is the best for the given mixture of Cauchy and Gaussian noise.

CONCLUSIONS

The goal of signal processing based on the PMD is to produce the signal with the minimum duration. To describe the signal duration in practice, the concept of the signal quasi-duration can be used. This concept is implemented by the quasi-duration functional and, in particular, by the quasi-duration objective function. The superfamily of the quasi-duration functional is proposed. It covers the families that include the q-smoothed median functional, q-smoothed root functional, q-smoothed logarithmic functional, the generalized Demidenko and Meshalkin functionals.

The general class of the minimum-duration filtering methods which depends on the three free-adjustable parameters is introduced. The myriad and meridian filtering methods occupy the intermediate positions in this class.

The potential of the minimum-duration filtering exceeds the potential of the median and average filtering. Theoretically, the minimum-duration filtering methods enable to filter the signal, which is destroyed in more than half length of the filter window, when the median filtering may fail.

By adjusting the free parameters, the proposed approach enables efficient processing of the signal which is distorted by noise of different types. Finding optimal values of α, β and q is the major problem in taking full advantage of the minimum-duration filtering.

REFERENCES

1. Gonzalez J. G. Optimality of the myriad filter in practical impulsive-noise environments / J. G. Gonzalez, G. R. Arce // IEEE Trans. on Signal Processing. – 2001. – Vol. 49, No. 2. – P. 438–441.
2. Nunez R. C. Fast and accurate computation of the myriad filter via branch-and-bound search / R. C. Nunez, J. G. Gonzalez, G. R. Arce // IEEE Trans. Signal Processing. – 2008. – Vol. 56, No. 7. – P. 3340–3346.
3. Aysal T. C. Meridian filtering for robust signal processing / T. C. Aysal, K. E. Barner // IEEE Trans. on Signal Processing. – 2007. – Vol. 55, No. 8. – P. 3949–3962.
4. Analysis of meridian estimator performance for non-Gaussian PDF data samples / [D. A. Kurkin, V. V. Lukin, A. A. Roenko, I. Djurovic] // Telecommunications and Radio Engineering. – 2010. – Vol. 69, No. 8. – P. 669–679.
5. Huber P. J. Robust estimation of a location parameter / P. J. Huber // Annals of Mathematical Statistics. – 1964. – Vol. 35, No. 1. – P. 73–101.
6. Huber P. J. Robust statistics / P. J. Huber. – New York : John Wiley and Sons, 1981. – 312 p.
7. Carrillo R. E. Generalized Cauchy distribution based robust estimation / R. E. Carrillo, T. C. Aysal, K. E. Barner // Acoustic, Speech and Signal Processing : IEEE International Conference ICASSP 2008, Las Vegas, 31 March – 4 April 2008 : proceedings. – IEEE, 2008. – P. 3389–3392. DOI: 10.1109/ICASSP.2008.4518378
8. Vovk S. M. A minimum-duration method for recovering finite signals / S. M. Vovk, V. F. Borul'ko // Radioelectronics and Communications Systems. – 1991. – Vol. 34. – P. 67–69.
9. Vovk S. M. Elimination of the measurement background by the minimum duration method / S. M. Vovk, V. F. Borulko // Radioelectronics and Communications Systems. – 1998. – Vol. 41. – P. 48–49.
10. Vovk S. M. Statement of a problem of definition of linear signals parameters in quasinormed space / S. M. Vovk, V. F. Borul'ko // Radioelectronics and Communications Systems. – 2010. – Vol. 53. – P. 367–375.
11. Titchmarsh E. C. The theory of functions / E. C. Titchmarsh. – New York : Oxford University Press, 1939. – 454 p.
12. Hartley R. Multiple View Geometry in Computer Vision / R. Hartley, A. Zisserman – Cambridge University Press, 2004. – 645 p.
13. Vovk S. Family of generalized Demidenko functionals for robust estimation / S. Vovk, V. Borulko // Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory: 17th International Seminar/Workshop DIPED'12, Tbilisi, 24–27 September 2012 : proceedings. – IEEE, 2012. – P. 151–154.
14. Демиденко Е. З. Оптимизация и регрессия / Е. З. Демиденко. – М. : Наука, 1989. – 296 с.
15. Meshalkin L. D. Some mathematical methods for the study of non-communicable diseases / L. D. Meshalkin // Uses of epidemiology in planning of health services. Proc. of the Sixth International Scientific Meeting, Primosten, August 29–September 3, 1971. – Belgrade. – 1973. – P. 250–2566.
16. Shevlyakov G. L. Robustness in data analysis: criteria and methods / G. L. Shevlyakov, N. O. Vilchevski. – Utrecht : VSP, 2002. – 310 p.
17. Vovk S. Dual Method of Minimum Spatial Extent / S. Vovk, V. Borulko // Mathematical Methods in Electromagnetic Theory: 14th International Conference MMET-2014, Dnipropetrovsk, Ukraine, 26–28 August 2014 : proceedings. – IEEE, 2014. – P. 144–147. DOI: 10.1109/MMET.2014.6928715.

Article was submitted 10.11.2015.
After revision 02.12.2015.

Борулько В. Ф.[1], Вовк С. М.[2]

[1]Канд. физ.-мат. наук, с.н.с., с.н.с. кафедры прикладной и компьютерной радиофизики Днепропетровского национального университета, Днепропетровск, Украина

[2]Канд. физ.-мат. наук, доцент, доцент кафедры автоматизированных систем обработки информации Днепропетровского национального университета, Днепропетровск, Украина

ФИЛЬТРАЦИЯ СИГНАЛОВ НА ОСНОВЕ ПРИНЦИПА МИНИМУМА ДЛИТЕЛЬНОСТИ

Методы мириадной и меридианной фильтрации известны как робастные методы обработки сигналов. Теория этих методов основана на обобщенном распределении Коши и критерии максимального правдоподобия. Мы представляем альтернативный подход, основанный на принципе минимума длительности, к обоснованию и обобщению методов мириадной и меридианной фильтрации. Предлагаемый подход показывает, что мириадная и меридианная фильтрация являются частными случаями методов фильтрации, которые выводятся из концепции «квазидлительность сигнала». Математически эта концепция реализуется через понятие функционала с помощью предложенного множества стоимостных функций. На этом фундаменте построено «суперсемейство» функционала квазидлительности и введен общий класс методов фильтрации, который зависит от трех свободно-настраиваемых параметров. Приводятся результаты числового моделирования для сравнения эффективности предложенных и стандартных методов для задачи фильтрации постоянного сигнала, который искажен смесью шумов Коши, Лапласа и Гаусса.

Ключевые слова: мириадная фильтрация, меридианная фильтрация, длительность.

Борулько В. Ф.[1], Вовк С. М.[2]

[1]Канд. фіз.-мат. наук, с.н.с., с.н.с. кафедри прикладної та комп'ютерної радіофізики Дніпропетровського національного університету, Дніпропетровськ, Україна

[2]Канд. фіз.-мат. наук, доцент, доцент кафедри автоматизованих систем обробки інформації Дніпропетровського національного університету, Дніпропетровськ, Україна

ФІЛЬТРАЦІЯ СИГНАЛІВ НА ОСНОВІ ПРИНЦИПУ МІНІМУМУ ТРИВАЛОСТІ

Методи міріадної й меридіанної фільтрації відомі як робастні методи обробки сигналів. Теорія цих методів заснована на узагальненому розподілі Коші та критерії максимальної правдоподібності. Ми надаємо альтернативний підхід, заснований на принципі мінімуму тривалості, до обгрунтування та узагальнення методів міріадної й меридіанної фільтрації. Запропонований підхід показує, что міріадна й меридіанна фільтрації є частинними випадками методів фільтрації, які виводяться з концепції «квазітривалість сигналу». Математично ця концепція реалізується через поняття функционалу за допомогою запропонованої множини вартісних функцій. На цьому фундаменті побудовано «суперсімейство» функціоналу квазітривалість та уведено загальний клас методів фільтрації, який залежить від трьох параметрів, що вільно налаштовуються. Приводяться результати числового моделювання для порівняння ефективності запропонованих та стандартніх методів для задачі фільтрації постійного сигналу, який є спотвореним сумішшю шумів Коші, Лапласа та Гаусса.

Ключові слова: міріадна фільтрація, меридіанна фільтрація, тривалість.

REFERENCES

1. Gonzalez J. G., Arce G. R. Optimality of the myriad filter in practical impulsive-noise environments, *IEEE Trans. on Signal Processing*, 2001, Vol. 49, No. 2, pp. 438–441.
2. Nunez R. C., Gonzalez J. G., Arce G. R. Fast and accurate computation of the myriad filter via branch-and-bound search, *IEEE Trans. Signal Processing*, 2008, Vol. 56, No. 7, P. 3340–3346.
3. Aysal T. C., Barner K. E. Meridian filtering for robust signal processing, *IEEE Trans. on Signal Processing*, 2007, Vol. 55, No. 8, P. 3949–3962.
4. Kurkin D. A., Lukin V. V., Roenko A. A., Djurovic I. Analysis of meridian estimator performance for non-Gaussian PDF data samples, *Telecommunications and Radio Engineering*, 2010, Vol. 69, No. 8, pp. 669–679.
5. Huber P. J. Robust estimation of a location parameter, *Annals of Mathematical Statistics*, 1964, Vol. 35, No. 1, pp. 73–101.
6. Huber P. J. Robust statistics. New York, John Wiley and Sons, 1981, 312 p.
7. Carrillo R. E., Aysal T. C., Barner K. E. Generalized Cauchy distribution based robust estimation, *Acoustic, Speech and Signal Processing : IEEE International Conference ICASSP 2008, Las Vegas, 31 March – 4 April 2008 : proceedings*. IEEE, 2008, pp. 3389–3392. DOI: 10.1109/ICASSP.2008.4518378
8. Vovk S. M., Borul'ko V. F. A minimum-duration method for recovering finite signals, *Radioelectronics and Communications Systems,* 1991, Vol. 34, pp. 67–69.
9. Vovk S. M., Borulko V. F. Elimination of the measurement background by the minimum duration method, *Radioelectronics and Communications Systems,* 1998, Vol. 41, pp. 48–49.
10. Vovk S. M., Borul'ko V. F. Statement of a problem of definition of linear signals parameters in quasinormed space, *Radioelectronics and Communications Systems,* 2010, Vol. 53, pp. 367–375.
11. Titchmarsh E. C. The theory of functions. New York, Oxford University Press, 1939, 454 p.
12. Hartley R., Zisserman A. Multiple View Geometry in Computer Vision. Cambridge University Press, 2004, 645 p.
13. Vovk S., Borulko V. Family of generalized Demidenko functionals for robust estimation, *Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory: 17th International Seminar/Workshop DIPED'12, Tbilisi, 24–27 September 2012 : proceedings*. IEEE, 2012, pp. 151–154.
14. Demidenko E. Z. Optimizacija i regressija. Moscow, Nauka, 1989, 296 p.
15. Meshalkin L. D. Some mathematical methods for the study of non-communicable diseases, *Uses of epidemiology in planning of health services. Proc. of the Sixth International Scientific Meeting, Primosten, August 29–September 3,* 1971. Belgrade, 1973, pp. 250–2566.
16. Shevlyakov G. L., Vilchevski N. O. Robustness in data analysis: criteria and methods. Utrecht, VSP, 2002, 310 p.
17. Vovk S., Borulko V. Dual Method of Minimum Spatial Extent, *Mathematical Methods in Electromagnetic Theory: 14th International Conference MMET-2014*. Dnipropetrovsk, Ukraine, 26–28 August 2014, proceedings, IEEE, 2014, P. 144–147. DOI: 10.1109/MMET.2014.6928715.

p-ISSN 1607-3274. Радіоелектроніка, інформатика, управління. 2016. № 1
e-ISSN 2313-688X. Radio Electronics, Computer Science, Control. 2016. № 1

РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ

РАДИОЭЛЕКТРОНИКА ТА ТЕЛЕКОММУНИКАЦИИ

RADIO ELECTRONICS AND TELECOMMUNICATIONS

УДК 638.235.231

Зінченко М. В.[1], Зіньковський Ю. Ф.[2]

[1]Канд. техн. наук, доцент кафедри радіоконструювання та виробництва радіоапаратури Національного технічного університету України «Київський політехнічний інститут», Київ, Україна
[2]Д-р техн. наук, професор, професор кафедри радіоконструювання та виробництва радіоапаратури Національного технічного університету України «Київський політехнічний інститут», Київ, Україна

ШИРОКОСМУГОВІ РОЗСІЮВАЧІ В ЗАДАЧАХ НЕЛІНІЙНОЇ РАДІОЛОКАЦІЇ

Вирішено завдання впровадження єдиного імітатора нелінійного розсіювача для визначення показників призначення нелінійних радіолокаторів. Показано, що для об'єктивного порівняння нелінійних радіолокаторів за показниками призначення у реальних умовах необхідно враховувати вплив на розсіюваний нелінійним об'єктом сигнал випромінювань вузькосмугових сторонніх джерел та відгуків структур «метал-окисел-метал». Штатні імітатори розсіювачів, що входять до комплекту нелінійних радіолокаторів, не можна використовувати у якості еталонних, оскільки вони є резонансними. Запропоновано використовувати імітатор нелінійного розсіювача на базі двозаходової плоскої спіральної антени з узгодженим нелінійним навантаженням, оскільки за рахунок широкосмуговості та еліптичної поляризації матимемо ефективне поглинання енергії зондуючого сигналу з подальшим перевипромінюванням вагомих за рівнем нелінійних продуктів у порівнянні з випромінюваннями вузькосмугових сторонніх джерел та спектральних складових відгуків структур «метал-окисел-метал». Проведені у «польових» умовах експериментальні дослідження підтвердили вказані переваги запропонованого імітатора над штатними калібрувальними розсіювачами. Вимірювання для нелінійних радіолокаторів максимальної відстані виявлення вибірки імітаторів на базі двозаходової плоскої спіральної антени з узгодженим нелінійним навантаженням показали високу відтворюваність параметрів розсіювачів, що дозволяє використовувати їх у якості еталонів.

Ключові слова: ближня радіолокація, нелінійний радіолокатор, зондуючий сигнал, імітатор нелінійного розсіювача, двозаходова плоска спіральна антена.

НОМЕНКЛАТУРА

НР – нелінійний радіолокатор;
НРс – нелінійний розсіювач;
МОМ – «метал-окисел-метал»;
ЗС – зондуючий сигнал;
ККД – коефіцієнт корисної дії;
НВЧ – надвисокі частоти;
ДС – діаграма спрямованості;
КСД – коефіцієнт спрямованої дії;
ПСА – плоска спіральна антена;
A – векторний потенціал поляризаційної структури поля випромінювання ПСА;
J – вектор комплексної амплітуди об'ємної щільності струму;
$G(R)$ – функція Гріна для вільного простору;
R – відстань спостереження;
V – об'єм випромінюючої системи
k – хвильове число;
μ_0 – абсолютна магнітна проникність середовища;
AR – рівень послаблення зондуючого сигналу в дБ;
BR – чутливість приймачів другої і третьої гармонік в дБ.

DOI 10.15588/1607-3274-2016-1-2

ВСТУП

Нелінійна радіолокація використовується у багатьох сферах людської діяльності: технічний захист інформації, технології військового призначення, неруйнівний контроль якості виробів тощо. Нелінійні радіолокатори (НР) призначені для виявлення «закладних» радіоелектронних пристроїв, складовими елементної бази котрих є напівпровідникові прилади. Функціонування НР безпосередньо пов'язане з ефектом перевипромінювання нелінійними розсіювачами (НРс) – антенними структурами з нелінійним навантаженням, у простір під час зондування нових спектральних складових, не характерних для спектру опромінюючого сигналу (кратні гармоніки чи комбінаційні частоти). За результатом аналізу характеристик прийнятих нелінійних продуктів сигналу відгуку опе-

ратор робить висновок щодо знаходження та ідентифікацію у досліджуваному середовищі НРс [1–3].

Як правило, у реальних умовах використання нелінійних радіолокаторів на корисний сигнал відгуку від шуканого НРс накладаються спектральні складові з тими ж частотами від випромінювань вузькосмугових сторонніх джерел та відгуків структур «метал-окисел-метал» (МОМ-структур).

Вплив зовнішніх завад на НРс може призводити до появи мультиплікативних складових в розсіяному сигналі, що може створити фіктивні сигнали відгуку навіть при відсутності ЗС НР. Частота завадового сигналу може виявитися близькою до частоти корисного розсіяного сигналу та потрапити у смугу приймача. Боротьба з зовнішніми завадами вирішується, як правило, обранням частоти ЗС, забезпеченням можливості її зміни та відповідними схемотехнічними та конструкторськими рішеннями [4–5].

Причина виникнення «хибних» відгуків від МОМ-структур пов'язана з тим, що слабкі металеві контакти, як правило, є квазінелінійними елементами з симетричною вольт-амперною характеристикою. При цьому контактні нелінійності є джерелом нелінійних продуктів, в основному, непарного порядку, а присутні гармоніки парного порядку за рівнем менші на 20 дБ і більше. Селекція шуканих НРс від МОМ-структур здійснюється: за відношенням рівнів прийнятих перевипромінювань другої та третьої гармонік частоти ЗС НР; за характером зміни амплітуди сигналу на виході приймача поблизу перевипромінюючого об'єкта; за реакцією об'єкта на вібродію [6–7].

Таким чином, ефективність використання нелінійних радіолокаторів безпосередньо пов'язана з мінімізацією впливу випромінювань вузькосмугових сторонніх джерел та відгуків структур «метал-окисел-метал».

1 ПОСТАНОВКА ЗАДАЧІ

Сертифікація нелінійних радіолокаторів передбачає випробовування за багатьма показниками їх якості, вагомим серед яких є максимальна відстань виявлення об'єкта пошуку. Об'єктивність порівняння зразків нелінійних радіолокаторів за цим показником вимагає впровадження єдиного імітатора НРс [8–10]. Штатний калібрувальний розсіювач, відносно якого виконується налаштування в «польових» умовах конкретного НР, у більшості випадків є резонансним та налаштованим на частоту ЗС НР. Таким чином, він ефективно поглинає енергію зондуючого сигналу певного НР і перевипро-мінює достатньо вагомі за рівнем нелінійні продукти у порівнянні з випромінюваннями вузькосмугових сторонніх джерел та спектральних складових відгуків МОМ-структур. Оскільки робоча частота ЗС для різних типів НР лежить у широкому діапазоні (600…1000) МГц, то штатні резонансні імітатори НРс по різному поглинатимуть опромінюючий сигнал однакової підведеної потужності від нелінійних радіолокаторів. Таким чином, для об'єктивного порівняння єдиний імітатор НРс повинен задовольняти наступним вимогам: бути широкосмуговим; мати кругову або еліптичну поляризацію; конструктивно здатним до високої відтворюваності параметрів; забезпечувати спрямовану передачу потужності; володіти високим коефіцієнтом корисної дії (ККД); забезпечувати заданий рівень узгодження з нелінійним навантаженням.

2 ОГЛЯД ЛІТЕРАТУРИ

В Україні та Росії за напрямком нелінійної радіолокації працювали групи дослідників під керівництвом Штейншлейгера В. Б., Вернигорова Н. С., Парватова Г. Н., Петрова Б. М., Горбачева О. О., Шифріна Я. С. та інших.

Провідними американськими дослідниками у нелінійній радіолокації стали Thomas H. Jones, Bruce R. Barsumian, Robert A. Rubega та інші. В Англії – відповідно James H. Stephen, John D. McCann, Steven John Holmes, Andrew Barry Stephen та інші.

Потужність неперервного ЗС НР в більшості випадків становить 10...850 мВт. В імпульсному режимі випромінювання пікова потужність в імпульсі 5...400 Вт. Деякі сучасні НР мають можливість зміни потужності ЗС. Чутливість приймачів сучасних НР лежить у межах від 10^{-15} до 10^{-11} Вт. В імпульсних вона трохи гірша, що пояснюється відповідною перевагою рівня пікової потужності імпульсних передавачів (приблизно на 35–40 дБ). У більшості НР використовуються приймачі з регульованою чутливістю. Діапазон регулювання цього параметра становить 30...50 дБ [6, 11].

Дальність дії більшості НР обмежена величиною близько 1 м. Обмеження відповідає варіанту роботи на відкритих площах або у великих необладнаних приміщеннях. Для офісних приміщень можливості виявлення ще скромніші, що пов'язано з високою концентрацією різних завад [6, 11, 12]. З поняттям максимальної дальності дії тісно пов'язана максимальна глибина виявлення об'єктів у досліджуваному середовищі. Для будівельних конструкцій вона може досягати лише декількох десятків сантиметрів.

Згідно з [6, 11] дальність виявлення НРс при імпульсному режимі набагато більша в порівнянні з неперервним випромінюванням ЗС НР.

Дальність дії нелінійних радіолокаторів може варіюватися не тільки їх енергетичним потенціалом і коефіцієнтом шуму прийомного пристрою, але і паразитними або побічними нелінійними ефектами [6]:

– паразитні бічні пелюстки діаграми спрямованості (ДС) випромінюючої антени НР провокують появу побічних гармонік від навколишніх радіоелектронних засобів;

– наявність у ЗС НР паразитних гармонічних складових, які, відбиваючись від поверхні, попадають у прийомний тракт НР;

– виникнення в досліджуваному об'єкті за певних умов ефектів, які якісно впливають на зміну властивостей його демаскуючих ознак.

Широкий спектр практичного використання ефекту нелінійного розсіювання в багатьох сферах людської діяльності сприяв появі численних фундаментальних досліджень за тематикою нелінійної радіолокації. Більшість проведених робіт присвячені задачі підвищення ефективності виявлення, ідентифікації та локалізації нелінійних розсіювачів на фоні різних завад. Сучасні нелінійні радіолокатори для пошуку, ідентифікації та локалізації нелінійних розсіювачів використовують первинні демаскуючі ознаки нелінійних розсіювачів, тобто всі можливі спостере-

жувані за допомогою відповідної апаратури явища та процеси в досліджуваному середовищі, які породжуються або зазнають певних змін за рахунок наявних нелінійностей характеристик напівпровідникових структур НРс.

Як правило, підвищення ефективності нелінійних радіолокаторів у теперішній час зводиться до збільшення потужності випромінюваного НВЧ-сигналу НР, підвищення чутливості приймачів, вибору оптимальних параметрів ЗС тощо. Все це вимагає вирішення досить складних схемотехнічних та конструкторських задач електромагнітної сумісності, забезпечення високої точності вихідних параметрів, завадостійкості тощо. При цьому виграш в більшості випадків незначний та не відповідає витратам [13].

Паспортні дані більшості представлених на ринку нелінійних радіолокаторів на показники ефективності використання (дальність дії, роздільна здатність, вибірковість тощо) важко співставити нормативам документа технічного захисту інформації НД ТЗІ 1.4.–002–08 «Радіолокатори нелінійні. Класифікація. Рекомендовані методи та засоби випробувань». При цьому випробування за нормативним документом необхідно здійснювати у лабораторних умовах. Запропонований в документі імітатор необхідно підлаштовувати під кожний НР, що теж додає незручностей та вносить вагому похибку.

3 МАТЕРІАЛИ І МЕТОДИ

Перспективними в якості антенних структур імітатора НРс є спіральні антени, оскільки вони є широкосмуговими та мають еліптичну поляризацію [14].

Плоскі спіральні антени (ПСА) виконуються зі спіралей двох видів: рівнокутних логарифмічних і арифметичних (архімедових). Гілки спіралей можуть бути або провідниками, розташованими на діелектричній основі, або виконуватися у вигляді щілин в провідній площині. Зазвичай плоскі спіралі мають дві гілки і залежно від фазових співвідношень в точці збудження можуть працювати в двох режимах: осьовому (направленому) і ненаправленому випромінюванні. Якщо дві гілки спіралі збуджуються в протифазі, то виникає режим осьового випромінювання, при якому головна пелюстка діаграми спрямованості розташована уздовж осі спіралі. Режим ненаправленого випромінювання, при якому поле максимальне в площині спіралі, має місце при синфазному збудженні її гілок.

Аналіз роботи плоскої антени, виконаної з архімедової спіралі, базується на твердженні, що її випромінювання визначається в основному тим витком, де струми в суміжних елементах спіралі майже синфазні. За межами основного випромінюючого витка існують додаткові витки, параметри яких кратні параметрам основного. Однак експериментальні дані показують, що ці смужки випромінюють лише малу частину енергії. При зміні довжини хвилі основний випромінюючий виток автоматично переміщується уздовж радіуса спіралі, зберігаючи постійність своєї електричної довжини.

Спіральні антени з коефіцієнтом перекриття по частоті від 1,5 до 10 дозволяють формувати односпрямовані ДС шириною (90°…180°) з коефіцієнтом спрямованої дії (КСД) (2...8).

Для ПСА вхідний опір і розподіл струму на провідниках розраховується із застосуванням методу узагальнених наведених ЕРС. Для аналізу поляризаційної структури поля випромінювання ПСА використовується метод векторного потенціалу [14]:

$$A = \frac{\mu_0}{4\pi}\int_V JG(R)dV,$$

де A – векторний потенціал у довільній точці спостереження, який визначається на підставі відомого (попередньо обчисленого) розподілу збуджуючого струму; J – вектор комплексної амплітуди об'ємної щільності стороннього електричного струму; $G(R) = \exp(-jkR)/R$ – функція Гріна для вільного простору; R – відстань між точками спостереження та інтегрування; V – об'єм, заповнений струмами випромінюючої системи; k – хвильове число; μ_0 – абсолютна магнітна проникність середовища.

Для поляризаційних досліджень двозаходової ПСА широко використовуються два методи, які визначені видом розкладання електромагнітної хвилі: на дві хвилі з ортогональними лінійними поляризаціями і на дві хвилі з круговими поляризаціями протилежного напрямку обертання [14].

Розглянемо в якості антенної структури імітатора НРс двозаходову ПСА з нелінійним навантаженням. Для проведення розрахунків обрано діапазон частот (0,8...3) ГГц. Двозаходова ПСА з максимальним радіусом 26 мм, виконана з фольгованого текстоліту і навантажена на діод типу КД-522А в точках А і В (рис. 1).

Геометричні параметри випромінюючої структури: кількість входів – 2; початковий радіус (22…26) мм; кількість витків 2…3. У смузі частот (0,8...3) ГГц діаграма спрямованості зазнає змін (рис. 2, 3). Коефіцієнт спрямованої дії на частоті 1 ГГц становить $(6{,}5...2{,}5)$ за ширини головної пелюстки ДС (90°…120°).

КСД на частоті 2 ГГц становить (2,0…0,5) дБ за ширини головної пелюстки ДС (120°…160°). На частотах понад 2 ГГц спостерігається зниження КСД, обумовлене розширенням головної пелюстки ДС і зростанням ширини випромінювання.

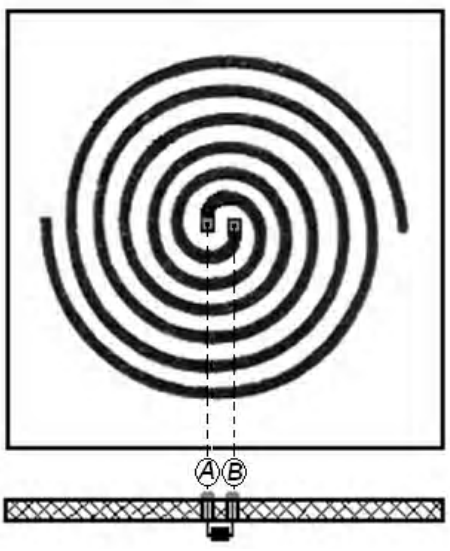


Рисунок 1 – Імітатор НРс на базі плоскої двозаходової спіральної антени

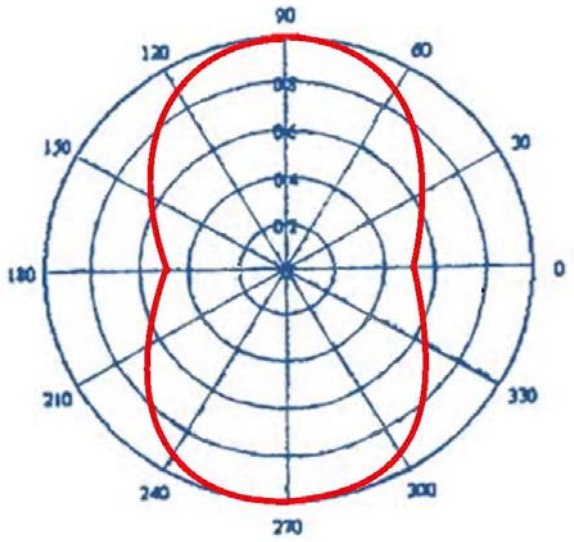


Рисунок 2 – ДС ПСА, розрахована в наближенні заданого струму в Matchcad 13 для частоти 1 ГГц

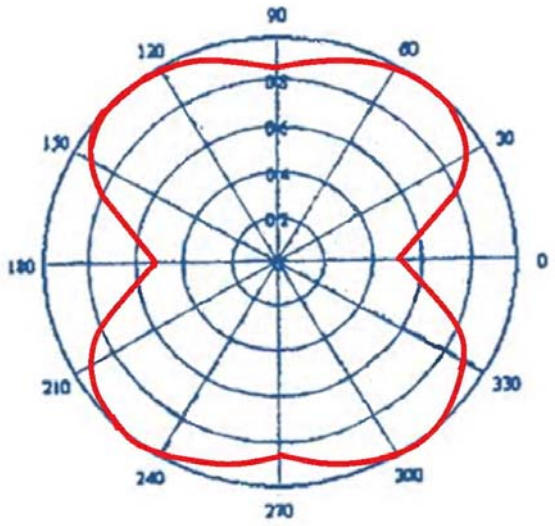


Рисунок 3 – ДС ПСА, розрахована в наближенні заданого струму в Matchcad 13 для частоти 2 ГГц

4 ЕКСПЕРИМЕНТ

Для підтвердження ефективності розробленого імітатора були проведені експериментальні дослідження у реальних умовах трьох зразків розсіювачів (одного на базі ПСА та двох на базі симетричного вібратора) нелінійним радіолокатором «NR-μ», антени якого мають кругову поляризацію (кругова поляризація антен є характерною для більшості НР, оскільки її застосування суттєво покращує спроможність виявлення радіоелектронних об'єктів), максимальна потужність сигналу зондування в імпульсі становить 250 Вт на частоті 848 МГц, а чутливість приймачів не перевищує – 140 дБ/Вт. Дослідження проводилися за схемою, представленою на рис. 4. Зазначимо, що в експерименті резонансні імітатори НРс на базі симетричного вібратора мали різну довжину пліч (на рис. 4 *а* довжина пліч штатного імітатора НРс співрозмірна з довжиною хвилі ЗС НР, а на рис. 4 *б* – на порядок менша).

Виміри проводилися для рівнів другої та третьої гармонік сигналу відгуку у шістьох точках згідно рис. 4. Причому основна площина досліджуваних зразків була розміщена перпендикулярно до вісі, на якій лежать точки спостереження 1, 2 та 3.

5 РЕЗУЛЬТАТИ

Результати дослідження співрозмірного з довжиною хвилі ЗС НР симетричного вібратора з узгодженим у навантаженні діодом КД522А (рис. 4*а*) у вигляді гістограм співвідношень рівнів другої та третьої гармонік сигналу відгуку в дБ, приведені на рис. 5, 6, де *AR* – рівень послаблення зондуючого сигналу в дБ, *BR* – чутливість приймачів другої і третьої гармонік в дБ (ці позначення використані і на рис. 7–10).

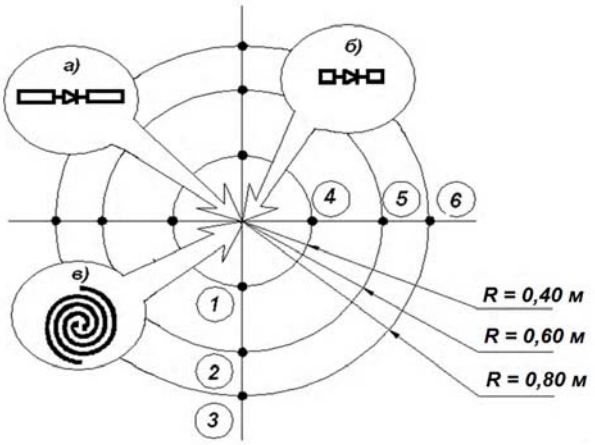


Рисунок 4 – Схема дослідження ефективності ідентифікації НРс

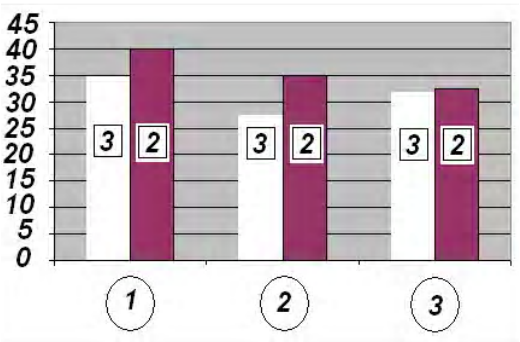


Рисунок 5 – Рівні гармонік, дБ, виміряних в точках 1, 2 і 3, згідно з рис. 4 при *AR*=0 дБ, BR=0 дБ

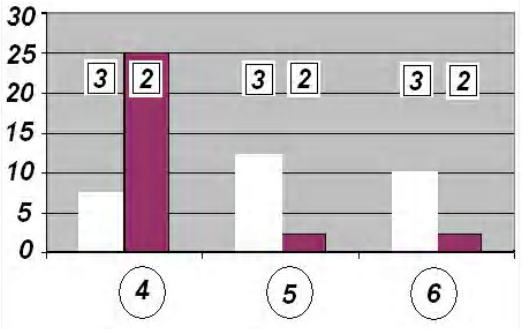


Рисунок 6 – Рівні гармонік, дБ, виміряних в точках 4, 5 і 6, згідно з рис. 4 при *AR*= –5 дБ, *BR*=0 дБ

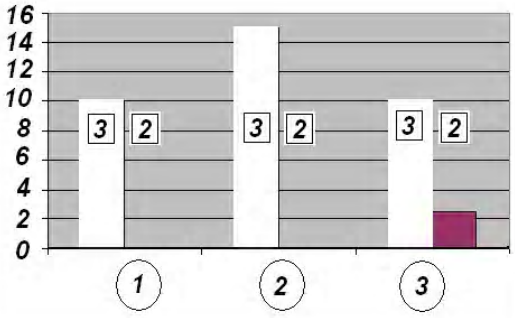


Рисунок 7 – Рівні гармонік, дБ, виміряних в точках 1, 2 і 3, згідно з рис. 4 при *AR*= –5 дБ, *BR*=0 дБ

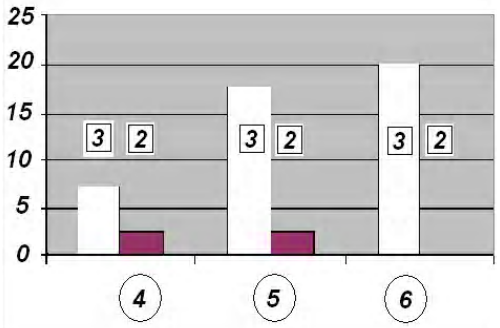


Рисунок 8 – Рівні гармонік, дБ, виміряних в точках 4, 5 і 6, згідно з рис. 4 при *AR*= –5 дБ, *BR*=0 дБ

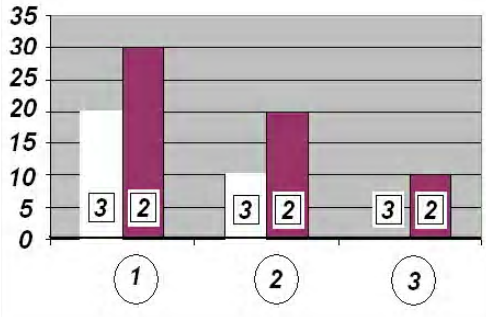


Рисунок 9 – Рівні гармонік, дБ, виміряних в точках 1, 2 і 3, згідно з рис. 4 при *AR*= –10 дБ, *BR*= –30 дБ

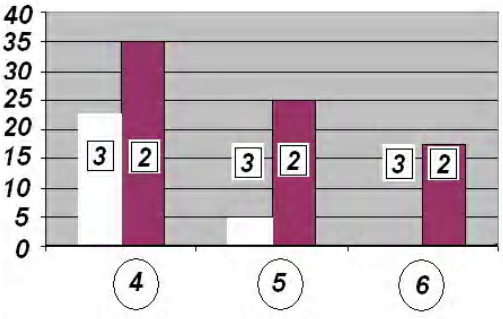


Рисунок 10 – Рівні гармонік, дБ, виміряних в точках 4, 5 і 6, згідно з рис. 4 при *AR*= –10 дБ, *BR*= –20 дБ

На рис. 7, 8 представлені дослідження для штатного НРс з «електрично малою» антеною з узгодженим у навантаженні діодом КД522А (рис. 4 *б*).

На рис. 9, 10 приведені результати дослідження двозаходової ПСА з узгодженим у навантаженні діодом КД522А (рис. 4 *в*).

У табл. 1, для п'яти різних типів серійних НР: y_1 – «ОНЕГА 3»; y_2 – «NR 900M»; y_3 – «NR 900E»; y_4 – «РОДНИК 23»; y_5 – «ЦИКЛОН М1А», приведені результати дослідження максимальної відстані виявлення зразків імітатора НРс на базі двозаходової ПСА з узгодженим у навантаженні діододом КД522А (виміри здійснювались перпендикулярно основній площині досліджуваних зразків).

6 ОБГОВОРЕННЯ

З рис. 9–10 видно, що імітатор на базі двозаходової ПСА з узгодженим нелінійним навантаженням за рахунок широкосмуговості та еліптичної поляризації перевипромінює досить вагомі рівні нелінійних продуктів сигналу відгуку у порівнянні з резонансними калібрувальними розсіювачами. Це дає можливість у реальних умовах чітко його ідентифікувати апаратурою нелінійної радіолокації на досить великих відстанях при будь-якій його просторовій орієнтації відносно дії ЗС, оскільки маємо ефективне поглинання енергії зондуючого сигналу з подальшим перевипромінюванням вагомих за рівнем нелінійних продуктів, що значимо переважають наявні випромінювання вузькосмугових сторонніх джерел та спектральних складових відгуків структур «метал-окисел-метал» [15, 16].

Значний вплив оточуючого середовища підтверджує хибна ідентифікація штатного резонансного НРс з «електрично малою» антеною з узгодженим нелінійним навантаженням як МОМ-структури, оскільки рівень третьої гармоніки суттєво перевищував рівень другої (див рис. 7, 8).

Чітка ідентифікація співрозмірного з довжиною хвилі ЗС НР симетричного вібратора з узгодженим навантаженням на великих для нелінійної радіолокації відстанях можлива лише за умови перпендикулярної орієнтації основної площини досліджуваного зразка дії ЗС НР (рис. 5). У випадку опромінення «збоку» зі збільшенням відстані значимим стає вплив оточуючого середовища, тому резонансний імітатор НРс ідентифікується за співвідношенням рівнів другої та третьої гармонік як МОМ-структура (рис. 6).

Дослідження максимальної відстані виявлення імітаторів НРс на базі двозаходової ПСА з узгодженим нелінійним навантаженням проводилися для восьми зразків з метою перевірки якості відтворювання їх параметрів (табл. 1).

На рис. 11 згідно результатів табл. 1 представлена кумулятивна функція розподілу максимальної відстані виявлення розробленого імітатора НРс серійними нелінійними радіолокаторами [17]. З якісного візуального аналізу представленої кумулятивної функції можна зробити висновок, що закон розподілу з високою довірчою вірогідністю є нормальним.

Перевірка гіпотези щодо виду функції розподілу за критерієм Шапіро-Уілка згідно ISO 5479-97 теж підтвердила, що розподіл у вибірці є нормальним, не відхиляється при рівні значимості $\alpha = 0,05$.

Нормальний закон розподілу та мінімальні розходження середніх значень максимальної відстані виявлення по зразкам широкосмугових імітаторів НРс (відповідність критеріям Фішера за довірчої ймовірності 99,8% [17]) свідчить про високу якість відтворення параметрів імітаторів НРс на базі двозаходової ПСА з узгодженим нелінійним навантаженням, а тому їх можна вважати за еталонні.

За середніми значеннями максимальної відстані виявлення еталонних зразків нелінійними радіолокаторами можна здійснити впорядкування типів НР за ефективністю використання. Так, кращим є «ЦИКЛОН М1А», а гіршим «ОНЕГА 3». Причому відносно середнього значення по всій вибірці значень максимальної відстані

Таблиця 1 – Дослідження максимальної відстані виявлення зразків імітатора НРс

Досліджуваний пристрій, на базі ПСА	Максимальна відстань виявлення об'єкта за видами НР, м					Середні значення по імітаторам НРс
	y_1	y_2	y_3	y_4	y_5	
Зразок №1	1,73	1,80	1,76	1,62	2,00	1,78
Зразок №2	1,68	1,76	1,83	1,59	2,05	1,78
Зразок №3	1,62	1,82	1,85	1,66	1,98	1,79
Зразок №4	1,65	1,73	1,74	1,71	1,96	1,76
Зразок №5	1,66	1,82	1,71	1,70	1,94	1,77
Зразок №6	1,71	1,75	1,73	1,57	2,03	1,76
Зразок №7	1,74	1,87	1,84	1,63	1,91	1,80
Зразок №8	1,69	1,79	1,75	1,65	1,96	1,77
Середні значення по НР	1,69	1,79	1,78	1,64	1,98	

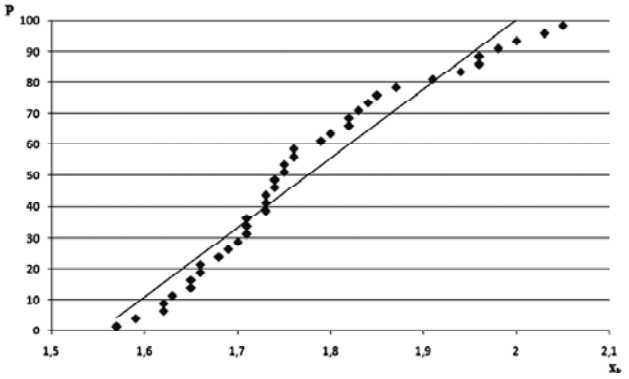


Рисунок 11 – Кумулятивна функція P розподілу максимальної відстані виявлення імітатора НРс на базі двозаходової ПСА серійними нелінійними радіолокаторами

виявлення еталонів (1,77 м), показник якості для «ЦИКЛОН М1А» на 11% кращий, а для «ОНЕГА 3» відповідно на 14% гірший.

ВИСНОВКИ

Сертифікація нелінійних радіолокаторів вимагає впровадження єдиного імітатора нелінійного розсіювача (НРс). Штатні калібрувальні розсіювачі, відносно яких виконується налаштування в «польових» умовах нелінійних радіолокаторів (НР), у більшості випадків є резонансними та налаштовані на конкретну частоту зондуючого сигналу (ЗС). Тому їх не можна обрати для об'єктивного порівняння приладів через широкий діапазон частот зондування (600…1000) МГц існуючих видів НР. Адже необхідно забезпечити ефективне поглинання енергії ЗС НР і перевипромінювання достатньо вагомих за рівнем нелінійних продуктів у порівнянні з випромінюваннями вузькосмугових сторонніх джерел та спектральних складових відгуків структур «метал-окисел-метал». Для об'єктивного порівняння в якості єдиного імітатора НРс доцільно використати двозаходову плоску спіральну антену (ПСА) з узгодженим нелінійним навантаженням, оскільки він задовольняє пред'явленим вимогам: широкосмуговий; має еліптичну поляризацію; конструктивно здатний до високої відтворюваності параметрів; забезпечує спрямовану передачу потужності; володіє високим коефіцієнтом корисної дії; забезпечує заданий рівень узгодження з нелінійним елементом. Експериментальні дослідження показали, що запропонований імітатор за рахунок широкосмуговості та еліптичної поляризації перевипромінює досить вагомі рівні нелінійних продуктів сигналу відгуку у порівнянні з резонансними калібрувальними розсіювачами. Що дає можливість у польових умовах чітко його ідентифікувати апаратурою нелінійної радіолокації на досить великих відстанях при будь-якій просторовій орієнтації відносно дії ЗС. Дослідження максимальної відстані виявлення різними НР зразків широкосмугового імітатора НРс підтвердили прогнозовані мінімальні розходження результатів, що свідчить про високу якість відтворення параметрів розсіювачів, а отже, можливість їх використання в якості еталонних. Це забезпечує отримання значень показників ефективності використання НР, які легко співставляються з нормативами документа технічного захисту інформації НД ТЗІ 1.4.–002–08 «Радіолокатори нелінійні. Класифікація. Рекомендовані методи та засоби випробувань».

СПИСОК ЛІТЕРАТУРИ

1. Вернигоров Н. С. Нелинейный локатор – эффективное средство обеспечения безопасности в области утечки информации / Н. С. Вернигоров // Конфидент. – 1996. – № 1. – С. 67.
2. Вернигоров Н. С. Практические применения нелинейного радиолокатора / Н. С. Вернигоров // Безопасность от А до Я. – 1998. – № 2–3. – С. 14–15.
3. Вернигоров Н. С. Процесс нелинейного преобразования и рассеяния электромагнитного поля электрически нелинейными объектами / Н. С. Вернигоров // Радиотехника и электроника. – 1997. – Т. 42, № 10. – С. 1181–1185.
4. Беляев В. В. Состояние и перспективы развития нелинейной радиолокации / В. В. Беляев, А. Т. Маюнов, С. Н. Разиньков // Зарубежная радиоэлектроника. Успехи современной радиоэлектроники. – 2002. – № 6. – С. 59–78.
5. Помехи в системах нелинейного зондирования / [А. А. Горбачев, С. В. Лавцов, С. П. Тараканков, Е. П. Чигин] // Радиотехника и электроника. – 1998. – Т. 43, № 1. – С. 71–76.
6. Хорошко В. А. Методы и средства защиты информации / В. А. Хорошко, А. А. Чекатков. – К. : «Юниор», 2003. – 504 с.
7. Распознавание нелинейных рассеивателей, содержащих несовершенные металлические контакты или полупроводниковые радиокомпоненты / [А. П. Колбанов, А. А. Потапов, Е. Е. Степанов, Е. П. Чигин] // Нелинейный мир. – 2005. – Т. 3, № 4. – С. 239–244.
8. Калабухов В. А. Нелинейная радиолокация: принципы сравнения / В. А. Калабухов, Д. В. Ткачев // Специальная техника. – 2001. – № 2. – С. 28.
9. Вернигоров Н. С. К вопросу о принципе сравнения в нелинейной радиолокации / Н. С. Вернигоров, Т. В. Кузнецов // ИНФОРМОСТ Радиоэлектроника и Телекоммуникации. – 2002. – № 3(21). – С. 7–14.
10. Авдеев В. Б. Методический аппарат для оценки эффективности средств нелинейной радиолокации и противорадиолокации / В. Б. Авдеев, С. Н. Панычев, Д. В. Сенькевич // Вестник Воронежского государственного технического университета. – 2007. – Т. 3, № 4. – С. 115–119.
11. Зыонг Дык Тхиен. Исследование возможностей и методов построения аппаратуры для нелинейной радиолокации: дис. … к.т.н. : 05.12.04 / Зыонг Дык Тхиен. – М, 2007. – 153 с.
12. Захаров А. В. Методика работы с различными моделями нелинейных локаторов / А. В. Захаров // Конфидент. Зашита информации. – 2001. – № 4. – С. 43–47.
13. Вернигоров Н. С. Влияние антенно-фидерного тракта нелинейного объекта на дальность обнаружения в нелинейной локации / Н. С. Вернигоров, В. Б. Харин // Радиотехника и электроника. – 1997. – Т. 42, № 12. – С. 1267.
14. Юрцев О. А. Спиральные антенны / О. А. Юрцев, А. В. Рунов, А. Н. Казарин. – М. : Сов. радио, 1974. – 224 с.

15. Козлов А. И Эффективная площадь рассеяния нелинейных отражателей / А. И. Козлов, Д. В. Колядов // Научный вестник МГТУ ГА. Радиофизика и радиотехника. – 2004. – № 79. – С. 36–40.
16. Зинченко М. В. Рассеивание плоских волн системой симметричных вибраторов с нелинейными нагрузками при воздействии нелинейного радиолокатора / М. В. Зинченко, Ю. Ф. Зиньковский // Известия высших учебных заведений. Радиоэлектроника. НТУУ «КПИ». – 2010. – Том 53, № 10. – С. 24–34.
17. Радченко С. Г. Математическое моделирование технологических процессов в машиностроении / С. Г. Радченко. – К. : ЗАО «Укрспецмонтажпроект», 1998. – 274 с.

Стаття надійшла до редакції 10.11.2015.
Після доробки 02.12.2015.

Зинченко М. В.[1], Зиньковский Ю. Ф.[2]

[1]Канд. техн. наук, доцент кафедры радиоконструирования и производства радиоаппаратуры Национального технического университета Украины «Киевский политехнический институт», Киев, Украина

[2]Д-р техн. наук, профессор, профессор кафедры радиоконструирования и производства радиоаппаратуры Национального технического университета Украины «Киевский политехнический институт», Киев, Украина

ШИРОКОПОЛОСНЫЕ РАССЕИВАТЕЛИ В ЗАДАЧАХ НЕЛИНЕЙНОЙ РАДИОЛОКАЦИИ

Решена задача внедрения единого имитатора нелинейного рассеивателя для определения показателей назначения нелинейных радиолокаторов. Показано, что для объективного сравнения нелинейных радиолокаторов по показателям назначения в реальных условиях необходимо учитывать влияние на рассеиваемый нелинейным объектом сигнал излучения узкополосных сторонних источников и откликов структур «металл-окисел-металл». Штатные имитаторы рассеивателей, входящие в комплект нелинейных радиолокаторов, нельзя использовать в качестве эталонных, поскольку они являются резонансными. Предложено использовать имитатор нелинейного рассеивателя на базе двозаходовой плоской спиральной антенны с согласованной нелинейной нагрузкой, поскольку за счет широкополосности и эллиптической поляризации будем иметь эффективное поглощение энергии зондирующего сигнала с последующим переизлучением весомых по уровню нелинейных продуктов по сравнению с излучениями узкополосных сторонних источников и спектральных составляющих откликов структур «металл-окисел-металл». Проведенные в «полевых» условиях экспериментальные исследования подтвердили указанные преимущества предложенного имитатора над штатными калибровочными рассеивателями. Измерения для нелинейных радиолокаторов максимального расстояния обнаружения выборки имитаторов на базе двозаходовой плоской спиральной антенны с согласованной нелинейной нагрузкой показали высокую воспроизводимость параметров рассеивателей, что позволяет использовать их в качестве эталонов.

Ключевые слова: ближняя радиолокация, нелинейный радиолокатор, зондирующий сигнал, имитатор нелинейного рассеивателя, двозаходовая плоская спиральная антенна.

Zinchenko M. V.[1], Zinkovskiy Yu. F.[2]

[1]PhD., Department of Radio Design and Electronic Radio Equipment Manufacture of the National Technical University of Ukraine «Kyiv Polytechnic Institute», Kiev, Ukraine

[2]DrSc., Professor, Department of Radio Design and Electronic Radio Equipment Manufacture of the National Technical University of Ukraine «Kyiv Polytechnic Institute», Kiev, Ukraine

BROADBAND SCATTERERS IN NONLINEAR RADAR

The problem of a single nonlinear scatterer simulator introduction for the detection of appointment parameters of nonlinear radars is solved. It is shown that for an objective comparison of nonlinear radars by appointment parameters in the real world it is necessary to take into account the narrowband off-site sources and response of structures «metal-oxide-metal» radiation effect on the dissipated by nonlinear object signal. The regular simulators of scatterers included in the assembly of nonlinear radar, can not be used as reference because they are resonant. It is proposed to use a simulator of a nonlinear scatterer based on flat double-spiral antenna with non-linear matched load, as by broadband and elliptical polarization be effective absorption of the energy of the probing signal with subsequent re-emission weighty in terms of non-linear products as compared with the radiation of a narrowband off-site sources and spectral components of the response of structures «metal-oxide-metal». Carried out in the «field» conditions, experimental studies have confirmed these benefits over the regular imitator of the proposed calibration scatterers. The measurement for nonlinear radar maximum distance detection of simulators sampling based on flat double-spiral antenna with matched nonlinear load shown a high reproducibility of parameters of scatterers, so one can use them as references.

Keywords: short-range radar, nonlinear radar, probing signal, the nonlinear scatterer simulator, flat double-spiral antenna.

REFERENCES

1. Vernigorov N. S. Nelineinyi lokator – effektivnoe sredstvo obespecheniia bezopasnosti v oblasti utechki informatsii, *Konfident,* 1996, No. 1, pp. 67.
2. Vernigorov N. S. Prakticheskie primeneniia nelineinogo radiolokatora, *Bezopasnost ot A do Ya.*, 1998, No. 2–3, pp. 14–15.
3. Vernigorov N. S. Protsess nelineynogo preobrazovaniia i rasseianiia elektromagnitnogo polia elektricheski nelineinymi obektami, *Radiotehnika i elektronika,* 1997, Vol. 42, No. 10, pp. 1181–1185.
4. Beliaev V. V., Maiunov A. T., Razinkov S. N. Sostoianie i perspektivy razvitiia nelineinoi radiolokatsii, *Zarubezhnaia radioelektronika. Uspehi sovremennoi radioelektroniki*, 2002, No. 6, pp. 59–78.
5. Gorbachev A. A., Lavtsov S. V., Tarakankov S. P., Chigin E. P. Pomehi v sistemah nelineinogo zondirovaniia, Radiotehnika i elektronika, 1998, Vol. 43, No. 1, pp. 71–76.
6. Horoshko V. A., Chekatkov A. A. Metody i sredstva zaschity informatsii. Kiev, «Yunior», 2003, 504 p.
7. Kolbanov A. P., Potapov A. A., Stepanov E. E., Chigin E. P. Raspoznavanie nelineinyh rasseivatelei, soderzhaschih nesovershennye metallicheskie kontakty ili poluprovodnikovye radiokomponenty, Nelineynyi mir, 2005, Vol. 3, No. 4, pp. 239–244.
8. Kalabuhov V. A. Tkachev D. V. Nelineynaia radiolokatsiia: printsipy sravneniia, *Spetsialnaia tehnika*, 2001, No. 2, pp. 28.
9. Vernigorov N. S., Kuznetsov T. V. K voprosu o printsipe sravneniia v nelineynoi radiolokatsii, *INFORMOST Radioelektronika i Telekommunikatsii*, 2002, No. 3(21), pp. 7–14.
10. Avdeev V. B., Panychev S. N., Senkevich D. V. Metodicheskii apparat dlia otsenki effektivnosti sredstv nelineynoi radiolokatsii i protivoradiolokatsii, *Vestnik Voronezhskogo gosudarstvennogo tehnicheskogo universiteta*, 2007, Vol. 3, No. 4, pp. 115–119.
11. Zyong Dyk Thien. Issledovanie vozmozhnostei i metodov postroeniia apparatury dlia nelineynoi radiolokatsii: dis. … k.t.n. : 05.12.04 / Zyong Dyk Thien. M, 2007, 153 p.
12. Zaharov A. B. Metodika raboty s razlichnymi modeliami nelineynyh lokatorov, *Konfident. Zashita informatsii*, 2001, No. 4, pp. 43–47.
13. Vernigorov N. S., Harin V. B. Vliianie antenno-fidernogo trakta nelineynogo obekta na dalnost obnaruzheniia v nelineynoi lokatsii, *Radiotehnika i elektronika*, 1997, Vol. 42, No. 12, pp. 1267.
14. Yurtsev O. A., Runov A. V., Kazarin A. N. Spiralnye antenny. Moscow, Sov. radio, 1974, 224 p.
15. Kozlov A. I., Kolyadov D. V. Effektivnaia ploschad rasseianiia nelineinyih otrazhatelei, *Nauchnyi vestnik MGTU GA. Radiofizika i radiotehnika*, 2004, No. 79, pp. 36–40.
16. Zinchenko M. V., Zinkovskii Yu. F. Rasseivanie ploskih voln sistemoi simmetrichnyih vibratorov s nelineynyimi nagruzkami pri vozdeystvii nelineynogo radiolokatora, Izvestiia vyisshih uchebnyh zavedenii. Radioelektronika. NTUU «KPI», 2010, Vol. 53, No. 10, pp. 24–34.
17. Radchenko S. G. Matematicheskoe modelirovanie tehnologicheskih protsessov v mashinostroenii. Kiev, ZAO «Ukrspetsmontazhproekt», 1998, 274 p.

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL AND COMPUTER MODELLING

УДК 519.766.4

Трухан С. В.[1], Бідюк П. І.[2]

[1]Аспірантка інституту прикладного системного аналізу НТУУ «КПІ», Київ, Україна
[2]Д-р техн. наук, професор кафедри математичних методів системного аналізу НТУУ «КПІ», Київ, Україна

МЕТОДИКА АНАЛІЗУ ЕКСТРЕМАЛЬНИХ ДАНИХ ТА ЇЇ ВИКОРИСТАННЯ ПРИ ОЦІНЮВАННІ ПАРАМЕТРІВ УЗАГАЛЬНЕНИХ ЛІНІЙНИХ МОДЕЛЕЙ

Запропонована методика аналізу екстремальних значень з метою її застосування при оцінюванні невідомих параметрів узагальнених лінійних моделей. В якості математичного апарату використано теорію екстремальних значень, яка є одним із розділів математичної статистики та пов'язана з дослідженням відхилень екстремальних значень від медіани у ймовірнісних розподілах. Також розглянуто методи наближення експериментальних даних до класу узагальнених екстремальних розподілів, методи оцінювання невідомих параметрів та вибору оптимального порогу для екстремальних значень. На основі фактичних статистичних даних із галузі страхування та запропонованого підходу побудовано моделі обробки екстремальних значень для подальшого застосування при оцінюванні прогнозних моделей. Прийнятним для подальшого використання виявилась модель з наближенням даних за допомогою узагальненого розподілу Парето. Це підтверджується незначною похибкою та максимальним наближенням емпіричної кривої до теоретичної функції щільності розподілу. Порівняння результатів оцінювання невідомих параметрів моделі за допомогою методу максимальної правдоподібності та байєсівського підходу показало, що байєсівські методи оцінювання є ефективним підгрунтям для розв'язання задачі вибору кращої моделі на основі множини отриманих альтернатив та значень апріорних параметрів. Можливість використання результатів застосування моделей екстремальних значень при побудові прогнозних узагальнених лінійних моделей є підставою для подальшого дослідження.

Ключові слова: теорія екстремальних значень, узагальнені лінійні моделі, поріг екстремального значення, метод максимальної правдоподібності, байєсівський підхід.

НОМЕНКЛАТУРА

GEV – Generalized extreme value;

GPD – Generalized Pareto distribution;

ММП – метод максимальної правдоподібності;

МКМЛ – метод Монте-Карло для марковських ланцюгів;

ТЕЗ – теорія екстремальних значень;

УЛМ – узагальнені лінійні моделі;

F – функція розподілу випадкової величини;

u – поріг випадкової величини;

$X_1,...,X_n$ – послідовність незалежних випадкових величин;

μ – параметр розподілу;

ξ – параметр форми розподілу;

σ – параметр масштабованості.

ВСТУП

У зв'язку з необхідністю розв'язання нових задач моделювання і прогнозування на основі великих обсягів вироджених вхідних даних, які не можна розв'язати з використанням існуючих методів, виникає потреба у розробці нових інтегрованих інформаційних систем, методів та підходів до обробки таких даних. Одним із таких підходів є ТЕЗ. Вона широко застосовується до розв'язання таких задач як регулювання структури портфелю активів у страхуванні, аналіз виникнення ризикових ситуацій у сфері фінансів та кредитування, прогнозуванні трафіку в галузі телекомунікацій.

Задачею теорії екстремальних значень є цілеспрямований аналіз та оцінювання ймовірності появи випадкових величин, пов'язаних з екстремальними, тобто рідкісними подіями. Екстремальні значення не є фіксованими величинами, це нові випадкові величини, які залежать від типу вихідного розподілу та об'ємів вибірки. Наприклад, в області страхування будь-якого майна рідкісною, але ймовірною подією є настання страхового випадку, яке повинно супроводжуватись виплатами страхових премій.

DOI 10.15588/1607-3274-2016-1-3

Саме тому для розв'язання задачі прогнозування страхових виплат пропонується ймовірнісна модель, яка будується із застосуванням теорії екстремальних значень. В свою чергу, одним із ключових моментів побудови адекватної моделі досліджуваного процесу є коректний вибір методу оцінювання параметрів математичних моделей за експериментальними (статистичними) даними. Для розв'язання задачі оцінювання невідомих параметрів моделі часто застосовують метод максимальної правдоподібності та байєсівський підхід. Останній дає можливість точніше оцінювати моделі в умовах невизначеності, а саме, коли статистичні дані мають різні типи розподілів ймовірностей, а також вибрати кращу модель із множини оцінених кандидатів. Перевагою даного підходу є можливість його застосування до обробки статистичних вибірок відносно малих розмірів, а також за наявності пропусків даних [4, 5]. Популярним і відносно універсальним є на сьогодні МКМЛ, який застосовують для оцінювання параметрів лінійних і нелінійних моделей [6–8].

1 ПОСТАНОВКА ЗАДАЧІ

У роботі ставиться за мету застосування теорії екстремальних значень для побудови комплексної моделі обробки екстремальних даних з метою створення УЛМ та оцінювання їх параметрів.

Для досягнення поставленої мети необхідно розв'язати такі задачі:

1) дослідити властивості розподілів екстремальних значень;

2) дослідити методи оцінювання невідомих параметрів моделей екстремальних значень, зокрема можливості використання байєсівського підходу, методу максимальної правдоподібності та ін.;

3) розробити комплексну модель обробки екстремальних значень;

4) навести приклади застосування комплексної моделі для обробки вироджених статистичних даних у страхуванні.

2 ОГЛЯД ЛІТЕРАТУРИ

На ранньому етапі створення статистичної теорії оцінювання найбільша увага приділялась розв'язанню задач наближення кривих розподілу до даних, а значно пізніше – розвитку теорії побудови статистичного висновку. На сьогодні теорія екстремальних значень є складовою частиною багатьох напрямів розвитку практичних наук, таких як гідрологія, астрономія, телекомунікації, економіка та ін. Перші історичні свідчення стосовно існування сімейства розподілів екстремальних значень пов'язані з роботою М. Бернуллі (1709 р.) стосовно визначення середньої тривалості життя. Перші спроби дослідження теорії екстремальних значень грунтувались на використанні нормального розподілу. У 1925 р. Тіппет обчислив ймовірності найбільших значень у нормально розподіленій вибірці із врахуванням різних об'ємів вибірки (до 1000 значень), а також оцінював середній розмах нормально розподілених вибірок (від 2 до 1000 значень). Таблиці Тіппета – це фундаментальний підхід до практичного застосування найбільших величин у вибірці з нормальним розподілом. Саме те, що більшість досліджень грунтувалось на нормальному розподілі, гальмувало розвиток теорії екстремальних значень. Фреше успішно дослідив перший тип розподілу екстремальних даних та отримав граничні розподіли найбільших величин вибірки, запропонував постулат стійкості. Використовуючи даний постулат Фреше та Тіппет винайшли два інші розподіли екстремальних значень та підкреслили повільну збіжність ряду границі розподілів найбільших величин із нормальної вибірки [1].

Проблема недостатніх об'ємів інформації часто зустрічається при дослідженні процесів економічного, фізичного, природничого походження і стає причиною труднощів при розв'язанні задач побудови моделей таких екстремальних даних. Тому виникає потреба у дослідженні інших джерел знань для пошуку оптимальних рішень стосовно обробки даних та побудови моделей. Наприклад, економісту потрібно знайти максимальне значення з деякої вибірки з виродженими даними. Звичайно існує кілька можливих способів, за якими експерт із знаннями досліджуваного процесу може надати інформацію, що має відношення до екстремальної поведінки і яка залежить від наявних даних. Але часто така інформація супроводжується наближеними вимірами, які відрізняються від дійсних значень та роблять хибними майбутні прогнози, що будуються на їх основі.

Тому, виходячи із актуальності задачі обробки екстремальних значень, роботу присвячено дослідженню та розробці комплексної моделі для опису екстремальних значень і оцінюванню невідомих параметрів УЛМ. Такі моделі широко використовують для аналізу страхових випадків, прогнозування продовження старих чи укладення нових страхових договорів, розробці тарифів та андерайтингу, а також у цільовому маркетингу.

3 МАТЕРІАЛИ І МЕТОДИ

Математичну модель екстремальних даних можна представити у вигляді [1]:

$$M_n = \max\{X_1,...X_n\}, \tag{1}$$

де $X_1,...,X_n$ – послідовність незалежних випадкових величин з функцією розподілу F. У виразі (1) величина M_n позначає максимум досліджуваного процесу на інтервалі часу n і має розподіл [1]:

$$\Pr\{M_n \le z\} = \Pr\{X_1 \le z,..., X_n \le z\} = \Pr\{X_1 \le z\} \times$$
$$\times ... \times \Pr\{X_n \le z\} = \{F(z)\}^n. \tag{2}$$

Функція F невідома, а тому розглядається наближена оцінка для F^n. Якщо послідовність констант $\{a_n > 0\}$ та $\{b_n > 0\}$ таких, що

$$\Pr\left\{\frac{M_n - b_n}{a_n} \le z\right\} \to F(a_n x + b_n x)^n \to G(z),$$

при $n \to \infty$, то G – невироджена функція розподілу, яка належить до одного з розподілів екстремальних значень, наприклад, до узагальненого розподілу екстремальних значень (Generalized extreme value – GEV):

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \tag{3}$$

де μ – параметр розподілу; σ – параметр масштабованості; ξ – параметр форми розподілу [2].

Відповідно до теореми про типи розподілів екстремальних значень виділяють три типи таких розподілів, а саме:

1) Розподіл Гумбела:

$$G(z)=\exp\left\{-\exp\left(-\left(\frac{z-b}{a}\right)\right)\right\},\ -\infty<z<\infty;$$

2) Розподіл Фреше:

$$G(z)=\begin{cases}0,\ z\le b;\\ \exp\left(-\left(\frac{z-b}{a}\right)^{-\alpha}\right),\ z>b;\end{cases}$$

3) Розподіл Вейбулла:

$$G(z)=\begin{cases}\exp\left(-\left(-\left(\frac{z-b}{a}\right)\right)^{\alpha}\right),\ z<b.\\ 1,\ z\ge b\end{cases}$$

Для всіх трьох випадків $a>0$, b – дійсне число. Для другої та третьої функції параметр $\alpha>0$. Ці три класи розподілів називають розподілами екстремальних значень, вони зображені на рис. 1.

З рис. 1 видно, що кожен з розподілів має свою форму поведінки хвоста. Наприклад, для розподілу Вейбула хвіст має кінцеву точку $z_{\sup}=\frac{\mu-\sigma}{\xi}$, а для розподілів Фреше та Гумбела $z_{\sup}=\infty$. Крім того, щільність розподілу Гумбела експоненціально затухає, тоді як щільність розподілу Фреше затухає поліноміально. Розподіл Гумбела є наближеним до класу таких відомих як нормальний, лог-нормальний та гамма – розподілів. Розподіл Фреше має тяжкий хвіст, який позначається як $E(X^r)=\infty$ для $r\ge\frac{1}{\xi}$ (що означає нескінченність дисперсії при $\xi\ge 1/2$).

В окремий клас виділяють узагальнений розподіл Парето (*Generalized Pareto Distribution – GPD*), який

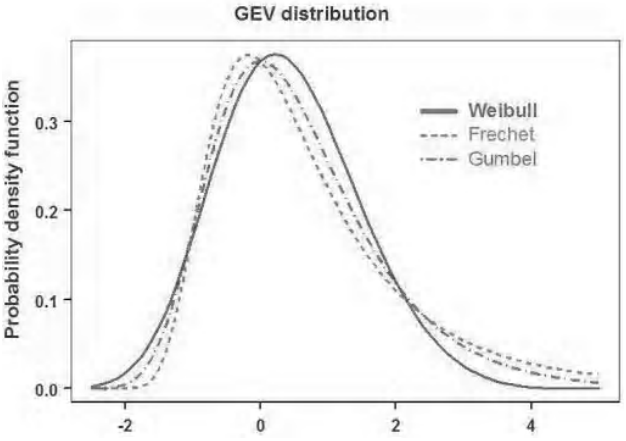


Рисунок 1 – Функції щільності розподілу для трьох типів розподілів

отримуємо за умови: X – це розподіл, що умовно перевищує деякий поріг u:

$$F_u(y)=\frac{F(u+y)-F(u)}{1-F(u)},\qquad(4)$$

де $u\to w_F=\sup\{x:F(x)<1\}$, що найчастіше зводиться до пошуку границі:

$$F_u(y)\approx G(y,\sigma_u,\xi),$$

де G – узагальнений розподіл Парето, еквівалентний виразу [2]:

$$G(y,\sigma,\xi)=1-\left(1+\xi\frac{y}{\sigma}\right)_+^{-1/\xi},$$

1) якщо $\xi>0$, то маємо довгий хвіст $x^{-1/\xi}$, що еквівалентно розподілу Парето;

2) якщо $\xi=0$ та спрямовуючи $\xi\to 0$, отримаємо $G(y,\sigma,0)=1-\exp\left(-\frac{y}{\sigma}\right)$, тобто експоненціальний розподіл з середнім σ;

3) якщо $\xi<0$, то кінцева верхня точка знаходиться на рівні $-\frac{\sigma}{\xi}$.

Також однією із переваг *GEV*-розподілів є інваріантність кожного з розподілів, які належать до даного класу.

Розглянемо методику обробки екстремальних значень. Для обробки статистичного ряду з n-незалежних, однаково розподілених змінних $X_1,...,X_n$ застосовується така послідовність дій.

1. Групування вибірок даних з n спостережень. Такі вибірки повинні містити від 50 до 100 значень.

2. Визначається максимум Z_i для кожного блоку i.

3. Наближення кожного блоку максимумів до *GEV*-розподілу.

Зазвичай за довжину блоку беруть величину першого року, але для зручності часто використовують дані річного максимуму Z_i i-го року.

Після апроксимації *GEV*-розподілом для кожного з річних максимумів розраховується функція квантилю [3, 4]:

$$z_p=\begin{cases}\mu-(\sigma/\xi)\left(1-(-\log(1-p))^{-\xi}\right),\ \ \xi\ne 0;\\ \mu-\sigma\log(-\log(1-p)),\ \ \xi=0.\end{cases}$$

Припустимо, що $y_p=-\log(1-p)$, тоді квантиль-функція матиме вигляд:

$$z_p=\begin{cases}\mu-(\sigma/\xi)\left(1-(y_p)^{-\xi}\right),\ \ \xi\ne 0;\\ \mu-\sigma\log(y_p),\ \ \xi=0;\end{cases}$$

Якщо зобразити z_p в залежності від $\log(y_p)$, то графік буде мати лінійний характер: при $\xi=0$.

Якщо $\xi < 0$, отримаємо випуклу криву з асимптотичною границею $(\mu - \sigma)/\xi$ при $p \to 0$, а при $\xi > 0$ отримаємо увігнутий графік без кінцевої границі.

Такий графік називається графіком повернення рівня (return level plot), він вважається інструментом або способом представлення згладженої моделі [3].

4. Виконується оцінювання параметрів моделі та розв'язується задача пошуку оптимальної довжини блоку.

Остання зводиться до пошуку співвідношення між величинами відхилення та дисперсії. Наприклад, коли довжина блоків незначна, то наближення розподілів до границь є поганим і призводить до відхилень у оцінюванні та екстраполяції. З іншого боку, великі блоки породжують значення з великими оцінками дисперсії.

Для оцінювання параметрів моделей часто використовується метод максимальної правдоподібності (ММП). Однак, умова регулярності оцінювання не задовольняється при застосуванні ММП до *GEV*-розподілів, тому що кінцева точка розподілів залежить від значення параметра. Це означає, що стандартні асимптотичні результати аналізу за методом максимальної правдоподібності недоречно застосовувати до *GEV*-розподілів. Цю проблему дослідив Сміт у 1985 році з такими результатами [3]:

– якщо $\xi > -0{,}5$, то оцінювання за ММП носить стандартний асимптотичний характер;

– якщо $-1 < \xi < 0{,}5$, то оцінки ММП можуть бути отримані, але не із заданими асимптотичними властивостями;

– якщо $\xi < -1$, то оцінки ММП вважаються неправдоподібними.

Окремий випадок: якщо $\xi < -0{,}5$, то це еквівалентно розподілу з дуже коротким обмеженим верхнім хвостом, який є рідкісним явищем для теорії екстремальних значень [5].

Логарифмічна функція правдоподібності для *GEV*-розподілів, коли $\xi \neq 0$, має вигляд:

$$l(\mu,\sigma,\xi) = -m\log\sigma - (1+1/\xi)\sum_{i=1}^{m}\log\left(1+\xi\frac{z_i-\mu}{\sigma}\right) - \sum_{i=1}^{m}\left(1+\xi\frac{z_i-\mu}{\sigma}\right)^{-1/\xi}$$

за умови, що $\left(1+\xi\frac{z_i-\mu}{\sigma}\right) > 0$ для $i = 1,...,m$. Як тільки остання умова не виконується, то функція правдоподібності дорівнює нулю і логарифмічна функція правдоподібності набуває значення нескінченності.

Для розподілу Гумбела $\xi = 0$ логарифмічна функція правдоподібності має вигляд:

$$l(\mu,\sigma) = -m\log\sigma - \sum_{i=1}^{m}\left(\frac{z_i-\mu}{\sigma}\right) - \sum_{i=1}^{m}\left(-\frac{z_i-\mu}{\sigma}\right). \quad (5)$$

Після використання методів чисельної оптимізації та максимізації виразу (5), отримуємо оцінку максимальної правдоподібності вигляду $\left(\hat{\mu},\hat{\sigma},\hat{\xi}\right)$ [3, 5].

5. Графічна перевірка наближення *GEV*-моделей.

Для обгрунтування екстраполяції *GEV*-моделей можна скористатись способами графічного аналізу даних.

Графік щільності розподілу. В основі даного графіка лежить порівняння емпіричної та апроксимуючої функцій щільності розподілу. Абсциса точки на графіку щільності розподілів є емпіричною функцією розподілу, у яку замість аргументу підставляють дані з вибірки, а ордината – це теоретична функція розподілу, куди аналогічно замість аргументу підставляють дані із статистичної вибірки. Функція емпіричного розподілу оцінюється в i-му упорядкованому блоці максимумів Z_i і має вигляд:

$$\widetilde{G}_i(Z_i) = i/(m+1).$$

Апроксимуюча функція щільності розподілу в тій самій точці виглядає так:

$$\hat{G}(Z_i) = \exp\left\{-\left(1+\hat{\xi}\left(\frac{z_{(i)}-\hat{\mu}}{\hat{\sigma}}\right)\right)^{-1/\hat{\xi}}\right\}.$$

Для того, щоб отримати найкраще наближення моделі необхідно задовольнити рівність $\widetilde{G}(Z_i) = \hat{G}(Z_i)$. За допомогою цього графіка на практиці часто вдається запобігти ефекту «виродженості». Тобто, коли множина точок $\left\{\widetilde{G}(Z_i),\ \hat{G}(Z_i)\right\},\ i = 1,...,m$ – лежить близько до першої діагоналі в той час, коли обидві функції є обмеженими в околі одиниці та значення абсциси z збільшуються.

Графік квантилів (*Q–Q plot*). Недоліком класичної методології оцінювання фінансових ризиків VaR є припущення про нормальність розподілу та наявність симетрії у розподілі. На практиці більшість економічних процесів асиметричні, а фінансові ряди мають вироджений хвіст. Саме графік квантилів дає можливість оцінити ступінь довіри для ряду параметричних моделей. Графік квантилів визначається як множина точок [4]:

$$\left\{X_{k,n},\ F^{-1}\left(\frac{n-k+1}{n}\right),\ \ k = 1,...,n\right\}.$$

Якщо параметрична надає прийнятне згладжування, то графік має лінійну форму. Тому графік дає можливість порівняти оцінені моделі та вибрати найкращу; оцінити як обрана модель апроксимує хвіст емпіричного розподілу. Тобто, якщо ряд апроксимується нормальним розподілом і емпіричні дані мають вироджений хвіст, то графік квантилів буде характеризувати криву на вершині правого кінця або на дні лівого кінця розподілу. Крім розглянутих вище видів графічного аналізу існують графік рівня процесу (return level plot) та середня функція ексцесу (mean excess function) [3, 4].

6. Визначення порогу екстремального значення.

Для забезпечення ефективнішого результату наближення екстремальних даних до одного з *GEV*-розподілів застосовують так звані порогові моделі. Нехай множина статистичних даних перевищує деякий поріг u, а $X_1,...,X_n$ – послідовність незалежних однаково розподі-

лених змінних з функцією розподілу F. Тоді умовна ймовірність визначається так:

$$F_u(y) = P(X \le u + y \mid X > u), \text{ або}$$

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)}.$$

Цей вираз дозволяє визначити ступінь наближення значень ймовірності для великих значень порогу u.

Задача вибору оптимального порога ідентична задачі визначення розміру блока. Обидві задачі спрямовані на визначення балансу між відхиленням та дисперсією. Низький рівень призводить до порушень асимптотичної апроксимації, а високий рівень забезпечує велику дисперсію.

Метод вибору порогу базується на основі середнього GPD розподілу. Якщо Y – випадкова змінна у *GPD*-розподілі з параметрами σ і ξ, коли $\xi < 1$, то математичне сподівання $E(Y) = \sigma/(1-\xi)$. В інших випадках середнє є нескінченністю.

Якщо модель є істинною відносно порогу u_0, то вона також істинна для всіх інших порогів u більших за u_0. Тобто для забезпечення високого рівня адекватності побудованої моделі достатньо знайти одне значення порогу, а всі інші припустити проміжними при оцінюванні невідомих параметрів моделі. Середнє для обох випадків визначається так [5]:

$$e(u_0) = E(X - u_0 / X > u_0) = \tilde{\sigma}_{u_0}/(1-\xi),$$

$$e(u) = E(X - u / X > u) = \tilde{\sigma}_u/(1-\xi) = (\tilde{\sigma}_{u_0} + \xi(u - u_0))/(1-\xi). \quad (6)$$

Оскільки $e(u) = E(X - u / X > u)$ – це лінійна функція від u, то враховуючи вираз (6), оцінювання величини порогу можна виконати за такою інструкцією [3, 10]:

1) побудувати графік кривої залишків, що відображають множину точок:

$$\left(u, \sum_{i=1}^{n_u} (x_i - u)/n_u \right), \; u < x_{\max},$$

де n_u – число дослідів, які перевищують u; $x_{\max}$ – верхня межа досліджуваного значення;

2) вибрати порогове значення, над яким графік приймає наближено лінійний характер стосовно u. Застосування довірчих інтервалів допомагає визначити цю точку.

Також, для визначення порогу екстремального значення використовують метод умовно прийнятного вибору, який базується на такому правилі: поріг встановлюється у тому регіоні, де хвіст становить 5–10% від усієї вибірки. Головне припущення: він не повинен бути більшим ніж 10–15%. На практиці 10% межу часто використовували у своїх дослідах Роко (2011), Макнейл і Фрей (2000) [10].

7. Оцінювання невідомих параметрів моделі.

Після кроку визначення порогу потрібно виконати оцінку невідомих параметрів узагальненого розподілу Парето. Як відомо серед методів оцінювання невідомих параметрів моделі поширеним є метод максимальної правдоподібності.

Нехай $y_1, ..., y_k$ – це значення k-залишків з порогу; тоді логарифмічна функція правдоподібності при $\xi \neq 0$:

$$l(\sigma, \xi) = -k \log \sigma - (1 + 1/\xi) \sum_{i=1}^{k} \log(1 + \xi \, y_i / \sigma),$$

коли $(1 + \xi \, y_i / \sigma) > 0$, а для будь-яких інших випадків $l(\sigma, \xi) = -\infty$.

При $\xi = 0$ логарифмічна функція правдоподібності:

$$l(\sigma) = -k \left(\log \sigma - \sigma^{-1} \sum_{i=1}^{k} y_i \right).$$

Другим поширеним методом оцінювання невідомих параметрів є байєсівський підхід. Перевагою байєсівського аналізу при застосуванні до моделей обробки екстремальних значень є його незалежність від регулярності припущень стосовно характеру початкового розподілу, як цього потребує метод максимальної правдоподібності. Практичне застосування байєсівського підходу до оцінювання невідомих параметрів було проілюстровано на прикладі узагальнених лінійних моделей [9, 10].

Крім того, даний підхід надає обґрунтовану альтернативу для випадків, коли припущення, необхідні для застосування методу максимальної правдоподібності та ймовірності зважених моментів не виконуються.

4 ЕКСПЕРИМЕНТИ

Експериментальне дослідження ефективності запропонованої методики виконано за допомогою фактичних статистичних даних. Об'єм статистичної вибірки складав 247 вимірів, які включають такі змінні: назва страхової компанії; грошовий еквівалент страхових виплат; статистичний рік; кількість договорів, які уклала конкретна страхова компанія; страхові платежі; кількість страхових випадків на рік. Основна залежна змінна – страхові виплати, яка відображає здійснення грошових переказів при настанні страхового випадку. Решта змінних, які включені до вибірки, є незалежними і беруться до уваги як фактори.

Для виконання попереднього аналізу статистичних даних та реалізації окремих кроків алгоритму обробки екстремальних значень використовувались такі програмні продукти: Microsoft Excel 2010; інструментальне середовище програмування R2.9.2 для статистичної обробки даних та роботи з графікою; економетричний пакет Eviews 8.0 для побудови моделей та попереднього оцінювання невідомих параметрів. В пакеті Eviews 8.0 використано такі модулі: розрахунок описових статистик, побудова УЛМ, метод максимальної правдоподібності для оцінювання параметрів моделі. В середовищі програмування R2.9.2 виконано інтеграцію модулів Rcmdr, extRemes, evdbayes та mcmcPack.

5 РЕЗУЛЬТАТИ

На рис. 2 відображено графік залежності страхових виплат від статистичного року. Різкі зміни величини «*Страхові виплати*» пояснюється коливаннями величини «Кількість страхових випадків» для відповідного періоду. На рис. 3 відображено значення описових статистик. Із рис. 3 помітно, що коефіцієнт асиметрії (Skewness) коливається в межах 2,839 до 8,664. А це в свою чергу свідчить про наявність «правого хвосту» в розподілі. Так, як параметр ексцессу (Kurtosis) має значення більше трьох, то розподіл є гостровершинним.

Рисунок 2 – Залежність страхових виплат від статистичного року

Також, попередній аналіз початкових даних свідчить про сильну виродженість вибірки, яка проявляється у вигляді шуму при побудові моделі, на прикладі рис. 4. Саме тому прийнято рішення про доречність попереднього логарифмування даних.

Аналіз описової статистики та візуальний аналіз логарифмованих даних (рис. 5) дають можливість припустити про наближення даних до *GEV*- або *GPD*-розподілу.

Відносно високий поріг вибирається з метою того, щоб зменшити зміщення моделі, а з іншої сторони – це буде означати, що лише декілька дослідів використовуються для оцінювання параметрів розподілу, тим самим гарантуючи збільшення оцінки дисперсії. Мета вибору величини порогу полягає в тому, щоб уникнути зміщення моделі. Згідно розглянутого вище методу визначення величини порогу для експерименту прийнято значення 6,65. Графік Mean Residual Life Plot відображає залежність порогу від середнього залишку для оціненої моделі. Він слугує важелем перевірки вибраного порогу. З рис. 6 видно, що після значення порогу 6 з'являються помітні відхилення від лінійності.

View	Proc	Object	Print	Name	Freeze	Sample	Sheet	Stats	Spec
			Q_CASES	Q_ARRANG	DAMAGES	CHARGES			
Mean			248.9717	42819.83	2779.190	47154.60			
Median			65.00000	2431.000	780.0000	13017.40			
Maximum			3800.000	2241084.	49575.00	557884.0			
Minimum			2.000000	6.000000	1.200000	469.8000			
Std. Dev.			526.4061	181679.5	6121.650	81935.75			
Skewness			4.241486	8.663552	4.644020	2.838897			
Kurtosis			24.51794	94.69488	29.05691	12.32584			
Jarque-Bera			5505.864	89621.68	7875.493	1226.855			
Probability			0.000000	0.000000	0.000000	0.000000			
Sum			61496.00	10576499	686459.9	11647185			
Sum Sq. Dev.			68167427	8.12E+12	9.22E+09	1.65E+12			
Observations			247	247	247	247			

Рисунок 3 – Описові статистики початкових даних

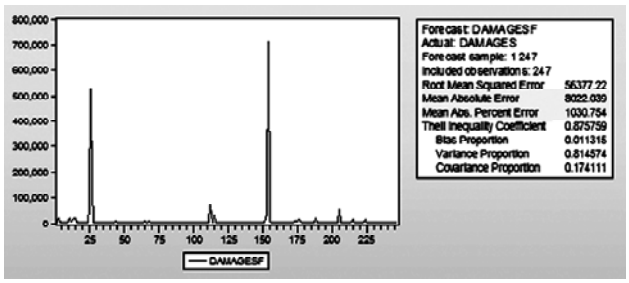


Рисунок 4 – Результати оцінювання моделі без попередньої обробки

Порівняльна характеристика параметрів розподілу представлена в табл. 1. Вона показує, що оптимальним є наближення даних за допомогою *GPD*-розподілу із незначною похибкою та максимальним наближенням емпіричної кривої до теоретичної функції щільності розподілу (рис. 7).

Параметри оцінювання побудованої моделі за допомогою байєсівського підходу зображено на рис. 8. По-

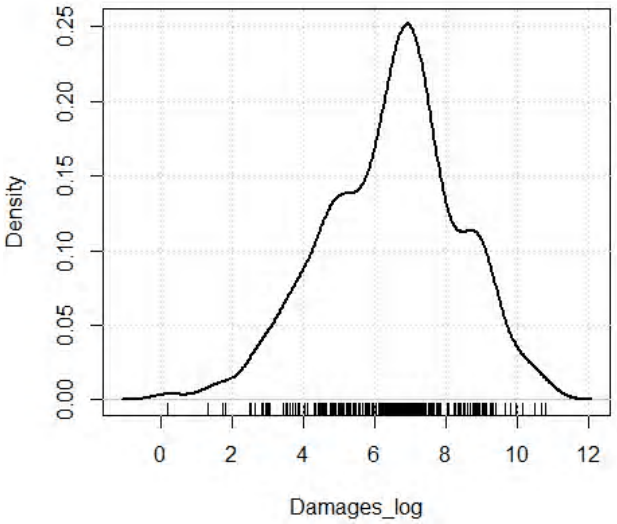


Рисунок 5 – Графік залежності логарифмованих страхових виплат від щільності розподілу

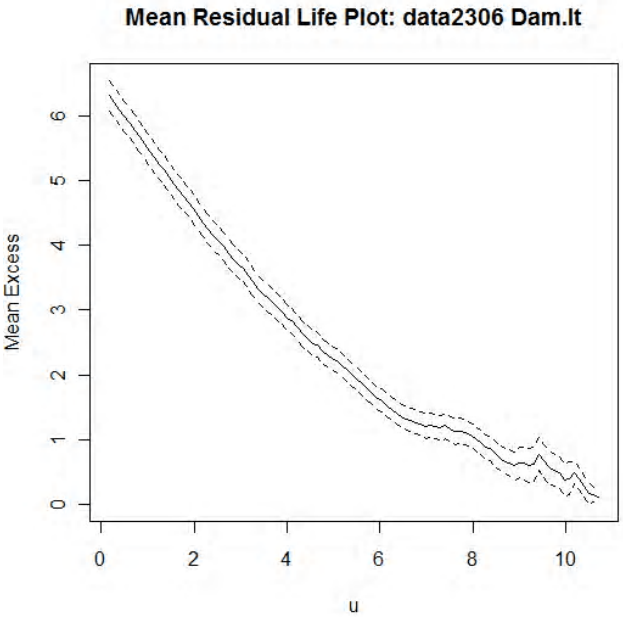


Рисунок 6 – Залежність значення порогу від середнього залишку GPD-моделі

Таблиця 1 – Порівняльна характеристика параметрів розподілів

Тип розподілу	Sigma		Xi		Log-likelihood	Excee-dance rate (per year)	Number of exceedances of threshold
	Maximum likelihood estimation	Std. error	Maximum likelihood estimation	Std. error			
GEV-розподіл	1,953	0,712	–0,650	0,095	487,812	–	–
GPD-розподіл	0,777	0,346	–0,541	0,206	146,369	183,364	124

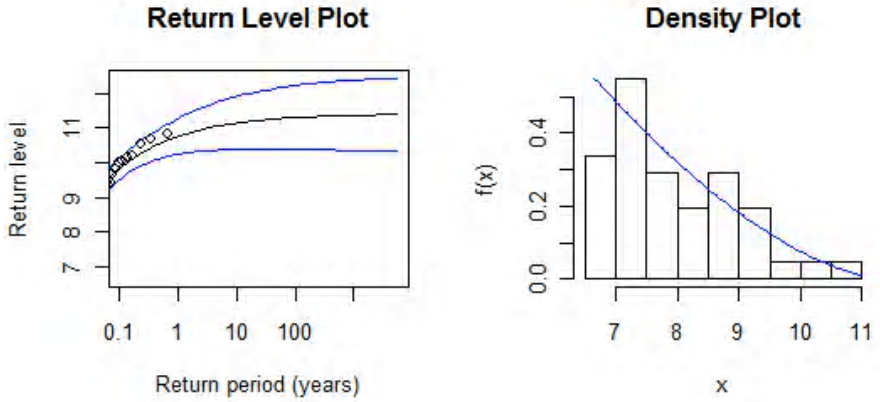


Рисунок 7 – Графічне представлення оціненої GPD-моделі

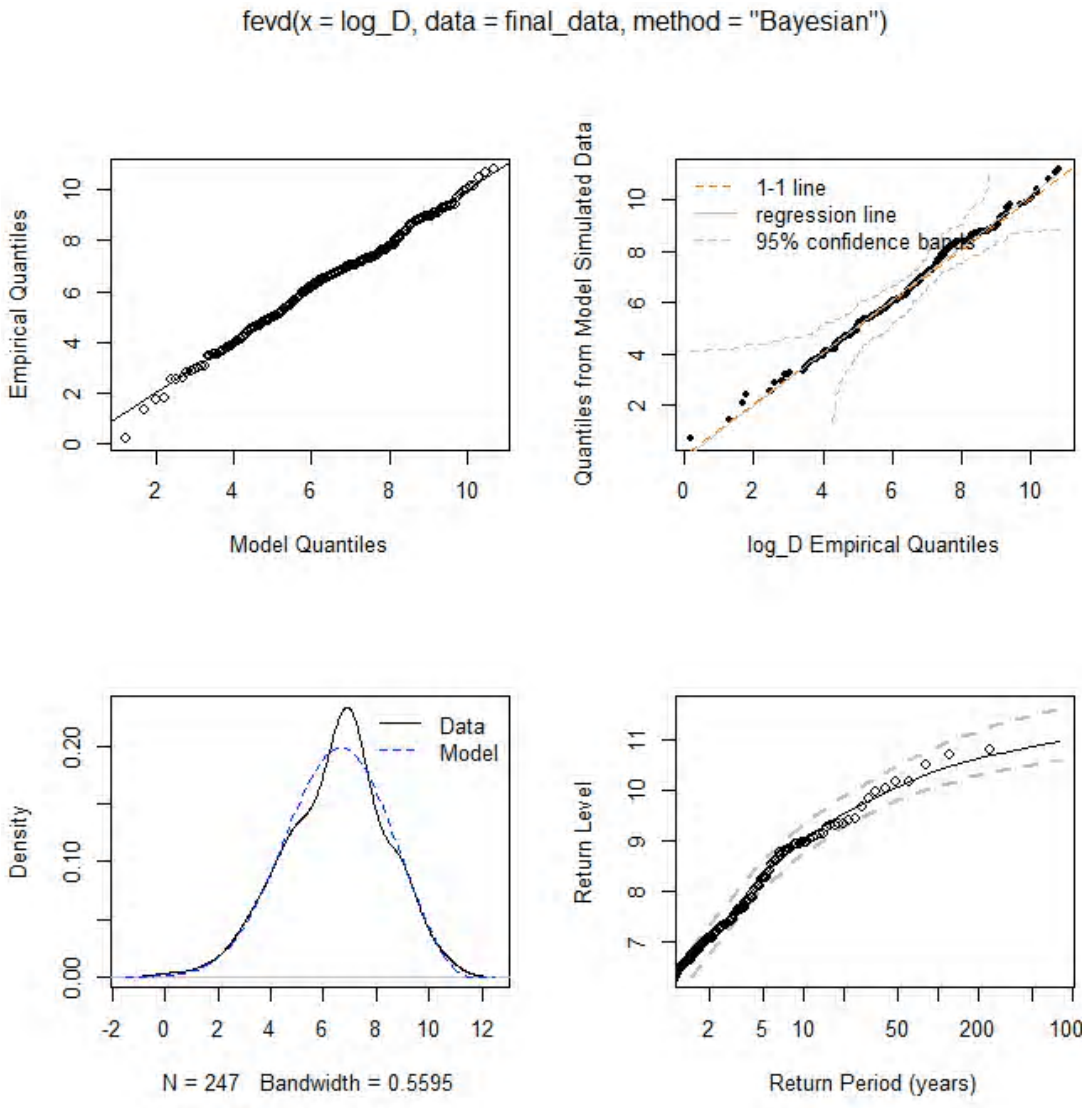


Рисунок 8 – Діагностика наближення моделі до одного з GEV-розподілів

рівнюючи графіки щільності розподілу для побудованої моделі та актуальної вибірки даних помітні значні покращення моделі у термінах належності до одного з GEV-розподілів. Числові значення оцінок невідомих параметрів моделі за допомогою байєсівської методології наведені на рис. 9. Слід зазначити, що на рис. 9 за параметр масштабованості відповідає змінна *scale*, а *shape* – параметр форми. На рис. 10 відображено результати обчислення параметрів апостеріорної вибірки згідно методу Монте-Карло. За допомогою функції «*ci*» було обчислено довірчі інтервали для відповідних параметрів та рівнів повернення (рис. 11). Графічне відображення апріорних оцінок параметрів за методом Монте-Карло та «trace-графіків» наведено на рис. 12.

Порівнюючи результати отриманих оцінок слід зауважити, що байєсівський підхід демонструє кращі результати ніж метод максимальної правдоподібності та сприяє обґрунтованому вибору кращої моделі із запропонованих GEV-розподілів, виходячи зі значень апріорних параметрів, а також алгоритмів вибору кращої моделі.

```
> postmode(fb)
  location      scale      shape
 5.8691196  1.9770899 -0.3610363
```

Рисунок 10 – Результати обчислення параметрів розподілу апріорної вибірки

```
fevd(x = log_D, data = final_data, method = "Bayesian")

[1] "Estimation Method used: Bayesian"


 Acceptance Rates:
log.scale     shape
0.2530506 0.1956391
fevd(x = log_D, data = final_data, method = "Bayesian")

[1] "Quantiles of MCMC Sample from Posterior Distribution"

              2.5% Posterior Mean      97.5%
location  5.5960245      5.8697199  6.1488656
scale     1.7911608      1.9873764  2.2171284
shape    -0.4251922     -0.3529839 -0.2779164
```

Рисунок 9 – Результати оцінювання параметрів моделі за допомогою байєсівської методології

```
> ci(fb)
fevd(x = log_D, data = final_data, method = "Bayesian")

[1] "Quantiles of MCMC Sample from Posterior Distribution"

[1] "Posterior Mean 100-year level: 10.394"

[1] "95% Confidence Interval: (10.0854, 10.8736)"


> ci(fb, type ="parameter")
fevd(x = log_D, data = final_data, method = "Bayesian")

[1] "Quantiles of MCMC Sample from Posterior Distribution"

              2.5% Posterior Mean      97.5%
location  5.5960245      5.8697199  6.1488656
scale     1.7911608      1.9873764  2.2171284
shape    -0.4251922     -0.3529839 -0.2779164
```

Рисунок 11 – Результати обчислення довірчих інтервалів для параметрів форми та масштабованості відповідно та квантилі статистики Монте-Карло для апріорних розподілів

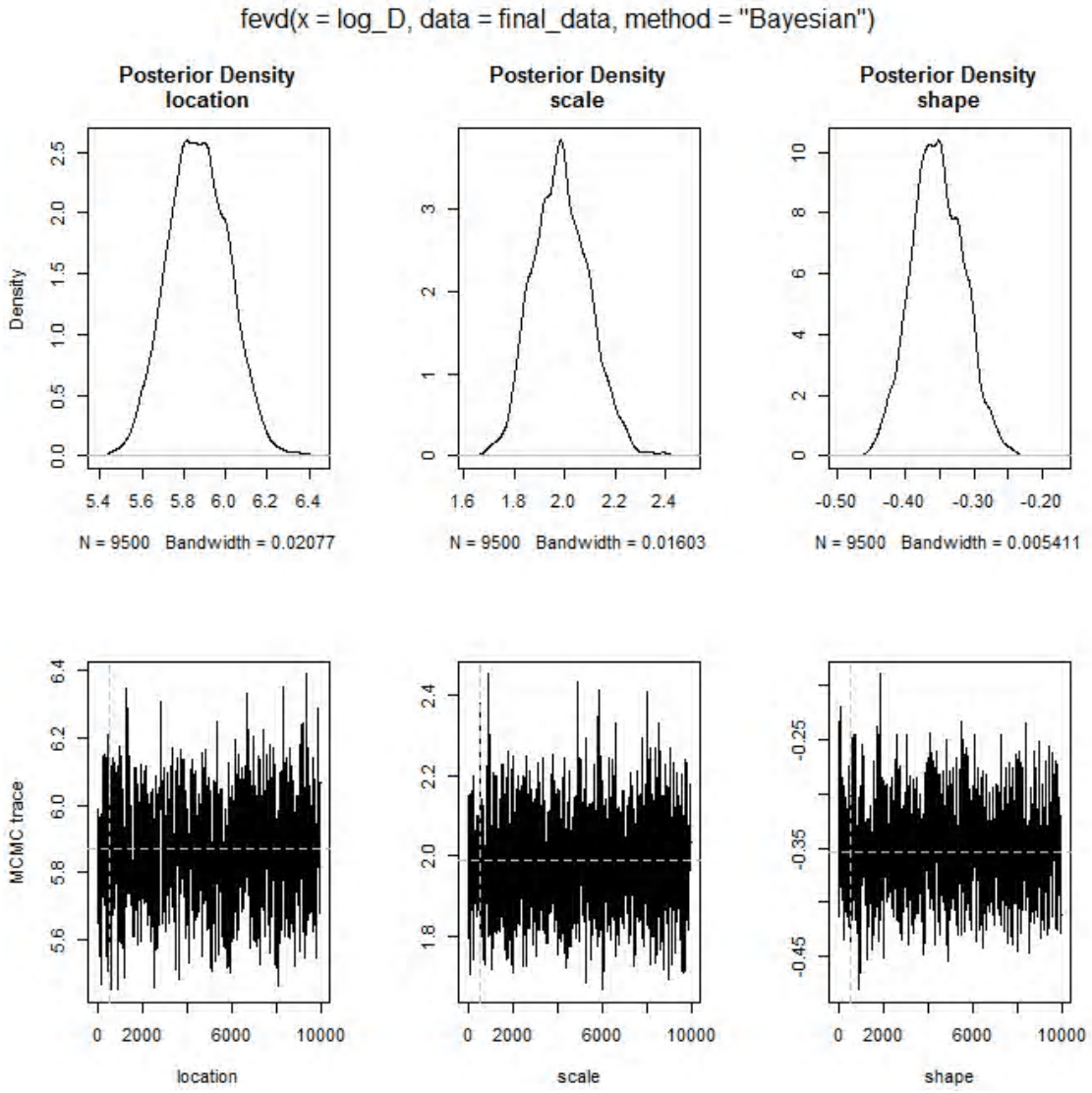


Рисунок 12 – Графічне відображення апріорних оцінок параметрів за методом Монте-Карло та «trace-графіків»

ОБГОВОРЕННЯ

В результаті використання запропонованої комплексної моделі обробки екстремальних статистичних даних вдалося успішно розв'язати проблему невиродженості даних у статистичній вибірці із застосуванням теорії екстремальних значень.

Для оцінювання невідомих параметрів побудованих моделей, які належать до класу GEV-розподілів можна успішно використовувати байєсівський підхід, оперуючи апріорними та апостеріорними розподілами параметрів, а також алгоритмами вибору кращої моделі. Залучення новітніх комбінованих методів до розв'язання задачі обробки екстремальних даних та оцінювання невідомих параметрів, вибору кращої моделі на основі алгоритмів зменшення порогів викидів відкриває нові можливості для дослідження особливостей методів математичного моделювання.

ВИСНОВКИ

Виконано дослідження щодо пошуку ефективної методики обробки екстремальних значень у статистичній вибірці. Запропоновано та експериментально доведено ефективність функціонування створеного багатокрокового підходу із використанням математичного апарату теорії екстремальних значень та методів оцінювання невідомих параметрів моделей. Розглянутий приклад свідчить про те, що запропонований комплексний підхід стосовно обробки екстремальних значень є ефективним та зручним інструментом аналізу вироджених масивів даних та моделювання актуарних процесів. Для оцінювання невідомих параметрів екстремальних моделей зручно використовувати байєсівський підхід, який надає можливість оперувати апріорними та апостеріорними розподілами параметрів і алгоритмами вибору кращої моделі.

Залучення новітніх комбінованих методів до обробки погано структурованих вироджених статистичних даних розкриває нові можливості щодо дослідження особливостей сучасних методик та математичних методів. Надалі необхідно дослідити можливість використання результатів застосування моделей екстремальних значень при побудові прогнозних УЛМ моделей. Застосування запропонованої процедури обробки екстремальних значень гарантує високу точність наближення даних до розподілів та уникнення шуму. Порівняння результатів оцінювання параметрів моделі за допомогою методу максимальної правдоподібності показало, що байєсівські методи оцінювання є кращим підгрунтям для розв'язання задачі вибору кращої моделі на основі множини отриманих альтернатив. Також можна зробити висновок, що сфера страхування, за умови належного менеджменту із застосуванням сучасних матема-

тичних методів обробки даних, оцінювання моделей та прогнозів може бути надійним джерелом стабілізації економіки країни у цілому.

ПОДЯКИ

Роботу виконано відповідно з тематичними планами наукових досліджень Національного технічного університету України «Київський політехнічний інститут». Дослідження виконано в рамках бюджетної НДР, реєстраційний № 0115U000356, тема № 2813–п НТУУ «КПІ»: «Розробка методології системного аналізу, моделювання та оцінювання фінансових ризиків». Страхові дані отримано за сприяння Ліги страхових компаній.

СПИСОК ЛІТЕРАТУРИ

1. Coles S. An Introduction to Statistical Modeling of Extreme Values / S. Coles. – London : Springer-Verlag, 2001. – P. 45–104.
2. Smith R. L. An overview of Extreme value theory / R. L. Smith. – Lausanne : Bernoulli Center, 2009.
3. Mallor F. An introduction to statistical modeling of extreme values / F. Mallor, E. Nualart, E. Omey // Hub research paper. – 2009. – No. 36. – P. 5–31.
4. Shumway R. H. Time series analysis and its applications / R. H. Shumway, D. S. Stoffer. – New York : Springer, 2006. – 598 p.
5. Romano A. Dynamic learning methods / A. Romano, G. Secundo – New York: Springer, 2009. – 190 p.
6. McCullagh P. Generalized Linear Models / P. McCullagh, J. Nelder.– New York : Chapman & Hall, 1989. – 526 p.
7. Tsay R. S. Analysis of financial time series / R. S. Tsay. – New Jersey : John Wiley & Sons, Inc., 2010. – 715 p.
8. Besag J. Markov Chain Monte Carlo for Statistical Inference / J. Besag. – Center for Statistics and the Social Sciences. – 2001. – No. 9. – 25 p.
9. Бідюк П. І. Оцінювання узагальнених лінійних моделей за байєсівським підходом в актуарному моделюванні / П. І. Бідюк, С. В. Трухан // Наукові Вісті НТУУ «КПІ». – 2014. – № 6. – С.49–55.
10. Beirlant J. Statistics of extremes: Theory and application / J. Beirlant. – New York : John Wiley & Sons, Inc., 2004. – 505 p.

Стаття надійшла до редакції 21.10.2015.
Після доробки 28.10.2015.

Трухан С. В.[1], Бидюк П. И.[2]
[1]Аспирантка института прикладного системного анализа НТУУ «КПИ», Киев, Украина
[2]Д-р техн. наук, профессор кафедры математических методов системного анализа НТУУ «КПИ», Киев, Украина

МЕТОДИКА АНАЛИЗА ЭКСТРЕМАЛЬНЫХ ДАННЫХ И ЕЕ ИСПОЛЬЗОВАНИЕ ПРИ ОЦЕНИВАНИИ ПАРАМЕТРОВ ОБОЩЕННЫХ ЛИНЕЙНЫХ МОДЕЛЕЙ

Предложена методика анализа экстремальных значений с целью ее использования при оценивании неизвестных параметров обобщенных линейных моделей. В качестве математического аппарата использована теория экстремальных значений, которая является одним разделов математической статистики и связана с исследованием отклонений экстремальных значений от медианы в вероятностных распределениях. Также рассмотрены методы приближения экстремальных данных к классу обобщенных экстремальных распределений, методы оценивания неизвестных параметров и выбора оптимального порога для экстремальных значений. На основе реальных статистических данных и исследуемого подхода построены модели обработки экстремальных значений для дальнейшего использования при оценивании прогнозных моделей. Допустимой для дальнейшего применения оказалась модель приближения данных с помощью обобщенного распределения Парето. Это обосновывается минимальной величиной погрешности, а также максимальным приближением эмпирической кривой к теоретической функции плотности распределения. Сравнение результатов оценивания неизвестных параметров модели с помощью метода максимального правдоподобия и байесовского подхода показало, что байесовские методы оценивания являются эффективным основанием для решения задачи выбора лучшей модели исходя из множества полученных альтернатив и значений априорных параметров. Для дальнейшего исследования целесообразно рассмотреть задачу применения моделей экстремальных значений при построении прогнозных обобщенных линейных моделей.

Ключевые слова: теория экстремальных значений, обобщенные линейные модели, порог экстремального значения, метод максимального правдоподобия, байесовский подход.

Trukhan S.[1], Bidyuk P.[2]
[1]Post-graduate student of Institute for Applied System Analysis, NTUU «KPI», Kyiv, Ukraine
[2]Dr. Sc., Professor at the Department of Mathematical methods for System Analysis, NTUU «KPI», Kyiv, Ukraine

METHODOLOGY OF EXTREME VALUES ANALYSIS AND ITS APPLICATION FOR PARAMETER ESTIMATION OF GENERALIZED LINEAR MODELS

The article deals with methodology of extreme values treatment for building and estimating unknown parameters of generalized linear models. As a mathematical tool for carrying out the research the extreme value theory was used that creates one of the directions in mathematical statistics, and is related to investigating the extreme deviations from the median values in probability distributions. Also, the methods of approximation statistical data to generalized extreme value distribution, the methods of estimating unknown parameters and selecting an optimal threshold for extreme value models are discussed. The models of treatment extreme values are constructed which are based on actual statistical data and approach is proposed for their future application for estimating predictive models. The model with generalized Pareto distribution turned out to be acceptable for further use, because it has minimum value of observation error and the best approximation of observed curve to theoretical density function. The comparison of evaluation unknown models' parameters using method of maximum likelihood and Bayesian approach leads to next conclusion. The Bayesian methods are efficient way to solve the problem of selection the best model, based on the received alternatives set and prior parameters values. In future studies it will be reasonable to consider the application of extreme value analysis to predicted generalized linear models.

Keywords: extreme value theory, generalized linear models, extreme value threshold, maximum likelihood method, Bayesian approach.

REFERENCES

1. Coles S. An Introduction to Statistical Modeling of Extreme Values. London, Springer-Verlag, 2001, pp. 45–104.
2. Smith R. L. An overview of Extreme value theory. Lausanne, Bernoulli Center, 2009.
3. Mallor F., Omey E. An introduction to statistical modeling of extreme values. Hub research paper, 2009, No. 36, pp. 5–31.
4. Shumway R. H., Stoffer D. S. Time series analysis and its applications. New York, Springer, 2006, 598 p.
5. Romano A., Secundo G. Dynamic learning methods. New York, Springer, 2009, 190 p.
6. McCullagh P., Nelder J. Generalized Linear Models. New York, Chapman & Hall, 1989, 526 p.
7. Tsay R. S. Analysis of financial time series. New Jersey, John Wiley & Sons, Inc., 2010, 715 p.
8. Besag J. Markov Chain Monte Carlo for Statistical Inference, Center *for Statistics and the Social Sciences*, 2001, No. 9, 25 p.
9. Bidyuk, P., Trukhan, S. Estimation of generalized linear models using Bayesian approach in actuarial modeling, *Naukovi Visti NTUU «KPI»*, 2014, No. 6, pp. 49–55.
10. Beirlant J. Statistics of extremes: Theory and application. New York, John Wiley & Sons, Inc., 2004, 505 p.

НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

НЕЙРОИНФОРМАТИКА И ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

NEUROINFORMATICS AND INTELLIGENT SYSTEMS

УДК 519.71

Кучеренко Е. И.[1], Глушенкова И. С.[2], Глушенков С. А.[3]

[1]Д-р техн. наук, профессор, профессор кафедры искусственного интеллекта Харьковского национального университета радиоэлектроники, Харьков, Украина

[2]Канд. техн. наук, доцент кафедры геоинформационных систем, оценки земли и недвижимого имущества Харьковского национального университета городского хозяйства имени А.Н. Бекетова, Харьков, Украина

[3]Аспирант кафедры геоинформационных систем, оценки земли и недвижимого имущества Харьковского национального университета городского хозяйства имени А.Н. Бекетова, Харьков, Украина

НЕЧЕТКОЕ РАЗБИЕНИЕ ОБЪЕКТОВ НА ОСНОВЕ КРИТЕРИЕВ ПЛОТНОСТИ

Решена задача разбиения по критериям плотности в нечетком пространстве состояний при пересечении признаков. Объектом исследования являлись процессы разбиения заданной выборки объектов на подмножества. Предмет исследования составляют методы и алгоритмы нечеткого разбиения объектов на основе критериев плотности в сложных системах. Цель работы: развитие метода горной кластеризации Ягера-Филева на основе нечетких представлений для повышения эффективности решений. Предложен нечеткий метод разбиения, основанный на вычислении плотности распределения интегральных признаков объектов в нечетком пространстве состояний, который, в отличие от существующих, дополнительно функционирует в нечетком пространстве состояний и признаков. Описаны и обоснованы этапы метода нечеткого разбиения признаков с использованием нечеткого расстояния Хемминга. Была создана программа моделирования плотности распределения признаков на основе разработанного метода. Выполнен эксперимент по определению принадлежности объекта при пересечении областей нечеткого распределения признаков и предоставление результатов в виде логического вывода и графического материала. Результаты эксперимента позволяют рекомендовать предложенный метод для использования на практике. Перспективой дальнейших исследований является исследование и алгоритмизация метода, его адаптация в пространстве признаков предметных областей.

Ключевые слова: кластеризация, расстояние Хемминга, горная кластеризация, нечеткая логика.

НОМЕНКЛАТУРА

ГИС – геоинформационные системы;

C_i – размер множества разбиения;

D – детерминированное состояние;

$\{d_{ij}\}$ – множество расстояний;

E – ребро графа;

$\widetilde{F}$ – нечеткое состояние;

G^t – граф;

Int – интегральное распределение признаков;

$\{K_i\}$ – множество разбиений;

m – количество уровней иерархии;

$\{O_{ij}\}$ – множество объектов;

$\{O_\alpha, O_\beta\}$ – объекты локализованного пространства;

$P(x, y)$ – вероятностное распределение плотностей;

$\{p_{ij\alpha}\}$ – параметры множества;

R_i – радиус множества разбиения;

$S_{ij\beta}$ – площадь распределения плотности;

U – матрица распределения плотностей признаков по объектам;

Z_i – центры множества разбиения;

$\{a_{ij}\}$ – пространственные координаты;

$\Delta\mu(S_{ij\beta})$ – величина дискретизации;

$(\delta(\widetilde{O}_\alpha), \delta(\widetilde{O}_\beta))$ – относительное расстояние Хемминга;

δ_i – расстояние Хемминга;

$\eta(\widetilde{O})$ – квадратичный индекс нечеткости;

$\mu_{o_\alpha}, \mu_{o\beta}$ – множество значений функции принадлежности;

$\mu_{\widetilde{A}_i}$ – функция принадлежности;

$\mu(S_{ij\beta})$ – функция принадлежности плотности распределения;

DOI 10.15588/1607-3274-2016-1-4

ρ_α – плотность распределения признаков;

$\{\rho_{ij\beta}\}$ – множество распределения плотности признаков.

ВВЕДЕНИЕ

Важным аспектом классификации объектов является представление, структурирование и анализ огромных массивов информации, которые составляют основу функционирования и развития сложных систем. При анализе многомерных распределенных объектов требуются универсальные и надежные подходы, направленные на минимизацию критериев на множестве ограничений предметной области. Особенно это актуально при реализации геоинформационных систем (ГИС) и технологий.

Проблема принятия решений в таких системах является не тривиальной задачей [1] и характеризуется неопределенностью, которая может быть снижена за счет применения нечетких (fuzzy) знание-ориентированных технологий. Существующие решения [2] на основе построения кластеров являются актуальными в ГИС и технологиях. Однако наличие свойств пересечения кластеров часто приводит к трудностям классификации объектов производственных систем, что приводит к технологиям нечеткого разбиения пространства по заданным критериям.

Целью настоящих исследований является разработка и совершенствование подходов к оптимизации и классификации таких объектов на основе развития методов и алгоритмов нечеткой кластеризации и разбиения, а также систем на их основе.

1 ПОСТАНОВКА ЗАДАЧИ

Кластерный анализ – это задача разбиения заданной выборки объектов на подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались [1]. Кластерный анализ представляет собой многомерную статистическую процедуру, выполняющую сбор данных, содержащих информацию о выборке объектов, и упорядочивающую объекты в сравнительно однородные группы нечетких разбиений. Множество разбиений $\{K_i\}, i \in I$ характеризуется:

– центрами $Z_i, i \in I$;

– размером $C_i - \{K_i\}, i \in I \,|\, C_i \neq \varnothing$;

– радиусами $R_i - \{K_i\}, i \in I \,|\, R_i \neq \varnothing$;

– множеством объектов $\{O_{ij}\}, j \in J$;

– множеством расстояний между кластерами $\{d_{ij}\}$, причем:

$$d(\widetilde{A}, \widetilde{A}) = 0;$$
$$d(\widetilde{A}, \widetilde{B}) = d(\widetilde{B}, \widetilde{A});$$
$$d(\widetilde{A}, \widetilde{C}) \le d(\widetilde{A}, \widetilde{B}) \circ d(\widetilde{B}, \widetilde{C});$$
$$d(\widetilde{A}, \widetilde{B}) \ge 0.$$

Множество объектов $\{O_{ij}\}, j \in J$ характеризуется множествами признаков распределения плотностей $\{\rho_{ij\beta}\}, \beta \in \mathrm{B}$, пространственных координат $\{a_{ij}\}$ и параметров $\{p_{ij\alpha}\}, \alpha \in \mathrm{A}$, причем $O_{ij} \in \{O_{ij}\},\ j \in J \,|\, \{a_{ij}\} \neq \varnothing, \{p_{ij}\} \neq \varnothing$.

Необходимо предложить методы и алгоритмы нечеткого разбиения объектов на основе критериев плотности в сложных системах и технологиях пространственно распределенных объектов предметных областей.

2 ОБЗОР ЛИТЕРАТУРЫ

Применение кластерного анализа в общем виде сводится к следующим этапам [3]:

– выборка объектов для кластеризации;

– определение множества переменных, по которым будут оцениваться объекты в выборке с нормализацией значений переменных;

– вычисление значений меры сходства между объектами;

– применение метода кластерного анализа для создания групп сходных объектов (кластеров);

– представление и интерпретация результатов анализа.

Для каждой пары объектов измеряется «расстояние» между ними – степень схожести. Существует множество метрик [3]: эвклидово расстояние; квадрат эвклидова расстояния; манхэттенское расстояние или расстояние Хемминга; расстояние Чебышева; степенное расстояние, которое является модификацией расстояния Эвклида.

Существуют следующие методы и алгоритмы кластеризации [3] (табл. 1):

– алгоритмы иерархической кластеризации;

– алгоритмы квадратичной ошибки;

– нечеткие алгоритмы;

– алгоритмы, основанные на теории графов;

Таблица 1 – Методы и алгоритмы кластеризации

Методы и алгоритмы кластеризации	Положительные качества	Отрицательные качества
алгоритмы иерархической кластеризации	представление результата в виде дендрограммы	необходима система полных разбиений
алгоритмы квадратичной ошибки	минимизация среднеквадратической ошибки разбиения	требуется задание количества кластеров
нечеткие алгоритмы	мягкое разбиение на кластеры	необходимо заранее знать количество кластеров
алгоритмы, основанные на теории графов	наглядность и возможность внесения различных модификаций	сложность подбора значащих коэффициентов
алгоритм выделения связных компонент	наглядность и возможность внесения различных модификаций	трудности управления количеством кластеров при помощи порога расстояния
алгоритм минимального покрывающего дерева	наглядность и возможность внесения различных модификаций	трудности управления количеством кластеров при помощи порога расстояния
послойная кластеризация	наглядность и возможность внесения различных модификаций	трудности управления количеством кластеров при помощи порога расстояния

– алгоритм выделения связных компонент;
– алгоритм минимального покрывающего дерева;
– послойная кластеризация.

Наличие множества методов и алгоритмов кластеризации не охватывает всей совокупности подходов и особенностей распределения признаков объектов, что свойственно распределению различной природы и требует дальнейших исследований.

3 МАТЕРИАЛЫ И МЕТОДЫ

Предлагаемые решения являются дальнейшим развитием метода горной кластеризации Ягера-Филева [4] на основе нечетких представлений.

Утверждение 1. Если задано множество площадей распределения плотностей признаков $\{\rho_{ij\beta}\}, \beta \in \mathrm{B}$, пространственных координат $\{a_{ij}\}$ и параметров $\{p_{ij\alpha}\}, \alpha \in \mathrm{A}$, а также задана плотность распределения признаков $\{\rho_{ij\beta}\}, \beta \in \mathrm{B}$ в виде $\rho_\alpha, \alpha \in \mathrm{A}$, для которых характерно $S_{ij}, i \in I, j \in J$, тогда критерием поиска плотности распределения признаков может быть

$$\rho_\alpha = \frac{\sum\limits_{i\in I,\ j\in J} \rho_{ij\beta} S_{ij}}{\sum\limits_{i\in I,\ j\in J} S_{ij}} \mu(S_{ij\beta}),\ \rho_{ij} > \rho^*, \qquad (1)$$

где ρ^* – порог плотности, $\mu(S_{ij\beta})$ – функция принадлежности плотности распределения (ρ_{ij}) по некоторой площади из $S_{ij\beta}$.

Особенностью метода является то, что он не требует задания количества кластеров и при числе признаков $(n = 2)$ поверхность распределения близка к горному рельефу. Кластеризация по горному методу не является нечеткой, однако, ее часто используют при синтезе нечетких правил на знаниях [4]. Особенностью горной кластеризации является следующее:

– на первом шаге горной кластеризации определяют точки, которые могут быть центрами кластеров;

– на втором шаге для каждой такой точки рассчитывается значение потенциала, показывающего возможность формирования кластера в ее окрестности. Плотность расположения объектов в окрестности потенциального центра кластера является функцией от значения его потенциала;

– на третьем шаге итерационно выбираются центры кластеров среди точек с максимальными потенциалами. Формируются также объекты кластеризации.

Важной особенностью, накладывающей ограничения на применение метода, является отсутствие возможности решений в нечетком пространстве состояний. Анализ возможных решений по совершенствованию вышеизложенного метода [4] позволил сформулировать следующее предположение: на первом этапе построение решетчатого гиперкуба следует дополнить функцией принадлежности распределения $\mu(S_{ij\beta})$.

Тогда этапы нового метода нечеткого разбиения признаков могут быть такие.

Этап 1. Формируем интегральное распределение признаков

$$Int = \frac{\sum\limits_{i\in I, j\in J, \beta\in \mathrm{B}} \rho_{ij\beta} S_{ij\beta}}{\sum\limits_{i\in I, j\in J, \beta\in \mathrm{B}} S_{ij\beta}}, \qquad (2)$$

исключаем из рассмотрения $\rho_{ij\beta}$, для которых $|I| \wedge |J| \wedge |\mathrm{B}| = 0$, и уточняем интегральные признаки (2)

$$Int' = \frac{\sum\limits_{i\in I, j\in J, \beta\in \mathrm{B}} \rho_{ij\beta} S_{ij\beta}}{\sum\limits_{i\in I, j\in J, \beta\in \mathrm{B}} S_{ij\beta}}. \qquad (3)$$

Этап 2. Формируем область $S_{ij\beta_x}$, для которой значение квадратичного индекса нечеткости [5] принимает вид

$$\eta(\widetilde{O}) = \frac{2}{\sqrt{n}} e(\widetilde{O}, \widetilde{\overline{O}}) \xrightarrow{a_{ij}} \min. \qquad (4)$$

Искомое значение (4) определяем в качестве первого с максимальной плотностью признаков потенциального центра разбиения $\{K_i\}, i = 1 | C_i \neq \varnothing$.

Этап 3. Формируем потенциальные центры разбиения. Для этого диапазоны изменения входных признаков разбиваем на n интервалов согласно (1), причем принимаем

$$\mu(S_{ij\beta}) = \mu(S_{i+1,\ j+1,\beta.}), \mu(S_{ij\beta)} = \exp(-k(x-u)^2), k > 0, \qquad (5)$$

где k – крутизна функции, u – центр гауссиана.

Параметры функции (5) являются элементами настройки.

Этап 4. Задав в (5) $\mu(S_{ij\beta}) = \mu(S_{ij\beta})_o$ и величину дискретизации $\Delta\mu(S_{ij\beta})$, находим итерационно, согласно (1), значения $x = x_o, ..., x_n$, таким образом, что

$$\rho_{ij} = \rho_o > \rho^*. \qquad (6)$$

Это определяет число и размер потенциальных пространств разбиения $\Delta S\alpha_{ij}$.

Этап 5. Уточняем из особенностей поверхности $\{S_{ij\beta}\}$ проекции значений $\{x_o,..,x_n\}$, $|x_o,..,x_n| = N$, что и определяет размер разбиения

$$C_i = x_{i+1-} x_i \qquad (7)$$

и радиус

$$R_i = \frac{x_{i+1} - x_i}{2} \qquad (8)$$

объектов.

Этап 6. Радиусы объектов уточняются на основе квадратичной нормы с учетом (6) и повторной реализацией этапов 1–5. Отметим, что $R_i \neq R_{i+1,\ C_i} \neq C_{i+1}$.

Этап 7. Осуществляем упорядочение признаков и формирование матрицы распределения плотностей признаков по объектам

$$U = \{K_i\}, i \in I\ |\ (C_i \neq \varnothing, O_{ij}, i \in I, j \in J),$$

$$if\ O_{ij} \in \{O_{ij}\}, j \in J\ |\ \{a_{ij}\} \neq \varnothing, \{p_{ij}\} \neq \varnothing. \qquad (9)$$

Этап 8. Останов.

Для реализации этапов 1–7 в знание-ориентированных технологиях целесообразно применить стратегию согласно [6], как решение

$$\{if\ x\ is\ \mu(x)\ then\ y\ is\ \mu(y)\}, \tag{10}$$

согласно

$$y' = \vee x' \wedge \mu(x,y) \tag{11}$$

с последующей (11) дефаззификацией [6].

Пусть существуют объекты $\{O_{\alpha,}\ O_{\beta}\} \subset \{O_{ij}\}, j \in J$ множества значений функций принадлежности, для которых

$$\mu_{o_\alpha} \cap \mu_{o_\beta} \neq \varnothing. \tag{12}$$

Существует некоторое пространство $S(x,y)$, которое принадлежит (12). Необходимо определить принадлежность пространства $S(x,y)$ одному из объектов из $\{O_{\alpha,}\ O_{\beta}\}$, если $\rho_{S(x,y)} \geq \rho^*$.

Утверждение 2. Если справедливо (12) и необходимо определить принадлежность пространства $S(x,y)$ одному из объектов из $\{O_{\alpha,}\ O_{\beta}\}$, при $\rho_{S(x,y)} \geq \rho^*$, то пространство $S(x,y)$ принадлежит одному из объектов $O_k \in \{O_{\alpha,}\ O_{\beta}\}$, согласно

$$(\delta(\widetilde{O}_\alpha), \delta(\widetilde{O}_\beta)) \xrightarrow{S(x,y),\ \rho_{S(x,y)} \geq \rho^*} \min. \tag{13}$$

Доказательство утверждения 2 очевидно, если принять в качестве критерия меру (13). В качестве критерия локализации пространства $S(x,y)$ на объектах $\{O_\alpha, O_\beta\}$ введем меру на основе нечеткого расстояния Хемминга [7]

$$\delta(\widetilde{O}_k, \widetilde{S}(x,y)) = \frac{d(\widetilde{O}_k, \widetilde{S}(x,y))}{n}, \tag{14}$$

которая справедлива при условии, что плоскость дискретизации функций принадлежности объектов в (14) определена в виде

$$n_{O_\alpha} = n_{O_\beta} = n. \tag{15}$$

Тогда, используя (14), (15), мы можем выявить свойство принадлежности (13) пространства $S(x,y)$ к объектам из $\{O_{\alpha,}\ O_{\beta}\}$.

Следствие утверждения 2. Если существует мера расстояния в виде расстояния Хемминга [14] причем

$$\{\mu_{\widetilde{A}_i}\} \cap \{\mu_{A_{i+1}}\} \neq \varnothing, \tag{16}$$

то следует рассматривать ряд случаев распределения плотностей.

Случай А. Пусть существует вероятностное распределение плотностей вещества на плоской поверхности

$$P(x,y) = e^{-k(x-a)^2}, \tag{17}$$

где a – центр гауссиана.

Определим пространство распределения плотностей на основе (17), где предложено формирование первого разбиения $S_{ij\beta_x}$ с центром разбиения

$$K_i \in \{K_i\}, i = 1 | C_i \neq \varnothing. \tag{18}$$

Тогда справедливо утверждение.

Утверждение 3. Если задана карта распределения с центрами a_1, a_2, находящимися в изолированном пространстве, и подвергающимися динамическим внешним факторам, то правило матричного отображения пространства имеет вид (19).

Приняв в (16) $a_1 = a_2$, уточнив (17), выполним анализ расположения точки $a(x,y) \in \widetilde{S}(x,y)$ согласно критерия

$$a_i \cap P(x,y) \neq \varnothing, \tag{19}$$

При нарушении (19), имеем возможное несоответствие согласно утверждения 3, что требует дальнейших исследований.

Рассмотрим случай, когда справедливо (18), (19).

Случай В. Приняв терм лингвистической переменной в виде $\mu_{A_1} \neq 0,\ \mu_{A2} \neq 0$, функции принадлежности определены в виде гауссианов:

$$\mu_{A_1} = e^{-k_1(x-a_1)^2},\ k_1 > 0, \tag{20}$$

$$\mu_{A_2} = e^{-k_2(x-a_2)^2},\ k_2 > 0. \tag{21}$$

Сформулируем утверждение 4.

Утверждение 4. Если существует (14), для которого справедливо (16), то нахождение минимального значения из

$$\alpha = \min(\delta(\mu_{\widetilde{A}_i}), \delta(\mu_{A_{i+1}})), i \in I \tag{22}$$

определяет принадлежность области $\widetilde{S}(x,y)$ соответствующему нечеткому разбиению при выполнении (16).

Справедливость утверждения 4 непосредственно следует из меры расстояния (14) и сущности операции пересечения функций принадлежности (16).

Используя положения (14), рассмотрим пересечение областей (20), (21), причем $k_1 > 0, k_2 > 0, k_1 < k_2, a_1 > a_2$, для которых справедливо (16).

Вычислениями определено, что $\delta_{A_1} < \delta_{A_2}$, тогда область $\widetilde{S}(x,y)$ принадлежит, согласно (22), разбиению (20).

Действительно, пространство $\widetilde{S}(x,y)$ попадает под влияние области с меньшим расстоянием Хемминга, что подтверждает справедливость (22) на функциях (20), (21).

Следствие 1 утверждения 4. В случае, если расстояния равны $\delta(14)$

$$\left| \{\mu_{\widetilde{A}_i}\} = \{\mu_{A_{i+1}}\} \right| \leq \varepsilon, \tag{23}$$

где ε – норма точности, то пространство $\widetilde{S}(x,y)$ не принадлежит ни к одному из источников.

Следствие 2 утверждения 4. Положения утверждения 4 справедливы при выполнении условия (6).

Тогда, учитывая положения утверждений 2–4, дополнительно к этапам 1–8, сформулируем дополнения:

5′ Если справедливо (12), то осуществляем уточнение нахождения дополнительных кластеров согласно (13)–(22);

5″ Осуществляем контроль разбиения согласно этапов 1–7.

4 ЭКСПЕРИМЕНТЫ

Для эксперимента была создана программа моделирования плотности распределения признаков на основе разработанного метода при условии $\{\mu_{\tilde{A}_i}\} \cap \{\mu_{A_{i+1}}\} \neq \varnothing$.

UML – диаграмма классов представлена на рис. 1.

Программа PL.v.0.1, которая реализована в среде Python [8].

5 РЕЗУЛЬТАТЫ

В качестве исходных данных используются параметры, определяющие вид функции и местоположение в пространстве (рис. 2):

– координаты проекций источников признаков влияния и искомой точки на ось OX;

– значения крутизны функций принадлежности распределения плотности признаков от соответствующих источников;

– значение ограничителя в расчетах (ПДК).

ПДК – минимальное значение плотности признаков для изменения состояния искомой точки. ПДК выполняет роль логического оператора, который определяет дальнейший сценарий работы программы. Возможны 3 сценария работы программы. Сценарий 1: ПДК больше значений множеств в искомой точке, тогда точка не изменяет свое состояние и не принадлежит не к одной точке. Сценарий 2: ПДК меньше одного из значения множеств. Сценарий 3: ПДК меньше значений двух функций принадлежности в искомой точке, тогда для определения множества, к которому принадлежит точка, выполняется расчет расстояния Хемминга.

Во время работы программы выполняется расчет принадлежности точки к зонам влияния исходных множеств. В случае влияния обоих множеств на искомую точку выполняется расчет расстояния Хемминга и формирования результата путем логического вывода. Помимо выдачи текстового результата работы, программа выводит изображение с наглядным отображением всех объектов.

Таким образом, в результате эксперимента подтверждена справедливость утверждения 4 (рис. 2).

6 ОБСУЖДЕНИЕ

В работе рассмотрены методы нечеткого разбиения признаков по критерию плотности. Сформулирован метод, который является развитием метода горной кластеризации Ягера-Филева, что позволяет развивать метод на случай нечеткого пространства состояний. Дана оценка эффективности методов, основанных на вероятностном и нечетком распределении плотности признаков. Определена перспективность нечеткого разбиения по отношению к вероятностному распределению.

Действительно, используя положения (14) рассмотрено пересечение областей нечеткого распределения признаков. Вычислениями определено, что в случае, когда $\delta_{A_1} < \delta_{A_2}$, область $\tilde{S}(x, y)$ принадлежит, согласно (22), разбиению (20).

Определено, что справедливо следствие утверждения 4 в случае, если расстояния равны (23).

Экспериментом подтверждена вычислительная сложность в виде полинома второго порядка

$$O(n) = A_0 + bx + cx^2, \tag{24}$$

для (24) справедливо

$$A_0 \neq 0, b \neq 0, c \neq 0.$$

Развитием метода является адаптация подходов к предметным областям путем дополнительного введения влияния различных внешних факторов на процесс нечеткого распределения пространства признаков.

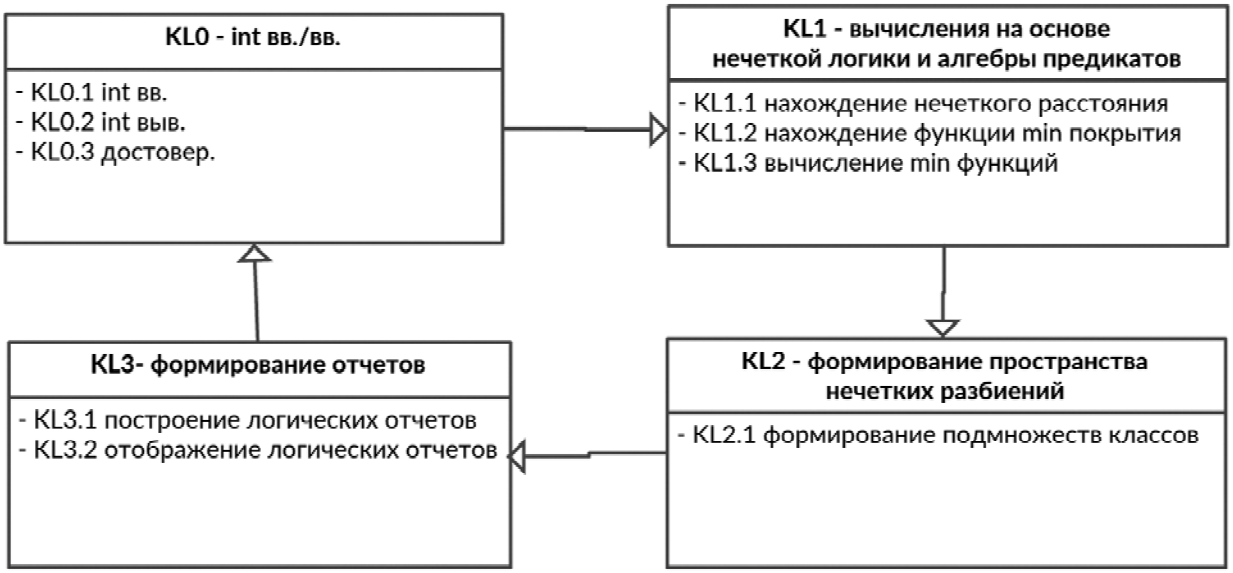
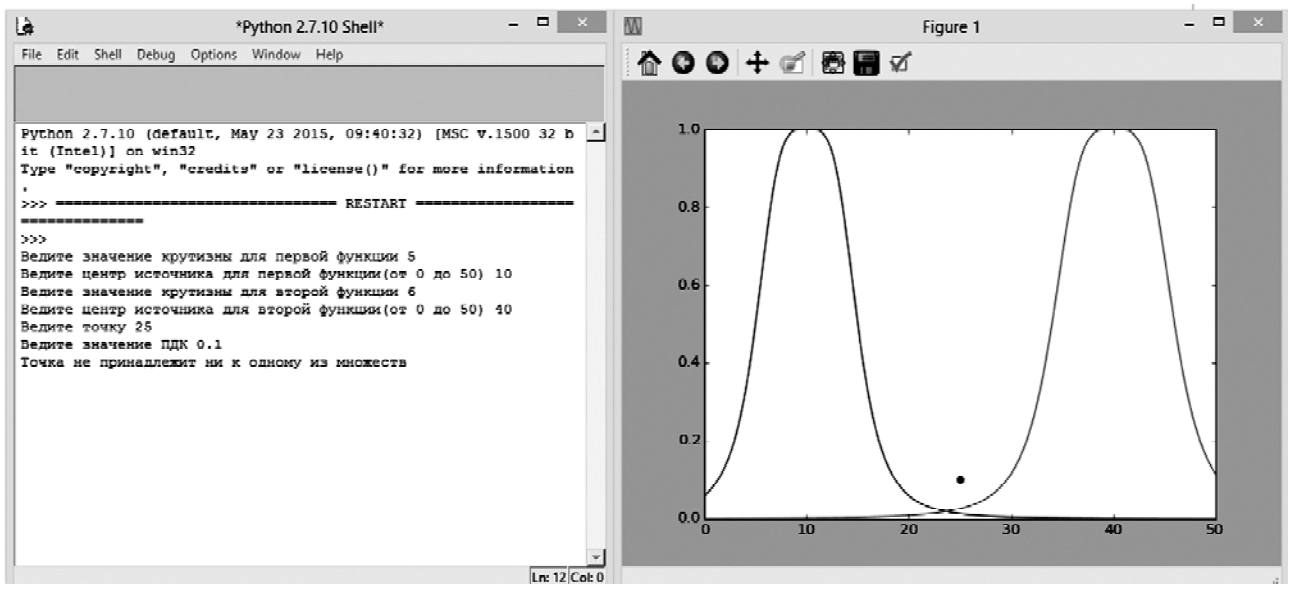
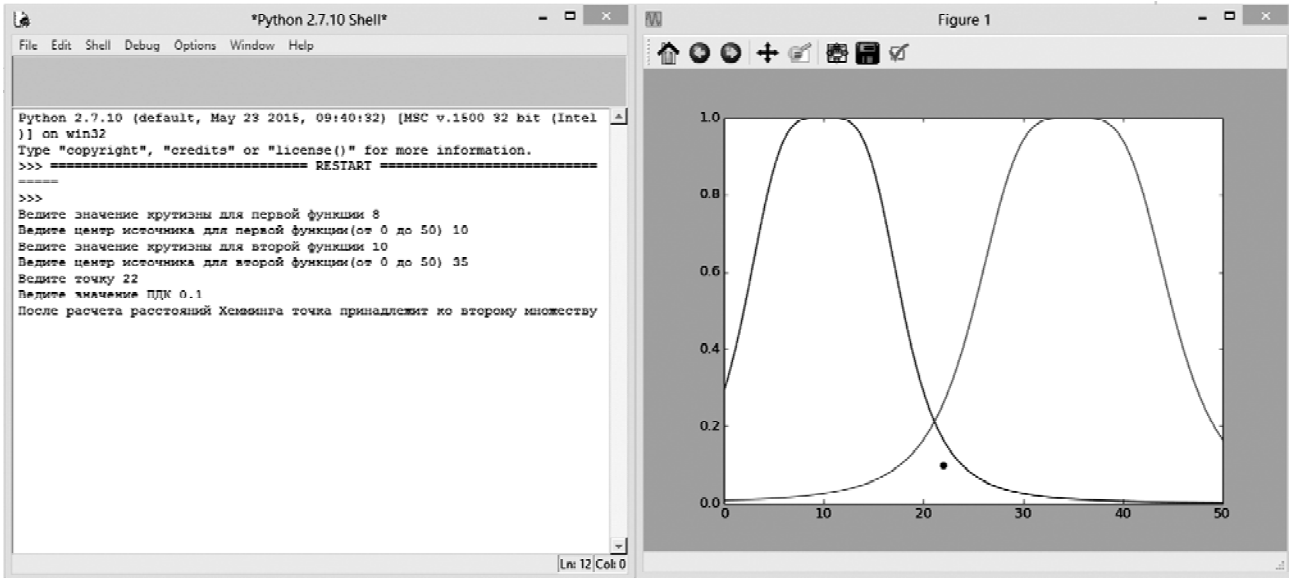


Рисунок 1 – UML диаграмма классов приложений





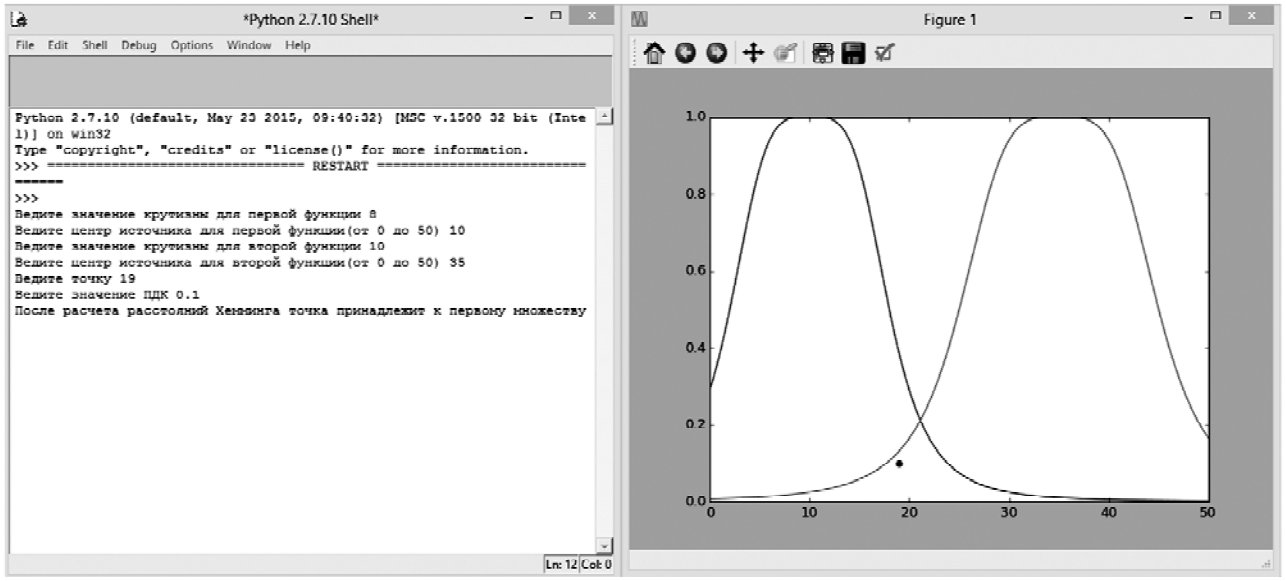


Рисунок 2 – Моделирование разбиений с различными сценариями

ВЫВОДЫ

Разработка знание-ориентированных интеллектуальных методов и моделей анализа сложных объектов является важной составляющей технологических пространственно распределенных процессов, функционирующих в условиях неопределенности. Знание-ориентированные методы направлены на моделирование и обработку детерминированных, вероятностных и нечетких знаний, как фактора повышения качества разбиений. Полученные результаты научных исследований позволили более полно, с высокой адекватностью реализовать разбиения по критериям плотности в нечетком пространстве состояний. Подход позволяет принципиально выделить нечеткую область при пересечении признаков. Таким образом, в работе предложено и исследовано:

1. Выполнен содержательный анализ методов и алгоритмов кластеризации объектов на множестве признаков. Определено, что наличие множества методов и алгоритмов кластеризации не охватывает всей совокупности подходов и особенностей распределения признаков плотности объектов, что свойственно распределению различной природы. В связи с этим важно рассмотрение подходов к нечеткому разбиению объектов на основе плотности их распределения в нечетком пространстве состояний.

2. Получил дальнейшее развитие новый нечеткий метод разбиения, основанный на вычислении плотности распределения интегральных признаков объектов в нечетком пространстве состояний, который, в отличие от существующих, дополнительно функционирует в нечетком пространстве состояний и признаков путем построения решетчатого гиперкуба с нечеткой функцией принадлежности, что позволяет рациональное разбиение признаков на множестве объектов.

3. Перспективой дальнейших исследований является исследование и алгоритмизация метода в задачах проектирования и эксплуатации систем, его адаптация в пространстве признаков предметных областей.

БЛАГОДАРНОСТИ

Работа выполнена в рамках исследований госбюджетной НИР «Нейро-фаззи системы для текущей кластеризации и классификации последовательностей данных в условиях их искривления отсутствующими и аномальными наблюдениями» (№ гос. регистрации 0113U000361). Авторами разработаны новые методы и модели, основанные на нечетком разбиении объектов на основе критериев плотности. Определены границы адекватности метода в программной среде ГИС и его вычислительная сложность, которая близка к квадратичной.

В рамках выполняемой НИР решены также задачи практической реализации и внедрения на реальных объектах.

СПИСОК ЛИТЕРАТУРЫ

1. Gan G. Data Clustering: theory, algorithms, and applications / G. Gan, C. Ma, J. Wu. – SIAM, Philadelphia, ASA, Alexandria, VA, 2007. – 466 p.
2. Cluster Analysis / [B. Everitt, S. Landau, M. Leese, D. Stahl]. – John Wiley &Sons Ltd, 2011. – 330 p.
3. Xu R. Clustering / R. Xu, D. C. Wunsch. John Wiley &Sons, Inc, 2009. – 358 p.
4. Yager R. Essentials of Fuzzy Modeling and Control / R. Yager, D. Filev. – USA : John Willey & Sons,1984. –387 p.
5. Борисов В. В. Нечеткие модели и сети / В. В. Борисов, В. В. Круглов, А. С. Федулов. – М. : Горячая линия, 2012. – 284 с.
6. Tsoukalas L. H. Fuzzy and Neural Approaches in Engineering / L. H. Tsoukalas, R. E. Uhrig. – New York : John Wiley&Sons.Inc, 1997. – 587 p.
7. Блейхут Р. Теория и практика кодов, контролирующих ошибки / Р. Блейхут. – М. : Мир, 1986. – 576 с.
8. Severance C. Python for Informatics [Electronic resource]. – Regime of access : http://do1.dr-chuck.com/py4inf/EN-us/book.pdf.

Статья поступила в редакцию 16.10.2015.
После доработки 28.10.2015.

Кучеренко Є. І.[1], Глушенкова І. С.[2], Глушенков С. А.[3]

[1]Д-р техн. наук, професор, професор кафедри штучного інтелекту Харківського національного університету радіоелектроніки, Харків, Україна

[2]Канд. техн. наук, доцент кафедри геоінформаційних систем, оцінки землі та нерухомого майна Харківського національного університету міського господарства імені О.М. Бекетова, Харків, Україна

[3]Аспирант кафедри геоінформаційних систем, оцінки землі та нерухомого майна Харківського національного університету міського господарства імені О.М. Бекетова, Харків, Україна

НЕЧІТКЕ РОЗБИТТЯ ОБ'ЄКТІВ НА ОСНОВІ КРИТЕРІЇВ ЩІЛЬНОСТІ

Розв'язано задачу розбиття за критеріями щільності в нечіткому просторі станів при перетині ознак. Об'єктом дослідження були процеси розбиття заданої вибірки об'єктів на підмножини. Предмет дослідження становлять методи й алгоритми нечіткого розбиття об'єктів на основі критеріїв щільності в складних системах. Мета роботи: розвиток методу гірської кластеризації Ягера-Філев на основі нечітких уявлень для підвищення ефективності рішень. Запропоновано нечіткий метод розбиття, заснований на обчисленні щільності розподілу інтегральних ознак об'єктів в нечіткому просторі станів, який, на відміну від існуючих, додатково функціонує в нечіткому просторі станів і ознак. Описано й обгрунтовано етапи методу нечіткого розбиття ознак із застосуванням нечіткої відстані Хеммінга. Було створено програму моделювання щільності розподілу ознак на основі розробленого методу. Виконано експеримент щодо визначення належності об'єкта при перетині областей нечіткого розподілу ознак та надання результатів у вигляді логічного виведення і графічного матеріалу. Результати експерименту дозволяють рекомендувати запропонований метод для використання на практиці. Перспективою подальших досліджень є дослідження та алгоритмізація методу, його адаптація в просторі ознак предметних областей.

Ключові слова: кластеризація, відстань Хеммінга, гірська кластеризація, нечітка логіка.

Kucherenko Ye. I.[1], Glushenkova I. S.[2], Glushenkov S. A.[3]

[1]Dr.Sc., Professor, Department of Artificial Intelligence, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

[2]PhD, Associate Professor, Department of Geoinformation Systems, Land Valuation and Real Property, O. M. Beketov National University of Urban Ekonomy in Kharkiv, Kharkov, Ukraine

[3]Postgraduate student, Department of Geoinformation Systems, Land Valuation and Real Property, O. M. Beketov National University of Urban Ekonomy in Kharkiv, Kharkiv, Ukraine

FUZZY PARTITIONING OF THE OBJECTS BASED ON THE CRITERIA OF DENSITY

The problem of the partition of the criteria in the fuzzy space density of states at the intersection of features. The object of research is the process of partitioning a given sample of objects into subsets. Subject of research methods and algorithms make fuzzy partition of objects based on the criteria density in complex systems. Objective: to develop a method of clustering mining Jager-Fileva based on fuzzy concepts to improve the effectiveness of the decisions. It was proposed fuzzy partitioning method based on the calculation of the density distribution of the integral attributes of the objects in a fuzzy space of conditions. The method, in contrast to existing, additionally operates in a fuzzy state space and features. Describe and justify the steps of the method of fuzzy partitioning features using fuzzy Hamming distance. It was created simulation program distribution density of features on the basis of this method. An experiments conducted to determine the affiliation of the object at the intersection of fuzzy areas of distribution and the provision of evidence of results in the form of inference and graphic material. The experimental results allow us to recommend the proposed method to be used in practice. Prospects for further research is to study and algorithmization method, its adaptation to the feature space domains.

Keywords: Clustering, Hamming distance, mountain clustering, fuzzy logic.

REFERENCES

1 Gan G., Ma C., Wu J. Data Clustering: theory, algorithms, and applicationsю SIAM, Philadelphia, ASA, Alexandria, VA, 2007, 466 p.

2. Everitt B., Landau S., Leese M., Stahl D. Cluster Analysis. John Wiley &Sons Ltd, 2011, 330 p.
3. Xu R., Wunsch D. C. Clustering. John Wiley &Sons, Inc, 2009, 358 p.
4. Yager R., Filev D. Essentials of Fuzzy Modeling and Control. USA, John Willey & Sons, 1984, 387 p.
5. Borisov V. V., Kruglov V. V., Fedulov A. S. Nechetkie modeli i seti. Moscow, Gorjachaja linija, 2012, 284 p.
6. Tsoukalas L. H., Uhrig R. E. Fuzzy and Neural Approaches in Engineering. New York, John Wiley&Sons.Inc, 1997, 587 p.
7. Blejhut R. Teorija i praktika kodov, kontrolirujushhih oshibki. Moscow, Mir, 1986, 576 p.
8. Severance C. Python for Informatics [Electronic resource]. Regime of access : http://do1.dr-chuck.com/py4inf/EN-us/book.pdf.

УДК 004.93

Олійник А. О.

Канд. техн. наук, доцент, доцент кафедри програмних засобів, Запорізький національний технічний університет, Запоріжжя, Україна

ВИДОБУВАННЯ ПРОДУКЦІЙНИХ ПРАВИЛ НА ОСНОВІ НЕГАТИВНОГО ВІДБОРУ

Вирішено завдання розробки математичного забезпечення для автоматизації видобування набору знань у вигляді продукційних правил з навчальних вибірок даних. Об'єктом дослідження є процес побудови моделей неруйнівного контролю якості. Предмет дослідження становлять методи видобування продукційних правил на основі негативного відбору для синтезу моделей контролю якості. Мета роботи: створення методу синтезу продукційних правил на основі негативного відбору, що полягає в обробці даних навчальної вибірки, яка характеризується істотною відмінністю кількості екземплярів, що відносяться до різних класів. Запропоновано метод синтезу продукційних правил на основі негативного відбору для випадку нерівномірного розподілу екземплярів класів вибірки, який при генерації набору детекторів використовує відому інформацію про екземпляри всіх класів вибірки, враховує інформацію про індивідуальну значущість ознак, як форму детектора використовує гіперкуб максимально можливого об'єму. Розроблений метод дозволяє виключати малозначущі і надлишкові ознаки з вибірки, скоротивши тим самим простір пошуку і час виконання методу, а також формувати набір детекторів з високими апроксимаційними й узагальнюючими здібностями. Запропонований метод за рахунок підвищення узагальнюючих властивостей синтезованих моделей шляхом скорочення числа детекторів і умов антецедентів також підвищує інтерпретабельність моделі, скорочує її розмірність (структурну і параметричну складність), обсяг використовуваної пам'яті і підвищує швидкодію моделі при послідовній реалізації обчислень. Проведено експерименти з дослідження властивостей запропонованого методу. Результати експериментів дозволяють рекомендувати запропонований метод для використання на практиці.

Ключові слова: вибірка, діагностування, модель контролю якості, негативний відбір, продукційне правило.

НОМЕНКЛАТУРА

E – помилка розпізнавання на навчальних даних ($S =< P, T >$);

E_t – помилка розпізнавання на тестових даних;

m – номер ознаки (характеристики) об'єкта;

M – кількість ознак вибірки S;

$N(p_{mn})$ – кількість екземплярів вибірки S, значення m-ї ознаки яких, належать n-му інтервалу діапазону її зміни;

$N_{\text{int}}(p_m)$ – кількість інтервалів, на які розбивається діапазон значень m-ї ознаки p_m;

$N(p_{mn}, t_l)$ – кількість екземплярів вибірки S, значення вихідного параметра T яких дорівнює t_l (належать l-му інтервалу діапазону його зміни t_l) за умови, що значення їх m-ї ознаки належить n-му інтервалу p_{mn};

$N_{\text{int}}(T)$ – кількість можливих значень (інтервалів, на які розбивається діапазон значень) вихідного параметра T;

N_{it} – кількість ітерацій роботи методу;

$N_{t_q = t'_0}$ – кількість екземплярів вибірки $S =< P, T >$, значення вихідного параметра t_q яких дорівнює t'_0;

$N_{t_q = t'_1}$ – кількість екземплярів вибірки $S =< P, T >$, значення вихідного параметра t_q яких дорівнює t'_1;

$N_{t, t_q = t'_1 / t_q = t'_0}$ – кількість екземплярів тестової вибірки, що розпізнані як «свої» ($t_q = t'_1$), але реально відносяться до класу «чужих» ($t_q = t'_0$);

$N_{t, t_q = t'_0}$ – кількість екземплярів тестової вибірки, що відносяться до класу «чужих» ($t_q = t'_0$);

$N_{t, t_q = t'_0 / t_q = t'_1}$ – кількість екземплярів тестової вибірки, що розпізнані як «чужі» ($t_q = t'_0$), але реально відносяться до класу «своїх» ($t_q = t'_1$);

$N_{t, t_q = t'_1}$ – кількість екземплярів тестової вибірки, що відносяться до класу «своїх» ($t_q = t'_1$);

P – набір вхідних характеристик (ознак) об'єктів у вибірці $S =< P, T >$;

p_{qm} – значення m-ї ознаки q-го екземпляра вибірки S;

$p_{m\min}$ – мінімальне значення m-ї ознаки у вибірці;

$p_{m\max}$ – максимальне значення m-ї ознаки у вибірці;

$P_{t, t_q = t'_1 / t_q = t'_0}$ – імовірність помилки віднесення до класу «своїх» ($t_q = t'_1$) за умови, що екземпляр реально відноситься до класу «чужих» ($t_q = t'_0$);

$P_{t, t_q = t'_0 / t_q = t'_1}$ – імовірність помилки віднесення до класу «чужих» ($t_q = t'_0$) за умови, що екземпляр реально відноситься до класу «своїх» ($t_q = t'_1$);

q – номер екземпляра (об'єкта) у вибірці S;

Q – кількість екземплярів вибірки S;

$\rho(p_{mn})$ – імовірність того, що значення ознаки p_m екземплярів вибірки S потрапить у n-й інтервал діапазону її зміни;

$\rho(p_{mn}, t_l)$ – умовна імовірність того, що значення вихідного параметра T буде дорівнює t_l (потрапить у l-й інтервал t_l) за умови, що m-а ознака p_m потрапить у n-й інтервал p_{mn};

$S =< P, T >$ – навчальна вибірка;

t – час роботи методу, мс;

$t_q \in T'$ – значення вихідного параметра q-го екземпляра;

T – множина значень вихідного параметра у вибірці $S =< P, T >$;

T' – множина можливих значень вихідного параметра T.

DOI 10.15588/1607-3274-2016-1-5

ВСТУП

У процесі побудови моделей прийняття рішень для неруйнівного контролю якості, технічного та медичного діагностування, розпізнавання образів [1–4] можуть виникати ситуації, коли велика частина інформації в навчальній вибірці даних відноситься до одного класу (наприклад, переважна більшість виробів відноситься до одного класу придатності) [5, 6].

У таких випадках для формалізації описів досліджуваних об'єктів або процесів доцільно синтезувати моделі на основі штучних імунних систем [7–9], що характеризуються можливостями навчання на основі екземплярів тільки одного класу, а також високим рівнем адаптації. Для вирішення задач, що характеризуються істотною відмінністю кількості екземплярів, що відносяться до різних класів, пропонується використовувати штучні імунні системи, що працюють на основі принципів негативного відбору [10–13], що передбачає побудову набору детекторів (обчислювальних елементів), здатних до розпізнавання невідомих екземплярів [14–16]. Такий підхід дозволяє виявляти аномалії або випадкові зміни в діагностованих об'єктах [7, 10], а також розпізнавати екземпляри чужих класів (класів об'єктів, екземпляри яких не представлені в навчальній вибірці) [8, 12, 15].

Проте відомі методи синтезу штучних імунних систем на основі негативного відбору [8–16] генерують надлишкову кількість детекторів (можливих рішень задачі), висувають значні вимоги до обчислювальних ресурсів ЕОМ, як правило, використовують інформацію тільки про один клас екземплярів («своїх», придатних і т.п.), не враховуючи при цьому дані про екземпляри інших класів. Отже, актуальною є розробка методів синтезу штучних імунних систем на основі негативного відбору, вільних від зазначених недоліків. Крім того, діагностичні моделі на основі штучних імунних систем характеризуються низьким рівнем узагальнення. Не дивлячись на те, що детектори (правила) імунної системи по окремості є легкими в сприйнятті і розумінні людиною, через низький рівень узагальнення, система детекторів має велику розмірність, і, отже, є складною для сприйняття й аналізу людиною, що в цілому призводить до зниження інтерпретабельності діагностичної моделі на основі імунних систем.

Метою роботи є створення методу синтезу продукційних правил на основі негативного відбору, що полягає в обробці даних навчальної вибірки, яка характеризується істотною відмінністю кількості екземплярів, що відносяться до різних класів.

1 ПОСТАНОВА ЗАДАЧІ

Нехай задана навчальна множина $S =< P, T >$. Набір P представляється у вигляді матриці $P = \left(p_{qm}\right)_{QM}$, $m = 1, 2, ..., M$, $q = 1, 2, ..., Q$. Набір значень вихідного параметра представляється у виді вектора $T = \left(t_q\right)_Q$, елементи $t_q \in T'$ якого приймаються значення з множини T'. У задачах неруйнівного контролю якості і розпізнавання образів множина T', як правило, складається з двох елементів $T' = \left\{t'_0, t'_1\right\}$, що визначають клас придатності виробу, наприклад при $t_q = t'_0$ q-й виріб вважається придатним, при $t_q = t'_1$ – некондиційним).

При цьому кількість екземплярів вибірки $S =< P, T >$ одного класу (наприклад, екземплярів класу некондиційних $t_q = t'_1$) істотно відрізняється від кількості екземплярів іншого класу, що визначається нерівністю (1):

$$0 \le N_{t_q = t'_0} << N_{t_q = t'_1}, \qquad (1)$$

де $N_{t_q = t'_0} + N_{t_q = t'_1} = Q$.

Тоді на основі заданої вибірки $S =< P, T >$ необхідно синтезувати набір $RB = \left\{rule_1, rule_2, ..., rule_{NR}\right\}$ продукцій $P_r \to T_r$, що дозволяє забезпечити прийнятний рівень похибки розпізнавання E, розрахованої як відношення кількості N_{er} неправильно розпізнаних за допомогою набору правил RB спостережень вибірки S до загальної кількості екземплярів Q (2):

$$E = \frac{N_{er}}{Q}. \qquad (2)$$

2 ОГЛЯД ЛІТЕРАТУРИ

Методи негативного відбору використовують процеси позитивної та негативної селекції, здійснювані під час дозрівання T-лімфоцитів, використовуються для класифікації та розпізнавання в задачах, де простір станів моделюється на основі наявних знань [10–13, 16]. В основі роботи моделі негативного відбору [10–13, 16] лежить поведінка T-клітин, що забезпечує терпимість імунної системи організму до власних клітин. При цьому T-клітини мають здатність розпізнавати практично будь-які (невідомі їм раніше) антигени. У термінах теорії імунної системи антигеном називають будь-який екземпляр, який може бути розпізнаний системою [7–9].

Основне завдання, яке вирішується за допомогою моделі негативного відбору, полягає у виявленні відмінностей між двома класами об'єктів і в проведенні подальшої двокласової класифікації [11, 16]. З погляду діагностування це завдання можна розглядати як завдання виявлення аномалій або випадкових змін у стані діагностованих об'єктів.

Чисельні детектори використовуються в тих випадках, коли стан системи можна представити у вигляді вектора ознак, значення яких нормалізовані і знаходяться в діапазоні від 0 до 1. Методи негативного відбору [10, 11, 16] засновані на використанні чисельного представлення детекторів, що призводить до більш компактного подання даних та прискорює генерацію детекторів, а також використовують класичні принципи класифікації для визначення належності детектора до придатних або дефектних екземплярів.

Проте використання чисельних детекторів у відомих методах негативного добору приводить до різних проблем, таких як неможливість задати заздалегідь розміри детекторів і їх кількість, неможливість передбачити збіжність методу, в результаті чого можливі ситуації, коли оптимальний набір детекторів так і не буде отриманий.

Потреба усунення недоліків відомих методів обумовлює необхідність розробки нового методу негативного відбору, здатного синтезувати набір детекторів за даними навчальної вибірки, що містить інформацію про екземпляри різних класів.

3 МАТЕРІАЛИ ТА МЕТОДИ

Як відзначено вище, відомі методи негативного відбору [8–16] мають такі недоліки, як генерація надлишкової кількості детекторів, використання інформації про екземпляри тільки одного класу, низька інтерпретабельність синтезованого набору рішень у вигляді детекторів. Крім того, більшість методів, заснованих на принципі негативного відбору, в якості детекторів використовують гіперсфери з фіксованим радіусом, який визначає область простору ознак, що покривається детектором. Вибір величини радіуса гіперсфери-детектора являє собою дуже складну задачу, оскільки при великих значеннях радіуса знижується точність розпізнавання, а при низьких значеннях збільшується кількість генерованих детекторів, що у свою чергу знижує узагальнюючі властивості синтезованої моделі у вигляді набору детекторів штучної імунної мережі.

Наявність зазначених недоліків обумовлює необхідність висування істотних вимог до обчислювальних ресурсів ЕОМ при використанні відомих методів негативного відбору, що, у свою чергу, знижує швидкість пошуку рішення і, у деяких випадках, не дозволяє знайти прийнятне рішення.

Для усунення зазначених недоліків доцільно використовувати метод синтезу продукційних правил на основі негативного відбору для випадку нерівномірного розподілу екземплярів класів вибірки, у якому пропонується:

– при генерації набору детекторів $AB = \{Ab_1, Ab_2, ..., Ab_{N_{Ab}}\}$ використовувати відому інформацію про екземпляри обох класів $T' = \{t'_0, t'_1\}$, що дозволить сформувати набір детекторів з великими апроксимаційними й узагальнюючими властивостями;

– використовувати інформацію про індивідуальну значущість V_m ознак p_m, що дозволить виключити малозначущі та надлишкові ознаки з вибірки $S =< P, T >$;

– як форму детектора використовувати гіперкуб максимально можливого обсягу. На відміну від відомих методів негативного відбору, у яких в якості форми детектора використовується гіперсфера, це дозволить виключити необхідність вирішення ресурсномісткої задачі пошуку оптимальних радіусів гіперсфер детекторів.

У розробленому методі на початковому етапі пропонується оцінювати значущість ознак p_m стосовно вихідного параметра T, що дозволить виявити і виключити з подальшого розгляду малозначущі ознаки, скоротивши тим самим простір пошуку і час виконання методу. Як відзначено вище, у цій роботі розглядається задача, у якій вихідна вибірка $S =< P, T >$ характеризується дискретною кількістю класів $T' = \{t'_0, t'_1\}$. Тому для оцінювання значущості V_m ознак p_m доцільно застосовувати критерії, що дозволяють виконувати оцінку інформативності ознак стосовно дискретного вихідного параметра T [2, 4, 17–22]. Як такий критерій пропонується використовувати ентропію ознаки [4, 17] як характеристику, що відображає ступінь невизначеності стану об'єкта й обчислюється за формулою (3):

$$V_m = -\sum_{n=1}^{N_{\text{int}}(p_m)} \left(\rho(p_{mn}) \sum_{l=1}^{N_{\text{int}}(T)} \rho(p_{mn}, t_l) \log_2 \rho(p_{mn}, t_l) \right), \quad (3)$$

де $\rho(p_{mn}) = \dfrac{N(p_{mn})}{Q}$; $\rho(p_{mn}, t_l) = \dfrac{N(p_{mn}, t_l)}{N(p_{mn})}$.

Ознаки p_m, значення індивідуальної інформативності яких нижче мінімально допустимої ($V_m < V_{\min}$), вважаються малоінформативними і виключаються з вибірки $S =< P, T >$.

Крім того пропонується оцінити взаємозв'язок ознак як інформативність однієї з них стосовно іншої, що дозволить виявити групи взаємозалежних ознак, у кожній з який залишити тільки одну високо інформативну ознаку, а інші ознаки, пов'язані з нею у групі можна виключити з подальшого розгляду, оскільки вони є надлишковими, ускладнюють процес синтезу діагностичних моделей і знижують їх інтерпретабельність. Для оцінювання інформативності ознак між собою V_{md} пропонується також використовувати ентропію ознак, використовуючи формулу (3) і вважаючи при цьому одну з ознак p_d вихідним параметром T (попередньо інтервал значень ознаки, що вважається вихідним параметром p_d, розбивається на $N_{\text{int}}(T)$ дискретних інтервалів).

Після оцінювання інформативності ознак стосовно інших ознак з вибірки S виключаються ті з них, для яких існують аналоги (у випадку, якщо значення взаємної інформативності ознак більше максимально припустимої $V_{md} > V_{\max}$).

Далі виконується побудова множини детекторів – структур, що дозволяють визначити, чи відноситься оцінюваний екземпляр до деякого класу. Важливо відзначити, що при використанні принципів негативного відбору детектори, що будуються на основі класу $T' = t'_1$, дозволяють виявити з невідомих екземплярів такі, які не відносяться до відповідного класу t'_1 [9, 11, 13]. Тому для формування множини детекторів у задачі розпізнавання, в якій вихідний параметр приймає два значення t'_1 (клас «своїх») і t'_0 (клас «чужих»), необхідно з вхідної вибірки $S =< P, T >$ сформувати вибірки S_0 і S_1 з екземплярів, що відносяться до класів t'_1 і t'_0, відповідно: $S_1 =< P, T = t'_1 >$ (вибірка «своїх» екземплярів) і $S_0 =< P, T = t'_0 >$ (вибірка «чужих» екземплярів).

Після цього виконується формування першого кандидата в детектори $Ab_1 =< Ab_{1\min}, Ab_{1\max} > \in AB_1$, де $Ab_{1\min} = \{Ab_{11\min}, Ab_{12\min}, ..., Ab_{1M\min}\}$ і $Ab_{1\max} = \{Ab_{11\max}, Ab_{12\max}, ..., Ab_{1M\max}\}$ – відповідно, набори мінімальних і максимальних значень m-х ознак кандидата в детектори Ab_1 ($Ab_{1m\min} = \min\limits_{q=1,2,...,Q_1} (p_{qm})$,

$Ab_{1m\max} = \max\limits_{q=1,2,...,Q_1}\left(p_{qm}\right)$, $m = 1,2,...,M$, Q_1 – кількість екземплярів у вибірці S_1), представленого у вигляді гіперкуба. Множина AB_1 детекторів Ab_k формується на основі набору «своїх» екземплярів S_1 і дозволяє виявляти серед невідомих екземплярів «чужі», тобто такі, які не відносяться до класу t_1'.

Потім для кожного q-го екземпляра s_q вибірки $S_1 =< P, T = t_1 >$ визначається його відповідність кандидату в детектори Ab_k за формулою (4):

$$eq\left(Ab_k, s_q\right) = \begin{cases} 1, \left(\sum\limits_{m=1}^{M}\left\{1 \mid \left(Ab_{km\min} < p_{qm}\right) \wedge \left(Ab_{km\max} > p_{qm}\right)\right\}\right) = M; \\ 0, \left(\sum\limits_{m=1}^{M}\left\{1 \mid \left(Ab_{km\min} < p_{qm}\right) \wedge \left(Ab_{km\max} > p_{qm}\right)\right\}\right) \neq M, \end{cases} \quad (4)$$

де сума $\sum\limits_{m=1}^{M}\left\{1 \mid \left(Ab_{km\min} < p_{qm}\right) \wedge \left(Ab_{km\max} > p_{qm}\right)\right\}$ визначає кількість відповідностей значень ознак p_{qm} q-го екземпляра кандидату Ab_k. Якщо $\left(\sum\limits_{m=1}^{M}\left\{1 \mid \left(Ab_{km\min} < p_{qm}\right) \wedge \left(Ab_{km\max} > p_{qm}\right)\right\}\right) = M$, то вважається, що екземпляр $s_q =< p_{qm}, t_q >$ відповідає кандидату в детектори Ab_k (знаходиться усередині простору гіперкуба з координатами $Ab_{1\min} = \left\{Ab_{11\min}, Ab_{12\min}, ..., Ab_{1M\min}\right\}$ і $Ab_{1\max} = \left\{Ab_{11\max}, Ab_{12\max}, ..., Ab_{1M\max}\right\}$).

Якщо існує хоча б один екземпляр $s_q =< p_{qm}, t_q = t_1' >\in S_1$, для якого $eq\left(Ab_k, s_q\right) = 1$, то вважається, що кандидат Ab_k активується при співставленні його з екземпляром s_q і, отже, не може бути детектором. Тому при виконанні умови (5)

$$\exists s_q \in S : eq\left(Ab_k, s_q\right) = 1 \quad (5)$$

відбувається етап до навчання кандидата Ab_k. Метою даного етапу є перетворення кандидата в детектори Ab_k таким чином, щоб у вибірці S_1 не існувало екземплярів, при зіставленні з якими відбувалася би активація детектора Ab_k. Для цього вибирається кортеж однієї з ознак $Ab_{km} =< Ab_{km\min}, Ab_{km\max} >$, за якими кандидат у детектори Ab_k збігається з екземпляром s_q. Далі перетворюється одне з граничних значень m-ї ознаки кандидата Ab_k: $Ab_{km\min} = p_{qm} + \eta_n\left(Ab_{km\max} - Ab_{km\min}\right)$, якщо $rnd > 0{,}5$ ($rnd = rand[0;1)$ – випадково згенероване число в діапазоні $[0;1)$), у противному випадку – $Ab_{km\max} = p_{qm} - \eta_n\left(Ab_{km\max} - Ab_{km\min}\right)$, у результаті чого об'єм гіперкуба Ab_k зменшується таким чином, що екземпляр s_q розташовується поза простором, описуваним кандидатом у детектори Ab_k. Коефіцієнт η_n задається користувачем як параметр методу в інтервалі $\eta_n \in (0;1]$. Чим більше значення даного коефіцієнта, тим більше відстань між екземплярами вибірки S_1 і гіперкубом детектора Ab_k.

Після перетворення граничних значень $Ab_{km} =< Ab_{km\min}, Ab_{km\max} >$ однієї з ознак кандидата в детектори Ab_k, він повторно перевіряється з кожним екземпляром вибірки S_1 на виконання умови (5). При виконанні умови (5) аналогічним чином відбувається повторне перетворення граничних значень однієї з ознак кандидата Ab_k. І так доти, поки буде виконуватися умова (5).

Після того, як у множині $S_1 =< P, T = t_1 >$ не залишиться екземплярів s_q, при зіставленні з якими активізується кандидат Ab_k, виконується етап оцінювання пристосованості кандидата Ab_k до узагальнення. При використанні принципів негативного відбору формується множина детекторів $AB_1 = \left\{Ab_1, Ab_2, ..., Ab_{N_{Ab}}\right\}$, що дозволяють з високою точністю визначити належність екземплярів s_q вибірки S до визначеного класу [10–16]. Тому як критерій оцінювання будемо використовувати характеристики (6) і (7), що дозволяють оцінити здатність детектора Ab_k до узагальнення даних:

$$G_1\left(Ab_k\right) = \frac{1}{M}\sum_{m=1}^{M}\frac{Ab_{km\max} - Ab_{km\min}}{p_{m\max} - p_{m\min}}, \quad (6)$$

$$G_2\left(Ab_k\right) = \frac{\prod\limits_{m=1}^{M}\left(Ab_{km\max} - Ab_{km\min}\right)}{\prod\limits_{m=1}^{M}\left(p_{m\max} - p_{m\min}\right)}, \quad (7)$$

де $p_{m\min} = \min\limits_{q=1,2,...,Q_1}\left(p_{qm}\right)$ та $p_{m\max} = \max\limits_{q=1,2,...,Q_1}\left(p_{qm}\right)$.

Критерії (6) і (7) відображають частину детектора, що покривається за допомогою, простору пошуку. Критерій $G_1\left(Ab_k\right)$ показує середню частку простору, що покривається детектором, у кожному з M вимірів простору ознак. Критерій $G_2\left(Ab_k\right)$ відображає об'ємну частину простору, що покривається.

Чим більше значення критеріїв $G_1\left(Ab_k\right)$ і $G_2\left(Ab_k\right)$, тим більш велику частину простору пошуку покриває детектор. Отже, якщо значення критерію оцінювання якості узагальнення $G\left(Ab_k\right)$ вище граничного $G_{\min}$ ($G\left(Ab_k\right) > G_{\min}$), то вважається, що кандидат Ab_k характеризується високими узагальнюючими здібностями, може бути детектором і додається в множину детекторів: $AB_1 = AB_1 \cup \left\{Ab_k\right\}$.

Створення нових кандидатів Ab_k здійснюється доти, поки не будуть досягнуті критерії закінчення пошуку. Як такі критерії можуть бути використані точність розпізнавання $E(S)$, досягнення максимально припустимої

кількості детекторів ($N_{Ab} = |AB| > N_{Ab\max}$), перевищення максимальне припустимого часу пошуку ($t > t_{\max}$).

Отриманий у результаті негативного відбору набір детекторів $AB_1 = \{Ab_1, Ab_2, ..., Ab_{N_{Ab}}\}$ описує область простору пошуку $\overline{S_1}$, комплементарну області простору, у якій розташована множина «своїх» екземплярів S_1. При цьому множина $AB_1 = \{Ab_1, Ab_2, ..., Ab_{N_{Ab}}\}$ характеризується високими апроксимаційними й узагальнюючими здібностями.

Аналогічним чином можна сформувати набір детекторів AB_0 для множини S_0. Проте, у задачах з нерівномірним розподілом екземплярів класів вибірки $S =< P, T >$, коли кількість екземплярів одного класу $Q_1 = N_{t_q = t_1'}$ істотно перевищує кількість екземплярів іншого класу $Q_0 = N_{t_q = t_0'}$ ($Q_0 << Q_1$), можуть виникнути проблеми з генерацією детекторів, що адекватно відображають простір екземплярів S_0 (можуть бути згенеровані детектори Ab_k у виді гіперкубів занадто великого обсягу, що не зможуть узагальнити дані генеральної сукупності). Це обумовлено невеликою (недостатньою) кількістю екземплярів у вибірці S_0 ($Q_0 << Q_1$), а іноді і майже повною їхньою відсутністю.

Тому в цій роботі при генерації детекторів для екземплярів S_0 пропонується використовувати інформацію про розміри детекторів, побудованих на основі вибірки S_1. При цьому детектори будуть відображати інформацію про наявність у гіперкубі екземплярів вибірки S_0 (а не про їх відсутність, як при класичному негативному відборі), і, отже, по суті будуть цілком відповідати детекторам, що побудовані раніше для вибірки S_1 на основі негативного відбору й містять інформацію про області простору пошуку, у яких не розташовуються екземпляри S_1.

Детектори $Ab_k^{(0)}$ вибірки S_0 генеруються таким чином, щоб їх центри відповідали координатам екземплярів $s_k =< p_{km}, t_k = t_0' > \in S_0$ вибірки S_0, а розміри граней їх гіперкубів відповідали аналогічним розмірам детекторів, створених на основі даних вибірки S_1. Отже, координати детектора $Ab_{km}^{(0)} =< Ab_{km\min}^{(0)}, Ab_{km\max}^{(0)} >$ визначаються за формулами (8)–(9):

$$Ab_{km\min}^{(0)} = p_{km} - \frac{1}{2}\Delta Ab_m, \quad (8)$$

$$Ab_{km\max}^{(0)} = p_{km} + \frac{1}{2}\Delta Ab_m, \quad (9)$$

де ΔAb_m – середня довжина граней детекторів $AB_1 = \{Ab_1, Ab_2, ..., Ab_{N_{Ab}}\}$, створених на основі множини S_1. Величину ΔAb_m пропонується обчислювати, виходячи з інформації про детектори $AB_1 = \{Ab_1, Ab_2, ..., Ab_{N_{Ab}}\}$, використовуючи формулу (10):

$$\Delta Ab_m = \frac{1}{N_{Ab}M}\left(\sum_{k=1}^{N_{Ab}}\sum_{m=1}^{M}\left(Ab_{km\max} - Ab_{km\min}\right)\right). \quad (10)$$

Потім виконується зіставлення згенерованих детекторів $Ab_k^{(0)}$ з екземплярами вибірки S_1, використовуючи формулу (4), при виконанні умови (5) відбувається перетворення детекторів $Ab_k^{(0)}$ аналогічно описаному вище етапу донавчання. Після цього обчислюється значення одного з критеріїв $G\left(Ab_k^{(0)}\right)$ оцінювання здатності детектора до узагальнення даних (6)–(7), і, у випадку, якщо його значення вище граничного, детектор $Ab_k^{(0)}$ додається в множину $AB_0 = AB_0 \cup \{Ab_k^{(0)}\}$.

У такий спосіб генерується набір детекторів AB_0, що описує, як і набір AB_1, область простору пошуку $\overline{S_1}$, комплементарну області розташування множини екземплярів S_1. Тому розпізнавальна модель може бути представлена у вигляді множини детекторів $AB = AB_0 \cup AB_1$, що дозволяють розпізнавати належність невідомих екземплярів $s_q' =< p_{qm}', t_q' - ? > \notin S$ до класу «чужих», тобто відносити їх до класу t_0': $t_q' = t_0'$.

З метою підвищення рівня інтерпретабельності отриманої розпізнавальної моделі представленої у вигляді набору $AB = \{Ab_1, Ab_2, ..., Ab_{N_{Ab}}\}$ детекторів, пропонується на основі набору AB сформувати множину продукційних правил $PR_r : P_r \to T_r$, у яких ліва частина P_r імплікації являє собою набір умов вигляду (11)

«Якщо $(p_1 \in [Ab_{k1\min}; Ab_{k1\max}]) \wedge (p_2 \in [Ab_{k2\min};$
$Ab_{k2\max}]) \wedge ... \wedge (p_M \in [Ab_{kM\min}; Ab_{kM\max}])$» (11)

а права частина T_r містить значення вихідного параметра T при виконанні r-го набору умов P_r (11).

При формуванні набору правил PR в антецедент правила P_r будемо включати тільки ті границі ознак, для яких вони не є граничними значеннями, тобто $Ab_{km\min} \neq \min_{q=1,2,...,Q}(p_{qm})$ і $Ab_{km\max} \neq \max_{q=1,2,...,Q}(p_{qm})$. Наприклад, для детектора вигляду $Ab_k = \{< 5,7 >, < 8, p_{2\max} >, < p_{3\min}, p_{3\max} >, < 4,6 >\}$ буде сформоване правило PR_k вигляду: Якщо $(p_1 > 5 \wedge p_1 < 7) \wedge (p_2 > 8) \wedge (p_4 > 4 \wedge p_4 < 6)$, то віднести екземпляр до класу «чужих» ($T \neq t_1$). Як видно, у правило не ввійшли в явному вигляді верхня границя ознаки p_2 і цілком ознака p_3, оскільки відповідні значення детектора знаходяться на мінімальній і максимальній границях ознак і не впливають на якість розпізнавання. Крім того, виключення таких значень із правила PR_k знижує його складність, підвищуючи в такий спосіб інтерпретабельність правила.

Використовуючи такий підхід, на основі кожного *k*-го детектора Ab_k виконується побудова свого правила, формуючи в такий спосіб множину PR, що складається з N_{Ab} продукційних правил $PR_r : P_r \to T_r$.

Таким чином, запропонований метод синтезу продукційних правил на основі негативного відбору для випадку нерівномірного розподілу екземплярів класів вибірки при генерації набору детекторів використовує відому інформацію про екземпляри всіх класів вибірки, враховує інформацію про індивідуальну значущість ознак, як форму детектора використовує гіперкуб максимально можливого обсягу, що дозволяє виключати малозначущі і надлишкові ознаки з вибірки, скоротивши тим самим простір пошуку і час виконання методу, а також формувати набір детекторів з високими апроксимаційними й узагальнюючими здібностями.

Запропонований метод за рахунок підвищення узагальнюючих властивостей синтезованих моделей шляхом скорочення кількості детекторів і умов антецедентів також підвищує інтерпретабельність моделі, скорочує її розмірність (структурну і параметричну складність), обсяг використовуваної пам'яті і підвищує швидкодію моделі при послідовній реалізації обчислень.

4 ЕКСПЕРИМЕНТИ

Для перевірки працездатності запропонованого методу синтезу продукційних правил на основі негативного відбору було розроблено комп'ютерну програму, що реалізує запропонований метод. За допомогою розробленого програмного забезпечення розв'язувалася задача діагностування лопаток газотурбінних авіаційних двигунів [23]. Лопатки характеризувалися значеннями спектрів потужності загасаючих коливань після ударного збудження, які використовувалися як вхідні ознаки. Експертно було визначено класи якості лопаток: придатні (кондиційні) і дефектні (некондиційні). Кожна лопатка була описана 10240 ознаками, що характеризували спектр потужності загасаючих коливань. Для скорочення простору пошуку на основі цих ознак були отримані штучні ознаки-згортки, у результаті чого сформовано набір P, який складається зі штучних 80 ознак.

Отримана вибірка спостережень $S =< P, T >$, вочевидь, не є статистично репрезентативною, оскільки не відображає реального розподілу частот класів (у генеральній сукупності придатних лопаток суттєво більше, ніж дефектних). При цьому дефектні лопатки ($t_q = t_1'$) у наявній вибірці S представляють типові випадки некондиційності, що забезпечує топологічну репрезентативність дефектних лопаток у вибірці. А всі можливі випадки класу придатних ($t_q = t_0'$) неможливо представити у вибірці з практичної точки зору. Тому виникає необхідність на основі наявної вибірки $S =< P, T >$ з нерівномірним розподілом екземплярів по класах побудувати діагностичну модель, що дозволяє виконувати технічне діагностування лопаток авіадвигунів на основі заданого набору вимірювань.

Вибірка $S =< P, T >$ містить 42 екземпляри, що характеризують дефектні лопатки, і 72 екземпляри придатних.

Запропонований метод синтезу продукційних правил на основі негативного відбору порівнювався з існуючими методами негативного відбору, що синтезують набір детекторів тільки на основі «своїх» екземплярів вибірки $S_1 \subseteq S$. Тому за допомогою запропонованого методу задача діагностування лопаток газотурбінних авіаційних двигунів розв'язувалася два рази:

– з використанням підвибірки ($S_1 \subseteq S$), що містить інформацію тільки про некондиційні екземпляри («своїх»);

– з використанням усієї вихідної вибірки $S =< P, T >$.

Як тестова вибірка S_t використовувалася вибірка, що містить інформацію про 273 екземпляри (261 екземпляр придатних виробів, що відносяться до класу $t_q = t_0'$, і 12 екземплярів дефектних виробів, що відносяться до класу $t_q = t_1'$).

Також виконано порівняння запропонованого методу з іншими методами, що дозволяють вирішувати задачі розпізнавання образів. Для цього розв'язувалася описана вище задача діагностування лопаток газотурбінних авіаційних двигунів. Досліджувалися властивості і характеристики таких розпізнавальних моделей:

– модель у вигляді набору продукційних правил, синтезованих за допомогою запропонованого методу на основі негативного відбору з урахуванням інформативності ознак;

– нейромережева модель прямого поширення, що складається з трьох шарів нейронів, побудована на основі методу зворотного поширення помилки. На першому шарі нейромережі знаходилося п'ять нейронів, на другому – три нейрони, на третьому – один нейрон. Нейрони першого і другого шару мали логістичну сигмоїдну функцію активації, третього – граничну функцію активації;

– модель у вигляді набору детекторів, побудована за допомогою методу негативного відбору з маскуванням детекторів [16].

При цьому використовувалася вся навчальна вибірка $S =< P, T >$ обсягом 114 екземплярів (42 екземпляри, що характеризують дефектні лопатки, і 72 екземпляри придатних) при побудові першої і другої моделі, і частина вибірки ($S_1 \subseteq S$) при побудові моделі на основі набору детекторів за допомогою методу негативного відбору з маскуванням, оскільки даний метод припускає роботу з екземплярами тільки одного класу.

5 РЕЗУЛЬТАТИ

Результати експериментів з порівняння запропонованого методу з іншими методами негативного відбору наведено в табл. 1. $P_{t,t_q=t_1'/t_q=t_0'}$ й $P_{t,t_q=t_0'/t_q=t_1'}$ помилки віднесення до класу «своїх» $t_q = t_1'$ («чужих», $t_q = t_0'$) за умови, що екземпляр реально відноситься до класу «чужих» $t_q = t_0'$ («своїх», $t_q = t_1'$) обчислюються за формулами (12) і (13), відповідно:

$$P_{t,t_q=t_1'/t_q=t_0'} = \frac{N_{t,t_q=t_1'/t_q=t_0'}}{N_{t,t_q=t_0'}}, \qquad (12)$$

$$P_{t,t_q=t'_0/t_q=t'_1}=\frac{N_{t,t_q=t'_0/t_q=t'_1}}{N_{t,t_q=t'_1}}. \quad (13)$$

Результати порівняння різних моделей при вирішенні задачі діагностування лопаток газотурбінних авіаційних двигунів [23] наведено в табл. 2, де N_{param} – критерій, що визначає параметричну складність моделі. Критерій N_{param} розраховувався як кількість параметрів моделі: загальна кількість параметрів $Ab_{km\min}$ і $Ab_{km\max}$ в моделях на основі продукцій і на основі набору детекторів [16], загальна кількість настроюваних параметрів (вагових коефіцієнтів) – у нейромережевій моделі.

6 ОБГОВОРЕННЯ

Як видно з табл. 1, прийнятне значення помилки розпізнавання E забезпечив метод з маскуванням детекторів [16] ($E=0{,}018$) і запропонований метод синтезу продукційних правил на основі негативного відбору (МСППНВ). Низькі значення помилки розпізнавання зазначених методів забезпечувалися за рахунок широкого покриття синтезованими детекторами області «своїх» екземплярів вибірки $S_1 \subseteq S$. При цьому запропонований метод МСППНВ, що синтезував набір детекторів на основі екземплярів усіх класів вибірки $S=<P,T>$, забезпечив більш прийнятні результати ($E=0{,}009$) у порівнянні з набором детекторів, синтезованим тільки з використанням екземплярів класу «своїх» $S_1 \subseteq S$ ($E=0{,}026$). Менш прийнятні значення помилки розпізнавання E показали метод з цензуруванням [13] ($E=0{,}070$) і модель V-Detector [14, 15] ($E=0{,}035$), що свідчить про недостатність покриття синтезованими детекторами області «своїх» екземплярів $S_1 \subseteq S$.

За результатами експериментів видно, що при використанні методу з цензуруванням [13] і моделі V-Detector [14, 15] генерується найбільша кількість детекторів N_{Ab} ($N_{Ab}=207$ і $N_{Ab}=41$, відповідно), що негативно впливає на час навчання t і витрати обчислювальних ресурсів комп'ютера. Метод з маскуванням детекторів [16] і запропонований метод синтезу продукційних правил на основі негативного відбору (при використанні вибірки $S_1 \subseteq S$) згенерували істотно меншу кількість детекторів ($N_{Ab}=19$ і $N_{Ab}=20$, відповідно), що свідчить про більш ефективну роботу цих методів. Зокрема, метод МСППНВ використовує апріорну інформацію про значущість ознак на початковому етапі і виключає з подальшого розгляду малозначущі і надлишкові ознаки, що дозволяє скоротити простір пошуку і створювати набір з невеликої кількості детекторів на основі високоінформативних ознак, що характеризується високими апроксимаційними й узагальнюючими здібностями.

Для аналізу узагальнюючих здібностей досліджуваних методів використовувалися критерії E_t, $P_{t,t_q=t'_1/t_q=t'_0}$ і $P_{t,t_q=t'_0/t_q=t'_1}$, що характеризують помилки розпізнавання й імовірності прийняття помилкових рішень на тестових даних. Як видно з табл. 1, помилки розпізнавання на тестових даних E_t у моделей, синтезованих за допомогою запропонованого методу МСППНВ, істотно нижче помилок моделей, побудованих за допомогою відомих методів [13]–[16] ($E_t=0{,}136$, $E_t=0{,}077$, $E_t=0{,}055$ для методів [13], [14, 15] і [16], відповідно). Це пояснюється використанням в якості критеріїв оцінювання кандидатів у детектори характеристик $G(Ab_k)$, що дозволяють оцінювати здатність детекторів до узагальнення даних. Запропонований метод синтезу продукційних правил на основі негативного відбору дозволив досягти рівнів помилки $E_t=0{,}037$ (при використанні частини вибірки $S_1 \subseteq S$) і $E_t=0{,}011$ (при використанні повної вибірки $S=<P,T>$).

Важливо відзначити, що в силу специфіки розв'язуваної задачі діагностування лопаток газотурбінних авіаційних двигунів дуже високу ціну має імовірність по-

Таблиця 1 – Результати експериментів з порівняння запропонованого методу з іншими методами негативного відбору

Метод	N_{Ab}	N_{it}	t, мс	E	$P_{t,t_q=t'_1/t_q=t'_0}$	$P_{t,t_q=t'_0/t_q=t'_1}$	E_t
Метод з цензуруванням [13]	207	50	27,3	0,070	0,126	0,333	0,136
Модель V-Detector [14, 15]	41	50	24,1	0,035	0,069	0,250	0,077
Метод з маскуванням детекторів [16]	19	14	13,2	0,018	0,054	0,083	0,055
Метод синтезу продукційних правил на основі негативного відбору МСППНВ (використовувалася вибірка $S_1 \subseteq S$)	20	12	12,1	0,026	0,038	0	0,037
Метод синтезу продукційних правил на основі негативного відбору МСППНВ (використовувалася вибірка $S=<P,T>$)	31	19	13,7	0,009	0,011	0	0,011

Таблиця 2 – Результати порівняння різних моделей

Модель	N_{param}	E	$P_{t,t_q=t'_1/t_q=t'_0}$	$P_{t,t_q=t'_0/t_q=t'_1}$	E_t
Модель у вигляді набору продукційних правил, синтезованих за допомогою запропонованого методу	652	0,009	0,011	0	0,011
Нейромережева модель прямого поширення	427	0,018	0,065	0,167	0,070
Модель у вигляді набору детекторів, побудована за допомогою методу негативного відбору з маскуванням детекторів [16]	804	0,018	0,054	0,083	0,055

милки $P_{t,t_q=t'_0/t_q=t'_1}$ віднесення до класу «чужих» ($t_q = t'_0$, придатних) за умови, що екземпляр реально відноситься до класу «своїх» ($t_q = t'_1$, дефектних). Це обумовлено тим, що віднесення дефектних лопаток авіадвигунів до класу придатних може коштувати людських життів. Як видно з табл. 1, запропонований метод МСППНВ, на відміну від існуючих, на тестових даних показав нульовий рівень імовірності помилки $P_{t,t_q=t'_0/t_q=t'_1}$, що свідчить про його високу ефективність для розв'язання подібних задач. Нульовий рівень імовірності помилки $P_{t,t_q=t'_0/t_q=t'_1}$ при використанні запропонованого методу пояснюється:

– високим рівнем покриття типових екземплярів класу $t_q = t'_1$ за допомогою згенерованого набору детекторів $AB = \{Ab_1, Ab_2, ..., Ab_{N_{Ab}}\}$, отриманого з використанням апріорної інформації про значущість ознак;

– високими узагальнюючими здібностями синтезованого набору детекторів, що обумовлено застосуванням в якості критерію оцінювання детекторів характеристик (6) і (7), що дозволяють оцінити здатність детекторів Ab_k до узагальнення даних.

Як видно з табл. 2, кількість параметрів N_{param} моделі у вигляді набору продукційних правил, синтезованих за допомогою запропонованого методу ($N_{param} = 652$), менше, ніж в аналогічній моделі, побудованої за допомогою методу негативного відбору з маскуванням детекторів [16] ($N_{param} = 804$). Це обумовлено тим, що при використанні запропонованого методу середній розмір згенерованих детекторів менше, оскільки в процесі негативного відбору використовується апріорна інформація про значущість ознак. Це дозволяє виявляти і виключати з подальшого розгляду малозначущі і надлишкові ознаки, що ускладнюють процес синтезу діагностичних моделей і знижують їх інтерпретабельність. Таким чином, порівняння значень різних критеріїв, представлених у табл. 2, показує, що модель у вигляді набору продукційних правил, синтезованих за допомогою запропонованого методу МСППНВ, є більш простою і зрозумілою в порівнянні з моделлю, створеною за допомогою методу [16]. Апроксимаційні й узагальнюючі здібності моделі, синтезованої на основі методу МСППНВ, також є більш високими, про що свідчать значення критеріїв E, E_t, $P_{t,t_q=t'_1/t_q=t'_0}$ і $P_{t,t_q=t'_0/t_q=t'_1}$.

Порівняння моделі на основі методу МСППНВ і нейромережевої моделі дозволяє зробити висновок про те, що модель, побудована за допомогою запропонованого методу, характеризується більш високими узагальнюючими й апроксимаційними здібностями (критерії E, E_t, $P_{t,t_q=t'_1/t_q=t'_0}$ і $P_{t,t_q=t'_0/t_q=t'_1}$). Однак, кількість параметрів N_{param} моделі на основі методу МСППНВ ($N_{param} = 652$) є дещо більшою, ніж у нейромережевій моделі ($N_{param} = 427$). Це пояснюється тим, що нейромережева модель представляється у вигляді множини нейронів, які пов'язані між собою певним чином і характеризуються ваговими коефіцієнтами як налагоджуваними параметрами. А кожен нейрон, по суті, відповідає деякій функції багатьох аргументів. При цьому така модель є досить складною для сприйняття людиною. Не дивлячись на більше значення критерію N_{param}, модель у виді набору продукційних правил, синтезованих за допомогою запропонованого методу, є більш інтерпретабельною у порівнянні з нейромережевою моделлю, оскільки продукційні правила вигляду «Якщо умова, то дія» є значно більш зрозумілими і зручними для сприйняття людиною, ніж набір коефіцієнтів, що відображаються ступінь зв'язків нейронів у нейромережевій моделі.

Таким чином, результати експериментів показали, що запропонований метод за рахунок використання апріорної інформації і виключення малозначущих і надлишкових ознак з вибірки дозволяє скорочувати простір пошуку і час виконання методу, а також синтезувати розпізнавальні моделі у вигляді набору детекторів з високими апроксимаційними й узагальнюючими здібностями. Крім того за рахунок підвищення узагальнюючих властивостей синтезованих моделей шляхом скорочення кількості детекторів і умов антецедентів підвищує інтерпретабельність моделі, скорочує її розмірність і, отже, обсяг використовуваної пам'яті.

ВИСНОВКИ

У роботі вирішено актуальне завдання автоматизації синтезу продукційних правил на основі негативного відбору для випадку нерівномірного розподілу екземплярів класів вибірки.

Наукова новизна роботи полягає в тому, що запропоновано метод синтезу продукційних правил на основі негативного відбору для випадку нерівномірного розподілу екземплярів класів вибірки, який при генерації набору детекторів використовує відому інформацію про екземпляри всіх класів вибірки, враховує інформацію про індивідуальну значущість ознак, як форму детектора використовує гіперкуб максимально можливого об'єму, що дозволяє виключати малозначущі і надлишкові ознаки з вибірки, скоротивши тим самим простір пошуку і час виконання методу, а також формувати набір детекторів з високими апроксимаційними й узагальнюючими здібностями. Запропонований метод за рахунок підвищення узагальнюючих властивостей синтезованих моделей шляхом скорочення числа детекторів і умов антецедентів також підвищує інтерпретабельність моделі, скорочує її розмірність (структурну і параметричну складність), обсяг використовуваної пам'яті і підвищує швидкодію моделі при послідовній реалізації обчислень.

Практична цінність отриманих результатів полягає в тому, що виконано експериментальне дослідження запропонованого методу і його порівняння з відомими аналогами, а також вирішено практичну задачу діагностування лопаток газотурбінних авіаційних двигунів.

Перспективи подальших досліджень полягають у застосуванні запропонованого підходу до видобування знань у вигляді набору продукційних правил з навчальних вибірок даних при синтезі нейро-нечітких моделей для вирішення практичних задач неруйнівного контролю якості.

ПОДЯКИ

Роботу виконано при частковій підтримці міжнародного проекту «Centers of Excellence for young RESearchers» (CERES) програми «Tempus» Європейської Комісії (реєстраційний номер 544137-TEMPUS-1-2013-1-SK-TEMPUS-JPHES).

СПИСОК ЛІТЕРАТУРИ

1. Ding S. X. Model-based fault diagnosis techniques: design schemes, algorithms, and tools / S. X. Ding. – Berlin : Springer, 2008. – 473 p. DOI: 10.1007/978-3-540-76304-8.
2. ASM handbook. – Vol. 17: Nondestructive evaluation and quality control. – Cleveland : ASM International, 1997. – 1607 p.
3. Diagnosis and fault-tolerant control / [M. Blanke, M. Kinnaert, J. Lunze, M. Staroswiecki]. – Berlin : Springer, 2006. – 672 p. DOI: 10.1007/978-3-662-05344-7.
4. Price C. Computer based diagnostic systems / C. Price. – London : Springer, 1999. – 136 p. DOI: 10.1007/978-1-4471-0535-0.
5. Denton T. Advanced automotive fault diagnosis / T. Denton. – London : Elsevier, 2006. – 271 p. DOI: 10.4324/9780080462585.
6. Ukil A. Intelligent Systems and Signal Processing in Power Engineering / A. Ukil. – Berlin : Springer, 2007. – 372 p. DOI: 10.1007/978-3-540-73170-2.
7. Ishida Y. Immunity-based systems: a design perspective / Y. Ishida. – Berlin : Springer, 2004. – 177 p. DOI: 10.1007/978-3-662-07863-1.
8. Segel L. A. Design principles for immune system and other distributed autonomous systems / L. A. Segel, I. R. Cohen. – New York : Oxford University Press, 2001. – 428 p.
9. Flower D. In silico immunology / D. Flower, J Timmis. – New York : Springer, 2007. – 451 p. DOI: 10.1007/978-0-387-39241-7.
10. A new cluster based real negative selection algorithm / W. Chen, T. Li, J. Qin [et al.] // Information and automation. – 2011. – Vol. 86. – P. 125–131. DOI: 10.1007/978-3-642-19853-3_18.
11. Esponda F. Negative representations of information / F. Esponda, S. Forrest, P. Helman // International Journal of Information Security. – 2009. – Vol. 8, № 5. – P. 331–345. DOI: 10.1007/s10207-009-0078-1.
12. Gonzalez F. The effect of binary matching rules in negative selection / F. Gonzalez, D. Dasgupta, J. Gomez // Genetic and Evolutionary Computation: conference GECCO-2003, Chicago, 9–11 July 2003: proceedings. – Berlin-Heidelberg : Springer-Verlag, 2003. – P. 195–206. DOI: 10.1007/3-540-45105-6_25.
13. Ong A. An adaptive anomaly detection system using data mining and artificial immune system / A. Ong. – London : King's College London, 2007. – 348 p.
14. Gonzalez F. A. Anomaly detection using real-valued negative selection / F. A. Gonzalez, D. Dasgupta // Journal of Genetic Programming and Evolvable Machines. – 2003. – Vol. 4, № 4. – P. 383-403. DOI: 10.1023/a:1026195112518.
15. Chmielewski A. Simple method of increasing the coverage of nonself region for negative selection algorithms / A. Chmielewski, S. T. Wierzchon // Computer Information Systems and Industrial Management Applications: 6th International Conference CISIM'07, Elk, 28–30 June 2007: proceedings. – Los Alamitos : IEEE Computer Society, 2007. – P. 155–160. DOI: 10.1109/cisim.2007.60.
16. Интеллектуальные информационные технологии проектирования автоматизированных систем диагностирования и распознавания образов : монография / [Субботин С. А., Олейник Ан. А., Гофман Е. А., Зайцев С. А., Олейник Ал. А.; под общ. ред. С. А. Субботина]. – Харьков : Компания СМИТ, 2012. – 318 с.
17. Clarke B. Principles and theory for data mining and machine learning / B. Clarke, E. Fokoue, H. H. Zhang. – New York : Springer, 2009. – 781 p. DOI: 10.1007/978-0-387-98135-2.
18. Bishop C.M. Pattern recognition and machine learning / C. M. Bishop. – New York : Springer, 2006. – 738 p. DOI: 10.1108/03684920710743466.
19. Analysis and design of intelligent systems using soft computing techniques / eds.: P. Melin, O. R. Castillo, E. G. Ramirez, J. Kacprzyk. – Heidelberg : Springer, 2007. – 855 p. DOI: 10.1007/978-3-540-72432-2.
20. Encyclopedia of machine learning / [eds. C. Sammut, G. I. Webb]. – New York : Springer, 2011. – 1031 p. DOI: 10.1007/978-0-387-30164-8.
21. Russel S. Artificial intelligence: a modern approach / S. Russel, P. Norvig. – New Jersey: Prentice Hall, 2009. – 1152 p.
22. Intelligent data analysis: an introduction / [eds. M. Berthold, D. J. Hand]. – New York: Springer Verlag, 2007. – 525 p.
23. Интеллектуальные средства диагностики и прогнозирования надежности авиадвигателей : монография / [Дубровин В. И., Субботин С. А., Богуслаев А. В., Яценко В. К.]. – Запорожье: ОАО «Мотор-Сич», 2003. – 279 с.

Стаття надійшла до редакції 09.01.2016.
Після доробки 26.01.2016.

Олейник А. А.

Канд. техн. наук, доцент, доцент кафедры программных средств, Запорожский национальный технический университет, Запорожье, Украина

ИЗВЛЕЧЕНИЕ ПРОДУКЦИОННЫХ ПРАВИЛ НА ОСНОВЕ НЕГАТИВНОГО ОТБОРА

Решена задача разработки математического обеспечения для автоматизации извлечения знаний в виде набора продукционных правил из обучающих выборок данных. Объектом исследования являлся процесс построения моделей неразрушающего контроля качества. Предмет исследования составляют методы извлечения продукционных правил на основе отрицательного отбора для синтеза моделей контроля качества. Цель работы: создание метода синтеза продукционных правил на основе множества детекторов, заключающегося в обработке данных обучающей выборки, характеризующейся существенным отличием числа экземпляров, относящихся к разных классам. Предложен метод синтеза продукционных правил на основе отрицательного отбора для случая неравномерного распределения экземпляров классов выборки, который при генерации набора детекторов использует известную информацию об экземплярах всех классов выборки, учитывает информацию об индивидуальной значимости признаков, в качестве формы детектора использует гиперкуб максимально возможного объема. Разработанный метод позволяет исключать малозначимые и избыточные признаки из выборки, сократив тем самым пространство поиска и время выполнения метода, а также формировать набор детекторов с высокими аппроксимационными и обобщающими способностями. Предложенный метод за счет повышения обобщающих свойств синтезируемых моделей путем сокращения числа детекторов и условий антецедентов также повышает интерпретабельность модели, сокращает ее размерность (структурную и параметрическую сложность), объем используемой памяти и повышает быстродействие модели при последовательной реализации вычислений. Разработано программное обеспечение, реализующее предложенный метод. Проведены эксперименты по исследованию свойств предложенного метода. Результаты экспериментов позволяют рекомендовать предложенный метод для использования на практике.

Ключевые слова: выборка, диагностирование, модель контроля качества, отрицательный отбор, продукционное правило.

Oliinyk A.

PhD., Associate Professor, Associate Professor of Department of Software Tools, Zaporizhzhya National Technical University, Zaporizhzhya, Ukraine

PRODUCTION RULES EXTRACTION BASED ON NEGATIVE SELECTION

The problem of mathematical support development is solved to automate the extraction knowledge as production rules from the training data samples. The object of study is the process of constructing models of non-destructive quality control. The subject of study are methods of production rules extraction based on negative selection for synthesis of quality control models. The purpose of the work is to develop a method of production rules synthesis on the basis of a set of detectors is in the handling of data of training sample, characterized by a substantial number of instances of distinction belonging to different classes. A method for the synthesis of production rules on the basis of negative selection in the case of uneven distribution of instances of the sample classes is proposed. The developed method allows to exclude irrelevant and redundant features from the sample, thereby reducing the search space and time of execution of the method, as well as generate a set of detectors with high approximation and generalization capability. The proposed method improves the generalizing properties of synthesized model and its interpretability. The software implementing proposed method is developed. The experiments to study the properties of the proposed method are conducted. The experimental results allow to recommend the proposed method for use in practice.

Keywords: sample, diagnostics, model of quality control, negative selection, production rule.

REFERENCES

1. Ding S. X. Model-based fault diagnosis techniques: design schemes, algorithms, and tools. Berlin, Springer, 2008, 473 p. DOI: 10.1007/978-3-540-76304-8.
2. ASM handbook. Vol. 17: Nondestructive evaluation and quality control. Cleveland, ASM International, 1997, 1607 p.
3. Blanke M., Kinnaert M., Lunze J., Staroswiecki M. Diagnosis and fault-tolerant control. Berlin, Springer, 2006, 672 p. DOI: 10.1007/978-3-662-05344-7.
4. Price C. Computer based diagnostic systems. London, Springer, 1999, 136 p. DOI: 10.1007/978-1-4471-0535-0.
5. Denton T. Advanced automotive fault diagnosis. London, Elsevier, 2006, 271 p. DOI: 10.4324/9780080462585.
6. Ukil A. Intelligent Systems and Signal Processing in Power Engineering. Berlin, Springer, 2007, 372 p. DOI: 10.1007/978-3-540-73170-2.
7. Ishida Y. Immunity-based systems: a design perspective. Berlin, Springer, 2004, 177 p. DOI: 10.1007/978-3-662-07863-1.
8. Segel L.A., Cohen I. R. Design principles for immune system and other distributed autonomous systems. New York, Oxford University Press, 2001, 428 p.
9. Flower D., Timmis J. In silico immunology. New York, Springer, 2007, 451 p. DOI: 10.1007/978-0-387-39241-7.
10. Chen W., Li T., Qin J. [et al.] A new cluster based real negative selection algorithm, *Information and automation*, 2011, Vol. 86, pp. 125–131. DOI: 10.1007/978-3-642-19853-3_18.
11. Esponda F., Forrest S., Helman P. Negative representations of information, *International Journal of Information Security*, 2009, Vol. 8, No. 5, pp. 331–345. DOI: 10.1007/s10207-009-0078-1.
12. Gonzalez F., Dasgupta D., Gomez J. The effect of binary matching rules in negative selection, *Genetic and Evolutionary Computation: conference GECCO-2003, Chicago, 9–11 July 2003: proceedings*. Berlin-Heidelberg, Springer-Verlag, 2003, pp. 195–206. DOI: 10.1007/3-540-45105-6_25.
13. Ong A. An adaptive anomaly detection system using data mining and artificial immune system. London, King's College London, 2007, 348 p.
14. Gonzalez F. A., Dasgupta D. Anomaly detection using real-valued negative selection, *Journal of Genetic Programming and Evolvable Machines*, 2003, Vol. 4, No. 4, pp. 383–403. DOI: 10.1023/a:1026195112518.
15. Chmielewski A., Wierzchon S. T. Simple method of increasing the coverage of nonself region for negative selection algorithms, *Computer Information Systems and Industrial Management Applications: 6th International Conference CISIM'07, Elk, 28–30 June 2007: proceedings*. Los Alamitos, IEEE Computer Society, 2007, pp. 155–160. DOI: 10.1109/cisim.2007.60.
16. Subbotin S. A., Olejnik An. A., Gofman E. A., Zajcev S. A., Olejnik Al. A.; pod obshh. red. S. A. Subbotina Intellektual'nye informacionnye tehnologii proektirovanija avtomatizirovannyh sistem diagnostirovanija i raspoznavanija obrazov: monografija. Har'kov, Kompanija SMIT, 2012, 318 p.
17. Clarke B. Fokoue E., Zhang H. H. Principles and theory for data mining and machine learning. New York, Springer, 2009, 781 p. DOI: 10.1007/978-0-387-98135-2.
18. Bishop C. M. Pattern recognition and machine learning. New York, Springer, 2006, 738 p. DOI: 10.1108/03684920710743466.
19. eds.: Melin P., Castillo O. R., Ramirez E. G., Kacprzyk J. Analysis and design of intelligent systems using soft computing techniques. Heidelberg, Springer, 2007, 855 p. DOI: 10.1007/978-3-540-72432-2.
20. eds. Sammut C., Webb G. I. Encyclopedia of machine learning. New York, Springer, 2011, 1031 p. DOI: 10.1007/978-0-387-30164-8.
21. Russel S., Norvig P. Artificial intelligence: a modern approach. New Jersey, Prentice Hall, 2009, 1152 p.
22. eds. Berthold M., Hand D. J. Intelligent data analysis: an introduction. New York, Springer Verlag, 2007, 525 p.
23. Dubrovin V. I., Subbotin S. A., Boguslaev A. V., Jacenko V. K. Intellektual'nye sredstva diagnostiki i prognozirovanija nadezhnosti aviadvigatelej : monografija. Zaporozh'e, OAO «Motor-Sich», 2003, 279 p.

УДК 004.93

Субботин С. А.

Д-р техн. наук, профессор, профессор кафедры программных средств Запорожского национального технического университета, Запорожье, Украина

СИНТЕЗ НЕЙРО-НЕЧЕТКИХ СЕТЕЙ С РАНЖИРОВАНИЕМ И СПЕЦИФИЧЕСКИМ КОДИРОВАНИЕМ ПРИЗНАКОВ ДЛЯ ДИАГНОСТИКИ И АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ ПО ПРЕЦЕДЕНТАМ

Решена задача автоматизации синтеза нейро-нечетких сетей для диагностирования и автоматической классификации по признакам. Предложен метод синтеза нейро-нечетких моделей по прецедентам, который определяет оценки взаимосвязи входных признаков, выделяет взаимосвязанные признаки, оценивает влияние признаков на выходной признак, а также выявляет взаимосвязанные термы, включает в модель наиболее важные признаки и термы, устраняет дубляж термов и признаков, использует различные варианты кодирования сигналов, а также переупорядочивает признаки, обеспечивая группировку признаков, формирует правила для прецедентов ранее не встречавшихся классов или существенно отличающихся от имеющихся правил своего класса, а уже имеющиеся правила корректирует на основе поступающих прецедентов с учетом числа ранее рассмотренных наблюдений. Предложенный метод позволяет существенно ускорить синтез нейро-нечетких моделей, обеспечивая приемлемую точность и более высокий уровень обобщения данных, снизить сложность и избыточность, а также повысить интерпретабельность нейромодели. Проведены эксперименты по решению практических задач диагностирования и автоматической классификации, подтвердившие работоспособность и применимость предложенного метода. Получена зависимость ошибки модели, синтезированной предложенным методом, от заданной разрядности признаков. Использование полученной зависимости позволяет на практике более рационально выбирать значение числа интервалов разбиения диапазонов значений признаков, обеспечивая приемлемую точность нейромодели.

Ключевые слова: выборка, нейро-нечеткая сеть, нечеткий вывод, обучение по прецедентам, диагностирование.

НОМЕНКЛАТУРА

ЭВМ – электронная вычислительная машина;

δ_j – длина интервала, на которые разбивается диапазон значений j-го признака;

μ_q – принадлежность к q-му классу;

$\mu^{q,k}$ – принадлежность экземпляра к k-му эталону (правилу) q-го класса;

$\mu_{q,k,i}$ – принадлежность экземпляра к k-му эталону (правилу) q-го класса по i-му признаку;

$\mu_{q,k,i,l}$ – принадлежность экземпляра к k-му эталону (правилу) q-го класса по l-му интервалу i-го признака;

$\mu_{i,l}(x^s)$ – функции принадлежности к l-му терму i-го признака;

$\omega_{q,k}$ – нормированный вес k-го правила q-го класса;

a, b, c, d – параметры функции принадлежности;

C – набор правил-эталонов;

$C^{(q,\,k)}$ – k-й эталон (правило) q-го класса;

$\overline{E}$ – ошибка модели;

E – ошибка сети для задач с дискретным выходом;

$E_{\text{об.}}$ – ошибка модели для обучающей выборки;

$E_{\text{тест.}}$ – ошибка модели для тестовой выборки;

f – критерий качества модели;

$F()$ – структура нейромодели;

$I^*(q,i.l)$ – значимость l-го интервала i-го признака относительно q-го класса;

i, j – номера признаков;

I_G – коэффициент обобщения нейромодели;

I_i – информативность j-го признака;

$I_{i,l}$ – информативность l-го интервала i-го признака;

I_w – число настраиваемых параметров нейромодели;

k – номер правила (эталона);

K – число классов;

l – номер разряда (интервала);

L – размерность кодирующего бинарного вектора – элемента сенсорной матрицы;

N – число входных признаков;

N' – число входных признаков, после отбора признаков;

opt – условное обозначение оптимума;

q – номер класса;

Q^q – число правил (эталонов) q-го класса в наборе C;

$R(x^s,C^{(q,k)})$ – расстояние от экземпляра x^s до k-го эталона q-го класса $C^{(q,k)}$;

$r_{i,j}$ – мера взаимосвязи входных признаков x_i, и x_j;

$r_{i,y}$ – мера важности для каждого входного признака относительно выходного признака;

S – объем выборки;

$S_{\text{об.}}$ – число экземпляров в обучающей выборке;

$S_{\text{тест.}}$ – число экземпляров в тестовой выборке;

$t_{\text{об.}}$ – время, затраченное на обучение неро-нечеткой модели;

w – параметры нейромодели;

$w^{(q,k)}$ – вес k-го правила q-го класса;

x – набор входных векторов прецедентов;

X – набор прецедентов;

x_j – j-й входной признак;

x^s – набор входных признаков s-го прецедента;

x^s_j – значение j-го входного s-го прецедента;

y – выходной признак;

$x_j^{\min}$, $x_j^{\max}$ – минимальное и максимальное значения признака x_j;

DOI 10.15588/1607-3274-2016-1-6

$x^s_{j,l}$ – значение l-го разряда j-го признака s-го экземпляра в специфическом кодовом представлении;

y^s – значение выходного признака для s-го прецедента.

ВВЕДЕНИЕ

Для обеспечения и поддержания высокого уровня качества, надежности и долговечности выпускаемых и эксплуатируемых изделий промышленного производства необходимо своевременно осуществлять их диагностирование [1].

Помимо методов физических измерений [1, 2], специфических для каждой конкретной задачи, ключевым элементом процесса диагностирования является диагностическая модель, построение которой требует либо полного знания физики диагностируемого объекта, либо привлечение человека-эксперта, либо использование методов машинного обучения [3].

Построение физических моделей [2], как правило, возможно только для хорошо изученных объектов диагностирования и специфически зависит от конкретной задачи. Такие методы непригодны для более широко встречающихся случаев, когда физика диагностируемого объекта является не полностью изученной, а сам объект подвергается воздействию внешних факторов [3].

Очевидными недостатками экспертных методов [4] является их чрезвычайная зависимость от субъективных знаний эксперта, его компетентности и ответственности, сложность извлечения и формализации знаний, а также сложность модификации и адаптации полученной модели принятия решений с учетом возможных изменений диагностируемого объекта, внешней среды.

Методы машинного обучения [5, 6], в отличие от методов построения физических моделей, не требуют знания физики диагностируемого объекта, а, в отличие от экспертных методов [4], не требуют привлечения человека-эксперта и способны автоматически извлекать знания из данных в процессе построения модели [3]. Поэтому данные методы целесообразно выбрать в качестве базиса для построения диагностических моделей.

Среди методов машинного обучения [5, 6] широкое распространение на практике в последнее время получили нейро-нечеткие сети [4, 7], которые способны обучаясь по прецедентам извлекать знания из данных, а также могут интегрировать в свою структуру экспертные знания и являются преобразуемыми в понятные для человека продукционные правила вида «если, то» [4].

Однако большинство известных методов построения нейро-нечетких моделей [4, 7] предполагают вовлечение пользователя в процесс создания сети и решение задач многомерной нелинейной оптимизации, характеризующихся неопределенностью выбора начальной точки поиска. Полученные диагностические модели в результате оказываются избыточными [3].

Объектом исследования данной работы является процесс построения диагностических моделей по прецедентам.

Предметом исследования данной работы являлись методы построения нейро-нечетких сетей для решения задач диагностирования.

Целью работы являлось создание метода синтеза диагностических моделей на основе нейро-нечетких сетей, свободного от указанных выше недостатков.

1 ПОСТАНОВКА ЗАДАЧИ

Пусть мы имеем исходную выборку $X = \langle x, y\rangle$ – набор S прецедентов о зависимости $y(x)$, $x = \{x^s\}$, $y=\{y^s\}$, $s = 1, 2, ..., S$, характеризующихся набором N входных признаков $\{x_j\}$, $j = 1, 2, ..., N$, и выходным признаком y. Каждый s-й прецедент представим как $\langle x^s, y^s\rangle$, $x^s=\{x^s_j\}$, $y^s \in \{1, 2, ..., K\}$, $K>1$.

Тогда задача синтеза нейромодели зависимости $y(x)$ будет заключаться в определении таких структуры $F()$ и значений параметров w нейромодели, при которых будет удовлетворен критерий качества модели $f(F(), w, \langle x, y\rangle) \to opt$ [3]. Обычно критерий качества обучения нейромоделей определяют как функцию ошибки модели:

$$\overline{E} = \frac{1}{2}\sum_{s=1}^{S}(y^s - F(w, x^s))^2 \to \min.$$

Для задач с дискретным выходом ошибку обученной сети можно характеризовать также формулой:

$$E = \frac{100\%}{S}\sum_{s=1}^{S}\left|y^s - F(w, x^s)\right| \to \min.$$

2 ЛИТЕРАТУРНЫЙ ОБЗОР

В настоящее время известное большое число различных архитектур нейро-нечетких сетей [4, 7]:

– нечеткие нейронные системы (fuzzy neural systems): в нейронных сетях применяются принципы нечеткой логики для ускорения процесса настойки или улучшения других параметров; нечеткая логика является лишь инструментом нейронных сетей и такая система не может быть интерпретирована в нечеткие правила, поскольку представляет собой «черный ящик»;

– конкурирующие нейро-нечеткие системы (concurrent neuro-fuzzy systems): нечеткая система и нейронная сеть работают над одной задачей, не влияя на параметры друг друга;

– параллельные нейро-нечеткие системы (cooperative neuro-fuzzy systems): настройка параметров выполняется с помощью нейронных сетей, после чего нечеткая система функционирует самостоятельно;

– интегрированные (гибридные) нейро-нечеткие системы (integrated neuro-fuzzy systems) – системы с тесным взаимодействием нечеткой логики и нейронных сетей. Данный тип нейро-нечетких сетей является наиболее подходящим для решения задач диагностирования.

Наиболее широко используемой на практике архитектурой интегрированных нейро-нечетких сетей являются сети Мамдани [8, 9], представляющие собой нейро-нечеткий классификатор. Однако методы их построения предполагают вовлечение пользователя в процесс создания сети (задание числа термов, видов их функций принадлежности, начальных значений параметров функций принадлежности термов, правил принятия решений), либо требуют решения задач кластер-анализа [10] набора прецедентов с последующей настройкой весов на основе градиентных методов оптимизации [11] с расчетом частных производных целевой функции на основе техники обратного распространения ошибки [4, 5].

Полученные диагностические модели в результате оказываются неоптимальными по структуре (зачастую избыточными), и, как следствие, оказываются менее точными и более сложными для восприятия и анализа человеком, что несколько осложняет их практическое применение [3].

Поэтому необходимо разработать метод, позволяющий получать диагностические модели, не требующие участия человека в процессе синтеза структуры и настройки параметров нейромодели, а также позволяющий снизить сложность получаемых моделей и повысить их интерпретабельность и обобщающие свойства.

3 МАТЕРИАЛЫ И МЕТОДЫ

Избыточность нейро-нечетких моделей, как правило, обусловлена тем, что набор используемых признаков является избыточным, а нечеткие термы формируются субъективно пользователем и могут излишне детализировать разбиение пространства признаков. Поэтому для снижения избыточности нейро-нечетких моделей предлагается в процессе синтеза сети определять оценки взаимосвязи входных признаков между собой, что позволит выделить взаимосвязанные признаки, оценивать влияние признаков на выходной признак (класс), а также выявлять взаимосвязанные термы, что позволит включать в модель наиболее важные признаки и термы, устранить дубляж термов и признаков путем оставления лишь части из них.

Для повышения вариативности (потенциала получения возможных решений) синтезируемых моделей предлагается использовать различные варианты кодирования сигналов, а также переупорядочивать признаки, обеспечивая тем самым группировку признаков по подобию с целью получения более легко воспринимаемых человеком диагностических моделей.

Для автоматизации формирования термов и выделения правил на основе заданного набора прецедентов предлагается реализовать процедуру адаптивного четкого кластер-анализа прецедентов, в процессе реализации которой формировать новые эталоны для прецедентов ранее не встречавшихся классов или существенно отличающихся от имеющихся эталонов своего класса, а уже имеющиеся эталоны корректировать на основе поступающих прецедентов с учетом числа ранее рассмотренных наблюдений.

На основе приведенных идей более формально изложим предлагаемый метод.

Этап инициализации. Задать обучающую выборку $<x, y>$. Определить минимальное и максимальное значения признаков:

$$x_j^{\min} = \min_{s=1,2,...,S}\{x_j^s\},\ x_j^{\max} = \max_{s=1,2,...,S}\{x_j^s\}.$$

Задать размерность кодирующего бинарного вектора – элемента сенсорной матрицы L. Очевидно, что должно быть $L>1$.

Этап кодирования сигналов. Заменить значение каждого признака в выборке бинарным кодом. Для этого возможно использовать интервальное или колоночное кодирование.

Вначале на основе заданного L определим длину интервала каждого j-го признака:

$$\delta_j = \frac{x_j^{\max} - x_j^{\min}}{L},\ j = 1, 2,...,N.$$

При интервальном кодировании диапазон значений каждого признака разбивается на заданное число интервалов равной длины, а значение признака, заменяется на бинарный вектор, элементы которого (разряды) соответствуют интервалам значений признака. В процессе кодирования определяется номер интервала, в который попало значение признака для конкретного экземпляра, после чего соответствующий элемент кодирующего вектора устанавливается равным единице, а остальные элементы – равными нулю:

$$x_{j,l}^s = \begin{cases} 0, l\delta_j < x_j^s - x_j^{\min}; \\ 1, (l-1)\delta_j \le x_j^s - x_j^{\min} \le l\delta_j; \\ 0, (l-1)\delta_j > x_j^s - x_j^{\min}, \end{cases}$$

$$j = 1, 2, ..., N, l = 1, 2, ..., L.$$

При колоночном кодировании диапазон значений каждого признака также разбивается на заданное число интервалов равной длины, значение признака, заменяется на бинарный вектор, элементы которого (разряды) соответствуют интервалам значений признака и определяется номер интервала, в который попало значение признака для конкретного экземпляра, однако все элементы кодирующего вектора, соответствующие данному интервалу и предыдущим ему интервалам устанавливается равным единице, а остальные элементы, соответствующие интервалам больших значений признака – равными нулю:

$$x_{j,l}^s = \begin{cases} 1, l\delta_j \le x_j^s - x_j^{\min}; \\ 0, l\delta_j > x_j^s - x_j^{\min}, \end{cases}$$

$$j = 1, 2, ..., N, l = 1, 2, ..., L.$$

Этап выделения правил. Каждый экземпляр выборки x^s в выбранном кодовом представлении $x^s = \{x_{j,l}^s\}$ представить в виде модифицированного продукционного правила $\{x_{i,l}^s\} \to y^s$.

Этап обобщения правил. Выполнить динамическое формирование правил-эталонов кластеров.

Для этого в начале задать пустой набор правил-эталонов $C = \varnothing$.

Затем просматривая экземпляры выборки (s=1,2,...,S) для текущего s-го экземпляра определить, имеется ли эталон его класса в наборе эталонов.

Если эталон для его класса отсутствует в C, то принять q=y^s, добавить эталон и занести в него s-й экземпляр: $C = C \cup C^{(q,1)}$, $C^{(q,1)}$=x^s: $C_{j,l}^{(q,1)} = \{x_{j,l}^s\}$, j=1, 2, ..., N; l=1, 2, ..., L, $w^{(q,1)}$=1.

В противном случае, если эталон класса для текущего s-го экземпляра отсутствует в C, то определить расстояние от экземпляра x^s до всех эталонов $R(x^s,C^{(q,k)})$, q=1, 2, ..., Q; k=1, 2, ..., Q^q:

$$R(x^s,C^{(q,k)})=\sqrt{\sum_{j=1}^{N}\sum_{l=1}^{L}\left(x_{j,l}^s-C_{j,l}^{(q,k)}\right)^2}\ .$$

Если расстояние от экземпляра x^s до ближайшего к нему эталона чужого класса меньше, чем расстояние от экземпляра до ближайшего эталона своего класса, т.е.

$$\min_{\substack{y^s\neq q,\\ k=1,2,..,Q^q}}\{R(x^s,C^{(q,k)})\}<\min_{\substack{y^s=q,\\ k=1,2,..,Q^q}}\{R(x^s,C^{(q,k)})\},$$

тогда сформировать новый эталон и занести в него экземпляр: $q=y^s$, $k=Q^q+1$, $C=C\cup C^{(q,k)}$, $C^{(q,\ k)}=x^s$: $C^{(q,\ k)}=\{x_{j,l}^s\}$, $w^{(q,k)}=1$; в противном случае – новый эталон получить как среднее между эталоном и экземпляром $C^{(q,\ k)}=(w^{(q,k)}C^{(q,\ k)}+x^s)/(w^{(q,k)}+1)$:

$$C_{j,l}^{(q,k)}=\frac{w^{(q,k)}C_{j,l}^{(q,k)}+x_{j,l}^s}{w^{(q,k)}+1},$$

j=1, 2, ..., N; l=1, 2, ..., L,
после чего принять: $w^{(q,k)}=w^{(q,k)}+1$.

Этап ранжирования кодовых представлений сигналов. При построении и анализе диагностических и распознающих моделей желательно выявить и понимать взаимосвязь признаков, а также зависимость принимаемых решений от описательных признаков. Для этого предлагается определить ряд показателей.

Определим меру взаимосвязи входных признаков относительно интервалов (сенсоров) одного уровня, используя формулу:

$$r_{i,j}=r_{j,i}=\frac{\sum_{s=1}^{S}\sum_{l=1}^{L}\sum_{t=1}^{L}\{1\mid x_{j,l}^s=x_{i,t}^s\}}{SL^2},\ i,j=1,2,...,N.$$

Признаки будут тем сильнее взаимосвязанными, чем больше экземпляров попадает в и интервалы (сенсоры) одного уровня.

Определим меру важности для каждого входного признака относительно выходного признака как

$$r_{i,y}=\frac{\sum_{s=1}^{S}\sum_{l=1}^{L}\sum_{t=1}^{K}\{1\mid x_{j,l}^s=1,y^s=t\}}{SLK},\ i=1,2,...,N.$$

Входной и выходной признаки будут тем сильнее взаимосвязанными, чем больше экземпляров попадает в и интервалы (сенсоры) одного класса.

После определения мер близости признаков ранжируем и перенумеруем признаки. В качестве первого признака целесообразно выбрать признак с наибольшим значением $r_{i,y}$. Затем в новый набор включим последовательно признаки, наиболее тесно связанные с предыдущим признаками, в порядке убывания значений $r_{i,j}$. Данный этап можно пропустить для стационарных сигналов, распределенных во времени.

Этап оценки значимости интервалов и признаков. Признаки и их интервалы значений могут иметь различную ценность при принятии решений.

При этом для каждого класса значимость конкретного интервала признака определяется тем, сколько экземпляров, попавших в этот интервал будет принадлежать к данному классу и тем, сколько экземпляров, попавших в этот интервал не будет принадлежать к данному классу. Для определения значимости интервалов относительно классов предлагается использовать показатель, выражаемый формулой:

$$I^*(q,i,l)=2I(q,i,l)-\sum_{p=1}^{K}I(p,i,l),$$

где $I(q,i,l)=\sum_{s=1}^{S}\sum_{i=1}^{N}\sum_{l=1}^{L}\left\{x_{i,l}^s\mid y^s=q\right\}, q=1,2,...,K; i=1,2,...,N;$

l=1, 2, ..., L.

На основе введенных показателей можно определить

– показатель информативности l-го интервала i-го признака:

$$I_{i,l}=\frac{1}{K}\sum_{l=1}^{K}I^*(q,i,l)\ ,\ i=1,2,...,N;\ l=1,2,...,L;$$

– показатель информативности j-го признака:

$$I_i=I_{i,l}=\frac{1}{L}\sum_{l=1}^{L}I_{i,l}=\frac{1}{KL}\sum_{l=1}^{L}\sum_{q=1}^{K}I^*(q,i,l),\ i=1,2,...,N.$$

Этап удаления малозначимых интервалов значений признаков. Все интервалы признаков, у которых значение $I_{i,l}$ =0, можно считать малозначимыми и удалить из дальнейшего рассмотрения. Среди оставшихся интервалов можно рассматривать как малозначимые те, для которых

$$I_{i,l}\le\left\lceil\frac{ES}{L}\right\rceil,$$

где E – заданное заранее значение максимально допустимой ошибки.

Этап удаления малозначимых признаков. Если среди признаков имеется такой, что все его интервалы признаны малозначимыми, то данный признак следует ислючить из рассмотрения.

Этап формирования термов признаков. На основе выделенных интервалов значений признаков возможно определить функции принадлежности к термам признаков на основе трапециевидной функции ($a\le b\le c\le d$):

$$\mu_{i,l}(x^s)=\begin{cases}0, x\le a;\\ \dfrac{x-a}{b-a}, a\le x\le b;\\ 1, b\le x\le c;\\ \dfrac{d-x}{d-c}, c\le x\le d\\ 0, d\le x.\end{cases}$$

При интервальном кодировании параметры функции принадлежности к l-му терму i-го признака будут инициализироваться по формулам:

$$a=(l-1)\delta+x_i^{\min},\ b=(l-1)\delta+x_i^{\min},\ c=l\delta+x_i^{\min},$$

$$d=l\delta+x_i^{\min}.$$

При колоночном кодировании параметры функции принадлежности к l-му терму i-го признака будут инициализироваться по формулам:

$$a=x_i^{\min},\ b=x_i^{\min},\ c=l\delta+x_i^{\min},\ d=l\delta+x_i^{\min}.$$

Этап формирования системы нечеткого вывода. На данном этапе задаются принципы преобразования значений функций принадлежности к термам в принадлежности к кластерам и классам, а также способ дефаззификации результата.

Принадлежности к сформированным эталонам определим по формулам:

$$\mu^{q,k}=\max_{i=1,2,\ldots,N}\{\mu_{q,k,i}\}\ \text{либо}\ \mu^{q,k}=\max_{i=1,2,\ldots,N}\left\{\frac{I_i\mu_{q,k,i}}{\sum_{j=1}^{N}I_j}\right\},$$

где

$$\mu_{q,k,i}=\max_{l=1,2,\ldots,L}\{\mu_{q,k,i,l}\}\ \text{либо}$$

$$\mu_{q,k,i}=\max_{l=1,2,\ldots,L}\left\{\frac{I_{i,l}\mu_{q,k,i,l}}{\sum_{p=1}^{L}I_{i,p}}\right\},$$

$$\mu_{q,k,i,l}=\mu_{i,l}(x^s)C_{i,l}^{(q,k)},$$

q=1, 2, ..., K, k = 1, 2, ..., Q^q, i = 1, 2, ..., N; l=1, 2, ..., L.

Принадлежности к классам определим формулами:

$$\mu_q=\max_{k=1,2,\ldots,Q^q}\left\{\omega_{q,k}\mu^{q,k}\right\},$$

где $\omega_{q,k}=\dfrac{w^{(q,k)}}{\sum_{p=1}^{Q^q}w^{(q,p)}}$.

Номер класса, соответственно, будет определяться по формуле:

$$y^s=\arg\max_{q=1,2,\ldots,K}\left\{\mu_q\right\}\ \text{или}\ y^s=\left\lceil\frac{\sum_{q=1}^{K}q\mu_q}{\sum_{q=1}^{K}\mu_q}\right\rceil.$$

Этап синтеза нейро-нечеткой сети. На основе выделенных термов и определенных функций системы нечеткого вывода можно сформировать структуру нейро-нечеткой сети в соответствии со схемой, приведенной на рис. 1.

На входной (первый) слой сети поступают входные сигналы – значения признаков распознаваемого экземпляра, которые далее поступают на второй слой, выполняющий фаззификацию. Выходы нейронов второго слоя представляют собой значения функций принадлежности распознаваемого экземпляра к термам признаков. Нейроэлементы третьего слоя сети комбинируют принадлежности к термам в принадлежности к эталонам кластеров, на основе которых нейроны четвертого слоя определяют принадлежности к классам. Единственный нейрон пятого (выходного) слоя осуществляет дефаззификацию и выдает на выходе расчетное значение номера класса распознаваемого экземпляра.

Этап настройки весов. На основе обучающей выборки оценить ошибку для построенной сети. Если ошибка является приемлемой, то завершить синтез сети, в противном случае выполнить процедуру обучения посредством решения задачи многомерной оптимизации параметров нечетких термов второго слоя, используя метод обратного распространения ошибки [4, 11] или эволюционные алгоритмы [3, 5].

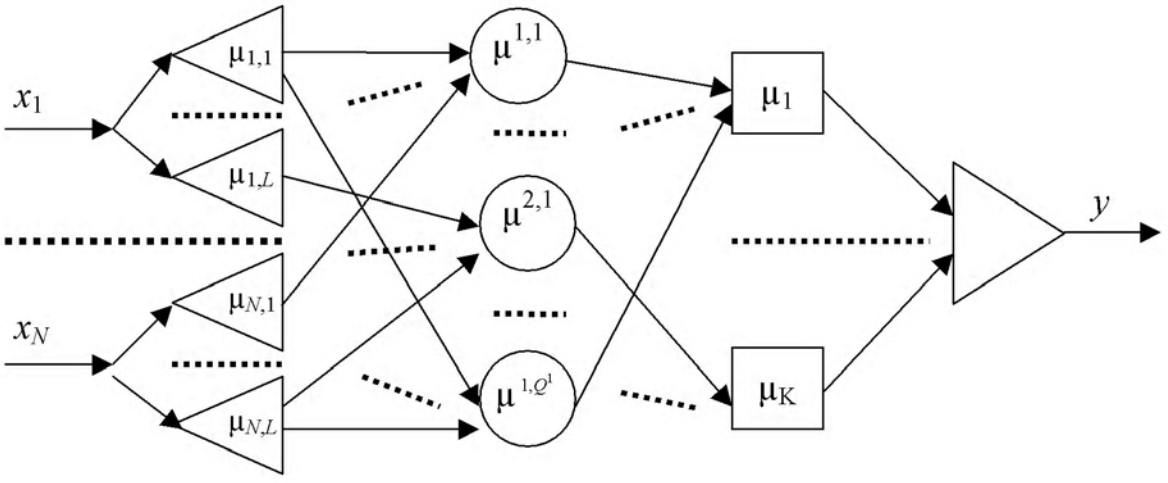


Рисунок 1 – Схема нейро-нечеткой сети

Если в результате обучения достигнут требуемый уровень точности (приемлемая ошибка), то завершить синтез сети, в противном случае – увеличить число интервалов L и перейти к этапу кодирования. Поскольку число интервалов L не должно превышать числа экземпляров в выборке S, то при получении на данном этапе $L \geq S$ следует прекратить дальнейший поиск и вернуть полученную модель с наилучшей точностью и наименьшим L.

4 ЭКСПЕРИМЕНТЫ

Для исследования практической применимости предложенного метода он был программно реализован как дополнение к Автоматизированной системе синтеза нейросетевых и нейро-нечетких моделей для неразрушающей диагностики и классификации образов по признакам [12].

Разработанное программное обеспечение использовалось для решения задач неразрушающего диагностирования и автоматической классификации, характеристики которых представлены в табл. 1.

Для каждой задачи на основе кластер анализа человеком выделялись нечеткие термы и правила, по которым определялась структура нейро-нечетких моделей, после чего на основе обратного распространения ошибки выполнялась настройка параметров моделей. Также для каждой задачи на основе выборки данных в автоматическом режиме на основе предложенного метода синтезировались нейро-нечеткие модели.

5 РЕЗУЛЬТАТЫ

Результаты проведенных вычислительных экспериментов представлены в табл. 2.

Здесь коэффициент обобщения определялся по формуле:

$$I_G = \frac{NS}{I_w},\ I_w \geq 1.$$

Чем больше значение коэффициента обобщения, тем выше уровень обобщения обучающей выборки данных нейро-нечеткой моделью.

С целью изучения зависимости ошибки модели E от заданной разрядности признаков L для различных задач (выборок данных различной размерности и природы [3]) получен график, приведенный на рис. 2.

На рис. 3 представлен график усредненной зависимости ошибки модели E от доли отобранных признаков в первоначальном наборе признаков N'/N для различных задач (выборок данных различной размерности и природы [3]).

На рис. 4 представлен график усредненной зависимости коэффициента обобщения модели I_G от доли отобранных признаков в первоначальном наборе признаков N'/N для различных задач (выборок данных различной размерности и природы [3]).

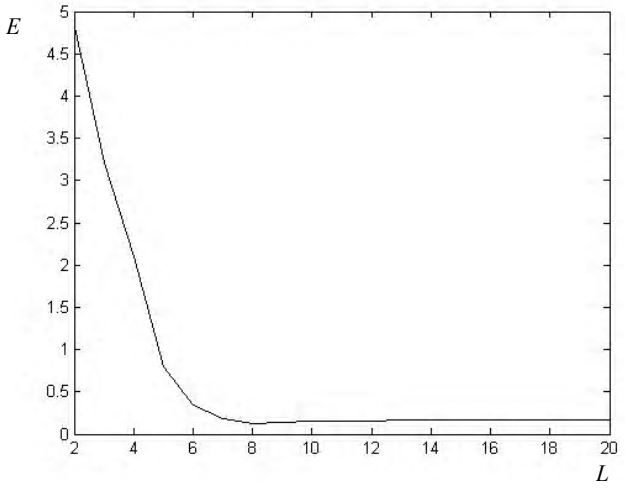


Рисунок 2 – График усредненной зависимости E от L

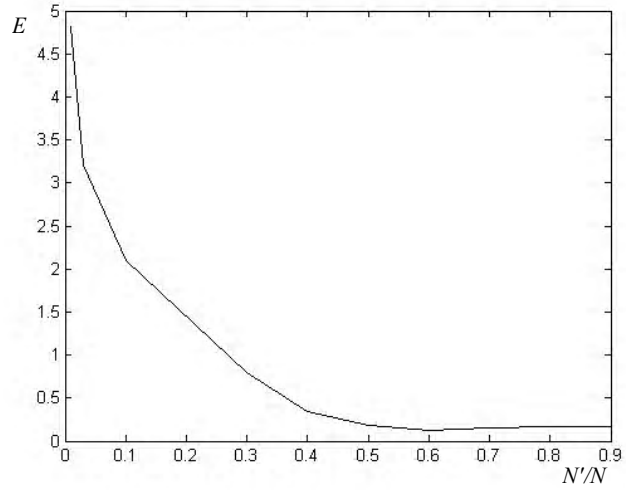


Рисунок 3 – График усредненной зависимости E от N'/N

Таблица 1 – Характеристики практических задач диагностирования и автоматической классификации

Название задачи	N	$S_{об.}$	$S_{тест.}$	K
Диагностирование лопаток газотурбинных авиадвигателей по спектрам свободных затухающих колебаний после ударного возбуждения [3]	100	16	16	2
Автоматическая классификация растительных объектов по коэффициентам спектральной яркости [13]	55	120	128	2
Диагностирование хронического обструктивного бронхита [14]	28	102	103	2

Таблица 2 – Результаты вычислительных экспериментов

Название задачи	Сеть Мамдани, заданная человеком и обученная методом обратного распространения ошибки				Сеть Мамдани, синтезированная на основе предложенного метода			
	$E_{об.}$, %	$E_{тест.}$, %	$t_{об.}$, с	I_G	E	$E_{тест.}$	$t_{об.}$, с	I_G
Диагностирование лопаток газотурбинных авиадвигателей [3]	0	0	13,92	0,8	0	0	12,75	1,33
Автоматическая классификация растительных объектов [13]	1,7	2,3	23,14	4,29	0	0,8	21,89	5
Диагностирование хронического обструктивного бронхита [14]	4,9	2,9	18,85	2,83	0	0,9	17,14	3,19

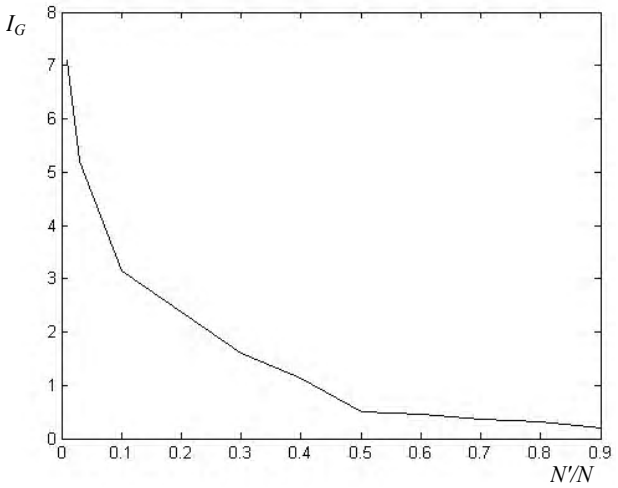


Рисунок 4 – График усредненной зависимости I_G от N'/N

6 ОБСУЖДЕНИЕ

Как видно из табл. 2, предложенный метод позволяет существенно ускорить синтез нейро-нечетких моделей, обеспечивая приемлемую точность и более высокий уровень обобщения данных моделью. Это позволяет рекомендовать предложенный метод для использования на практике.

Полученная зависимость ошибки модели *E* от заданной разрядности признаков *L* для выборок данных различной размерности и природы, график которой приведен на рис. 2, показывает, что при начальном значении *L*=2 ошибка также имеет наибольшее значение, которое постепенно снижается до некоторого значения, после чего начинается постепенный рост ошибки. Изначально высокое значение ошибки объясняется недостаточной разрядностью кодирования вследствие большой потери информации при разбиении на интервалы значений признаков из-за малого числа интервалов. Постепенное убывание ошибке по мере роста значения *L* объясняется увеличением объема информации, доступного для отображения в модели. Дальнейший рост ошибки объясняется ухудшением обобщающих свойств модели вследствие избыточности ее структуры и параметров, а также излишней детализации разбиения признакового пространства.

Представленный на рис. 3 график показывает, что при небольшом числе используемых признаков относительно размера исходного набора признаков ошибка построенной модели является большей, чем при использовании большего числа признаков. Однако при приближении числа используемых в модели признаков к числу признаков исходном наборе убывание ошибки становится незначительным. Таким образом, отбор признаков должен быть направлен на поиск оптимального соотношения N'/N, обеспечивающего приемлемый уровень ошибки.

График, изображенный на рис. 4, показывает, что при небольшом числе используемых признаков относительно размера исходного набора признаков коэффициент обобщения построенной модели является большим, чем при использовании большего числа признаков. Таким образом, отбор признаков должен быть направлен также на поиск оптимального соотношения N'/N, обеспечивающего не только приемлемый уровень ошибки, но и высокий уровень обобщения модели.

ЗАКЛЮЧЕНИЕ

В работе решена актуальная задача автоматизации синтеза нейро-нечетких сетей для диагностирования и автоматической классификации по признакам.

Научная новизна работы состоит в том, что предложен метод синтеза нейро-нечетких моделей по прецедентам, который в процессе синтеза сети определяет оценки взаимосвязи входных признаков между собой, выделяет взаимосвязанные признаки, оценивает влияние признаков на выходной признак, а также выявляет взаимосвязанные термы, включает в модель наиболее важные признаки и термы, устраняет дубляж термов и признаков путем оставления лишь части из них, использует различные варианты кодирования сигналов, а также переупорядочивает признаки, обеспечивая группировку признаков, формирует правила для прецедентов ранее не встречавшихся классов или существенно отличающихся от имеющихся правил своего класса, а уже имеющиеся правила корректирует на основе поступающих прецедентов с учетом числа ранее рассмотренных наблюдений.

Предложенный метод позволяет существенно ускорить синтез нейро-нечетких моделей, обеспечивая приемлемую точность и более высокий уровень обобщения данных, снизить сложность и избыточность, а также повысить интерпретабельность нейромодели.

Практическая значимость результатов работы состоит в том, что проведены эксперименты по решению практических задач диагностирования и автоматической классификации, подтвердившие работоспособность и применимость предложенного метода. Получена зависимость ошибки модели, синтезированной предложенным методом, от заданной разрядности признаков. Использование полученной зависимости позволяет на практике более рационально выбирать значение числа интервалов разбиения диапазонов значений признаков, обеспечивая приемлемую точность нейромодели. Получены зависимости ошибки синтезированной модели и коэффициента обобщения от доли используемых в модели признаков в исходном наборе признаков.

Перспективы дальнейших исследований заключаются в том, чтобы изучить зависимость оптимального значения числа интервалов разбиения диапазонов значений признаков от информативности набора признаков и размера набора признаков.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке международного проекта «Centers of Excellence for young RESearchers» программы «Темпус» Европейской Комиссии (№ 544137-TEMPUS-1-2013-1-SK-TEMPUS-JPHES) в рамках госбюджетных научно-исследовательских тем Запорожского национального технического университета «Интеллектуальные информационные технологии диагностирования и автоматической классификации» и «Интеллектуальные методы диагностирования систем управления удаленными техническими объектами».

СПИСОК ЛИТЕРАТУРЫ

1. Биргер И. А. Техническая диагностика / И. А. Биргер. – М. : Машиностроение, 1978. – 240 с.
2. Ding S. X. Model-based fault diagnosis techniques: design schemes, algorithms, and tools / S. X. Ding. – Berlin: Springer, 2008. – 473 p.
3. Интеллектуальные информационные технологии проектирования автоматизированных систем диагностирования и распознавания образов : монография / [С. А. Субботин, Ан. А. Олейник, Е. А. Гофман, С. А. Зайцев, Ал. А. Олейник] ; под ред. С. А. Субботина. – Харьков : Компания СМИТ, 2012. – 318 с.
4. 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.
5. Computational intelligence: a methodological introduction / [R. Kruse, C. Borgelt, F.Klawonn et. al.]. – London: Springer-Verlag, 2013. – 488 p. DOI: 10.1007/978-1-4471-5013-8_1
6. Bishop C. M. Pattern recognition and machine learning / C. M. Bishop. – New York : Springer, 2011. – 740 p.
7. Buckleya J. J. Fuzzy neural networks: a survey / J. J. Buckleya, Y. Hayashi // Fuzzy sets and systems. – 1994. –Vol. 66, Issue 1. – P. 1–13.

8. Mamdani E. H. An experiment in linguistic synthesis with fuzzy logic controller / E. H. Mamdani, S. Assilian // International journal of man-machine studies. – 1975. – Vol. 7, № 1. – P. 1–13.
9. Chai Y. Mamdani Model based Adaptive Neural Fuzzy Inference System and its Application / Y. Chai, L. Jia, Z. Zhang // International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:3, No:3, 2009. –P. 663–670.
10. Леоненков А. В. Нечеткое моделирование в среде MATLAB и fuzzyTECH / А. В. Леоненков. – СПб.: БХВ-Петербург, 2003. – 736 с.
11. Химмельблау Д. Прикладное нелинейное программирование / Д. Химмельблау. – М.: Мир, 1974. – 534 с.
12. Комп'ютерна програма «Автоматизована система синтезу нейромережних та нейро-нечітких моделей для неруйнівної діагностики та класифікації образів за ознаками» : свідоцтво про реєстрацію авторського права на твір № 35431 / С. О. Субботін. – Держ. департамент інтелектуальної власності. – № 34011 ; заявл. 21.04.10 ; зареєстр. 21.10.10.
13. Шама Є. О.Побудова класифікатора рослинних об'єктів за допомогою нейронних мереж / Є. О. Шама , С. О. Субботін, С. В. Морщавка // Радіоелектроніка, інформатика, управління. – 2013. – № 1. – С. 55–61.
14. Колісник Н. В. Моделювання імунопатогенезу хронічного обструктивного бронхіту за допомогою нейромереж / Н. В. Колісник, С. О. Суботин // Сучасні проблеми біології, екології та хімії : II Міжнародна конференція, Запоріжжя, 1–3 жовтня 2009 р. : збірка матеріалів. – Запоріжжя : ЗНУ. – С. 124–125.

Статья поступила в редакцию 17.01.2016.
После доработки 10.02.2016.

Субботін С. О.
Д-р техн. наук, професор, професор кафедри програмних засобів Запорізького національного технічного університету, Запоріжжя, Україна

СИНТЕЗ НЕЙРО-НЕЧІТКИХ МЕРЕЖ З РАНЖИРУВАННЯМ І СПЕЦИФІЧНИМ КОДУВАННЯМ ОЗНАК ДЛЯ ДІАГНОСТИКИ Й АВТОМАТИЧНОЇ КЛАСИФІКАЦІЇ ЗА ПРЕЦЕДЕНТАМИ

Вирішено задачу автоматизації синтезу нейро-нечітких мереж для діагностування й автоматичної класифікації за ознаками. Запропоновано метод синтезу нейро-нечітких моделей за прецедентами, що визначає оцінки взаємозв'язку вхідних ознак, виділяє взаємозалежні ознаки, оцінює вплив ознак на вихідну ознаку, а також виявляє взаємозалежні терми, включає в модель найбільш важливі ознаки і терми, усуває дубляж термів і ознак, використовує різні варіанти кодування сигналів, а також переупорядковує ознаки, забезпечуючи групування ознак, формує правила для прецедентів класів, що раніше не зустрічалися, або таких, що істотно відрізняються від наявних правил свого класу, а уже наявні правила коректує на основі прецедентів, що надходять, з урахуванням числа раніше розглянутих спостережень. Запропонований метод дозволяє істотно прискорити синтез нейро-нечітких моделей, забезпечуючи прийнятну точність і більш високий рівень узагальнення даних, знизити складність і надмірність, а також підвищити інтерпретабельність нейромоделі. Проведено експерименти з вирішення практичних задач діагностування й автоматичної класифікації, що підтвердили працездатність і застосовність запропонованого методу. Отримано залежність помилки моделі, синтезованої запропонованим методом, від заданої розрядності ознак. Використання отриманої залежності дозволяє на практиці більш раціонально вибирати значення кількості інтервалів розбиття діапазонів значень ознак, забезпечуючи прийнятну точність нейромоделі.

Ключові слова: вибірка, нейро-нечітка мережа, нечітке виведення, навчання за прецедентами, діагностування.

Subbotin S. A.
Dr.Sc., Professor, Professor of department of software tools, Zaporizhzhya National Technical University, Zaporizhzhya, Ukraine

THE NEURO-FUZZY NETWORK SYNTHESIS WITH THE RANKING AND SPECIFIC ENCODING OF FEATURES FOR THE DIAGNOSIS AND AUTOMATIC CLASSIFICATION ON PRECEDENTS

The problem of automation synthesis of neuro-fuzzy networks for diagnostics and automatic classification on features is solved. The method of neuro-fuzzy model synthesis on precedents is proposed.

It evaluates the relationship of input features, extracts related features, evaluates the impact of features on the output feature and identify related terms, includes the most important features and terms in the model, eliminates dubbing of the terms and features, uses different types of signal encoding, and also arranges features providing their grouping, forms a rules for precedents of previously non-experienced classes or significantly different from the existing rules of their class, adjust the existing rules based on incoming precedents, given the number previously considered observations. The proposed method allows to significantly accelerate the synthesis of neuro-fuzzy models, providing acceptable accuracy and a higher level of generalization of data, reduce complexity and redundancy, as well as increase of interpretability of neural model. The experiments on solution of practical problems of diagnosis and automatic classification are conducted. They confirmed the efficiency and applicability of the proposed method. The dependence of error of the model synthesized by the proposed method from the specified feature bitness is obtained. Using the obtained dependence allows to more rational choose the value of the number of divisions of the range of feature values in practice providing a reasonably accurate neural model.

Keywords: sample, neuro-fuzzy network, fuzzy inference, training on precedents.

REFERENCES

1. Birger I. A. Tekhnicheskaya diagnostika. Moscow. Mashinostroyeniye, 1978, 240 p.
2. Ding S. X. Model-based fault diagnosis techniques: design schemes, algorithms, and tools. Berlin, Springer, 2008, 473 p.
3. Subbotin S. A., Oleynik An. A., Gofman Ye. A., Zaytsev S. A., Oleynik Al. A. ; pod red. S. A. Subbotina. Intellektual'nyye informatsionnyye tekhnologii proyektirovaniya avtomatizirovannykh sistem diagnostirovaniya i raspoznavaniya obrazov : monografiya. Khar'kov, Kompaniya SMIT, 2012, 318 p.
4. Subbotin S. O. Podannya y obrobka znan' u systemakh shtuchnoho intelektu ta pidtrymky pryynyattya rishen' : navch. posibnyk. Zaporizhzhya, ZNTU, 2008, 341 p.
5. Kruse R., Borgelt C., Klawonn F. et. al. Computational intelligence: a methodological introduction. London: Springer-Verlag, 2013, 488 p. DOI: 10.1007/978-1-4471-5013-8_1
6. Bishop C. M. Pattern recognition and machine learning. New York, Springer, 2011, 740 p.
7. Buckleya J. J., Hayashi Y. Fuzzy neural networks: a survey, *Fuzzy sets and systems*, 1994, Vol. 66, Issue 1, pp. 1–13.
8. Mamdani E. H. Assilian S. An experiment in linguistic synthesis with fuzzy logic controller, *International journal of man-machine studies*, 1975, Vol. 7, No. 1, pp. 1–13.
9. Chai Y., Jia L., Zhang Z. Mamdani Model based Adaptive Neural Fuzzy Inference System and its Application, International Journal of Computer, Electrical, Automation, Control and Information Engineering, Vol:3, No:3, 2009, pp. 663–670.
10. Leonenkov A. V. Nechotkoye modelirovaniye v srede MATLAB i fuzzyTECH. Sankt-Peterburg, BKHV-Peterburg, 2003, 736 p.
11. Khimmel'blau D. Prikladnoye nelineynoye programmirovaniye. Moscow, Mir, 1974, 534 p.
12. Subbotin S. O. Komp'yuterna prohrama «Avtomatyzovana systema syntezu neyromerezhevykh ta neyro-nechitkykh modeley dlya neruynivnoyi diahnostyky ta klasyfikatsiyi obraziv za oznakamy» : cvidotstvo pro reyestratsiyu avtors'koho prava na tvir № 35431. Derzh. departament intelektual'noyi vlasnosti. № 34011 ; zayavl. 21.04.10 ; zareyestr. 21.10.10.
13. Shama Ye. O., Subbotin S. O., Morshchavka S. V. Pobudova klasyfikatora roslynnykh ob'yektiv za dopomohoyu neyronnykh merezh, *Radioelektronika, informatyka, upravlinnya*, 2013, No. 1, pp. 55–61.
14. Kolisnyk N. V., Subotyn S. O. Modelyuvannya imunopatohenezu khronichnoho obstruktyvnoho bronkhitu za dopomohoyu neyromerezh. Suchasni problemy biolohiyi, ekolohiyi ta khimiyi : II Mizhnarodna konferentsiya, Zaporizhzhya, 1–3 zhovtnya 2009 r. : zbirka materialiv. Zaporizhzhya, ZNU, pp. 124–125.

УДК 004.93

Фазылов Ш. Х.[1], Мирзаев Н. М.[2], Мирзаев О. Н.[3]

[1]Д-р техн. наук, профессор, заведующий лабораторией Центра разработки программных продуктов и аппаратно-программных комплексов при Ташкентском Университете Информационных Технологий, Ташкент, Узбекистан
[2]Канд. техн. наук, старший научный сотрудник Центра разработки программных продуктов и аппаратно-программных комплексов при Ташкентском Университете Информационных Технологий, Ташкент, Узбекистан
[3]Младший научный сотрудник Центра разработки программных продуктов и аппаратно-программных комплексов при Ташкентском Университете Информационных Технологий, Ташкент, Узбекистан

ПОСТРОЕНИЕ РАСПОЗНАЮЩИХ ОПЕРАТОРОВ В УСЛОВИЯХ ВЗАИМОСВЯЗАННОСТИ ПРИЗНАКОВ

Рассмотрена задача распознавания образов, заданных в пространстве взаимосвязанных признаков. Предложен новый подход к построению модели распознающих операторов, учитывающий взаимосвязанность заданных признаков. При этом построение модели осуществлено для распознающих операторов типа потенциальных функций. Основная идея предлагаемого подхода заключается в формировании независимых подмножеств взаимосвязанных признаков и выделении предпочтительной модели зависимости для каждого подмножества сильносвязанных признаков. Анализ полученных результатов показывает, что рассмотренные распознающие операторы используются в тех случаях, когда между признаками объектов, принадлежащих к одному и тому же классу, существует некоторая зависимость. При слабом выражении этой зависимости используется классическая модель распознающих операторов. Основным преимуществом предложенных распознающих операторов является улучшение точности и значительное сокращение объема вычислительных операций при распознавании неизвестных объектов, что позволяет применить их при построении распознающих систем, работающих в режиме реального времени.

Ключевые слова: распознавание образов, модель распознающих операторов, потенциальная функция, зависимость признаков, подмножество сильносвязанных признаков, репрезентативный признак, предпочтительная модель зависимости.

НОМЕНКЛАТУРА

$\Im$ – множество допустимых объектов;

$\widetilde{S}^m$ – множество выделенных допустимых объектов ($\widetilde{S}^m \subset \Im$, $\widetilde{S}^m = \{S_1,...,S_i,...,S_m\}$);

$\widetilde{S}^q$ – множество контрольных объектов ($\widetilde{S}^q \subset \Im, \widetilde{S}^q = \{S_1^{'},...,S_q^{'}\}$);

$\widetilde{K}_j$ – множество эталонных объектов, принадлежащих к классу K_j ($\widetilde{K}_j \subset \widetilde{S}^m$);

$\left|\widetilde{K}_j\right|$ – мощность множества объектов $\widetilde{K}_j$;

V_o – обучающая выборка ($V_0 = \{< S_i; \chi(S_i) >: S_i \in \widetilde{S}^m\}$);

V_k – контрольная выборка ($V_k = \{< S_u; \chi(S_u) >: S_u \in \widetilde{S}^q\}$);

$\chi(S_i)$ – индикатор объекта S_i ($\chi(S_i) = -1, \text{при } S_i \in K_1$, $\chi(S_i) = +1, \text{при } S_i \in K_2$);

X – пространства n-мерных признаков $\left(\mathrm{X} = (x_1,...,x_i,...,x_n)\right)$;

A – произвольный алгоритм распознавания ($A = B \circ C$);

B – распознающий оператор, $B(I_0, \widetilde{S}^q) = (b_1, b_2,..., b_q)$;

C – решающее правило $C(b_i) = \lambda_i (\lambda_i \in \{-1, 0, +1\})$;

$\widetilde{\pi}$ – вектор параметров распознающего оператора;

$\widetilde{\pi}_0$ – вектор параметров распознающего оператора, вычисленных по выборке V_o;

$\varphi(\widetilde{\pi})$ – функционал качества распознающего оператора $B(\widetilde{\pi})$;

$n^{'}$ – число подмножеств сильносвязанных признаков;

Ω_q – q-е подмножество сильносвязанных признаков;

Y – пространство $n^{'}$-мерных репрезентативных признаков ($Y = (y_1,..., y_{n^{'}})$).

ВВЕДЕНИЕ

В настоящее время все более широкий круг специалистов уделяет внимание проблеме распознавания образов, и число научных публикаций по данной тематике постоянно растет. Это связано с тем, что в последние годы распознавание образов находит все большее применение в науке, технике, производстве и повседневной жизни.

В проблеме распознавания образов центральное место занимает задача построения распознающих алгоритмов. Для рассмотрения разнообразных постановок этой задачи можно выделить три параметра, от способа задания которых во многом определяется метод решения. К ним относятся, во-первых, вид описания распознаваемых объектов (независимые и взаимосвязанные признаки), во-вторых, способ предъявления обучающей информации (единственная выборка и последовательность выборок) и, в-третьих, тип используемой модели, на основе которой формируется алгоритм распознавания (см. раздел 2).

Такой подход дает возможность выделить отдельные подклассы задач распознавания образов, описанных в пространстве взаимосвязанных признаков. Однако решение этих задач исследовано недостаточно (см. раздел 2).

Целью данной работы является разработка модифицированной модели распознающих операторов типа потенциальных функций [1–3], направленная на приспособление данной модели к решению задач распознавания образов, описанных в пространстве взаимосвязанных признаков.

DOI 10.15588/1607-3274-2016-1-7

Для достижения этой цели необходимо решить следующие задачи:

– проанализировать существующие модели распознающих операторов и определить круг решаемых задач;

– разработать модель алгоритмов распознавания, основанных на оценке взаимосвязанности признаков;

– провести экспериментальные исследования для оценки эффективности разработанных распознающих операторов.

Объектом исследования являются распознающие операторы типа потенциальных функций. Предмет исследования – модифицированные распознающие операторы, основанные на оценке взаимосвязанности признаков.

Основная идея предлагаемой модели распознающих операторов заключается в поиске некоторых зависимостей между признаками, характеризующими объекты, принадлежащие к одному и тому же классу.

В научном плане результаты данной работы представляют собой новое решение научной задачи, связанной с построением распознающих алгоритмов при условии большой размерности пространства признаков. Практическая значимость полученных результатов заключается в том, что разработанные алгоритмы и программы могут быть применены при решении прикладных задач в условиях большой размерности пространства признаков (например, при идентификации человека по фотопортрету).

1 ПОСТАНОВКА ЗАДАЧИ

Для простоты, но без ограничения общности, рассмотрим задачу распознавания образов в случае двух классов. Пусть набор объектов $\widetilde{S}^q = \{S_1^{'},...,S_q^{'}\}$ ($\widetilde{S}^q \in \{S\}$) задан в пространстве признаков X большой размерности ($X = (x_1,...x_i,...,x_n)$, n >200).

Основная задача заключается в построении такого распознающего оператора B, который с применением решающего правила C (см. формула 2 в разделе 2) позволяет вычислить значения характеристической функции $\hat{\lambda}_i = \chi(S_i)$, $i = \overline{1,q}$ по выборке V_o:

$$B(\widetilde{S}^q) = \|b_i\|_q, \ \|b_i\|_q = (b_1,...,b_i,...,b_q)^T,$$

$$C(\|b_i\|_q) = (\hat{\lambda}_1,...,\hat{\lambda}_i,...,\hat{\lambda}_q)^T, \ \hat{\lambda}_i \in \{-1,0,1\}.$$

Здесь $\hat{\lambda}_i$ интерпретируется следующим образом: если $\hat{\lambda}_i \in \{-1,+1\}$ ($\hat{\lambda}_i = -1$ – объект $S_i^{'}$ входит в класс K_1, $\hat{\lambda}_i = +1$ – объект $S_i^{'}$ входит в класс K_2), то $\hat{\lambda}_i$ есть значение характеристической функции $\lambda(S_i^{'})$ на допустимом объекте $S_i^{'}$, вычисленное оператором B; если $\hat{\lambda}_i = 0$, то считается, что оператор B не смог определить значение характеристической функции $\lambda(S_i^{'})$.

2 ОБЗОР ЛИТЕРАТУРЫ

В течение длительного времени подавляющее большинство приложений теории распознавания образов было связано с плохо формализованными областями – медициной, геологией, социологией, химией и т. д. Поэтому на первом этапе развития распознавания появилось множество алгоритмов, которые носили характер проектов различных технических устройств или алгоритмов для решения конкретных прикладных задач. Ценность разработанных алгоритмов определялась, прежде всего, достигнутыми экспериментальными результатами [2, 4].

Второй этап развития теории распознавания образов характеризуется переходом от отдельных алгоритмов к построению моделей – семейства алгоритмов для единого описания методов решения классификационных задач. Потребность в синтезе моделей алгоритмов распознавания образов определялась необходимостью фиксировать класс алгоритмов при выборе оптимальной или хотя бы приемлемой процедуры решения конкретной задачи.

На данном этапе развития Ю. И. Журавлев показал [2], что произвольный алгоритм распознавания можно представить как последовательное выполнение операторов B (распознающий оператор) и C (решающее правило):

$$A = B \circ C. \quad (1)$$

Из (1) следует, что каждый алгоритм A можно разделить на два последовательных этапа. На первом этапе распознающий оператор B осуществляет перевод допустимого объекта S_i в числовую оценку b_i: $B(S_i) = b_i$.

На втором этапе по числовой оценке b_i решающее правило C определяет принадлежность объекта S_i к классам K_1 и K_2:

$$C(\mu(S_i)) = C(b_i), \ C(b_i) = \hat{\lambda}_i,$$

$$\hat{\lambda} = \begin{cases} -1, \text{если } \mu(S_i) < -c; \\ 0, \text{если } c \le \mu(S_i) \le c; \\ +1, \text{если } c < \mu(S_i), \end{cases} \quad (2)$$

где c – параметр решающего правила.

К настоящему времени построено и изучено несколько типов моделей. К их числу относятся [1–12]:

– модели, основанные на использовании разделяющих функций;

– модели, основанные на использовании математической статистики и теории вероятности;

– модели, основанные на использовании аппарата математической логики;

– модели, основанные на использовании принципа потенциалов;

– модели, основанные на использовании алгоритмов вычисления оценок.

Анализ литературных источников, в частности [1–12], показывает, что разработанные модели алгоритмов распознавания, в основном, ориентированы на решение задач, где объекты описаны в пространстве независимых признаков (или зависимость между признаками достаточно слабая).

На практике часто встречаются прикладные задачи распознавания образов, заданных в пространстве признаков большой размерности. При решении подобных

задач предположение о независимости признаков достаточно часто не выполняется [4, 13]. Следовательно, остается недостаточно решенным вопрос по созданию распознающих алгоритмов, которые могут быть применены для решения прикладных задач распознавания при больших размерностях признакового пространства и наличии взаимосвязанности признаков.

3 МАТЕРИАЛЫ И МЕТОДЫ

Для решения задачи, сформулированной в разделе 1, предлагается подход, который является логическим продолжением работ академика Ю. И. Журавлева и его учеников. На базе этого подхода разработана модель модифицированных распознающих операторов, основанных на выявлении независимых подмножеств взаимосвязанных признаков и выделении предпочтительной модели зависимости для каждого подмножества сильносвязанных признаков. В качестве исходной модели рассматриваются распознающие операторы типа потенциальных функций [1–3].

Предлагаемая модель распознающих операторов включает следующие основные этапы.

1. *Выделение подмножеств сильносвязанных признаков*. На этом этапе определяется система «независимых» подмножеств признаков, состав которой будет зависеть от параметра n'. В зависимости от способа задания меры близости между подмножествами сильносвязанных признаков (Ω_p и Ω_q) и функционала качества классификационного анализа можно получить разнообразные процедуры выделения независимых множеств сильносвязанных признаков. Процедура выделения подмножеств сильносвязанных признаков более подробно рассмотрена в [14].

2. *Формирование набора репрезентативных признаков*. В результате выполнения данного этапа получаем сокращенное пространство признаков, размерность которого намного меньше исходного ($n' < n$). Далее сформированное пространство признаков обозначим через Y ($Y = (y_1,..., y_{n'})$).

Процедура выделения набора репрезентативных признаков более подробно рассмотрена в [15].

3. *Определение моделей зависимости в каждом подмножестве признаков для класса* K_j ($j = \overline{1,l}$). Пусть x_i – произвольный признак, принадлежащий подмножеству Ω_q. Предполагается, что элементы Ω_q линейно упорядочены по индексу признаков (т. е. $x_i < x_j$, если $i < j$). Далее, нулевым элементом (x_0) подмножества Ω_q считается y_q, остальные элементы обозначаются через x_i ($N_q = |\Omega_q|; i = 1,..., N_q - 1$).

Пусть модель зависимости в Ω_q имеет вид

$$y_q = F(\bar{c}, x_i),\ x_i \in \Omega_q \setminus y_q,$$

где $\bar{c}$ – вектор неизвестных параметров, x_i – произвольный признак, принадлежащий подмножеству Ω_q, F – функция из некоторого заданного класса $\{F\}$.

Вычисленные значения вектора неизвестных параметров $\bar{c}$ определяют модель зависимости в подмножестве признаков Ω_q для класса $K_j (j = \overline{1,l})$. В зависимости от задания параметрического вида $F(\bar{c}, x)$ и метода определения $\bar{c}$ получаем разнообразные модели зависимости во множестве признаков Ω_q $(q = \overline{1,n'})$.

4. *Выделение предпочтительных моделей зависимости.* Поиск предпочтительной модели зависимости в Ω_q осуществляется на основе оценки доминированности рассматриваемой модели, которая разделяет объекты, принадлежащие множеству $\tilde{S}^m$, на два подмножества K_1 и K_2 [16]:

$$T_{qi} = \frac{L_1 \sum\limits_{S \in \tilde{K}_2} |y_q - F_q(\bar{c}, x_i)|}{L_2 \sum\limits_{S \in \tilde{K}_1} |y_q - F_q(\bar{c}, x_i)|}, \quad L_1 = |\tilde{K}_1|,\ L_2 = |\tilde{K}_2|.$$

Чем больше величина T_{qi}, тем больше отдается предпочтение i-ой модели зависимости в подмножестве Ω_q. Если несколько моделей получают одинаковое предпочтение, то выбирается любая из них.

В результате выполнения данного этапа определяется предпочтительная модель зависимости для подмножества признаков Ω_q. Далее рассматриваются только эти модели зависимости.

5. *Определение функции различия между объектами* S_u *и* S. Пусть заданы два объекта S_u и S в пространстве $X = (x_1,..., x_i,..., x_n)$:

$$S_u = (a_{1u},..., a_{nu}) \ \text{ и } \ S = (b_1,..., b_n).$$

На данном этапе определяется расстояние между этими объектами по всем подмножествам $\Omega_q (q = \overline{1,n'})$:

$$d(S_u, S) = \sum_{q=1}^{n'} \tau_q d_q(S_u, S),$$

$$d_q(S_u, S) = |\phi_q - \psi_q|,\ \phi_q = (a_{ui} + b_i)/2,$$

$$\psi_q = (F_q(\bar{c}, a_{ui_q}) + F_q(\bar{c}, b_{i_q}))/2,$$

где τ_q – параметр распознающего оператора.

6. *Определение функции близости между объектами* S_u *и* S. На этом этапе функция близости между объектами S_u и S определяется как потенциальная функция второго типа, например [1, 3],

$$U(S_u, S) = \frac{1}{1 + \xi d(S_u S)},$$

где ξ – параметр распознающего оператора.

7. *Вычисление оценки принадлежности к классу* K_j. Оценка принадлежности объекта S к классу K_j $(j = 1,2)$ определяется как функция оценки близости:

$$\mu(S) = \sum_{S_u \in \tilde{K}_2} \gamma_u U(S_u, S) - \sum_{S_u \in \tilde{K}_1} \gamma_u U(S_u, S),$$

где γ_u – параметр распознающего оператора.

Таким образом, определена модель распознающих операторов типа потенциальных функций, основанных на оценке взаимосвязанности признаков. Произвольный оператор B из этой модели полностью определяется заданием набора параметров $\tilde{\pi}$ [17]. Совокупность всех распознающих операторов из предлагаемой модели обозначаем через $B(\tilde{\pi}, S)$. Определение наилучшего распознающего оператора в рамках рассмотренной модели осуществляется в пространстве параметров $\tilde{\pi}$. Наилучший оператор $B(\tilde{\pi}_o, S)$ выбирается на основе поиска минимального значения функционала качества распознающего оператора:

$$\varphi(\tilde{\pi}) = \frac{1}{m} \sum_{S \in V_o} \theta\{\|\chi(S) - C(B(\tilde{\pi}, S))\|\}, \quad \theta(x) = \begin{cases} 1, \text{ при } x = 0; \\ 0, \text{ при } x \neq 0, \end{cases} \quad (3)$$

где $\|\cdot\|$ – норма булевого вектора.

4 ЭКСПЕРИМЕНТЫ

Экспериментальное исследование работоспособности предложенной модели распознающих операторов осуществлено на примере решения модельной задачи.

Исходные данные распознаваемых объектов для модельной задачи сгенерированы в пространстве зависимых признаков. Количество классов в данном эксперименте равно двум. Объем исходной выборки – 1100 реализаций (по 550 реализаций для объектов каждого класса). Количество признаков в модельном примере равно 20. Число подмножеств сильносвязанных признаков – 5.

В качестве испытуемых моделей распознающих операторов были выбраны: классическая модель распознающих операторов типа потенциальных функций (B_1), модель распознающих операторов, основанных на вычислении оценок (B_2), и модель (B_3), предлагаемая в настоящей работе. Сравнительный анализ перечисленных моделей распознающих операторов при решении рассмотренной задачи проведен по следующим критериям:

– точность распознавания объектов контрольной выборки;

– время, затраченное на обучение;

– время, затраченное на распознавание объектов из контрольной выборки.

Для вычисления указанных критериев при решении рассматриваемой задачи, в целях исключения удачного (или неудачного) разбиения исходной выборки V на две равные части V_o и V_k ($V = V_o \bigcup V_k$, V_o – выборка для обучения, V_k – выборка для контроля), используется метод скользящего контроля [18]. В этом методе исходная выборка объектов случайным образом разбивается на 110 непересекающихся блоков, включающих по 10 объектов каждый. При этом требуется, чтобы во всех блоках сохранилась пропорция по количеству объектов, принадлежащих к разным классам. В результате получается, что в каждом блоке по 5 объектов каждого класса. Процесс скользящего контроля по этим блокам включает несколько шагов. На каждом шаге выбирают 10 из 110 блоков в качестве обучающей выборки, и на этой выборке обучаются распознающие операторы с заданными параметрами. Обученный таким образом распознающий оператор проверяется на остальных 100 блоках (контрольной выборке). В результате каждой проверки определяется и фиксируется оценка качества распознающих операторов по указанным критериям. При выполнении каждого очередного шага для оценки качества распознающих операторов из контрольной и обучающей выборок выбирают по одному блоку и меняют их местами. В целях исключения повторного использования объектов обучающей выборки соответствующие блоки маркируются, и при выборе кандидатов для включения в обучающую выборку они не участвуют. После завершения процедуры скользящего экзамена точность распознавания и временные показатели определялись как средние. Эксперименты проводились на компьютере Pentium IV Dual Core 2,2 GHz с объемом оперативной памяти 1 Gb.

5 РЕЗУЛЬТАТЫ

Основным результатом данной статьи является модель распознающих операторов, основанных на оценке взаимосвязанности признаков (см. раздел 3). Построение данной модели выполнено в рамках модели алгоритмов типа потенциальных функций и заключалось в выявлении независимых подмножеств взаимосвязанных признаков с последующим выделением предпочтительной модели зависимости для каждого подмножества сильносвязанных признаков. Данная модель распознающих операторов имеет большое значение при решении задачи распознавания образов в условиях взаимосвязанности признаков и позволяет расширить область применения метода потенциальных функций.

В целях проверки работоспособности предложенной модели операторов проведено экспериментальное исследование при решении модельной задачи. Как уже отмечалось ранее, модельная задача решалась с применением классической модели распознающих операторов типа потенциальных функций (B_1), модели распознающих операторов, основанных на вычислениях оценок, (B_2) и предлагаемой модели (B_3). Точность распознавания в процессе обучения (см. формула 3 в разделе 3) для B_1 равна 94,8%, для B_2 – 97,6%, и, наконец, для B_3 – 99,1%. Результаты решения рассматриваемой задачи с применением B_1, B_2 и B_3 в процессе контроля приведены в табл. 1.

Сравнение этих результатов показывает (см. табл. 1), что предложенная модель распознающих операторов $|B_3|$ позволила повысить точность распознавания объектов, описанных в пространстве взаимосвязанных признаков (более чем на 10% выше, чем по B_1 и B_2). Это объясняется тем, что в моделях B_1 и B_2 не учитывается взаимосвязанность признаков. Однако для модели B_3

Таблица 1 – Результаты решения задачи с помощью различных распознающих операторов

Распознающий оператор	Время (в сек.)		Точность распознавания (в %)
	обучения	распознавания	
B_1	4,141	0,014	80,5
B_2	6,127	0,012	82,7
B_3	9,041	0,002	93,4

имеет место некоторое увеличение времени обучения за счет реализации дополнительной процедуры формирования независимых подмножеств взаимосвязанных признаков.

6 ОБСУЖДЕНИЕ

Разработанные операторы отличаются от традиционных распознающих операторов типа потенциальных функций тем, что они основаны на оценке взаимосвязанности признаков. Поэтому эти алгоритмы целесообразно использовать в тех случаях, когда между признаками обнаруживается какая-нибудь зависимость. Несомненно, что эта зависимость должна отличаться для объектов каждого класса. Это позволяет описать объекты каждого класса индивидуальной моделью. Если зависимость между признаками слаба, то используется классическая модель распознающих операторов (например, модель, рассмотренная в [2, 3]). Следовательно, предложенная модель распознающих операторов не является альтернативной моделям распознающих операторов типа потенциальных функций, а только дополняет их.

В случае, когда между признаками всех рассматриваемых объектов обнаруживается достаточно сильная зависимость, то в процессе формирования набора репрезентативных признаков (описанных на первом и втором этапах) исключаются признаки, повторяющие одну и ту же информацию, что обеспечивает выбор признаков, достаточно хорошо представляющих все те признаки, которые не содержатся в данном наборе [4].

Результаты приведенного экспериментального исследования показывают, что предложенная модель распознающих операторов позволяет решать задачу распознавания образов более точно в условиях взаимосвязанности признаков. При этом показатель времени, затраченного на распознавание объектов, намного меньше, чем тот же показатель B_1 и B_2 (см. табл. 1). Это связано с тем, что в предложенных распознающих операторах для распознавания объекта используются только предпочтительные модели зависимости, что обусловило повышение скорости распознавания объектов. Поэтому эту модель можно использовать при разработке систем распознавания, функционирующих в режиме реального времени. Вместе с тем необходимо отметить тот факт, что время, израсходованное на обучение алгоритма, увеличилось, т. к. для построения оптимального распознающего оператора требуется оптимизировать большее число параметров, чем при использовании традиционной модели распознающих операторов, в частности B_1 и B_2.

ВЫВОДЫ

1. Анализ существующих моделей распознающих операторов показал, что в основном они ориентированы на решение прикладных задач распознавания объектов с независимыми признаками. В большинстве прикладных задач распознавания, встречающихся в науке, технике и производстве, рассматриваемые образы характеризуются взаимосвязанными признаками. Хотя разработаны различные алгоритмы решения задач распознавания для таких образов, они оказались малоэффективными с точки зрения точности и сложности вычисления.

2. Предложен новый подход для построения модели распознающих операторов и на основе этого подхода построена модифицированная модель распознающих операторов типа потенциальных функций. Предложенная модель операторов опирается на оценку взаимосвязанности признаков. Она позволяет расширить модели распознающих операторов типа потенциальных функций и увеличивает область применения метода потенциальных функций. Данная модель ориентирована на решение задачи распознавания объектов, заданных в пространстве взаимосвязанных признаков.

3. Результаты решения модельной задачи показали, что предложенная модель распознающих операторов улучшает точность и значительно сокращает число вычислительных операций посредством распознавания неизвестного объекта, заданного в пространстве взаимосвязанных признаков. Она может быть использована при составлении различных программных комплексов, ориентированных на решение прикладных задач распознавания.

БЛАГОДАРНОСТИ

Работа выполнена в рамках гранта А5-ФА-Ф022 Государственного комитета по координации развития науки и технологий Республики Узбекистан.

СПИСОК ЛИТЕРАТУРЫ

1. Айзерман М. А. Метод потенциальных функций в теории обучения машин / М. А. Айзерман, Э. М. Браверманн, Л. И. Розоноэр. – М. : Наука, 1970. – 348 с.
2. Журавлев Ю. И. Избранные научные труды / Ю. И. Журавлев. – М. : Магистр, 1998. – 420 с.
3. Журавлев Ю. И. Распознавание. Математические методы. Программная система. Практические применения / Ю. И. Журавлев, В. В. Рязанов, О. В. Сенько. – М. : Фазис, 2006. –159 с.
4. Камилов М. М. Современное состояние вопросов построения моделей алгоритмов распознавания / М. М. Камилов, Н. М. Мирзаев, С. С. Раджабов. // Научный журнал: Химическая технология. Контроль и управление. – 2009. – № 2. – С. 21–27.
5. Загоруйко Н. Г. Прикладные методы анализа данных и знаний / Н. Г. Загоруйко. – Новосибирск : ИМ СО РАН, 1999.
6. Лбов Г. С. Логические решающие функции и вопросы статистической устойчивости решений / Г. С. Лбов, Н. Г. Старцева. – Новосибирск : Изд-во ИМ СО РАН, 1999. – 211 с.
7. Шлезингер М. Десять лекций по статистическому и структурному распознаванию / М. Шлезингер, В. Главач. – К. : Наукова думка, 2004. – 545 с.
8. Потапов А. С. Распознавание образов и машинное восприятие / А. С. Потапов. – СПб. : Политехника, 2007. – 548 с.
9. Duda R. O. Pattern Classification, Second Edition / R. O. Duda, P. E. Hart, D. G. Stork. – New York : John Wiley, Inc., 2001. – 680 p.
10. Vapnik V. Statistical Learning Theory / V. Vapnik. – New York : John-Wiley Sons, Inc., 1998. – 732 p.
11. Theodoridis S. Pattern Recognition: Theory and Applications, 4th edition / S. Theodoridis, K. Koutroumbas. – New York : Academic Press, 2009. – 957 p.
12. Dougherty G. Pattern Recognition and Classification: An Introduction / G. Dougherty. – New York: Springer, 2013. – 196 p.
13. Фазылов Ш. Х. Построение модели алгоритмов вычисления оценок с учетом большой размерности признакового пространства / Ш. Х. Фазылов, Н. М. Мирзаев, С. С. Раджабов // Вестник СГТУ. – Саратов, 2012. – № 1 (64), Вып. 2. – С. 274–279.
14. Мирзаев Н. М. Алгоритмы выделения подмножеств сильносвязанных признаков / Н. М. Мирзаев // Вопросы кибернетики : сб. науч. тр. – Ташкент : ИМИТ АН РУз, 2008. – Вып. 177. – С. 99–104.
15. Мирзаев О. Н. Выделение репрезентативных признаков при построении алгоритмов распознавания / О. Н. Мирзаев // Проблемы информатики и энергетики. – Ташкент, 2008. – № 6. – С. 23–27.

16. Модель распознающих операторов, основанных на принципе ближайшего соседа, в условиях взаимосвязанности признаков / Ш. Х. Фазылов, Н. М. Мирзаев, С. С. Раджабов, И. К. Каримов // Информатика и системы управления. – Благовещенск, 2012. – № 4. – С. 34–42.
17. Мирзаев Н. М. О параметризации моделей алгоритмов распознавания, основанных на оценке взаимосвязанности признаков / Н. М. Мирзаев, С. С. Раджабов, Т. С. Жумаев // Проблемы информатики и энергетики. – Ташкент, 2008. – № 2–3. – С. 23–27.
18. Vorontsov K. V. Computable Combinatorial Overыtting Bounds / K. V. Vorontsov, A. I. Frey, E. A. Sokolov // Journal of Machine Learning and Data Analysis. – 2013. – Т. 1, № 6. – С. 734–743.

Статья поступила в редакцию 09.11.2015.
После доработки 02.12.2015.

Фазилов Ш. Х.[1], Мірзаєв Н. М.[2], Мірзаєв О. Н.[3]
[1]Д-р техн. наук, професор, завідувач лабораторії Центру розробки програмних продуктів і апаратно-програмних комплексів при Ташкентському Університеті Інформаційних Технологій, Ташкент, Узбекистан
[2]Канд. техн. наук, старший науковий співробітник Центру розробки програмних продуктів і апаратно-програмних комплексів при Ташкентському Університеті Інформаційних Технологій, Ташкент, Узбекистан
[3]Молодший науковий співробітник Центру розробки програмних продуктів і апаратно-програмних комплексів при Ташкентському Університеті Інформаційних Технологій, Ташкент, Узбекистан

ПОБУДОВА РОЗПІЗНАЮЧИХ ОПЕРАТОРІВ В УМОВАХ ВЗАЄМОПОВ'ЯЗАНОСТІ ОЗНАК

Розглянуто задачу розпізнавання образів, заданих у просторі взаємозалежних ознак. Запропоновано новий підхід до побудови моделі розпізнавальних операторів, що враховує взаємозв'язок заданих ознак. При цьому побудова моделі здійснена для розпізнавальних операторів типу потенційних функцій. Основна ідея пропонованого підходу полягає у формуванні незалежних підмножин взаємозалежних ознак і виділенні кращої моделі залежності для кожної підмножини сильнозв'язаних ознак. Аналіз отриманих результатів показує, що розглянуті розпізнавальні оператори використовуються в тих випадках, коли між ознаками об'єктів, що належать до доного класу, існує деяка залежність. При слабкому вираженні цієї залежності використовується класична модель розпізнавальних операторів. Основною перевагою запропонованих розпізнавальних операторів є поліпшення точності і значне скорочення обсягу обчислювальних операцій при розпізнаванні невідомих об'єктів, що дозволяє застосувати їх при побудові розпізнавальних систем, які працюють у режимі реального часу.

Ключові слова: розпізнавання образів, модель розпізнавальних операторів, потенційна функція, залежність ознак, підмножина сильнозв'язаних ознак, репрезентативна ознака, краща модель залежності.

Fazilov Sh. Kh.[1], Mizraev N. M.[2], Mizraev O. N.[3]
[1]D. Sc., Professor, Head of the Laboratory of the Centre for development of software and hardware-program complexes at Tashkent University of information technologies, Tashkent, Uzbekistan
[2]PhD, Senior Researcher at the Centre for Development of Software and Hardware-program Complexes at Tashkent University of information technologies, Tashkent, Uzbekistan
[3]Junior researcher at the Centre for Development of Software and Hardware-program Complexes at Tashkent University of information technologies, Tashkent, Uzbekistan

BUILDING OF RECOGNITION OPERATORS IN CONDITION OF FEATURES' CORRELATIONS

The problem of recognizing of patterns given in the space of correlated features is considered. The new approach to the building of model of recognition operators, which considers the correlation of given features, is proposed. The building of the model is carried out for potential function type recognition operators. The main idea of the proposed approach is formation of uncorrelated subsets of strongly correlated features and extracting preferred correlation model for each of subsets of strongly correlated features. Analysis of the results shows that the considered recognition operators are used in cases when there is a certain correlation between objects belonging to the same class. When the expression of this relationship is weak, classical model of recognition operators is used. The main advantage of the proposed recognition operators is to improve the accuracy and the significant reduction in the volume of computational operations in recognition of unknown objects, which allows them to use when building recognition systems working in real time.

Keywords: pattern recognition, model of recognition operators, potential function, features' correlations, subset of strongly correlated features, representative feature, preferred correlation model.

REFERENCES

1. Ayzerman M. A., Braverman E. M., Ayzerman M. A., Rozonoer L. I. Metod potentsialnyh funktsiy v teorii obucheniya mashin. Moscow, Nauka, 1970, 348 p.
2. Zhuravlev Yu.I Izbranniye nauchniye trudy. Moscow, Magistr, 1998, 420 p.
3. Zhuravlev Yu. I., Ryazanov V. V., Senko O. V. Raspoznavaniya. Matematicheskiye metody. Programmnaya sistema. Prakticheskiye primeneniya. Moscow, Fazis, 2006, 159 p.
4. Kamilov M. M., Mirzayev N. M., Radjabov S. S. Sovremennoye sostoyaniye voprosov postroyeniya modeley algoritmov raspoznavaniya, *Nauchniy zhurnal: Himicheskaya tehnologiya. Kontrol i upravleniye,* 2009, No. 2, pp. 21–27.
5. Zagoruyko N.G. Prikladniye metody analiza dannyh i znaniy. Novosibirksk, IM SO RAN, 1999.
6. Lbov G. S., Startseva N. G. Logicheskiye reshayushiye funktsii i voprosy statisticheskoy ustoychivosti resheniy. Novovsibirsk, Izd-vo IM SO RAN, 1999, 211 p.
7. Shlezinger M., Glavach V. Desyat lektsiy po statisticheskomu i strukturnomu rasposnavaniyu. Kiev, Naukova dumka, 2004, 545 p.
8. Potapov A. S. Raspoznavaniye obrazov i mashinnoye vospriyatiye. Sankt-Peterburg, Politehnika, 2007, 548 p.
9. Duda R. O. Hart P. E. Stork D. G. Pattern Classification, Second Edition. New York, John Wiley, Inc., 2001, 680 p.
10. Vapnik V. Statistical Learning Theory. New York, John-Wiley Sons, Inc., 1998, 732 p.
11. Theodoridis S., Koutroumbas K. Pattern Recognition: Theory and Applications, 4th edition. New York, Academic Press, 2009, 957 p.
12. Dougherty G. Pattern Recognition and Classification: An Introduction. New York, Springer, 2013, 196 p.
13. Fazylov Sh. H., Mirzayev N. M., Radjabov S. S. Postroyeniye modeli algoritmov vychisleniya otsenok s uchyotom bolshoy razmernosti priznakovogo prostranstva, *Vestnik SGTU*, Saratov, 2012. – No. 1 (64), Vyp. 2, pp. 274–279.
14. Mirzaev N. M. Algoritmy vydeleniya podmnozhestv silnosvyazannyh priznakov, *Voprosy kibernetiki: Sb.nauch.tr.* Tashkent, IMIT AN RUz, 2008, Vyp.177, pp. 99–104.
15. Mirzaev O. N. Vydeleniye perpezentativnyh priznakov pri postroyenii algoritmov raspoznavaniya, *Problemy informatiki i energetiki*. Tashkent, 2008, No. 6, pp. 23–27.
16. Fazylov Sh. H., Mirzayev N. M., Radjabov S. S., Karimov I. K. Model raspoznayushih operatorov, osnovannyh na printsipe blijayshego soseda, v usloviyah vzaimosvyazannosti priznakov, *Informatika i sistemy upravleniya*. Blagoveshensk, 2012, No. 4, pp. 34–42.
17. Mirzayev N. M., Radjabov S. S., Zhumayev T. S. O parametrizatsii modeley algoritmov raspoznavaniya, osnovannyh na otsenke vzaimosvyazannosti priznakov, *Problemy informatiki i energetiki*, Tashkent, 2008, No. 2–3, pp. 23–27.
18. Vorontsov K.V., Frey A. I., Sokolov E. A. Computable Combinatorial Overыtting Bounds, *Journal of Machine Learning and Data Analysis*, 2013, Vol. 1, No. 6, pp. 734–743.

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

ПРОГРЕССИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

PROGRESSIVE INFORMATION TECHNOLOGIES

УДК 681.324

Бабаков Р. М.

Канд. техн. наук, доцент, доцент кафедры компьютерных наук Донецкого национального технического университета, г. Красноармейск, Украина

ПРОМЕЖУТОЧНАЯ АЛГЕБРА ПЕРЕХОДОВ В МИКРОПРОГРАММНОМ АВТОМАТЕ

Решена задача формализации задания микропрограммного автомата, в структуре которого часть автоматных переходов реализуется неканоническим способом. Предложен новый подход к организации функции переходов микропрограммного автомата, в соответствии с которым функция переходов представляется в виде семейства частичных функций, каждая из которых определена лишь на части области определения функции переходов автомата и соответствует некоторому подмножеству автоматных переходов.

С учетом предложенного подхода традиционное представление автомата в виде многоосновной алгебры претерпевает ряд изменений. Во-первых, взаимная независимость функций переходов и выходов, образующих сигнатуру алгебры, позволяет рассматривать их отдельно друг от друга, что приводит к представлению автомата в виде двух алгебр: алгебры переходов, сигнатура которой содержит только функцию переходов, и алгебры выходов, сигнатура которой содержит только функцию выходов. Во-вторых, представление функции переходов в виде множества частичных функций приводит к замене алгебры переходов множеством подалгебр переходов, в каждой из которых сигнатура образована частичной функцией переходов.

На примере микропрограммного автомата со счетчиком показано, что закон преобразования кодов состояний в рамках некоторого подмножества переходов может быть задан некоторой алгебраической функцией (операцией переходов), использующей скалярную интерпретацию кодов состояний структурного автомата. Скалярную интерпретацию кодов состояний совместно с операцией переходов предлагается представлять в виде т.н. промежуточной алгебры переходов, изоморфной соответствующим подалгебрам переходов абстрактного и эквивалентного ему структурного автоматов.

Ключевые слова: микропрограммный автомат, частичная функция переходов, промежуточная алгебра переходов, автомат на счетчике.

НОМЕНКЛАТУРА

МПА – микропрограммный автомат;

S_A – абстрактный автомат;

A – множество состояний абстрактного автомата;

Z – входной алфавит абстрактного автомата;

W – выходной алфавит абстрактного автомата;

G_A – абстрактная алгебра, задающая абстрактный автомат;

$\mathcal{A}_A$ – носитель абстрактной алгебры;

$\mathcal{F}_A$ – сигнатура абстрактной алгебры;

δ – функция переходов абстрактного автомата;

λ – функция выходов абстрактного автомата;

G_δ – абстрактная алгебра переходов;

$\mathcal{A}_\delta$ – носитель абстрактной алгебры переходов;

$\mathcal{F}_\delta$ – сигнатура абстрактной алгебры переходов;

G_λ – абстрактная алгебра выходов;

$\mathcal{A}_\lambda$ – носитель абстрактной алгебры выходов;

$\mathcal{F}_\lambda$ – сигнатура абстрактной алгебры выходов;

$a_н$ – начальное состояние автомата;

D_δ – область определения функции переходов абстрактного автомата;

δ_i – частичная функция переходов абстрактного автомата;

N_δ – количество частичных функций переходов δ_i;

B – количество переходов автомата;

a_i – текущее состояние абстрактного автомата;

z_j – входной сигнал абстрактного автомата;

a_k – состояние перехода абстрактного автомата;

D_{δ_i} – область определения функции δ_i;

G_{δ_i} – частичная абстрактная подалгебра переходов;

DOI 10.15588/1607-3274-2016-1-8

$\mathcal{A}_{\delta_i}$ – носитель частичной абстрактной подалгебры переходов G_{δ_i};

$\mathcal{F}_{\delta_i}$ – сигнатура частичной абстрактной подалгебры переходов G_{δ_i};

A_{δ_i} – подмножество состояний, встречающихся в кортежах функции δ_i;

Z_{δ_i} – подмножество входных сигналов, встречающихся в кортежах функции δ_i;

S_S – структурный автомат;

G_S – структурная алгебра переходов;

$\mathcal{A}_S$ – носитель структурной алгебры переходов;

$\mathcal{F}_S$ – сигнатура структурной алгебры переходов;

d – структурная функция переходов;

l – структурная функция выходов;

X – множество структурных входных сигналов;

Y – множество структурных выходных сигналов;

G_d – структурная алгебра переходов;

$\mathcal{A}_d$ – носитель структурной алгебры переходов;

$\mathcal{F}_d$ – сигнатура структурной алгебры переходов;

G_l – структурная алгебра выходов;

$\mathcal{A}_l$ – носитель структурной алгебры выходов;

$\mathcal{F}_l$ – сигнатура структурной алгебры выходов;

D_d – область определения структурной функции переходов;

d_i – частичная структурная функция переходов;

N_d – количество частичных структурных функций переходов;

D_{d_i} – область определения функции d_i;

F – мощность множества Z;

L – мощность множества X;

G – мощность множества W;

N – мощность множества Y;

$K_S(a)$ – структурный код состояния автомата;

$K_S(z)$ – структурный код входного сигнала;

T – структурный код текущего состояния автомата;

R – разрядность кода T;

КС$_{d_i}$ – комбинационная схема, реализующая структурную функцию переходов d_i;

МХ – мультиплексор;

КС$_{Inc}$ – схема управления мультиплексором МХ;

РП – регистр памяти;

СФМО – схема формирования микроопераций;

СТ – инкрементный счетчик состояний;

$K_I(a_i)$ – промежуточный код состояния a_i;

G_I – промежуточная алгебра переходов;

$\mathcal{A}_I$ – носитель промежуточной алгебры переходов;

$\mathcal{F}_I$ – сигнатура промежуточной алгебры переходов;

O_{Inc} – операция инкремента;

$K_I(A_{\delta_i})$ – множество кодов $K_I(a_i)$, используемых в функции δ_i.

ВВЕДЕНИЕ

Важным компонентом современных вычислительных систем является устройство управления, одним из способов реализации которого является микропрограммный автомат [1]. Наблюдаемый сегодня рост сложности алгоритмов, интерпретируемых МПА, способствует увеличению аппаратурных затрат в логической схеме автомата. В данном аспекте актуальной является задача снижения аппаратурных затратах в схеме МПА, одно из решений которой состоит в разработке новых структур МПА и их формальных математических моделей. Классические математические модели абстрактного и структурного автоматов подходят для описания предложенной В. М. Глушковым канонической структуры МПА, однако при использовании иных структур данные модели в некоторых случаях не отражают внутренние информационные процессы, ограничивая возможности проектировщика цифровых устройств в вопросах оптимизации логической схемы синтезируемого устройства.

Объектом исследования в работе выступает микропрограммный автомат, предметом исследования – математическая модель МПА, предполагающая специальное представление функции переходов автомата. Цель работы состоит в разработке и формализации такой модели и достигается путем анализа и обобщения свойств одной из известных структур МПА – автомата на счетчике.

1 ПОСТАНОВКА ЗАДАЧИ

Основной научной задачей, решаемой в данной работе, является разработка формальных математических моделей абстрактного и структурного автоматов, использующих представление функции переходов автомата в виде множества частичных функций. Модели должны обеспечивать выражение закономерностей преобразования информации в части функции переходов структурного автомата, подчеркивая эквивалентность данной функции с функцией переходов тождественного абстрактного автомата. Подтверждение состоятельности предлагаемых моделей должно быть продемонстрировано на примере одной из известных структур МПА.

2 ОБЗОР ЛИТЕРАТУРЫ

Современная теория автоматов имеет тесную связь с теорией универсальных алгебр. Алгебраический подход находит применение в теории автоматов, способствуя установлению связи теории автоматов с другими областями математики. Математический аппарат универсальных алгебр лежит в основе алгебраической теории автоматов, изучающей дискретные автоматы с абстрактно-алгебраической точки зрения [2, 3].

Базовым объектом изучения в теории универсальных алгебр является алгебраическая система, определяемая как абстрактный объект

$$G =< \mathcal{A}, \mathcal{F}, \mathcal{P} >, \quad (1)$$

состоящий из трех множеств: непустого множества $\mathcal{A}$, множества операций $\mathcal{F} = \{\mathcal{F}_0,...,\mathcal{F}_\xi,....\}$, определенных на $\mathcal{A}$, и множества предикатов (отношений) $\mathcal{P} = \{\mathcal{P}_0,...,\mathcal{P}_\eta...\}$, заданных на $\mathcal{A}$ [4]. Под операцией

здесь понимается функциональное отношение любой арности, заданное на множестве $\mathcal{A}$. Под предикатом понимается функция на множестве $\mathcal{A}$, значения которой, называемые модальностями или степенями истинности, лежат в особом множестве, состоящем из двух элементов – «истина» и «ложь» [4, 5].

Множество $\mathcal{A}$ называют носителем системы *G*. Набор множеств $< \mathcal{F}, \mathcal{P} >$ образует сигнатуру или структуру алгебраической системы. В случае конечности носителя данные множества также конечны. Сигнатуру с пустым множеством отношений называют функциональной, с пустым множеством операций – предикатной [6].

Алгебраическая система (1) называется *универсальной алгеброй* или просто *алгеброй*, если ее сигнатура функциональна, и *моделью* (реляционной системой), если ее сигнатура предикатна [4]. Таким образом, алгеброй является следующий двухэлементный объект [4, 5]:

$$G =< \mathcal{A}, \mathcal{F} >. \qquad (2)$$

Иногда носитель алгебры представляется несколькими совокупностями разнородных объектов, объединение которых в единое множество является неестественным [7]. В этом случае оказывается удобным представлять носитель в виде семейства непересекающихся множеств. Пусть $\mathcal{A} = \{\mathcal{A}_1, ..., \mathcal{A}_k\}$ – множество основ (носителей), $\mathcal{F} = \{f_1, ..., f_m\}$ – сигнатура, причем $f_i : \mathcal{A}_i^1 \times \mathcal{A}_i^2 \times ... \times \mathcal{A}_i^{k_i} \to \mathcal{A}_j$. Тогда система (2) называется многоосновной (*k*-основной) алгеброй [5, 7]. Многоосновная алгебра имеет несколько носителей, а каждая операция сигнатуры является соответствием, действующим из прямого произведения некоторых носителей в некоторый носитель [8].

В теории автоматов абстрактный и структурный автоматы иногда рассматриваются как алгебры, поскольку их структура как математических объектов подобна структуре алгебраической системы [9]. Представление автомата в виде алгебры дает возможность использовать алгебраические приемы при решении таких задач, как декомпозиция автоматов, их классификация, описание поведения и другие [3].

3 МАТЕРИАЛЫ И МЕТОДЫ

Дадим собственную алгебраическую интерпретацию конечных автоматов.

Абстрактный автомат.

Как известно, конечный абстрактный автомат S_A представляется совокупностью трех конечных множеств – состояний, входных и выходных слов, и двух детерминированных функций – переходов и выходов: $S_A =< A, Z, W, \delta, \lambda >$ [10]. В таком виде автомат может рассматриваться как трехосновная алгебра с двумя операциями [7]. Будем называть алгебру, отождествляемую с абстрактным автоматом, *абстрактной алгеброй* G_A:

$$G_A =< \mathcal{A}_A, \mathcal{F}_A >, \qquad (3)$$

$$\mathcal{A}_A = \{A, Z, W\}, \qquad (4)$$

$$\mathcal{F}_A = \{\delta, \lambda\}. \qquad (5)$$

Фактически, выражения (3)–(5) перечисляют объекты, которые должны быть определены для задания абстрактного автомата. При этом способ задания объектов не имеет значения и может быть различным [2].

Функции δ и λ реализуют различные соответствия и могут рассматриваться отдельно друг от друга. Это оказывается полезным, например, в тех случаях, когда выходные реакции автомата неизвестны или не представляют интереса, а объектом исследования является доступная наблюдению эволюция его внутренних состояний под воздействием входных сигналов. В работах [3, 7, 11] упоминаются т.н. *автоматы без выходов* (полуавтоматы, X-автоматы, редукты) – объекты, задаваемые только с помощью множества состояний, входного алфавита и функции переходов.

Абстрактный автомат без функции выходов есть двухосновная алгебра:

$$G_\delta = \ < \mathcal{A}_\delta, \mathcal{F}_\delta > =< \{A, Z\}, \ \{\delta\} >, \qquad (6)$$

без функции переходов – трехосновная алгебра

$$G_\lambda = \ < \mathcal{A}_\lambda, \mathcal{F}_\lambda > =< \{A, Z, W\}, \{\lambda\} >. \qquad (7)$$

Выражение (7) справедливо для автомата Мили. Для автомата Мура необходимо исключить компонент *Z*:

$$G_\lambda = \ < \mathcal{A}_\lambda, \mathcal{F}_\lambda > =< \{A, W\}, \{\lambda\} >. \qquad (8)$$

Будем называть алгебру (6) *абстрактной алгеброй переходов*, алгебры (7) и (8) – *абстрактной алгеброй выходов*. Тогда абстрактный автомат есть двойка

$$S_A =< G_\delta, G_\lambda >. \qquad (9)$$

В случае выделения на множестве состояний специального начального состояния (инициальный автомат) в выражение (9) может быть добавлен компонент $a_н \in A$:

$$S_A =< G_\delta, G_\lambda, a_н >. \qquad (10)$$

Как известно, в абстрактном автомате функция *d* есть соответствие

$$\delta : (D_\delta \subseteq A \times Z) \to A. \qquad (11)$$

Будучи представленной в виде (11), функция δ считается, в общем случае, частично определенной на множестве $A \times Z$. Исходя из математического определения соответствия, даваемого в [12], функция (11) есть множество кортежей

$$< a_i, z_j, a_k > \in \delta, \qquad (12)$$

каждый из которых соответствует одному автоматному переходу из состояния $a_i \in A$ под воздействием входного сигнала $z_j \in Z$ в состояние $a_k \in A$.

Разобьем некоторым образом множество кортежей вида (12) на несколько непустых, попарно непересекающихся подмножеств, максимально возможное число которых равно числу переходов автомата. Каждое такое подмножество переходов автомата, будем называть *частичной функцией переходов* δ_i, понимая здесь под термином «частичная» ее неполное определение на области определения D_δ функции переходов δ:

$$\delta_i : (D_{\delta_i} \subseteq D_\delta) \to A. \qquad (13)$$

В этом случае функция переходов δ есть множество подмножеств кортежей (12), то есть множество частичных функций переходов:

$$\delta = \{\delta_1, ..., \delta_{N_\delta}\}, \quad (14)$$

где $1 \le N_\delta \le B$. При этом $\bigcup_{i=1}^{N_\delta} \delta_i = \delta$.

Подобное разбиение может быть выполнено и для функции выходов.

Свяжем с каждой частичной функцией переходов d_i собственную алгебру вида

$$G_{\delta_i} = <\mathcal{A}_{\delta_i}, \mathcal{F}_{\delta_i}> = <\{A_{\delta_i}, Z_{\delta_i}\}, \{\delta_i\}>. \quad (15)$$

В данном выражении $A_{\delta_i} \subseteq A$ есть множество всех состояний, присутствующих в кортежах вида (12) функции δ_i. Иными словами, в A_{δ_i} входят те и только те состояния, переходы из которых или переходы в которые реализуются функцией δ_i. Множество $Z_{\delta_i} \subseteq Z$ есть множество входных сигналов, присутствующих в кортежах вида (12) функции δ_i. Поскольку $A_{\delta_i} \subseteq A$, $Z_{\delta_i} \subseteq Z$ и для любых $a \in A_{\delta_i}$ и $z \in Z_{\delta_i}$ выполняется равенство $\delta_i(a, z) = \delta(a, z)$, выражение (15) есть *подалгебра* алгебры (6) [4]. Будем называть алгебру (15) *абстрактной подалгеброй переходов*. В алгебраической теории автоматов понятию подалгебры соответствует понятие *подавтомата* [2, 3].

Множество частичных функций переходов приводит ко множеству абстрактных подалгебр переходов, отождествляемому с абстрактной алгеброй переходов (6):

$$G_\delta = \{G_{\delta_1}, ..., G_{\delta_{N_\delta}}\}. \quad (16)$$

Компоненты $\mathcal{A}_\delta$ и $\mathcal{F}_\delta$ алгебры G_δ формируются путем объединения соответствующих компонентов всех подалгебр:

$$\mathcal{A}_\delta = \bigcup_{i=1}^{N_\delta} \mathcal{A}_{\delta_i} = \{\bigcup_{i=1}^{N_\delta} A_{\delta_i}, \bigcup_{i=1}^{N_\delta} Z_{\delta_i}\}, \mathcal{F}_\delta = \bigcup_{i=1}^{N_\delta} \mathcal{F}_{\delta_i} = \{\bigcup_{i=1}^{N_\delta} \delta_i\}.$$

Отметим, что в то время как частичные функции δ_i образуют *разбиение*, т.е. $\forall(i, j \in [1, N_\delta], i \ne j): \delta_i \cap \delta_j = \varnothing$, семейства множеств A_{δ_i} и Z_{δ_i} образуют на множествах A и Z соответствующие *покрытия*, что в общем случае допускает $\forall(i, j \in [1, N_\delta]): A_{\delta_i} \cap A_{\delta_j} \ne \varnothing$ и $Z_{\delta_i} \cap Z_{\delta_j} \ne \varnothing$. Присутствие одного и того же состояния $a \in A$ в носителях нескольких подалгебр означает, что переходы в данное состояние и/или из него выполняются с помощью разных частичных функций δ_i.

По аналогии с (16), в виде множества подалгебр представляются абстрактная алгебра выходов (7) автомата Мили и абстрактная алгебра выходов (8) автомата Мура.

Структурный автомат.

Модель структурного автомата $S_S = < X, Y, A, d, l >$ также является трехосновной алгеброй, носитель $\mathcal{A}_S$ которой образован конечными множествами структурного автомата, а сигнатура $\mathcal{F}_S$ – структурной функцией переходов d и структурной функцией выходов l:

$$\mathcal{A}_S = \{X, A, Y\}, \quad (17)$$

$$\mathcal{F}_S = \{d, l\}. \quad (18)$$

Отметим, что в выражении (17) множество состояний A отличается от одноименного множества в выражении (4) тем, что содержит не абстрактные, а структурные состояния, представленные своими структурными кодами [10].

Назовем алгебру, отождествляемую со структурным автоматом, *структурной алгеброй* G_S:

$$G_S = <\mathcal{A}_S, \mathcal{F}_S> = <\{X, A, Y\}, \{d, l\}>. \quad (19)$$

Разделение функций переходов и выходов дает две отдельные алгебры: *структурную алгебру переходов:*

$$G_d = <\mathcal{A}_d, \mathcal{F}_d> = <\{A, X\}, \{d\}> \quad (20)$$

и *структурную алгебру выходов* (выражение (21) – для автомата Мили, (22) – для автомата Мура):

$$G_l = <\mathcal{A}_l, \mathcal{F}_l> = <\{A, X, Y\}, \{l\}>, \quad (21)$$

$$G_l = <\mathcal{A}_l, \mathcal{F}_l> = <\{A, Y\}, \{l\}>. \quad (22)$$

При этом структурный автомат может рассматриваться как двойка

$$S_S = < G_d, G_l > . \quad (23)$$

Согласно модели $S_S = < X, Y, A, d, l >$, функция d есть соответствие

$$d: (D_d \subseteq X \times A) \to A. \quad (24)$$

В структурном автомате все элементы множеств абстрактного автомата представляются в виде слов в структурном (двоичном) алфавите. Предположение о том, что память структурного автомата строится в базисе триггеров D-типа, дает возможность отождествлять результат структурной функции переходов (структурный код состояния перехода) со множеством функций возбуждения элементов памяти структурного автомата.

При $|X| = L$ и $\lceil \log_2(|A|) \rceil = R$ выражение (24) принимает вид:

$$d: (D_d \subseteq x_1 \times x_2 \times ... \times x_L \times e_1 \times e_2 \times ... \times e_R) \to e_1 \times e_2 \times ... \times e_R, \quad (25)$$

где каждый из элементов x и e есть множество букв структурного (двоичного) алфавита, т.е. множество {0, 1} [9]. Фактически, (25) есть соответствие:

$$d: (D_d \subseteq \{0,1\}^{L+R}) \to \{0,1\}^R, \quad (26)$$

элементами которого являются кортежи, состоящие из (L+2R) букв $q \in \{0, 1\}$:

$$< q_1, ..., q_{L+2R} > \in D_d. \quad (27)$$

Структурная функция переходов (24), являясь множеством кортежей (27), может быть представлена в виде множества *частичных структурных функций* переходов:

$$d = \{d_1, ..., d_{N_d}\}, \quad (28)$$

$$d_i: (D_{d_i} \subseteq D_d) \to A. \quad (29)$$

Отметим, что в выражении (28) максимально возможное количество частичных функций N_d определяется числом кортежей (27). Если в случае абстрактного автомата имеем $1 \le N_\delta \le B$, то для значения N_d дело обстоит иначе.

В процессе кодирования каждому символу z_f входного алфавита абстрактного автомата ставится в соответствие вектор вида $< x_1, ..., x_L >$, образованный значениями переменных $x_1, ..., x_L$, соответствующих одноименным входным структурным сигналам. Поскольку способ кодирования элементов множества Z может быть произвольным, возможны ситуации, когда в векторе $< x_1, ..., x_L >$, соответствующем входному символу $z_f \in Z$, некоторая переменная x_l может принимать произвольное значение [13]. В этом случае каждому переходу, выполняющемуся под воздействием входного сигнала z_f, в структурной функции d соответствуют два вектора вида (27), в первом из которых компонент, соответствующий x_l, равен нулю, во втором – единице.

В общем случае количество фиктивных переменных в коде символа z_f может быть больше единицы. Такие ситуации часто возникают в процессе структурного синтеза МПА, заданных в виде ГСА, когда переход из состояния в состояние зависит лишь от части логических условий, отождествляемых со структурными входными сигналами. В работе [13] предлагается каждый структурный входной сигнал условно представлять заданным на трехзначном множестве {0, 1, –}, где элемент «–» соответствует произвольному значению из {0, 1}. Будем полагать, что каждая из функций d_i есть множество кортежей вида (27) с трехзначным представлением структурных входных сигналов. Тогда, как и для абстрактных частичных функций, имеем $1 \le N_d \le B$. Понятно, что при $N_d = B$ каждому переходу автомата будет сопоставлена отдельная функция вида (29), в то время как при $N_d = 1$ все переходы автомата реализуются одной общей структурной функцией переходов.

Установим теперь формальную связь между абстрактной и структурной алгебрами переходов. Однотипность алгебраических систем характеризуется возможностью установления между ними различных отображений, таких как *гомоморфизм*, *изоморфизм* и др. [4]. Воспользуемся понятием *гомоморфизма автоматов*, которое В. М. Глушков определяет следующим образом [11]. Гомоморфизм ψ автомата $A(Z_A, A_A, W_A, \delta_A, \lambda_A)$ в автомат $B(Z_B, A_B, W_B, \delta_B, \lambda_B)$ есть совокупность трех однозначных отображений: $\psi_1 : Z_A \to Z_B$, $\psi_2 : A_A \to A_B$, $\psi_3 : W_A \to W_B$, удовлетворяющих для любых элементов $a \in A_A$ и $z \in Z_A$ соотношениям:

$$\psi_1(\delta_A(a,z)) = \delta_B(\psi_1(a), \psi_2(z)), \qquad (30)$$

$$\psi_3(\lambda_A(a,z)) = \lambda_B(\psi_1(a), \psi_2(z)). \qquad (31)$$

Хотя МПА в части функции переходов может быть задан как абстрактной алгеброй (3), так и структурной алгеброй (19), установление гомоморфизма между данными алгебрами, с формальной точки зрения, невозможно. Причиной является несоответствие носителей (4) и (17), а точнее – различные мощности множеств, образующих носители. Тот факт, что в общем случае $(|Z| = F) \ne (|X| = L)$ и $(|W| = G) \ne (|Y| = N)$, не позволяет установить между элементами соответствующих множеств взаимно однозначное соответствие, обязательное для гомоморфизма.

Преобразуем структурную алгебру (19) следующим образом.

1. Пусть носитель $\mathcal{A}_S$ содержит следующие три множества

– множество $K_S(Z) = \{K_S(z_1), \ldots, K_S(z_F)\}$ структурных кодов входных сигналов абстрактного автомата, где каждый элемент $K_S(z_f)$ есть вектор $< x_1, ..., x_L >$ значений структурных входных сигналов, определенных на трехэлементном алфавите {0, 1, –};

– множество $K_S(A) = \{K_S(a_1), \ldots, K_S(a_M)\}$ структурных кодов состояний абстрактного автомата, заданных векторами $< e_1, ..., e_R >$, $e_r \in \{0, 1\}$;

– множество $K_S(W) = \{K_S(w_1), \ldots, K_S(w_G)\}$ структурных кодов выходных сигналов абстрактного автомата, заданных векторами $< y_1, ..., y_N >$, $y_n \in \{0, 1\}$.

2. Структурные функции переходов d и выходов l, образующие сигнатуру $\mathcal{F}_S$, определим выражениями (32) и (33), (34) соответственно:

$$d: \ (D_d \subseteq K_S(Z) \times K_S(A)) \to K_S(A). \qquad (32)$$

$$l: \ (D_d \subseteq K_S(Z) \times K_S(A)) \to K_S(W). \qquad (33)$$

$$l: \ (D_d \subseteq K_S(A)) \to K_S(W). \qquad (34)$$

Выражение (33) справедливо для автомата Мили, (34) – для автомата Мура.

Тогда структурная алгебра (19) принимает следующий вид:

$$G_S = < \mathcal{A}_S, \mathcal{F}_S > = < \{K_S(Z), K_S(A), K_S(W)\}, \{d, l\} >. \qquad (35)$$

Структурные алгебры переходов и выходов определяются аналогично (20)–(22) следующим образом:

$$G_d = < \mathcal{A}_d, \mathcal{F}_d > = < \{K_S(A), K_S(Z)\}, \{d\} >, \qquad (36)$$

$$G_l = < \mathcal{A}_l, \mathcal{F}_l > = < \{K_S(A), K_S(Z), K_S(W)\}, \{l\} >, \qquad (37)$$

$$G_l = < \mathcal{A}_l, \mathcal{F}_l > = < \{K_S(A), K_S(W)\}, \{l\} >. \qquad (38)$$

Структурная алгебра (35) отличается от (19) тем, что формальными аргументами и значениями ее операций являются не отдельные структурные сигналы, а элементы носителей абстрактной алгебры, представленные в виде векторов элементарных структурных сигналов. Это позволяет установить гомоморфизм абстрактной алгебры (3) в структурную алгебру (35). Поскольку между носителями структурной алгебры (35) и носителями абстрактной алгебры (3) могут быть установлены однозначные соответствия $K_S(A) \to A$, $K_S(Z) \to Z$, $K_S(W) \to W$, гомоморфизм алгебр (3) и (35) является *изоморфизмом* (взаимно однозначным гомоморфизмом, [11]). Очевидно, что существуют также изоморфизмы между алгебрами переходов (6) и (36), а также между алгебрами выходов: (7) и (37) для автомата Мили, (8) и (38) для автомата Мура.

Функция переходов (32) есть множества кортежей следующего вида:

$$< K_S(a_i), K_S(z_j), K_S(a_k) > \in D_d \quad (39)$$

и может быть представлена в виде семейства частичных функций переходов (28), причем

$$d_i : (D_{d_i} \subseteq D_d) \to K_S(A). \quad (40)$$

Изоморфизм алгебр (3) и (35) устанавливает, в том числе, отображение между элементами соответствующих носителей данных алгебр. Сопоставляя каждому компоненту любого кортежа (12) соответствующий ему элемент из носителей структурной алгебры, можно получить тождественный кортеж (39). Следовательно, каждой частичной функции переходов δ_i абстрактного автомата, определяемой выражением (13) и являющейся некоторым подмножеством множества кортежей (12), может быть поставлена в соответствие частичная функция переходов d_i, определяемая выражением (40) и являющаяся подмножеством множества кортежей (39). При этом изоморфизм абстрактной и структурной алгебр возможен при любом разбиении функций абстрактного автомата на множества частичных функций. Данное утверждение также справедливо как алгебр переходов (6) и (36) и для алгебр выходов (7)/(8) и (37)/(38).

По аналогии с (15), с*труктурной подалгеброй переходов* будем называть алгебру вида:

$$G_{d_i} = < \mathcal{A}_{d_i}, \mathcal{F}_{d_i} > = < \{K_S(A_{d_i}), K_S(Z_{d_i})\}, \{d_i\} > . \quad (41)$$

Здесь $K_S(A_{d_i}) \subseteq K_S(A)$ – множество структурных кодов состояний, присутствующих в кортежах (39) частичной структурной функции d_i; $K_S(Z_{d_i}) \subseteq K_S(Z)$ – множество структурных кодов входных сигналов, используемых для кодирования входных сигналов в кортежах (39) функции d_i. Структурную алгебру переходов (35), по аналогии с (16), представим в виде множества структурных подалгебр переходов:

$$G_d = \{G_{d_1}, ..., G_{d_{N_d}}\}. \quad (42)$$

Как и в случае абстрактного автомата, компоненты структурной алгебры переходов (42) формируются объединением элементов соответствующих компонентов своих подалгебр. Аналогичным образом в виде множества своих подалгебр могут быть представлены структурные алгебры выходов (37) и (38).

4 ЭКСПЕРИМЕНТЫ

Рассмотрим отдельно функцию переходов автомата. В соответствии с предложенным В.М. Глушковым каноническим методом структурного синтеза функция переходов представляется в векторном виде и задается системой канонических скалярных уравнений [10]. Система скалярных уравнений строится известным способом [10, 13], при котором для каждой пары $< K_S(a_i), K_S(z_j) >$ кортежа (39), соответствующего одному переходу автомата, формируется терм, присутствие которого в том или ином уравнении системы определяется элементом

В структурной теории автоматов известен класс устройств, называемых автоматами на счетчике или С-автоматами [14]. В их основе лежит принцип специального кодирования состояний, при котором на множестве состояний заданной ГСА выделяются т.н. линейные последовательности состояний, внутри которых коды состояний задаются в естественном (последовательном) порядке.

Применим принцип представления функции переходов автомата в виде множества частичных функций к автомату на счетчике. Пусть произвольный абстрактный автомат задан алгеброй (3), а реализующий его автомат на счетчике – алгеброй (35). Функции переходов данных автоматов образуют сигнатуры абстрактной алгебры переходов (6) и структурной алгебры переходов (36) соответственно. В силу изоморфизма алгебр (3) и (35), алгебры переходов (6) и (36) также изоморфны:

$$G_\delta \leftrightarrow G_S. \quad (43)$$

Тот факт, что в С-автомате часть переходов реализуется с использованием инкрементного счетчика, часть – в виде системы канонических уравнений, позволяет представить функцию d в виде двух частичных функций d_1 и d_2, образующих сигнатуры двух подалгебр переходов:

$$G_{d_1} = < \mathcal{A}_{d_1}, \mathcal{F}_{d_1} > = < \{K_S(A_{d_1}), K_S(Z_{d_1})\}, \{d_1\} >; \quad (44)$$

$$G_{d_2} = < \mathcal{A}_{d_2}, \mathcal{F}_{d_2} > = < \{K_S(A_{d_2}), K_S(Z_{d_2})\}, \{d_2\} > . \quad (45)$$

Представление функции d в виде двух частичных функций предполагает аналогичное представление абстрактной функции переходов δ. Частичные абстрактные функции переходов δ_1 и δ_2 образуют сигнатуры двух абстрактных подалгебр переходов:

$$G_{\delta_1} = < \mathcal{A}_{\delta_1}, \mathcal{F}_{\delta_1} > = < \{A_{\delta_1}, Z_{\delta_1}\}, \{\delta_1\} >; \quad (46)$$

$$G_{\delta_2} = < \mathcal{A}_{\delta_2}, \mathcal{F}_{\delta_2} > = < \{A_{\delta_2}, Z_{\delta_2}\}, \{\delta_2\} >. \quad (47)$$

Из изоморфизма (43) вытекают изоморфизмы соответствующих подалгебр:

$$G_{\delta_1} \leftrightarrow G_{d_1}; \quad (48)$$

$$G_{\delta_2} \leftrightarrow G_{d_2}. \quad (49)$$

Структурная схема С-автомата, отражающая представление структурной функции переходов в виде двух частичных функций, приведена на рис. 1.

Здесь схема КС_{d_1} реализует частичную структурную функцию переходов d_1 и представляет собой R-разрядный инкрементор. Структурный код T текущего состояния, представленный набором структурных переменных $< T_1, ..., T_R >$, в котором $T_r \in \{0, 1\}$, интерпретируется на входе КС_{d_1} как R-разрядное беззнаковое целое, представленное в структурном (двоичном) алфавите. На выходе инкрементора формируется структурный код состояния перехода, обозначенный символом d_1 и отождествляемый со значением одноименной функции переходов.

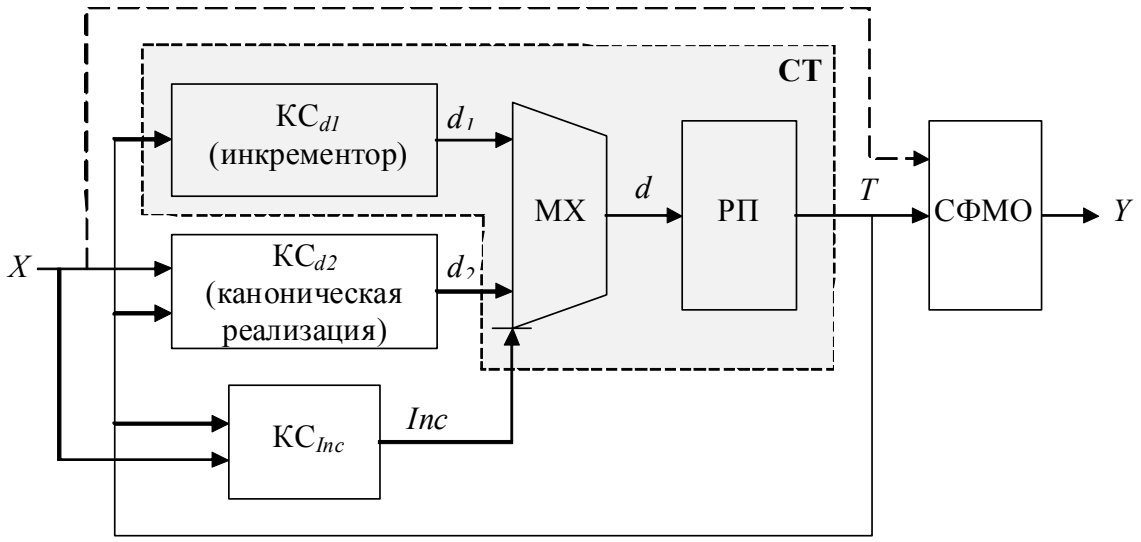


Рисунок 1 – Структурная схема С-автомата

Частичная функция d_2 представлена комбинационной схемой КС$_{d_2}$. Данная схема строится по известной методике [14] и реализует систему канонических уравнений, соответствующую функции d_2 На вход КС$_{d_2}$ поступают структурный код текущего состояния $T =< T_1, ..., T_R >$ и структурный код входного слова, представленный набором $X =< x_1, ..., x_L >$, в котором $x_l \in \{0, 1, -\}$. Код, формируемый на выходе КС$_{d_2}$, обозначен символом d_2 и отождествляется с значением одноименной функции переходов.

Поскольку каждая функция d_i, представленная в виде логической схемы, определена на всех наборах аргументов, но при этом отдельный автоматный переход реализуется всегда какой-то одной частичной функцией, на этапе структурного синтеза возникает задача организации выбора нужной частичной функции при каждом переходе автомата. В С-автомате данная задача решается мультиплексированием одновременно формируемых значений двух частичных функций. На рис. 1 схема КС$_{Inc}$ вырабатывает сигнал *Inc*, управляющий мультиплексором МХ. Формируемый на выходе мультиплексора структурный код d состояния перехода отождествляется со значением структурной функции переходов автомата и поступает в регистр памяти РП. Отметим, что в структуре С-автомата, приведенной в [14], блокам КС$_{d_1}$, МХ и РП соответствует инкрементный счетчик СТ.

Увеличение числа переходов, реализуемых функцией d_1, приводит, в общем случае, к снижению аппаратурных затрат в схеме формирования переходов С-автомата по сравнению с канонической структурой МПА. В связи с этим С-автоматы рекомендуется использовать для линейных ГСА, в которых не менее 75% вершин являются операторными [14].

Не смотря на то, что каждая частичная функция реализует собственное подмножество переходов, одно и то же состояние автомата может встречаться в кортежах нескольких частичных функций. Например, переход в состояние a_k может выполняться с помощью частичной функции δ_1, переход из этого состояния – с помощью частичной функции δ_2. Если изоморфизмы (48) и (49) задавать независимо друг от друга, возможны ситуации, когда один и тот же структурный код присвоен двум различным состояниям, либо одному и тому же состоянию присвоено два различных структурных кода. Сказанное относится также и к структурным кодам входных сигналов. Подобные неоднозначности делают невозможным последующий синтез логической схемы автомата.

Решение данной проблемы заключается во *взаимосвязанном задании изоморфизмов* (48) и (49), при котором:

– каждое состояние $a_i \in A$ получает единственный и уникальный структурный код $K_S(a_i)$ в соответствующих носителях всех структурных подалгебр, в которых оно встречается;

– каждый входной сигнал $z_j \in Z$ получает единственный и уникальный структурный код $K_S(z_j)$ в соответствующих носителях всех структурных подалгебр, в которых он встречается.

В алгебраической форме данные требования могут быть выражены требованием существования изоморфизма (43). Данный изоморфизм устанавливает, в том числе, взаимно однозначные соответствия между каждым состоянием $a_i \in A$ и его структурным кодом $K_S(a_i) \in K_S(A)$, а также между каждым входным сигналом $z_j \in Z$ и его структурным кодом $K_S(z_j) \in K_S(Z)$. В результате за каждым состоянием и входным сигналом закрепляется единственный структурный код независимо от количества используемых подалгебр.

Применяемая в С-автомате операция инкремента является арифметической операцией и определена для скалярных величин. В процессе синтеза С-автомата на этапе формирования линейных последовательностей состояний каждое состояние $a_i \in A_{\delta_1}$ получает уникальный «скалярный» код $K_I(a_i)$ из множества натуральных чисел. «Скалярные» коды назначаются всем состояниям, переходы в которые или переходы из которых

выполняются с помощью операции инкремента. Таким образом, результатом кодирования всех состояний множества A_{δ_1} является множество скалярных кодов $K_I(A_{\delta_1})$. В то же время состояния, отсутствующие в кортежах функции δ_1, «скалярных» кодов не получают.

Обозначим операцию инкремента символом O_{Inc}. Заметим, что множество $K_I(A_{\delta_1})$ совместно с операцией O_{Inc} можно рассматривать как алгебру

$$G_I =< \mathcal{A}_I, \mathcal{F}_I >=< K_I(A_{\delta_1}), O_{Inc} > . \quad (50)$$

Здесь операция O_{Inc} формально представляется множеством кортежей вида $< K_I(a_i), K_I(a_j) > \in O_{Inc}$. Компонент $K_I(a_i) \in K_I(A_{\delta_1})$ есть «скалярный» код исходного состояния, компонент $K_I(a_j) \in K_I(A_{\delta_1})$ – «скалярный» код состояния перехода. Очевидно, что подобный кортеж не содержит информации о входном сигнале, под воздействием которого выполняется переход, и в таком виде не может быть поставлен в соответствие кортежам частичных функций δ_1 и d_1.

Введем в алгебру (50) дополнительный носитель Z_{δ_1}, идентичный одноименному носителю из абстрактной подалгебры G_{δ_1}. В результате получим двухосновную алгебру:

$$G_I =< \mathcal{A}_I, \mathcal{F}_I >=< \{K_I(A_{\delta_1}), Z_{\delta_1}\}, \{O_{Inc}\} > , \quad (51)$$

в которой операция O_{Inc} задается множеством кортежей вида:

$$< K_I(a_i), z_j, K_I(a_k) > \in O_{Inc}, \quad (52)$$

где $K_I(a_i), K_I(a_j) \in K_I(A_{\delta_1})$, $z_j \in Z_{\delta_1}$.

Теперь, учитывая возможность установления попарных взаимно однозначных соответствий между носителями A_{δ_1}, $K_I(A_{\delta_1})$ и $K_S(A_{d_1})$, а также между носителями Z_{δ_1} и $K_S(Z_{d_1})$, каждому кортежу вида (12) в функции δ_1 и соответствующему ему кортежу вида (39) в функции d_1 может быть поставлен в соответствие единственный кортеж вида (52). Это позволяет задать операцию O_{Inc} таким множеством кортежей (52), чтобы в рамках алгебры (51) она выполняла преобразования «скалярных» кодов состояний по тому же закону, что и функции δ_1 и d_1 в своих подалгебрах, реализуя, таким образом, те же автоматные переходы, что и функции δ_1 и d_1.

Все это позволяет говорить о взаимном изоморфизме трех алгебр: алгебры (51), абстрактной подалгебры G_{δ_1} и структурной подалгебры G_{d_1}:

$$G_{\delta_1} \leftrightarrow G_I \leftrightarrow G_{d_1} . \quad (53)$$

Данное выражение содержит три изоморфизма: $G_{\delta_1} \leftrightarrow G_{d_1}$, $G_{\delta_1} \leftrightarrow G_I$ и $G_I \leftrightarrow G_{d_1}$. Учитывая требование их одновременного существования, будем для краткости рассматривать данные изоморфизмы как единый изоморфизм (53).

Если структурные коды состояний, встречающиеся в кортежах частичной функции d_2, могут задаваться произвольно (из допустимого множества структурных кодов), то в случае функции d_1 структурные коды должны задаваться таким образом, чтобы сделать возможным их преобразование с помощью операции инкремента. Задание «скалярных» кодов состояний в естественном (инкрементном) порядке есть задание транзитивного замыкания для отношения инкремента на некотором множестве скалярных чисел. На произвольном множестве структурных кодов, для которых неизвестна их «скалярная» интерпретация, отношение инкремента не может быть определено. По этой причине выбор «скалярных» кодов состояний множества A_{δ_1} должен предшествовать выбору их структурных кодов. В свою очередь, задание структурных кодов состояний множества A_{δ_2} выполняется после задания структурных кодов состояний множества A_{δ_1}, а синтез схем КС_{d_2} и КС_{Inc} возможен лишь после задания структурных кодов всех состояний автомата.

5 РЕЗУЛЬТАТЫ

Применение принципа представления функции переходов автомата в виде множества частичных функций позволяет, в случае автомата на счетчике, сформировать следующий порядок задания алгебр:

1. Задание абстрактных подалгебр переходов (46) и (47) в соответствии со сформированными линейными последовательностями состояний.
2. Задание алгебры (51), изоморфной подалгебре (46).
3. Задание структурной подалгебры (44), изоморфной алгебре (51).
4. Задание структурной подалгебры (45), изоморфной подалгебре (44).

Учитывая то, что алгебра (51) играет в С-автомате вспомогательную роль при построении структурной алгебры переходов по заданной абстрактной алгебре переходов, назовем алгебру G_I *промежуточной алгеброй переходов*, понимая под нижним индексом «*I*» термин «*intermediate*» – «промежуточный». «Скалярные» коды состояний, образующие носитель промежуточной алгебры, назовем *промежуточными кодами состояний*. Операцию, образующую сигнатуру промежуточной алгебры, назовем *операцией переходов*.

В алгебраической форме эквивалентность абстрактного автомата и реализующего его структурного С-автомата выражается существованием следующих изоморфизмов:

1. Изоморфизм (53) промежуточной алгебры G_I с соответствующими ей абстрактной и структурной подалгебрами G_{δ_1} и G_{d_1}. Данный изоморфизм выражает реализацию части переходов автомата с помощью операции инкремента.
2. Изоморфизм (43) абстрактной и структурной алгебр переходов. Данный изоморфизм обеспечивается совместным заданием изоморфизмов (48) и (49), при котором сохраняются уникальность и однозначность структурных кодов состояний и входных сигналов независимо от способа разбиения функции переходов на частичные функции.

Если хотя бы один из данных изоморфизмов не может быть задан, то не может быть построен С-автомат, эквивалентный заданному абстрактному автомату. Таким образом, каждый из изоморфизмов является необходимым условием существования С-автомата, а достаточным условием является одновременное существование обоих изоморфизмов. Сведем изоморфизмы (53) и (43) в систему (54), которую будем рассматривать в качестве *математической модели автомата на счетчике*:

$$\begin{cases} G_{\delta_1} \leftrightarrow G_I \leftrightarrow G_{d_1}; \\ G_{\delta} \leftrightarrow G_d. \end{cases} \quad (54)$$

Процесс построения системы (54) включает выполнение следующих этапов структурного синтеза С-автомата: формирование линейных последовательностей состояний, кодирование состояний, формирование таблицы переходов [14].

После задания системы (54) следует завершающий этап структурного синтеза – реализация в заданном элементном базисе логической схемы автомата (комбинационных схем КС_{d_2}, КС_{Inc} и схемы формирования микроопераций) в соответствии со сформированными множеством структурных кодов состояний и разбиением функции переходов на частичные функции. Таким образом, метод структурного синтеза С-автомата может быть сокращенно представлен двумя этапами:

1. Задание системы изоморфизмов (54).
2. Реализация логической схемы автомата в заданном элементном базисе.

6 ОБСУЖДЕНИЕ

Представление функции переходов С-автомата в виде двух частичных функций наряду с использованием промежуточной алгебры переходов отражает совмещение в одной структуре двух различных способов реализации переходов автомата – канонического и инкрементного. Разбиение функции переходов на частичные функции в случае канонической структуры МПА не имеет смысла, поскольку в этом случае каждая частичная функция будет лишь реализовывать некоторый фрагмент единой системы канонических уравнений, а требуемая при этом реализация выбора частичной функции путем мультиплексирования приведет к увеличению аппаратурных затрат в схеме устройства.

Приведенный пример использования промежуточной алгебры на базе операции инкремента позволяет предположить возможность формирования сигнатуры промежуточной алгебры из других операций, например, операции декремента. Подобный подход может способствовать снижению аппаратурных затрат в схеме проектируемого микропрограммного автомата за счет использования в схеме формирования переходов таких комбинационных схем, аппаратурные затраты в которых не зависят или зависят незначительно от количества реализуемых автоматных переходов.

ВЫВОДЫ

В настоящей работе предложены новые математические модели абстрактного и структурного автоматов, в основе которых лежит представление функции переходов в виде множества частичных функций. Результатом исследований является представление процессов преобразования кодов состояний в структурном автомате в виде промежуточной алгебры переходов, описывающей закон преобразования кодов состояний для некоторого подмножества автоматных переходов.

Использование различных операций переходов позволяет выделять на функции переходов структурного автомата различные законы преобразования кодов состояний, что может способствовать снижению аппаратурных затрат в логической схеме автомата в сравнении с другими известными структурами МПА. Вопрос целесообразности применения предложенного подхода для оптимизации аппаратурных затрат в логической схеме микропрограммного автомата не является очевидным из предложенных моделей и требует самостоятельного исследования.

СПИСОК ЛИТЕРАТУРЫ

1. Баранов С. И. Синтез микропрограммных автоматов / С. И. Баранов. – Л. : Энергия, 1979. – 232 с.
2. Алгебраическая теория автоматов, языков и полугрупп / под ред. М. Арбиба ; пер. с англ. – М. : Статистика, 1975. – 335 с.
3. Плоткин Б. И. Элементы алгебраической теории автоматов: учеб. пособие для вузов / Б. И. Плоткин, Л. Я. Гринглаз, А. А. Гварамия. – М. : Высшая школа, 1994. – 191 с.
4. Мальцев А. И. Алгебраические системы / А. И. Мальцев. – М. : Наука, 1970. – 392 с.
5. Плоткин Б. И. Универсальная алгебра, алгебраическая логика и базы данных / Б. И. Плоткин. – М. : Наука, гл. ред. физ.-мат. лит., 1991. – 448 с.
6. Судоплатов С. В. Элементы дискретной математики : учебник / С. В. Судоплатов, Е. В. Овчинникова. – М. : ИНФРА-М, Новосибирск : Изд-во НГТУ, 2002. – 280 с.
7. Богомолов А. М. Алгебраические основы теории дискретных систем / А. М. Богомолов, В. Н. Салий. – М. : Наука-Физматлит, 1997. – 368 с.
8. Новиков Ф. А. Дискретная математика для программистов / Ф. А. Новиков. – СПб. : Питер, 2000. – 304 с.
9. Кудрявцев В. Б. Введение в теорию автоматов / В. Б. Кудрявцев, С. В. Алешин, А. С. Подколзин. – М. : Наука, 1985. – 320 с.
10. Глушков В. М. Синтез цифровых автоматов / В. М. Глушков. – М. : Физматгиз, 1962. – 476 с.
11. Глушков В. М. Абстрактная теория автоматов / В. М. Глушков // Успехи математических наук. – 1961. – Т. XVI, Вып. 5. – С. 3–62.
12. Трахтенброт Б.А. Конечные автоматы (поведение и синтез) / Б. А. Трахтенброт, Я.М. Бардзинь. – М. : Наука, 1970. – 400 с.
13. Шиханович Ю. А. Введение в современную математику (начальные понятия) / Ю. А. Шиханович. – М. : Наука, 1965. – 376 с.
14. Баркалов А. А. Синтез устройств управления на программируемых логических устройствах / А. А. Баркалов. – Донецк : ДонНТУ, 2002. – 262 с.

Статья поступила в редакцию 24.09.2015.
После доработки 05.10.2015.

Бабаков Р. М.

Канд. техн. наук, доцент, доцент кафедри комп'ютерних наук Донецького національного технічного університету, м. Красноармійськ, Україна

ПРОМІЖНА АЛГЕБРА ПЕРЕХОДІВ В МІКРОПРОГРАМНОМУ АВТОМАТІ

Вирішено задачу формалізації завдання мікропрограмного автомата, в структурі якого частина автоматних переходів реалізується неканонічним шляхом. Запропоновано новий пілхід до організації функції переходів мікропрограмного автомата, відповідно до якого функція переходів представляється у вигляді сімейства часткових функцій, кожна з яких визначена лише на частині області визначення функції переходів автомата і відповідає певній підмножині автоматних переходів.

З урахуванням запропонованого підходу традиційне представлення автомата у вигляді багатоосновної алгебри матиме певні зміни. По-перше, взаємна незалежність функцій переходів і виходів, що утворюють сигнатуру алгебри, дозволяє розглядати їх окремо одна від іншої, що призводить до представлення автомата у вигляді двох алгебр: алгебри переходів, сигнатура якої містить лише функцію переходів, і алгебри виходів, сигнатура якої містить лише функцію виходів. По-друге, представлення функції переходів у вигляді множини часткових функцій виводить до заміни алгебри переходів множиною підалгебр переходів, в кожній з яких сигнатура утворена частковою функцією переходів.

На прикладі мікропрограмного автомата з лічильником показано, що закон перетворення кодів станів в рамках певної підмножини переходів може бути заданий алгебраїчною функцією (операцією переходів), що використовує скалярну інтерпретацію кодів станів структурного автомата. Скалярну інтерпретацію кодів станів разом з операцією переходів пропонується представляти у вигляді так званої проміжної алгебри переходів, яка ізоморфна відповідним підалгебрам переходів абстрактного і еквівалентного йому структурного автоматів.

Ключові слова: мікропрограмний автомат, часткова функція переходів, проміжна алгебра переходів, автомат із лічильником.

Babakov R. M.

PhD, Associate Professor of the Department of Computer Sciences of the Donetsk National Technical University, Krasnoarmiysk, Ukraine

INTERMEDIATE ALGEBRA OF TRANSITIONS IN MICROPROGRAM FINAL-STATE MACHINE

The problem of formalization of representation of final-state machine, where the part of automaton transition is realized in non-canonical way, is solved. A new approach for organization of the function of transitions of the final-state machine is proposed. According to it the function of transitions is represented as a family of partial functions, each of which is defined only on the part of the domain of the function of transitions, and corresponds to a subset of the automaton transitions.

According to the proposed approach the traditional representation of the final-state machine as a polybasic algebra is changed. First, the mutual independence of the function of transitions and function of outputs that form the signature of algebra, allows us to consider them separately from each other. This way leads to presentation of the final-state machine as two algebras: the algebra of transitions whose signature contains only the function of transitions and algebra of outputs whose signature contains only the function of outputs. Second, the representation of the function of transitions in the form of a set of partial functions leads to the replacement of the algebra of transitions by the set of subalgebras of transitions, a signature of each of which is formed by partial function of transitions.

The example of the final-state machine with a counter shows that the law of transformation of codes of states within a certain subset of transitions can be set by an algebraic function (operation of transitions) using scalar interpretation of codes of states of the structural final-state machine. The representation of scalar interpretation of codes of states and the operation of transitions as so-called intermediate algebra of transitions isomorphic to both according subalgebras of transitions of the abstract and equivalent structural final-state machines is proposed.

Keywords: microprogram final-state machine, partial function of transitions, intermediate algebra of transitions, final-state machine with counter.

REFERENCES

1. Baranov S. I. Sintez mikroprogramnih avtomatov. Leningrad, Energiya, 1979, 232 p.
2. Algebraicheskaya teoriya avtomatov, yazikov i polugrupp. Pod red. M. Arbiba. Per. s angl. Moscow, Statistika, 1975, 335 p.
3. Plotkin B. I., Gringlaz L. Y., Gvaramiya A. A. Elementi algebraicheskoy teorii avtomatov: ucheb. posobie dlya vuzov. Moscow, Vish. shk., 1994, 191 p.
4. Maltsev A. I. Algebraicheskie sistemi. Moscow, Nauka, 1970, 392 p.
5. Plotkin B. I. Universalnaya algebra, algebraicheskaya logika i bazi dannih. Moscow, Nauka. Gl. red. fiz.-mat. lit., 1991, 448 p.
6. Sudoplatov S. V., Ovchinnikova E. V. Elementi diskretnoy matematiki: uchebnik. Moscow, INFRA-M, Novosibirsk, Izdatelstvo NGTU, 2002, 280 p.
7. Bogomolov A. M., Saliy V. N. Algebraicheskie osnovi teorii diskretnih sistem. Moscow, Nauka, Fizmatlit, 1997, 368 p.
8. Novikov F. A. Diskretnaya matematika dlya programmistov. SPb, Piter, 2000, 304 p.
9. Kudryavtsev V. B., Alyoshin S. V., Podkolzin A. S. Vvedenie v teoriju avtomatov. Moscow, Nauka, 1985, 320 p.
10. Glushkov V. M. Sintez tsifrovih avtomatov. Moscow, Fizmatgiz, 1962, 476 p.
11. Glushkov V. M. Abstraktnaya teoriya avtomatov, *Uspehi matematicheskih nauk,* 1961, Vol. XVI, Issue 5, pp. 3–62.
12. Trahtenbrot B. A., Bardzin Y. M. Konechnie avtomati (povedenie i sintez). Moscow, Nauka, 1970, 400 p.
13. Shihanovich U. A. Vvedenie v sovremennuju matematiku (nachalnie ponyatiya). Moscow, Nauka, 1965, 376 p.
14. Barkalov A. A. Sintez ustroystv upravleniya na programmiruemih logicheskih ustroystvah. Donetsk, DonNTU, 2002, 262 p.

УДК 004.9

Бісікало О. В.[1], Висоцька В. А.[2]

[1]Д-р техн. наук, професор, декан факультету комп'ютерних систем і автоматики Вінницького національного технічного університету, Вінниця, Україна

[2]Канд. техн. наук, доцент кафедри «Інформаційні системи та мережі» Національного університету «Львівська політехніка», Львів, Україна

ВИЯВЛЕННЯ КЛЮЧОВИХ СЛІВ НА ОСНОВІ МЕТОДУ КОНТЕНТ-МОНІТОРИНГУ УКРАЇНОМОВНИХ ТЕКСТІВ

Вирішено завдання розробки алгоритмічного забезпечення процесів контент-моніторінгу для розв'язання задачі визначення ключових слів україномовного тексту. Розглянуто формальне обгрунтування методу контент-моніторінгу тексту за допомогою стеммера Портера, в основу модифікації стемінгу покладено відомі результати класифікації морфемної і словотвірної структури дериватів української мови, виявлення закономірностей комбінаторики афіксів, моделювання структурної організації дієслів і суфіксальних іменників, а також морфонологічних модифікацій у процесі словозміни дієслова та словозміні і словотворенні прикметників української мови. Проведено декомпозицію методу та розроблено алгоритмічне забезпечення його основних структурних складових за результатами контент-аналізу тексту. Теоретично виявлено способи покращення показників ефективності пошуку ключових слів, зокрема щільності ключовиків у тексті. На основі розробленого програмного забезпечення отримано результати експериментальної апробації запропонованого методу контент-моніторінгу для визначення ключових слів в наукових текстах технічного профілю. Виявлено, що для обраної експериментальної бази зі 100 робіт найкращих результатів за критерієм щільності досягає метод аналізу статті без початкової обов'язкової інформації і без списку літератури, але із перевіркою уточнених заблокованих слів та уточненого тематичного словника.

Ключові слова: текст, україномовний, алгоритм, контент-моніторінг, ключові слова, контент-аналіз, стеммер Портера, лінгвістичний аналіз, синтаксичний аналіз.

НОМЕНКЛАТУРА

ІТ – інформаційні технології;

СЕКК – система електронної контент-комерції;

е-бізнес – електронний бізнес;

Е-комерція – електронна комерція;

ПЗ – програмне забезпечення;

$X = \{x_1, x_2 \ldots, x_{n_X}\}$ – множина вхідних даних $x_i \in X$ з різних інформаційних ресурсів або від модераторів при $i = \overline{1, n_X}$;

$C = \{c_1, c_2 \ldots, c_{n_C}\}$ – множина комерційного контенту $c_r \in C$ при $r = \overline{1, n_C}$;

C_0 – сформований комерційний контент;

C_1 – відфільтрований комерційний контент;

C_2 – відформатований комерційний контент;

C_3 – комерційний контент з визначеною множиною ключових слів;

$< U_C, U_G, U_K >$ – набір критеріїв для текстового контенту X;

$U_C = \{U_{C1}, U_{C2}, \ldots, U_{Cn_C}\}$ – множина критеріїв створення комерційного контенту;

$U_G = \{U_{G1}, U_{G2}, \ldots, U_{Gn_G}\}$ – множина критеріїв збирання комерційного контенту (фільтри);

$U_K = \{U_{K1}, U_{K2}, U_{K3}, U_{K4}\}$ – множина критеріїв визначення ключових слів в контенті;

U_{K1} – унікальність термів – іменників, словосполучень іменників або прикметника з іменником серед множини слів контенту;

U_{K2} – частота появи ключових слів комерційного контенту;

U_{K3} – кількість знаків без пробілів для $Noun \in U_{K1}$ при $Unicity \geq 80$;

U_{K4} – критерій формування множини ключових слів;

$T = \{t_1, t_2 \ldots, t_{n_T}\}$ – час $t_p \in T$ транзакції формування контенту при $p = \overline{1, n_T}$;

$\alpha_0 : (X, U_C, T) \to C_0$ – оператор створення контенту – відображення даних з різних джерел у контент, який відрізняється актуальністю;

$\alpha_1 : (X, U_G, T) \to C_0$ – оператор збирання контенту – відображення даних від авторів у контент, який відрізняється достовірністю та актуальністю;

$\alpha_2 : (C_0, T, U_B) \to C_1$ – оператор виявлення дублювання контенту – відображення контенту в новий стан, який відрізняється унікальністю;

$\alpha_3 : (C_1, U_{FR}, T) \to C_2$ – оператор форматування контенту – відображення контенту в новий стан, який відмінний від попереднього форматом подання;

$\alpha_4 : (C_2, U_K, T) \to C_3$ – оператор виявлення ключових слів контенту – відображення контенту в новий стан, який відрізняється наявністю множини ключових слів, що загально описують його зміст.

ВСТУП

Активний розвиток мережі Інтернет сприяє зростанню потреб в отриманні оперативних даних виробничого/стратегічного характеру і реалізації нових форм інформаційного обслуговування через сучасні ІТ е-бізнесу [1–3]. Документована інформація, підготовлена відповідно до потреб користувачів, є інформаційним продуктом або комерційним контентом, наприклад, електронний матеріал Інтернет-видавництва, маркетингові дослідження,

DOI 10.15588/1607-3274-2016-1-9

консалтингові послуги тощо. Дії для забезпечення користувачів комерційним контентом є інформаційною послугою. Інтернет-ринок є сукупністю економічних, правових, організаційних і програмних відносин з продажу інформаційних продуктів/послуг між виробниками, постачальниками та користувачами [1–3].

Комерційний контент визначають як:

– вміст інформаційних ресурсів в СЕКК;

– об'єкт бізнес-процесів в СЕКК, наприклад, стаття, ПЗ, книга тощо;

– структурована та логічно завершена множина даних, що є об'єктом взаємовідносин між користувачем та СЕКК;

– набір електронних даних без наперед визначеної структури;

– дані комерційного призначення, що неподільні в часі;

– основний чинник формування області діяльності, функціонування та призначення СЕКК.

Сьогодні е-комерція є об'єктивною реальністю та перспективним бізнес-процесом. Інтернет є бізнес-середовищем, а комерційний контент є товаром з найбільшим попитом у ньому та основним об'єктом процесів електронної контент-комерції. Комерційний контент можна зразу замовити, оформити, оплатити і отримати on-line як товар. Через Інтернет продають весь спектр комерційного контенту – наукові та публіцистичні статті, музика, книги, фільми, фото, ПЗ тощо. Відомими корпораціями, які реалізують електронну контент-комерцію, є Google через Play Market, Apple – Apple Store, Amazon – Amazon.com [1].

Більшість рішень та досліджень зроблено на рівні реальних прикладних проектів, а сучасні СЕКК побудовані за закритим принципом як разові проекти та орієнтовані на реалізацію комерційного контенту, створеного за їх межами. Тому для проектування, створення, впровадження та супроводу СЕКК потребують розробки загальні методи та інформаційні технології формування, управління та супроводу комерційного контенту. З огляду на важливість для функціонування СЕКК ключових слів об'єктом дослідження обрано процес виявлення ключових слів в україномовних текстах у режимі реального часу, предмет дослідження – методи та моделі контент-моніторінгу таких текстів.

1 ПОСТАНОВКА ЗАДАЧІ

Нехай україномовний текстовий контент X з різних джерел інформації у вигляді $X = \{x_1, x_2 \ldots, x_{n_X}\}$ має стати основою відфільтрованого контенту C_1, відформатованого контенту C_2 та його модифікації C_3 з визначеною множиною ключових слів $KeyWords \in U_{K4}$. За відомими критеріями $< U_C, U_G, U_K >$ потрібно визначити оператор виявлення ключових слів комерційного контенту $\alpha_4 : (C_2, U_K, T) \to C_3$ та експериментально перевірити параметр частоти появи ключових слів комерційного контенту U_{K2} за різними режимами роботи алгоритмічного забезпечення запропонованого методу контент-моніторінгу.

2 ОГЛЯД ЛІТЕРАТУРИ

Сталою сучасною тенденцією можна вважати постійний ріст темпів виробництва текстового контенту в Інтернет-просторі. Цей процес є об'єктивним і позитивним, але виникла проблема – прогрес у галузі виробництва текстового контенту призводить до пониження загального рівня інформованості потенційного користувача Інтернет-простору [1–3]. Крім збільшення обсягів текстового контенту до масштабів, яке унеможливлює його безпосереднє опрацювання та помітно гальмує його поширення виникає низка специфічних проблем (табл. 1).

Негативні чинники у формуванні текстового контенту ускладнюють процес пошуку необхідних даних при скануванні різних джерел інформації. Збільшення фізичного обсягу та змінність співвідношення актуальності/динаміки контентних потоків (наслідок систематичного або нерегулярного оновлення) призводить до виникнення дублювання, інформаційного шуму та надмірності результатів пошуку контенту. Охоплення та узагальнення великих динамічних потоків контенту, які неперервно генерують в Інтернет-джерелах, вимагає якісно нових методів/підходів пошуку – таких як контент-моніторинг (рис. 1) на основі аналізу ключових слів [1–32]. Вхідною інформацією для контент-моніторингу є текст на природній мові як послідовність символів, вихідна інформація – це таблиці розділів, речень і лексем аналізованого тексту. Контент-моніторинг є програмним засобом автоматизації знаходження найбільш важливих складових в потоках контенту за допомогою алгоритмів стемінгу [1–32]. Це змістовний аналіз потоків контенту з метою постійного отримання необхідних якісних/кількісних зрізів на протязі наперед не визначеного проміжку часу.

Мета роботи полягає у створенні алгоритмічного забезпечення методу контент-моніторингу україномовних текстів на основі стеммера Портера та його застосуванні для виявлення значущих ключових слів. Для досягнення

Таблиця 1 – Основні негативні чинники у формуванні текстового контенту

Назва	Основна причина	Рішення
Інформаційний шум	Структурованість масивів контенту.	Фільтри, контент-моніторинг, аналіз сайту, контент-аналіз.
Паразитичний контент	Поява в якості додатків.	Фільтри, контент-моніторинг, контент-аналіз.
Нерелевантність контенту	Невідповідність потребам користувачів.	Створення анотованої бази даних, пошукових образів первинного контенту та їх кластеризація, контент-аналіз.
Дублювання контенту	Дублювання в джерелах.	Контент-аналіз, сканери і фільтри на базі статистики та критеріїв.
Навігація в потоці контенту	Швидкий ріст обсягу і поширення контенту.	Аналіз сайту, фільтри, контент-моніторинг, контент-аналіз.
Надмірність пошуку	Дублювання і нерелевантність.	Анотований пошук, контент-аналіз та реферування.

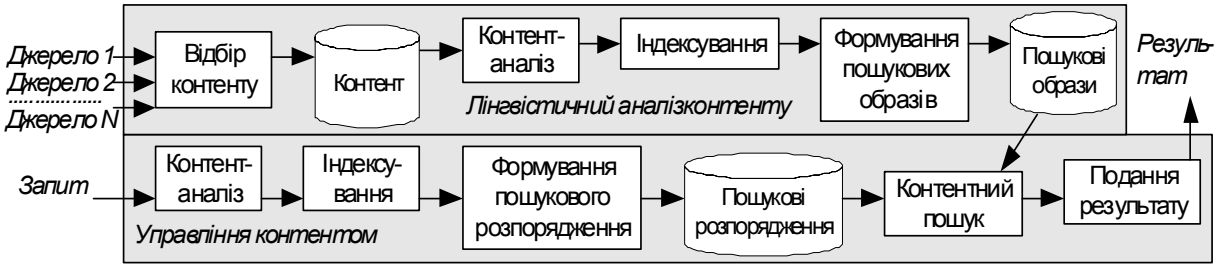


Рисунок 1 – Структурна схема процесу контент-моніторингу текстових масивів даних

мети пропонується розв'язати такі задачі дослідження: провести контент-аналіз текстової інформації; забезпечити визначення множини ключових слів; провести лінгвістичний аналіз текстового контенту; розробити синтаксичний аналізатор текстового контенту.

3 МАТЕРІАЛИ І МЕТОДИ

Головною складовою контент-моніторингу є контентний пошук та контент-аналіз тексту. Контент-аналіз призначений для пошуку контенту в масиві даних за змістовими лінгвістичними одиницями (алг. 1). Одиниця рахунку є кількісною мірою одиниці аналізу, що дозволяє реєструвати частоту (регулярність) появи ознаки категорії аналізу в тексті (кількість певних слів або їх поєднань, рядків, друкованих знаків, сторінок, абзаців, авторських аркушів, площа тексту тощо).

Алгоритм 1. Контент-аналіз текстового контенту.

Етап 1. Визначення набору критеріїв $< U_C, U_G >$ для текстового контенту X.

Крок 1. Формування набору критеріїв як тип джерела (форум, електронна пошта, Інтернет-газета, чат, Інтернет-журнал); тип контенту (стаття, е-лист, банер, коментарій); учасники комунікації (відправник, одержувач, реципієнт).

Крок 2. Визначення розміру (мінімальний обсяг або довжина), частоти появи, способу/місця розповсюдження та час появи контенту.

Крок 3. Фільтрування згідно сформованого набору критеріїв контентного потоку та зберігання ідентифікованого релевантного контенту X.

Етап 2. Контент-аналітичний відбір. Формування вибіркової сукупності контенту X' за критеріями обмеженої вибірки $< U_C, U_G >$ з більшого масиву X.

Етап 3. Виявлення змістовних одиниць аналізу $< U'_C, U'_G >$ текстового комерційного контенту X' (словосполучення, речення, тема, ідея, автор, персонаж, соціальна ситуація, частина тексту, кластеризована за змістом категорії аналізу) за модифікованим алгоритмом Потера. Вимоги до вибору лінгвістичної одиниці аналізу: велика для інтерпретації значення; мала, щоб не інтерпретувати багато значень; легко ідентифікується; кількість одиниць велика для проведення вибірки.

Етап 4. Виділення одиниць рахунку аналізу текстового контенту X'.

Крок 1. Якщо одиниці рахунку $< U_C, U_G >$ збігаються з одиницями аналізу $< U'_C, U'_G >$, то знаходять частоти появи виділеної змістовної одиниці, інакше перейти до кроку 2.

Крок 2. Модератор на основі аналізованого контенту висуває та доповнює одиниці рахунку $< U_C, U_G >$, наприклад, протяжність текстів; площа тексту, заповнена змістовними одиницями; кількість рядків (абзаців, знаків, колонок тексту); розмір/вид файлу; кількість рисунків з певним змістом/сюжетом тощо.

Етап 5. Порівняння змістовних одиниць аналізу $< U'_C, U'_G >$ з одиницями $< U_C, U_G >$.

Крок 1. Класифікація за угрупованнями із оціненням ваги змістовних категорій в загальному обсязі тексту. Класифікатором є загальна таблиця, в яку зведені всі категорії аналізу і одиниці аналізу. Фіксують одиниці виразу категорій.

Крок 2. Статистичні розрахунки зрозумілості та атрактивності контенту.

Етап 6. Розроблення інструменту контент-аналізу.

Крок 1. Створення закодованого протоколу контенту X' для компактності подання даних та швидкого порівняння результатів аналізу різного контенту.

Крок 2. Заповнення протоколу контенту X' властивостями (автор, час видання, обсяг тощо).

Крок 3. Заповнення протоколу контенту X' підсумками його аналізу (кількість вживання в ньому певних одиниць аналізу і висновки щодо категорій аналізу). Протокол кожного контенту X' заповнюється на основі підрахунку даних всіх його реєстраційних карток.

Етап 7. Розроблення таблиці контент-аналізу. Тип таблиці визначають у вигляді системи скоординованих і субординованих категорій аналізу: кожна категорія (питання) передбачає ряд ознак (відповідей), за якими квантифікується зміст тексту X'.

Етап 8. Розроблення кодувальної матриці контент-аналізу.

Крок 1. Якщо обсяг вибірки і 100 одиниць, то аналізується набір матричних листів, інакше виконати крок 2.

Крок 2. Якщо вибірка < 100 одиниць, то проводиться двовимірний аналіз. В цьому випадку для кожного контенту X' формується кодувальна матриця.

Етап 9. Проведення аналізу тексту X' згідно створених кодувальних матриць.

Етап 10. Інтерпретація результатів $\alpha_0 : (X, U_C, T) \to C_0$ та $\alpha_1 : (X, U_G, T) \to C_0$. Виявляють і оцінюють характеристики контенту X' на основі статистичного набору підрахованих коефіцієнтів за певний період часу на визначену категорію. Охоплює всі здобуті фрагменти тексту C_0, висновки спираються не на частину результатів, а враховуються всі без винятку. Фільтрування $\alpha_3 : (C_1, U_{FR}, T) \to C_2$ та форматування $\alpha_3 : (C_1, U_{FR}, T) \to C_2$ комерційного контенту.

Застосування контент-аналізу при моніторингу Інтернет-джерел даних дозволяє автоматизувати процес знаходження найбільш важливих складових в потоках контенту при відборі даних з цих джерел. Це усуває дублювання контенту, інформаційний шум, паразитичний контент, надмірність результатів пошуку тощо. Даний метод застосовують в подальших етапах формування контенту для отримання більш точного релевантного результату – створення унікального комерційного контенту, який користується попитом серед користувачів СЕКК.

З метою реалізації контент-аналізу текстових масивів даних для формування множини ключових слів було розроблено програму на основі стеммера Потера, адаптованого до української мови (алг. 2), а також таблиці основ основних тематичних слів для подальшої рубрикації, текстів, що досліджуються (табл. 2). Блок-схему алгоритму наведено на рис. 2.

Таблиця 2 – Основні складові програми формування ключових слів

№	Назва	Пояснення
1	Filter	список потенційних ключовиків з аналізованого тексту з розрахунком відносної частоти їх появи в тексті
2	Input	вхідний текст для аналізу та визначення ключовиків
3	Format	тестувальний відформатований вхідний текст
4	Parser	парсер, адаптований до української
5	Stemer	правила стеммера Потера, адаптований до української
6	Object	класи об’єктів для стеммера Потера
7	Resvoc	список ключовиків (частоти вживання яких в тексті попали у визначений діапазон згідно алг. 2 на рис. 2 та відповідають тематичному словнику Thematic.txt)
8	Thematic	тематичний словник (формується модератором)
9	Vocab	список слів з аналізованого тексту з розрахунком абсолютної частоти їх появи в тексті

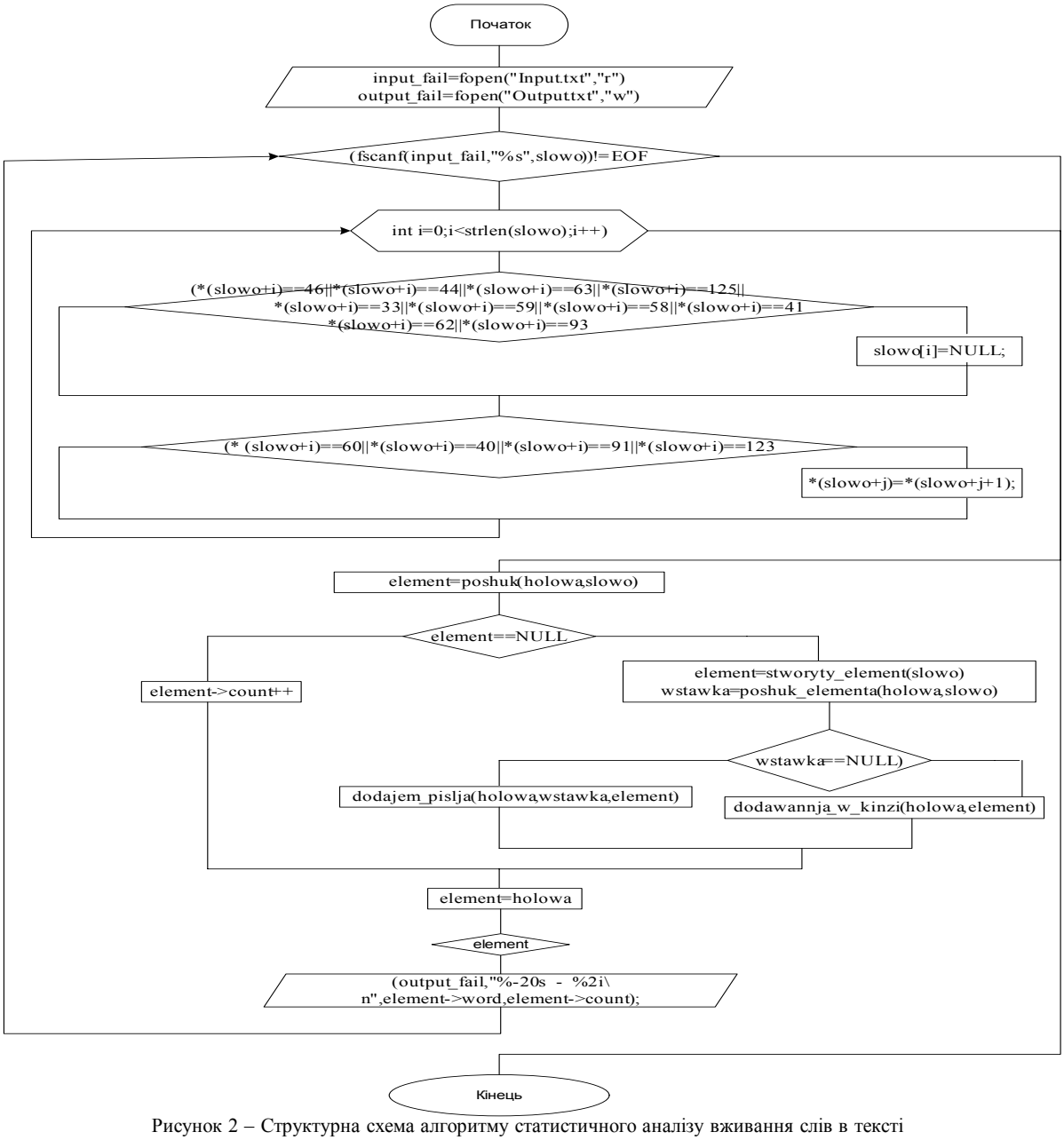


Рисунок 2 – Структурна схема алгоритму статистичного аналізу вживання слів в тексті

Алгоритм 2. Визначення множини ключових слів

Етап 1. В *Input* зберегти текст, який необхідно дослідити.

Етап 2. Відформатувати вхідний текст (однакові апострофи, забрати зайві символи, які не входять в абетку, окрім службових як пробіл, апостроф). В *Format* зберегти текст, який відформатовано.

Етап 3. При необхідності редагувати тематичний словник *Thematic*.

Етап 4. Запустити процес визначення множини ключових слів.

Крок 1. Будуємо спочатку алфавітно-частотний словник (абсолютні частоти) – *Vocab*.

Крок 2. Будуємо потім з *Vocab* алфавітно-частотний словник (відносні частоти) слів, тобто список слів за алфавітом та їх відносні частоти відносно загального обсягу тексту.

Крок 3. Будуємо скорочений список слів, частоти яких відповідають умовам формування ключовиків $U_K = \{U_{K1}, U_{K2}, U_{K3}, U_{K4}\}$, тобто список потенційних ключовиків – *Filter*.

Крок 4. Звіряємо сформований скорочений список з *Filter* зі списком *Thematic* та відповідно формуємо новий список входжень потенційних ключовиків з *Filter* в *Thematic* – список ключовиків в *Resvoc*.

Цей алгоритм не враховує пошук ключовиків по основах, у зв'язку з цим – результати дослідження текстів на формування ключових слів були негативні, зокрема:

– відсутні взагалі ключові слова у вихідному файлі (знайдені слова не відповідали вимогам до ключових слів – не попадали в діапазон частоти вживання в тексті);

– в список ключових слів попали службові слова, дієприкметники, дієслова, як ніяк не можуть бути ключовими словами (некоректно прописана база правил заблокованих слів);

– були присутні декілька слів з одною основою, але з різними флексіями (некоректно прописана база правил визначення основ, наприклад, пошук – пошуковими, користувач – користувачам, рейтинг – високорейтингового, рейтингу, контент – контентного, інформація – інформаційний, або були присутні граматичні помилки).

Тому був розроблений інший алгоритм знаходження множини ключових слів з врахуванням основ тематичних слів (рис. 3), та розміщений у відкритому доступі за адресою http://victana.lviv.ua/index.php/kliuchovi-slova.

4 ЕКСПЕРИМЕНТИ

Лінгвістичною базою для експериментального дослідження обрано 100 наукових публікацій Вісника Національного університету «Львівська політехніка» серії «Інформаційні системи та мережі» (http://science.lp.edu.ua/sisn), двох номерів 783 (http://science.lp.edu.ua/SISN/SISN-2014) та 805 (http://science.lp.edu.ua/sisn/vol-cur-805-2014-2). Аналіз статистики функціонування системи виявлення множини ключових слів із 100 наукових статей було проведено у два етапи, зокрема:

1. Проаналізувати всі статті із перевіркою загальних заблокованих слів та тематичного словника.

2. Проаналізувати всі статті із перевіркою уточнених заблокованих слів та уточненого тематичного словника (з більшою кількістю запуску системи формується множина невідомих слів (відсутніх і в тематичному словнику і в множині заблокованих).

Окрім того на кожному етапі перевірка відбувалась в два кроки для кожної статті: аналіз всієї статті (рис. 4а) та аналіз статті без початку (назва, автори, удк, анотації двома мовами, авторські ключові слова двома мовами, місце роботи авторів) і без списку літератури (рис. 4б) для того, щоб визначити похибки точності формування множини ключових слів.

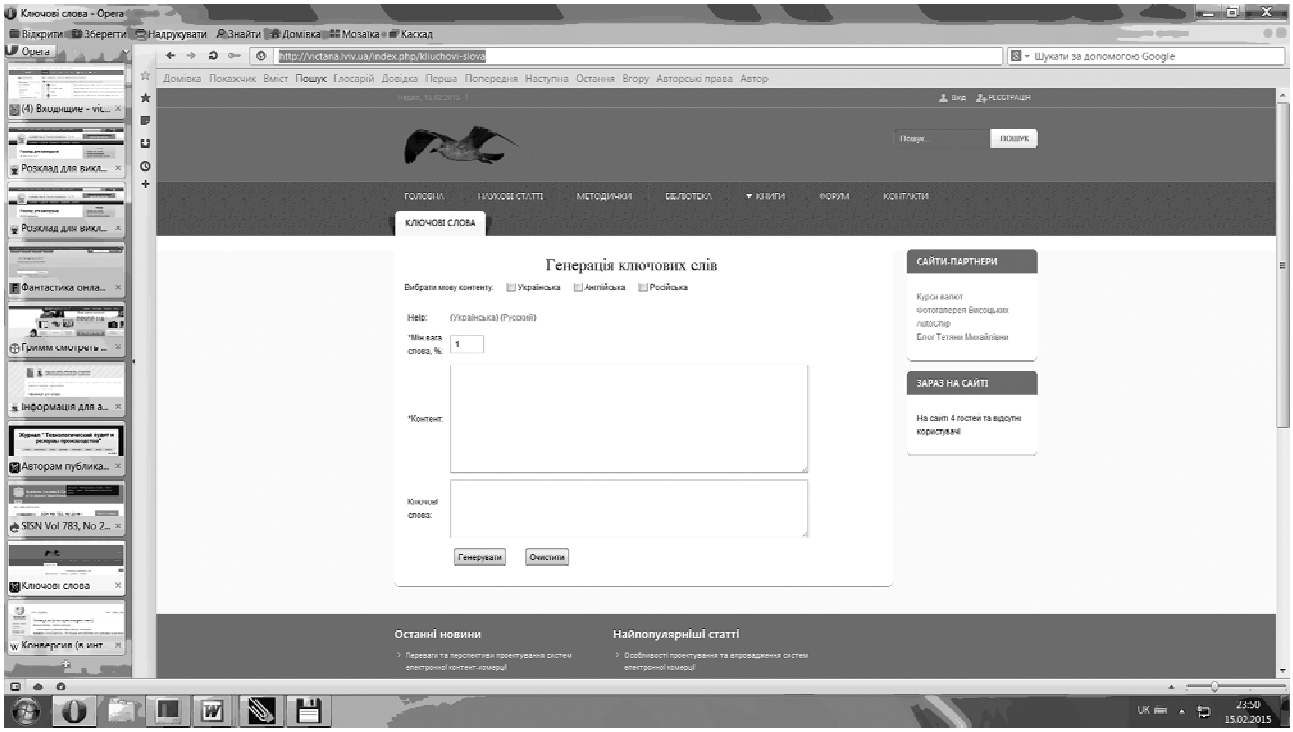
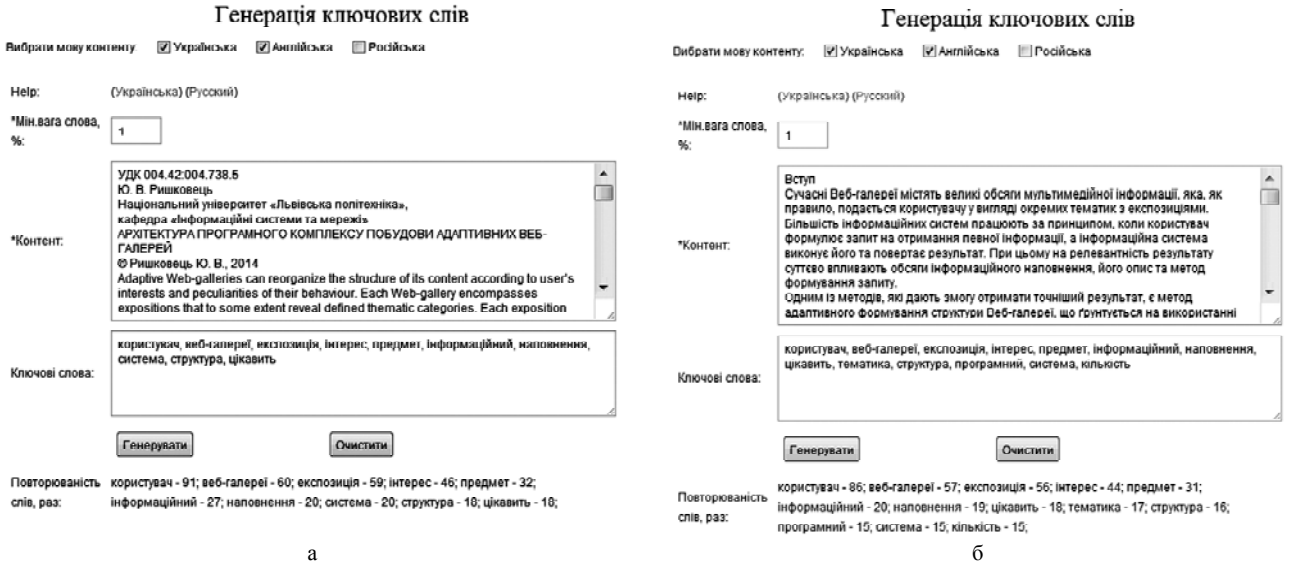


Рисунок 3 – Інформаційний ресурс визначення ключових слів з тексту



а б

Рисунок 4 – Результати перевірки статті: а – приклад аналізу всієї статті, б – приклад аналізу статті без початку і без списку літератури

5 РЕЗУЛЬТАТИ

Аналіз статистики здійснювався за принципом порівняння множини авторських ключових слів (визначені та прописані в статті самими авторами цих робіт), множини ключових слів визначених за першим та другим етапами з різними вагами слів (але більше, за визначене в опції **Мін.вага слова, %* в межах [1,5]) з повними та скороченими текстами робіт (табл. 3) при середньому арифметичному значенні авторських ключових словосполучень / слів біля 5 (4,77), які в середньому утворені з 10 (9,82) слів. Вага слова розраховується як відносна частота появи основи цього слова у всьому тексті. В табл. 4 присутні такі позначення, як *A* (всього ключових слів, визначених системою при заданій вазі слова), *B* (змістовних слів зі списку утворених, тобто без невідомих абревіатур, дієслів, службових слів тощо), *C* (збіг слів з визначеними автором статті), *D* (точність збігу знайдених ключовиків з авторським ключовими словами), *E* (додаткові ключові слова, визначені системою, але не визначені автором статті).

Таблиця 3 – Статистичні дані досліджених обсягів текстів статей

Назва обсягу статті	Крок 1		Крок 2	
	Всього	Середнє арифметичне	Всього	Середнє арифметичне
Сторінок	956	9,56	828	8,28
Абзаців	16497	164,97	15263	152,63
Рядків	42553	425,53	36965	369,65
Слів	345580	3455,8	291247	2912,47
Знаків	2327209	23272,09	1974773	19747,73
Знаків та пробілів	2674889	26748,89	2265917	22659,17

6 ОБГОВОРЕННЯ

На рис. 5 наведена порівняльна діаграма відсотків вживання знайдених системою ключових слів в відфільтрованому тексті (без початку (назва, автори, удк, анотації двома мовами, авторські ключові слова двома мовами, місце роботи авторів) і без списку літератури) Per_f та первинному авторському тексті Per_0 без уточнення модератором тематичного словника через поповнення заблокованих слів.

Отримані середні значення для 100 текстів $\overline{Per_f} = 0{,}28$ та $\overline{Per_0} = 0{,}19$ показують, що така фільтрація наукових статей покращує щільність ключовиків у 1,48 раз або на 47,83 відсотка. На рис. 6 наведена порівняльна діаграма відсотків вживання знайдених системою ключових слів в відфільтрованому тексті (без початку (назва, автори, удк, анотації двома мовами, авторські ключові слова двома мовами, місце роботи авторів) і без списку

Таблиця 4 – Статистичні дані досліджених змісту текстів статей

Назва	Вага слова	Етап 1					Етап 2				
		A	*B*	*C*	*D*	*E*	*A*	*B*	*C*	*D*	*E*
Крок 1	≥ 1	5,46	3,92	2,51	2,08	1,74	7,43	7,03	3,27	3	4,18
	≥ 2	1,08	0,88	0,63	0,59	0,26	2,67	2,64	1,65	1,54	1,12
	≥ 3	0,41	0,38	0,22	0,21	0,16	1,21	1,2	0,85	0,79	0,41
	≥ 4	0,15	0,13	0,09	0,09	0,04	0,46	0,45	0,33	0,31	0,15
	≥ 5	0	0	0	0	0	0	0	0	0	0
Крок 2	≥ 1	6,51	5,02	2,68	2,23	2,37	8,35	7,78	3,25	2,91	4,99
	≥ 2	1,34	1,11	0,74	0,72	0,39	3,12	3,07	1,81	1,67	1,43
	≥ 3	0,51	0,45	0,29	0,27	0,17	1,42	1,4	0,93	0,85	0,54
	≥ 4	0,19	0,17	0,12	0,12	0,05	0,73	0,72	0,45	0,42	0,31
	≥ 5	0,11	0,1	0,06	0,06	0,04	0,33	0,32	0,25	0,23	0,1

літератури) Per_f^v та первинному авторському тексті Per_0^v з врахуванням уточнення модератором тематичного словника через поповнення заблокованих слів. Отримані середні значення для 100 текстів $\overline{Per_f^v} = 0{,}34$ та $\overline{Per_0^v} = 0{,}25$ показують, що фільтрація з одночасною модерацією тематичного словника покращує щільність ключовиків у 1,35 раз або на 35,44%.

На рис. 7 наведена порівняльна діаграма відсотків вживання знайдених системою ключових слів в початковому первинному авторському тексті без уточнення модератором тематичного словника через поповнення заблокованих слів (Per_0) та з врахуванням уточнення модератором тематичного словника через поповнення заблокованих слів (Per_0^v).

Порівняння значень $\overline{Per_0} = 0{,}19$ та $\overline{Per_0^v} = 0{,}25$ демонструє ефективність модерації тематичного словника у початковому тексті – щільність ключовиків збільшується у 1,34 раз або на 34,33 відсотка. На рис. 8 наведена порівняльна діаграма відсотків вживання знайдених системою ключових слів в відфільтрованому авторському тексті без уточнення модератором тематичного словника через поповнення заблокованих слів (Per_f)

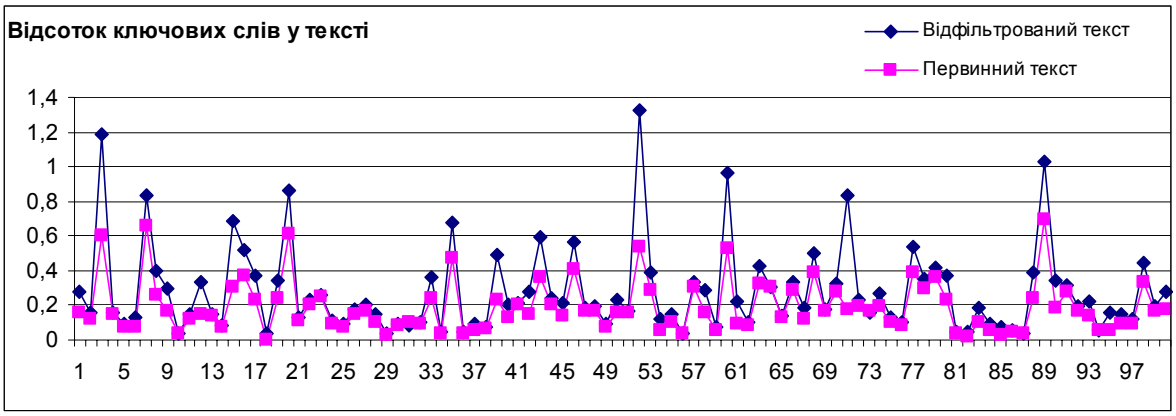


Рисунок 5 – Результати перевірки статей без уточнення модератором тематичного словника

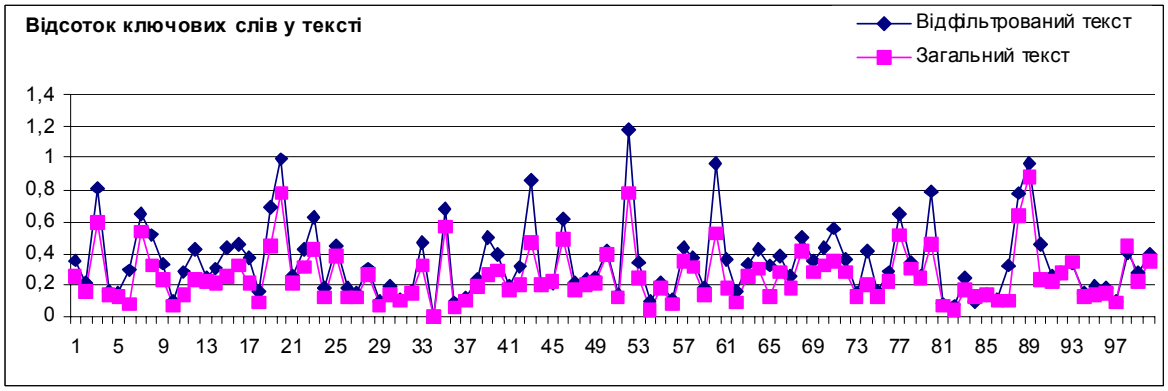


Рисунок 6 – Результати перевірки статей з врахуванням уточнення модератором тематичного словника

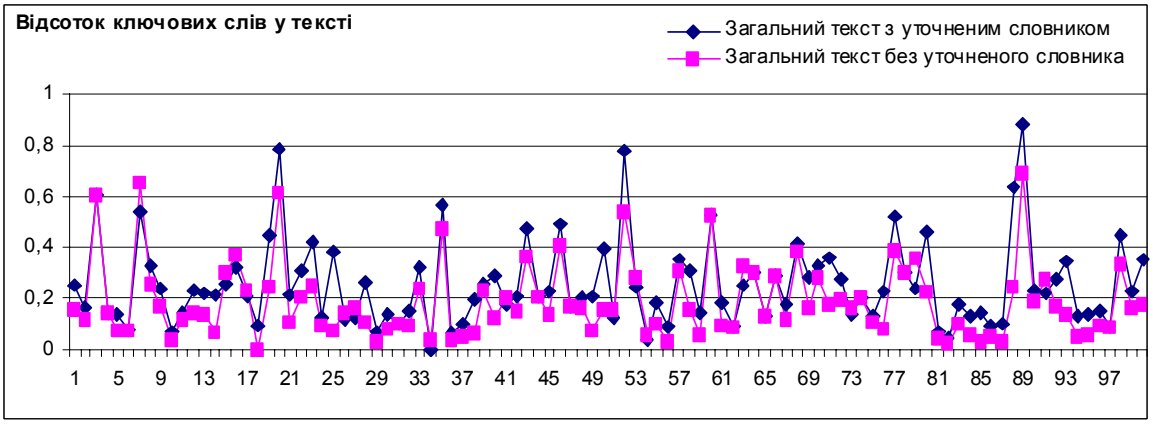


Рисунок 7 – Результати перевірки первинних авторських статей з різними словниками

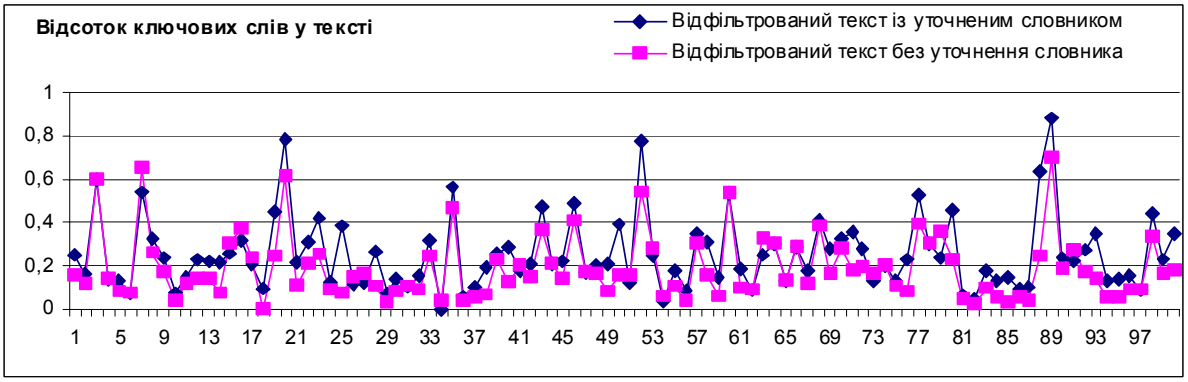


Рисунок 8 – Результати перевірки відфільтрованих статей з різними словниками

та з врахуванням модерації тематичного словника (Per_f^v).Порівняння значень $\overline{Per_f} = 0{,}28$ та $\overline{Per_f^v} = 0{,}34$ демонструє ефективність модерації тематичного словника у відфільтрованому тексті – щільність ключовиків збільшується у 1,23 раз або на 23,14 відсотка.

ВИСНОВКИ

У статті наведено теоретичне та експериментальне обґрунтування методу контент-моніторінгу україномовного тексту на основі стемінгу Портера. Метод спрямовано на автоматичне виявлення значущих ключових слів україномовного тексту за рахунок запропонованого формального підходу до реалізації стемінгу україномовного контенту. Проведено декомпозицію методу контент-моніторінгу на взаємопов'язані складові контент-аналізу текстової інформації та визначення множини ключових слів. Розроблено алгоритмічне забезпечення основних структурних складових запропонованого методу, а основу якого покладено адаптований до української мови алгоритм (стеммер) Портера. Теоретично виявлено способи покращення показників ефективності пошуку ключових слів, зокрема щільності ключовиків у тексті. Експериментальне дослідження 100 наукових публікацій з двох номерів (783 та 805) Вісника Національного університету «Львівська політехніка» серії «Інформаційні системи та мережі» (http://science.lp.edu.ua/sisn) продемонструвало позитивний вплив фільтрації тексту статті та модерації тематичного словника на визначення ключових слів. Виявлено, що для технічних наукових текстів експериментальної бази найкращих результатів досягає метод аналізу статті без початку (назва, автори, удк, анотації двома мовами, авторські ключові слова двома мовами, місце роботи авторів) і без списку літератури із перевіркою уточнених заблокованих слів та уточненого тематичного словника – для нього середнє значення щільності ключовиків у тексті досягає $\overline{Per_f^v} = 0{,}34$, що на 81% більше за аналогічне значення щільності первинного тексту $\overline{Per_0} = 0{,}19$. Потребує подальшого експериментального дослідження визначення ключових слів для інших категорій текстів – наукових гуманітарного про-філю, художніх, публіцистичних тощо.

ПОДЯКИ

У статті розв'язана науково-практична задача автоматичного виявлення значущих ключових слів та рубрикації україномовного контенту в Інтернет-системах на основі попереднього опрацювання відповідної текстової інформації. Роботу виконано в рамках спільних наукових досліджень кафедри інформаційних систем та мереж Національного університету «Львівська політехніка» на тему «Дослідження, розроблення і впровадження інтелектуальних розподілених інформаційних технологій та систем на основі ресурсів баз даних, сховищ даних, просторів даних та знань з метою прискорення процесів формування сучасного інформаційного суспільства», а також кафедри автоматики та інформаційно-вимірювальної техніки Вінницького національного технічного університету у межах діяльності науково-дослідного центру прикладної та комп'ютерної лінгвістики. Результати досліджень здійснювались у рамках держбюджетних науково-дослідних робіт за темами «Розробка методів, алгоритмів і програмних засобів моделювання, проектування та оптимізації інтелектуальних інформаційних систем на основі Web-технологій «ВЕБ» та «Інтелектуальна інформаційна технологія образного аналізу тексту та синтезу інтегрованої бази знань природно-мовного контенту». Наукові дослідження провадилися також в рамках ініціативної тематики досліджень кафедри ІСМ Національного університету «Львівська політехніка» на тему «Розроблення інтелектуальних розподілених систем на основі онтологічного підходу з метою інтеграції інформаційних ресурсів».

СПИСОК ЛІТЕРАТУРИ

1. Берко А. Системи електронної контент-комерції / А. Берко, В. Висоцька, В. Пасічник. – Л. : НУЛП, 2009. – 612 с.
2. Математична лінгвістика / [В. Висоцька, В. Пасічник, Ю. Щербина, Т. Шестакевич]. – Л. : «Новий Світ-2000», 2012. – 359 с.
3. Найефективніші методи залучення потенційних клієнтів [Електронний ресурс] / Центр ресурсів якості трафіку оголошень, Google AdWords. – Режим доступу: http://www.google.com/intl/uk_ALL/ads/adtrafficquality/advertisers/best-practices-for-generating-leads.html. – Назва з титул. екрану.
4. Нечеткий поиск в тексте и словаре [Електронний ресурс]. – Режим доступу: http://habrahabr.ru/post/114997/. – Назва з титул. екрану.
5. Реализации алгоритмов. Расстояние Левенштейна [Електронний ресурс]. – Режим доступу: http://ru.wikibooks.org/wiki/Реализации_алгоритмов/Расстояние_Левенштейна. – Назва з титул. екрану.
6. Задача о расстоянии Дамерау-Левенштейна [Електронний ресурс]. – Режим доступу: http://neerc.ifmo.ru/wiki/index.php?title=%D0%97%D0%B0%D0%B4%D0%B0%D1%87%D0%B0_%D0%BE_%D1%80%D0%B0%D1%81%D1%82%D0%BE%D1%8F%D0%BD%D0%B8%D0%B8_%D0%94%D0%B0%D0%BC%D0%B5%D1%80%D0%B0%D1%83-%D0%9B%D0%B5%D0%B2%D0%B5%D0%BD%D1%88%D1%82%D0%B5%D0%B9%D0%BD%D0%B0. – Назва з титул. екрану.
7. Насонов Д. Функция Левенштейна [Електронний ресурс] / Д. Насонов. – Режим доступу: http://rain.ifmo.ru/cat/data/theory/unsorted/levenshtein-2006/article.pdf. – Назва з титул. екрану.

8. Левенштейн, который сравнивает строки [Електронний ресурс] / Веб-разработка. – Режим доступу: http://dayte2.com/levenshtein. – Назва з титул. екрану.
9. Вычисление расстояния Левенштейна между двумя строками [Електронний ресурс]. – Режим доступу: http://wm-help.net/lib/b/book/827961078/78. – Назва з титул. екрану.
10. Стеммер Потера [Електронний ресурс]. – Режим доступу: http://labs.abcvg.com/stemmer/index.php. – Назва з титул. екрану.
11. Moseichuk V. Porter stemming algorithm for Ukrainian languages [Electronic resource] / V. Moseichuk. – Access mode: http://www.marazm.org.ua/document/stemer_ua/. – Title from the screen.
12. Стемінг [Електронний ресурс]. – Режим доступу: https://uk.wikipedia.org/wiki/Стемінг. – Назва з титул. екрану.
13. Russian stemming algorithm [Electronic resource]. – Access mode: http://snowball.tartarus.org/algorithms/russian/stemmer.html. – Title from the screen.
14. Porter stemmer – реализация алгоритма стеммера Портера для русского языка на чистом функциональном языке Clojure [Електронний ресурс]. – Режим доступу: https://github.com/allaud/porter-stemmer. – Назва з титул. екрану.
15. The Porter Stemming Algorithm – Porter's homepage. [Електронний ресурс]. – Режим доступу: http://tartarus.org/~martin/PorterStemmer/. – Назва з титул. екрану.
16. The Porter Stemming Algorithm – Project «Snowball» [Electronic resource]. – Access mode: http://snowball.tartarus.org/algorithms/porter/stemmer.html. – Title from the screen.
17. The English (Porter2) stemming algorithm – Project «Snowball» [Electronic resource]. – Access mode: http://snowball.tartarus.org/algorithms/english/stemmer.html. – Title from the screen.
18. Porter M. F. An algorithm for suffix stripping [Electronic resource] / M. F. Porter // Program. – 1980. – Т. 14, № 3. – С. 130–137. – Access mode: http://telemat.det.unifi.it/book/2001/wchange/download/stem_porter.html. – Title from the screen.
19. Willett P. The Porter stemming algorithm: then and now [Electronic resource] / P. Willett // Program: Electronic Library and Information Systems. – 2006. – В. 3, Т. 40. – С. 219–223. – ISSN 0033-0337. – Access mode: http://eprints.whiterose.ac.uk/1434/. – Title from the screen.
20. Сеник М. Вільний алгоритм стемінгу для української мови [Електронний ресурс] / М. Сеник. – Режим доступу: http://www.senyk.poltava.ua/projects/ukr_stemming/stemming_about.html. – Назва з титул. екрану.
21. Сеник М. Інструмент для пошуку слів з однаковими закінченнями [Електронний ресурс] / М. Сеник. – Режим доступу: http://www.senyk.poltava.ua/projects/ukr_stemming/word_by_ending.html. – Назва з титул. екрану.
22. Сеник М. Статичне дерево закінчень [Електронний ресурс] / М. Сеник. – Режим доступу: http://www.senyk.poltava.ua/projects/ukr_stemming/ukr_endings.html#dyn. – Назва з титул. екрану.
23. Сеник М. Демо стемінгу для української мови [Електронний ресурс] / М. Сеник. – Режим доступу: http://www.senyk.poltava.ua/projects/ukr_stemming/demo.html. – Назва з титул. екрану.
24. Вероятностный морфологический анализатор русского и украинского языков [Електронний ресурс]. – Режим доступу: http://www.keva.ru/stemka/stemka.html. – Назва з титул. екрану.
25. Стемминг [Електронний ресурс]. – Режим доступу: https://ru.wikipedia.org/wiki/Стемминг. – Назва з титул. екрану.
26. Lovins J. B. Development of a stemming algorithm / J. B. Lovins // Mechanical Translation and Computational Linguistics 11:22–31. – 1968.
27. Jongejan, B. Automatic training of lemmatization rules that handle morphological changes in pre-, in- and suffixes alike [Electronic resource] / B. Jongejan, H. Dalianis // In the Proceeding of the ACL-2009, Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Singapore, August 2–7, 2009, pp. 145–153. – Access mode: http://www.aclweb.org/anthology/P/P09/P09-1017.pdf. – Title from the screen.
28. Вірогідний морфологічний аналізатор російської та української [Електронний ресурс]. – Режим доступу: http://www.keva.ru/stemka/stemka.html. – Назва з титул. екрану.
29. Модуль Drupal для стемінга українською. Новий модуль для алгоритму Стема для Українського пошуку з виділенням коренів [Електронний ресурс]. – Режим доступу: http://drupal.ua/node/1170. – Назва з титул. екрану.
30. Стемінг Портера для української мови [Електронний ресурс]. – Режим доступу: http://www.marazm.org.ua/document/stemer_ua/. – Назва з титул. екрану.
31. Hardcoded stemmer for Ukrainian [Electronic resource]. – Access mode: https://github.com/vgrichina/ukrainian-stemmer. – Title from the screen.
32. Perestoronin P. Стеммер Портера для русского языка [Електронний ресурс] / P. Perestoronin. – Режим доступу: http://blog.eigene.in/post/49598738049/snowball. – Назва з титул. екрану.

Стаття надійшла до редакції 23.12.2015.
Після доробки 04.01.2016.

Бисикало О. В.[1], Высоцкая В. А.[2]
[1]Д-р техн. наук, профессор, декан факультета компьютерных систем и автоматики Винницкого национального технического университета, Винница, Украина
[2]Канд. техн. наук, доцент кафедры «Информационные системы и сети» Национального университета «Львовская политехника», Львов, Украина

ВЫЯВЛЕНИЕ КЛЮЧЕВЫХ СЛОВ НА ОСНОВЕ МЕТОДА КОНТЕНТ-МОНИТОРИНГА УКРАИНОЯЗЫЧНЫХ ТЕКСТОВ

Решена задача разработки алгоритмического обеспечения процессов контент-мониторинга для решения задачи определения ключевых слов русскоязычного текста. Рассмотрено формальное обоснование метода контент-мониторинга текста с помощью Стеммер Портера, в основу модификации стемминг положены известны результаты классификации морфемнои и словообразовательной структуры дериватов украинского языка, выявление закономерностей комбинаторики аффиксов, моделирование структурной организации глаголов и суфиксальних существительных, а также морфонологичных модификаций в процессе словоизменения глагола и словоизменении и словообразовании прилагательных украинского языка. Проведения декомпозиции метода и разработано алгоритмическое обеспечение его основных структурных составляющих по результатам контент-анализа текста. Теоретически обнаружены способы улучшения показателей эффективности поиска ключевых слов, в том числе плотности ключевиков в тексте. На основе разработанного программного обеспечения получены результаты экспериментальной апробации предложенного метода контент-мониторинга для определения ключевых слов в научных текстах технического профиля. Выявлено, что для выбранной экспериментальной базы из 100 работ лучших результатов по критерию плотности достигает метод анализа статьи без начальной обязательной информации и без списка литературы, но с проверкой уточненных заблокированных слов и уточненного тематического словаря.

Ключевые слова: текст, украиноязычный, алгоритм, контент-мониторинг, ключевые слова, контент-анализ, Стеммер Портера, лингвистический анализ, синтаксический анализ.

Bisikalo O. V.[1], Vysotska V. A.[2]
[1]F.D., professor, Dean of Faculty for Computer Systems and Automation, Vinnytsia National Technical University, Vinnytsia, Ukraine
[2]Phd, associate professor of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine

IDENTIFYING KEYWORDS ON THE BASIS OF CONTENT MONITORING METHOD IN UKRAINIAN TEXTS

The task of developing algorithmic providing processes of content monitoring for the problem solution of determining a keyword in Ukrainian text is solved. The formal justification of content monitoring in text using Porter stemmer is considered. The basis of the stemming

modification is the known results of morpheme and word building structure derivatives classification in Ukrainian language, affix combinatorics patterns identification, modeling the structural organization of verbs and suffixal nouns and morphonological modifications in the verb inflection and word formation and inflection of adjectives in Ukrainian language. The method decomposition is conducted and the algorithmic software of its basic structural components of the text content analysis results is developed. Theoretically means to improve the performance indicators of keywords search are identified, including keyword density in text. Based on the software obtained results of experimental testing of the proposed method of content monitoring to keywords identification in scientific texts of technical profile are developed. It is detected that the chosen experimental base of 100 works the article analysis method the without the initial required information and without the reference list reaches the best results for the density criterion, but with the specified blocked words and qualifying thematic dictionary verification.

Keywords: text, a Ukrainian, algorithm, content monitoring, keywords, content analysis, Porter stemmer, linguistic analysis, parsing.

REFERENCES

1. Berko A., Vysotska V., Pasichnyk V. Systemy elektronnoyi kontent-komertsiyi. Leningrad, NULP, 2009, 612 p.
2. Vysotska V., Pasichnyk V., Scherbyna J., Shestakevych T. Matematychna linhvistyka. Leningrad, Novyy Svit-2000, 2012, 359 p.
3. Nayefektyvnishi metody zaluchennya potentsiynyh kliyentiv [Electronic resource]. Tsentr resursiv yakosti trafiku oholoshen, Google AdWords. Access mode: http://www.google.com/intl/uk_ALL/ads/adtrafficquality/advertisers/best-practices-for-generating-leads.html. Title from the screen.
4. Nechetkyy poysk v tekste y slovare [Electronic resource]. Access mode: http://habrahabr.ru/post/114997/. Title from the screen.
5. Realyzatsyy alhorytmov. Rasstoyanye Levenshteyna [Electronic resource]. Access mode: http://ru.wikibooks.org/wiki/Реализации_алгоритмов/Расстояние_Левенштейна. Title from the screen.
6. Zadacha o rasstoyanyy Damerau-Levenshteyna [Electronic resource]. Access mode: http://neerc.ifmo.ru/wiki/index.php?title=%D0%97%D0%B0%D0%B4%D0%B0%D1%87%D0%B0_%D0%BE_%D1%80%D0%B0%D1%81%D1%81%D1%82%D0%BE%D1%8F%D0%BD%D0%B8%D0%B8_%D0%94%D0%B0%D0%BC%D0%B5%D1%80%D0%B0%D0%B0%D1%83-%D0%9B%D0%B5%D0%B2%D0 %B5%D0%BD%D1%88%D1%82%D0%B5%D0%B9%D0%BD%D0%B0. Title from the screen.
7. Nasonov D. Funktsyya Levenshteyna [Electronic resource]. Access mode: http://rain.ifmo.ru/cat/data/theory/unsorted/levenshtein-2006/article.pdf. Title from the screen.
8. Levenshteyn, kotory sravnivaet stroki [Electronic resource]. Web development. Access mode: http://dayte2.com/levenshtein. – Title from the screen.
9. Vychislenie rasstoyaniya Levenshteyna mezhdu dvumya strokami [Electronic resource]. Access mode: http://wm-help.net/lib/b/book/827961078/78. Title from the screen.
10. Porter stemmer [Electronic resource]. Access mode: http://labs.abcvg.com/stemmer/index.php. Title from the screen.
11. Moseichuk V. Porter stemming algorithm for Ukrainian languages [Electronic resource]. Access mode: http://www.marazm.org.ua/document/stemer_ua/. Title from the screen.
12. Steming [Electronic resource]. Access mode: https://uk.wikipedia.org/wiki/Стемінг. Title from the screen.
13. Russian stemming algorithm [Electronic resource]. Access mode: http://snowball.tartarus.org/algorithms/russian/stemmer.html. Title from the screen.
14. Porter stemmer-realizatsiya algoritma stemmera Portera dlya russkogo yazyka na chistom funktsionalnom yazyke Clojure [Electronic resource]. Access mode: https://github.com/allaud/porter-stemmer. Title from the screen.
15. The Porter Stemming Algorithm-Porter's homepage [Electronic resource]. Access mode: http://tartarus.org/~martin/PorterStemmer/. Title from the screen.
16. The Porter Stemming Algorithm – Project «Snowball» [Electronic resource]. Access mode: http://snowball.tartarus.org/algorithms/porter/stemmer.html. – Title from the screen.
17. The English (Porter2) stemming algorithm – Project «Snowball» [Electronic resource]. Access mode: http://snowball.tartarus.org/algorithms/english/stemmer.html. Title from the screen.
18. Porter M. F. An algorithm for suffix stripping [Electronic resource], Program, 1980,Vol. 14, No. 3, pp. 130–137. Access mode: http://telemat.det.unifi.it/book/2001/wchange/download/stem_porter.html. Title from the screen.
19. Willett P. The Porter stemming algorithm: then and now [Electronic resource], Program: Electronic Library and Information Systems, 2006, B. 3, Vol. 40, pp. 219–223. ISSN 0033-0337. – Access mode: http://eprints.whiterose.ac.uk/1434. Title from the screen.
20. Senyk M. Vilnyy alhorytm steminhu dlya ukrayinskoyi movy [Electronic resource]. Access mode: http://www.senyk.poltava.ua/projects/ukr_stemming/stemming_about.html. Title from the screen.
21. Senyk M. Instrument dlya poshuku sliv z odnakovymy zakinchennyamy [Electronic resource], Access mode: http://www.senyk.poltava.ua/projects/ukr_stemming/word_by_ending.html. Title from the screen.
22. Senyk M. Statychne derevo zakinchen [Electronic resource]. Access mode: http://www.senyk.poltava.ua/projects/ukr_stemming/ukr_endings.html#dyn. Title from the screen.
23. Senyk M. Demo steminhu dlya ukrayinskoyi movy [Electronic resource]. Access mode: http://www.senyk.poltava.ua/projects/ukr_stemming/demo.html. Title from the screen.
24. Veroyatnostny morfologichesky analizator russkogo i ukrainskogo yazykov [Electronic resource]. Access mode: http://www.keva.ru/stemka/stemka.html. Title from the screen.
25. Steming [Electronic resource]. Access mode: https://ru.wikipedia.org/wiki/Стемминг. Title from the screen.
26. Lovins J. B. Development of a stemming algorithm, Mechanical Translation and Computational Linguistics, 11:22–31. – 1968.
27. Jongejan B., Dalianis H. Automatic training of lemmatization rules that handle morphological changes in pre-, in- and suffixes alike [Electronic resource], In the Proceeding of the ACL-2009, Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Singapore, August 2–7, 2009, pp. 145–153. Access mode: http://www.aclweb.org/anthology/P/P09/P09-1017.pdf. Title from the screen.
28. Virohidnyy morfolohichnyy analizator rosiyskoyi ta ukrayinskoyi [Electronic resource]. – Access mode: http://www.keva.ru/stemka/stemka.html. – Title from the screen.
29. Modul Drupal dlya steminha ukrayinskoyu. Novyy modul dlya alhorytmu Stema dlya Ukrayinskoho poshuku z vydilennyam koreniv [Electronic resource]. – Access mode: http://drupal.ua/node/1170. – Title from the screen.
30. Steminh Portera dlya ukrayinskoyi movy [Electronic resource]. – Access mode: http://www.marazm.org.ua/document/stemer_ua/. – Title from the screen.
31. Hardcoded stemmer for Ukrainian [Electronic resource]. – Access mode: https://github.com/vgrichina/ukrainian-stemmer. – Title from the screen.
32. Perestoronin, P. Stemmer Portera dlya russkogo yazyka [Electronic resource]. – P. Perestoronin // Access mode: http://blog.eigene.in/post/49598738049/snowball. – Title from the screen.

УДК 681.518

Гурко А. Г.[1], Плахтеев А. П.[2], Плахтеев П. А.[3]

[1]Канд. техн. наук, доцент, доцент кафедры автоматизации и компьютерно-интегрированных технологий Харьковского национального автомобильно-дорожного университета, Харьков, Украина
[2]Канд. техн. наук, доцент, доцент кафедры автоматизации и компьютерно-интегрированных технологий Харьковского национального автомобильно-дорожного университета, Харьков, Украина
[3]Инженер, младший научный сотрудник кафедры автоматизации и компьютерно-интегрированных технологий Харьковского национального автомобильно-дорожного университета, Харьков, Украина

ПОВЫШЕНИЕ ТОЧНОСТИ ОЦЕНКИ СОСТОЯНИЯ ДИНАМИЧНЫХ ОБЪЕКТОВ КОМПЛЕКСОМ MATLAB-ARDUINO ПРИ ПРОЕКТИРОВАНИИ КИБЕР-ФИЗИЧЕСКИХ СИСТЕМ

Решена задача повышения эффективности взаимодействия MATLAB и Arduino при проектировании кибер-физических систем путем внесения изменений в реализацию стандартного протокола обмена со стороны Arduino. Предложено при запросе MATLAB на чтение данных с первого аналогового порта Arduino выполнять аналогово-цифровое преобразование данных со всех требуемых аналоговых портов с дальнейшей последовательной передачей полученных данных в MATLAB, что позволяет повысить качество управления динамичными процессами за счет уменьшения области неопределенности состояния многомерной быстродействующей системы управления. Кроме того, предложено функции предварительной обработки показаний датчиков выполнять средствами Arduino, что повышает гибкость кибер-физической системы за счет возможности изменения аппаратного обеспечения без изменения программного кода и протокола обмена MATLAB. Формальное описание взаимодействия MATLAB и Arduino позволяет реализовать протокол обмена с использованием беспроводных интерфейсов, микропроцессорных устройств и платформ, не совместимых с Arduino.

Ключевые слова: кибер-физическая система, MATLAB, Arduino, протокол обмена, период опроса.

НОМЕНКЛАТУРА

CPS – кибер-физическая система;
e – ошибка управления;
k – момент дискретизации по времени;
K_p – коэффициент усиления пропорционального канала регулятора;
T_i, T_d – постоянные времени интегрирующего и дифференцирующего каналов регулятора;
T_R – период чтения данных;
T_s – период дискретизации по времени;
T_W – период записи данных;
u – управляющее воздействие;
val – символьная переменная.

ВВЕДЕНИЕ

В результате взаимодействия встроенных систем управления и сетевых технологий возникла новая технология управления объектами и процессами – кибер-физическая система (cyber-physical system, CPS), которая характеризуется тесной интеграцией и координацией между вычислительными и физическими процессами при помощи сетевых технологий [1]. CPS широко внедряются в самые разнообразные отрасли производства: транспорт, энергетику, машиностроение, робототехнику, металлургию, а также в биомедицинские системы.

Проектирование CPS требует наличия более надежных моделей физических процессов, протекающих в системах управления [2, 3, 4]. От того, как модель соотносится с реальностью, зависит работоспособность CPS.

В промышленной практике аппаратные и программные составляющие многих систем управления до сих пор разрабатываются отдельно, без учета их взаимодействия между собой и с физическим миром. И уже после разработки системы управления, проверки ее на моделях, устраняется влияние различного рода неопределенностей путем использования специальных методов настройки. Этот процесс является трудоемким и дорогостоящим, а с усложнением систем – практически неосуществимым.

Альтернативным подходом к проектированию CPS является использование технологии модельно-ориентированного проектирования [4]. Ее суть заключается в том, что вместо физических прототипов в специализированном программном пакете создается имитационная модель объекта управления, на которой отрабатывается и совершенствуется разрабатываемый алгоритм управления. В случае достижения алгоритмом поставленной задачи выполняется автоматическая генерация программного кода для целевой платформы. Чаще всего такое проектирование проводят в средах LabVIEW компании National Instruments и MATLAB/Simulink компании Mathworks [5, 6]. Возможности этих продуктов существенно расширяются при использовании связи с физическим миром, для чего в них предусмотрены коммуникационные функции. Различные датчики и исполнительные устройства могут подключаться через COM порт или USB. Это позволяет перейти от имитационных моделей к гибридным, в которых сочетаются как модели сложных объектов, так и реальные физические устройства.

Однако использование указанных функций при необходимости реализации протоколов обмена вызывает определенные трудности, и до недавнего времени было доступно лишь специалистам узкого профиля, поэтому встречалось довольно редко.

Ситуация изменилась с появлением платформы Arduino, которая существенно упрощает разработку CPS. В частности, унифицированные функции взаимодействия с сервером на Arduino получили поддержку в MATLAB путем разработки приложения «Arduino

DOI 10.15588/1607-3274-2016-1-10

Support from MATLAB» [7]. Данная работа направлена на повышение эффективности взаимодействия MATLAB и платформы Arduino при гибридном (модель/устройство) подходе к проектированию систем управления.

1 ПОСТАНОВКА ЗАДАЧИ

Привлечение MATLAB позволяет не только упростить процедуру проектирования, но и решать достаточно сложные задачи управления, используя ПЭВМ с установленным MATLAB в качестве ПЭВМ верхнего уровня, в то время как платформа Arduino решает задачи по получению и преобразованию информации о значениях контролируемых параметров системы управления и реализует протокол обмена доступа к серверу (рис. 1).

Реализация текстового обмена вида запрос/ответ между MATLAB и Arduino осуществляется по последовательному USB интерфейсу с использованием механизма виртуальных COM-портов (рис. 2) при помощи команд, сходных с командами среды Arduino. Известны реализации серверных протоколов [7], где для повышения гибкости управления каждая команда оперирует одним из 14 цифровых или одним из 6 аналоговых выводов. В текстовых запросах/ответах каждому порту (pin 0…13) Arduino ставится в соответствие символьное значение a…n переменной val (табл. 1).

Основные команды, которые содержит стандартный текстовый протокол обмена MATLAB и Arduino, реализованный в MATLAB и в скетче adiosrv_UNO.ino для Arduino (настройка линий цифрового ввода-вывода, собственно цифровой ввод-вывод, аналоговый ввод и псевдоаналоговый вывод ШИМ), содержат код команды (функции) и строку параметров (рис. 3):

< function><pin>

<function><pin><val>

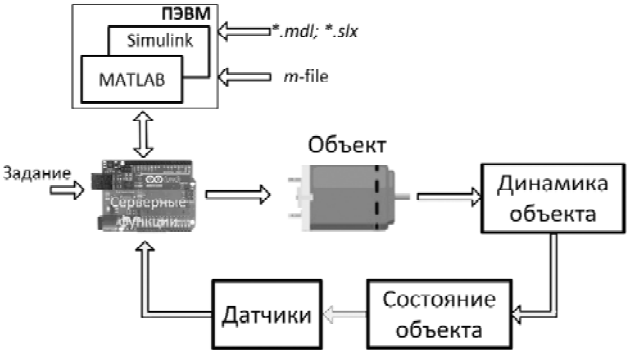


Рисунок 1 – Структура системы управления с Arduino

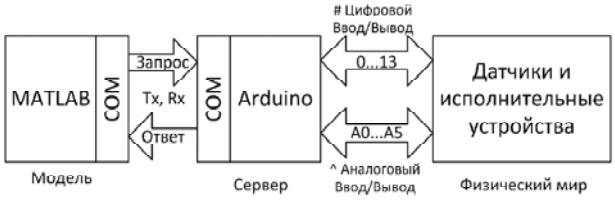


Рисунок 2 – Структурная схема взаимодействия MATLAB и Arduino

function∈ {‘0’…‘4’} = {0x30…0x34} = {48…52} – код запроса/ команды MATLAB;

pin∈{‘a’…’n’} = {0x61…0x6E} = {97 … 110} – символьные обозначения цифровых линий pin0 … pin13;

pin∈{‘a’ …’f’} = {0x61 … 0x66} = {97 … 102} – символьные обозначения аналоговых линий A0 … A5;

val∈{0…255} – байтовые значения псевдоаналогового вывода (ШИМ).

val∈{‘0’,’1’} – бинарное значение для задания режима или состояния линии цифрового ввода-вывода.

Например, команда ‘0e1’ устанавливает цифровой порт 4 (‘e’) в состояние вывода; команда ‘3b’ дает указание на выполнение аналого-цифрового преобразования сигнала с A1 (‘b’), а команда ‘4fz’ выдает через pin 5 (‘f’) параметр ШИМ 122 (‘z’ в ASCII равен 122).

Однако использование рассмотренного стандартного протокола обмена MATLAB – Arduino выявило следующие недостатки, которые затрудняют его применение при исследовании и управлении динамичными объектами в режиме реального времени.

1) Непостоянство периода опроса портов в зависимости от загруженности ПЭВМ. В то же время при реализации обратной связи информацию от датчиков о состоянии объекта необходимо получать со строгой привязкой ко времени.

Кроме того, последовательный опрос нескольких сигналов, определяющих состояние многомерной системы, приводит к значительному временному сдвигу получаемых отсчетов (τ_1, τ_2 на рис. 4), который также существенно зависит от скорости работы и загрузки компьютера и изменяется во времени (среднее время записи/ считывания с одного порта составляет 15 мс). Это затрудняет оценку состояния быстро меняющихся, динамичных процессов. Подобная проблема была также описана в [8–11]. Практически целесообразно, чтобы информация от всех портов поступала одновременно за одно обращение, либо, при последовательном чтении, соответствовала состояниям объекта в предельно близкие моменты времени.

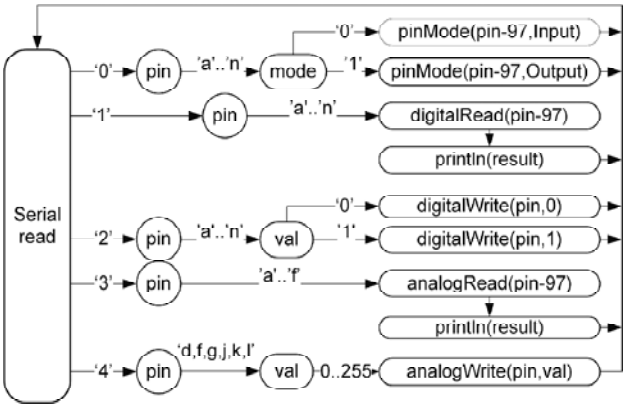


Рисунок 3 – Графическое представление протокола обмена MATLAB – Arduino

Таблица 1 – Соответствие номера порта символьной переменной val

val	a	b	c	d	e	f	g	h	i	j	k	l	m	n
pin	0	1	2	3	4	5	6	7	8	9	10	11	12	13

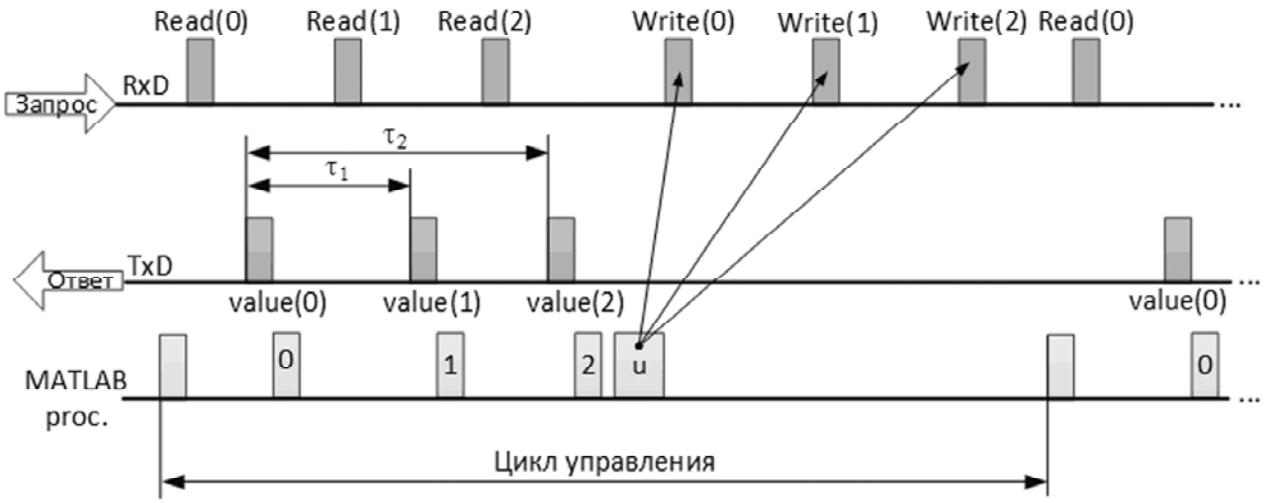


Рисунок 4 – Задержка определения состояния объекта при последовательном опросе портов Arduino

2) При использовании датчиков возникает необходимость предварительной обработки их показаний. Выполнять ее средствами MATLAB нецелесообразно, т.к. в этом случае имеет место жесткая привязка программы MATLAB к аппаратной части системы управления.

2 ОБЗОР ЛИТЕРАТУРЫ

Совместное использование MATLAB/Simulink и Arduino при проектировании кибер-физических систем в литературе освещено достаточно широко. Так в [9, 10, 12, 13] приведены примеры систем управления скоростью вращения вала двигателя постоянного тока как сервопривода различных объектов. В работе [14] представлено микропроцессорное устройство контроля параметров трехфазной сети по текущей информации от датчиков тока и напряжения. Устройство реализовано на плате Arduino Galileo. Для анализа точности выполняемых расчетов в среде MATLAB/Simulink разработана компьютерная модель микропроцессорного устройства, учитывающая дискретные преобразования аналоговых входных сигналов токов и напряжений по времени и уровню, измеренных непрерывными датчиками в трехфазной сети.

В [15] описана система управления уровнем жидкости в резервуаре на основе аппарата нечеткой логики. При этом для синтеза нечеткого регулятора использовались возможности MATLAB, а плата Arduino служила для получения данных от датчиков и генерации ШИМ-сигнала для управления насосом.

В [16] предложена система управления движением мобильного робота на основе Arduino, для программирования которой используются возможности MATLAB/Simulink.

Однако приведенные выше, и подавляющее большинство других публикаций посвящены описанию процедуры проектирования и принципов работы соответствующих систем управления и практически не рассматривают процессы, возникающие при взаимодействии Arduino и MATLAB. Результаты экспериментов с платой Arduino для оценки реализуемой ею точности измерения временных интервалов представлены в [11]. Однако в этой работе не предлагается методов повышения эффективности использования Arduino для управления объектами в реальном масштабе времени.

3 МАТЕРИАЛЫ И МЕТОДЫ

Рассмотренные выше недостатки стандартного протокола обмена между MATLAB и Arduino снижают эффективность использования гибридного подхода к проектированию систем управления. В связи с этим ниже предлагается модифицированный протокол обмена между MATLAB и Arduino, позволяющий устранить указанные недостатки.

Синхронизацию считываемых с портов данных предлагается осуществлять следующим образом. При получении запроса от MATLAB на чтение данных с первого аналогового порта (рис. 5), Arduino производит опрос всех требуемых аналоговых входов, тем самым формируется выборка данных, которые затем, при поступлении соответствующего запроса (функция ‘3’) передаются в MATLAB. И хотя в Arduino опрос осуществляется также последовательно и занимает определенное время (считывание значения с аналогового входа занимает 14 тактов частоты синхронизации АЦП с периодом 8мкс, тогда Тацп = 148 = 112 мкс), это значительно меньше, чем при программном опросе с помощью MATLAB (около 17 мс).

На рис. 5 жирными (красными) стрелками показано прохождение потоков данных между Arduino, MATLAB и в приложениях *.m, *.ino.

Функцию предварительной обработки показаний с датчиков предлагается возложить на платформу Arduino, тем более что она имеет значительные вычислительные возможности. Так, после выполнения аналогово-цифрового преобразования данных со всех требуемых аналоговых портов, Arduino производит обработку полученного кода и определяет численные значения контролируемых параметров в соответствующих единицах измерения, а затем передает полученные данные из Arduino на ПЭВМ через последовательный порт. Приведенная процедура позволяет более гибко подходить к выбору технических средств получения информации о состоянии физических параметров системы управления. При этом нет необходимости вносить изменения в программу MATLAB.

С учетом всего сказанного выше в данном параграфе, соответствующий рис. 5 фрагмент модернизированного протокола обмена MATLAB-Arduino отображается рис. 6.

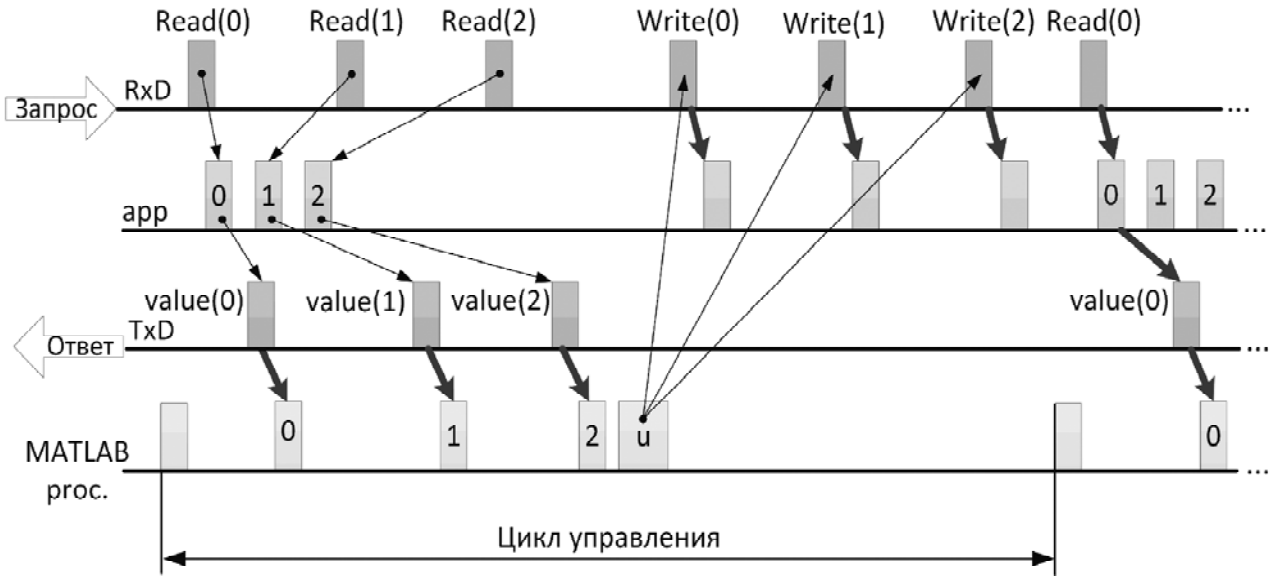


Рисунок 5 – Диаграмма опроса/ответа обмена при синхронизации считывания аналоговых сигналов

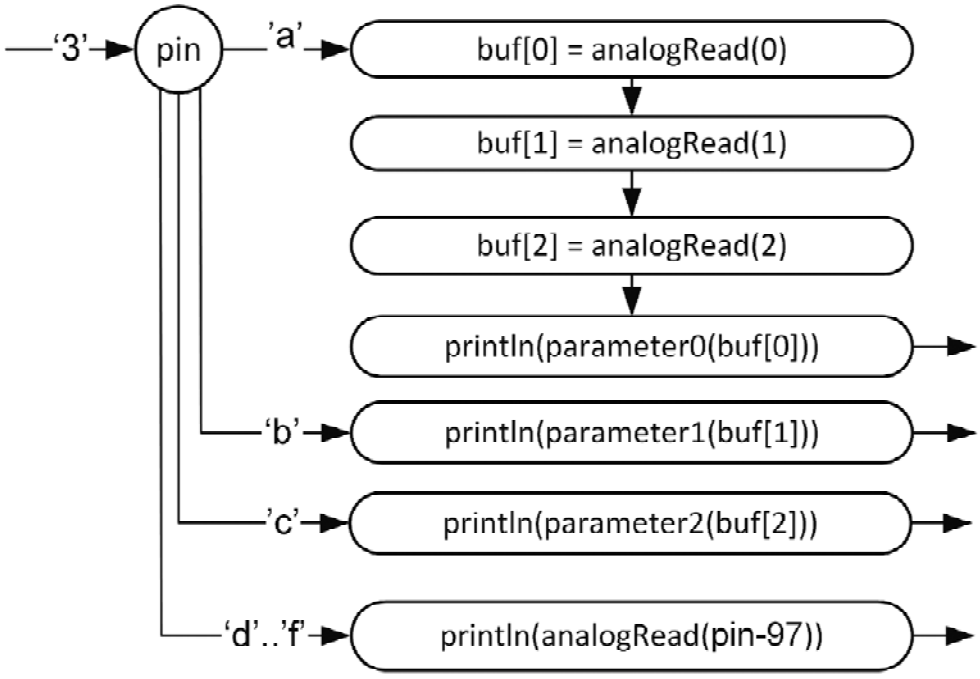


Рисунок 6 – Фрагмент модифицированного протокола обмена MATLAB-Arduino

На рис. 6 функция ‘3a’ запрашивает результат аналогово-цифрового преобразования сигнала с входа A0 (‘a’), но последовательно выполняется преобразование для A0, A1, A2 и результаты сохраняются в буфере памяти (buf[0..2]) и впоследствии подвергаются обработке с помощью различных, в общем случае, функций parameter0(buf[0]), parameter1(buf[1]), parameter2(buf[2]). Преобразование обработанного кода buf[0] в текстовый формат и отправка по последовательному интерфейсу в ПЭВМ реализует функция println(parameter0(buf[0])). По запросам MATLAB ‘3b’ и ‘3c’ возвращаются результаты преобразования buf[1] и buf[2]. В то же время по запросам ‘3d’..‘3f’ выполняется аналого-цифровое преобразование сигналов с входов A3..A5 и возвращаются результаты в текстовом виде – println(analogRead(pin-97)).

Например, если ‘a’ … ‘c’ (A0..A2) – входы для датчиков углового положения с разными характеристиками, то при указанной организации протокола в MATLAB возвращаются результаты преобразования напряжение-угол, скрывающие отличия датчиков. Аналогично можно организовать обработку для 1..6 произвольных аналоговых входов.

При необходимости жесткой цикличности опроса портов, например, при считывании информации с датчиков, функции отсчета временных интервалов целесообразно, как и буферизации данных, целесообразно возложить на Arduino. В этом случае MATLAB будет получать поток данных с переменной цикличностью, но массив полученных данных привязан к временной шкале.

4 ЭКСПЕРИМЕНТЫ

Для подтверждения наличия описанных в п.1 недостатков стандартного протокола обмена MATLAB-Arduino, а также проверки эффективности предлагаемой модификации протокола проведен ряд экспериментов на стенде, предназначенном для управления двигателем постоянного тока (ДПТ) посредством MATLAB и Arduino. Стенд (рис. 7) состоит из двигателя постоянного тока DC напряжением питания 5..12 В, ПЭВМ с MATLAB 2011b, Arduino Uno (MCU) с микроконтроллером Atmega328 с 32 кБ флеш-памяти (Flash), 1кБ ЭСППЗУ (EEP ROM) и 2 кБ SRAM, электронного ключа (ES), датчика скорости вращения двигателя (S). В качестве ПЭВМ использовался ноутбук с процессором Intel Core i3-330M 2,13 ГГц, ОЗУ 3ГБ и операционной системой Windows 7 Professional x64.

Задание желаемой скорости Vз вращения осуществляется напряжением U(Vз) = 0..5В с выхода потенциометра, а в качестве датчика скорости используется датчик Холла, частота импульсов f(V) на выходе которого определяется скоростью V вращения электродвигателя с 12-полюсным магнитом на валу. Преобразователь частота-напряжение (f/U) формирует напряжение U(V), пропорциональное скорости вращения двигателя. Управление двигателем производится посредством ШИМ-сигнала (PWM), параметры которого определяются на основе ПИД закона (1):

$$u_k = (u_{k-1} + K_p(e_k - e_{k-1}) + \frac{K_p T_s}{T_i} e_k + \frac{K_p T_d}{T_s}(e_k - 2e_{k-1} + e_{k-2})) \cdot \frac{255}{1023}. \quad (1)$$

При расчете параметров ПИД-регулятора использовались встроенные возможности MATLAB.

Взаимодействие MATLAB с Arduino осуществлялось на основе приведенного в п. 3 модифицированного протокола обмена.

5 РЕЗУЛЬТАТЫ

Рис. 8 отображает осциллограммы обмена между MATLAB и Arduino UNO посредством стандартного протокола. Период обмена включает: посылку MATLAB команды чтения аналогового входа A0 на вход последовательного обмена Arduino (RxD); прием значения (value), передаваемого по линии TxD Arduino; посылку команды аналогового вывода на pin 9, на которую ответ не передается. На рис. 8 видно изменение периода опроса D, зависящее от загрузки компьютера. Горизонтальная развертка составляет 2,5 мс/деление, вертикальная – 5В/деление, т.е. период T_1 составляет 12,5 мс, а период T_2 – 10 м/с.

Результаты серии экспериментов с обращением к аналоговым портам Arduino через MATLAB свидетельствуют о том, что основной вклад в вариацию периодов обмена вносит период чтения T_R (рис. 8), который, при обращении к одному порту, изменяется случайным образом в диапазоне $T_R = 8 \div 14$мс. В то же время, период записи T_W изменяется не так заметно и в среднем составляет около 3 мс. С увеличением количества задействованных портов период их опроса увеличивается практически пропорционально.

Динамику системы управления двигателем при действии внешней нагрузки отображает рис. 9, где U_x =U(Vз)/5 ·1023 – результат преобразования напряжения U(Vз), снимаемого с переменного резистора, в код 0..1023; U_v = U(V)/5 ·1023 – результат преобразования напряжения U(V), снимаемого с датчика скорости в код 0..1023; pwm – параметр ШИМ в диапазоне 0..255; int – выход интегрирующего канала ПИД-регулятора.

6 ОБСУЖДЕНИЕ

Результаты экспериментальных исследований подтверждают преимущества предложенного в данной работе протокола. Реализация процедуры синхронизации считываемых с портов данных позволяет на два порядка (с 15÷17 мс до 112 мкс) уменьшить время работы с аналоговыми портами Arduino. Это, а также уменьшение зависимости системы управления от непостоянства периода опроса портов Arduino, приводит к уменьшению неопределенности относительно области возможных состояний объекта. Очевидно, что при управлении инерционными объектами полученный результат не окажет заметного влияния на качество управления, однако при управлении быстро меняющимися, динамичными процессами, использование предложенного протокола дает ощутимый эффект.

ВЫВОДЫ

Распространению кибер-физических систем управления способствует применение современных методов модельно-ориентированного проектирования. При этом важную роль играет использование недорогих, относительно простых в использовании и в то же время эффективных периферийных устройств, обеспечивающих коммуникацию между моделью объекта или процесса управления и физическим миром. Одной из таких устройств является платформа Arduino.

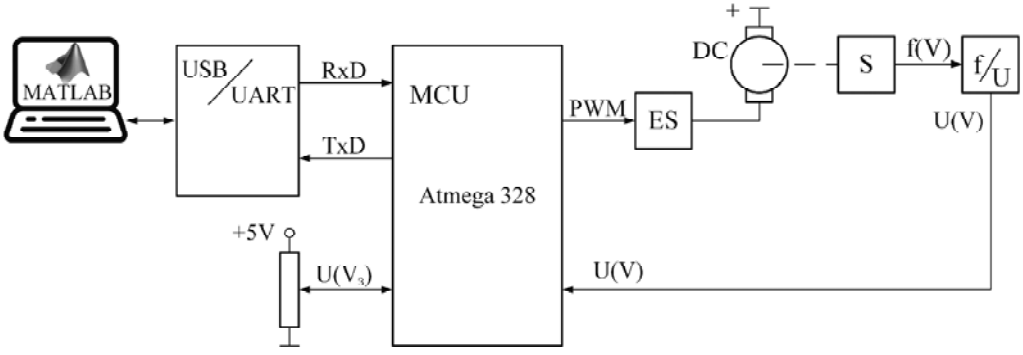


Рисунок 7 – Структурная схема стенда

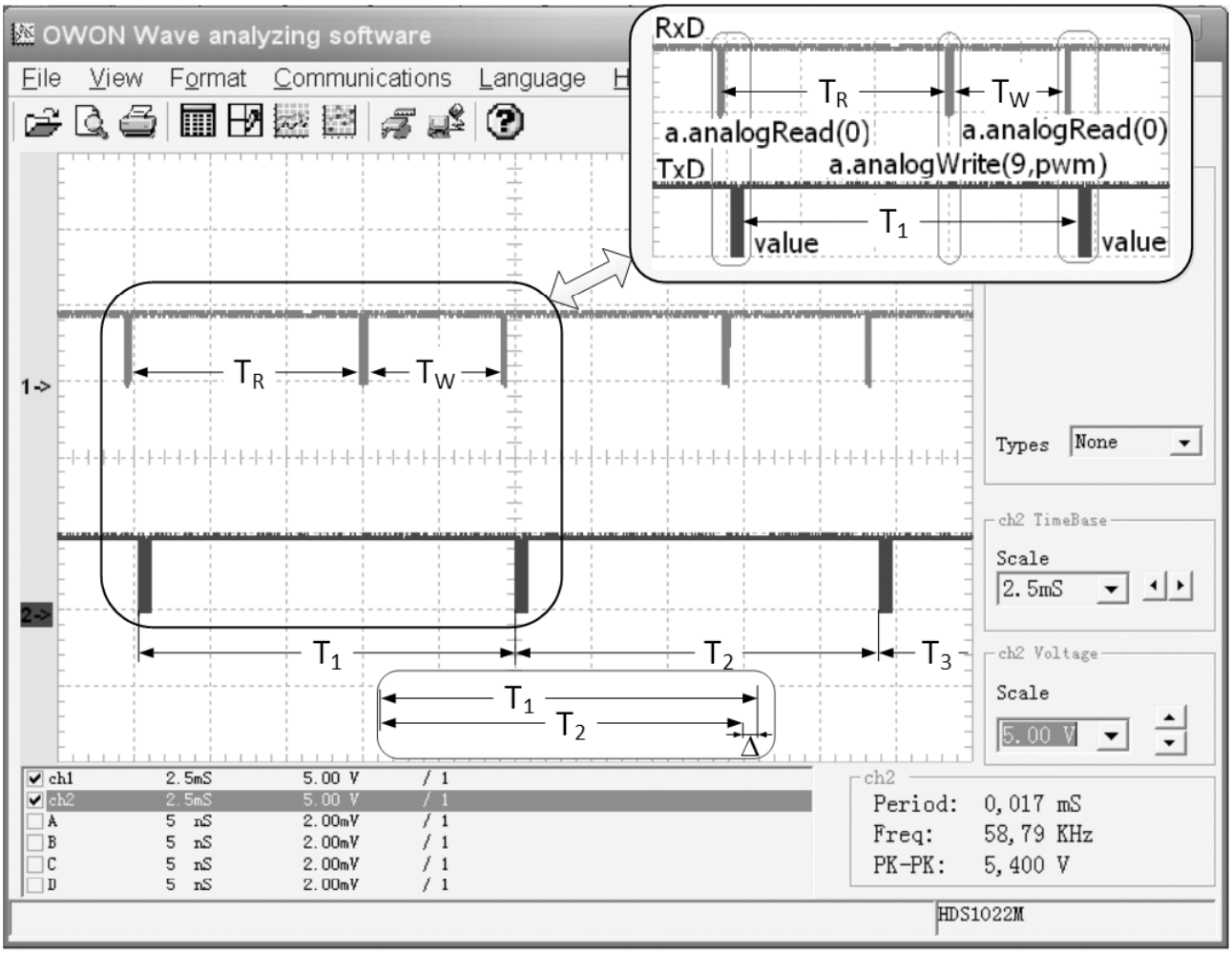


Рисунок 8 – Осциллограммы обмена между MATLAB и Arduino

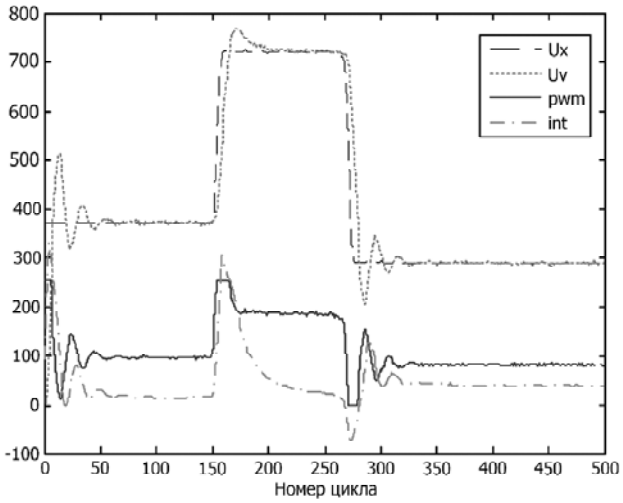


Рисунок 9 – Поведение системы управления двигателем

Однако использование стандартного протокола Arduino для обмена данными между ПЭВМ и физическим миром имеет определенные недостатки, связанные с непостоянством периода времени между получением выборки данных, невозможностью параллельного получения данных о состоянии объекта, а также нерациональным использованием ресурсов ПЭВМ.

Предложенный в данной работе протокол обмена данными между MATLAB и Arduino позволяет оптимизировать процесс взаимодействия системы управления с объектом и повысить качество управления динамичными многомерными процессами за счет уменьшения запаздывания в определении состояния многомерных объектов при последовательном обращении к аналоговым портам Arduino.

Выполнение предварительной обработки показаний датчиков средствами Arduino не влияет непосредственно на качество управления, однако повышает гибкость CPS за счет возможности изменения аппаратного обеспечения без изменения программного кода MATLAB.

Рассмотренный протокол может быть модифицирован для применения с другими средами моделирования, например, LabVIEW, а также реализован на основе иных платформ (Raspberry и проч.) и микроконтроллерах известных семейств – PIC, MSP430, MCS51, ARM, MISP и т.д. Последовательный обмен может быть построен с использованием беспроводных Bluetooth (HC05 и др.), WiFi (ESP8266 и др), XBee и прочих модемов. Это расширяет область применения предлагаемых подходов.

СПИСОК ЛИТЕРАТУРЫ

1. Sangiovanni-Vincentelli A. Taming Dr. Frankenstein: Contract-Based Design for Cyber-Physical Systems / A. Sangiovanni-

Vincentelli, W. Damm, R. Passerone // European Journal of Control. – 2012. – Vol. 18, № 3. – P. 217–238. DOI:10.3166/ejc.18.217-238.
2. Черняк Л. Киберфизические системы на старте / Л. Черняк // Открытые системы. СУБД. – 2014. – № 02 (198) – С. 10–13.
3. Park Kyung-Joon Cyber-physical systems: Milestones and research challenges / Kyung-Joon Park, Rong Zheng, Xue Liu // Computer Communications – 2012. – № 36. – P. 1–7. DOI 10.1016/j.comcom.2012.09.006.
4. Introduction to Embedded Systems – A Cyber-Physical Systems Approach / E. A. Lee and S. A. Seshia. Second Edition. [Electronic resource]. – Access mode: http://LeeSeshia.org, 2015.
5. Jensen J. C. A model-based design methodology for cyber-physical systems / J. C. Jensen, D. H. Chang, E. A. Lee // Wireless Communications and Mobile Computing Conference (IWCMC): 7th International conference, Istanbul, 4–8 July 2011: proceedings. – IEEE, 2011. – P. 1666–1671. DOI: 10.1109 IWCMC.2011.5982785.
6. Модельно-ориентированное проектирование программного обеспечения для встраиваемых систем в среде Matlab/Simulink / [Г. К. Топораш, А. В. Мазур, Д. А. Ковальчук, А. А. Пушкин] // Автоматизація технологічних і бізнес-процесів. – 2014. – Т. 17, № 17. – С. 26–29.
7. Практические стратегии для перехода на модельно-ориентированное проектирование встроенных приложений / [Д. Венси, Д. Эрик, К. Ларри, Р. Винод] // Компоненты и технологии. – 2011. – Т. 10, №. 123. – С. 172–180.
8. Arduino Support from MATLAB [Electronic resource]. – Access mode: http://www.mathworks.com/hardware-support/arduino-matlab.html.
9. Al-Busaidi A. M. Development of an educational environment for online control of a biped robot using MATLAB and Arduino / A. M. Al-Busaidi // Mechatronics: 9th France-Japan & 7th Europe-Asia Congress on and Research and Education in Mechatronics (REM), 21–23 Nov. 2012, 13th Int'l Workshop on. – IEEE, 2012. – P. 337–344. DOI: 10.1109/MECATRONICS.2012.6451030.
10. Barber R. Control Practices using Simulink with Arduino as Low Cost Hardware / R. Barber, M. Horra. J. Crespo. // The 10th IFAC Symposium Advances in Control Education, August 28-30, 2013: proceedings. – University of Sheffield, Sheffield, UK, 2013. – P. 250–255. DOI: 10.3182/20130828-3-UK-2039.00061.
11. D'Ausilio A. Arduino: A low-cost multipurpose lab equipment / Alessandro D'Ausilio // Behavior Research Methods. – 2012. – Vol. 44, Issue 2. – P. 305–313. DOI 10.3758/s13428-011-0163-z
12. Online Performance Optimization of a DC Motor Driving a Variable Pitch Propeller / [R. Cohen, D. Miculescu, K. Reilley and other] // arXiv preprint arXiv:1310.0133. – 2013.
13. Bawa D. Fuzzy control based solar tracker using Arduino Uno / D. Bawa, C.Y. Patil // International Journal of Engineering and Innovative Technology (IJEIT). – 2013. – Vol. 2, Issue 12. – P. 179–187.
14. Омельченко Е. Я. Устройство контроля трехфазной сети на основе Arduino-совместимых микроконтроллеров / Е. Я. Омельченко, А. В. Белый и др. // Электротехника. – 2014. – Т. 1, № 1. – С. 17–22.
15. Krivić S. Design and implementation of fuzzy controller on embedded computer for water level control / S. Krivić, M. Hujdur, A. Mrzić, S. Konjicija // MIPRO 2012: 35th International Convention, Opatija, Croatia, 21–25 May 2012: proceedings. – IEEE, 2012. – P. 1747–1751.
16. Gartseev I. B. A low-cost real-time mobile robot platform (ArEduBot) to support project-based learning in robotics & mechatronics / I. B. Gartseev, L. F. Lee, V. N. Krovi // 2nd International Conference on Robotics in Education (RiE 2011), Vienna, Austria, September 15–16, 2011: proceedings [Electronic resource]. – Access mode: http://mechatronics.eng.buffalo.edu/publications/conference/ICRE2011_IlyaLeeKrovi_ArEduBot.pdf

Статья поступила в редакцию 28.10.2015.
После доработки 09.11.2015.

Гурко О. Г.[1], Плахтєєв А. П.[2], Плахтєєв П. А.[3]

[1]Канд. техн. наук, доцент, доцент кафедри автоматизації та комп'ютерно-інтегрованих технологій Харківського національного автомобільно-дорожнього університету, Харків, Україна

[2]Канд. техн. наук, доцент, доцент кафедри автоматизації та комп'ютерно-інтегрованих технологій Харківського національного автомобільно-дорожнього університету, Харків, Україна

[3]Інженер, молодший науковий співробітник кафедри автоматизації та комп'ютерно-інтегрованих технологій Харківського національного автомобільно-дорожнього університету, Харків, Україна

ПІДВИЩЕННЯ ТОЧНОСТІ ОЦІНКИ СТАНУ ДИНАМІЧНИХ ОБ'ЄКТІВ КОМПЛЕКСОМ MATLAB-ARDUINO ПРИ ПРОЕКТУВАННІ КІБЕР-ФІЗИЧНИХ СИСТЕМ

Розв'язано задачу підвищення ефективності взаємодії MATLAB і Arduino при проектуванні кібер-фізичних систем шляхом внесення змін у реалізацію стандартного протоколу обміну з боку Arduino. Запропоновано при запиті MATLAB на читання даних з першого аналогового порту Arduino виконувати аналогово-цифрове перетворення даних з усіх необхідних аналогових портів з подальшою послідовною передачею отриманих даних в MATLAB, що дозволяє підвищити якість управління динамічними процесами за рахунок зменшення області невизначеності стану багатовимірної швидкодіючої системи управління. Крім того, запропоновано функції попередньої обробки показань датчиків виконувати засобами Arduino, що підвищує гнучкість кібер-фізичної системи за рахунок можливості зміни апаратного забезпечення без зміни програмного коду і протоколу обміну MATLAB. Формальний опис взаємодії MATLAB і Arduino дозволяє реалізувати протокол обміну з використанням бездротових інтерфейсів, мікропроцесорних пристроїв та платформ, не сумісних з Arduino.

Ключові слова: кібер-фізична система, MATLAB, Arduino, протокол обміну, період опитування.

Gurko A. G.[1], Plakhteev A. P.[2], Plakhteev P. A.[3]

[1]PhD, Associate professor, Associate professor of department of automation and computer-integrated technologies, Kharkiv National Automobile and Highway University, Kharkiv, Ukraine

[2]PhD, Associate professor, Associate professor of department of automation and computer-integrated technologies, Kharkiv National Automobile and Highway University, Kharkiv, Ukraine

[3]Engineer, Junior researcher of department of automation and computer-integrated technologies, Kharkiv National Automobile and Highway University, Kharkiv, Ukraine

ACCURACY INCREASE OF DYNAMIC OBJECTS STATE ESTIMATION BY A COMPLEX MATLAB-ARDUINO WHEN CYBER-PHYSICAL SYSTEMS DESIGNING

The problem of accuracy increase of MATLAB and Arduino interaction when cyber-physical systems designing by making changes into the implementation of the standard protocol of Arduino has been solved. It has been proposed to carry out the analog-digital conversion of all required analog pins with further sequential data transmission to MATLAB when requesting to read data from the first analog pin of Arduino.

It allows improving control quality of dynamic processes by reducing the uncertainty range of a high-speed multi-dimensional control system state. In addition, it has been offered to perform the preprocessing of sensors data by means of Arduino that increases the flexibility of cyber-physical systems due to the possibility of the hardware changing without MATLAB software and protocol changing. A formal description of MATLAB and Arduino interaction allows realizing the communications protocol using wireless interfaces as well as microprocessor units and platforms, which are not compatible with Arduino.

Keywords: cyber-physical system, MATLAB, Arduino, communications protocol, polling period.

REFERENCES

1. Sangiovanni-Vincentelli A., Damm W., Passerone R. Taming Dr. Frankenstein: Contract-Based Design for Cyber-Physical Systems, *European Journal of Control*, 2012, Vol. 18, No. 3, pp. 217–238. DOI:10.3166/ejc.18.217-238.
2. Chernjak L. Kiberfizicheskie sistemy na starte, *Otkrytye sistemy. SUBD,* 2014, No. 02 (198), pp. 10–13.
3. Park, Kyung-Joon, Rong Zheng, Xue Liu Cyber-physical systems: Milestones and research challenges, *Computer Communications*, 2012, No. 36, pp. 1–7. DOI 10.1016/j.comcom.2012.09.006.
4. Lee E. A. and Seshia S. A. Introduction to Embedded Systems – A Cyber-Physical Systems Approach. Second Edition. [Electronic resource]. Access mode: http://LeeSeshia.org, 2015.
5. Jensen J. C., Chang D. H., Lee E. A. A model-based design methodology for cyber-physical systems, *Wireless Communications and Mobile Computing Conference (IWCMC): 7th International conference, Istanbul, 4–8 July 2011: proceedings. IEEE*, 2011, P. 1666–1671. DOI: 10.1109/IWCMC.2011.5982785.
6. Toporash G. K., Mazur A. V., Koval'chuk D. A., Pushkin A. A. Model'no-orientirovannoe proektirovanie programmnogo obespechenija dlja vstraivaemyh sistem v srede Matlab/Simulink, *Avtomatizacija tehnologichnih i biznes-procesiv*, 2014, Vol. 17, No. 17, pp. 26–29.
7. Vensi D., Jerik D., Larri K., Vinod R. Prakticheskie strategii dlja perehoda na model'no-orientirovannoe proektirovanie vstroennyh prilozhenij, *Komponenty i tehnologii*, 2011, Vol. 10, No. 123, pp. 172–180.
8. Arduino Support from MATLAB [Electronic resource]. Access mode: http://www.mathworks.com/hardware-support/arduino-matlab.html.
9. Al-Busaidi A. M. Development of an educational environment for online control of a biped robot using MATLAB and Arduino, *Mechatronics: 9th France-Japan & 7th Europe-Asia Congress on and Research and Education in Mechatronics (REM), 21–23 Nov. 2012, 13th Int'l Workshop on. – IEEE, 2012*, pp. 337–344. DOI: 10.1109/MECATRONICS.2012.6451030.
10. Barber R., Horra M., Crespo J. Control Practices using Simulink with Arduino as Low Cost Hardware, *The 10th IFAC Symposium Advances in Control Education, August 28–30, 2013: proceedings. – University of Sheffield, Sheffield, UK, 2013,* pp. 250–255. DOI: 10.3182/20130828-3-UK-2039.00061.
11. D'Ausilio A. Arduino: A low-cost multipurpose lab equipment, *Behavior Research Methods*, 2012, Vol. 44, Issue 2, pp. 305–313. DOI 10.3758/s13428-011-0163-z
12. Cohen R., Miculescu D., Reilley K., Pakmehr M., Feron E. Online Performance Optimization of a DC Motor Driving a Variable Pitch Propeller, *arXiv preprint arXiv:1310.0133,* 2013.
13. Bawa D., Patil C. Y. Fuzzy control based solar tracker using Arduino Uno, *International Journal of Engineering and Innovative Technology (IJEIT),* 2013, Vol. 2, Issue 12, pp. 179–187.
14. Omel'chenko E. Ja., Belyj A. V. i dr Ustrojstvo kontrolja trehfaznoj seti na osnove Arduino-sovmestimyh mikrokontrollerov, *Jelektrotehnika*, 2014, Vol. 1, No. 1, pp. 17–22.
15. Krivić S., Hujdur M., Mrzić A., Konjicija S. Design and implementation of fuzzy controller on embedded computer for water level control, *MIPRO 2012: 35th International Convention, Opatija, Croatia, 21–25 May 2012: proceedings. IEEE, 2012*, pp. 1747–1751.
16. Gartseev I. B., Lee L. F., Krovi V. N. A low-cost real-time mobile robot platform (ArEduBot) to support project-based learning in robotics & mechatronics, *2nd International Conference on Robotics in Education (RiE 2011), Vienna, Austria, September 15–16, 2011: proceedings [Electronic resource]*. Access mode: http://mechatronics.eng.buffalo.edu/publications/conference/ICRE2011_IlyaLeeKrovi_ArEduBot.pdf

UDC 004.93

Lukac M.[1], Abdiyeva K.[2], Kameyama M.[3]

[1]Dr., Assistant Professor of Department Computer Science, Nazarbayev University, Astana, Kazakhstan
[2]Post-graduate Student, ROSE Laboratory, Nanyang Technological University, Singapore
[3]Dr., Professor of the Graduate School of Information Sciences, Tohoku University, Sendai, Japan

EVALUATION OF COMPONENT ALGORITHMS IN AN ALGORITHM SELECTION APPROACH FOR SEMANTIC SEGMENTATION BASED ON HIGH-LEVEL INFORMATION FEEDBACK

In this paper we discuss certain theoretical properties of the algorithm selection approach to the problem of semantic segmentation in computer vision. High quality algorithm selection is possible only if each algorithm's suitability is well known because only then the algorithm selection result can improve the best possible result given by a single algorithm. We show that an algorithm's evaluation score depends on final task; i.e. to properly evaluate an algorithm and to determine its suitability, only well formulated tasks must be used. When algorithm suitability is well known, the algorithm can be efficiently used for a task by applying it in the most favorable environmental conditions determined during the evaluation. The task dependent evaluation is demonstrated on segmentation and object recognition. Additionally, we also discuss the importance of high level symbolic knowledge in the selection process. The importance of this symbolic hypothesis is demonstrated on a set of learning experiments with a Bayesian Network, a SVM and with statistics obtained during algorithm selector training. We show that task dependent evaluation is required to allow efficient algorithm selection. We show that using symbolic preferences of algorithms, the accuracy of algorithm selection can be improved by 10 to 15% and the symbolic segmentation quality can be improved by up to 5% when compared with the best available algorithm.

Keywords: algorithm selection, algorithm suitability, computer vision.

NOMENCLATURE

ALE is Automated Labeling Environment;
BN is Bayesian Network;
CPMC is Constrained Parametric Min-Cuts for Automatic Object Segmentation;
FFT is Fast Fourier Transform;
HOG is Histogram of Oriented Gradients;
IA is Iterative Analysis;
MSER is Maximally Stable Extremal Regions;
SIFT is Scale Invariant Feature Transform;
SDS is Simultaneous Detection and Segmentation;
SVM is Support Vector Machine.

INTRODUCTION

The algorithm selection problem has been introduced by Rice [1] and since has been used in various applications. Recently it has been applied to machine vision and image processing [2, 3]. While in general the algorithm selection process works well [4–7], for more complex problem spaces, problems that are related to feature selection, evaluation and algorithm suitability have been recorded and reported [3, 8].

Algorithm selection is in general seen as a secondary solution to a problem because to select best algorithm from a set of available algorithms several preconditions must be satisfied: knowledge about the problem, knowledge about the algorithms, algorithm suitability and distinctive features for each algorithm must be known. Consequently, algorithm selection is neither easy to apply nor the least expensive solution. However, for complex problems that have large feature spaces including problems that deal with real-world situations and environment, algorithm selection is a viable alternative. The concept behind algorithm selection is in the algorithm separation: an algorithm that would deal with the problem successfully for all combinations of environmental conditions will be too complex but a set of more specific algorithms for subsets of conditions will provide better and cheaper solutions when applied on a case by case basis.

To obtain result improvement from a case by case selected set of algorithms, high quality selection mechanism with a minimal precision of selection is required: for a set of inputs, the selected algorithms must be such that the cumulative result is better than the best of the available algorithms. This implies that the algorithm selection mechanism must be able to select the best algorithm as often as possible.

A reliable algorithm selection implies that the set of available algorithms have been evaluated in a very strict setting and in a task dependent manner. As will be shown, task specific evaluation provides data that can be used for algorithm selection because only such evaluation results can be used to predict reliably algorithm results on new untested input data. This means that evaluation of a single algorithm cannot be seen as a holistic process but rather as a precise and specific process that is not generalizable.

Finally, the algorithm selection presented in this paper is situated within the framework for high level image understanding. We show that unlike standard feature-only based algorithm selection approaches, the high level symbolic description greatly improves the accuracy of algorithm selection as well as the final result of high level understanding.

1 PROBLEM STATEMENT

In this paper we analyze several problems of the algorithm selection:

– the impact of the high level symbolic understanding of image content on the accuracy of algorithm selection;

– the impact of algorithm evaluation on the algorithm selection process;

– the impact of feature for object recognition evaluation on the algorithm selection process.

DOI 10.15588/1607-3274-2016-1-11

2 REVIEW OF LITERATURE

Algorithm selection was introduced by Rice [1] in the context of selection of scheduling algorithm in computer operating system. Since then it has been applied to various problems and fields of research in different ways and granularity. In image processing and computer vision the algorithm selection has been used to determine the best algorithm for segmentation of artificially generated images of noisy geometrical shapes [4]. In [7] algorithm selection was used for determining the best algorithm for the segmentation of biological cell images and [5] used algorithm selection in a performance predicting framework.

For segmentation of more complex natural images [2] proposed an algorithm selection approach using machine learning and composition: final segmentation was created from partial segmentation from best algorithms for different regions of the image. The method showed that despite results with high accuracy of selection the final result was only as good as the best available algorithm. Finally, a more specific approach was used to select parameters in single algorithm for segmentation in [9].

In [10] uses depth information to estimate whole image properties such as occlusions, background and foreground isolation and point of view estimation to determine type of objects in the image. All the modules of this approach are processed in parallel and integrated in a final single step. An airport apron analysis is performed in [11] where the authors use motion tracking and understanding inspired by cognitive vision techniques. Finally, the image understanding can also be approached from a more holistic approach such as for instance in [12] where the intent is only to estimate the nature of the image and distinguish between mostly natural or artificial content.

Currently there is a large amount of work combining segmentation and recognition and some of them are [13, 14]. In [15] uses an interleaved object recognition and segmentation in such manner that the recognition is used to seed the segmentation and obtain more precise detected objects contours. In [16] objects are detected by combining part detection and segmentation in order to obtain better shapes of objects. More general approaches such as [17] build a list of available objects and categories by learning them from data samples and reducing them to relevant information using some dictionary tool. However this approach does not scale to arbitrary size because the labels are not structured and ultimately require complete knowledge of the whole world.

3 MATERIALS AND METHODS

In [8] an alternative approach to image understanding was proposed: an algorithm selection platform with verification of the high-level symbolic interpretation of the image content was proposed. This platform is used as basis of research in this paper and is shown in Fig. 1.

The platform works in two distinct modes and integrates both the algorithm selection from features and algorithm selection from high-level feedback. Initially, the input image is processed (box 1) by algorithm selected using the algorithm selection mechanism (box 3) that uses only the image features (Loop 1). The resulting high-level description of the image (obtained from the object recognition), is verified for logical contradictions (box 4) both on the context, on the part-level, on the location and on the relative size level. If the verification does not detect any high-level symbolic contradiction the processing stops and outputs the current high level description. If however a logical contradiction is detected, a hypothesis that solves the contradiction is generated (box 5). The image region that corresponds to the contradiction and to the hypothesis is used to extract local features, to determine local context information and to estimate attributes of the possible objects located in the selected image region. These three sources of information are used in the meta level to estimate what other algorithm should be used to correct the contradiction (Loop 2). This second loop is iterated over all contradictions until all contradictions are resolved. For the rest of this paper the presented system will be referred to as Iterative Analysis (IA).

Notice that the proposed system uses a twofold processing convergence. First convergence of the approach is to obtain a non-contradictory high-level description (contradiction resolution). The second convergence is the match between a description without contradiction and a set of algorithms (algorithm matching). The proposed approach thus combines processing quality with the meta-processing algorithm matching. This approach thus enables to exploit each algorithm's strongest points on an application, image features and image content basis.

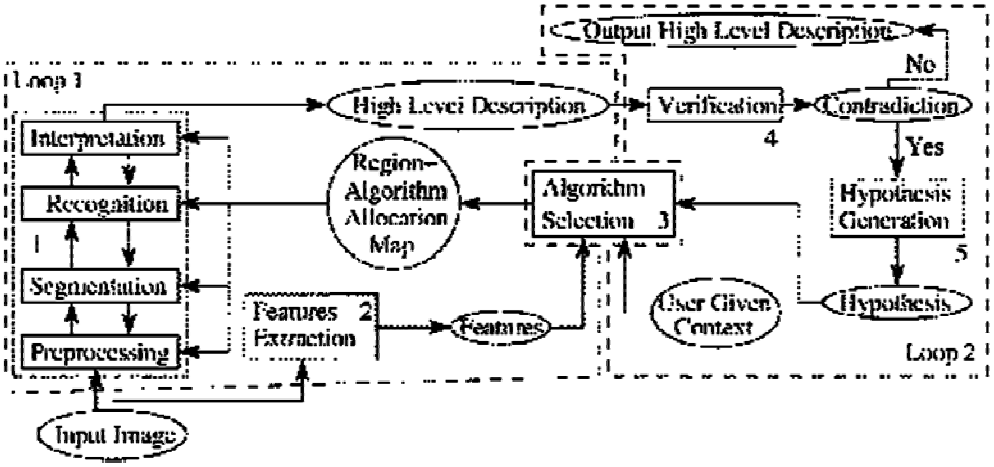


Figure 1 – Algorithm selection platform with verification of the high-level symbolic interpretation of the image content

The concept behind the processing in box 1, Figure 1 is that each algorithm used is a network of various component algorithms. Box 1 shows the general classical robotic sequential processing that uses four components processing levels: the preprocessing, segmentation, recognition and interpretation. However as in this paper the algorithms used are performing the semantic segmentation the interpretation is obtained by a single common algorithm. Also the selection is not limited to these four processing blocks but rather is intended to accommodate various algorithm networks.

As a final note some specific information about the selection process is required. In the initial loop of the IA processing, features extracted from the input images are FFT coefficients, Gabor features, wavelets, gist, color average, intensity average, edges, covariant features, SIFT, HOG, MSER and textures. All these features are transformed into a histogram and are concatenated into a single vector per image (or per region) of 5000 values.

For all loops after the initial one, the hypothesis is represented as a set of attributes. These attributes are obtained using the regprops function in Matlab. These attributes have been discretized in order to simplify the representation but to allow discrete representation of each of the available hypotheses.

In this paper the platform uses algorithms performing semantic segmentation: first segment an image and recognize regions as objects. The result of such processing is fed to the interpretation and verification according to the above platform description.

4 EXPERIMENTS

In order to assess an algorithm processing quality, it is necessary to evaluate its performance with respect to some training data set and ground truth. Each evaluation experiment was designed using real algorithm selection data. The algorithms used in our classification task are the ALE [11, 18], CPMC with recognition [14] and the SDS [19]. The three algorithms have similar performance results shown in Table 1. Here the numbers given in the original papers may vary due to different set up, initialization and training conditions of the original and this experiments.

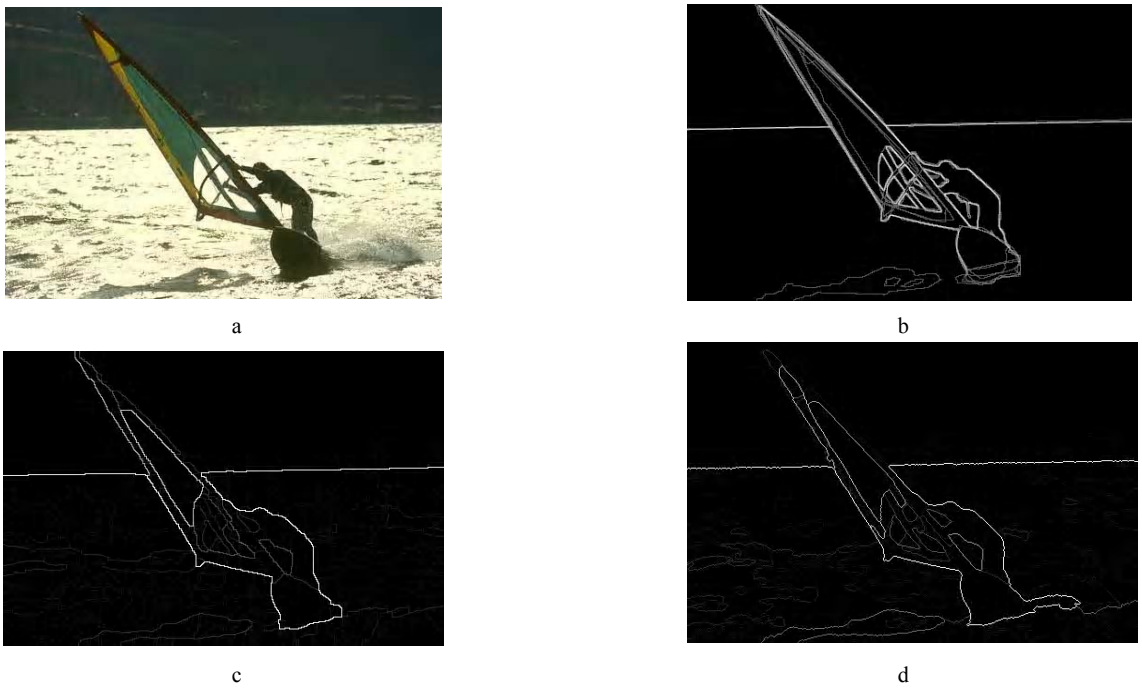
Consequently most of the algorithms that perform the semantic segmentation task first segments an image using some well-known segmentation algorithm and then apply the object recognition (there are other algorithms that are not using this order such as [15]).

Let us assume that an algorithm is evaluated for the quality of segmentation i.e. it evaluates whole image segmentation by comparing the result of processing to a human provided ground truth. Figure 2a shows an example of input image, Fig. 2b – human generated ground truth and Fig. 2c–Fig. 2d – the result of a segmentation algorithm. Fig. 2b – Fig. 2c have also their f-value shown in the parentheses. F-value is one of the standard measures used to determine the accuracy of computer generated segmentation [20]. According to the f-measure in this case of evaluation the algorithm generating the result shown in Fig. 2c is superior (closer when pixel-to-pixel comparison is done with human segmentation in Fig. 2b) to the algorithm which result is shown in Fig. 2d.

Now let's look at the same algorithms in the task of semantic segmentation. In semantic segmentation and input

Table 1 – Results of semantic segmentation algorithms on the VOC2012 validation dataset

Name	Result
ALE	47.8%
CPMC	48.3%
SDS	49.9%



a b c d

Figure 2 – Example of algorithmic Segmentation: a – input image, b – human ground truth (0.9), c – result of algorithm from [21] (0.92), d – result of algorithm from [22] (0.87)

image is segmented and then each region is labeled form a set of available object label set. In the task of semantic segmentation, the two best algorithms for image segmentation shown in Fig. 2c, Fig. 2d will not have the same f-values. In fact, algorithms with much lower f-value f= 0.77 (in the task of segmentation and with result shown in Fig. 3b) will have much higher resulting score because the regions obtained from the detected regions are more precise for object detection and labeling.

The reason for such change of the score is possible because in image segmentation the algorithm's result is evaluated by comparing the obtained boundaries with the ground truth generated by human. However, in the case of semantic segmentation the evaluation is made first by determining the boundaries of the target object and then the detection of the correct object is tested. This means that segmenting the whole image and comparing it to a set of human generated ground truth will result in more variation because even humans will not generally agree on how to segment a whole scene. This is because the evaluation is done with respect to a human segmentation that depends on feeling and intuition. However, when segmenting an image to determine object boundary the disparity between humans is much smaller. The semantic segmentation can be automatically judged on whether or not the correct object was correctly detected. Consequently despite the fact that some segmentation might be close enough to a human like segmentation it might not be well suited for the segmentation of a particular object.

Thus for two different tasks, the score of the final evaluation of a same algorithm might not be the same and the algorithm that had a good result in one tasks will have much lower result score for another task. But a statistical evaluation of algorithms might not be sufficient to determine advantages and disadvantages precisely enough. Figure 4 shows the standard model of robotics where multiple processes are formed into a set of consecutive algorithmic steps. The combination of algorithms can result in non-linear result that would not be observed otherwise. Thus it is necessary to evaluate the component algorithms as well so that individual suitabilities can be determined and impact on the result of the entire computation.

Similarly to the segmentation study a change of result can be obtained in recognition. Various features have different accuracy and ability to detect and recognize an object.

Using various features for detection (using the bag of words recognition model) it can be shown that depending on the region used to extract the feature descriptors and on the features extracted, the recognition accuracy will change. For instance assume that a segmentation algorithm such as [21, 25] is used for segmentation. The results of the segmentation are boundaries that indicate main regions of the image where the features for recognition should be extracted and the recognition model should be applied. Depending on what features are extracted the accuracy will change depending on the image. In some cases there will be no detection and in some other cases the detection will be a success.

Figure 5 shows the results of calculating the bounding box after two different features (SIFT and HOG) have been used for object recognition. In this case we extract features form the whole images. The features and the descriptors extracted are used to recognize a motor-bike and then the same feature descriptors are used to generate a bounding box. The idea behind this experiment is to assess the importance of a region in recognition of a motor-bike given that segmentation occurred prior to recognition.

The bounding box method determination is shown in Fig. 6 and Fig. 7.

Following the standard bag of words object recognition method a model of the object being detected is available as a set of histograms of feature clustered centers. Input image is first used as input for feature extraction, the feature descriptors are then clustered into k centers and a new histogram is constructed with bins corresponding to the k centers. Once the histogram is obtained, it is compared to all histograms in the model database and four closest matches are saved. Finally features corresponding to four best matching bins (each from one of the model histograms) are used to determine which descriptors and consequently which key points are used to determine the bounding box (Figure 7).

a b

Figure 3 – Example of two algorithms for segmentation with lower f-value:
a – result of segmentation of input image from Figure 2a using the algorithm from [23], b – result of segmentation of input image from Figure 2a using the algorithm from [24]

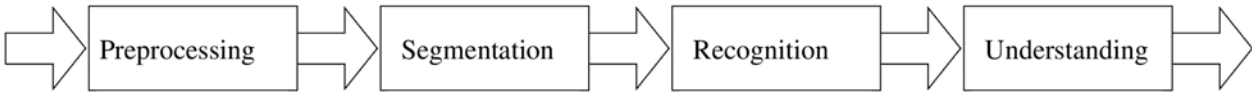
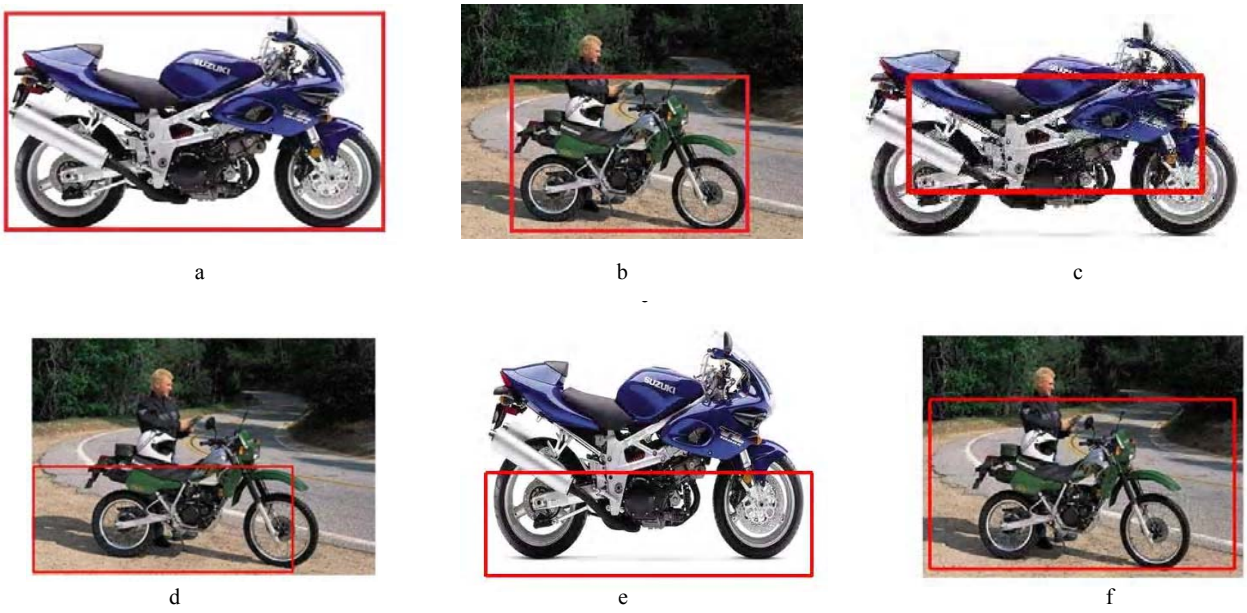


Figure 4 – Typical example of processing required for semantic segmentation



a b c

d e f

Figure 5 – Comparison of Bounding boxes obtained from SIFT and HOG features: a – Input Image 1 and Bounding Box by Human, b – Input Image 2 and Bounding Box by Human, c – Bounding box from SIFT, d – Bounding box from SIFT, e – Bounding Box from HOG, f – Bounding Box from HOG

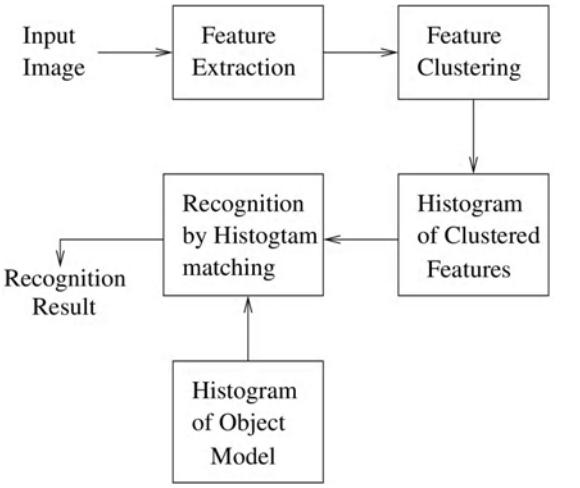


Figure 6 – Schema of Bag-of-words recognition algorithm

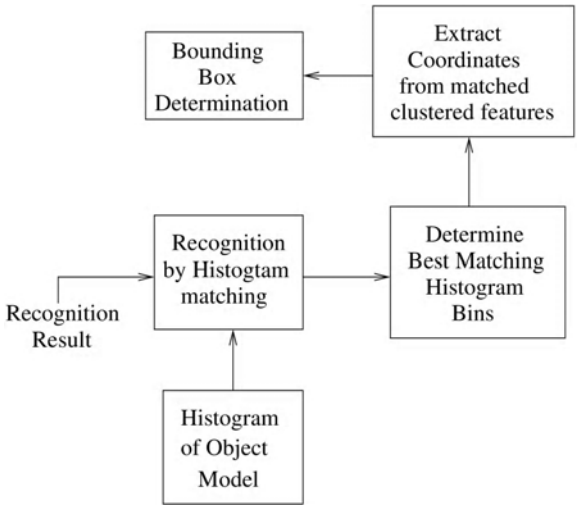


Figure 7 – Schema of Bounding Box extraction from a successful object recognition using Bag-of-words recognition algorithm

Using the method for determining the bounding box as an evaluation of the feature descriptor relevance to the motor-bicycle model, it can be seen that having different regions used to extract the feature descriptors would have a significant impact on object detection. For instance if only the upper left region (Region 1 in Fig. 8) of the bicycle would be contained in a single region no detection would happen if the HOG features would be used. Also if the SIFT features would be used it is possible that positive detection might not occur as not enough of the significant SIFT descriptors for successful detection are contained in the Region 1 only. On the other hand if the bottom of the motorcycle would be contained in a single region (Region 3 in Fig. 8) the HOG features would not be able to detect the motorcycle.

Consequently, using various features for only recognition or for semantic segmentation can have considerably different results as both the segmentation and the recognition are sensitive and difficult operations. Their evaluation is thus highly task dependent.

In the software platform previously introduced the algorithm selection is iterated through several iterations. The stopping condition for the processing of the image is either no more improvement is possible due to having tried all available algorithms or no more improvement is possible as the new hypothesis generated is the same as the previous one.

Initially, the algorithms are selected using only the image features but after the first processing loop the hypothesis generated is used for algorithm selection. The features have been successfully used for algorithm selection in various approaches, however in general such algorithm selector is limited due to the fact that many algorithms are designed for particular symbolic and semantic context.

The semantic segmentation results in a set of symbolically labeled regions and thus analyzing the obtained regions by various algorithms it is possible to conclude that various algorithms have affinities for different objects.

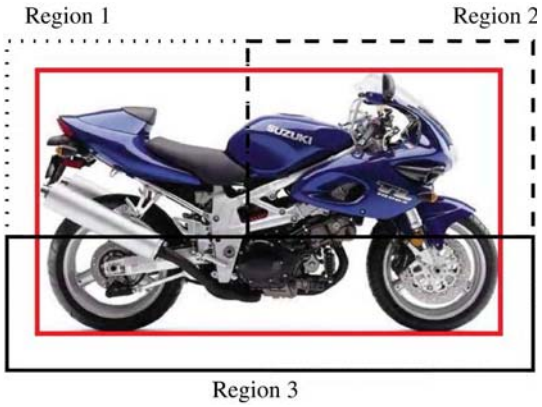


Figure 8 – Example of three different regions obtained as a result of a possible segmentation

Such affinity for particular objects can be due to the following reasons:

– The environment in which the particular object is captured has particular interaction with the object that is favorable to be detected by a particular algorithm.

– The object itself has a set of features that a particular algorithm is better suited for detection and segmentation.

Consequently we asked: what is the impact of symbolic information (content related) on the accuracy of algorithm selection?

To answer the above question we conducted a set of experiments. The experiments were carried using the VOC2012 [26] database. The dataset used is not the standard VOC2012 validation set but a reduced one in order to allow applying our platform. This means that only images where multiple objects to be segmented are present. The dataset is thus reduced and contains only ~300 images out of ~1500 images contained in the VOC2012 dataset.

5 RESULTS

The platform was initially designed to use Bayesian Network (BN) because the probabilistic inference is well suited to deal with missing variables. Consequently and because the two different modes of algorithm selection (features only and features with high level description), a single trained BN can be used. However selection of algorithms using Bayesian Network is still problematic and thus two alternative algorithm selectors were used for comparison. These two selection mechanisms are SVM and Statistics from training.

In a first experiment we compared the BN and the SVM because both of these algorithm selectors work on similar principles. Both BN and SVM are used in the initial and all further iterations of the IA platform. In the first iterations only features are used to select algorithm while in all further loops the features from the contradiction region as well as the hypothesis is used. The main difference between using the BN and SVM is that SVM requires incomplete input information imputation [28] while the BN is well suited to handle missing input information by design. This means that for the first iteration, the SVM is provided with average values of the hypothesis in order for the input vector has the desired and fixed length.

The results of comparison of the precision of the BN and of the SVM are shown in Table 2.

The problem of using the Bayesian Network is that it requires discrete input information. However most of the features extracted from input image are continuous and unbounded. Consequently it is required to cluster the input information and only then use it as input to the BN. This however has in most of the cases a dramatic influence on the performance of the probabilistic algorithm selection.

As can be seen the impact of hypothesis attributes is significant in the case of SVM, however in the case of BN it is difficult to evaluate as the overall precision is too low. The general increase of algorithm selection using the features and attributes compared to the selection using only features is up to 10% of accuracy.

The final evaluation of the high level information (feedback) in our system is the usage of statistical accuracy of each algorithm. The accuracy represents the percentile average of the f-measure of each semantic segmentation algorithm. Table 3 shows the accuracy of semantic segmentation for each of the three used algorithms: ALE [18], CPMC [14] and SDS [19]. Each columns shows average accuracy for each of the categories of objects that are to be recognized and segmented and overall average accuracy in the bottom row. The statistical information provided was obtained by evaluating the VOC2012 validation data set that contains approximately 1500 images.

The last column in Table 3 shows for each class of objects the best algorithm based on the statistical accuracy of each algorithm. This means that in the platform and during iterations all but the first one, for each hypothesis algorithm will be selected only using the best algorithm listed in the rightmost column. Using this approach we evaluated the proposed Iterative Analysis method described in this paper. The result comparison is shown in Table 4. It shows average precision for each algorithm for the test dataset.

Table 2 – Comparison of BN and SVM in selection accuracy using only features vs. features and attributes

Task	Algorithm	Features	Features+Attributes
2-class	SVM	58%	65%
2-class	BN	46%	55%
3-class	SVM	43%	46%
3-class	BN	39%	43%

Table 3 – Average accuracy of different algorithms on the training set

Accuracy for each class (intersection/union measure)	Algorithm Names			Best Algorithm
	ALE	SDS	CPMC	
Background	71.711	84.937	83.098	SDS
Aeroplane	52.096	60.927	64.404	CPMC
Bicycle	27.558	26.823	17.965	ALE
Bird	36.684	56.239	50.783	SDS
Boat	38.656	47.003	45.036	SDS
Bottle	43.643	48.465	41.605	SDS
Bus	65.787	70.559	69.104	SDS
Car	58.338	60.723	60.733	CPMC
Cat	63.789	59.847	56.524	ALE
Chair	24.001	20.815	11.663	ALE
Cow	64.853	42.112	52.842	ALE
Dining Table	41.339	38.694	19.406	ALE
Dog	55.190	51.535	48.995	ALE
Horse	58.998	43.653	43.899	ALE
Motorbike	56.909	52.300	52.858	ALE
Person	49.107	61.649	46.707	SDS
Potted Plant	31.408	37.360	40.563	CPMC
Sheep	53.563	51.829	49.285	ALE
Sofa	38.598	22.375	49.208	ALE
Train	53.910	56.288	58.319	CPMC
Average Accuracy	48.473	50.089	47.048	

Table 4 – Results of semantic segmentation accuracy on the test data set

Accuracy for each class (intersection/ union measure)	Algorithm's Name				Best Algorithm
	ALE	SDS	CPMC	IA	
Background	54.878	80.061	77.478	62.157	SDS
Aeroplane	0.000	0.000	0.000	0.000	--
Bicycle	26.799	31.913	13.515	27.624	SDS
Bird	22.070	37.042	59.947	21.932	CPMC
Boat	0.000	0.000	0.000	0.000	--
Bottle	37.445	50.990	39.280	50.226	SDS
Bus	44.212	12.034	71.156	45.412	CPMC
Car	52.788	34.924	31.873	56.241	IA
Cat	63.939	65.552	62.707	63.802	SDS
Chair	19.113	22.355	7.800	19.014	SDS
Cow	33.093	0.000	0.000	30.991	ALE
Dining Table	39.155	50.907	23.997	40.169	SDS
Dog	60.085	49.253	49.827	59.148	ALE
Horse	46.406	27.761	27.155	47.128	IA
Motorbike	61.154	28.477	33.949	61.697	IA
Person	46.362	63.940	46.068	57.947	SDS
Potted Plant	25.762	36.391	25.045	23.245	SDS
Sheep	69.008	66.129	27.191	69.008	IA
Sofa	29.672	17.062	11.806	29.702	IA
Train	43.602	0.000	28.651	51.174	IA
TVmonitor	31.320	62.904	53.201	37.091	SDS
Average Accuracy	38.422	35.128	32.935	40.653	IA

6 DISCUSSION

Notice that not all algorithms tested have an average score in all categories: this is due to the fact for the images that contained the object cow was not detected not even once by neither SDS nor CPMC. Moreover observe that our approach IA is best only in few categories but in most of the categories is relatively close to the best one. As a result of using the statistical information for algorithm selection the IA approach results in the best semantic segmentation.

As a final comment on the importance of high level image description and content understanding, Figure 9 shows the results of three different semantic segmentation algorithms (Fig. 9c–Fig. 9e) and the result obtained by IA platform (Fig. 9f) that uses features and features and hypothesis attributes for algorithm selection. In the experiment illustrated in Fig. 9 the input image is shown in Fig. 9a. The first algorithm selected generated the result shown in Fig. 9c. The obtained semantic segmentation was analyzed for shape, proximity, position and relative size contradiction [27] and a hypothesis solving the contradiction is generated.

Using this hypothesis a new algorithm (Fig. 9e) was selected and the two results of semantic segmentation are merged. The result is shown in Fig. 9f. Notice the replacement of the chair (red region) from the initial result without removing any part of the sofa (green region).

CONCLUSION

In this paper we described some theoretical properties of algorithm selection. In particular we discussed the importance of the proper evaluation and the importance of hypothesis in the algorithm selection. The results show that for algorithms that are context sensitive – and most of algorithms used in real world application are context sensitive – the iterative approach proposed in this paper improves the overall computer vision and image understanding. The high level information was demonstrated to be very important – using only the statistics on the class level segmentation accuracy the algorithm selection approach provides best results and outperforms all the used algorithms.

Several extensions to this work are planned. The statistical information obtained during testing is not precise enough and thus will be explored in combination of features from the contradiction regions for increased accuracy of algorithm selection. The features used so far in the algorithm selection also require accuracy improvement by finding richer features. Such features have been recently obtained by the use of convolutional neural networks and we plan to integrate them into the IA platform. Finally, the hypothesis used is a simple object label obtained from measured properties such as objects

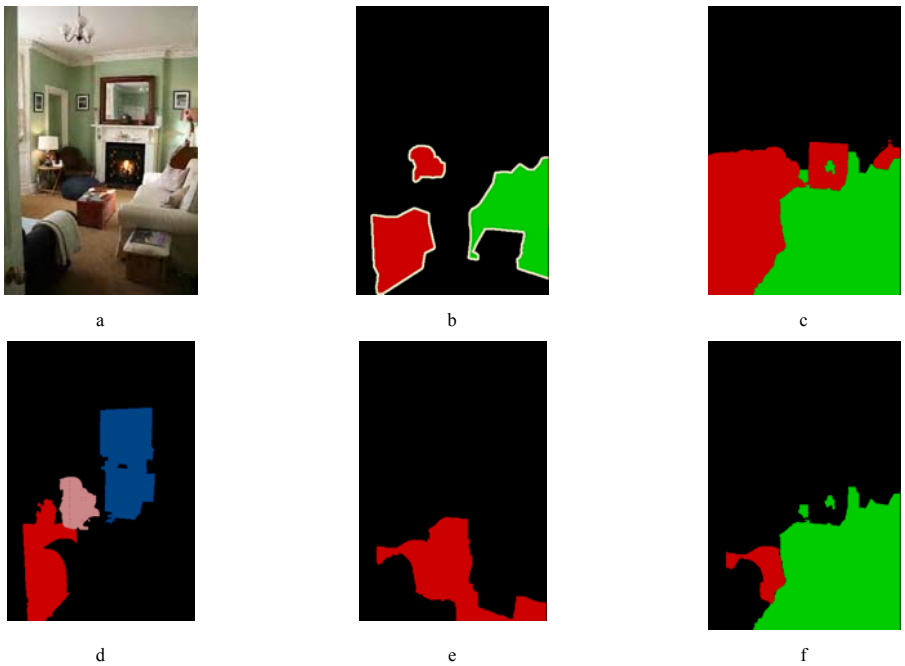


Figure 9 – An example of different stages of processing an input image using the algorithm selection platform: a – Input Image, b – Ground Truth, c – ALE Result, d – SDS Result, e – CPMC Result, f – IA Result

proximity, relative size and so on: to increase the accuracy of hypothesis generation a deeper semantic model connecting more object attributes and the objects with the environment of the world is to be build and used in close future.

REFERENCES

1. Rice J. The algorithm selection problem / J. Rice // Advances in Computers. – 1976. – Vol. 15. – P. 65–118.
2. Lukac M. Machine learning based adaptive contour detection using algorithm selection and image splitting / M. Lukac, R. Tanizawa, M. Kameyama // Interdisciplinary Information Sciences. – 2012. – Vol. 18, № 2. – P. 123–134.
3. Lukac M. Natural image understanding using algorithm selection and high level feedback / M. Lukac, M. Kameyama, K. Hiura // SPIE Intelligent Robots and Computer Vision XXX: algorithms and Techniques. – 2013. DOI: 10.1117/12.2008593
4. Zhang Y. Optimal selection of segmentation algorithms based on performance evaluation / Y. Zhang and H. Luo // Optical Engineering. – 2000. – Vol. 39, № 6. – P. 1450–1456.
5. Yong X. Optimal selection of image segmentation algorithms based on performance prediction / X. Yong, D. Feng, Z. Rongchun // Proceedings of the Pan-Sydney Area Workshop on Visual Information Processing (VIP2003). – 2003. – P. 105–108.
6. Yu L. Feature selection for high-dimensional data: A fast correlation-based filter solution / L. Yu, H. Liu // Proceedings of the 20th International Conference on Machine Learning. – 2004. – P. 856–863.
7. Takemoto S. Algorithm selection for intracellular image segmentation based on region similarity / S. Takemoto, H. Yokota // Ninth International Conference on Intelligent Systems Design and Applications. –2009. – P. 1413–1418. DOI: 10.1109/ISDA.2009.205
8. Lukac M. Bayesian-network-based algorithm selection with high level representation feedback for real-world intelligent systems / M. Lukac, and M. Kameyama // Information Technology in Industry. – 2015. – Vol. 3, № 1. – P. 10–15.
9. Peng B. Parameter selection for graph cut based image segmentation / B. Peng, V. Veksler // In Proceedings of the British Conference on Computer Vision. – 2008. – P. 16.1–16.10. DOI: 10.5244/C.22.16
10. Hoiem D. Closing the loop on scene interpretation / D. Hoiem, A. A. Efros, M. Hebert // Proc. Computer Vision and Pattern Recognition (CVPR). – 2008. – P. 1–8. DOI: 10.1109/CVPR.2008.4587587
11. Ferryman J. Automated scene understanding for airport aprons / [J. Ferryman, M. Borg, D. Thirde and other] // Proceedings of 18th Australian Joint Conference on Artificial Intelligence. – 2005. – P. 593–603. DOI: 10.1007/11589990_62
12. Oliva A. Modeling the shape of the scene: a holistic representation of the spatial envelope / A. Oliva, A. Torralba // International Journal of Computer Vision. – 2001. – Vol. 42, № 3. – P. 145–175.
13. Ladicky L. Graph cut based inference with co-occurrence statistics / L. Ladicky, C. Russell, P. Kohli, and P. Torr // In Proceedings of the 11th European conference on Computer vision. –2010. – P. 239–253. DOI: 10.1007/978-3-642-15555-0_18
14. Carreira J. Object recognition by sequential figure-ground ranking / J. Carreira, F. Li, C. Sminchisescu // International Journal of Computer Vision. –2012. – Vol. 98, № 3. –P. 243–262.
15. Leibe B. Robust object detection with interleaved categorization and segmentation / B. Leibe, A. Leonardis, B. Schiele // International Journal of Computer Vision. – 2008. – Vol. 77. – P. 259–289.
16. Finding animals: Semantic segmentation using regions and parts / [Arbelaez P., Hariharan B., Gu C. and other] // International Conference on Computer Vision and Pattern Recognition. – 2012. – P. 3378–3385. DOI: 10.1109/CVPR.2012.6248077
17. Li L.-J. Towards total scene understanding: classification, annotation and segmentation in an automatic framework / L.-J. Li, R. Socher, L. Fei-Fei // Computer Vision and Pattern Recognition (CVPR). –2009. – P. 2036–2043. DOI:10.1109/CVPR.2009.5206718
18. Ladicky L. Inference methods for crfs with co-occurrence statistics / L. Ladicky, C. Russell, P. Kohli, P. Torr // International Journal of Computer Vision. – 2013. – Vol. 103, № 2. – P. 213–225.
19. Hariharan B. Simultaneous detection and segmentation / B. Hariharan, P. Arbelaez, R. Girshick, J. Malik // European Conference on Computer Vision (ECCV). – 2014. – P. 297–312. DOI: 10.1007/978-3-319-10584-0_20
20. Martin M. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics / M. Martin, C. Fowlkes, D. Tal, J. Malik // Proceedings of 8th International Conference on Computer Vision. – 2001. – P. 416–423. DOI: 10.1109/ICCV.2001.937655
21. Arbelaez P. Contour detection and hierarchical image segmentation / P. Arbelaez, M. Maire, C. Fowlkes, J. Malik // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2011. – Vol. 33, № 5. – P. 898–916.
22. Arbelaez P. Boundary extraction in natural images using ultrametric contour maps / P. Arbelaez // Computer Vision and Pattern Recognition Workshop. – 2006. – P. 182–190. DOI: 10.1109/CVPRW.2006.5.
23. Ren X. Multi-scale improves boundary detection in natural images / X. Ren // Proceedings of the 10th European Conference on Computer Vision. –2008. – P. 533–545. DOI: 10.1007/978-3-540-88690-7_40
24. Dollar P. Supervised learning of edges and object boundaries / P. Dollar, Z. Tu, S. Belongie // IEEE Computer Vision and Pattern Recognition (CVPR). – 2006. – P. 1964–1971. DOI: 10.1109/CVPR.2006.298
25. Using contours to detect and localize junctions in natural images / [M. Maire, P. Arbelaez, C. Fowlkes, J. Malik] // Conference on Vision and Pattern Recognition. – 2008. – P. 1–8. DOI: 10.1109/CVPR.2008.4587339
26. The pascal visual object classes (VOC) challenge / [M. Everingham, L. Van Gool, C. K. I. Williams and other] // International Journal of Computer Vision. – 2010. – Vol. 88, № 2. – P. 303–338.
27. Lukac M. Bayesian-network-based algorithm selection with high level representation feedback for real-world information processing / M. Lukac M. Kameyama // IT in Industry. – 2015. – Vol. 3, № 1. – P. 10–15.
28. Handling missing values in support vector machine classifier / [K. Pelckmans, J. De Brabanter, J.A.K. Suykens and other] // Neural Networks. – 2005. – Vol. 18, № 5–6. – P. 684–692.

Article was submitted 22.09.2015.
After revision 06.10.2015.

Лукач М.[1], Абдиева К.[2], Камеяма М.[3]
[1]Д-р философии, ассистент кафедры компьютерных наук, Университет им. Назарбаева, Астана, Казахстан
[2]Аспирант, лаборатория ROSE, Наньянский технологический Университет, Сингапур
[3]Д-р наук, профессор, профессор школы информатики, Университе Тохоку, Сендай, Япония

ОЦЕНИВАНИЕ КОМПОНЕНТНЫХ АЛГОРИТМОВ ДЛЯ ВЫБОРА АЛГОРИТМА СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ НА ОСНОВЕ ОБРАТНОЙ СВЯЗИ С ВЫСОКИМ УРОВНЕМ ИНФОРМАЦИИ

Обсуждаются некоторые теоретические свойства подхода по выбору алгоритма для решения проблемы семантической сегментации в компьютерном зрении. Высококачественный выбор алгоритма возможен, только если пригодность каждого алгоритма хорошо известна, потому что только тогда результат выбора алгоритма может улучшить наилучший возможный результат, полученный одним алгоритмом. Показано, что оценка алгоритма зависит от конечной задачи; т.е. для того чтобы правильно оценивать алгоритм и определить его пригодность, необходимо использовать только хорошо сформулированные задачи. Когда пригодность алгоритма известна алгоритм может быть эффективно использован для задачи, применяясь в наиболее благоприятных условиях, определяемых в ходе оценивания. Оценивание, зависящее от задачи, продемонстрировано на сегментации и распознавании объектов. Кроме того, обсуждается важность символического знания высокого уровня в процессе отбора. Важность этой символической гипотезы продемонстрировано на наборе экспериментов по по обучению байесовской сети и SVM, а также с помощью статистических данных, полученные во время обучения селектора алгоритма. Показано, что для выбора эффективного алгоритма требуется оценивание, зависящее от

задачи. Показано, что используя символические предпочтения алгоритмов, точность выбора алгоритма может быть улучшена на 10–15%, а качество символической сегментации может быть улучшено до 5% по сравнению с наилучшим доступным алгоритмом.

Ключевые слова: выбор алгоритма, пригодность алгоритма, компьютерное зрение.

Лукач М.[1], Абдієв К.[2], Камеяма М.[3]
[1]Д-р філософії, асистент кафедри комп'ютерних наук, Університет ім. Назарбаєва, Астана, Казахстан
[2]Аспірант, лабораторія ROSE, Наньянський технологічний Університет, Сінгапур
[3]Д-р наук, професор, професор школи інформатики Університам Тохоку, Сендай, Японія

ОЦІНЮВАННЯ КОМПОНЕНТНИЙ АЛГОРИТМІВ ДЛЯ ВИБОРУ АЛГОРИТМА СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ НА ОСНОВІ ЗВОРОТНОГО ЗВ'ЯЗКУ З ВИСОКИМ РІВНЕМ ІНФОРМАЦІЇ

Показано, що оцінка алгоритму залежить від кінцевого завдання; тобто для того щоб правильно оцінювати алгоритм і визначити його придатність, необхідно використовувати тільки добре сформульовані завдання. Коли придатність алгоритму відома, алгоритм може бути ефективно використаний для завдання, застосовуючись у найбільш сприятливих умовах, обумовлених у ході оцінювання. Оцінювання, залежне від завдання, продемонстровано на сегментації і розпізнаванні об'єктів. Крім того, обговорюється важливість символічного знання високого рівня у процесі відбору. Важливість цієї символічної гіпотези продемонстровано на наборі експериментів з навчання байєсівської мережі та SVM, а також за допомогою статистичних даних, отриманих під час навчання селектора алгоритму. Показано, що для вибору ефективного алгоритму потрібно оцінювання, залежне від завдання. Показано, що використовуючи символічні переваги алгоритмів, точність вибору алгоритму може бути поліпшена на 10–15%, а якість символічної сегментації може бути покращена до 5% у порівнянні з найкращим доступним алгоритмом.

Ключові слова: вибір алгоритму, придатність алгоритму, комп'ютерний зір.

REFERENCES

1. Rice J. The algorithm selection problem, *Advances in Computers*, 1976, Vol. 15, pp. 65–118.
2. Lukac M., Tanizawa R., Kameyama M. Machine learning based adaptive contour detection using algorithm selection and image splitting, *Interdisciplinary Information Sciences*, 2012, Vol. 18, No. 2, pp. 123–134.
3. Lukac M., Kameyama M., Hiura K. Natural image understanding using algorithm selection and high level feedback, *SPIE Intelligent Robots and Computer Vision XXX: algorithms and Techniques*, 2013. DOI: 10.1117/12.2008593
4. Zhang Y., Luo H. Optimal selection of segmentation algorithms based on performance evaluation, *Optical Engineering*, 2000, Vol. 39, No. 6, pp. 1450–1456.
5. Yong X., Feng D., Rongchun Z. Optimal selection of image segmentation algorithms based on performance prediction, *Proceedings of the Pan-Sydney Area Workshop on Visual Information Processing (VIP2003)*, 2003, pp. 105–108.
6. Yu L., Liu H. Feature selection for high-dimensional data: A fast correlation-based filter solution, *Proceedings of the 20th International Conference on Machine Learning*, 2004, pp. 856–863.
7. Takemoto S., Yokota H. Algorithm selection for intracellular image segmentation based on region similarity, *Ninth International Conference on Intelligent Systems Design and Applications*. 2009, pp. 1413–1418. DOI: 10.1109/ISDA.2009.205
8. Lukac M., Kameyama M. Bayesian-network-based algorithm selection with high level representation feedback for real-world intelligent systems, *Information Technology in Industry,* 2015, Vol. 3, No. 1, pp. 10–15.
9. Peng B., Veksler V. Parameter selection for graph cut based image segmentation, *In Proceedings of the British Conference on Computer Vision*, 2008, pp. 16.1–16.10. DOI: 10.5244/C.22.16
10. Hoiem D., Efros A. A., Hebert M. Closing the loop on scene interpretation / D. Hoiem, // Proc. Computer Vision and Pattern Recognition (CVPR), 2008, P. 1–8. DOI: 10.1109/CVPR.2008.4587587
11. Ferryman J. Borg M., Thirde D., Fusier F., Valentin V., Bremond F., Thonnat M., Aguilera J., Kampel M. Automated scene understanding for airport aprons, Proceedings of 18th Australian Joint Conference on Artificial Intelligence, 2005, pp. 593–603. DOI: 10.1007/11589990_62
12. Oliva A., Torralba A. Modeling the shape of the scene: a holistic representation of the spatial envelope, *International Journal of Computer Vision*, 2001, Vol. 42, No. 3, pp. 145–175.
13. Ladicky L., Russell C., Kohli P., and Torr P. Graph cut based inference with co-occurrence statistics, *In Proceedings of the 11th European conference on Computer vision*, 2010, pp. 239–253. DOI: 10.1007/978-3-642-15555-0_18
14. Carreira J. Li F., Sminchisescu C. Object recognition by sequential figure-ground ranking, *International Journal of Computer Vision*, 2012, Vol. 98, No. 3, pp. 243–262.
15. Leibe B. Leonardis A., Schiele B. Robust object detection with interleaved categorization and segmentation, *International Journal of Computer Vision,* 2008, Vol. 77, pp. 259–289.
16. Arbelaez P. Hariharan B., Gu C., Gupta S., Bourdev L., Malik J. Finding animals: Semantic segmentation using regions and parts, *International Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3378–3385. DOI: 10.1109/CVPR.2012.6248077
17. Li L.-J., Socher R., Fei-Fei L. Towards total scene understanding: classification, annotation and segmentation in an automatic framework, *Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 2036–2043. DOI:10.1109/CVPR.2009.5206718
18. Ladicky L., Russell C., Kohli P., Torr P. Inference methods for crfs with co-occurrence statistics, *International Journal of Computer Vision*, 2013, Vol. 103, No. 2, pp. 213–225.
19. Hariharan B., Arbelaez P., Girshick R., Malik J. Simultaneous detection and segmentation, *European Conference on Computer Vision (ECCV)*, 2014, pp. 297–312. DOI: 10.1007/978-3-319-10584-0_20
20. Martin M., Fowlkes C., Tal D., Malik J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, *Proceedings of 8th International Conference on Computer Vision*, 2001, P. 416–423. DOI: 10.1109/ICCV.2001.937655
21. Arbelaez P., Maire M., Fowlkes C., Malik J. Contour detection and hierarchical image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011,Vol. 33, No. 5, pp. 898–916.
22. Arbelaez P. Boundary extraction in natural images using ultrametric contour maps, *Computer Vision and Pattern Recognition Workshop*, 2006, pp. 182–190. DOI: 10.1109/CVPRW.2006.5.
23. Ren X. Multi-scale improves boundary detection in natural images, *Proceedings of the 10th European Conference on Computer Vision*, 2008, pp. 533–545. DOI: 10.1007/978-3-540-88690-7_40
24. Dollar P., Z. Tu, S. Belongie Supervised learning of edges and object boundaries, IEEE Computer Vision and Pattern Recognition (CVPR), 2006, pp. 1964–1971. DOI: 10.1109/CVPR.2006.298
25. Maire M. Arbelaez P., Fowlkes C., Malik J. Using contours to detect and localize junctions in natural images, *Conference on Vision and Pattern Recognition*, 2008, P. 1–8. DOI: 10.1109/CVPR.2008.4587339
26. Everingham M., Van Gool L., Williams C. K. I., Winn J., Zisserman A. The pascal visual object classes (VOC) challenge, *International Journal of Computer Vision*, 2010, Vol. 88, No. 2, pp. 303–338.
27. Lukac M., Kameyama M. Bayesian-network-based algorithm selection with high level representation feedback for real-world information processing, *IT in Industry*, 2015, Vol. 3, No. 1, pp. 10–15.
28. Pelckmans K., De Brabanter J., Suykens J.A.K., De Moor B. Handling missing values in support vector machine classifier, *Neural Networks*, 2005, Vol. 18, No. 5–6, pp. 684–692.

p-ISSN 1607-3274. Радіоелектроніка, інформатика, управління. 2016. № 1
e-ISSN 2313-688X. Radio Electronics, Computer Science, Control. 2016. № 1

УПРАВЛІННЯ У ТЕХНІЧНИХ СИСТЕМАХ

УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ СИСТЕМАХ

CONTROL IN TECHNICAL SYSTEMS

УДК 629.7

Хацько Н. Е.

Канд. техн. наук, старший преподаватель кафедры систем и процессов управления Национального технического университета «Харьковский политехнический институт», Харьков, Украина

РЕШЕНИЕ ТЕРМИНАЛЬНОЙ ЗАДАЧИ УПРАВЛЕНИЯ С ИСПОЛЬЗОВАНИЕМ ИНФОРМАЦИИ ИНЕРЦИАЛЬНОГО БЛОКА

Рассматривается задача терминального управления летательным аппаратом по информации бесплатформенной инерциальной навигационной системы, в которой за заданное время требуется перевести аппарат в заданное пространственное положение с требуемым значением вектора конечной скорости. Построено программное управление для невозмущенного движения летательного аппарата. Синтезировано управление летательного аппарата по текущему значению его вектора состояния, определенного инерциальным навигационным блоком. Получены выражения для оценки точности попадания в конечную точку с учетом ошибки акселерометра. Проанализирована точность выведения летательного аппарата в конечную точку замкнутой системой управления при использовании информации от инерциальных датчиков с известным уровнем ошибок.

На примере упрощенной модели проведены компьютерное моделирование движения летательного аппарата и анализ точности приведения вектора состояния в заданную конечную точку в зависимости от уровня внешних возмущений и погрешности входной информации. Проведены эксперименты, позволяющие выявить влияние численного значения глубины прогноза, использованного в синтезе управления по методу преследования ведущей точки, на точность решения терминальной задачи. Сформулированы рекомендации для проектирования алгоритмов автоматического управления движением в комплексе с проектированием информационно-измерительной системы.

Ключевые слова: система управления, бесплатформенная инерциальная навигационная система, ошибка измерения вектора состояния, динамическое возмущение, точность терминального управления.

НОМЕНКЛАТУРА

БИНС – бесплатформенная инерциальная навигационная система;

ИКАО – (от англ. International Civil Aviation Organization) – международная организация гражданской авиации;

ИИСНС – интегрированая инерциально-спутниковая навигационная система;

ЛА – летательный аппарат;

МЭМС – микро-электро-механическая система;

САУП – система автоматического управления;

τ – глубина прогноза, интервал времени;

$a(t)$ – ускорение ЛА;

δa – постоянная погрешность ускорения;

b_i – коэффициенты полинома;

$f(t)$ – возмущающее воздействие;

f^* – максимальная величина воздействия;

$r(t)$ – координата объекта;

$r^*(t)$ – программная (эталонная) траектория;

$\delta r(t)$ – координатная ошибка;

r_0 – координата объекта в начальный момент времени;

r_T – координата объекта в конечный момент времени;

$\dot{r}_0$ – скорость объекта в начальный момент времени;

$\dot{r}_T$ – скорость объекта в конечный момент времени;

T –общее время движения по траектории;

t –текущее время;

t' – переменная локального времени;

$u(t, r, \dot{r})$ – закон управления ЛА;

$\delta u(t)$ – ошибка управления.

ВВЕДЕНИЕ

Неотъемлемым элементом систем автоматического управления полетом (САУП) современных летательных аппаратов (ЛА) является бортовая навигационная система, обеспечивающая получение информации о местоположении, скорости и ориентации аппарата в про-

DOI 10.15588/1607-3274-2016-1-12

странстве. Учитывая зональные требования к точности самолетовождения, регламентированные нормативными документами ИКАО [1], бортовое оборудование должно обеспечивать высокоточную навигацию, особенно на критических участках движения, таких, как взлет, посадка, маневры в горизонтальной и вертикальной плоскости.

Эффективным способом достижения требуемой точности самолетовождения является использование так называемых интегрированных инерциально-спутниковых навигационных систем (ИИСНС), получивших широкое распространение благодаря успехам спутниковой радионавигации [2]. Принцип работы таких систем основан на комплексировании информации, получаемой от инерциальных датчиков (гироскопов и акселерометров), и спутниковой информации о текущих координатах и скорости ЛА. При этом регулярное поступление и использование достоверных спутниковых измерений позволяет сколь угодно долго удерживать ошибки определения навигационных параметров практически в окрестности нуля. Однако, как показано в [3], даже при полностью развернутых спутниковых орбитальных группировках в течение суток в некоторых областях Земного шара могут встречаться интервалы времени, более 15 минут, когда спутниковая информация вследствие особого взаимного расположения спутников в пространстве (геометрического фактора) становится не пригодной для высокоточной навигации. В этих условиях пилотирование осуществляется по информации исключительно инерциальной подсистемы, роль которой исполняет бортовая бесплатформенная инерциальная навигационная система (БИНС).

Вышесказанное определяет актуальность и место рассматриваемой в данной статье проблемы в контексте современной технологии управления полетом ЛА.

1 ПОСТАНОВКА ЗАДАЧИ

Рассматривается задача управления ЛА по информации БИНС, а именно – терминальная задача, в которой за заданное время требуется перевести аппарат в заданное пространственное положение с требуемым значением вектора конечной скорости.

Полагаем, что в начальный момент времени положение и скорость ЛА известны. Необходимо с помощью управляющих воздействий за заданное время перевести ЛА в требуемую точку пространства с требуемой конечной скоростью. При этом информация о текущем значении вектора состояния ЛА доступна с некоторой ошибкой, соответствующей принятой модели ошибок инерциального датчика.

Для удобства изложения принципиального хода решения задачи и некоторых аналитических результатов рассмотрим одномерный случай, для которого невозмущенное управляемое движение ЛА описывается уравнением

$$\ddot{r}(t) = u(t), \qquad (1)$$

где $r(t)$ – координата объекта, $u(t)$ – искомое управляющее ускорение. В начальный ($t = 0$) и конечный ($t = T$) моменты времени состояние управляемого объекта характеризуется параметрами

$$r(0) = r_0,\ \dot{r}(0) = \dot{r}_0\ , \qquad (2)$$

$$r(T) = r_T,\ \dot{r}(T) = \dot{r}_T\ . \qquad (3)$$

Время движения по траектории определено величиной T.

С учетом получаемой информации о векторе состояния необходимо синтезировать закон управления $u(t, r, \dot{r})$, обеспечивающий в силу (1), (2) выполнение условий (3).

2 ОБЗОР ЛИТЕРАТУРЫ

На практике часто возникают задачи выполнения полета ЛА в заданную точку пространства с заданными параметрами движения за отведенное время. Это терминальные задачи, решению которых посвящены работы [4–7]. Понятие терминального управления формулируется как управление, цель которого заключается в переводе объекта управления в заданное конечное состояние в заданный момент времени [8]. Наилучшим решением терминальной задачи в алгоритмическом смысле является получение явных аналитических зависимостей для управляющих функций [5]. Современная техника управления базируется на теории обратной связи и анализе линейных систем [9]. Контроль текущих параметров состояния ЛА и формирование управления, корректирующего полет, повышает эффективность работы САУП. Обычно процесс управления включает следующие основные этапы:

– выработку эталонной траектории движения ЛА к заданной цели, то есть получение требуемого состояния ЛА через заданный промежуток времени в идеальных условиях полета;

– получение информации о текущем состоянии ЛА на шаге работы САУП;

– анализ полученной информации и выработку управляющих воздействий для продолжения движения ЛА.

Система управления является соединением отдельных элементов в определенную конфигурацию, имеющую причинно-следственные связи между элементами и обеспечивающую заданные характеристики [10], поэтому точность системы управления зависит от точности работы всех элементов системы. Навигационный модуль следует рассматривать как элемент системы управления.

Основным инструментом для получения информации о текущем векторе состояния ЛА чаще всего является ИИСНС, объединяющая измерители различной физической природы – спутниковую и инерциальную навигационную систему – в единый интегрированный навигационный комплекс. Для повышения точности определения навигационных и угловых параметров ориентации в ИИСНС используют различные способы комплексирования информации: раздельные, слабо связные, жестко связные и глубоко интегрированные [11]. Но, независимо от степени интеграции данных, при задержке получения достоверной информации от бортовой спутниковой навигационной аппаратуры потребителя, БИНС работает в автономном режиме. В этом случае только полученные от нее данные могут быть использованы для формирования оптимального управляющего воздействия на ЛА.

Следует отметить, что в последнее время существует устойчивая тенденция к удешевлению инерциальных систем. Уже 20 лет стремительно развивается технология

микро-электро-механических систем (МЭМС). Если ранее точность датчиков, построенных с использованием этой технологии была низкой, и их можно было использовать только в автомобилях и бытовых приборах, то в последнее время точность МЭМС-акселерометров повысилась до навигационной и точность МЭМС-гироскопов также находится на высоком уровне [12]. Ведущие мировые компании уже наладили выпуск навигационных модулей с МЭМС-датчиками для малой и средней авиации.

Известно [13], что в ходе функционирования БИНС накапливает ошибки оценивания навигационных параметров, вызванные ошибками измерений инерциальных датчиков. Связь ошибок БИНС с ошибками датчиков чрезвычайно сложная, т.к. характер накопления ошибок БИНС существенно зависит также и от маневров, совершаемых ЛА, в процессе функционирования БИНС. Поэтому исследовать в общем виде влияние ошибок БИНС в целом и инерциальных датчиков, в частности, на точность управления пространственным движением ЛА удается только путем численного моделирования. В то же время, для выяснения некоторых особенностей взаимодействия задач управления и идентификации вектора состояния ЛА представляется необходимым провести аналитические исследования на упрощенной модели управляемого процесса.

В данной статье проводится исследование точности достижения конечных условий в задаче терминального управления движением объекта, решаемой с помощью замкнутой САУП по информации инерциальных датчиков.

3 МАТЕРИАЛЫ И МЕТОДЫ

В основу решения задачи терминального управления полагается некоторая программно заданная траектория, соединяющая краевые точки, для которой согласно концепции обратных задач динамики [5] определяется программное управление $u^*(t)$.

Зададим программную траекторию, как экстремаль функционала $\int_0^T \ddot{r}^2(t)\cdot dt = \int_0^T u^2(t)\cdot dt$ [14]. В этом случае

$$r^*(t) = b_3\cdot t^3 + b_2\cdot t^2 + b_1\cdot t + b_0, \qquad (4)$$

а коэффициенты b_i, $i=\overline{0,3}$ определяются из условия согласования решения (4) с краевыми условиями (2), (3):

$$b_0 = r_0,$$
$$b_1 = \dot{r}_0,$$
$$b_2 = \frac{3(r_T - r_0)}{T^2} - \frac{\dot{r}_T + 2\dot{r}_0}{T},$$
$$b_3 = \frac{2(r_0 - r_T)}{T^3} + \frac{\dot{r}_T + \dot{r}_0}{T^2}. \qquad (5)$$

Таким образом, располагая уравнением программной траектории (4), из (1) находим программное управление в виде:

$$u^*(t) = 6b_3t + 2b_2$$

или, после подстановки (5):

$$u^*(t) = \left(12\frac{r_0 - r_T}{T^3} + 6\frac{\dot{r}_T + \dot{r}_0}{T^2}\right)\cdot t + 6\frac{r_T - r_0}{T^2} - 2\frac{\dot{r}_T + 2\dot{r}_0}{T}. \qquad (6)$$

В дальнейшем, учитывая предполагаемый идеальный характер движения, программное управление (6) и порождаемую им программную траекторию (4) будем называть эталонными.

Найденное программное управление соответствует разомкнутой схеме управления и, по понятным причинам, не может использоваться на практике. Для осуществления движения вдоль программной траектории в условиях возможных параметрических и динамических возмущений предлагается воспользоваться алгоритмом синтеза по методу преследования ведущей точки [5].

Будем считать, что в каждый момент времени доступны измерения координаты и скорости $\hat{r}(t)$, $\dot{\hat{r}}(t)$. В этих условиях решается задача перевода за некоторое время τ, в дальнейшем называемое глубиной прогноза, вектора состояния $(r,\dot{r})$ из точки $(\hat{r}(t),\dot{\hat{r}}(t))$ в точку $(r^*(t+\tau),\dot{r}^*(t+\tau))$. Как видно из постановки задачи, целевая точка выбирается на эталонной траектории (4), опережает текущую точку по времени на τ и к моменту времени T, согласно процедуре построения траектории (4), приходит в требуемое конечное положение, обеспечивая решение исходной задачи терминального управления.

Итак, дополнительно учитывая ограниченное по модулю возмущение $|f|\le f^*$, математическая модель задачи синтеза имеет вид:

$$\ddot{r}(t') = u(t') + f(t'),$$

где $t'\in[0,\tau]$ – переменная локального времени, с краевыми условиями:

$$r(0) = \hat{r}(t), \qquad \dot{r}(0) = \dot{\hat{r}}(t),$$
$$r(\tau) = r^*(t+\tau), \quad \dot{r}(\tau) = \dot{r}^*(t+\tau). \qquad (7)$$

Соединяя начальную и конечную точки (7) кубическим полиномом от переменной локального времени t', получим

$$r(t') = a_3t'^3 + a_2t'^2 + a_1t' + a_0,$$

где

$$a_0 = \hat{r}(t), \quad a_1 = \dot{\hat{r}}(t),$$
$$a_2 = 3\frac{r^*(t+\tau) - \hat{r}(t)}{\tau^2} - \frac{\dot{r}^*(t+\tau) + 2\dot{\hat{r}}(t)}{\tau}$$
$$a_3 = 2\frac{\hat{r}(t) - r^*(t+\tau)}{\tau^3} + \frac{\dot{r}^*(t+\tau) + \dot{\hat{r}}(t)}{\tau^2},$$

для управления получим:

$$u(t') = 6a_3t' + 2a_2. \qquad (8)$$

Ограничивая временной интервал использования управления (8) моментом локального времени $t'=0$, в качестве закона синтеза окончательно получим:

$$u(t,\hat{r},\hat{\dot{r}})=6\cdot\frac{r^*(t+\tau)-\hat{r}(t)}{\tau^2}-2\cdot\frac{\dot{r}^*(t+\tau)+2\cdot\hat{\dot{r}}(t)}{\tau}, \quad (9)$$

в котором использованы как фактические текущие значения оценки вектора состояния $r(t),\dot{r}(t)$, так и его прогнозируемые эталонные значение $r^*(t+\tau),\dot{r}^*(t+\tau)$, представляющие собой известную функцию времени.

Для анализа эффективности выбранной схемы синтеза исследуем замкнутую систему управления с точки зрения возможности реализации эталонной траектории.

Будем считать, что фактическое значение переменной $r(t)$, соответствующее решению возмущенного уравнения движения, отличается от эталонного значения $r^*(t)$, задаваемого (4), на величину $\delta r(t)=r(t)-r^*(t)$. Представим закон синтеза управления (9) в виде:

$$u(t,r,\dot{r})=u^*(t)+\delta u(\delta r,\delta\dot{r}),$$

где второе слагаемое выражается как

$$\delta u(t)=-\frac{6}{\tau^2}\delta r(t)-\frac{4}{\tau}\delta\dot{r}(t). \quad (10)$$

В этих условиях уравнение для ошибки

$$\delta\ddot{r}(t)=\delta u+f(t)$$

с учетом (10) приводится к виду

$$\delta\ddot{r}(t)+\frac{4}{\tau}\delta\dot{r}(t)+\frac{6}{\tau^2}\delta r(t)=f(t) \quad (11)$$

и имеет решение

$$\delta r(t)=e^{\frac{-2}{\tau}t}\left(C_1\cos\frac{\sqrt{2}}{\tau}t+C_2\sin\frac{\sqrt{2}}{\tau}t\right)+\delta r^f(t), \quad (12)$$

в котором коэффициенты C_1, C_2 определяются из начальных условий $\delta r(0)$, $\delta\dot{r}(0)$, а вынужденная составляющая

$$\delta r^f(t)=\frac{\tau}{\sqrt{2}}\int_0^t e^{\frac{-2}{\tau}\theta}\sin\left(\frac{\sqrt{2}}{\tau}\cdot\theta\right)\cdot f(t-\theta)\cdot d\theta \quad (13)$$

соответствует свертке импульсной переходной функции инерциального звена второго порядка, каковым является (11), и функции $f(t)$.

Для получения грубой оценки $\delta r^f(t)$ при $t\to\infty$ без конкретизации функциональной зависимости $f(t)$, воспользуемся соотношениями $|f(t)|\le f^*$ и $\left|\sin\frac{\sqrt{2}}{\tau}\cdot\theta\right|\le 1$, где f^* – максимальная величина воздействия. Откуда следует, что для любого $t\ge 0$ имеет место неравенство

$$|\delta r^f(t)|<\frac{\tau}{\sqrt{2}}\int_0^t e^{\frac{-2}{\tau}\theta}\cdot f^*\cdot d\theta=\frac{\tau^2}{2\sqrt{2}}f^*\left(1-e^{\frac{-2}{\tau}\cdot t}\right)\underset{t\to\infty}{\to}\frac{\tau^2}{2\sqrt{2}}f^*. \quad (14)$$

Получить для (13) более точную оценку, соответствующую нестрогому неравенству, затруднительно, поэтому закономерен вопрос о достижимости, или о степени «грубости» полученной верхней оценки для $|\delta r^f(t)|$. Легко видеть, что в частном случае, когда $f(t)=f^*=Const$, вынужденная составляющая принимает вид $\delta r^f(t)=\frac{\tau^2}{6}f^*$ и отличается от оценки из (14) менее чем в два раза. Этот факт говорит о том, что использованные упрощения при получении оценки $|\delta r^f(t)|$ не приводят к существенному искажению результата.

Таким образом, из (12) следует, что при идеальных измерениях вектора состояния управление (9) при $f\equiv 0$ и не нулевых начальных условиях обеспечивает асимптотическую устойчивость движения относительно эталонной траектории, а при не нулевом, но ограниченном возмущении $|f(t)|\le f^*$, точка гарантированно не отклоняется от эталонной траектории более чем на величину $\frac{\tau^2}{2\sqrt{2}}f^*$.

Перейдем к анализу влияния ошибки измерения вектора состояния на точность работы замкнутой системы управления и, в частности, на терминальную точность решения задачи управления.

Пусть оценка вектора состояния $(r,\dot{r})$ осуществляется с ошибкой. Будем считать, что исходным измерением является величина $a(t)=\ddot{r}(t)+\delta a$, соответствующая сумме истинного ускорения и постоянной погрешности δa, а оценка вектора состояния $(\hat{r},\hat{\dot{r}})$ производится, согласно принципу инерциального счисления, путем двойного интегрирования измерений. Рассмотренный случай характерен для работы БИНС, в которой погрешность δa складывается из погрешности собственно акселерометра и ошибки компенсации вектора ускорения свободного падения, обусловленной погрешностью гироскопов. В некоторых режимах движения ЛА такую совокупную ошибку можно считать постоянной.

Итак, с учетом $\ddot{\hat{r}}(t)=a(t)$ для оценки $\hat{r}$, $\hat{\dot{r}}$, совпадающей в начальный момент времени с фактическими значениями одноименных переменных, имеем:

$$\hat{\dot{r}}(t)=\dot{r}(t)+\delta a\cdot t,$$

$$\hat{r}(t)=r(t)+\delta a\cdot\frac{t^2}{2}.$$

Таким образом, после подстановки (10), (11) в закон управления (9), для замкнутой системы уравнение ошибки реализации эталонного движения принимает вид:

$$\delta\ddot{r}(t)+\frac{4}{\tau}\delta\dot{r}(t)+\frac{6}{\tau^2}\delta r(t)=-\frac{4\delta a}{\tau}\cdot t-\frac{3\delta a}{\tau^2}\cdot t^2+f(t). \quad (15)$$

Полагая, что в начальный момент времени $\delta r(0)=0, \delta\dot{r}(0)=0$, уравнение (15) имеет решение

$$\delta r(t)=-\frac{\delta a}{2}t^2+\frac{\delta a}{6}\tau^2-e^{-\frac{2}{\tau}t}\frac{\delta a}{3}\tau^2\left(\frac{1}{2}\cos\frac{\sqrt{2}}{\tau}t+\frac{1}{\sqrt{2}}\sin\frac{\sqrt{2}}{\tau}t\right)+\delta r^f(t). \quad (16)$$

В отличие от (12), где $\delta r(t)$ является ограниченной по модулю величиной, в (16) абсолютное значение $\delta r(t)$ неограниченно растет с увеличением времени t, что объясняется накапливаемым характером ошибки идентификации вектора состояния в БИНС.

Аналогичный вывод справедлив и для ошибки терминальной скорости:

$$\delta\dot{r}_T^{\delta a}(\tau)=\frac{\delta a}{\sqrt{2}}\tau\cdot e^{-\frac{2}{\tau}T}\cdot\sin\frac{\sqrt{2}}{\tau}\cdot T-\delta a\cdot T. \quad (17)$$

4 ЭКСПЕРИМЕНТЫ

Был проведен вычислительный эксперимент с использованием разработанного программного комплекса, имитирующего работу САУП [15].

Для проверки правильности аналитических зависимостей (16), (17) выбрано устойчивое прямолинейное движение, при котором изменяется только одна координата. В ходе эксперимента интегрировались как параметры реального движения, так и оценки алгоритма БИНС.

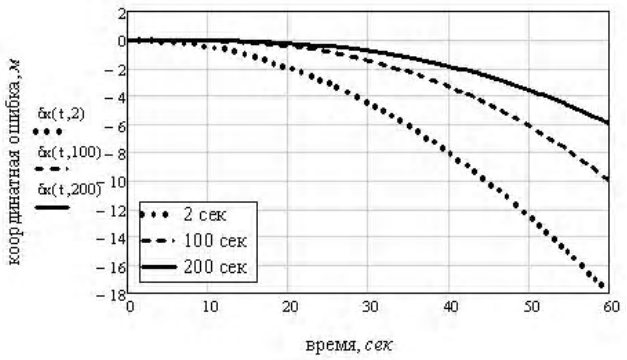


Рисунок 1 – Координатная ошибка при использовании закона управления с различными значениями

$$\tau\in\{2\,\text{сек}, 100\,\text{сек}, 200\,\text{сек}\}$$

5 РЕЗУЛЬТАТЫ

Эксперимент на малом интервале времени равном 100 сек., в котором зависимость $\delta r(t)$ при $f(t)\equiv 0$ и заданном уровне ошибок измерения акселерометра δa получена для различных значений глубины прогноза τ. Результаты эксперимента демонстрирует рис. 1.

Второй эксперимент проводился с целью демонстрации возможности нивелирования влияния ошибок измерений путем выбора параметра глубины прогноза τ на фиксированном интервале движения T. Поведение зависимости (16) как функции параметра τ показано на рис. 2.

6 ОБСУЖДЕНИЯ

Поведение координатной ошибки во время движения в терминальную точку (рис. 1) подтверждает вывод о том, что ошибка управления (16) неограниченно возрастает с увеличением времени t из-за накопления ошибки определения вектора состояния БИНС, причем, зависимость ошибки управления от времени t близка к квадратичной.

Приведенные на рис. 1 результаты примечательны также тем, что отражают зависимость скорости нарастания ошибки управления от глубины прогноза τ, являющейся свободным параметром закона управления (9). Из уравнения (15) видно, что чем больше τ, тем меньше при прочих равных условиях влияние δa на динамику δr. Так, при $\delta a=0{,}08\,\text{м/с}^2$ и терминальной продолжительности $T=60$ с, модуль терминальной ошибки $|\delta r(T)|$ для параметра τ, равного 2 с, 100 с и 200 с, достигает значений 18 м, 10 м и 6 м соответственно (рис. 1). Указанная закономерность сохраняется и при других значениях T.

Во втором эксперименте рассматривалась координатная ошибка $\delta r(t)$ при фиксированном времени движения ($T=\text{const}$), что позволило представить (16) в виде:

$$\delta r_T^{\delta a}(\tau)=-\frac{\delta a}{2}T^2+\frac{\delta a}{6}\tau^2-e^{-\frac{2}{\tau}T}\frac{\delta a}{3}\tau^2\left(\frac{1}{2}\cos\frac{\sqrt{2}}{\tau}T+\frac{1}{\sqrt{2}}\sin\frac{\sqrt{2}}{\tau}T\right). \quad (18)$$

Для двух вариантов терминальной продолжительности T=100 сек и T=200 сек параметр τ менялся от 0,01 сек до 10^4 сек при отсутствии динамического возмущения

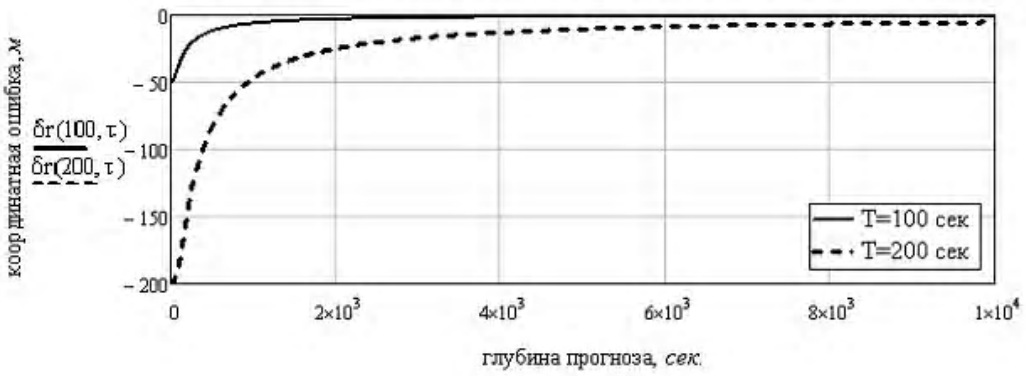


Рисунок 2 – Зависимость координатной ошибки в терминальной точке от глубины прогноза

($f(t) \equiv 0$) и ненулевых ошибках измерения δa. Обе кривые на рис. 2 стремятся к оси абсцисс, что показывает монотонное уменьшение терминальной ошибки при увеличении τ. При этом, анализируя (18), несложно показать, что для любого конечного T имеет место $\lim_{\tau \to \infty} (\delta r_T(\tau)) = 0$. Аналогичный вывод справедлив и для ошибки терминальной скорости:

$$\delta \dot{r}_T^{\delta a}(\tau) = \frac{\delta a}{\sqrt{2}} \tau \cdot e^{-\frac{2}{\tau} T} \cdot \sin \frac{\sqrt{2}}{\tau} \cdot T - \delta a \cdot T.$$

Таким образом, при пренебрежимо малых динамических возмущениях (f^* – мало) параметр τ, входящий в закон управления (9), целесообразно выбирать большим. Такой выбор позволяет уменьшить влияние ошибки измерений ускорения на терминальную точность управления. Однако данный вывод не может быть окончательным, поскольку в условиях ненулевых динамических возмущений увеличение параметра τ, как следует из (13), приводит к увеличению ошибки реализации эталонной траектории и, как следствие, к уменьшению точности выполнения терминальных условий.

ВЫВОДЫ

Проведено исследование точности функционирования замкнутой системы управления с учетом динамического возмущения и погрешности измерений параметров движения инерциальной навигационной системой. Решена задача терминального управления. Закон управления включает в себя программную составляющую управления, соответствующую невозмущенному движению вдоль эталонной траектории, и поправку, осуществляющую синтез по методу преследования ведущей точки. Синтез управления осуществляется по информации, полученной путем интегрирования измерений фактического ускорения и содержащей ошибку, характерную для акселерометров. В результате анализа терминальной точности функционирования замкнутой системы получены следующие выводы-рекомендации:

– при увеличении постоянной времени передаточной функции замкнутой системы для ошибки влияние ошибок измерений уменьшается;

– для уменьшения влияния динамического возмущения на терминальную точность следует уменьшать коэффициент усиления передаточной функции замкнутой системы для ошибки;

– при выборе варьируемого параметра в законе синтеза по методу преследования ведущей точки следует руководствоваться следующими соображениями: если информационно-измерительная система обеспечивает получение координат и скорости объекта с высокой точностью, то выбираемый параметр прогноза должен быть минимальным. При использовании датчиков низкой точности параметр прогноза следует увеличить. Окончательное решение проблемы настройки алгоритма синтеза проводить с использованием моделирования типовых режимов работы.

В данной работе сделана попытка проанализировать требования к замкнутой системе управления с позиций влияния на точность как динамического возмущения, так и ошибок измерения вектора состояния. Показано, что влияние упомянутых факторов прямо противоположно. Сформулированные рекомендации могут быть полезны для проектирования алгоритмов автоматического управления движением в комплексе с проектированием информационно-измерительной системы.

СПИСОК ЛИТЕРАТУРЫ

1. Performance Based Navigation. Doc 9613. – Montreal : ICAO, 2008. – 290 p.
2. Ревнивых С. Г. Тенденции развития глобальной спутниковой навигации / С. Г. Ревнивых // Международная конференция по интегрированным навигационным системам: XIX международная конференция, Санкт-Петербург, 28–30 мая 2012 г.: сборник докладов. – СПб. : ГНЦ РФ ЦНИИ «Электроприбор», 2012. – С. 266–276.
3. Разработка и испытание интегрированной инерциально-спутниковой навигационной системы НСИ-2000MTG с расширенной доступностью спутниковых измерений / [Т. Н. Вахитов, А. Б. Колчев, К. Ю. Счастливец, В. Б. Успенский, П. В. Ларионов, А. А. Фомичев] // Международная конференция по интегрированным навигационным системам: XIX международная конференция, Санкт-Петербург, 28–30 мая 2012 г. : сборник докладов. – СПб. : ГНЦ РФ ЦНИИ «Электроприбор», 2012. – С. 226–235.
4. Крутько П. Д. Алгоритмы терминального управления линейных динамических систем / П. Д. Крутько // Известия РАН. Теория и системы управления. – 1998. – № 6. – С. 33–45.
5. Крутько П. Д. Обратные задачи динамики управляемых систем. Линейные модели / П. Д. Крутько. – М. : Наука, 1987. – 304 с.
6. Батенко А. П. Системы терминального управления / А. П. Батенко. – М. : Радио и связь, 1984. –160 с.
7. Salychev O. S. Applied Inertial Navigation: Problems and Solutions / O. S. Salychev. – M. : BMSTU Press, 2004. – 304 p.
8. Галлиулин А. С. Методы решения обратных задач динамики / А. С. Галиуллин. – М. : Наука, 1986. – 224 с.
9. Дорф Р. Современные системы управления / Р. Дорф, Р. Бишоп. – М. : Лаборатория Базовых Знаний, 2002. – 832 с.
10. Теория управления. Терминология. – М. : Наука, 1988. – Вып. 107. – 56 с.
11. Неусыпин К. А. Современные системы и методы наведения, навигации и управления летательными аппаратами / К. А. Неусыпин. – М. : МГОУ, 2009. – 500 с.
12. Barbour N. M. Inertial Navigation Sensors [Electronic resource] / N. M. Barbour // NATO RTO Lecture Series, RTO-EN-SET-116, Low-Cost Navigation Sensors and Integration Technology, 2010. – Access mode: http://ftp.rta.nato.int/public//PubFullText/RTO/EN/RTO-EN-SET-116-2010///EN-SET-116(2010)-02.pdf
13. Голован А. А. Математические основы навигационных систем. Часть I. Математические модели инерциальной навигации / А. А. Голован, Н. А. Парусников. – М. : Изд-во МГУ. 2011. – 132 с.
14. Успенский В. Б. Теоретические основы гиросилового управления ориенацией космического летательного аппарата / В. Б. Успенский. – Харьков : НТУ «ХПИ», 2006. – 328 с.
15. Хацько Н. Е. Исследование возможности управления полетом по данным инерциальных датчиков низкого класса точности / Н. Е. Хацько // Проблеми машинобудування. – Харків : ІПМАШ. – 2013. – Т.16, № 5. – С. 32–35.

Статья поступила в редакцию 07.12.2015.
После доработки 14.12.2015.

Хацько Н. Є.

Канд. техн. наук, старший викладач кафедри систем і процесів управління Національного технічного університету «Харківський політехнічний інститут», Харків, Україна

РОЗВ'ЯЗАННЯ ТЕРМІНАЛЬНОЙ ЗАДАЧІ УПРАВЛІННЯ З ВИРИСТАННЯМ ІНОФРМАЦІЇ ІНЕРЦІАЛЬНОГО БЛОКА

Розглянуто задачу термінального управління літальним апаратом за інформацією безплатформної інерціальної навігаційної системи, в якій за заданий час потрібно перевести апарат у задане просторове положення з необхідним значенням вектора кінцевої швидкості. Побудовано програмне управління для незбуреного руху літального апарата. Синтезовано функцію управління літального апарату за поточним значенням його вектора стану, визначеного інерціальним навігаційним блоком. Отримано вирази для оцінки точності вирішення термінальної задачі з урахуванням похибки акселерометра. Проаналізовано точність виведення літального апарату в кінцеву точку замкненою системою управління при використанні інформації від інерціальних датчиків з відомим рівнем похибок.

На прикладі спрощеної моделі проведено комп'ютерне моделювання руху літального апарату. Доведено, що точність приведення вектора стану в задану кінцеву точку залежить від рівня збурень і похибки вхідної інформації. Проведено чисельні експерименти, що дозволяють виявити вплив чисельного значення глибини прогнозу, використаного в синтезі управління за методом переслідування провідною точки, на точність рішення термінальної задачі. Сформульовано рекомендації для проектування алгоритмів автоматичного управління рухом в комплексі з проектуванням інформаційно-вимірювальної системи.

Ключові слова: система управління, безплатформна інерціальна навігаційна система, похибка виміру вектору стану, динамічне збурення, точність термінального управління.

Khatsko N. E.

PhD, Senior Lecturer of Systems and processes of control department, Narional Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine

APPROACH TO SOLVING THE TERMINAL CONTROL PROBLEM ON CONDITION OF IMPRECISE MEASURING OF STATE VECTOR

The article considers the problem of terminal control of an aircraft using the data from the strapdown inertial navigation system, specifically the case when the aircraft is moved to a specific spatial position having the required vector of terminal velocity. Builded program model for not altered motion of aircraft. Developed software for aircraft control, based on current state-vector taken from inertial navigation system. Method of analysis of aircraft positioning accuracy in the final location of a closed-loop control system that is based on the information taken from inertial sensors that have a deliberately known level of errors. Deduced formulas for estimation of aircraft final position taking into account accelerometer errors.

On the base of a simplified model the analysis of the accuracy of the state vector bringing to a given endpoint is performed, depending on the input information level of the perturbation and inaccuracy of input information.

Conducted experiments demonstrate dependency of software accuracy on accuracy of terminal control problem solution.

The following recommendations were obtained after performing analysis of accuracy of a close-loop system functioning:

– With time, the impact of errors from inertial navigation system decreases;

– To reduce impact of dynamic altered motion onto the finale state, we should decrease coefficient of gain transfer function;

– When choosing a variable parameter in the law of Synthesis method of leading point persecution, you should be guided with the following considerations: if the information-measuring system provides the coordinates and velocity of the object with high accuracy, the selected parameter of the forecast should be minimal. When using low-accuracy sensors, the value of forecast parameter should be increased.

Keywords: control system, strapdown inertial navigation system, the measurement the error of the state vector measurement, dynamical disturbance, the accuracy of the terminal management.

REFERENCES

1. Performance Based Navigation. Doc 9613. Montreal: ICAO, 2008, 290 p.
2. Revnivykh S. G. Tendentsii razvitiya globalnoy sputnikovoy navigatsii, *Mezhdunarodnaya konferentsiya po integrirovannym navigatsionnym sistemam: XIX mezhdunarodnaya konferentsiya, Sankt-Peterburg, 28–30 maya 2012 g.: sbornik dokladov*. SPb, GNTs RF TsNII «Elektropribor», 2012, pp. 266–276.
3. Vakhitov T. N., Kolchev A. B., Schastlivets K. Yu., Uspenskiy V. B., Larionov P. V., Fomichev A. A. Razrabotka i ispytanie integrirovannoy inertsialno-sputnikovoy navigatsionnoy sistemy NSI-2000MTG s rasshirennoy dostupnostyu sputnikovykh izmereniy, *Mezhdunarodnaya konferentsiya po integrirovannym navigatsionnym sistemam: XIX mezhdunarodnaya konferentsiya, Sankt-Peterburg, 2–30 maya 2012 g.: sbornik dokladov.* SPb, GNTs RF TsNII «Elektropribor», 2012, pp. 226–235.
4. Krutko P. D. Algoritmy terminalnogo upravleniya lineynykh dinamicheskikh sistem, *Izvestiya RAN. Teoriya i sistemy upravleniya*, 1998, No. 6, pp. 33–45.
5. Krutko P. D. Obratnye zadachi dinamiki upravlyaemykh sistem. Lineynye modeli. Moscow, Nauka, 1987, 304 p.
6. Batenko A. P. Sistemy terminalnogo upravleniya. Moscow, Radio i svyaz, 1984, 160 p.
7. Salychev O. S. Applied Inertial Navigation: Problems and Solutions. Moscow, BMSTU Press, 2004, 304 p.
8. Galliulin A. S. Metody resheniya obratnykh zadach dinamiki. Moscow, Nauka, 1986, 224 p.
9. Dorf R., Bishop R. Sovremennye sistemy upravleniya. Moscow, Laboratoriya Bazovykh Znaniy, 2002, 832 p.
10. Teoriya upravleniya. Terminologiya. Moscow, Nauka, 1988, Vyp. 107, 56 p.
11. Neusypin K. A. Sovremennye sistemy i metody navedeniya, navigatsii i upravleniya letatelnymi apparatami. Moscow, MGOU, 2009, 500 p.
12. Barbour N. M. Inertial Navigation Sensors [Electronic resource], *NATO RTO Lecture Series, RTO-EN-SET-116, Low-Cost Navigation Sensors and Integration Technology, 2010.* Access mode: http://ftp.rta.nato.int/public//PubFullText/RTO/EN/RTO-EN-SET-116-2010///EN-SET-116(2010)-02.pdf
13. Golovan A. A., Parusnikov N. A. Matematicheskie osnovy navigatsionnykh sistem. Chast I. Matematicheskie modeli inertsialnoy navigatsii. Moscow, Izd-vo MGU, 2011, 132 p.
14. Uspenskiy V. B. Teoreticheskie osnovy girosilovogo upravleniya orienatsiey kosmicheskogo letatelnogo apparata. Kharkov, NTU «KhPI», 2006, 328 p.
15. Khatsko N. Ye. Issledovanie vozmozhnosti upravleniya poletom po dannym inertsialnykh datchikov nizkogo klassa tochnosti, *Problemi mashinobuduvannya*. Kharkiv, IPMASh, 2013, vol.16, No. 5, pp. 32–35.

Наукове видання

Радіоелектроніка, інформатика, управління

№ 1/2016

Науковий журнал

Головний редактор – д-р фіз.-мат. наук В. В. Погосов

Заст. головного редактора – д-р техн. наук С. О. Субботін

Комп'ютерне моделювання та верстання С. В. Зуб
Редактор англійських текстів С. О. Субботін

Оригінал-макет підготовлено у редакційно-видавничому відділі ЗНТУ

Свідоцтво про державну реєстрацію
КВ № 6904 від 29.01.2003.

Підписано до друку 05.04.2016. Формат 60×84/8.
Папір офс. Різогр. друк. Ум. друк. арк. 12,56.
Тираж 300 прим. Зам. № 253.

69063, м. Запоріжжя, ЗНТУ, друкарня, вул. Жуковського, 64

Свідоцтво суб'єкта видавничої справи
ДК № 2394 від 27.12.2005.