

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Том 25

2019

№ 12

САПР

КОМПЬЮТЕРНАЯ ГРАФИКА

МЕТОДЫ ПРОГРАММИРОВАНИЯ

ОПЕРАЦИОННЫЕ СИСТЕМЫ И СРЕДЫ

ТЕЛЕКОММУНИКАЦИИ
И ВЫЧИСЛИТЕЛЬНЫЕ СЕТИ

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

НЕЙРОСЕТИ И
НЕЙРОКОМПЬЮТЕРЫ

СТРУКТУРНЫЙ СИНТЕЗ

ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ
СИСТЕМЫ

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

ОПТИМИЗАЦИЯ И МОДЕЛИРОВАНИЕ

ИТ В ОБРАЗОВАНИИ

ГИС

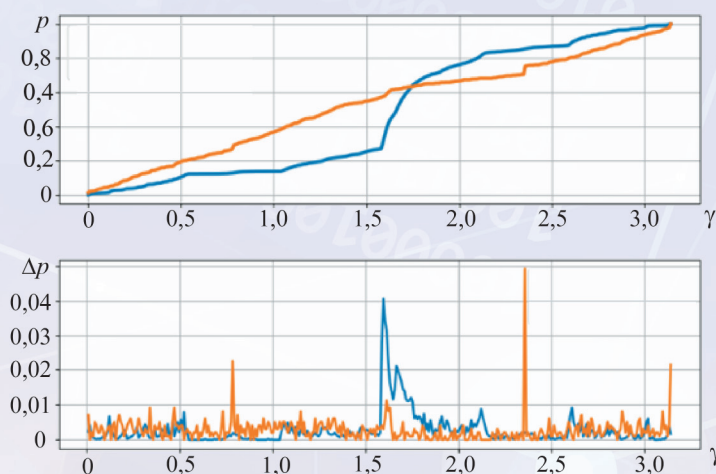


Рис. 5. Графики функций $p(\gamma)$ и $\Delta p(\gamma)$ при $\Delta\gamma = 0,01\pi$ ПКВ ландшафтной выборки тестовых функций Розенброка (синий) и Экли (оранжевый): — Rosenbrock; — Ackley

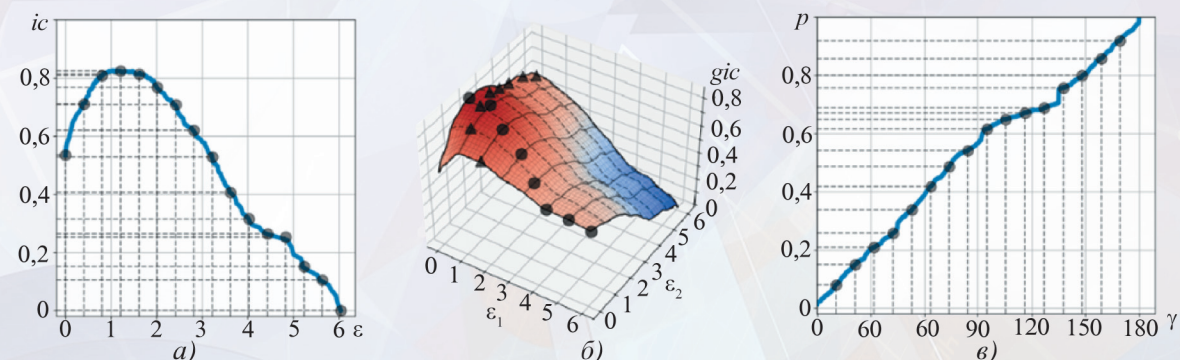


Рис. 6. Графики характеризующих функций методов информационного содержания (а), обобщенного информационного содержания (б) и секторизации карты вариативности (в); точками обозначены значения, используемые в качестве компонент вектора ХП при $|C| = 16$

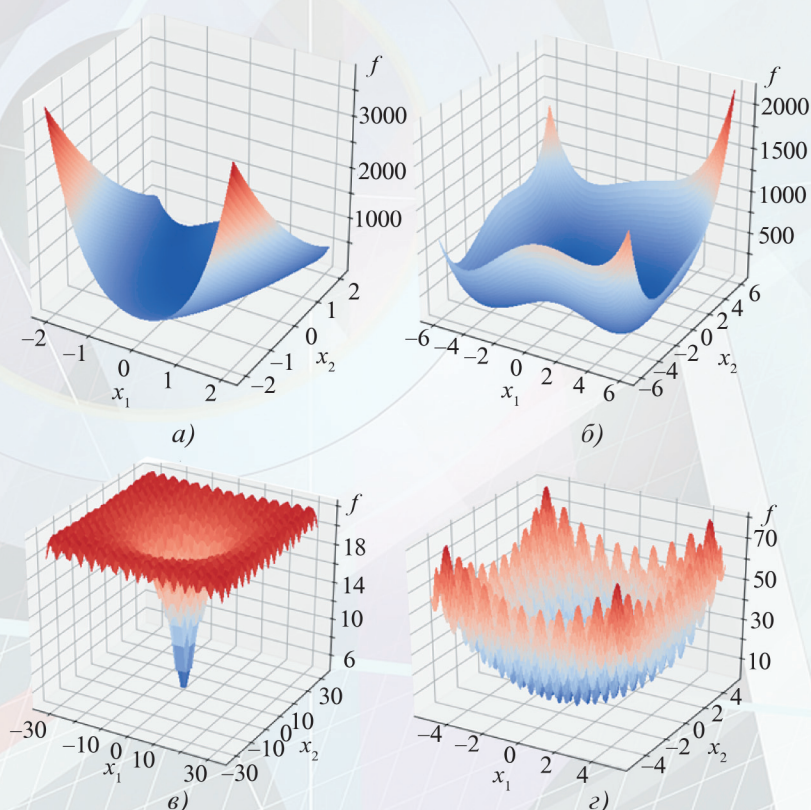


Рис. 7. Ландшафты целевых функций задач оптимизации $q_1(a)$, $q_2(b)$, $q_3(v)$ и $q_4(z)$

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Том 25
2019
№ 12

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с ноября 1995 г.

DOI 10.17587/issn.1684-6400

УЧРЕДИТЕЛЬ

Издательство "Новые технологии"

СОДЕРЖАНИЕ

МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ

Агасиев Т. А. Метод секторизации полной карты вариативности целевой функции для оценки характерных признаков задачи оптимизации 707

Салахов А. З. Исследование рассеяния электромагнитных полей в биообъектах методом компьютерного моделирования 718

ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ И СЕТИ

Абрамов А. Г., Евсеев А. В. Концептуальные аспекты создания в Российской Федерации национальной исследовательской компьютерной сети нового поколения 724

Холопов Ю. А., Нгуен Ван Хиеу. Реализация контроллера активной синхронной измерительной подсистемы цифровой системы управления 733

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Грибова В. В., Шалфеева Е. А. Обеспечение жизнеспособности систем, основанных на знаниях 738

НЕЙРОСЕТЕВЫЕ ТЕХНОЛОГИИ

Соловьев Р. А., Тельпухов Д. В., Кустов А. Г., Исаева Т. Ю., Волков А. А. Применение методов модулярной арифметики при разработке аппаратных реализаций нейронных сетей 747

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

Костин В. Н. Оценка эффективности физической защиты информации критически важных объектов на основе марковских цепей 757

Указатель статей, опубликованных в журнале "Информационные технологии" в 2019 г. 766

Главный редактор:

СТЕМПОВСКИЙ А. Л.,
акад. РАН, д. т. н., проф.

Зам. главного редактора:

ИВАННИКОВ А. Д., д. т. н., проф.
ФИЛИМОНОВ Н. Б., д. т. н., с.н.с.

Редакционный совет:

БЫЧКОВ И. В., акад. РАН, д. т. н.
ЖУРАВЛЕВ Ю. И.,
акад. РАН, д. ф.-м. н., проф.
КУЛЕШОВ А. П.,
акад. РАН, д. т. н., проф.
ПОПОВ Ю. С.,
акад. РАН, д. т. н., проф.
РУСАКОВ С. Г.,
чл.-корр. РАН, д. т. н., проф.
РЯБОВ Г. Г.,
чл.-корр. РАН, д. т. н., проф.
СОЙФЕР В. А.,
акад. РАН, д. т. н., проф.
СОКОЛОВ И. А.,
акад. РАН, д. т. н., проф.
СУЕТИН Н. В., д. ф.-м. н., проф.
ЧАПЛЫГИН Ю. А.,
акад. РАН, д. т. н., проф.
ШАХНОВ В. А.,
чл.-корр. РАН, д. т. н., проф.
ШОКИН Ю. И.,
акад. РАН, д. т. н., проф.
ЮСУПОВ Р. М.,
чл.-корр. РАН, д. т. н., проф.

Редакционная коллегия:

АВДОШИН С. М., к. т. н., доц.
АНТОНОВ Б. И.
БАРСКИЙ А. Б., д. т. н., проф.
ВАСЕНИН В. А., д. ф.-м. н., проф.
ВАСИЛЬЕВ В. И., д. т. н., проф.
ВИШНЕКОВ А. В., д. т. н., проф.
ДИМИТРИЕНКО Ю. И., д. ф.-м. н., проф.
ДОМРАЧЕВ В. Г., д. т. н., проф.
ЗАБОРОВСКИЙ В. С., д. т. н., проф.
ЗАРУБИН В. С., д. т. н., проф.
КАРПЕНКО А. П., д. ф.-м. н., проф.
КОЛИН К. К., д. т. н., проф.
КУЛАГИН В. П., д. т. н., проф.
КУРЕЙЧИК В. В., д. т. н., проф.
ЛЬВОВИЧ Я. Е., д. т. н., проф.
МАРТЫНОВ В. В., д. т. н., проф.
МИХАЙЛОВ Б. М., д. т. н., проф.
НЕЧАЕВ В. В., к. т. н., проф.
ПОЛЕЩУК О. М., д. т. н., проф.
САКСОНОВ Е. А., д. т. н., проф.
СОКОЛОВ Б. В., д. т. н., проф.
ТИМОНИНА Е. Е., д. т. н., проф.
УСКОВ В. Л., к. т. н. (США)
ФОМИЧЕВ В. А., д. т. н., проф.
ШИЛОВ В. В., к. т. н., доц.

Редакция:

БЕЗМЕНОВА М. Ю.

Информация о журнале доступна по сети Internet по адресу <http://novtex.ru/IT>.
Журнал включен в систему Российского индекса научного цитирования и базу данных RSCI на платформе Web of Science.
Журнал входит в Перечень научных журналов, в которых по рекомендации ВАК РФ должны быть опубликованы научные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

INFORMATION TECHNOLOGIES

INFORMACIONNYYE TEHNOLOGII

Vol. 25
2019
No. 12

THEORETICAL AND APPLIED SCIENTIFIC AND TECHNICAL JOURNAL

Published since November 1995

ISSN 1684-6400

CONTENTS

MODELING AND OPTIMIZATION

- Agasiev T. A.** Sectorization Method of Full Variability Map Analysis for Characteristic Features Assessment of Global Optimization Problems 707
- Salahov A. Z.** The Study of the Scattering of Electromagnetic Fields in Biological Objects by Computer Simulation 718

COMPUTING SYSTEMS AND NETWORKS

- Abramov A. G., Evseev A. V.** Conceptual Aspects of Creating the New Generation National Research Computer Network in the Russian Federation 724
- Holopov Yu. A., Nguyen Van Hieu.** Implementation of the Active Synchronous Measuring Subsystem Controller in Digital Control System 733

INTELLIGENT SYSTEMS AND TECHNOLOGIES

- Gribova V. V., Shalfeeva E. A.** Ensuring of Viability of Systems Based on Knowledge 738

NEUROTECHNOLOGIES

- Solovyev R. A., Telpukhov D. V., Kustov A. G., Isaeva T. U., Volkov A. A.** Application of Modular Arithmetic Methods in the Development of Hardware Implementations of Neural Networks 747

APPLICATION INFORMATION SYSTEMS

- Kostin V. N.** Assessment of Efficiency of Physical Security of Information of Crucial Objects on the Basis of Markov Chains 757

Editor-in-Chief:

Stempkovsky A. L., Member of RAS,
Dr. Sci. (Tech.), Prof.

Deputy Editor-in-Chief:

Ivannikov A. D., Dr. Sci. (Tech.), Prof.
Filimonov N. B., Dr. Sci. (Tech.), Prof.

Chairman:

Bychkov I. V., Member of RAS,
Dr. Sci. (Tech.), Prof.
Zhuravljov Yu. I., Member of RAS,
Dr. Sci. (Phys.-Math.), Prof.
Kuleshov A. P., Member of RAS,
Dr. Sci. (Tech.), Prof.
Popkov Yu. S., Member of RAS,
Dr. Sci. (Tech.), Prof.
Rusakov S. G., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.
Ryabov G. G., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.
Soifer V. A., Member of RAS,
Dr. Sci. (Tech.), Prof.
Sokolov I. A., Member of RAS,
Dr. Sci. (Phys.-Math.), Prof.
Suetin N. V.,
Dr. Sci. (Phys.-Math.), Prof.
Chaplygin Yu. A., Member of RAS,
Dr. Sci. (Tech.), Prof.
Shakhnov V. A., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.
Shokin Yu. I., Member of RAS,
Dr. Sci. (Tech.), Prof.
Yusupov R. M., Corresp. Member of RAS,
Dr. Sci. (Tech.), Prof.

Editorial Board Members:

Avdoshin S. M., Cand. Sci. (Tech.), Ass. Prof.
Antonov B. I.
Barsky A. B., Dr. Sci. (Tech.), Prof.
Vasenin V. A., Dr. Sci. (Phys.-Math.), Prof.
Vasiliev V. I., Dr. Sci. (Tech.), Prof.
Vishnekov A. V., Dr. Sci. (Tech.), Prof.
Dimitrienko Yu. I., Dr. Sci. (Phys.-Math.), Prof.
Domrachev V. G., Dr. Sci. (Tech.), Prof.
Zaborovsky V. S., Dr. Sci. (Tech.), Prof.
Zarubin V. S., Dr. Sci. (Tech.), Prof.
Karpenko A. P., Dr. Sci. (Phys.-Math.), Prof.
Kolin K. K., Dr. Sci. (Tech.)
Kulagin V. P., Dr. Sci. (Tech.), Prof.
Kureichik V. V., Dr. Sci. (Tech.), Prof.
Ljovich Ya. E., Dr. Sci. (Tech.), Prof.
Martynov V. V., Dr. Sci. (Tech.), Prof.
Mikhailov B. M., Dr. Sci. (Tech.), Prof.
Nechaev V. V., Cand. Sci. (Tech.), Ass. Prof.
Poleschuk O. M., Dr. Sci. (Tech.), Prof.
Saksonov E. A., Dr. Sci. (Tech.), Prof.
Sokolov B. V., Dr. Sci. (Tech.)
Timonina E. E., Dr. Sci. (Tech.), Prof.
Uskov V. L. (USA), Dr. Sci. (Tech.)
Fomichev V. A., Dr. Sci. (Tech.), Prof.
Shilov V. V., Cand. Sci. (Tech.), Ass. Prof.

Editors:

Bezmenova M. Yu.

Complete Internet version of the journal at site: <http://novtex.ru/IT>.

According to the decision of the Higher Certifying Commission of the Ministry of Education of Russian Federation, the journal is inscribed in "The List of the Leading Scientific Journals and Editions wherein Main Scientific Results of Theses for Doctor's or Candidate's Degrees Should Be Published"

Т. А. Агасиев, аспирант, e-mail: agtaleh@mail.ru,
МГТУ им. Н. Э. Баумана, Москва

Метод секторизации полной карты вариативности целевой функции для оценки характерных признаков задачи оптимизации

Методы ландшафтного анализа разработаны для оценки различных характерных признаков целевой функции в задаче оптимизации. Точность оценок во многом зависит от выбранного плана эксперимента, т. е. от числа и расположения в пространстве поиска точек, которые формируют дискретное представление ландшафта целевой функции. В статье рассмотрены методы анализа карты вариативности целевой функции, которые позволяют получить оценки, наиболее устойчивые к вариациям плана эксперимента. Недостатком этих методов является высокая чувствительность значений компонент вектора характерных признаков задачи оптимизации к масштабу значений приращений ее целевой функции. Это может отрицательно сказаться на обобщающей способности алгоритма классификации задач по наиболее значимым для оптимизации признакам целевой функции. В связи с этим в работе предложен метод секторизации карты вариативности целевой функции, позволяющий получить оценки характерных признаков, устойчивые к изменению масштаба значений целевой функции. Предложена методика проведения вычислительного эксперимента для сравнительного исследования эффективности методов анализа карты вариативности. Сформулированы критерии эффективности алгоритмов оценки характерных признаков целевой функции. Получены результаты, указывающие на целесообразность применения предложенных методов ландшафтного анализа для классификации задач оптимизации по различным признакам. Качество классификации оптимизационных задач во многом определяет эффективность интеллектуализации подсистем параметрической оптимизации в САПР.

Ключевые слова: глобальная оптимизация, целевая функция, характерные признаки задачи оптимизации, ландшафтный анализ, ландшафтная выборка, метод информационного содержания, метод секторизации, карта вариативности

Введение

Современные системы автоматизированного проектирования и инженерного анализа включают в себя программы, реализующие в том числе алгоритмы непрерывной глобальной оптимизации. Рассмотрим задачу q глобальной оптимизации:

$$\min_{X \in D_X} f(X) = f(X^*) = f^*, \quad (1)$$

где $X \in \mathbb{R}^n$ — n -мерный вектор варьируемых параметров (n — размерность задачи); $D_X \subset \mathbb{R}^n$ — область поиска; $f(X) \in \mathbb{R}^1$ — целевая функция; X^*, f^* — искомые оптимальный вектор X и значение функции $f(X)$.

При решении практически значимых задач оптимизации в САПР широко применяют методы предварительного анализа задачи q для эффективного распределения вычислительных

ресурсов (Shan S., Wang G. G. Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions) [1]. Результатом анализа свойств задачи является вектор $C = (c_1, \dots, c_q)$ характерных признаков (ХП) этой задачи (Kerschke P. Comprehensive Feature-Based Landscape Analysis of Continuous and Constrained Optimization Problems Using the R-Package flacco) [2]. К числу ХП задачи q относят и ХП ее целевой функции. Будем рассматривать только ХП целевой функции $f(X)$ задачи q .

Различают априорные и апостериорные ХП [3] (Агасиев Т. А., Карпенко А. П. Современные техники глобальной оптимизации. Обзор). Априорные ХП прямо вытекают из постановки задачи оптимизации, например, ее размерность n . Апостериорные ХП, в отличие от априорных, требуют предварительного вычисления значений целевой функции в некоторых

точках области поиска D_X в целях последующей экспертной и/или автоматической оценки полученных результатов. Точки для вычислений выбирают в соответствии с некоторым методом планирования эксперимента (например, на основе полного факторного эксперимента (Cavazuti M. Design of experiments)) [4]. В условиях высокой вычислительной сложности целевой функции целесообразно применять адаптивные методы планирования эксперимента (Li G., Aute V., Azarm S. An accumulative error based adaptive design of experiments for offline metamodeling) [5]. После определения значений ХП целевой функции результаты вычислительного эксперимента используют на этапе инициализации алгоритма оптимизации, например, для построения начальной аппроксимирующей модели целевой функции (Forrester A. I. J., Keane A. J. Recent advances in surrogate-based optimization) [6].

Методы автоматической оценки значений апостериорных ХП целевой функции включают методы *ландшафтного анализа* (Mersmann O. et al. Exploratory landscape analysis) [7]. Эти методы применяют, например, для различения и классификации задач оптимизации при исследовании эффективности алгоритмов их решения. Собранный информацию используют для прогнозирования эффективности оптимизационного алгоритма при решении задач оптимизации на основании предварительно вычисленного вектора ХП ее целевой функции. Это позволяет выбирать и настраивать оптимизационный алгоритм перед решением практически значимых задач (Hoos H. N., Neumann F., Trautmann H. Automated Algorithm Selection and Configuration) [8].

В большинстве случаев для получения оценок ХП целевой функции методами ландшафтного анализа необходима *ландшафтная выборка* — набор точек $X_i \in D_X$, $i \in [1:m]$, выбранных в соответствии с некоторым методом планирования эксперимента. Искомые оценки ХП зависят от числа и расположения точек ландшафтной выборки, определяющей дискретное представление ландшафта целевой функции. Известно большое число методов ландшафтного анализа, оценивающих различные ХП целевой функции [9—11]. Общим недостатком этих методов является высокая чувствительность значений искомых ХП к изменению ландшафтной выборки. Для более полного анализа особенностей целевой функции на основе ландшафтной выборки в нашей работе [12] сформулировано понятие *карты вариативности* целевой функции, а также предложен ме-

тод составления так называемой *полной карты вариативности* (ПКВ) ландшафтной выборки целевой функции.

Оценки ХП целевой функции $f(X)$ зависят как от ландшафтной выборки, так и от используемых методов составления и анализа карты вариативности этой выборки. В работе [12] предложено два метода анализа карты вариативности: *модифицированный метод информационного содержания* и *метод обобщенного информационного содержания*. Оценки ХП на основе этих методов в значительной степени зависят от константы Липшица исследуемой целевой функции $f(X)$. Высокая чувствительность методов анализа карты вариативности к константе Липшица может понизить точность оценки величин, характеризующих ключевые для алгоритма оптимизации особенности функции $f(X)$, например, степень ее мультимодальности. В данной работе предложен *метод секторизации карты вариативности*, позволяющий получить более устойчивые оценки значимых для алгоритма оптимизации ХП целевой функции.

В работе ключевым понятием является *характеризующая функция* карты вариативности. Характеризующей называем функцию, значения и аргументы которой используют в качестве компонент вектора ХП целевой функции. В данной работе предлагаем использовать только значения характеризующей функции при определенных значениях ее аргументов. Алгоритм оценки ХП целевой функции на основании значений характеризующей функции метода анализа карты вариативности называем алгоритмом *характеризации* целевой функции. В работе [12] рассмотрены одноименные характеризующие функции методов информационного и обобщенного информационного содержания. В данной работе для оценки ХП целевой функции методом секторизации предложена характеризующая *функция плотности сектора* карты вариативности. Целью работы является выбор и обоснование эффективности метода анализа карты вариативности для составления вектора ХП целевой функции. В связи с этим в работе также предложена методика оценки эффективности алгоритмов характеризации, которые являются фундаментом для синтеза интеллектуальных подсистем параметрической оптимизации путем сбора и обобщения опыта решения оптимизационных задач.

В п. 1 рассмотрены метод построения ПКВ ландшафтной выборки и общая схема методов анализа карты вариативности, кратко изложе-

ны модифицированный и обобщенный метод информационного содержания. В п. 2 представлен предлагаемый метод секторизации для анализа карты вариативности ландшафтной выборки. В п. 3 изложена методология проведения вычислительного эксперимента и сформулированы критерии эффективности алгоритма характеристики. В п. 4 на основе указанных методов анализа карты вариативности предложены векторы ХП целевой функции различной размерности. Выполнено сравнительное исследование эффективности классификации задач оптимизации по различным признакам на основе составленных векторов ХП.

1. Полная карта вариативности. Методы ее построения и анализа

Общая схема метода оценки ХП целевой функции на основе результатов анализа ПКВ ландшафтной выборки имеет следующий вид:

- 1) составляем ландшафтную выборку $X_i \in D_X$, $i \in [1:m]$, целевой функции $f(X)$;
- 2) строим ПКВ ландшафтной выборки;
- 3) в соответствии с выбранным методом анализа ПКВ рассчитываем значения характеризующей функции;
- 4) составляем искомый вектор ХП целевой функции $f(X)$ на основе этих значений.

Тип характеризующей функции и значения ее аргументов, используемые для расчета значений компонент вектора ХП целевой функции, определяются алгоритмом анализа карты вариативности. На основе значений характеризующей функции соответствующего метода можно предложить большое число алгоритмов оценки ХП целевой функции.

Ниже приведено краткое описание предложенного в работе [12] метода составления ПКВ целевой функции. Затем рассмотрен модифицированный метод информационного содержания *ICoFiS* [13], адаптированный для анализа карты вариативности [12]. Также представлен метод обобщенного информационного содержания [12], который позволяет более точно управлять разбиением карты вариативности на подобласти с учетом ее симметричности.

Полная карта вариативности [12]. Картой вариативности называем сформированный по некоторым правилам набор пар значений приращений целевой функции между соседними точками ландшафтной выборки.

Строим полную карту вариативности следующим образом.

1. Составляем для каждой точки X_i , $i \in [1:m]$, ландшафтной выборки наиболее пригодные для рассмотрения блоки точек по следующей схеме:

а) обозначим $d_{a,b}$ расстояние между двумя точками X_a и X_b . Находим ближайшие к X_i точки X_{p_k} , $k \in [1:K]$, где K — число точек, для которых выполнено неравенство

$$d_{i,p_k} < \max_{i_1} \min_{i_2} d_{i_1,i_2}, \quad i_1, i_2 \in [1:m], \quad i_1 \neq i_2; \quad (2)$$

б) если $K > 1$, то составляем блоки t_k , $k \in [1:K-1]$, из точек $X_{p_k}, X_i, X_{p_{k+1}}$ и точек $X_{p_{k+1}}, X_i, X_{p_k}$. Включаем в рассмотрение блок t_k в случае, если прямые $X_{p_k}X_i$ и $X_iX_{p_{k+1}}$ образуют угол $\alpha_k > \pi/2$. Если $K \leq 1$, либо если угол $\alpha_k > \pi/2$ для всех найденных точек, то отсутствуют подходящие точки для составления блока.

2. Для всех составленных блоков t_j , $j = 1, 2, \dots$, вычисляем разности δ_1^j, δ_2^j соответствующих значений целевой функции $f(X)$ между точками X_1^j, X_2^j и точками X_2^j, X_3^j с учетом евклидова расстояния между ними:

$$\delta_1^j = \frac{f_2^j - f_1^j}{\|X_2^j - X_1^j\|}, \quad \delta_2^j = \frac{f_3^j - f_2^j}{\|X_3^j - X_2^j\|}. \quad (3)$$

Обозначим l число блоков t_j , составленных в результате описанных выше действий. Полученные значения δ_1^j, δ_2^j , $j \in [1:l]$, образуют ПКВ ландшафтной выборки.

ПКВ удобно представить в виде множества точек в двухмерной системе координат $0\delta_1\delta_2$. Пример графического представления ПКВ ландшафтной выборки приведен на рис. 1 для

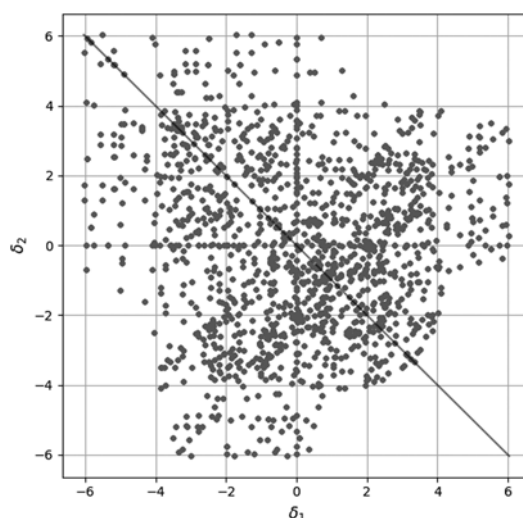


Рис. 1. Полная карта вариативности ландшафтной выборки, полученной методом полного факторного эксперимента для тестовой функции Экли

тестовой функции Экли [14]. Далее в работе, говоря о ПКВ, подразумеваем в том числе и ее графическое представление. Границы ПКВ определяются минимальными и максимальными значениями величин $\delta_1^j, \delta_2^j, j \in [1:L]$, и зависят от константы Липшица целевой функции $f(X)$.

По ПКВ ландшафтной выборки целевой функции $f(X)$ можно судить о степени ее мульти-модальности. Если точки $\delta_1^j, \delta_2^j, j \in [1:L]$, равномерно рассредоточены по ПКВ, то ландшафт функции $f(X)$ имеет большое число локальных экстремумов. Если большая часть этих точек сосредоточена в центральной области ПКВ, то, вероятнее всего, функция $f(X)$ является гладкой и имеет малое число локальных экстремумов [12].

Модифицированный метод информационного содержания [12]. Схема метода соответствует общей схеме методов расчета ХП, представленной выше. В данном случае преобразуем значения (3) $\delta_1^j, \delta_2^j, j \in [1:L]$, ПКВ в символьный набор $s_1^j, s_2^j, j \in [1:L]$, используя формулы

$$s_1^j(\varepsilon) = \begin{cases} \bar{1}, & \delta_1^j < -\varepsilon; \\ 0, & |\delta_1^j| \leq \varepsilon; \\ 1, & \delta_1^j > \varepsilon; \end{cases} \quad s_2^j(\varepsilon) = \begin{cases} \bar{1}, & \delta_2^j < -\varepsilon; \\ 0, & |\delta_2^j| \leq \varepsilon; \\ 1, & \delta_2^j > \varepsilon, \end{cases} \quad (4)$$

где $\varepsilon \in [0: \max(\delta_1^j, \delta_2^j, j \in [1:L])]$ — параметр, определяющий чувствительность метода и разбиение ПКВ на подобласти; символами $\{\bar{1}, 0, 1\}$ обозначаем отрицательные, нейтральные и положительные изменения значений функции $f(X)$ соответственно. Набор символов $s_1^j, s_2^j, j \in [1:L]$, используем для вычисления искомого ХП целевой функции $f(X)$ на основании значений функции информационного содержания ic (рис. 2), которая в данном случае является характеризующей

$$ic(\varepsilon) = - \sum_{a \neq b} p_{ab}(\varepsilon) \log_6 p_{ab}(\varepsilon); \quad (5)$$

$$p_{ab}(\varepsilon) = \frac{n_{ab}(\varepsilon)}{m-1}, \quad (6)$$

где $p_{ab}(\varepsilon)$ — оценка вероятности обнаружения блока типа ab в символьном наборе; $n_{ab}(\varepsilon)$ — число блоков t_j типа ab в этом наборе. Тип блока t_j определяют значения $a, b \in \{\bar{1}, 0, 1\}$ символов s_1^j, s_2^j , соответствующих значениям δ_1^j, δ_2^j этого блока [12].

Координаты наиболее информативных точек характеризующей функции $ic(\varepsilon)$ используем в качестве компонент вектора ХП целевой функции $f(X)$, например, точки перегиба,

точки, в которых достигаются максимальные и минимальные значения этой функции и т. д. Различные способы определения таких информативных точек и их интерпретация предложены в работе [13].

Метод обобщенного информационного содержания [12]. Схема метода также соответствует общей схеме методов расчета ХП, представленной выше. В данном случае разбиваем карту вариативности на подобласти, каждая из которых соответствует некоторому типу блоков t_j , определяемому символами

$$s_1^j(\varepsilon_1, \varepsilon_2) = \begin{cases} \bar{1}, & \delta_1^j < -\varepsilon_2; \\ 0, & -\varepsilon_2 \leq \delta_1^j \leq \varepsilon_1; \\ 1, & \delta_1^j > \varepsilon_1; \end{cases} \quad (7)$$

$$s_2^j(\varepsilon_1, \varepsilon_2) = \begin{cases} \bar{1}, & \delta_2^j < -\varepsilon_1; \\ 0, & -\varepsilon_1 \leq \delta_2^j \leq \varepsilon_2; \\ 1, & \delta_2^j > \varepsilon_2; \end{cases} \quad j \in [1:L].$$

Множество символов $s_1^j, s_2^j, j \in [1:L]$, используем для оценки искомого ХП целевой функции

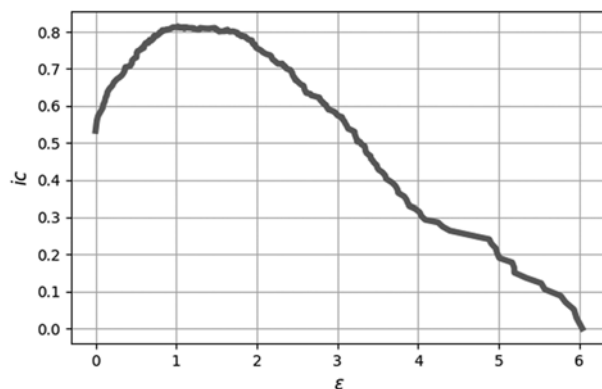


Рис. 2. Функция информационного содержания ландшафтной выборки тестовой функции Экли

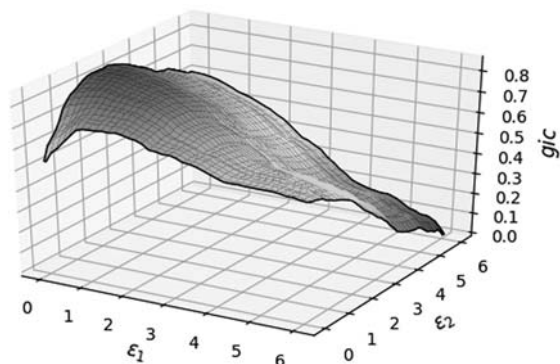


Рис. 3. Функция обобщенного информационного содержания ландшафтной выборки тестовой функции Экли

ции $f(X)$ на основании значений характеризующей функции обобщенного информационного содержания gic (рис. 3)

$$gic(\varepsilon_1, \varepsilon_2) = - \sum_{a \neq b} p_{ab}(\varepsilon_1, \varepsilon_2) \log_6 p_{ab}(\varepsilon_1, \varepsilon_2),$$

$$a, b \in \{\bar{1}, 0, 1\}. \quad (8)$$

Здесь $p_{ab}(\varepsilon_1, \varepsilon_2)$ — оценка вероятности обнаружения блока типа ab , которую рассчитываем по модифицированной формуле (6):

$$p_{ab}(\varepsilon_1, \varepsilon_2) = \frac{\sum_{k \in N_{ab}} \alpha_k}{\sum_{j=1} \alpha_j}, \quad a, b \in \{\bar{1}, 0, 1\}, \quad (9)$$

где N_{ab} — множество индексов блоков t_j типа ab ; α_j — угол, образуемый прямыми X_1^j, X_2^j и X_2^j, X_3^j , проходящими через точки блока t_j . Если $\alpha_j = \pi$ для всех рассматриваемых блоков $t_j, j \in [1:J]$, то получаем исходную формулу (6) оценки величины p_{ab} .

По аналогии с модифицированным методом информационного содержания используем координаты наиболее информативных точек характеризующей функции $gic(\varepsilon_1, \varepsilon_2)$ в качестве компонент вектора ХП целевой функции $f(X)$. Примеры таких точек приведены в п. 4.

2. Метод секторизации полной карты вариативности

Введем понятие *сектора* карты вариативности. Сектором называем область карты вариативности, ограниченную двумя полупрямыми, расположенными под углами $-\pi/4 - \gamma$ и $-\pi/4 + \gamma$ относительно положительной полуоси δ_1 . Параметр $\gamma \in [0; \pi]$ определяет площадь рассма-

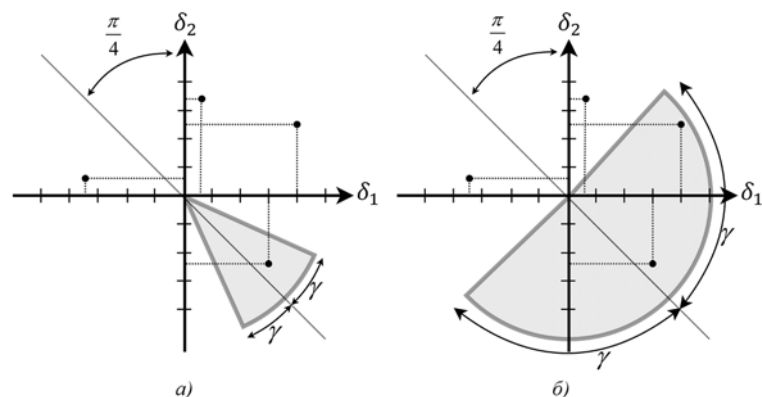


Рис. 4. Секторы карты вариативности ландшафтной выборки целевой функции, соответствующие значениям $\gamma = \pi/6$ (а) и $\gamma = \pi/2$ (б)

триваемой области, ограниченной величиной, равной расстоянию $\max_{j \in [1:J]} \sqrt{(\delta_1^j)^2 + (\delta_2^j)^2}$ до наиболее отдаленной от центра карты вариативности точки. На рис. 4 изображены примеры секторов карты вариативности, соответствующих различным значениям угла γ .

Определим характеризующую функцию *плотности сектора* карты вариативности

$$p(\gamma) = \frac{\sum_{k \in N_\gamma} \alpha_k}{\sum_{j=1} \alpha_j}. \quad (10)$$

Здесь N_γ — множество индексов блоков t_j , расположенных в секторе карты вариативности, определяемом углом γ . Функция $p(\gamma)$ является монотонно возрастающей в пределах $\gamma \in [0; \pi]$ и позволяет оценить вероятность обнаружения какого-либо блока t_j в указанном секторе. Значения этой функции пропорциональны числу точек, обнаруженных в секторе, и лежат в диапазоне $[0; 1]$.

Схема метода секторизации ПКВ ландшафтной выборки также соответствует общей схеме методов расчета ХП целевой функции (см. п. 1):

- 1) составляем ландшафтную выборку $X_i \in D_X, i \in [1:m]$, целевой функции $f(X)$;
- 2) строим ПКВ ландшафтной выборки;
- 3) рассчитываем соответствующие значения характеризующей функции $p(\gamma)$;
- 4) составляем искомый вектор ХП целевой функции $f(X)$ на основе этих значений.

Примеры значений параметра γ , используемых для расчета компонент вектора ХП на основе значений функции $p(\gamma)$, рассмотрены в п. 4.

На рис. 5 (см. вторую сторону обложки) приведены примеры графиков функций $p(\gamma)$ и $\Delta p(\gamma) = p(\gamma) - p(\gamma - \Delta\gamma)$, где $\Delta\gamma$ — шаг дискретизации. Эти графики позволяют визуально оценить отличия в распределении точек по секторам ПКВ ландшафтной выборки мультимодальной функции Экли и унимодальной функции Розенброка. Наибольший прирост значений функции $p(\gamma)$, соответствующих ландшафтной выборке мультимодальной функции Экли, наблюдается в окрестностях значений $\pi/4$ и $3\pi/4$ параметра γ . Повышенная концентрация точек в данных областях ПКВ свидетельствует о наличии большого числа блоков типа $\bar{1}\bar{1}$ или $\bar{1}1$ (в терминах метода *ICoFiS*), типичных

для областей целевой функции $f(X)$ с большим числом локальных экстремумов. Аналогичные графики для унимодальной функции Розенброка показывают, что максимальный прирост значений величины $p(\gamma)$ приходится на окрестность значения параметра $\gamma = \pi/2$. Это обусловлено тем, что в ландшафтной выборке целевой функции, имеющей малое число локальных экстремумов, превалируют "гладкие" блоки, соответствующие типам I и II [12].

Таким образом, распределение точек по секторам ПКВ в значительной степени зависит от существенных для метода оптимизации особенностей ландшафта целевой функции $f(X)$, например, от степени ее мультимодальности. Напротив, расстояния от этих точек до центральной области ПКВ очевидным образом не влияют на значения функции $p(\gamma)$ и в меньшей степени отражают ключевые признаки ландшафта целевой функции $f(X)$. Эти расстояния зависят, главным образом, от приращений целевой функции в соответствующей области пространства поиска D_X . Инвариантность оценок ХП целевой функции, полученных на основе значений функции $p(\gamma)$, к масштабу значений $\delta_1^j, \delta_2^j, j \in [1:J]$, потенциально улучшает обобщающую способность метода классификации задач оптимизации.

3. Методика оценки эффективности методов анализа ПКВ

Вектор ХП $C = (c_1, ..., c_{|C|}) \in D_C$ может иметь различную длину в зависимости от числа рассматриваемых значений характеризующей функции и принадлежит области допустимых значений D_C этой функции. Тип характеризующей функции определяется методом анализа карты вариативности, тогда как состав и размерность вектора ХП зависят от способа реализации этого метода в алгоритме характеристики целевой функции. Таким образом, метод анализа карты вариативности порождает

класс алгоритмов характеристики, каждый из которых определяет свое пространство ХП.

Об эффективности метода анализа карты вариативности судим по эффективности соответствующих алгоритмов характеристики. Эффективность алгоритма характеристики определяем эффективностью классификации оптимизационных задач на основе тех векторов ХП, состав и размерность которых установлены этим алгоритмом. Заметим, что эффективность классификации во многом зависит от устойчивости выбранных значений характеризующей функции как к изменению ландшафтной выборки, так и к незначительным изменениям ландшафта целевой функции.

Для исследования эффективности представленных методов анализа карты вариативности предлагаем алгоритмы характеристики, построенные по следующим правилам.

- Вектор ХП в случае использования одномерной характеризующей функции представляет собой набор ее значений в точках, равномерно распределенных в диапазоне варьирования аргумента этой функции.
- В случае если характеризующая функция имеет несколько аргументов, в качестве компонент вектора ХП выбираем максимальные значения этой функции по одному из аргументов при фиксированных значениях остальных.

Величина $|C|$ определяет размерность множества D_C допустимых векторов ХП. Алгоритмы характеристики целевой функции для методов информационного содержания, обобщенного информационного содержания и секторизации карты вариативности представлены в табл. 1. Обозначим $D_C^{ic}, D_C^{gic}, D_C^p$ множества допустимых векторов, определяемые соответствующими алгоритмами.

На рис. 6 (см. вторую сторону обложки) в качестве примера для тестовой функции Экли представлены графики характеризующих функций, точками на которых обозначены значения, используемые в качестве элементов вектора $C = (c_1, ..., c_{|C|})$.

Таблица 1

Алгоритмы характеристики целевой функции

Алгоритм оценки ХП	Элементы вектора ХП $C = (c_1, ..., c_{ C })$
Алгоритм информационного содержания	$ic \left(\varepsilon_i = \varepsilon_{\min} + \frac{i}{ C -1} (\varepsilon_{\max} - \varepsilon_{\min}) \right), i \in [0: C -1]$
Алгоритм обобщенного информационного содержания	$\max_{\varepsilon_1} gic(\varepsilon_1, \varepsilon_2 = \varepsilon_i), \max_{\varepsilon_2} gic(\varepsilon_1 = \varepsilon_i, \varepsilon_2), i \in \left[0: \left\lfloor \frac{ C }{2} - 1 \right\rfloor \right]$
Алгоритм секторизации карты вариативности	$p \left(\gamma_i = \frac{i}{ C +1} \pi \right), i \in [1: C]$

Общая схема исследования эффективности алгоритма характеристики целевой функции с использованием тестовых задач оптимизации имеет следующий вид.

1. Для каждой из тестовых задач генерируем некоторое число ландшафтных выборок $\mathbb{X}_j, j \in [1:M]$, где $\mathbb{X}_j = \{X_{i,j}, i \in [1:m]\}$ — ландшафтная выборка из m точек, случайным образом расположенных в области поиска D_X .

2. Для каждой из указанных ландшафтных выборок составляем ПКВ и рассчитываем значения компонент вектора ХП $C = (c_1, \dots, c_{|C|})$ на основе выбранных исследуемым алгоритмом значений характеризующей функции. В результате получаем множество $C_r, r \in [1:R]$, точек в пространстве ХП.

3. Назначаем метки h_r векторам C_r и тем самым получаем размеченную выборку $\{(C_1, h_1), \dots, (C_R, h_R)\}$.

4. Некоторым образом оцениваем качество кластерной структуры $(C_r, h_r)_{r=1}^R$ (см. п. 4).

Размеченная выборка $(C_r, h_r)_{r=1}^R$ затем может быть использована для построения приближенного отображения $C \mapsto \hat{h}$, где $\hat{h}$ — метка класса задачи оптимизации, спрогнозированная на основе ХП ландшафтной выборки целевой функции. Качество кластерной структуры выборки $(C_r, h_r)_{r=1}^R$ позволяет судить об эффективности метода характеристики без построения модели классификации и использования спрогнозированных меток $\hat{h}_r, r \in [1:R]$. В качестве метки h_r вектора C_r можно использовать, например, номер тестовой задачи оптимизации, для ландшафтной выборки которой этот вектор вычислен. Набор H используемых меток зависит от выбранного критерия эффективности алгоритма характеристики (см. п. 4).

4. Критерии эффективности алгоритма характеристики

Обозначим T тип векторов ХП:

$$T = \langle n, g(Y), \{Y_1, \dots, Y_n\} \rangle. \quad (11)$$

Здесь n — размерность векторов ХП этого типа; $g(Y)$ — используемая характеризующая функция, где Y — вектор ее аргументов; $\{Y_1, \dots, Y_n\} \subset R^n$ — множество значений аргументов, для которых вычисляем значения функции $g(Y)$, используемые в качестве компонент векторов ХП данного типа.

Например, в качестве компонент вектора ХП можно использовать значения функции $ic(\varepsilon)$

в точках $\varepsilon_1 = 0$ и $\varepsilon_2 = \arg \max_{\varepsilon} ic(\varepsilon)$. Тогда вектор $C = (ic(\varepsilon_1), ic(\varepsilon_2))$ имеет тип $T = \langle 2, ic(\varepsilon), \{\varepsilon_1, \varepsilon_2\} \rangle$.

Пусть указанным образом определен тип T вектора ХП. Тогда запись вида $C \in T$, где C — вектор ХП, означает, что этот вектор имеет тип T . Запись вида $C \in D_C$ значит, что значения компонент этого вектора принадлежат допустимому множеству D_C .

Предлагаем для численной оценки эффективности алгоритма характеристики использовать индекс силуэта [15] качества кластерной структуры выборки $(C_r, h_r)_{r=1}^R$, вычисляемый по формуле

$$S(T) = \frac{1}{R} \sum_{r=1}^R \frac{d_r^{out}(T) - d_r^{in}(T)}{\max(d_r^{out}(T), d_r^{in}(T))} \in [-1; 1], \quad (12)$$

где d_r^{in} — среднее расстояние от вектора C_r принадлежащего рассматриваемому кластеру, до других векторов этого же кластера; d_r^{out} — среднее расстояние от вектора C_r до векторов ближайшего кластера. Максимальное значение величины $S(T)$, равное 1, соответствует кластерной структуре с четко выделенными непересекающимися кластерами.

Индекс силуэта $S(T)$ оценивает как "компактность" кластеров — среднее расстояние между точками C_r одного кластера, так и отдаленность центров этих кластеров друг от друга — среднее расстояние между точками C_r соседних кластеров.

Обозначим D_T некоторое множество типов векторов $C \in D_C$, составленных на основе значений характеризующей функции $ic(\varepsilon)$, $gic(\varepsilon_1, \varepsilon_2)$ или $p(\gamma)$. Наилучшим является алгоритм характеристики, устанавливающий такой тип $T^* \in D_T$ векторов ХП, который обеспечивает максимальное значение индекса силуэта $S(T)$ выборки $(C_r, h_r)_{r=1}^R$:

$$\max_{T \in D_T} S(T) = S(T^*). \quad (13)$$

Набор H используемых меток h_r зависит от выбранного критерия эффективности алгоритма характеристики, устанавливающего тип вектора $C \in D_C$. Определим следующие критерии эффективности алгоритма характеристики.

Критерий 1. Точки, соответствующие векторам C_r , полученным на основе различных ландшафтных выборок данной целевой функции, находятся "близко" друг к другу в пространстве ХП. Здесь метки $h_r \in H^1 = \{h^{1,1}, h^{1,2}, \dots\}$ соответствуют номеру тестовой задачи оптимизации, для ландшафтной выборки которой

вычислен вектор C_r . В этом случае число кластеров соответствует числу тестовых задач оптимизации.

Критерий 2. Точки, соответствующие векторам C_r , полученным на основе различных ландшафтных выборок "похожих" целевых функций, находятся "близко" друг к другу в пространстве ХП. Здесь метки $h_r \in H^2 = \{h^{2,1}, h^{2,2}\}$ обозначают степень мультимодальности целевой функции тестовой задачи оптимизации, для ландшафтной выборки которой вычислен вектор C_r . Метку $h^{2,1}$ ставим в соответствие вектору C_r в случае, если он характеризует ландшафтную выборку целевой функции с малым числом локальных минимумов. Метку $h^{2,2}$ используем в случае, если вектор C_r характеризует ландшафтную выборку с большим числом локальных минимумов. Таким образом, в данном случае выборка $(C_r, h_r)_{r=1}^R$ содержит два кластера.

5. Вычислительный эксперимент

В данном пункте представлены тестовые задачи, использованные для проведения вычислительного эксперимента. Выполнено сравнительное исследование эффективности предложенных алгоритмов характеристики по схеме, описанной в п. 3. Для визуализации результатов оценки эффективности выполнено исследование алгоритмов, устанавливающих тип двумерных векторов ХП.

Для проведения вычислительного эксперимента используем тестовые задачи оптимизации [14]:

- задачи q_1, q_2 с малым числом локальных минимумов в области поиска D_X , составленные на основе тестовых функций Розенброка и Химмельблау соответственно;
- задачи q_3, q_4 с большим числом локальных минимумов в области поиска D_X , составленные на основе тестовых функций Экли и Растригина соответственно.

Ландшафты целевых функций тестовых задач оптимизации $q_1—q_4$ в области поиска D_X изображены на рис. 7 (см. вторую сторону обложки).

Вычисляем значения критериев эффективности алгоритма характеристики, который устанавливает тип T векторов ХП, по следующей схеме.

1. Для каждой из тестовых задач $q_1—q_4$ генерируем $M = 100$ различных ландшафтных выборок, каждая из которых состоит из $m = 100$ точек, случайным образом расположенных в области D_X .

2. Для каждой ландшафтной выборки составляем ПКВ и рассчитываем координаты векторов ХП на основе значений характеризующих функций из табл. 1. Получаем множество $\{C_r, r \in [1:400]\}$ точек в пространстве ХП: по 100 точек для каждой из тестовых задач $q_1—q_4$.

3. Ставим метки из наборов H^1 и H^2 в соответствие векторам C_r , получаем две размеченные выборки: $(C_r, h_r^1)_{r=1}^{400}$ и $(C_r, h_r^2)_{r=1}^{400}$.

4. Рассчитываем значения S_1 и S_2 индекса $S(T)$ для выборок $(C_r, h_r^1)_{r=1}^{400}$ и $(C_r, h_r^2)_{r=1}^{400}$ соответственно. Оцениваем эффективность алгоритма, устанавливающего тип вектора $C = (c_1, \dots, c_{|C|})$, по критерию 1 с использованием величины S_1 , по критерию 2 — с использованием величины S_2 .

Для корректного сравнения эффективности методов анализа карты вариативности на основе эффективности порожденных ими алгоритмов характеристики необходимо оценивать качество классификации задач оптимизации в пространствах ХП высокой размерности. На рис. 8 представлены результаты оценки качества кластерных структур $(C_r, h_r^1)_{r=1}^{400}$ и $(C_r, h_r^2)_{r=1}^{400}$ в 64-мерных пространствах ХП для ландшафтных выборок различного размера.

На рис. 8 видно, что индекс силуэта кластерной структуры, а значит, и эффективность алгоритма характеристики ожидаемо растут вместе с ростом числа точек ландшафтной выборки. Из рис. 8, а видно, что наиболее эффективными по критерию 1 являются алгоритмы характеристики на основе методов обобщенного информационного содержания и секторизации карты вариативности. По критерию 2 эффективность алгоритма на основе метода секторизации карты вариативности значительно превосходит эффективность других сравниваемых алгоритмов (рис. 8, б). Это может свидетельствовать о потенциально большей обобщающей способности алгоритма классификации задач оптимизации в случае использования алгоритмов характеристики на основе метода секторизации карты вариативности.

Двумерный вектор ХП. Для визуальной оценки эффективности классификации задач оптимизации составим двумерные векторы $C = (c_1, c_2)$ на основе значений характеризующих функций $ic(\epsilon)$, $gic(\epsilon_1, \epsilon_2)$ и $p(\gamma)$ (табл. 2)

На рис. 9 (см. третью сторону обложки) в качестве примера для ландшафтной выборки тестовой функции Экли представлены графики соответствующих характеризующих функций. Точками обозначены значения, которые используем в качестве элементов c_1, c_2 вектора ХП.

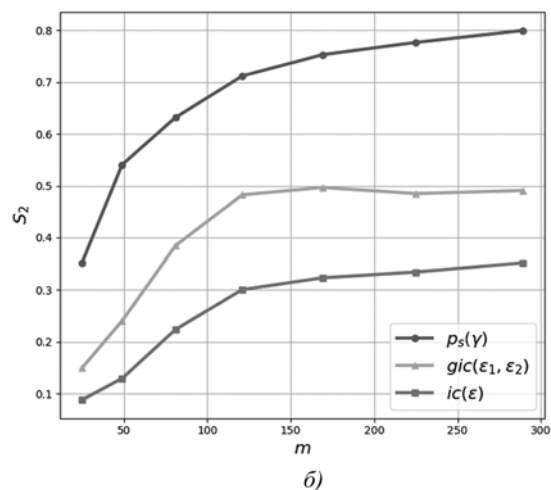
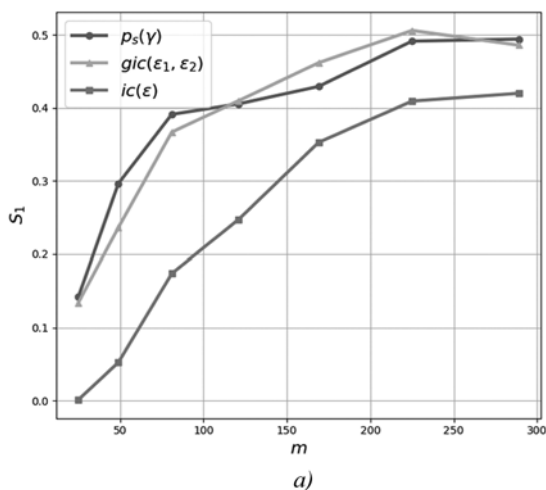


Рис. 8. Зависимости индекса силуэта $S(T)$ кластерных структур $(C_r, h_r^1)_{r=1}^{400}$ (а) и $(C_r, h_r^2)_{r=1}^{400}$ (б) от размера ландшафтной выборки в 64-мерных пространствах ХП

Для каждой сгенерированной ландшафтной выборки рассчитываем значения компонентов c_1, c_2 из табл. 1. Получаем три множества векторов $C_r^{ic}, C_r^{gic}, C_r^p, r \in [1:R]$, соответствующих различным методам анализа карты вариативности. На рис. 10 изображены эти множества вместе с метками $h_r \in H^1$, обозначенными различными цветами.

Из рис. 10, а и б (см. третью сторону обложки) видно, что при использовании векторов C_r^{ic}, C_r^{gic} группы точек, соответствующих ландшафтным выборкам тестовых задач оптимизации q_1 – q_4 , расположены в пространстве ХП в порядке роста степени мультимодальности целевых функций этих задач. При использовании вектора C_r^p (рис. 10, в, см. третью сторону обложки) наблюдаем явное разделение вдоль оси c_1 получившихся кластеров точек по степени мультимодальности целевых функций соответствующих им задач q_1 – q_4 : слева расположены кластеры, соответствующие задачам q_1, q_2 с малым числом локальных экстремумов; справа — кластеры, соответствующие задачам q_3, q_4 с большим числом локальных экстремумов.

Из табл. 2 следует, что наибольшая эффективность классификации оптимизационных задач по типу их целевой функции (критерий 1) достигается при использовании вектора ХП, составленного на основе значений функции $gic(\varepsilon_1, \varepsilon_2)$. В случае, если необходимо классифицировать задачи оптимизации по степени мультимодальности их целевой функции, то наиболее эффективным по критерию 2 является вектор ХП, составленный на основе значений функции $p(\gamma)$. Эффективность метода ландшафтного анализа можно значительно по-

высить, используя большее число точек характеризующих функций и большее число точек ландшафтной выборки.

В табл. 3 представлены значения индекса силуэта $S(T)$ качества кластерных структур, полученных в результате применения меток из наборов H^1 и H^2 к множествам векторов $C_r^{ic}, C_r^{gic}, C_r^p, r \in [1:R]$.

Таблица 2

Формулы оценки значений компонентов вектора ХП

Алгоритм оценки ХП	Элементы вектора ХП	
	c_1	c_2
Алгоритм информационного содержания	$ic(\varepsilon = 0)$	$\max_{\varepsilon} ic(\varepsilon)$
Алгоритм обобщенного информационного содержания	$\max_{\varepsilon_1} gic(\varepsilon_1, \varepsilon_2 = 0)$	$\max_{\varepsilon_2} gic(\varepsilon_1 = 0, \varepsilon_2)$
Алгоритм секторизации карты вариативности	$p(\gamma = \pi/6)$	$p(\gamma = \pi/3)$

Таблица 3

Результаты оценки величины $S(T)$

Алгоритм	Индекс силуэта S	
	метки H_r^1	метки H_r^2
Алгоритм информационного содержания	0,53	0,57
Алгоритм обобщенного информационного содержания	0,55	0,60
Алгоритм секторизации карты вариативности	0,34	0,67

Заключение

Методы ландшафтного анализа используют для оценки ХП целевой функции задачи оптимизации. В работе рассмотрены методы ландшафтного анализа целевой функции на основе карты вариативности ландшафтной выборки. Для анализа карты вариативности предложен метод ее секторизации, позволяющий по сравнению с методами информационного и обобщенного информационного содержания получить оценки более значимых для алгоритма оптимизации особенностей целевой функции. Высокая обобщающая способность алгоритмов классификации оптимизационных задач по наиболее значимым признакам является необходимой основой для интеллектуализации подсистем САПР параметрической оптимизации.

Искомые оценки ХП зависят от расположения точек ландшафтной выборки, используемого метода анализа карты вариативности и алгоритма характеристики, определяющего состав и размерность результирующего вектора ХП. Предложена методика сравнительного исследования эффективности указанных методов анализа карты вариативности на основе оценок эффективности порожденных ими алгоритмов характеристики. Результаты проведенного по данной методике вычислительного эксперимента свидетельствуют об универсальности и высокой эффективности метода секторизации карты вариативности по сравнению с другими методами ее анализа.

В целях дополнительного обоснования сделанных выводов составлены двухмерные векторы ХП на основе значений характеризующих функций рассмотренных методов, что позволило визуализировать полученные результаты в пространстве ХП целевой функции.

Исследование показало, что алгоритмы характеристики целевой функции на основе метода секторизации карты вариативности являются наиболее пригодными для классификации задач оптимизации по различным признакам.

Список литературы

1. Shan S., Wang G. G. Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions // *Structural and Multidisciplinary Optimization*. 2010. Vol. 41, N. 2. P. 219–241. DOI: 10.1007/s00158-009-0420-2.
2. Kerschke P. Comprehensive Feature-Based Landscape Analysis of Continuous and Constrained Optimization Problems Using the R-Package flacco // *arXiv preprint arXiv:1708.05258*. 2017.
3. Агасиев Т. А., Карпенко А. П. Современные техники глобальной оптимизации: Обзор // *Информационные технологии*. 2018. Т. 6, № 24. С. 363–370. DOI: 10.17587/it.24.370-386.
4. Cavazzuti M. Design of experiments // *Optimization Methods*. Springer, Berlin, Heidelberg, 2013. P. 13–42.
5. Li G., Aute V., Azarm S. An accumulative error based adaptive design of experiments for offline metamodeling // *Structural and Multidisciplinary Optimization*. 2010. Vol. 40, N. 1. pp. 137–155. DOI: 10.1007/s00158-009-0395-z.
6. Forrester A. I. J., Keane A. J. Recent advances in surrogate-based optimization // *Progress in Aerospace Sciences*. 2009. Vol. 45, N. 1–3. P. 50–79. DOI: 10.1016/j.paerosci.2008.11.001.
7. Mersmann O., Bischl B., Trautmann H., Preuss M., Weihs C., Rudolph G. Exploratory landscape analysis // *Proceedings of the 13th annual conference on Genetic and evolutionary computation*. ACM, 2011. P. 829–836. DOI: 10.1145/2001576.2001690.
8. Hoos H. H., Neumann F., Trautmann H. Automated Algorithm Selection and Configuration (Dagstuhl Seminar 16412) // *Dagstuhl Reports*. 2017. Vol. 6. P. 34–74.
9. Kerschke P., Preuss M., Hernández C., Schütze O., Sun J. Q., Grimme C., Rudolph G., Bischl B., Trautmann H. Cell mapping techniques for exploratory landscape analysis // *EVOLVE-A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation V*. Springer, Cham. 2014. P. 115–131. DOI: 10.1007/978-3-319-07494-8_9.
10. Flamm C., Hofacker I. L., Stadler P. F., Wolfinger M. T. Barrier trees of degenerate landscapes // *Zeitschrift für physikalische Chemie*. 2002. Vol. 216, N. 2. P. 155. DOI: 10.1524/zpch.2002.216.2.155.
11. Kerschke P., Preuss M., Wessing S., Trautmann H. Detecting funnel structures by means of exploratory landscape analysis // *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*. ACM. 2015. P. 265–272. DOI: 10.1145/2739480.2754642.
12. Агасиев Т. А. Карта Вариативности Целевой Функции Для Анализа Характерных Признаков Задачи Глобальной Оптимизации // *Приборостроение*. 2019 (в печати).
13. Muñoz M. A., Kirley M., Halgamuge S. K. Exploratory landscape analysis of continuous space optimization problems using information content // *IEEE Transactions on Evolutionary Computation*. 2015. Vol. 19, N. 1. P. 74–87. DOI: 10.1109/TEVC.2014.2302006.
14. Momin J., Yang X. S. A literature survey of benchmark functions for global optimization problems // *Journal of Mathematical Modelling and Numerical Optimisation*. 2013. Vol. 4, N. 2. P. 150–194. DOI: 10.1504/IJMMNO.2013.055204.
15. Guerra L., Robles V., Bielza C., Larrañaga P. A comparison of clustering quality indices using outliers and noise // *Intelligent Data Analysis*. 2012. Vol. 16, N. 4. P. 703–715. DOI: 10.3233/IDA-2012-0545.

Sectorization Method of Full Variability Map Analysis for Characteristic Features Assessment of Global Optimization Problems

Landscape analysis methods are designed to assess various characteristic features of objective function of optimization problem. The accuracy of estimations mostly depends on the chosen design of experiment, i.e. on the number and location of points in the search space. The paper considers methods of variability map analysis that allow to obtain estimations that are more stable to variations of the experimental design. The drawback of these methods is the high sensitivity of resulting values of characteristic features to the scale of objective function values variations. This may adversely affect the generalizing ability of the algorithm of problems classification by the most significant objective function features. In relation to this, the paper presents sectorization method of variability map analysis which allows to obtain estimations stable to the scale of objective function changes. A technique of conducting the computational experiment is proposed for comparative research of the efficiency of different methods of variability map analysis. Efficiency criteria for algorithms of characteristic features assessment are formulated. The results of experiment demonstrate relevance of using the proposed methods of landscape analysis for classifying optimization problems by different features. The quality of optimization problems classification largely defines the efficiency of intellectualization of the parametric optimization subsystems in CAD.

Keywords: global optimization, characteristic features of problem, exploratory landscape analysis, landscape sample, information content method, sectorization method, variability map

DOI: 10.17587/it.25.707-717

References

1. Shan S., Wang G. G. Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions, *Structural and Multidisciplinary Optimization*, 2010, vol. 41, no. 2, pp. 219–241. DOI: 10.1007/s00158-009-0420-2.
2. Kerschke P. Comprehensive Feature-Based Landscape Analysis of Continuous and Constrained Optimization Problems Using the R-Package flacco, *arXiv preprint arXiv:1708.05258*, 2017.
3. Agsiev T. A., Karpenko A. P. Modern Technologies of Global Optimization. Review, *Information Technologies*, 2018, vol. 6, no 24, pp. 363–370 (in Russian). DOI: 10.17587/it.24.370-386.
4. Cavazzuti M. Design of experiments, *Optimization Methods*, Springer, Berlin, Heidelberg, 2013, pp. 13–42.
5. Li G., Aute V., Azarm S. An accumulative error based adaptive design of experiments for offline metamodeling, *Structural and Multidisciplinary Optimization*, 2010, vol. 40, no. 1, pp. 137–155. DOI: 10.1007/s00158-009-0395-z.
6. Forrester A. I. J., Keane A. J. Recent advances in surrogate-based optimization, *Progress in Aerospace Sciences*, 2009, vol. 45, no. 1–3, pp. 50–79. DOI: 10.1016/j.paerosci.2008.11.001.
7. Mersmann O., Bischl B., Trautmann H., Preuss M., Weihs C., Rudolph G. Exploratory landscape analysis, *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, ACM, 2011, pp. 829–836. DOI: 10.1145/2001576.2001690.
8. Hoos H. H., Neumann F., Trautmann H. Automated Algorithm Selection and Configuration (Dagstuhl Seminar 16412), *Dagstuhl Reports*, 2017, vol. 6, pp. 34–74.
9. Kerschke P., Preuss M., Hernández C., Schütze O., Sun J. Q., Grimme C., Rudolph G., Bischl B., Trautmann H. Cell mapping techniques for exploratory landscape analysis, *EVOLVE-A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation V*, Springer, Cham, 2014, pp. 115–131. DOI: 10.1007/978-3-319-07494-8_9.
10. Flamm C., Hofacker I. L., Stadler P. F., Wolfinger M. T. Barrier trees of degenerate landscapes, *Zeitschrift für physikalische chemie*. 2002, vol. 216, no. 2, pp. 155. DOI: 10.1524/zpch.2002.216.2.155.
11. Kerschke P., Preuss M., Wessing S., Trautmann H. Detecting funnel structures by means of exploratory landscape analysis, *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*, ACM, 2015, pp. 265–272. DOI: 10.1145/2739480.2754642.
12. Agasiev T. A. Variability map of objective function for analysis of global optimization problem characteristic features. Herald of the Bauman Moscow State Technical University, Series Instrument Engineering, 2019, no. 2, pp. 4–21 (in Russian). DOI: 10.18698/0236-3933-2019-2-4-21.
13. Muñoz M. A., Kirley M., Halgamuge S. K. Exploratory landscape analysis of continuous space optimization problems using information content, *IEEE Transactions on Evolutionary Computation*, 2015, vol. 19, no. 1, pp. 74–87. DOI: 10.1109/TEVC.2014.2302006.
14. Momin J., Yang X. S. A literature survey of benchmark functions for global optimization problems, *Journal of Mathematical Modelling and Numerical Optimisation*, 2013, vol. 4, no. 2, pp. 150–194. DOI: 10.1504/IJMMNO.2013.055204.
15. Guerra L., Robles V., Bielza C., Larrañaga P. A comparison of clustering quality indices using outliers and noise, *Intelligent Data Analysis*, 2012, vol. 16, no. 4, pp. 703–715. DOI: 10.3233/IDA-2012-0545.

А. З. Салахов, аспирант, e-mail: alisalahov@gmail.com
Азербайджанский технический университет, г. Баку

Исследование рассеяния электромагнитных полей в биообъектах методом компьютерного моделирования

За последние годы функциональная возможность мобильных средств связи растет в геометрической прогрессии. В связи с тем, что популярность мобильных телефонов и продолжительность их использования увеличивается с каждым днем, беспокойство по поводу безопасности электромагнитного излучения от них становится все более обоснованным, и изучение влияния электромагнитного излучения на пользователя представляет интерес и научную ценность. Исследование данного влияния связано с созданием адекватной модели биологического объекта и с изучением рассеяния электрических волн в среде с меняющимися электрофизическими параметрами. Приняв во внимание невозможность проведения такого рода исследования непосредственно на человеке, можно говорить о компьютерном моделировании биологической среды. В данном случае речь идет о многослойной модели с произвольными параметрами удельной проводимости, диэлектрической проницаемости и исследовании распространения и поглощения электромагнитного излучения в нем.

Ключевые слова: напряженность электрического поля, удельный коэффициент поглощения, диэлектрическая проницаемость, компьютерное моделирование

Введение

В связи с широким распространением мобильных средств связи, а также многочисленных коммуникационных устройств, входящих в их состав, возрастают опасения по поводу безопасности электромагнитных волн, излучаемых ими. В данной работе проведен анализ влияния электромагнитного излучения на организм человека, рассмотрена предыстория болезней, вызванных мобильными системами связи. Показаны сравнительные графики напряженности электрического поля и удельной поглощаемой мощности внутри слоев модели.

В данной статье исследуется многослойная электродинамическая модель, приближенная к анатомической структуре головы человека. Расчеты рассеяния электромагнитных полей и поглощения электромагнитного поля внутри этой модели выполнены с помощью компьютерного моделирования. Модели, построенные на основе многослойной структуры, схожей с анатомическими особенностями человека, наиболее информативны в плане исследования влияния электромагнитных излучений на человека.

Цель и задачи. Основной целью проведенных исследований является повышение достоверности распределения уровней поглощаемой мощности в биологической среде в результате воздействия электромагнитного поля. Исследование проводится путем разработки адекватных компьютерных и математических моделей биологической среды и изучения воздействия электромагнитного излучения на них.

Особенности исследования рассеяния электромагнитных волн в биообъектах

Следует отметить, что негативное влияние на живые организмы может оказывать как излучение низкого уровня, так и излучения высокого уровня, поглощаемое в сравнительно малых областях биологической ткани с более высокой удельной проводимостью. Антенны сотовых телефонов должны быть хорошо согласованы с передатчиком, обеспечивая эффективное излучение сигнала, в противном случае реактивная мощность поля ближней зоны приведет к росту уровня поглощаемой мощности внутри биологической среды [1]. В связи с этим в настоящее время исследование рассеяния электромагнитных полей и поглощаемой мощности является достаточно актуальной задачей в плане контроля значений поглощаемой мощности внутри тканей человека. Поскольку проведение экспериментальных работ непосредственно на человеке не представляется возможным, необходимо изучение вопроса влияния электромагнитных излучений на биообъекты с использованием компьютерного трехмерного моделирования [2].

В большинстве случаев после длительного использования беспроводных средств связи у владельцев появляются такие жалобы, как головные боли, напряжение в области уха, приступы внезапной усталости, потеря концентрации, нарушение сна и т. д. [3]. Основным эффектом воздействия электромагнитного излучения частотой выше 1 МГц считается эффект нагре-

вания области, подвергнутой излучению [4]. Также предполагается, что электромагнитное излучение способствует изменению потенциалов мембран клеток нервной системы, а это может привести к нарушениям нервной системы [5, 6]. Все эти эффекты возникают вследствие поглощения энергии электромагнитного поля биологической тканью человека.

На сегодняшний день основным параметром, характеризующим уровень поглощения электромагнитной энергии в биологической среде, является SAR (Specific Absorption Rates) — удельный коэффициент поглощения (Вт/кг) [7]:

$$SAR = \frac{\sigma |E|^2}{\rho},$$

где E — напряженность электрического поля; ρ — плотность среды; σ — удельная проводимость. Как видно из формулы, интенсивность поглощения SAR зависит от напряженности электрического поля, плотности среды и ее удельной проводимости.

Для параметра удельного поглощения SAR по международным стандартам установлены нормы для вещества массой 1 и 10 г. Наибольший интерес представляет область головы, так как при использовании мобильных телефонов голова пользователя располагается в непосредственной близости от источника излучения. В технической документации на средства связи указывается значение SAR при работе устройства на максимальной мощности. Надо отметить что уровень SAR в зоне со слабым уровнем приема намного выше, чем в зоне с умеренным уровнем приема сигнала от базовой станции.

В работе [8] приведены результаты эксперимента при участии добровольцев, в котором проводилось исследование по выявлению влияния на человека сотового телефона в зависимости от его расположения по отношению к человеческому телу. Некоторые эффекты воздействия, к примеру головная боль, были отмечены уже после нескольких минут использования при уровне SAR свыше 0,5 Вт/кг.

В СВЧ диапазоне магнитные свойства среды настолько малозначительны, что их можно не учитывать.

Исследование данного влияния опирается на создание адекватной модели биологического объекта и изучение рассеяния электрических волн, а также их возбуждения в среде с меняющимися электрофизическими параметрами [9]. Для наиболее детального исследования влияния электромагнитных волн,

излучаемых средствами мобильной связи, была разработана максимально приближенная к анатомическим свойствам человека шестислойная модель головы (рис. 1).

Наиболее значимые электрофизические параметры биологической ткани для исследуемой шестислойной модели головы человека представлены в табл. 1.

Простой расчет электромагнитного поля внутри биологической среды в основном выполняется методом компьютерного моделирования этой среды с различными значениями диэлектрической проницаемости и удельной проводимости [10].

Взаимодействие многослойной структуры с электромагнитной волной можно сформулировать следующим образом. Представим, что в системе координат X, Y, Z расположена многослойная структура со слоями определен-

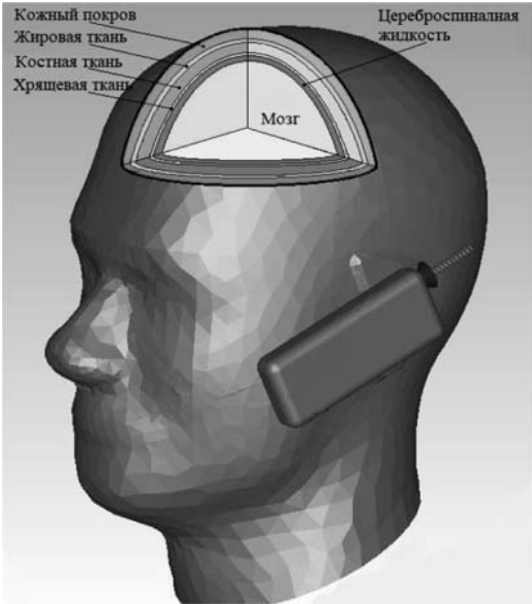


Рис. 1. Шестислойная модель головы человека

Таблица 1
Электрофизические параметры слоев модели

Тип ткани	Граничный радиус модели	Диэлектрическая проницаемость ϵ_r , Ф/м	Плотность тока j , А/м ²
Кожный покров	9,0	41,11	0,327
Жировая ткань	8,9	79,1	0,423
Костная ткань	8,76	40,7	0,25
Хрящевая ткань	8,35	20,9	0,247
Цереброспинальная жидкость	8,3	10	0,266
Мозговая ткань	8,1	40,7	0,25

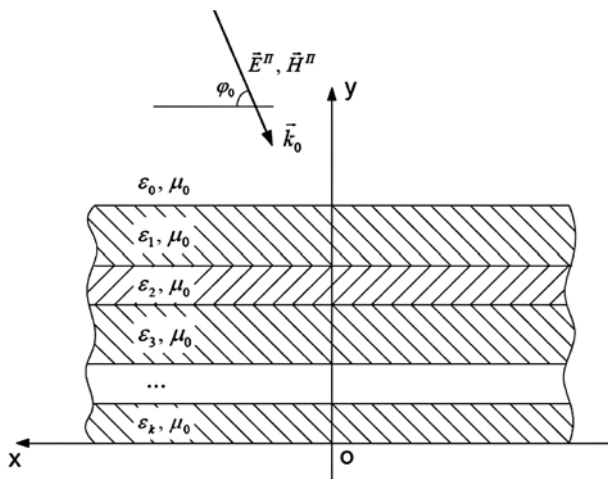


Рис. 2. Иллюстрация взаимодействия электромагнитной волны с многослойной структурой

ной толщины, каждый из которых состоит из изотопного однородного диэлектрика с диэлектрической проницаемостью ϵ_i и магнитной проницаемостью μ_0 . На эту структуру из пространства воздействует электромагнитная волна.

Задача заключается в определении зависимости показателей отражения и проникновения электромагнитных волн в среду с различными электрофизическими свойствами и геометрическими параметрами, а также в анализе распределения вектора напряженности электрического поля (рис. 2).

Расчет напряженности электрического поля и удельной поглощаемой мощности в исследуемой среде

При условии произвольного изменения удельной проводимости и диэлектрической проницаемости среды расчет электрического тока поляризации с объемной плотностью проводится универсальным методом интегральных уравнений. При линейной поляризации падающего поля $\vec{E}^II = \vec{i}_z \vec{E}_z^II$ в области пространства V определенного слоя объемная плотность электрического тока может быть записана в виде

$$\vec{J}^{\text{Эпол}} = i\omega\epsilon_0(\epsilon - 1)\vec{E},$$

где $\vec{J}^{\text{Эпол}}$ — объемная плотность электрического тока; $\vec{E}$ — напряженность электрического поля; ϵ — диэлектрическая проницаемость материала; ω — частота электромагнитных колебаний:

$$\epsilon = \epsilon' - i\epsilon'', \quad \epsilon'' = \frac{\sigma}{\epsilon_0\omega},$$

σ — удельная проводимость среды; ϵ' — вещественная, ϵ'' — мнимая части комплексной диэлектрической проницаемости.

Вторичное электрическое поле, создаваемое токами поляризации, можно записать в виде

$$E^p(p) = -\frac{1}{4\omega\epsilon_0} \iint J^{\text{Эпол}}(p') H_0^{(2)}(p, p') dx' dy'.$$

Если добавить первичное (внешнее) поле и заменить ток поляризации, то получим интегральное уравнение относительно напряженности электрического поля в определенном объеме материала:

$$E_z(p) + \frac{i}{4} \iint (\epsilon(y') - 1) E_z(x', y') \times \\ \times H_0^{(2)}(x', y, x, y) dx' dy' = E_z^II(x, y).$$

Зависимость поля от координаты x при падении волны на объект исследования можно выразить как простую экспонентную

$$E(x, y) = E(x) \exp(i\beta x),$$

где $\beta = \cos\phi$. Применив интегральную функцию Хенкеля

$$H_0^{(2)}(p) = \frac{i}{\pi} \int_{\chi=-\infty}^{\infty} \frac{\exp(-i\chi(x-x') - \sqrt{\chi^2 - 1}|y-y'|) d\chi}{\sqrt{\chi^2 - 1}},$$

где χ — диэлектрическая восприимчивость (способность вещества поляризоваться под действием электрического поля), получим интегральное уравнение для напряженности электрического поля в слоях относительно оси y :

$$E_z(y) - \frac{1}{2} \int_y (\epsilon(y') - 1) \frac{\exp(\sqrt{\beta^2 - 1}|y - y'|)}{\sqrt{\beta^2 - 1}} dy'.$$

В таком случае для коэффициентов отражения электромагнитной волны R и его преодоления T можем записать следующие выражения:

$$R = -\frac{i}{2\sin\phi_0} \int_y (\epsilon(y') - 1) \vec{E}_z(y') \exp(i\sin\phi_0 y') dy';$$

$$T = -\frac{i}{2\sin\phi_0} \int_y (\epsilon(y') - 1) \vec{E}_z(y') \exp(-i\sin\phi_0 y') dy' + 1.$$

Таким образом, решив интегральные уравнения, можем рассчитать уровни коэффициентов поглощаемой мощности, распределение напряженности электрического поля в тех или иных точках, находящихся в структуре ис-

следуемой среды, а также значения коэффициентов преодоления и отражения. Плотность мощности, поглощенной в ткани за счет колебаний дипольных молекул и наличия ионной проводимости, можно определить так:

$$p_t = \frac{\sigma}{2} |E|^2 = \frac{\omega \epsilon''}{2} |E|^2.$$

Используя значения удельной проводимости и диэлектрической проницаемости слоистой структуры в определенном диапазоне частот, можно исследовать свойства структуры модели, имитирующей биологическую ткань (кожный покров, жировая ткань, костная ткань, хрящевая ткань цереброспинальная жидкость, мозговая ткань). Выявлено, что значения удельной поглощаемой мощности в ткани зависят от изменения в ней концентрации жидкости. По причине изменения концентрации жидкости в ткани изменяется также значение ее удельной проводимости, которое также оказывает заметное влияние на уровень напряженности электрического поля в ткани. Поскольку плотность поглощаемой мощности зависит от наличия потерь в диэлектрике, она пропорциональна удельной проводимости ткани, а также является функцией напряженности электрического поля.

На рис. 3 показаны графики уровней плотности поглощаемой мощности и напряженности электрического поля для многослойной среды при плотности потока мощности внешней волны 1 мВт/см². Из кривых на графике видно, что при проводимости жировой ткани $\sigma = 147$ См/м, соответствующей верхней границе проводимости этой ткани, наибольшее поглощение мощности происходит в поверхностных жировых тканях, а при значении удельной проводимости $\sigma = 55,6$ См/м, соответствующей нижней границе проводимости для радиоча-

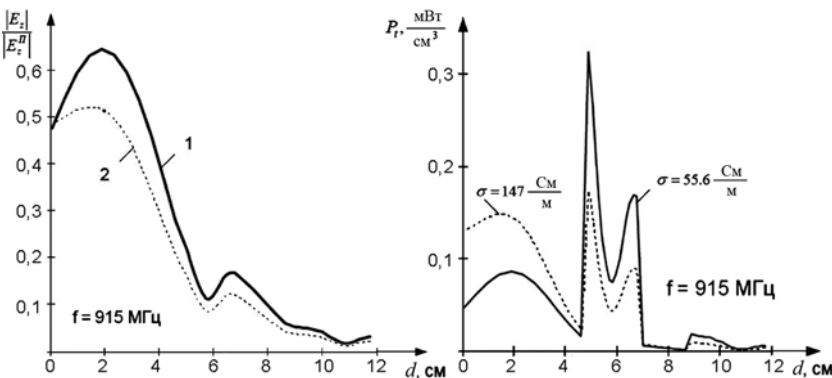


Рис. 3. Рассеяние удельной поглощаемой мощности и напряженности электрического поля в многослойной биологической среде

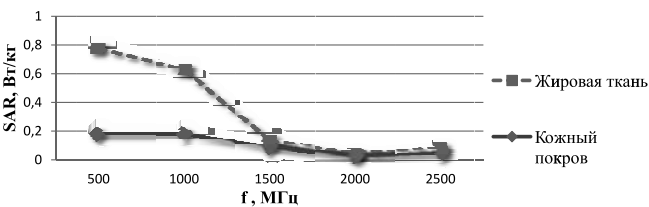


Рис 4. Зависимость SAR от частоты в первых двух слоях биологической ткани

стоты $f = 915$ МГц, основное поглощение происходит в цереброспинальной жидкости.

Это объясняется тем, что при исследовании многослойной структуры максимальное значение удельной поглощаемой мощности всегда достигается в слоях, для которых характерно высокое содержание жидкости. Также отметим, что коэффициент удельной поглощаемой мощности в различных тканях может изменяться в довольно широких пределах частот излучения.

На рис. 4 показаны зависимости коэффициента удельной поглощаемой мощности от частоты для представляющих наибольший интерес первых двух слоев шестислойной модели головы человека (кожи и жировой ткани).

Результаты проведенных расчетов, полученные с помощью компьютерного моделирования

Наиболее полное представление о распределении напряженности электрического поля и SAR в конкретном слое электрофизической модели может дать график, характеризующий распределение SAR в выбранном слое сферического образования для диапазона углов θ и ϕ от 0 до 180° в сферической системе координат (рис. 5).

Данные зависимости для первых двух слоев многослойной модели приведены на рис. 6—9 (рис. 6, 8 см. четвертую сторону обложки). Расчетные данные получены для частоты 915 МГц.

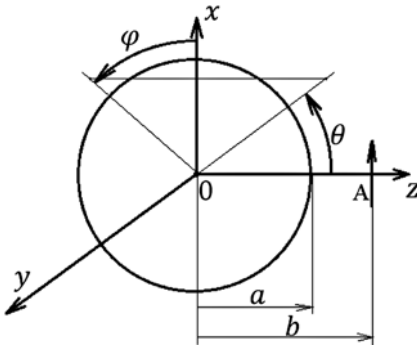


Рис. 5. Модель в сферической системе координат: А — вибраторная антенна; а — радиус модели; b — расстояние до антенны; θ и ϕ — углы в системе координат

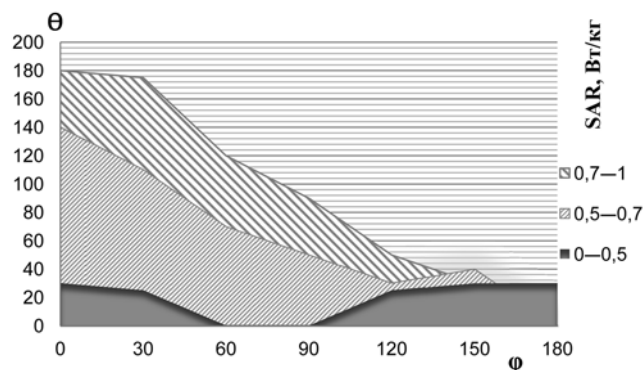


Рис. 7. Карта уровней распределения SAR в первом слое (кожа) для диапазона углов θ и ϕ от 0 до 180°

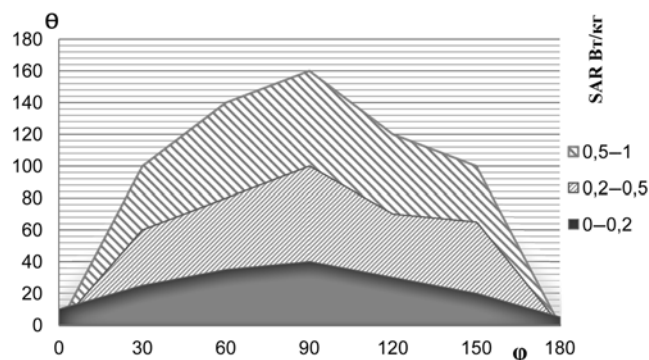


Рис. 9. Карта уровней распределения SAR во втором слое (жировая ткань) для диапазона углов θ и ϕ от 0 до 180°

Как видно из графиков, значение SAR меняется в зависимости от расположения антенны по отношению к модели. Также отметим, что удаление источника излучения от модели также существенно меняет картину рассеяния SAR и уровни ее пиковых значений.

В исследованиях данного типа используется моделирование с помощью многослойной структуры. Данное моделирование имеет особое преимущество и широко применяется для выполнения оценки рассеяния SAR [11]. В работе проведено исследование с использованием шестислойной модели головы человека. Проведена оценка рассеяния SAR с учетом изменения угла излучения электромагнитной волны. Расчеты рассеяния электромагнитных волн и рас-

Таблица 2

Значения SAR в зависимости от расположения антенны

SAR	Расстояние расположения антенны и модели головы, см			
	0	1	2	5
Максимальное SAR на 10 г/см ³ , Вт/кг	0,512	0,449	0,319	0,142
Усредненное SAR на 10 г/см ³ , Вт/кг	0,099	0,075	0,054	0,028

пределения SAR выполнено программой FEKO. Данная модель, а также результаты значений SAR в различных сечениях модели показаны на рис. 10 (см. третью сторону обложки).

В табл. 2 приведены результаты расчетов усредненного значения максимального поглощения SAR в 10 г/см³ биологической ткани при различных позициях расположения источника излучения по отношению к биообъекту.

Из полученных результатов видно, что на расстоянии 5 см от объекта (что соответствует 0,15 λ длины волны) значение SAR уменьшается в максимальной степени, а именно значение SAR уменьшается в 3,6 раза по сравнению с тем случаем, когда антенна размещена вблизи от головы пользователя. Среднее значение SAR уменьшилось в 3,5 раза для такого же случая.

Заключение

Таким образом, в работе представлены результаты расчета для модели головы человека, приближенной к анатомическим особенностям человека по форме и электрофизическим параметрам в виде сложной многослойной структуры, позволяющей получить достоверные результаты рассеяния полей, а также значения SAR в различных областях модели. Полученные результаты показывают, что максимальное значение удельной поглощаемой мощности при использовании мобильного телефона ниже порогового значения, установленного международными стандартами. Также можно отметить, что увеличив расстояние между сотовым телефоном и пользователем, можно ощутимо снизить уровень поглощения электромагнитного излучения. Удаление телефона от головы на 5 см приводит к снижению как усредненного, так и максимального значения поглощения более чем в 3 раза. Нужно отметить, что при этом не происходит смещение областей, где достигаются максимальные значения SAR.

Результаты данного исследования могут быть применены в разработках средств мобильной связи, благодаря чему можно достаточно достоверно прогнозировать распределение внутренних электромагнитных полей, а также оптимизировать конфигурацию антенны сотового телефона для получения минимального уровня излучений.

Полученные данные показывают, что с помощью разработки методов компьютерного моделирования взаимодействия электромагнитных волн с биообъектами можно опреде-

лить влияние высокочастотных излучений на организм человека. Исследуемый метод создает базу и для будущих теоретических и практических исследований в области предотвращения влияния электромагнитных излучений от мобильных телефонов на организм человека.

Список литературы

1. Салахов А. З. Антенны сотовых телефонов как излучатель // Петербургский журнал электроники. 2017. № 4 (89). С. 35—40.
2. Салахов А. З. Вспомогательные методы для вычисления поглощения мощности при высокочастотном излучении с помощью моделирования головы человека // Биомедицинская радиоэлектроника. 2018. № 11. С. 32—39.
3. Вершинин А. Е., Авдонина Л. А. Влияние сотовых телефонов на здоровье человека // Вестник Пензенского государственного университета. 2015. № 3 (11). С. 175—179.
4. Жаворонков Л. П., Петин В. Г. Влияние электромагнитных излучений сотовых телефонов на здоровье // Радиация и риск. 2016. Т. 25, № 2. С. 43—56.
5. Гайлус Т. Г. Влияние электромагнитного поля мобильного телефона на биоэлектрическую активность мозга человека // Биомедицинская радиоэлектроника. 1998. № 4. С. 36—42.

6. Горшунова Н. К., Тарасов А. Н. Влияние ЭМИ и ЭМИ мобильных телефонов на процессы старения и вегетативную нервную систему // Фундаментальные исследования. 2005. № 7. С. 84—85.

7. Адел Зин Елдин Мохамед Моусса. Контроль параметров мобильных средств связи по распределению удельной массовой мощности поглощения излучения телом человека: автореф. дис. канд. техн. наук. Казань: КГЭУ, 2005. 14 с.

8. Wilen J., Sandstrom M., Hansson Mild K. Subjective symptoms among mobile telephone users—a consequence of absorption of radiofrequency fields // Bioelectromagnetics. University of Washington, Seattle, Washington, 2003. P. 152—159.

9. Beard B. B., Kainz W., Onishi T. et al. Comparisons of Computed Mobile Phone Induced SAR in the SAM Phantom to that in Anatomically Correct Models of the Human Head // IEEE Transaction on Electromagnetic Compatibility. 2006. Vol. 48, N. 2. P. 397—407.

10. Keshvari J. et al. The Effect of Increase in Dielectric Values on Specific Absorption Rate (SAR) in Eye and Head Tissues Following 900, 1800 and 2450 MHz Radio Frequency (RF) Exposure // Physics in Medical and Biology. 2006. Vol. 51, N. 6. P. 1463—1477.

11. Amer I. A., Abdalgabar O. J., Vdovina N. V. Investigation of the Effects of Electromagnetic Radiation from a Cell Phone Antenna on a Human Head Model // Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics. 2018. Vol. 18, N. 1. P. 163—170.

A. Z. Salahov, Postgraduate Student, e-mail: alisalahov@gmail.com,
Azerbaijan Technical University, Baku

The Study of the Scattering of Electromagnetic Fields in Biological Objects by Computer Simulation

In recent years, the functionality of mobile communications is growing exponentially. Due to the fact that the popularity of mobile phones and the increase in the duration of their use is growing every day, concerns about the safety of electromagnetic radiation from them become more and more justified. Given all this research, the effect of electromagnetic radiation on a user is of particular interest and scientific value. The study of this effect is associated with the creation of an adequate model of a biological object and the study of the scattering of electric waves in a medium with varying electrophysical parameters. Taking into account the impossibility of conducting this kind of research directly on a person, it is a matter of talking about computer modeling of the biological environment. In this case, we are talking about a multilayer model with arbitrary parameters of specific conductivity, dielectric constant and the study of the propagation and absorption of electromagnetic radiation in it.

Keywords: electric field strength, specific absorption coefficient, dielectric constant, computer simulation

DOI: 10.17587/it.25.718-723

References

1. Salahov A. Z. Cell phone antennas as a radiator, *Peterburgskij Zhurnal Jelektroniki*, 2017, no. 4 (89), pp. 35—40 (in Russian).
2. Salahov A. Z. Auxiliary methods for calculating the power absorption in high-frequency radiation using human head modeling, *Biomeditsinskaja radiojelektronika*, 2018, no. 11, pp. 32—39 (in Russian).
3. Vershinin A. E., Avdonina L. A. Influence of cell phones on human health, *Vestnik Penzenskogo gosudarstvennogo universiteta*, 2015, no. 3 (11), pp. 175—179 (in Russian).
4. Zhavoronkov L. P., Petin V. G. The Effect of Electromagnetic Radiation of Cell Phones on Health, *Radiacija i Risk*, 2016, vol. 25, no. 2, pp. 43—56 (in Russian).
5. Gajlus T. G. The influence of the electromagnetic field of a mobile phone on the bioelectric activity of the human brain, *Biomeditsinskaja Radiojelektronika*, 1998, no. 4, pp. 36—42 (in Russian).
6. Gorshunova N. K., Tarasov A. N. The influence of EMI and EMI of mobile phones on the aging process and the autonomic nervous system, *Fundamental'nye issledovaniya*, 2005, no. 7, pp. 84—85 (in Russian).

7. Adel Zin Eldin Mohamed Moussa. Control of parameters of mobile communication devices on the distribution of the specific mass absorption power of radiation by the human body, *avtoref. dis. kand. tehn.*, KGJEU. Kazan', 2005, 14 p. (in Russian).

8. Wilen J., Sandstrom M., Hansson Mild K. Subjective symptoms among mobile telephone users—a consequence of absorption of radiofrequency fields, *Bioelectromagnetics*, University of Washington, Seattle, Washington, 2003, p. 152—159.

9. Beard B. B., Kainz W., Onishi T. et al. Comparisons of Computed Mobile Phone Induced SAR in the SAM Phantom to that in Anatomically Correct Models of the Human Head, *IEEE Transaction on Electromagnetic Compatibility*, 2006, vol. 48, no. 2, pp. 397—407.

10. Keshvari J. et al. The Effect of Increase in Dielectric Values on Specific Absorption Rate (SAR) in Eye and Head Tissues Following 900, 1800 and 2450 MHz Radio Frequency (RF) Exposure, *Physics in Medical and Biology*, 2006, vol. 51, no. 6, pp. 1463—1477.

11. Amer I. A., Abdalgabar O. J., Vdovina N. V. Investigation of the Effects of Electromagnetic Radiation from a Cell Phone Antenna on a Human Head Model, *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*, 2018, vol. 18, no. 1, pp. 163—170.

А. Г. Абрамов^{1, 2}, канд. физ.-мат. наук, доц., зам. директора, e-mail: abramov@runnet.ru,
А. В. Евсеев², директор, e-mail: evseev@runnet.ru,

¹ Санкт-Петербургский политехнический университет Петра Великого,

² Филиал Федерального государственного автономного образовательного учреждения дополнительного профессионального образования "Центр реализации государственной образовательной политики и информационных технологий" в г. Санкт-Петербурге

Концептуальные аспекты создания в Российской Федерации национальной исследовательской компьютерной сети нового поколения

Обсуждается современное состояние мировых научно-образовательных сетей, приводятся их ключевые характеристики, функции и решаемые задачи, даются краткая историческая справка и общие сведения о текущем уровне развития научно-образовательных сетей в России, обосновывается необходимость создания в стране единой научно-образовательной сети с концептуальным представлением ее целей и основных функций с учетом национальных приоритетов в сфере науки и образования, рассмотрением архитектурных компонентов сети и некоторых ожидаемых результатов.

Ключевые слова: национальная научно-образовательная сеть, NREN, НИКС, RUNNet, телекоммуникационная инфраструктура, научная сетевая связность, сетевые сервисы, цифровая платформа

Введение

Наличие телекоммуникационной инфраструктуры специального назначения, которая соответствует передовым достижениям отрасли, используется в интересах сферы науки и образования, обеспечивая эффективное сетевое взаимодействие целевых пользователей, связность с зарубежными научно-образовательными сетями, доступ в глобальное информационно-телекоммуникационное пространство, является одной из значимых характеристик уровня экономического и технологического развития государства и стратегически необходимым условием для стабильного прогресса и выхода на мировой уровень достижений в самых разных областях науки и технологий [1—4].

Результативное использование информационно-коммуникационных технологий (ИКТ) в научных исследованиях, научно-технических проектах и в образовательном процессе, способное внести значимый вклад в решение задачи построения в стране цифровой экономики, невозможно без сбалансированного развития компонентов национальной цифровой

платформы (ЦП), среди которых особая роль отводится современной распределенной инфраструктуре для хранения и обработки научных данных, высокоскоростной сети передачи данных, предназначенной для обеспечения передачи и прозрачного доступа к ним, а также экосистеме прикладных сервисов для хранения, обработки, анализа, передачи, поиска, публикации и других операций с данными [4—6].

В Указе Президента от 7 мая 2018 г. № 204 "О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года" обозначен ряд задач, решение которых непосредственно связано с оперативным внедрением и повсеместным использованием цифровых технологий, созданием в стране передовой инфраструктуры научных исследований и разработок, использующей глобальную инфраструктуру высокоскоростной передачи, обработки и хранения больших объемов данных.

К основным направлениям реализации государственной политики в области научно-технологического развития в реализуемой Стратегии научно-технологического развития Российской Федерации (СНТР) отнесено соз-

дание условий для проведения исследований и разработок, соответствующих современным принципам организации научной, научно-технической и инновационной деятельности и лучшим практикам. Это направление в том числе обеспечивается посредством развития инфраструктуры и поддержки функционирования центров коллективного пользования научно-технологическим оборудованием (ЦКП), поддержки создания и развития уникальных научных установок (УНУ), в том числе класса "мегасайенс", крупных исследовательских инфраструктур на территории страны, доступа исследовательских групп к национальным и зарубежным информационным ресурсам, участия российских ученых в международных проектах, обеспечивающих доступ к новым компетенциям и ресурсам исходя из национальных интересов страны.

Решение целой группы задач, достижение результатов и ключевых показателей принятых к исполнению в 2018 г. национальных проектов (НП), в первую очередь, национального проекта "Наука" и связанных с ним федеральных проектов, требует прорывного развития и широкого использования инфраструктурных и сервисных возможностей специализированной ЦП — отраслевой сети передачи данных как одного из компонентов Единой цифровой платформы науки и высшего образования Минобрнауки России (ЕЦП).

Данная ЦП обеспечит интеграцию существующих и создаваемых для решения задачи по развитию передовой инфраструктуры научных исследований и разработок установок класса "мегасайенс" и других УНУ, ЦКП, а также планируемых к созданию во исполнение планов развития научной и научно-производственной кооперации научно-образовательных центров (НОЦ), научных центров мирового уровня (НЦМУ), выполняющих исследования и разработки по приоритетам НТР, в том числе в целях хранения, анализа и обмена научно-технической информацией в процессе проведения научных исследований и разработок и при реализации совместных проектов.

Накопленный мировой опыт в области создания и совершенствования глобальных телекоммуникационных сетей подчеркивает особую роль в этом процессе инфраструктурных решений, функционирующих в интересах науки и образования. В большинстве стран мира в разные временные периоды были созданы и поступательно модернизируются ответственные за решение обозначенных выше и отдельных локальных задач специализированные отраслевые сети, за которыми закреплен

общепризнанный термин — *национальная научно-образовательная сеть* (англ. — National Research and Education Network, NREN).

Положением о Министерстве науки и высшего образования Российской Федерации к сфере его деятельности отнесены в том числе функции по оказанию государственных услуг и управлению государственным имуществом в области высшего образования, научной, научно-технической и инновационной деятельности, включая деятельность *национальной исследовательской компьютерной сети нового поколения* (далее — НИКС).

НИКС должна выступать в роли национальной научно-образовательной сети России и выполнять обсуждаемые далее функции NREN. На текущем этапе развития мировых научно-образовательных сетей НИКС следует рассматривать не только и не столько в качестве отраслевой сети передачи данных, а как современную инфраструктурно-сервисную ЦП, способствующую выполнению задач, достижению целей и национальных приоритетов, определенных в документах стратегического планирования, повышению эффективности и результативности научной и образовательной деятельности.

1. Научно-образовательные сети в России и в мире

1.1. Современное состояние мировых научно-образовательных сетей. Основные характеристики и функции

Сегодня NREN существуют (находясь на разных ступенях развития) в большинстве стран мира в качестве неотъемлемой части национальной ИКТ-инфраструктуры, обычно координируются государственными органами управления наукой и образованием и представляют страну в международных проектах, при реализации которых интенсивно используются современные средства телекоммуникаций, сетевые технологии, а также инфраструктурные и прикладные сервисы.

В иерархии телекоммуникационных сетей науки и образования национальная научно-образовательная сеть:

- взаимодействует с зарубежными NREN и наднациональными сетевыми консорциумами;
- управляет научно-образовательной сетью национального масштаба;
- осуществляет верхнеуровневое присоединение региональных научно-образовательных сетей;

- предоставляет услуги локальным сетям целевых пользователей (научных и образовательных организаций).

Характерные особенности научно-образовательных сетей в значительной степени определяются потребностями их целевых пользователей [2, 5, 6]:

- *повышенные требования к параметрам качества обслуживания QoS* (высокоскоростная сетевая связность для научных установок класса "мегасайенс", генераторов больших данных, сетей суперкомпьютерных центров; готовность к сверхвысоким пиковым нагрузкам; надежность передачи данных);
- *адаптируемость* (прозрачный доступ пользователей к интерфейсам управления инфраструктурой в целях ее оперативной настройки под текущие проекты BoD, L2/L3 VPN и др.);
- *инновационность* (NREN как платформа для разработки, апробации и внедрения новых перспективных ИКТ и сервисов);
- *поддержка глобальных проектов* (оперативное взаимодействие с зарубежными NREN для участия в крупных международных ИТ-проектах, реализуемых в интересах науки и образования — GNA, edu-iX, EOSC, EUDAT, EGI, PRACE, OpenAIRE и др.);
- *федеративная и межфедеративная аутентификация и авторизация* (внедрение технологий аутентификации и авторизации, предоставляющих пользователям повсеместный прозрачный доступ к ресурсам на базе согласованных на международном уровне механизмов — eduGAIN, InCommon, RA21, AARC, ResearcherID, ORCID, umbrellaID и др.);
- *мобильность* (свободный и прозрачный авторизованный доступ пользователей к ресурсам NREN и собственным сетевым ресурсам и сервисам независимо от местоположения с использованием Wi-Fi роуминга);
- *обучение и обмен опытом* (в рамках регулярно проводимых тематических научно-технологических конференций, обучающих семинаров и вебинаров).

Традиционными функциями и решаемыми задачами NREN являются [5, 7]:

- *организация национальной сетевой связности* (создание и обеспечение функционирования телекоммуникационной сети национального масштаба, связывающей подключенных к ней пользователей внутри страны);
- *взаимодействие с зарубежными NREN и сетевыми консорциумами* (обеспечение международной научной связности в партнерстве с зарубежными NREN и наднациональными сетевыми консорциумами);

- *сотрудничество с региональными научно-образовательными сетями* (организация верхнеуровневой связности участников региональных сетей; совместное управление на уровне принятия решений в целях снижения финансовых затрат за счет консолидации ресурсов и контроля расходов);
- *организация связности с коммерческим сегментом сети Интернет* (обеспечение централизованного доступа в сеть Интернет, снижение затрат пользователей на платные услуги и сервисы, привлечение внебюджетных доходов);
- *квалифицированная поддержка пользователей* (функционирование оператора NREN в качестве центра компетенций в области ИКТ, поддержка пользователей по предоставляемым услугам и сервисам);
- *управление операционными аспектами сервисов* (обеспечение жизненного цикла используемых в сети программно-аппаратных решений, техническое обслуживание и модернизация);
- *управление финансами* (финансовое планирование, управленческий учет и планирование на основе анализа ключевых показателей деятельности и достигнутых результатов, специальное ценообразование);
- *публичность и продвижение* (представление интересов страны в международных сетевых консорциумах и проектах, адресная коммуникация с текущими и потенциальными пользователями).

Среди наиболее развитых на сегодня NREN мира можно выделить отраслевые телекоммуникационные сети США — Internet2, ESnet (Energy Sciences Network) и NISN (NASA Integrated Services Network), Канады (CANARIE), Великобритании (JANET), Германии (DFN), Франции (RENATER), Италии (GARR), Испании (RedIris), Нидерландов (SURFnet), Швейцарии (SWITH), Австралии (AARNet), Китая (CERNET), Японии (SINET), Кореи (KOREN) и др.

В целях повышения эффективности реализации международных проектов и расширения взаимной кооперации в разные периоды времени были созданы наднациональные сетевые объединения в виде консорциумов научно-образовательных сетей, масштабные из которых — это GÉANT (Европа), NORDUnet (Скандинавия), RedClara (Латинская Америка), Asia@Connect/TEIN и APAN (Азиатско-Тихоокеанский регион), AfricaConnect (Африка).

Глобальный сетевой консорциум GÉANT (Gigabit European Academic Network Technology) представляет собой общеевропейскую высокоскоростную опорную транспортную IP-сеть

(100 Гбит/с), функционирующую в интересах научно-образовательного сообщества и обеспечивающую связность более 10 тыс. образовательных и научных организаций из 43 стран Европы и Западной Азии; число пользователей превышает 50 млн человек [7, 8]. GÉANT развивает широкий спектр сетевых сервисов, включая транзит IP-трафика и организацию выделенных сетей, виртуальные и облачные ресурсы и сервисы, сервисы федеративной аутентификации, службы коммуникаций в режиме реального времени, средства мониторинга и устранения неполадок при работе сети и др.

Следует отметить, что цели, задачи и функции NREN в процессе эволюции могут уточняться, расширяться и трансформироваться, откликаясь на возрастающие потребности сферы науки и образования, оперативно реагируя на интенсивное развитие средств телекоммуникаций, лавинообразный рост объемов генерируемых, передаваемых и обрабатываемых научных данных, разработку и широкое внедрение новых высокотехнологичных ИКТ и сервисов.

1.2. Историческая справка и современное состояние научно-образовательных сетей в Российской Федерации

В Российской Федерации с начала 1990-х гг. при адресной государственной поддержке в рамках исполнявшихся федеральных целевых программ, государственных научных программ, межведомственных программ и проектов создавались отраслевые телекоммуникационные сети, используемые в нуждах области науки и образования [2, 9–11].

Научно-образовательное телекоммуникационное пространство исторически строилось в виде ряда независимых сетей, в том числе проблемно-ориентированных, которые в значительной степени базировались на собственной сетевой инфраструктуре и международных каналах. Сети по сути решали и продолжают решать аналогичные задачи, при этом имеет место дублирование сетевой инфраструктуры и бюджетного финансирования.

По состоянию на 2019 г. относительно полноценно функционируют лишь наиболее крупные научно-образовательные сети — федеральная университетская компьютерная сеть RUNNet (Russian UNiversity Network), сеть системы региональных отделений РАН RASNet (Russian Academy of Sciences Network) и отраслевая сеть НИЦ "Курчатовский институт", связывающая организации, кото-

рые специализируются в области ядерной физики.

Созданная в 1994 г. сеть RUNNet [1–4, 10] имеет точки присутствия в десятках российских субъектов, напрямую предоставляя услуги более 100 организациям высшего образования и науки, среди которых — университеты особого статуса, федеральные, национальные исследовательские и опорные университеты, а также ряд научно-исследовательских институтов, организаций культуры и здравоохранения в Москве и Санкт-Петербурге.

Магистральная инфраструктура RUNNet объединяет опорные узлы и каналы связи между ними, формирующие опорную сеть, которая простирается от Амстердама до Новосибирска, а ее работу обеспечивают международные, федеральные и региональные телекоммуникационные узлы.

Федеральные узлы расположены в Москве и Санкт-Петербурге, объединяются внутри городов высокоскоростными каналами связи с разветвленной инфраструктурой и полным резервированием. Узлы между городами связаны четырьмя каналами (4×10 Гбит/с), организованными на базе физически независимых магистралей. Федеральные узлы связаны магистральными каналами с региональными узлами, узлами доступа, с научно-образовательными сетями, сетями коммерческих операторов и сетью Интернет.

Региональные узлы RUNNet расположены в городах всех федеральных округов (в том числе в Барнауле, Екатеринбурге, Нижнем Новгороде, Новосибирске, Перми, Самаре, Саратове, Томске, Уфе, Хабаровске, Челябинске), представляют собой магистральные узлы с размещенным на них телекоммуникационным оборудованием и используются для подключения расположенных в субъектах образовательных и научных организаций. Каналы связи между узлами имеют пропускную способность от 1 до 10 Гбит/с.

Функционирование магистральной инфраструктуры RUNNet за пределами страны основывается на 25-летнем сотрудничестве с сетью NORDUnet (используемой для доступа к ресурсам зарубежных NREN) [6] и налаженном сетевом взаимодействии с ведущими Tier1-операторами (для доступа в публичные сети и сеть Интернет, 30 Гбит/с в сумме). Международные узлы сети расположены на площадках сети NORDUnet (Стокгольм), институтов NIKHEF (Амстердам) и CSC (Хельсинки). RUNNet имеет два подключения к зарубежным NREN — в Хельсинки (к GÉANT) и в Стокгольме (к NORDUnet), с пропускной

способностью 10 Гбит/с каждое и с возможностью расширения.

RUNNet эксплуатирует и систематически развивает решения, базирующиеся на высокопроизводительных DWDM-сетях, участвует в обмене сетевым трафиком на шести IX-узлах (включая AMS-IX, MSK-IX, DATA-IX, SPB-IX; 41 Гбит/с в сумме), имеет более 25 прямых пиринговых соединений (175 Гбит/с в сумме).

Следует заметить, что после реорганизации в 2018 г. ФГАУ ГНИИ ИТТ "Информика", на протяжении многих лет являвшегося оператором RUNNet, и разделения Министерства образования и науки Российской Федерации на два федеральных органа исполнительной власти функции управления сетью были переданы ФГАОУ ДПО ЦРГОП и ИТ (подведомственному Минпросвещения России), для которого эта деятельность не является профильной.

Сегмент академической сети RASNet [11] в Московском регионе, оператором которого является МСЦ РАН, включает в себя опорную сеть в городе Москве, которая построена с использованием собственных волоконно-оптических линий связи (протяженностью более 100 км), с ядром на базе узлов в МСЦ РАН и на двух технологических площадках взаимодействия операторов связи, объединенных каналами с пропускной способностью 10 Гбит/с. Узлы используются для взаимодействия с другими научно-образовательными и отраслевыми сетями, научными центрами страны, а также для обмена сетевым трафиком и IP-транзита в сети Интернет.

Пользователями RASNet являются более 100 научных организаций, решающих с ее помощью задачи хранения, анализа и обмена научно-технической информацией, доступа к суперкомпьютерным центрам и развития научно-производственной кооперации. В состав RASNet входят сети трех региональных научных центров РАН (в Троицке, Пущино и Черноголовке).

Отдельные сегменты академической сети в Сибири и на Дальнем Востоке находятся под управлением региональных отделений РАН. Оператором сети Сибирского отделения РАН является Институт вычислительных технологий СО РАН [12], а сети Дальневосточного отделения РАН — Вычислительный центр ДВО РАН [13], который является ответственным арендатором канала передачи данных (200 Мбит/с), связывающего между собой отраслевые сети Московского, Сибирского и Дальневосточного регионов.

Можно констатировать таким образом, что на текущий момент единая национальная научно-образовательная сеть в стране отсутствует, вследствие чего имеют место весьма невы-

сокий уровень сетевого взаимодействия целевых пользователей, параллельное бюджетное финансирование идентичных инфраструктурных решений, а также в значительной мере ограничиваются потенциал международного научно-технологического сотрудничества и возможности полноправного участия российских организаций в международных проектах в области науки и образования (несмотря на многолетнее сотрудничество с отдельными зарубежными сетями и содержательный опыт участия в глобальных сетевых инициативах).

В последнем квартале 2018 г. ФГАОУ ДПО ЦРГОП и ИТ по заданию Минпросвещения России был сформирован комплекс организационно-технических решений, включающих накопленные в ходе управления сетью RUNNet результаты деятельности, в целях передачи управления проектом в профильное ведомство (Минобрнауки России) с сохранением целостности ключевых элементов сетевой инфраструктуры и сервисов и обеспечением их бесперебойного функционирования.

В соответствии с принятыми решениями Минобрнауки России в 2019 г. стартовали работы по созданию НИКС на основе интеграции сетей RUNNet и RASNet с передачей функций управления объединенной сетью, а также имущества, обязательств и персонала сети RUNNet в МСЦ РАН. После формирования ядра НИКС запланированы к реализации мероприятия по присоединению к ней региональных академических сетей СО РАН, ДВО РАН и других региональных и кампусных сетей с действующей в интересах научных и образовательных организаций локальной инфраструктурой доступа.

2. Цели и основные функции создаваемой НИКС

Ключевой целью создаваемой НИКС является предоставление научным и образовательным организациям страны возможностей для выполнения исследований и разработок по приоритетным направлениям НТР, участия в крупных российских и международных научных проектах, базирующихся на использовании устойчивой и отвечающей современным требованиям отраслевой информационно-телекоммуникационной сети, интегрированной в инфраструктуру мировых NREN.

НИКС будет ответственна за формирование и обеспечение функционирования единого научно-образовательного информационного пространства в целях эффективного использования создаваемой передовой инфраструктуры научных исследований и разработок, иннова-

ционной деятельности, в том числе для обеспечения глобальной доступности и возможностей распределенной работы с большими данными, генерируемыми научными установками класса "мегасайенс", другими крупными УНУ, а также на базе создаваемой сети НОЦ.

НИКС предоставит надежный высокоскоростной доступ к отраслевым и публичным сетям, расположенным в них ресурсам и сервисам, способствуя реализации информационного взаимодействия между организациями-участниками НОЦ при совместном выполнении ими научных и научно-производственных проектов для достижения целей НП, обеспечения решения задач СНТР и пространственного развития.

Предполагается, что НИКС будут делегированы отдельные функции разрабатываемой в рамках НП "Наука" ЕЦП и ее составных частей, в концепциях реализации которых явным образом зафиксирована необходимость создания и эксплуатации НИКС в качестве сетевой инфраструктуры российской науки и высшего образования.

На базе НИКС будет обеспечена отраслевая сетевая связность и предоставлен расширяемый набор специализированных сервисов, реализованы интерфейсы управления собственными ресурсами сети для возможности интеграции с цифровыми системами управления ЕЦП, в том числе с использованием развиваемых технологий федеративной аутентификации [14, 15].

Следует специально отметить, что НИКС лишена ограничений публичных сетей, эксплуатируемых коммерческими операторами связи, которые ориентированы на предоставление услуг на базе стандартных технологических решений с организацией доступа к научным ресурсам и проектам через глобальную сеть Интернет (и лишь при наличии принципиальной возможности). В сравнении с ними НИКС обеспечивает своим пользователям целый ряд существенных преимуществ, в том числе более дешевую стоимость услуг, лучшую производительность и характеристики качества, уникальную связность с международным научно-образовательным ИКТ-пространством, специализированные сервисы и адресную техническую поддержку.

3. Особенности архитектуры создаваемой НИКС: телекоммуникационная инфраструктура и сервисы

Функционирование НИКС основано на эксплуатации и поступательном совершенствовании глобальной гетерогенной отраслевой сети

передачи данных, обеспечивающей прямое сетевое взаимодействие целевых пользователей, сетевых субъектов и внешних генераторов больших научных данных. НИКС объединяет информационно-телекоммуникационные сети конечных пользователей и отдельные региональные сети в рамках общего адресного пространства с использованием современных сетевых технологий.

Архитектура НИКС проектируется, создается и модернизируется в интересах стратегических НП и программ сферы образования и науки с учетом ожидаемого в ближайшие годы многократного роста объемов генерируемых научных данных, которые потребуют передачи по сетям и распределенной обработки участниками устойчивых исследовательских коллабораций.

В архитектурном и функциональном плане НИКС подразделяется на *инфраструктурную* и *сервисную* составляющие. *Телекоммуникационная инфраструктура* НИКС, включающая в себя магистральную (опорную) инфраструктуру и инфраструктуру доступа, на начальном этапе построения будет использовать объединенные и оптимизированные инфраструктурные компоненты сетей RUNNet и RASNet и других интегрируемых в нее сетей. Транспортная основа магистральной инфраструктуры — это опорная сеть, обеспечивающая магистральную связность между всеми федеральными округами, а также с зарубежными NREN и сетевыми консорциумами. Работа магистральной инфраструктуры организуется на базе различающихся предназначением и составом оборудования опорных узлов сети, создаваемых в крупных городах страны и за рубежом, а также каналов связи между ними.

Опорные узлы федерального уровня являются ключевыми элементами сети, создаются и функционируют в целях обеспечения эффективного межсетевого обмена данными между целевыми пользователями, отказоустойчивости и оптимизации использования магистральной инфраструктуры с учетом поставленных задач СНТР, НП "Наука" и пространственного развития.

Федеральные опорные узлы связываются высокоскоростными магистральными каналами друг с другом с помощью физически независимых магистралей с полным резервированием (в соответствии с топологией и уровнем утилизации отдельных сегментов), с региональными опорными узлами сети, узлами доступа, научными и отраслевыми сетями иных ведомств, сетями других телекоммуникационных операторов и с глобальной сетью Интернет.

Пропускная способность каналов связи в составе магистральной инфраструктуры может

варьироваться от 10 до 100 Гбит/с в зависимости от потребностей пользователей (в первую очередь, базовых организаций объектов передовой инфраструктуры научных исследований и разработок, имеющих повышенные требования к характеристикам сетей передачи данных).

Региональные опорные узлы создаются в городах, являющихся значимыми научными и образовательными центрами, представляют собой оборудованные телекоммуникационные узлы, которые задействуются для подключения к магистральной инфраструктуре расположенных в субъекте целевых пользователей, в том числе участников НОЦ, НЦМУ, ЦК НТИ и др., а также объектов научной инфраструктуры коллективного пользования (ЦКП, УНУ).

Инфраструктура доступа представляет собой совокупность каналов связи и размещенного на узлах доступа телекоммуникационного оборудования, которые обеспечивают постоянный высоконадежный доступ пользователей к магистральной инфраструктуре НИКС. Узлы доступа могут располагаться в серверных помещениях организаций-пользователей, на технологических площадках взаимодействия операторов, телекоммуникационных узлах операторов связи. Скорость подключения пользователей к магистральной инфраструктуре может существенно различаться (в зависимости от потребностей, технических возможностей и экономических аспектов).

Управление интегрируемыми в структуру НИКС региональными и отраслевыми сетями, кампусными сетями пользователей осуществляется их операторами и телекоммуникационными службами самостоятельно во взаимодействии со специалистами НИКС.

Связность с зарубежными NREN и публичными сетями (включая глобальную сеть Интернет) организуется через федеральные и международные узлы НИКС посредством резервируемых высокоскоростных магистральных каналов связи с использованием ресурсных возможностей зарубежных NREN и сетевых консорциумов, верхнеуровневых сервис-провайдеров и международных технологических площадок взаимодействия операторов связи.

В целях повышения уровня сетевой связности НИКС может подключаться к расположенным внутри страны и за рубежом точкам обмена трафиком (IX), участвовать в прямом междосетевом обмене с сетями крупных операторов связи.

НИКС предоставляет пользователям систематически пополняемый набор специализированных сервисов, принимает участие в их развитии и совершенствовании, ориентируясь на достижения мировых NREN. Сервисы можно подразделить на четыре группы: инфраструк-

турные сервисы, сервисы телематических служб, прикладные и профессиональные сервисы.

К базовым *инфраструктурным сервисам* НИКС относятся:

- обеспечение функционирования отраслевой сети передачи данных, реализация сетевой связности в интересах сферы науки и образования страны, включая цели и задачи национальных проектов и программ;
- предоставление целевым пользователям возможностей удаленного контролируемого доступа к объектам создаваемой и совершенствуемой научной инфраструктуры коллективного пользования (установки класса "мегасайенс", ЦКП, УНУ, СКЦ);
- объединение на основе инфраструктурно-сервисной ЦП сети ресурсов организаций в целях развития научной и научно-производственной кооперации (НОЦ, НЦМУ, ЦК НТИ);
- обеспечение высокопроизводительного междоузлового взаимодействия с зарубежными NREN, доступа пользователей к функционирующим в них научным и образовательным ресурсам и сервисам;
- предоставление возможностей удаленного доступа к международным исследовательским проектам и массивам научных данных, в том числе класса "большие данные" (LHC, XFEL, ESRF, FAIR, ITER, SuperKEKB, NOvA, XENON, LIGO, PAO, SKA и др.);
- предоставление доступа к публичным сетям, включая сеть Интернет;
- осуществление IP-транзита трафика для региональных научно-образовательных сетей и операторов связи, аффилированных с научными и образовательными организациями;
- создание отдельных виртуальных частных сетей (VPN), функционирующих изолированно от публичной сети Интернет;
- организация каналов передачи данных по запросу с заданными параметрами качества SLA для совместного выполнения научных проектов;
- создание наложенных проблемно-ориентированных сетей, обеспечивающих сетевую связность организаций, вовлеченных в исследования по общим научным тематикам, в том числе сетей для распределенной обработки научных данных;
- сервисы сетевой и информационной безопасности (защита от DDoS-атак, фильтрация сетевого трафика, обмен данными в защищенном контуре и др.);
- сервисные решения на базе технологий программно-конфигурируемых сетей (SDN), виртуализация физических сетевых ресурсов;

- создание тестовых ИКТ-площадок, лабораторных стендов по запросу (апробация, тестирование и запуск в рабочую эксплуатацию перспективных IT-технологий и прикладных сервисов);
- облачные сервисы модели IaaS (управление средой облачных вычислений, предоставление виртуальных машин по запросу, индивидуальная конфигурация вычислительных мощностей, сервисы виртуальных блочных устройств, эластичного хранения объектных данных, виртуальных сетей).

К специализированным *сервисам телематических служб* относятся:

- сайт центра управления сетью (NOC) с личными кабинетами пользователей, развитым информационным и вспомогательным функционалом, включая обработку заявок по сетевым инцидентам, предоставление сведений об используемых ресурсах, загрузке каналов и др.;
- Local Internet Registry, LIR (управление AS, сетями IPv4/IPv6 адресов);
- сервисы тестирования характеристик сетевой инфраструктуры, измерения пропускной способности каналов передачи данных;
- служебные сервисы мониторинга и контроля работоспособности магистральной сетевой инфраструктуры и отдельных сервисов;
- сервисы блокировки доступа к запрещенным законодательством ресурсам;
- традиционные сетевые сервисы (DNS, NTP, электронная почта, обмен сообщениями, цифровые сертификаты и др.).

К *прикладным сервисам*, в развитии и эксплуатации которых могут задействоваться ресурсы и компетенции НИКС, относятся:

- сервисы системы управления научной инфраструктурой коллективного пользования (списки оборудования ЦКП и УНУ, мониторинг доступности и загрузки, средства предоставления оперативной аналитики и отчетности);
- облачные сервисы модели PaaS (вычислительные кластеры и разделяемые хранилища данных по запросу, виртуальные рабочие столы, совместная разработка и поддержка IT-проектов, разработка документов);
- облачные сервисы для проведения совместных исследований в удаленном доступе модели SaaS (предоставление доступа к прикладному программному обеспечению по запросу, запуск проблемно-ориентированных пакетов на вычислительных кластерах с контролем доступа и использования ресурсов);

- облачные сервисы работы с научными данными (архивное и временное хранение, виртуализация, обмен данными, совместная работа, цифровые права и управление ими);
- сервисы визуализации и проведения численных расчетов, "научные блокноты" (интерактивные среды программирования, создания и обмена документами с развитыми возможностями коллективной работы);
- сервисы идентификации участников научно-технического взаимодействия и повсеместного безбарьерного доступа к научным ресурсам на базе технологий федеративной аутентификации (обеспечение мобильности, повсеместного доступа к ресурсам, интеграция с зарубежными системами идентификации, участие в международных проектах eduroam, eduGAIN);
- сервисы научных коммуникаций в режиме реального времени для проведения конференций, семинаров, совещаний, персонального профессионального общения, научные социальные сети (вебинары, видеоконференцсвязь, стриминговые трансляции, VoIP и др.);
- сервисы мониторинга, анализа и визуализации статистики по степени использования телекоммуникационной инфраструктуры сети пользователями для целей обмена научными данными.

Профессиональные сервисы предполагают оказание консультационных услуг и технической поддержки пользователей, проектное управление при решении задач развития сети, апробации и внедрения новых технологий и сервисов.

Заключение

Создание НИКС с нормативным закреплением за ней статуса национальной научно-образовательной сети Российской Федерации и успешная реализация мероприятий по ее дальнейшему развитию будут способствовать укреплению позиций российской науки в приоритетных научных областях при реализации совместных проектов в рамках внутрироссийского и международного сотрудничества научных и научно-технологических коллабораций, предоставит возможности для прагматически обоснованной полномасштабной интеграции участников научно-технического процесса в международное научное ИКТ-пространство, глобальную инфраструктуру мировых NREN. Это обеспечит доступ к новым компетенциям, идеям и передовым технологиям, внесет свой вклад в формирование

интернациональной научной среды, устойчивой кооперации с мировым сообществом с учетом национальных интересов.

На базе НИКС будет реализована надежная высокоскоростная научная связность ведущих исследовательских центров, научных и образовательных организаций для эффективного обмена данными с контролем использования и оптимизацией затрат на сеть передачи данных, эксплуатируемую в интересах создаваемой географически распределенной передовой инфраструктуры исследований и разработок и способствующую достижению целей и ключевых показателей НП "Наука".

Список литературы

1. **Абрамов А. Г., Евсеев А. В.** Сеть RUNNet: навстречу современным вызовам сферы телекоммуникаций в науке и образовании // Информатизация образования и науки. 2017. № 1. С. 100—115.
2. **Ижванов Ю. Л.** Научно-образовательные компьютерные сети. Прошлое, настоящее и тенденции развития // Образовательные ресурсы и технологии. 2017. № 2. С. 17—25.
3. **Абрамов А. Г., Евсеев А. В.** RUNNet как национальная научно-образовательная сеть России: цели, основные задачи, телекоммуникационная инфраструктура и сервисы // Информатизация образования и науки. 2018. № 4. С. 3—15.
4. **Abramov A. G., Evseev A. V.** RUNNet: infrastructural and service basis of the national research and education network of the Russian Federation / Proc. VIII Int. Conf. "Distributed Computing and Grid-technologies in Science and Education" (GRID 2018), Dubna, September 10—14, 2018. P. 52—57. [Urn:nbn:de:0074-2267-5](https://arxiv.org/abs/1809.00074).
5. **Allocchio C., Balint L., Berkhout V., Bersee J., Izhvanov Y. et al.** A History of international research networking: the people who made it happen. N. Y.: Wiley-VCH, 2010. 317 p.
6. **Lehtisalo K.** The History of NORDUnet: Twenty-five years of networking cooperation in the Nordic countries (2005). URL: <http://www.nordu.net/history/book.html>.
7. **GEANT:** The Case for NRENs. A Repository of Resources to Support Funding, Advocacy and the Advancement of National and Regional R&E Networks. URL: <https://www.caseforrens.org>.
8. **Абрамов А. Г.** Панъевропейский научно-образовательный сетевой консорциум GEANT: особенности инфраструктуры, ключевые проекты и сервисы // Информационные технологии. 2018. № 8. С. 546—553.
9. **Васенин В. А.** Российские академические сети и Internet (Состояние, проблемы, решения). М.: РЭФИА, 1997. 173 с.
10. **Гугель Ю. В., Ижванов Ю. Л., Абрамов А. Г.** Федеральная университетская сеть RUNNet: прорыв в третье десятилетие // Труды XVIII объединенной конференции "Интернет и современное общество" (IMS-2015), Санкт-Петербург, 23-25 июня 2015 г. СПб.: Университет ИТМО, 2015. С. 249—259.
11. **Кулагин М. В., Серебряков В. А.** Информационное пространство РАН (Проекты и реализация, 1998—2013) // Научный сервис в сети Интернет: труды XVIII Всероссийской научной конференции (19—24 сентября 2016 г., г. Новороссийск). М.: ИПМ им. М. В. Келдыша, 2016. С. 194—222.
12. **Шокин Ю. И., Федотов А. М., Белов С. Д., Жижимов О. Л., Никульцев В. С., Чубаров Л. Б.** Мультисервисная корпоративная сеть передачи данных Сибирского отделения РАН (МКС ПД СО РАН). Сеть Интернет Сибирского отделения РАН СО РАН. Новосибирск: Издательский центр ИВТ СО РАН, 2005. 39 с.
13. **Ханчук А. И., Наумова В. В., Сорокин А. А.** Корпоративная сеть ДВО РАН: высокотехнологичная интеграция научных подразделений // Вестник РАН. 2008. № 4. С. 298—303.
14. **Овсянников А. П., Савин Г. И., Шабанов Б. М.** Удостоверяющие федерации научно-образовательных сетей // Программные продукты и системы. 2012. № 4. С. 3—7.
15. **Абрамов А. Г., Васильев И. В., Порхачев В. А.** Развитие инфраструктуры аутентификации и авторизации для удостоверяющей федерации в рамках проектов eduGAIN и eduoam на базе сети RUNNet // ИТНОУ: Информационные технологии в науке, образовании и управлении. 2017. № 4. С. 56—64.

A. G. Abramov^{1,2}, Ph.D., Assistant Professor, Deputy Director, e-mail: abramov@runnet.ru,

A. V. Evseev², Director, e-mail: evseev@runnet.ru,

¹ Peter the Great St.Petersburg Polytechnic University,

² St. Petersburg branch of FPAEI CVE CRSEPIIT, St. Petersburg, 199034, Russian Federation

Conceptual Aspects of Creating the New Generation National Research Computer Network in the Russian Federation

The current status of the world's research and education networks is discussed, their key characteristics, functions and tasks are presented, a brief historical background and general information about the current level of development of research and education networks in Russia are given, the necessity of creating a unified research and education network in the country is substantiated with a conceptual presentation of its objectives and core functions taking into account national priorities in science and education, the consideration of the architectural components of the network and some of the expected results.

Keywords: National Research and Education Network, NREN, RUNNet, telecommunication infrastructure, research network connectivity, network services, digital platform

DOI: 10.17587/it.25.724-733

References

1. Abramov A. G., Evseev A. V. Network RUNNet: towards the state-of-the-art challenges in the field of telecommunications in science and education, *Informatizatsiya Obrazovaniya i Nauki*, 2017, vol. 1, pp. 100–115 (in Russian).
2. Izhvanov Yu. L. Research and education computer networks. Past, present and development trends, *Obrazovatel'nye Resursy i Tekhnologii*, 2017, vol. 2, pp. 17–25 (in Russian).
3. Abramov A. G., Evseev A. V. RUNNet as a national research and education network of Russia: goals, main tasks, telecommunication infrastructure and services, *Informatizatsiya Obrazovaniya i Nauki*, 2018, vol. 4, pp. 3–15 (in Russian).
4. Abramov A. G., Evseev A. V. RUNNet: infrastructural and service basis of the national research and education network of the Russian Federation. Proc. VIII Int. Conf. "Distributed Computing and Grid-technologies in Science and Education" (GRID 2018), Dubna, September 10–14, 2018. pp. 52–57. Urn:nbn:de:0074-2267-5.
5. Allocchio C., Balint L., Berkhout V., Bersee J., Izhvanov Y. et al. *A History of international research networking: the people who made it happen*, N. Y., Wiley-VCH, 2010. 317 p.
6. Lehtisalo K. The History of NORDUnet: Twenty-five years of networking cooperation in the Nordic countries (2005), available at: <http://www.nordu.net/history/book.html>.
7. GEANT: The Case for NRENs. A Repository of Resources to Support Funding, Advocacy and the Advancement of National and Regional R&E Networks, available at: <https://www.casefornrens.org>.
8. Abramov A. G. Pan-European research and education network consortium GEANT: infrastructure features, key projects and services, *Informatsionnyye Tehnologii*, 2018, vol. 8, pp. 546–553 (in Russian).
9. Vasenin V. A. *Russian academic networks and Internet (Status, problems, solutions)*, Moskva, REFIA, 1997. 173 p. (in Russian).
10. Gugel Yu. V., Izhvanov Yu. L., Abramov A. G. Federal University Network RUNNet: breakthrough in the third decade. Proceedings of the XVIII Joint Conference "The Internet and Modern Society" (IMS-2015), St. Petersburg, June 23–25, 2015, ITMO University, pp. 249–259 (in Russian).
11. Kulagin M. V., Serebryakov V. A. Information space of RAS (Projects and implementation, 1998–2013). Scientific service on the Internet: works of the XVIII All-Russian Scientific Conference (September 19–24, 2016, Novorossiysk), Moscow, Keldysh IAM, 2016, pp. 194–222 (in Russian).
12. Shokin Yu. I., Fedotov A. M., Belov S. D., Zhizhimov O. L., Nikultsev V. S., Chubarov L. B. *Multiservice corporate data network of the Siberian Branch of the Russian Academy of Sciences (MCS DN SB RAS). The Internet of the Siberian Branch of the RAS SB RAS*. Novosibirsk, Publishing Center of ICT SB RAS, 2005. 39 p. (in Russian).
13. Khanchuk A. I., Naumova V. V., Sorokin A. A. Corporate network of the Far Eastern Branch of the Russian Academy of Sciences: high-tech integration of scientific departments, *Vestnik RAN*, 2008, vol. 4, pp. 298–303 (in Russian).
14. Ovsyannikov A. P., Savin G. I., Shabanov B. M. Identity federations of research and education networks, *Programmye Produkty i Sistemy*, 2012, vol. 4, pp. 3–7 (in Russian).
15. Abramov A. G., Vasilyev I. V., Porhachev V. A. Development of the authentication and authorization infrastructure for the identity federation within the eduGAIN and eduroam projects based on the RUNNet network, *ITNOU: Informatsionnyye Tekhnologii v Nauke, Obrazovanii i Upravlenii*, 2017, vol. 4, pp. 56–64 (in Russian).

УДК 004.031.2

DOI: 10.17587/it.25.733-737

Ю. А. Холопов, вед. инженер,

Институт точной механики и вычислительной техники им. С. А. Лебедева

Российской академии наук,

Нгуен Ван Хиеу, магистрант, e-mail: vanhieu942010@gmail.com

Московский физико-технический институт (национальный исследовательский университет)

Реализация контроллера активной синхронной измерительной подсистемы цифровой системы управления

Вынесение из центрального процессора функций управления вводом параметров упрощает разработку и отладку цифровых систем управления. В статье описан вариант построения контроллера периферии с общим для всех интерфейсов управления датчиками генератора временных меток. Показаны преимущества предложенной реализации.

Ключевые слова: цифровая система управления, активная синхронная измерительная подсистема, исполнительная подсистема

Введение

В настоящее время большинство цифровых систем управления (ЦСУ) построено на основе вычислителей с архитектурой универсальной ЭВМ. В таких системах управлением периферии

ей — датчиками и исполнительными устройствами — занимается центральный процессор (ЦП).

Наиболее точная картина состояния объекта управления (ОУ) может быть получена при фиксации всех его параметров одновременно. Программное управление датчиками проводит-

ся последовательно, что не позволяет фиксировать состояние ОУ по единому событию. Решение проблемы предлагается в данной статье.

1. Анализ задачи

В любой ЦСУ набор датчиков всегда статичен или заранее определен, следовательно, гибкость управления измерениями не требуется, и есть возможность вынести функцию управления датчиками из ЦП в некоторый автомат, который будет читать значения с выходов датчиков и пересылать их в память ЦП, выполняя функции активной измерительной подсистемы (ИП) ЦСУ.

Существуют два способа организации работы активной ИП: асинхронный и синхронный.

Асинхронная ИП с максимальной скоростью опрашивает датчики и переносит параметры состояния объекта от каждого из них в общую с ЦП память. В памяти всегда содержатся обновленные данные, которые опрашиваются ЦП асинхронно: и относительно момента их обновления, и относительно фаз обновления в памяти параметров из соседних измерительных каналов.

В синхронной ИП параметры датчиков фиксируются в один и тот же момент времени один раз в начале каждого цикла регулирования. В памяти ЦП параметры обновляются также одновременно по всем каналам измерения. В предлагаемой статье рассматривается реализация контроллера ИП ЦСУ, работающего в активном синхронном режиме.

2. Общая структура цифровой системы управления

Основными подсистемами цифровых систем управления (ЦСУ) являются: расчетная (ЦП), измерительная и исполнительная (рис. 1).

Цифровые системы управления работают циклически, в дискретно-непрерывном режиме. В каждом цикле управления параметры состояния объекта управления фиксируются датчиками (Д) и пересылаются измерительной подсистемой в память расчетной подсистемы, которая доступна и центральному процессору и ИП. На основе информации о состоянии объекта процессор рассчитывает значения параметров управления, необходимые для выполнения поставленной системой управления верхнего уровня или директивной подсистемой задачи. Затем исполнительная подсистема пересылает



Рис. 1. Структурная схема цифровой системы управления

параметры управления из памяти ЦП исполнительным устройствам (ИУ), которые преобразуют параметры управления в непосредственные воздействия на объект управления [1].

3. Интерфейсы активной синхронной измерительной подсистемы

Как следует из структуры ЦСУ, в модуле измерительной подсистемы должны присутствовать два типа интерфейсов: интерфейс с датчиками и интерфейс с памятью ЦП.

3.1. Интерфейс с датчиками

В ЦСУ типажность первичных интерфейсов для подключения датчиков невелика. Как правило, используются интерфейсы: I²C, UART, SPI, CAN и т. п. Каждый из первичных интерфейсов может использоваться в ЦСУ несколько раз.

Исходными данными для проектирования контроллера интерфейсов с датчиками (далее — модуль внешних интерфейсов, или МВИ) являются их спецификации. Основная задача МВИ: управлять этими интерфейсами в стиле активной синхронной периферийной подсистемы.

Покажем, что структура ИП может быть оптимизирована за счет разделенной реализации функций: логической и временной последовательностей событий управления интерфейсами.

Спецификация любого интерфейса задает последовательность событий управления присоединенным устройством. Переключения машины состояний интерфейса происходят или по временной метке, или по сочетанию внешнего логического события и временной метки.

По опыту авторов, до 60 % оборудования, предназначенного для сопряжения с первичными интерфейсами, приходится на схемы анализа и синтеза временных интервалов. Поэтому возникает желание перенести часть такого оборудования в общий генератор временных меток. Временные события в такой системе обрабатываются не с точностью до одного периода высокочастотного системного синхросигнала, а с необходимой практической точностью, обычно это 1/16 или 1/32 часть обрабатываемого временного интервала.

Основная идея оптимизации структуры МВИ состоит в том, что для генерации временных меток для разных интерфейсов будет использован единый генератор — многоканальный таймер, а логические события формируются индивидуально в соответствии со спецификацией конкретного интерфейса.

Поясним преимущества предлагаемого способа построения модуля внешних интерфейсов, для чего рассмотрим два варианта его реализации.

В первом из них для каждого интерфейса используется отдельный автономный контроллер, который запускается по временной метке начала цикла регулирования и автономно реализует обе последовательности интерфейсных событий.

Структурная схема такого варианта модуля показана на рис. 2.

Модуль внешних интерфейсов состоит из индивидуальных блоков B_i управления первичным интерфейсом каждого типа датчика ST_i . Каждый из этих блоков формирует на выходе набор измеренных параметров состояния $P[i]$ и сигнал их готовности $Ready[i]$, управляющий записью параметров в память контроллера. Все модули используют общий сигнал синхронизации clk , формируемый генератором G . Сетку временных событий интерфейса формирует локальный таймер TM_i . Запуск измерения во всех модулях происходит по единому сигналу $START$, который формируется от блока деления частоты. В современных ЦСУ типовой диапазон частот сигналов: clk — 1...200 МГц, $START$ — 50...1000 Гц.

Для канала записи параметров в память активность индивидуальных блоков управления носит случайный характер, и существует проблема конфликта сигналов готовности $Ready[i]$ от разных модулей: они могут формироваться одновременно. Для разрешения конфликта необходимо использовать дополнительную логику в интерфейсе с буферной памятью контроллера.

Рассмотрим второй вариант реализации группового контроллера, структурная схема которого показана на рис. 3.

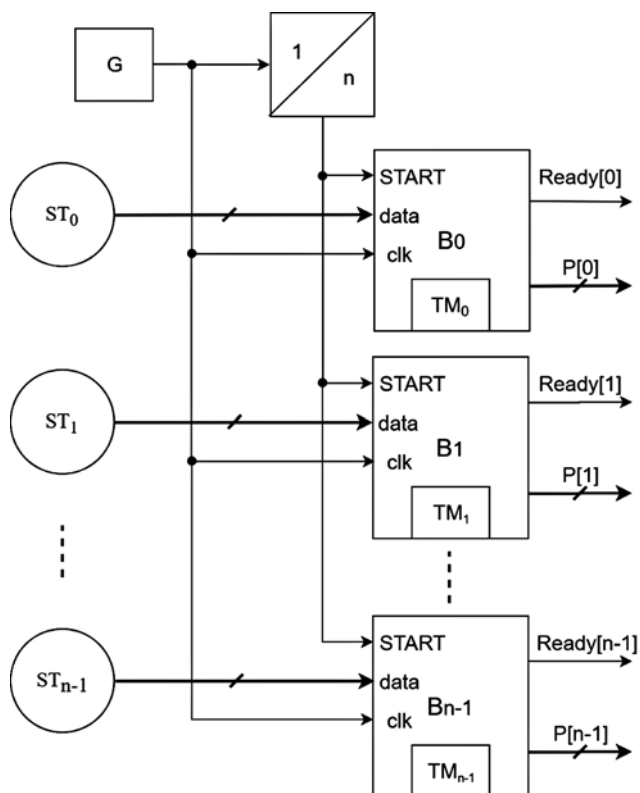


Рис. 2. Модуль внешних интерфейсов с локальными таймерами TM_i

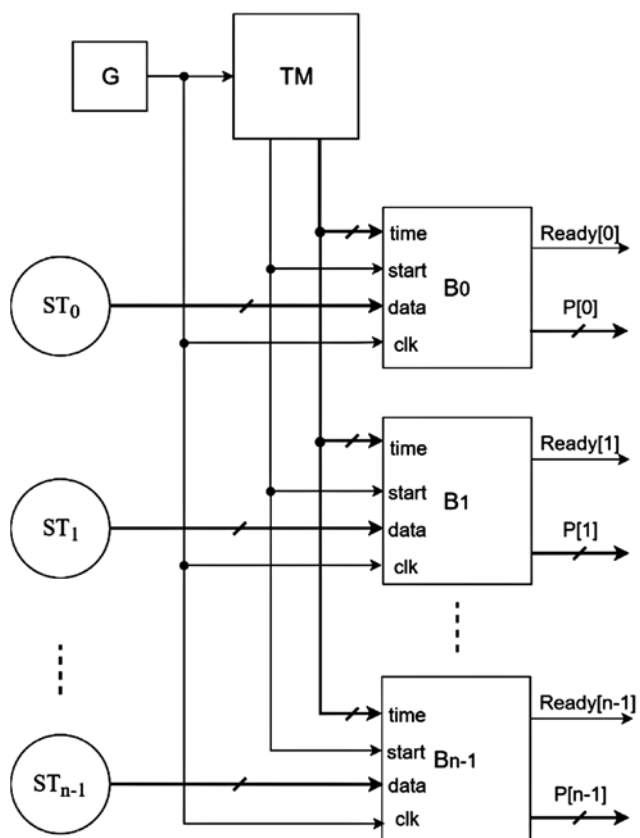


Рис. 3. Модуль внешних интерфейсов с общим таймером TM

Модуль внешних интерфейсов, как и прежде, строит из индивидуальных блоков B_i , которые управляют однотипными интерфейсами и синхронизируются общим сигналом синхронизации от генератора G . Но теперь блок деления частоты генерирует не только метку **START** — начало измерения, но и набор промежуточных временных меток, которые используются для управления внешними интерфейсами. Наличие в контроллере единой сетки временных меток позволяет управлять в разных каналах измерения фазой формирования сигналов готовности данных. Такая реализация позволяет не только сэкономить до 60 % оборудования на временных операциях, но и дополнительно исключить логику разрешения конфликтов доступа к буферной памяти контроллера. Теперь возникновением готовности в каждом измерительном канале можно управлять так, чтобы разнести их активность во времени. Режим работы синхронной ИП с общей памятью ЦП и измерительной подсистемы изменяется. Обслуживание случайных, конфликтующих запросов от измерительных каналов в асинхронной ИП трансформируется в синхронной ИП в режим обращения по расписанию.

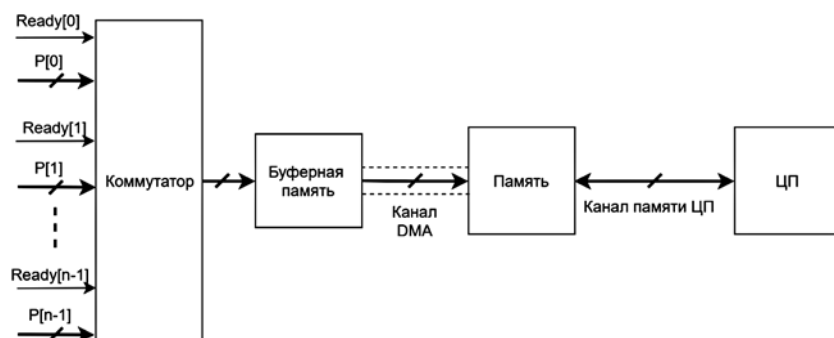


Рис. 4. Схема режима прямого доступа к памяти DMA

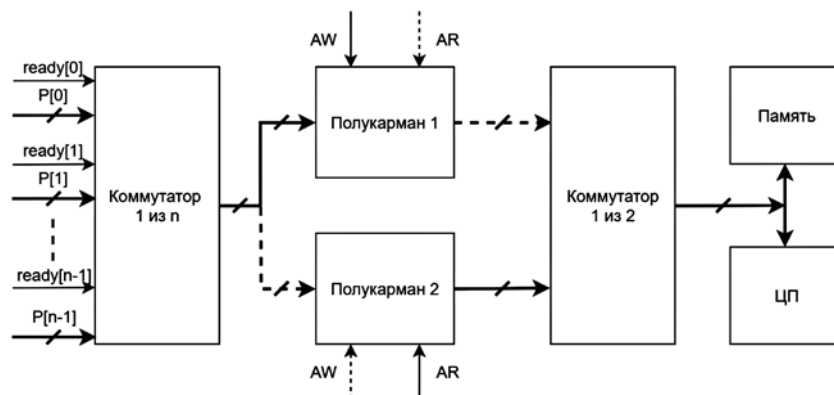


Рис. 5. Механизм "полукарманов"

Может создаться впечатление, что предлагаемая технология касается только подключения тривиальных измерительных датчиков. Однако опыт показывает применимость данного подхода, например, для синхронизации каналов систем технического зрения в целях последующей согласованной обработки кадров из разных каналов [2].

3.2. Интерфейс с памятью ЦП

В каждом цикле работы ЦСУ после выполнения фазы измерения все параметры состояния ОУ должны быть переданы в память ЦП. В контроллере активной синхронной ИП эту функцию выполняет интерфейс с памятью ЦП.

Рассмотрим два варианта реализации этого интерфейса. Первый из них — это традиционный режим "прямого доступа к памяти" (Direct memory access — DMA). Схема такого варианта показана на рис. 4.

DMA — режим обмена данными между устройствами или же между устройством и основной памятью, в котором ЦП не участвует. После окончания измерительной фазы цикла управления все параметры состояния зафиксированы в регистрах МВИ, и есть возможность пересылать их в память ЦП одной групповой операцией, через канал DMA.

Режим DMA рационально использовать для подключения подсистемы ввода к высокопроизводительным микропроцессорам, так как в их окружении часто присутствуют скоростные интерфейсы, реализованные на базе DMA-обменов, например каналы PCI-E. В простых однокристальных вычислителях взаимодействие ИП с ЦП легче реализовать через механизм "полукарманов" (рис. 5).

Принцип работы механизма "полукарманов" состоит в следующем. В каждом цикле регулирования одновременно выполняются два типа операций обмена между МВИ, ЦП и памятью параметров: запись в память вновь измеренных параметров, которые будут использоваться в следующем цикле, и чтение в ЦП параметров состояния, сформированных для текущего цикла. В буфере используются два банка однопортовой памяти. В ци-

кле регулирования из одного банка происходит (только) чтение параметров, в то время как в другой банк осуществляется запись вновь измеренных параметров. При записи в банк адреса формирует контроллер ИП, в режиме чтения банк подключается в отрезок адресного пространства ЦП. В следующем цикле регулирования функции банков меняются [3].

Такая реализация позволяет исключить конфликт между процессом записи новых параметров в память ЦП и процессом чтения предыдущих параметров из этой памяти. ЦП в любой момент времени может опрашивать подключенный в адресное пространство буфер в монопольном режиме.

Выводы

Предлагаемый синхронный способ взаимодействия ЦП с датчиками в ЦСУ реализует упорядоченный стиль управления внешними интерфейсами и имеет следующие преимущества в сравнении с традиционным — асинхронным:

- сокращает затраты на контроллеры индивидуальных первичных интерфейсов за счет использования единого генератора временных меток;
- отсутствуют конфликты обращения модулей первичных интерфейсов к буферной памяти, что сокращает аппаратные затраты и время формирования пакета состояния;
- ускоряется работа измерительной подсистемы: для пересылок параметров состояния в память ЦП используются только групповые обмены.

Список литературы

1. Поляков К. Ю. Основы теории цифровых систем управления: учеб. пособ. СПб.: Изд. СПбГМТУ, 2006. С. 10—12.
2. Пунь А. Б., Гаврилов Д. А., Щелкунов Н. Н., Fortunatov A. A. Алгоритм адаптивной бинаризации объектов в видеопоследовательности в режиме реального времени // Успехи современной радиоэлектроники. 2018. № 8. С. 40—48.
3. Ле Ба Чунг, Нгуен Тхань Чунг. Особенности подключения периферийных устройств к микропроцессору в высокودинамичной системе управления // Сб. XXI Междунар. науч.-практ. конф. "Техника и технология: новые перспективы развития. Москва, 20 мая 2016 г.". М.: Спутник+, 2016. С. 15—18.

Yu. A. Holopov, Leader Engineer,

Lebedev Institute of Precision Mechanics and Computer Engineering,

Nguyen Van Hieu, Postgraduate, e-mail: vanhieu942010@gmail.com,

Moscow Institute of Physics and Technology (National Research University)

Implementation of the Active Synchronous Measuring Subsystem Controller in Digital Control System

The removal of parameter input control functions from the CPU simplifies the development and debugging of digital control systems. The article describes a variant of designing a peripheral controller, for all sensor control interfaces, a timestamp generator. The advantages of the proposed implementation are shown.

Keywords: digital control system, active synchronous measuring subsystem, executive subsystem

DOI: 10.17587/it.25.733-737

References

1. Poljakov K. Ju. Fundamentals of the theory of digital control systems, SPb., Publishing house of SPbGMTU, 2006, pp. 10—12 (in Russian).
2. Pun' A. B., Gavrilov D. A., Shhelkunov N. N., Fortunatov A. A. *Uspehi sovremennoj radioelektroniki*, 2018, no. 8, pp. 40—48 (in Russian).

3. Le Ba Chung, Nguyen Thanh Trung. Features of connection of peripheral devices to the microprocessor in a highly dynamic control system, *Sbornik XXI Mezhdunarodnoj nauchno-prakticheskoy konferencii "Tehnika i tehnologija: novye perspektivy razvitiya. Moskva, 20 maja 2016 g."*, Sputnik+, 2016, pp. 15—18 (in Russian).

В. В. Грибова, д-р техн. наук, ст. науч. сотр,
зам. директора по научной работе, e-mail: gribova@iacp.dvo.ru,
Е. А. Шалфеева, канд. техн. наук, доц., ст. науч. сотр, e-mail: shalf@dvo.ru,
Институт автоматизации и процессов управления ДВО РАН, Владивосток

Обеспечение жизнеспособности систем, основанных на знаниях¹

Рассматривается модель жизнеспособности систем, основанных на знаниях, конструируемых с помощью онтолого-ориентированной среды развития, где каждый компонент формируется в терминах явно представленной онтологии предметной области с использованием адаптируемой инструментальной поддержки.

Ключевые слова: интеллектуальные системы, база знаний, основанные на знаниях системы, онтолого-ориентированный алгоритм, жизнеспособность, декларативный инструментарий, среда развития

Введение

Перед разработчиком любой программной системы (ПС) стоит задача не только создать программный продукт, отвечающий всем текущим требованиям к нему, но и заложить в него механизмы, обеспечивающие его неизбежное последующее сопровождение, вызываемое изменениями требований пользователей к интерфейсу и функциональности, знаний предметной области, условий эксплуатации.

На процесс сопровождения приходится значительная доля трудозатрат в процессе жизненного цикла программной системы, по оценкам специалистов — от 50 до 80 %. Поэтому необходимо проектировать ПС, закладывая механизмы для упрощения этого процесса.

Известными механизмами, используемыми в программной инженерии для упрощения сопровождаемости ПС, являются [1–4]:

- архитектурные решения (разделение на слабо сцепленные компоненты с логически понятными функциями);
- декларативное представление компонентов ПС (например, с помощью структурных и графических редакторов) и использование "методов 4GL", смещающих усилия "от программирования к конструированию";

- средства автоматизации построения проектных моделей и автоматической генерации по ним фрагментов кода (генерации схем СУБД, генерации программного кода);
- средства связывания компонентов друг с другом (например, классов, использующих друг друга, или хранимых таблиц БД с визуализируемыми результатами запросов);
- использование типовых архитектурных решений, "навязываемых" так называемыми оболочками (от СУБД типа MSAccess до платформы типа IC);
- введение в процесс жизненного цикла вида деятельности "управление требованиями" (с инструментальной поддержкой, как в Rational Rose для RUP);
- разделение компетенций (между дизайнером интерфейса, специалистом по сетевому взаимодействию, проектировщиком базы данных, наполнителями баз, специалистами по "большим данным", программистами и другими типами разработчиков).

Поскольку системы, основанные на знаниях (СОЗ), — особый класс систем, дополнительным архитектурным компонентом которых является база знаний, и команда разработчиков пополняется экспертами предметной области и когнитологами, проблема сопровождаемости СОЗ стоит особо остро [5–7]. Указанная специфика накладывает необходимость использования дополнительных специализированных

¹Работа выполнена при частичной финансовой поддержке РФФИ (проекты № 18-07-01079 и 19-07-00244).

механизмов обеспечения их сопровождаемости и жизнеспособности.

Целью работы является представление обобщенной архитектурной модели сопровождаемой СОЗ и совокупности средств обеспечения ее жизнеспособности.

1. Понятия сопровождаемости и жизнеспособности программных систем

С появлением сложных программных систем, на разработку которых стали требоваться усилия, измеряемые человеко-годами, стала актуальна проблема их сопровождаемости.

Традиционно *сопровождаемость* ПС означает возможность ее исправления, переструктурирования (рефакторинга), адаптации (к аппаратуре и системному программному обеспечению, к новым видам человеко-машинных интерфейсов, к новому типу пользователей) и расширения (новыми функциями по желанию пользователей). Но следующие десятилетия обозначили новое полезное свойство систем — жизнеспособность как устойчивость (сохранение работоспособности и полезности) ПС к изменениям окружающей среды [8] и эволюционируемость как способность приспосабливаться к изменчивости требований в течение "жизни" с наименее возможной стоимостью, поддерживая архитектурную целостность [9].

Будем называть *жизнеспособностью* устойчивость (сохранение работоспособности) ПС к некоторым изменениям среды функционирования и способность к развитию (эволюционируемость) в течение жизни, которые обеспечиваются через адаптивность (автоматическое изменение) к изменениям в среде и адаптируемость (интерактивное изменение) ПС в ответ на новые требования.

Адаптируемость реализуется через:

- декларативное представление компонентов, обеспечивающее, с одной стороны, программно-обрабатываемое их представление, с другой, наглядное и понятное, соответствующее системе терминов (системе понятий, языку общения) разработчика;
- наличие структурных и графических редакторов для поддержки разработки всех компонентов ПС;
- автоматизацию разработки компонентов (генерацию схем СУБД, генерацию про-

граммного кода) с проверкой их архитектурной целостности.

Адаптивность обеспечивается механизмами:

- машинного обучения, которые на основе наборов входных данных улучшают и совершенствуют алгоритмы обработки данных;
- обратной связи с пользователем, окружающей средой, обеспечивающих настройку интерфейса и сценария взаимодействия с ПС.

Таким образом, жизнеспособность расширяет традиционное понятие сопровождаемости, прежде всего, наличием механизмов адаптивности.

2. Потребность разных классов программных систем в развитии

Затраты на фазу сопровождения в жизненном цикле ПС занимают в среднем около 50 % [10, 11], имеются отчеты про долю до 80...90 % [12], хотя известно, что некоторым типам ПС (таким как встроенное программное обеспечение) сопровождение не требуется [1].

Потребуется ли модифицирование в связи с изменениями условий эксплуатации, требований пользователей, предметной области и др., зависит от назначения ПС, от уникальности (рыночная или заказная) и от типа обрабатываемой информации.

Если в качестве категорий ПС взять:

- прикладные ПС (в том числе СОЗ), создаваемые и приобретаемые для поддержки профессиональной деятельности специалистов или для улучшения обслуживания клиентов, а также для обучения,
- оболочки для создания прикладных ПС и
- инструментальные ПС,

то оценка потребности модифицирования приблизительно такова.

В процессе эксплуатации как прикладных ПС, так и инструментальных средств для их разработки появляется потребность в:

- исправлении ошибок;
- добавлении пользовательских функций (расширение);
- адаптации (новые приборы, устройства, платформы);
- изменении пользовательского интерфейса (ПИФ) специалистов;
- уточнении формата или состава информации (улучшение).

В случае с прикладными СОЗ, связанными с решением традиционных интеллектуальных

задач (диагностика, планирование, прогноз и т. д.), акценты смещаются. Здесь реже характерна потребность в добавлении новых пользовательских функций ("расширении"), скорее ожидается изменчивость знаний (появление новых методов диагностики, выявление новых факторов, влияющих на ситуации, и т. п.) и методов решения, объяснительной компоненты и пользовательского интерфейса. Причина в том, что постановки многих известных задач — диагностики, планирования и т. д. — достаточно устойчивы, а новые интеллектуальные задачи чаще всего являются конкретизацией/уточнением известных или их комбинацией. Развиваемость (модифицируемость) прикладных ПС также зависит и от инструментария для их разработки.

Разным категориям ПС свойственны и разные подходы к развитию (модернизации).

Для программных систем с заранее предопределенной архитектурой (обеспечиваемой фреймворками и оболочками) сопровождение и объединение разных компонентов заметно проще, поскольку проблема адаптируемости к устройствам, платформам, гибкость ПИФ, формата хранимой и другой информации "перекладывается" на отмеченные инструменты.

Сами фреймворки, оболочки, инструментальные среды подвергаются адаптации и расширению в рамках своих языковых сред (более низкого уровня), причем необходимые пользователю компоненты инструментария могут быть как "собственными", так и заимствованными или интегрированными из других инструментов (так, в Rational Rose-инструментарии "проектируются" собственно объектная модель, модель данных, компонентная модель, т. е. архитектурно существенные элементы", а "пользовательский интерфейс проектируется в среде Delphi" с преобразованием в программный код "в модели Rational Rose").

Подход к сопровождению прикладных программных систем схож с сопровождением средств разработки, т. е. инструментальных систем.

3. Особенности программных систем, основанных на знаниях

Системами, основанными на знаниях, называют класс таких программных систем искусственного интеллекта, в которых предметные знания представлены в явном виде

и отделены от прочих знаний системы. При разработке таких систем первична постановка задачи, знания, метод ее решения; типичное специальное требование предъявляется к представлению информации — все должно быть в форме, понятной пользователям без посредников: данные, знания, результат решения задач и ожидаемое объяснение результата.

Спецификация задачи на основе баз знаний есть предикат $P(x, k, y)$, где $x \in X$, $k \in K(X, Y)$, $y \in Y$. Особенностью спецификации задачи является то, что для нее не является справедливым утверждение о существовании решения [13]. Предполагается, что существует "правильная" база знаний $k^* \in K(X, Y)$, такая что $\forall x \in X \exists y \in Y P(x, k^*, y)$; однако предположение это не может быть доказано. В общем случае эта "правильная" база знаний неизвестна, а для любой другой базы знаний утверждение о существовании решения не обязано быть справедливым [13]. Но в предположении справедливости "слабого" утверждения (о существовании "правильной" базы знаний) можно разрабатывать алгоритм решения задачи (реализацию на алгоритмическом языке частично определенного функционального отображения $A: \langle X, K(X, Y) \rangle \rightarrow Y$) [13].

Такие задачи в современных СОЗ решаются "с применением эвристик, включая эмпирическую индукцию, аналогию и дедукцию" [14], часто с использованием более одного метода имитации интеллектуальной деятельности человека [15]. Современные СОЗ предлагают удобный ПИФ, они могут получать данные от современных средств измерения и часто интегрируются со статистическими и другими программными компонентами, не использующими формализованные знания (в том числе со стандартными пакетами прикладных программ) [16, 17].

Жизнеспособность для системы, основанной на знаниях, — это сохранение работоспособности и полезности в условиях изменчивости предметной области (чаще — знаний, реже — онтологии, среды функционирования и пользовательского интерфейса).

Поскольку жизнеспособность СОЗ проявляется, прежде всего, в условиях изменчивости знаний предметной области, а "способность к адаптации в соответствии с изменением множества фактов и знаний" считается одним из аспектов интеллекта [14], то для большинства предметных областей, связанных с решением интеллектуальных задач, подразумевается эво-

люционируемость знаний. Более того, в предметных областях, где важны влияние факторов и событий на состояние системы (объекта, ситуации), их изменение во времени, влияние индивидуальных характеристик системы и одних ее процессов на другие, развиваемость/эволюционируемость баз знаний — главный "вызов" современных "условий" по отношению к СОЗ.

Развиваемость баз знаний (БЗ) позволяет надеяться на получение "эталонной" базы знаний $k^* \in K(X, Y)$, от актуальности (соответствия современному представлению) и качества которой зависит успех применения СОЗ (для получения объяснения $y \in Y$ при любых $x \in X$).

4. Модель жизнеспособности систем, основанных на декларативных знаниях

Актуальность знаний (базы знаний) достигается тремя основными способами [3, 7, 18]: адаптируемостью (способностью к интерактивному изменению базы знаний), адаптивностью с использованием методов машинного обучения СОЗ (могут применяться средства индуктивного формирования знаний по отобраным прецедентам, средства выявления знаний из "больших данных"), либо их комбинацией.

"Успешность" адаптируемости (интерактивного изменения) зависит от онтологии предметной области, которая является системой понятий, удовлетворяющей всех участников разработки, и от ПИФ редактора БЗ, который должен соответствовать требованиям и ожиданиям экспертов предметной области. При этом онтология — структурная основа как инструментов для экспертов (средств редактирования), так и программных компонентов СОЗ (решателей интеллектуальных задач и ПИФ).

Таким образом, традиционная архитектура СОЗ — "база знаний + база фактов + решатель интеллектуальной задачи + ПИФ интеллектуальный" [14] — расширяется новым компонентом: "онтология + база знаний + база фактов + онтолого-ориентированный решатель задачи + ПИФ интеллектуальный".

Архитектурные свойства систем, основанных на декларативных знаниях. В онтологии предметной области определены типы утверждений (позволяющих решать задачи в предметной области) и ограничения на интерпретацию смысла понятий (терминов). На-

пример, для диагностики процессов одним из самых распространенных типов предложений (утверждений) является <класс отклонений- k , признак- j , диапазон значений- kj признака- j >; предложения типа "необходимое условие существования процесса", "вариант процесса изменения значений признака" и т. д.).

БЗ содержит предметные знания, которые представляются в явном виде. Они формируются вручную или индуктивно, в том числе по обучающим выборкам из архивов и баз данных (индуктивное обобщение в машинном обучении [14], байесовские классификаторы, алгоритмы кластеризации и обучение с подкреплением [19, 20]).

База фактов содержит набор фактов (утверждений), наблюдаемых или объективно измеренных в рассматриваемой ситуации, относительно которой решается задача.

Явное представление (и группирование) в онтологии всех структурных типов утверждений (и отделение онтологии от базы знаний) диктует необходимость "замены" решателей-рассуждателей, интерпретирующих продукционные правила, на специализированные алгоритмы на основе онтологий.

Онтолого-ориентированный алгоритм (ontological reasoner) выполняет для поиска либо опровержения гипотез обход (декларативной) базы знаний. Такой онтолого-ориентированный решатель задач "перебирает" утверждения БЗ каждого типа, относящегося к гипотезе. Он сопоставляет эти предложения входной информации (об объекте). Алгоритм обхода БЗ и сопоставления ее предложений элементам структуры объектов предметной области в общем случае несложен, так как число "правил сопоставления" определяется типами предложений БЗ, соответствующих решаемой задаче.

Реализация "обхода" декларативной БЗ в основном состоит в поиске значений элементов "оперативной" информации о ситуации в предметной области, относительно которой требуется принять решение, сравнения их с областями (возможных) значений соответствующих понятий (терминов) БЗ в целях выдвижения гипотез на основе декларативно представленных отношений между гипотезами и этими понятиями (терминами) в БЗ. (Пример: реализация "обхода" декларативной БЗ диагностики обычно состоит в поиске значений признаков у диагностируемого объекта и сравнения их с областями (ожидаемых) значений с целью отклонить или подтвердить соот-

ветствующие гипотезы о классах отклонений). Таким образом "структурная" сложность онтолого-ориентированного алгоритма, как правило, невелика. Его сложность определяется:

- числом и сложностью типов аксиом (предложений) онтологии;
- длиной цепочек причинно-следственных отношений (ПСО) между искомыми/проверяемыми гипотезами и наблюдениями — элементами описаний объектов предметной области (ПрОбл);
- числом наблюдений (элементов описаний) объектов и/или числом ограничений/целевых условий. Например, в задаче проектирования условия будут касаться размеров и разных свойств конструируемого объекта (моста, аппаратного комплекса...); в задаче лечения ограничение = "объект должен стать здоров"). На продолжительность обработки (вычисления) влияет объем БЗ.

Такие решатели способны стать повторно-используемыми, поскольку, во-первых, набор традиционно решаемых задач (диагностики, прогноза, управления, планирования....) известен [21, 22], во-вторых, традиционный набор типов связей в предметных областях с ПСО (для решения задач диагностики, прогноза, управления состоянием объекта) [21, 22] ограничен. Повторное использование состоит в выборе задачи в соответствии с классификацией задач и их уточнениями в соответствии с особенностями предметной области (дополнительными типами отношений в онтологии), поиске решателя задачи и, возможно, адаптации его к формату результата работы (объяснению).

Использование онтологии для формирования БЗ, данных (фактов), а также структуры объяснения дает возможность генерировать ПИФ в терминах (понятиях), понятных специалисту, а также, как правило, генерировать сценарий диалога, соответствующий структуре объяснения результатов.

СОЗ как разновидность ПС может также:

- включать в себя базы данных для хранения справочной или другой информации;
- принимать на вход файлы или базы данных с оперативной информацией (о ситуации, относительно которой требуется принять решение);
- создавать в качестве результата файлы или базы данных с результатами работы и объяснением принятого решения.

Таким образом, в составе жизнеспособной СОЗ должны быть:

- база знаний, выраженная в терминах онтологии;
- онтолого-базированный решатель (умеющий выдвигать и объяснять гипотезы);
- онтолого-зависимый адаптивный и адаптируемый ПИФ.

Наиболее характерные свойства СОЗ, относящиеся к жизнеспособности, таковы:

- регулярная обновляемость знаний;
- допустимость усовершенствования метода принятия решения;
- допустимость изменения или добавления функций (например, формирования дополнительных результатов);
- адаптивность ПИФ пользователя в связи с изменением входных данных и формируемых результатов функций;
- допустимость расширения онтологии (добавления понятий, отношений).

Изменения в онтологии, как правило, ведут к корректировке всех компонентов СОЗ; в этом случае целесообразно говорить о новой версии СОЗ.

Онтологический подход к средствам сопровождения и развития СОЗ. Онтология чаще соответствует классу решаемых задач или одному классу задач в одной предметной области. Для конструирования прикладных СОЗ может потребоваться не одна онтология, а их взаимосвязанная совокупность, в том числе для нескольких смежных классов задач (например, диагностика неполадок и исправление неполадок).

Инструментарий для конструирования прикладных систем, обрабатывающих декларативные знания, основан на онтологии предметной области, поскольку алгоритмы обрабатывают знания и другую информацию предметной области. Обрабатываемая информация — формализованные знания о законах предметной области, вводимые через ПИФ факты, хранимые в базах данных архивные данные и документы. Онтологические соглашения о правилах сопоставления фактов знаниям должны быть известны разработчикам алгоритмов решения задач. Знания и другая информация должны формироваться в единых понятиях предметной области (составляющих ее терминологический справочник) и в соответствии со структурой, определяемой онтологией ПрОбл, включающей онтологии знаний, входных данных и ожидаемых результатов.

Рассмотрим инструментарий, позволяющий создавать компоненты СОЗ на базе онтологии ПрОбл, специализированной среды развития СОЗ.

Специализированная среда развития СОЗ должна, как минимум, предоставлять:

- редактор БЗ;
- редактор БД;
- редактор структуры результатов решения и объяснения;
- библиотеку программных решателей, соответствующих классам решаемых задач (и их компонентов — программных единиц);
- средства поиска и выбора повторно используемых компонентов (ПИК) — БЗ и программных решателей;
- средства интеграции БЗ и решателей;
- специализированный редактор для ПИФ.

Примечание. В ситуации, когда не исключается возможность внесения изменений в онтологию ПрОбл, после построения БЗ среда развития СОЗ должна предоставлять инструмент редактирования онтологии. Внесение дополнительных понятий или новых связей между понятиями в онтологию не нарушает работоспособности программных решателей. (Онтолого-ориентированный решатель обладает таким свойством, поскольку обрабатывает в БЗ только известные ему типы предложений.)

Для формирования программных решателей понадобятся:

- средства кодирования новых программных единиц (ПрЕд) для решателя или их новых версий;
- средства каталогизации ПрЕд как ПИК;
- средства интеграции ПИК и новых ПрЕд в новые решатели или их новые версии.

Примечание. Возможность создания программных единиц для решателя может оказаться излишней (в среде развития), если готовый решатель вырабатывает правильные гипотезы в соответствии с постановкой задачи, а "наращивание" дополнительной функциональности не предполагается.

Ввиду важности развиваемости и эволюционируемости БЗ следует всерьез рассматривать только такие СОЗ, которые интегрированы с СУБЗ, т. е. программные компоненты для развития баз знаний, например, средства проверки баз знаний, оценивания их качества на эталонных архивах решенных задач — часть "интегрированной архитектуры". Тогда среда развития может предоставлять:

- средства проверки баз знаний;
- средства оценивания качества баз знаний на эталонных архивах решенных задач;
- средства индуктивного формирования баз знаний или фрагментов БЗ, интегрируемых с ней.

Вышеперечисленным свойствам жизнеспособной СОЗ соответствуют характерные свойства специализированной среды ее развития:

- поддержка обновления знаний;
- поддержка изменения компоновки онтолого-ориентированного решателя (в связи с заменой компонента, реализующего другую стратегию принятия решения или метода получения результата или в связи с добавлением компонента, реализующего дополнительную функцию, например, формирования дополнительного результата);
- поддержка усовершенствования ПИФ пользователя в связи с изменением функций;
- поддержка изменения в онтологии;
- поддержка усовершенствования ПИФ эксперта (и ПИФ пользователя) в связи с обновлением онтологии;
- поддержка кодирования новых версий программных единиц онтолого-ориентированного решателя или поддержка изменения кода ПрЕд онтолого-ориентированного решателя (т. е. адаптации обхода декларативной БЗ в связи с обновлением онтологии знаний).

Пример реализации сред развития СОЗ (в рамках онтологического подхода). Онтология чаще соответствует классу решаемых задач или одному классу задач в одной предметной области. Для конструирования прикладных СОЗ может потребоваться не одна онтология, а их взаимосвязанная совокупность, в том числе для нескольких смежных классов задач (например, диагностика неполадок и исправление неполадок).

Примером инструментария для формирования сред развития (для производства СОЗ) является реализованный на платформе IACPaas комплекс инструментов для обработки информационных ресурсов, формируемых по явно представляемой онтологии (на платформе используется термин "метайнформация") [23]. Онтология задает правила порождения информации, правила и ограничения ее интерпретации. Структуру порождаемой информации и ряд ограничений интерпретации определяет пользователь платформы IACPaas для

формирования собственных информационных ресурсов (когнитолог — для сообщества экспертов, специалистов, пользователей).

СОЗ является частным случаем прикладных программных сервисов на платформе IACPaaS (сервисов IACPaaS).

Комплекс инструментов для формирования информационных компонентов СОЗ таков:

- IACPaaS-Редактор онтологии;
- IACPaaS-Редактор БЗ, генерирующийся по онтологии с самоадаптирующимся ПИФ (при изменении онтологии);
- IACPaaS-Редактор БД (с самоадаптирующимся ПИФ).

Этот комплекс инструментов также содержит средства для управления программными компонентами СОЗ:

- "Мастер" формирования декларативной части ПрЕд (IACPaaS-агентов);
- конструктор решателя из корневого IACPaaS-агента и новых агентов (их версий), представляемого только декларативной частью;
- генератор заготовок программного кода для новых IACPaaS-агентов (новых версий);
- загрузчик байт-кода в декларативную часть IACPaaS-агента;
- средства тестирования ПрЕд и подготовки их для библиотеки ПИК (агентов).

IACPaaS-агенты кодируются на языке Java, для написания исходных кодов и создания байт-кода агента рекомендована среда разработки (IDE) Java Development Kit (JDK) для версии Java SE 8.

Среда развития СОЗ подразумевает кодирование новых версий агентов (когда надо учесть использование новых терминов или новых связей терминов) или кодирование новых IACPaaS-агентов (когда надо внести изменение в процесс принятия решения). Она может быть локальной для программиста (или размещенной на платформе IACPaaS).

Примечание. На платформе IACPaaS "Мастер" формирования декларативной части ПрЕд и Конструктор уникального решателя являются редакторами, генерируемыми по соответствующей метаинформации — шаблонам декларирования ПрЕд и решателя (из онтологии технологии IACPaaS). Самоадаптирующийся ПИФ для формирования данных (и знаний) и просмотра результата работы СОЗ предоставляет три способа визуализации, один из которых является графовым.

Таким образом, построение СОЗ с помощью онтолого-ориентированных сред развития принципиально увеличивает их жизнеспособность, существенно повышая роль и долю средств управления (для внесения изменений в декларативные компоненты) по отношению к средствам сопровождения (для внесения изменений в исходный код).

5. Обеспечение жизнеспособности СОЗ разными инструментами сред развития IACPaaS

Вышеописанные характерные свойства СОЗ, относящиеся к жизнеспособности, обеспечиваются в средах развития, реализуемых на платформе IACPaaS. Рассмотрим это на примере свойства *Обновляемость знаний и поддержка обновления знаний*.

Если требуется обновление базы знаний в связи с получением новых знаний (утверждений), естественно модифицировать их вручную (и оценить согласованность с имеющимися фактами). Для этого процесса (в среде развития) требуются редакторы для всех БЗ.

На платформе IACPaaS редактор Информационных Ресурсов (штатный) всегда является частью среды развития. В настоящее время разрабатывается средство проверки качества новой версии БЗ — проверки неухудшения правильности решения множества эталонных задач при замене версии БЗ (согласно важности монотонного усовершенствования баз знаний [24]). Способ проверки неухудшения правильности знаний: подача на вход решателю, интегрированному с новой версией БЗ, входных условий эталонной задачи и сравнение полученного объяснения с выходным результатом эталонной задачи.

Если требуется обновление БЗ в связи с получением "фактов" (прецедентов, решений), не согласованных со знаниями (когда прецедент содержит правильный результат решения задачи, который не соответствует результату, получаемому с помощью СОЗ), то эффективно автоматически (индуктивными методами) формировать новую версию БЗ (и оценить согласованность работы СОЗ по обновленной БЗ с имеющимися прецедентами). (Пример такого обновления знаний: из медучреждения через определенный промежуток времени поступает правильный результат диагностики или лечения, он сравнивается с результатом (объяснением гипотез) от СОЗ: не противоречат ли они друг другу.)

Для реализации процесса монотонного усовершенствования БЗ (на основе методов индуктивного формирования фрагментов БЗ [24]) требуются:

- средства индуктивного формирования знаний для каждой решаемой интеллектуальной задачи (с известной постановкой: диагностика, планирование, прогноз...);
- средство поддержки выбора прецедентов (решенных правильно задач в одной постановке);
- средство проверки правильности (качества) новой версии БЗ (тот же — вышеописанный).

Способ проверки наличия (у СОЗ) свойства "Обновляемость знаний" — проверка наличия и работоспособности вышеперечисленного инструментария.

На платформе IASPaas штатный (самоадаптирующийся) редактор всегда есть, остальные инструменты добавляются в среду развития по мере готовности.

Заключение

Предложенная модель жизнеспособности СОЗ основана на построении такой системы с помощью онтолого-ориентированной среды развития, где каждый компонент СОЗ формируется в терминах (понятиях) единой явно представленной онтологии ПрОбл с использованием инструментальной поддержки.

Онтолого-ориентированная среда развития реализует механизмы адаптивности таких систем. Она развивается и эволюционирует, когда создаваемые в ней прикладные СОЗ нуждаются в адаптации к условиям функционирования и уточнению знаний ПрОбл, а также когда "базовый инструментарий" нуждается в адаптации к новой аппаратуре и системному ПО, к новым видам человеко-машинных интерфейсов. Эти механизмы адаптивности обеспечивают возможность усовершенствования каждого компонента силами узких специалистов, гарантируя целостность всей СОЗ.

Список литературы

1. Pressman R. S. Software Engineering: Practitioner's Approach. 7th ed. McGraw-Hill, 2010. 930 p.
2. Dehaghani S. M. H., Hajrahimi N. Which factors affect software projects maintenance cost more? // *Acta Informatica Medica*. 2013. N. 21(1). P. 63–66.
3. Додонов А. Г., Ландэ Д. В. Живучесть информационных систем. Киев: Наукова Думка, 2011. 256 с.

4. Черников Б. В. Управление качеством программного обеспечения. М.: ИД "ФОРУМ": ИНФРА-М, 2012. 240 с.
5. Musen M. The Protégé Project: A Look Back and a Look Forward // *AI Matters*. 2015 Jun; N. 1(4). P. 4–12.
6. Mortensen J., Minty E., Januszyk M., Sweeney T., Rec-tor A., Noy N., Musen M. A. Using the wisdom of the crowds to find critical errors in biomedical ontologies: A study of SNOMED CT // *Journal of the American Medical Informatics Association*. 2015. N. 22(3). P. 640–648.
7. Кряжич О. А. Обеспечение жизнеспособности информации во времени при ее обработке в СППР // *Математичні машини і системи*. 2015. № 2. С. 170–176.
8. Izurieta C., Bieman J. M. A multiple case study of design pattern decay, grime, and rot in evolving software systems // *Software Quality Journal*. 2013. Vol. 21, N. 2. P. 289–323.
9. Breivold H. P., Crnkovic I., Eriksson P. J. Analyzing software evolvability // *Computer Software and Applications*, 2008. COMPSAC'08. 32nd Annual IEEE International. IEEE. 2008. P. 327–330.
10. Брауде Э. Технология разработки программного обеспечения. СПб.: Питер, 2004. 655 с.
11. Гласс Р., Нуазо Р. Сопровождение программного обеспечения. М.: Мир, 1983. 158 с.
12. Islam M., Katiyar V. Development of a software maintenance cost estimation model: 4th GL perspective // *International Journal of Technical Research and Applications*. 2014. Vol. 2, Iss. 6. P. 65–68.
13. Грибова В. В., Клещев А. С. Какой должна быть парадигма решения задач на основе баз знаний? // *Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2014): Матер. IV Междунар. науч.-техн. конф. (Минск, 20–22 февраля 2014 года) / редкол. В. В. Голенков (отв. ред.) [и др.]*. Минск: БГУИР, 2014. С. 131–136.
14. Финн В. К. Об интеллектуальном анализе данных // *Новости искусственного интеллекта*. 2004. № 3. С. 3–18.
15. Гаврилова Т. А., Кудрявцев Д. В., Муромцев Д. И. Инженерия знаний. Модели и методы: учебник. СПб.: Лань, 2016. 324 с.
16. Рыбина Г. В., Блохин Ю. М., Ивашенко М. Г. Интеллектуальная технология построения интегрированных экспертных систем // *Искусственный интеллект и принятие решений*. 2011. № 3. С. 48–57.
17. Рыбина Г. В. Обучающие интегрированные экспертные системы: некоторые итоги и перспективы // *Искусственный интеллект и принятие решений*. 2008. № 1. С. 22–46.
18. Грибова В. В., Клещев А. С. Парадигма разработки управляемых интеллектуальных систем // *Системы управления и информационные технологии*. 2016. № 3(65). С. 32–38.
19. Николенко С. И., Тулупьев А. Л. Самообучающиеся системы. М.: Изд-во МЦНМО, 2009. 288 с.
20. Golencov V. V., Gulyakina N. A., Grakova N. V., Nikulenkova V. Y., Ereemeev A. P., Tarasov V. B. From training intelligent systems to training their development means // *Proceedings of the International Conference OSTIS-2018*. Minsk. 2018. P. 81–98.
21. Clancey W. J. Heuristic Classification // *Artificial Intelligence*. 1985. N.27. P. 289–350.
22. Клещев А. С., Шалфеева Е. А. Онтология задач интеллектуальной деятельности // *Онтология проектирования*. 2015. Т. 5, № 2(16). С. 179–205.
23. Грибова В. В., Клещев А. С., Москаленко Ф. М., Тимченко В. А., Шалфеева Е. А. Расширяемый инструментарий для создания жизнеспособных систем с базами знаний // *Программная инженерия*. 2018. Т. 9, № 8. С. 339–348.
24. Клещев А. С., Черняховская М. Ю., Шалфеева Е. А. Парадигма автоматизации интеллектуальной профессиональной деятельности. Часть 1. Особенности интеллектуальной профессиональной деятельности // *Онтология проектирования*. 2013. № 3(9). С. 53–69.

Ensuring of Viability of Systems Based on Knowledge

The model of viability of knowledge based systems designed by means of the ontology-oriented development environment is considered in the article. In the conditions of high labor costs on software maintenance the maintenance problem for knowledge-based systems is particularly acute. Their architecture includes the knowledge base created by domain experts and cognitologists. It imposes need of use of additional specialized mechanisms of providing maintenance and viability. The generalized architectural model of an accompanied knowledge-based system and set of means of ensuring of her viability is presented in the paper. It is shown that a viable system should include: a knowledge base formalized in terms of ontology, an ontology-based solver and ontology-dependent adaptable GUI. The structure of instrumental specialized environment of development for creation components of any knowledge-based system is defined. Such environment of development is ontology-oriented. It evolves together with enhancement of knowledge, expansion of ontology and in process of adaptation to operating conditions. These mechanisms of adaptability provide a possibility of improvement of each component by narrow experts, guaranteeing integrity of all system. It is shown how different aspects of viability are being provided in the environments of development implemented on the IACPaaS platform.

Keywords: intellectual systems, knowledge base, knowledge base system, ontology-oriented algorithm, viability, declarative tools, development environment

DOI: 10.17587/it.25.738-746

References

1. Pressman R. S. Software Engineering: Practitioner's Approach, McGraw-Hill, 2010, 930 p.
2. Dehaghani S. M. H., Hajrahimi N. *Acta Informatica Medica*, 2013, no. 21(1), pp. 63–66.
3. Dodonov A. G., Landje D. V. Viability of information systems, Kiev, Naukova dumkab, 2011, 256 p. (in Russian).
4. Chernikov B. V. Software quality management, Moscow, ID "FORUM": INFRA-M, 2012, 240 p. (in Russian).
5. Musen M. *AI Matters*, 2015 Jun, no. 1(4), pp. 4–12.
6. Mortensen J., Minty E., Januszyk M., Sweeney T., Rec-tor A., Noy N., Musen M. A. *Journal of the American Medical Informatics Association*, 2015, no. 22(3), pp. 640–648.
7. Krjzhich O. A. *Matematichni mashini i sistemi*, 2015, no. 2, pp. 170–176 (in Russian).
8. Izurieta C., Bieman J. M. *Software Quality Journal*, 2013, vol. 21, no. 2, pp. 289–323.
9. Breivold H. P., Crnkovic I., Eriksson P. J. *Computer Software and Applications*, 2008, COMPSAC'08, 32nd Annual IEEE International, pp. 327–330.
10. Braude E. J., Bernstein M. E. Software Engineering: Modern approaches (2nd edition), John Wiley & Sons, Inc., 2010, 800 p.
11. Glass R. L., Noiseux R. Software Maintenance Guide-book, Prentice Hall, 1981, 193 p. (in Russian).
12. Islam M., Katiyar V. *International Journal of Technical Research and Applications*, 2014, vol. 2, iss. 6, pp. 65–68.
13. Gribova V. V., Kleshhev A. S. *Open Semantic Technologies for Intelligent Systems (OSTIS-2014): materialy IV mezhdunar. nauch.-tehn. konf.* (Minsk, 20–22 fevralja 2014 goda), Minsk, Publishing house of BGUIR, 2014, pp. 131–136 (in Russian).
14. Finn V. K. *Novosti Iskusstvennogo intellekta*, 2004, no. 3, pp. 3–18 (in Russian).
15. Gavrilova T. A., Kudrjavec D. V., Muromcev D. I. Engineering of knowledge. Models and methods: Textbook, SPb., Lan', 2016, 324 p. (in Russian).
16. Rybina G. V., Blohin Ju. M., Ivashhenko M. G. *Iskusstvennyj Intellekt i Prinjatje Reshenij*, 2011, no. 3, pp. 48–57 (in Russian).
17. Rybina G. V. *Iskusstvennyj Intellekt i Prinjatje Reshenij*, 2008, no. 1, pp. 22–46 (in Russian).
18. Gribova V. V., Kleshhev A. S. *Sistemy Upravljenija i Informacionnye Tehnologii*. 2016, no. 3(65), pp. 32–38 (in Russian).
19. Nikolenko S. I., Tulup'ev A. L. *Samoobuchajushiesja sistemy*, Moscow, Publishing house of MCNMO, 2009, 288 p. (in Russian).
20. Golenkov V. V., Gulyakina N. A., Grakova N. V., Nikulenka V. Y., Ereemeev A. P., Tarasov V. B. *Proceedings of the International Conference OSTIS-2018*, Minsk, 2018, pp. 81–98.
21. Clancey W. J. *Artificial Intelligence*, 1985, no. 27, pp. 289–350.
22. Kleshhev A. S., Shalfeeva E. A. *Ontologija proektirovanija*, 2015, vol. 5, no. 2(16), pp. 179–205 (in Russian).
23. Gribova V., Kleshev A., Moskalenko Ph., Timchenko V., Shalfeeva E. A. *Programmnaja Inzhenerija*, 2018, vol. 9, no. 8, pp. 339–348 (in Russian).
24. Kleshev A. S., Chernyakhovskaya M. Y., Shalfeeva E. A. *Ontologija proektirovanija*, 2013, no. 3(9), pp. 53–69 (in Russian).

Р. А. Соловьев, д-р техн. наук, вед. науч. сотр., turbo@ippm.ru,
Д. В. Тельпухов, д-р техн. наук, зав. отделом, turbo@ippm.ru,
А. Г. Кустов, инженер-исследователь, turbo@ippm.ru,
Т. Ю. Исаева, канд. техн. наук, науч. сотр., turbo@ippm.ru,
А. А. Волков, инженер-исследователь, turbo@ippm.ru,
Институт проблем проектирования в микроэлектронике РАН

Применение методов модулярной арифметики при разработке аппаратных реализаций нейронных сетей¹

Обсуждается использование методов модулярной арифметики для реализации нейронных сетей на аппаратном уровне в СБИС и ПЛИС. Рассматриваются широкоизвестные мобильные нейронные сети, которые имеют высокую точность и подходят для реализации "в железе". Рассмотрены и проанализированы основные трудности при их реализации в базе модулярной арифметики. Предложен ряд методов для решения этих проблем, таких как использование сверток со значением шага больше 1 вместо используемых ранее MaxPooling слоев, использование нестандартных активаций, содержащих только операции сложения, вычитания и умножения, а также эффективный алгоритм реализации операции округления. Дополнительно предложен полный маршрут проектирования и переноса нейронной сети MobileNet на аппаратный уровень в модулярном базисе.

Ключевые слова: нейронные сети, модулярная арифметика, система остаточных классов, операция округления, аппаратная реализация

Введение

Нейронные сети хорошо справляются с множеством задач, связанных с классификацией и обработкой изображений, аудио- и видеоданных, в некоторых случаях даже лучше человека [1, 2]. Большинство современных архитектур имеют в своем составе сверточные (convolution) блоки (например, VGG [1], Inception [3], ResNet [4], U-Net [5]), поэтому такие сети называются сверточными (convolutional neural nets (CNN)). Вычислительная сложность во время классификации так велика, что с вычислениями плохо справляются даже мощные процессоры общего назначения CPU. Для полноценной работы с современными нейросетями используют мощные и дорогие GPU (видеокарты) [6]. Особенно эта проблема актуальна при обработке видеoinформации в реальном времени.

Некоторые структуры нейросетей при очень высокой точности классификации изображений обладают свойствами, которые позволяют легко переносить их на аппаратную платформу. Существует направление, связанное с проектированием нейронных сетей для использования в мобильных устройствах: MobileNets [7], SqueezeNet [8]. Эти варианты отличаются малым числом весов и относительно небольшим числом арифметических операций. Однако они также выполняются на программном уровне и используют для работы вычисления с плавающей точкой. К сожалению, в ряде случаев, например, в задачах обработки видеoinформации в реальном времени, даже при использовании мобильных сетей не всегда удается обеспечить непрерывную обработку видеопотока со скоростью 30 кадров в секунду без использования существенной оптимизации.

Для использования нейросетей на этапе, когда у нас есть обученная модель в реальном устройстве, можно применить набор оптимизаций, чтобы ускорить выполнение вычислений в несколько раз. Для этого уже существует

¹Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 17-07-00404.

ряд методик, например, компрессия весов [9] или вычисления на малобитных данных [10].

Поскольку потребности в аппаратуре для работы с нейронными сетями постоянно растут, то требуется разработка специальных аппаратных блоков для использования в СБИС и ПЛИС для ускорения расчетов. Ускорение расчетов может достигаться за счет:

- аппаратной реализации свертки. Аппаратный блок для вычисления свертки работает быстрее, чем свертка, выполненная на программном уровне;
- перехода от вычислений с плавающей запятой к фиксированной запятой;
- уменьшения размерности вычислений с сохранением приемлемой точности;
- сокращения части нейронной сети с сохранением точности классификации;
- модификации структуры нейронной сети с незначительным уменьшением точности (или даже без нее) с увеличением скорости работы и уменьшением размера аппаратной реализации и хранимых весов.

Дополнительное ускорение расчетов или снижение потребляемой энергии при расчетах с использованием нейронных сетей, как показали современные исследования, может достигаться за счет использования модулярной арифметики.

В статье [11] предлагается использовать вариацию рекурсивной модулярной арифметики [26] для представления весов и карт признаков. Для этого массивный 48-битный МАС-юнит заменяется на большой набор 4-битных модулярных МАС-юнитов, которые реализованы в виде Lookup-таблиц в ПЛИС. За счет существенного уменьшения размерности модулей до 4 бит удалось значительно увеличить тактовую частоту устройства со 100...200 МГц до 400 МГц, а также существенно увеличить производительность устройства в целом.

В статье [12] предлагается усовершенствованный МАС-Unit на базе модулярной арифметики и системы модулей вида $\{2^{l_1}, 2^{l_2} - 1, \dots, 2^{l_N} - 1\}$. Немного уменьшая точность сети за счет округления весов и промежуточных вычислений, авторам удалось увеличить скорость работы в 1,5...2,4 раза и сократить потребление энергии в 3—9 раз. Точность классификации нейронной сети AlexNet при этом упала всего с 79 до 75 %.

Авторы работы [13] представили архитектуру сверточной нейронной

сети с вычислениями в системе остаточных классов с применением модулей специального вида, за счет чего достигли увеличения производительности на 37 % по сравнению с двоичной реализацией.

Указанные исследования показывают значительный потенциал модулярной арифметики при проектировании аппаратной реализации нейронных сетей. В данной статье исследуются методы, которые подойдут для реализации нейронных сетей в небольших специализированных микроэлектронных устройствах, в которых использование GPU является нерентабельным по причине существенного энергопотребления и высокой стоимости.

1. Модулярная арифметика для вычислений на базе нейронных сетей.

Преимущества и недостатки

Типовая структура сверточной нейронной сети приведена на рис. 1. Позже появились некоторые усовершенствованные структуры с разветвлениями [3, 4], но суть осталась та же: размер изображения от слоя к слою уменьшается, а число фильтров растет. В конце сверточной сети образуется набор признаков, которые подаются на классификационный слой (или слои), и выходные нейроны сигнализируют вероятность принадлежности изображения к конкретному классу.

Для ускорения вычислений обычно переходят от вычислений на программном уровне к аппаратной реализации, а также от вычислений с плавающей точкой к фиксированной точке [14, 15]. В данном случае открываются существенные возможности с точки зрения использования полезных свойств модулярной арифметики [26—28] при проектировании устройства, реализующего нейронную сеть:

1) модулярная арифметика плохо справляется с вычислениями на базе плавающей точки, однако при работе с фиксированной точкой,

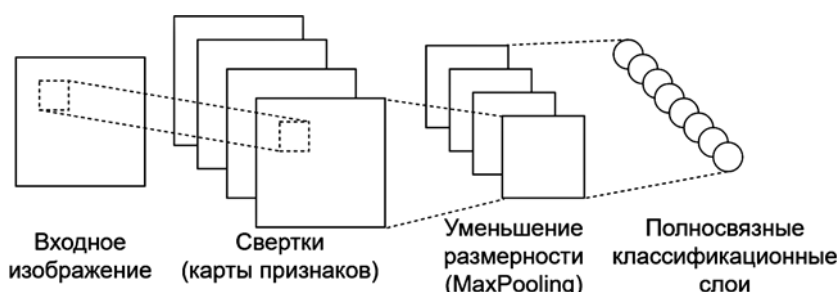


Рис. 1. Типовая структура сверточной нейронной сети

где заранее известен динамический диапазон вычислений, может быть подобран подходящий модулярный базис;

2) веса предобученной нейронной сети известны заранее, поэтому их можно хранить сразу в модулярном виде на базе известного модулярного базиса, т. е. не потребуются аппаратная реализация прямого преобразователя для весов, объем которых очень большой;

3) основные вычислительные блоки нейронной сети — сверточные и полносвязные слои, которые требуют 90 и более процентов вычислительной мощности [16], — имеют в своем составе только операции сложения и умножения. Именно на этих операциях модулярная арифметика покаывает свое преимущество перед позиционной;

4) входные данные в нейронную сеть подаются в виде цветного изображения небольшого размера. Затем над этими данными выполняется большой набор арифметических операций по мере движения данных от слоя к слою по нейронной сети. Результат при этом в случае классификации — это небольшой набор выходных чисел. Это означает, что прямые и обратные преобразователи будут выполнять совсем немного вычислений по сравнению с основным модулярным трактом.

Однако есть несколько проблем при использовании некоторых типов слоев, которые могут привести к уменьшению производительности при использовании модулярной арифметики:

1) в нейронных сетях для уменьшения размера промежуточных карт признаков используется блок MaxPooling или AvgPooling. Их главная проблема с точки зрения модулярной арифметики в первом случае — наличие операции Max, которая является частным случаем сравнения чисел и представляет собой сложную операцию в модулярной арифметике и требует отдельного блока большого размера. Во втором случае требуется вычисление среднего, что влечет за собой операцию деления на 4, которая тоже может вызывать трудности. Однако эта проблема может быть решена использованием вместо MaxPooling сверточного блока с параметром strides, равным 2, что уже широко используется в современных сетях, например в MobileNet. В случае использования сверточного блока вместо MaxPooling остаются только операции сложения и умножения;

2) слои типа Batch Normalization содержат деление, но после сокращения констант остаются только одна операция сложения и одна умножения. Дополнительно в случае если слой Batch Normalization идет сразу после сверточного бло-

ка, а именно так он располагается в большинстве современных нейронных сетей, его можно удалить, пересчитав веса у сверточного слоя;

3) основную проблему представляют слои активации нейронов, которые присутствуют почти после всех сверточных слоев и добавляют нелинейность. Самые популярные активации в современных сетях — Rectified linear unit (RELU) из-за его простоты с точки зрения имплементации в позиционной арифметике и Sigmoid или Softmax, которые обычно используются на выходных слоях сети:

$$\text{RELU} \rightarrow f(x) = \begin{cases} 0, & x < 0; \\ x, & x \geq 0; \end{cases}$$

$$\text{Sigmoid} \rightarrow f(x) = \frac{1}{1 + e^{-x}}.$$

RELU крайне проста для позиционной арифметики, но из-за операции сравнения становится сложной в модулярной арифметике. По-другому сравнение с нулем называется операцией определения знака [17]. Sigmoid вычислительно сложна в обоих случаях из-за наличия функции e^{-x} и деления.

Одним решением проблемы может быть разработка функции активации, которая содержит только "дружественные" для модулярной арифметики операции, либо создание эффективной аппаратной реализации одной из существующих активаций. Одной из таких активаций является RELU. Для определения знака можно использовать эффективные алгоритмы [18]. Однако более эффективным методом является подход с разработкой функций активации для модулярной арифметики.

2. Нестандартные функции активации

В работе были изучены функции активации, которые содержат только операции сложения и умножения. Экспериментальным путем было показано, что функция активации $f(x) = x^2 - kx$ при некоторых значениях k хорошо сходится. Была построена сеть MobileNet, в которой все слои RELU заменены на предложенную функцию, и показано, что для задачи классификации автомобилей на наборе Open Images Dataset [24] нейронная сеть обучается и показывает точность классификации до 93 % (рис. 2). Точность на валидационных данных чуть выше, чем на тренировочном наборе из-за того, что к входным данным применялись очень сильные аугментации.

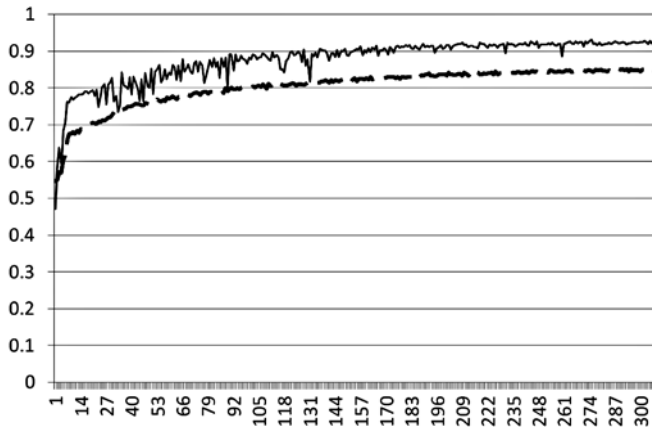


Рис. 2. Процесс тренировки MobileNet с функциями активации $f(x) = x^2 - 50x$. По горизонтали номер эпохи, по вертикали — точность классификации. Штриховая линия — точность на тренировочных данных, сплошная линия — точность на валидационных данных

Предложенная функция активации позволяет избавиться от операции сравнения при реализации нейронной сети и тем самым устранить узкое место для аппаратной реализации на базе модулярной арифметики.

3. Переход от вычислений с плавающей точкой к вычислениям с фиксированной точкой

В нейронных сетях вычисления традиционно проводятся с плавающей точкой на GPU (быстро) или CPU (медленно), например, с использованием типа float32. При реализации на аппаратном уровне вычисления с плавающей точкой работают медленнее, чем с фиксированной точкой из-за сложностей с контролем мантиссы и показателя степени при различных операциях.

Рассмотрим первый слой нейросети типа Convolution (рис. 3, см. третью сторону обложки) — в большинстве сверточных нейросетей этот слой является основным.

На входе слоя находится двумерная матрица (исходное изображение, значения которой находятся на интервале $[0; 1]$).

Известно также, что если $a \in [-1; 1]$ и $b \in [-1; 1]$, то произведение $ab \in [-1; 1]$.

Формула для расчета конкретного пикселя в позиции (i, j) второго слоя с учетом того, что используется Convolution с размером 3×3 :

$$n_{ij} = b + w_{00}p_{i-1,j-1} + w_{01}p_{i-1,j+0} + w_{02}p_{i-1,j+1} + \\ + w_{10}p_{i+0,j-1} + w_{11}p_{i+0,j+0} + w_{12}p_{i+0,j+1} + \\ + w_{20}p_{i+1,j-1} + w_{21}p_{i+1,j+0} + w_{22}p_{i+1,j+1}.$$

Поскольку веса $w_{i,j}$ и смещение b известны, то можно рассчитать потенциальный минимум mn и максимум mx значений на входе второго слоя. Пусть $M = \max(|mx|, |mn|)$. Если разделить $w_{i,j}$ и b на значение M , то можно гарантировать, что при любой конфигурации входных данных значение на втором слое не превысит 1. Назовем M коэффициентом редукции слоя. На втором слое получается такая же ситуация, как и на первом, а именно на входе слоя значение из промежутка $[-1; 1]$ и рассуждения можно повторить.

Можно легко показать, что для нейронной сети на последнем слое после всех редукций весов позиция максимума на последнем нейроне не изменится, т. е. нейронная сеть будет работать эквивалентно нейронной сети без редукции с точки зрения вычислений с плавающей точкой.

Выполнив редукцию на каждом слое, мы можем перейти от вычислений с плавающей точки к вычислениям с фиксированной точкой, поскольку мы точно знаем диапазон значений на каждом этапе вычислений. Для представления чисел размерности N бит будем использовать следующую нотацию:

$$xb = \lfloor x \cdot 2^N \rfloor.$$

Если $z = x + y$, тогда сложение $z' = xb + yb = \lfloor x \cdot 2^N \rfloor + \lfloor y \cdot 2^N \rfloor = \lfloor (x + y) \cdot 2^N \rfloor = \lfloor z \cdot 2^N \rfloor = \lfloor zb \rfloor$. Умножение $z' = xb \cdot yb = \lfloor x \cdot 2^N \rfloor \cdot \lfloor y \cdot 2^N \rfloor = \lfloor (x \cdot y) \cdot 2^N \cdot 2^N \rfloor = \lfloor z \cdot 2^N \cdot 2^N \rfloor = \lfloor zb \cdot 2^N \rfloor$, т. е. после умножения требуется разделить результат на 2^N , чтобы получить реальное значение, или же просто выполнить сдвиг на N позиций.

Если перебирать все возможные входные изображения и ориентироваться на потенциальный минимум и максимум, то коэффициенты редукции могут оказаться очень большими, и точность будет теряться от слоя к слою довольно быстро, что может потребовать большой разрядности фиксированной точки для хранения весов и промежуточных результатов вычислений. Чтобы этого не произошло, можно использовать все (или некоторую часть) изображения тренировочного набора как наиболее вероятные, чтобы найти максимальные и минимальные значения на каждом слое. Как показали эксперименты, использование тренировочного набора позволяет сильно уменьшить коэффициенты редукции. При этом желательно брать коэффициенты с небольшим запасом, например, увеличивая значение максимума на несколько процентов.

Однако при некоторых условиях возможно переполнение и выход за границы рассчитанного диапазона. Для этого в аппаратной реализации требуется детектор таких случаев и замена переполненных значений на максимум для данного слоя. Впрочем, этого можно достигнуть лишь небольшой модификацией блока выполнения свертки.

При вычислениях с фиксированной точкой с ограниченной разрядностью весов и промежуточных вычислений неизбежно возникают ошибки округления, которые накапливаются от слоя к слою и могут привести к некорректной работе нейронной сети. Некорректной работой мы считаем ошибку классификации при сравнении с математической моделью, а не с реальным ответом.

При использовании модулярной арифметики основной проблемой при вычислениях становится операция округления после всех операций сложений и умножений в сверточном блоке. Если же постараться избежать этой операции, то размерность чисел и требуемый динамический диапазон будут сильно расти от слоя к слою, и потребуется слишком большой набор модулей, что сделает применение модулярной арифметики менее эффективным. Для операции округления в модулярной арифметике существуют специальные быстрые методы [19–22].

4. Метод проектирования нейронной сети на базе модулярной арифметики

Рассмотрим последовательность действий, требуемых для подготовки аппаратной реализации нейронной сети на конкретном примере. Возьмем в качестве нейронной сети известную и хорошо зарекомендовавшую себя на практике сеть MobileNet [7]. Она имеет большое число слоев (около 100), не содержит разветвлений и отличается небольшим числом весов, что позволяет хранить их в памяти даже для небольших аппаратных устройств. Точность классификации на наборе ImageNet достигает 70,6 %.

Всего MobileNet включает в себя восемь типов слоев:

1) ZeroPadding2D — добавляет ободок из нулей вокруг изображения. Слой требуется для выполнения свертки таким образом, чтобы размер выходной карты признаков совпадал с размером входной карты признаков;

2) Conv2D ($\text{kernel} = 3 \times 3$) — двумерная свертка с ядром размера 3×3 ;

3) Conv2D ($\text{kernel} = 1 \times 1$) — двумерная свертка с ядром размера 1×1 ;

4) DepthwiseConv2D — упрощенная версия Conv2D (меньше сложений, умножений и соответственно меньше матриц с весами);

5) Activation (Relu6) — функция активации. Является усовершенствованной функцией RELU, в ней появляется дополнительное ограничение сверху. Все значения, большие 6, приравниваются 6. Это не дает потенциальным максимальным значениям уходить далеко от 0 и очень полезно для представления чисел с фиксированной точкой;

6) GlobalAvgPooling — слой, который используется в нейронной сети только один раз, является предпоследним слоем. У него нет весов, и он находит среднее значение от карт признаков предыдущего слоя. Генерирует вектор признаков изображения (длина вектора обычно от 512 до 4096);

7) Dense (Fully Connected) — последний полносвязный слой сети. Выполняет классификацию изображения. В нашей задаче он состоит из двух нейронов. Первый сигнализирует о том, что объект отсутствует, второй — о том, что объект присутствует. На выходе этого слоя используется активация Softmax, которая возвращает вероятности наличия искоемых объектов;

8) BatchNormalization — слой, выполняющий нормализацию и идущий в связке с каждым сверточным слоем.

Шаг 1. Работа начинается с подготовки тренировочного набора данных. Для простоты положим, что нам требуется разделить изображения на два класса по какому-то признаку, например, определить, есть ли на изображении человек или нет, что может быть полезно для камеры слежения или мобильных телефонов.

Шаг 2. На данном шаге требуется выбрать размер входного изображения для нейронной сети, от которого будет зависеть ее конфигурация. Типовые размеры изображений подходящие для MobileNet: 128×128 , 160×160 , 192×192 и 224×224 пикселей.

Шаг 3. Тренируем сеть на нашем наборе данных, используя любой подходящий фреймворк: tensorflow, pytorch, caffe, keras и т. д. Цель — получить веса для сети, которые представляют собой набор чисел с плавающей точкой.

Шаг 4. Связки слоев Conv2D + BatchNormalization и DepthwiseConv2D + BatchNormalization можно сократить, заменив одним Conv2D и одним DepthwiseConv2D слоями, пересчитав веса по формулам метода BatchNorm Fusion. На этом этапе MobileNet уменьшится

со 100 до 70 слоев. Сеть будет работать быстрее, а значения на выходе сети будут эквивалентны полной версии.

Замечание 1: после редукции слоев сеть уже не подходит для обучения, точнее обучение будет гораздо менее эффективно, чем на полной сети.

Замечание 2: в начальной версии сети у слоев Conv2D и DepthwiseConv2D не было смещения (bias), после редукции он у них появится.

Шаг 5. Значение 6 у функции активации RELU(6) можно заменить на 1, тогда значения карт признаков после каждой активации будут нормироваться на интервале от 0 до 1, так же как и у входного изображения. Сделать это не сложно. Для этого достаточно все веса только первого слоя разделить на 6, затем смещение у каждого слоя разделить на 6 и заменить все RELU(6) в сети на функцию RELU(1). Позиция максимума на последнем слое не изменится, т. е. классификация будет выполнена аналогично начальной модели.

Шаг 6. На данном этапе у нас получилось нормировать входное изображение и все значения промежуточных карт признаков на промежутке от 0 до 1, однако веса и смещения в общем случае могут превышать эти значения. Поэтому нам требуется найти максимум и минимум для них, чтобы определить, сколько значений старших разрядов нам потребуется хранить дополнительно. Для этого можно использовать формулы

$$\begin{aligned} \text{Conv}W &= \lceil \log_2 \max(\text{weights}) \rceil \text{Conv}B = \\ &= \lceil \log_2 \max(\text{bias}) \rceil, \end{aligned}$$

где *weights* — все значения весов на всех слоях; *bias* — все значения смещений на всех слоях; *ConvW* — число дополнительных битов, которые потребуются для хранения каждого веса; *ConvB* — число дополнительных битов, которые потребуются для хранения каждого смещения.

Если сеть обучена на большом наборе данных и имеет хорошую точность классификации, то *ConvW* и *ConvB* не должны быть очень большими по сравнению с размерностью карт признаков и входных изображений. На нашей тестовой сети было рассчитано *ConvW* = 7 и *ConvB* = 3.

Шаг 7. Далее для перехода к вычислениям с фиксированной точкой необходимо определить, сколько битов *N* достаточно хранить, чтобы вычисления не теряли точность, т. е. классификатор выдавал бы одинаковые ре-

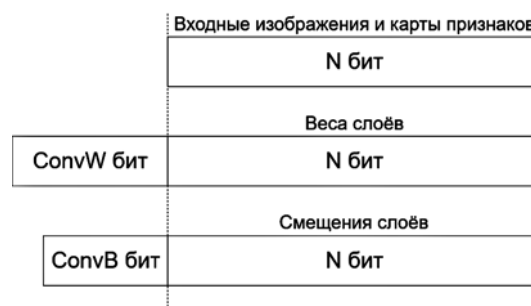


Рис. 4. Одинаковая размерность весов, смещений и карт признаков

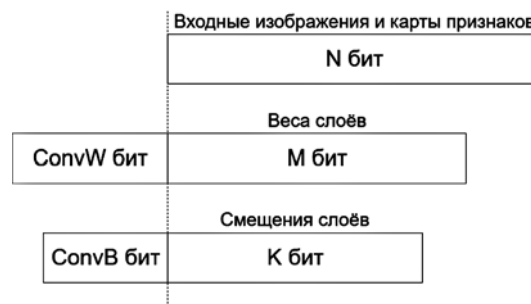


Рис. 5. Разная размерность весов, смещений и карт признаков

зультаты при вычислениях с плавающей и фиксированной точкой (рис. 4).

Можно немного усложнить задачу и найти оптимальные размерности отдельно для карт признаков, весов и смещений (рис. 5).

Для этого можно использовать алгоритм следующего типа. Устанавливаем переменную *N* в некоторое большое значение, на котором сеть заведомо будет работать точно. Пропускаем все валидационные изображения через сеть, и сравниваем результат классификации с математической моделью с плавающей точкой. Если точность выше требуемой, уменьшаем *N* до тех пор, пока не найдем *N*, где точность ниже заданной. Затем пробуем отдельно уменьшить размерность весов *M* и смещений *K*, пока не найдем оптимум. Эта операция известна как квантование с калибровкой (quantization with calibration) [25].

Для разных датасетов для тестовой сети MobileNet из этого исследования значения *N*, *M* и *K* могут немного различаться. Значения варьировались $N \in [10, 12]$, $M \in [9, 11]$, $K \in [8, 10]$ (без учета знака чисел).

Шаг 8. Поскольку планируется выполнять операцию округления после каждого сверточного блока, нам требуется найти максимальное значение промежуточных вычислений, которое может быть достигнуто для заданных параметров. Стратегия выполнения округления может быть разной. Например, можно вы-

полнять округление после каждой элементарной операции сверточного блока: 9 умножений и 9 сложений (включая смещение). Но так как в модулярной арифметике операция округления является вычислительно затратной, то лучше выполнять ее после расчета карты признаков для следующего слоя или даже один раз за несколько слоев. Поэтому во всей сети требуется найти максимальное число сложений и умножений, которое будет выполнено, прежде чем будет выполнена операция округления. Если округление выполняется после каждого отдельного сверточного слоя, то рассчитать размерность динамического диапазона D можно по следующей формуле:

$$D = \lceil \log_2 S \rceil + N + (ConvW + M),$$

где S — требуемое число сложений. Округление после каждого сверточного слоя нужно делать на значение 2^M бит, аналогичное сдвигу на M позиций влево в позиционной записи числа.

Шаг 9. В зависимости от вычисленного динамического диапазона выбирается модулярный базис.

Шаг 10. После выбора системы модулей для представления чисел требуется преобразовать все веса и смещения сначала в двоичный вид и запись с фиксированной точкой и затем в модулярный вид на базе выбранной системы модулей и сохранить в ОЗУ устройства.

Шаг 11. MobileNet после редукции имеет структуру, состоящую из последовательности блоков "свертка" + "активация", т. е. в модулярном виде потребуется набор операций модульного канала, состоящих только из сложений и умножений. Рекомендуется сначала выполнить операцию активации RELU (1), т. е. сравнение с 0 и 1, чтобы уменьшить размерность в операции округления. Операции сравнения и округления можно выполнять параллельно с модульными вычислениями, как только будут готовы результаты вычислений первых пикселей для карт признаков. Иными словами, для вычисления результатов для всей сети потребуется набор из трех последовательных блоков операций:

- массивный набор быстрых модульных вычислений на сверточном блоке, включающий только параллельные умножения и сложения малой размерности;
- операция сравнения с 0 и 1;
- операция сдвига на M позиций в двоичной записи числа.

5. Метод округления чисел с фиксированной точкой в модулярной арифметике

За основу алгоритма для округления чисел в аппаратной реализации нейронной сети были взяты два метода из публикаций [23] и [20]. В первой статье предлагается метод для реализации операции расширения модулярного базиса (base extension) на основе Lookup-таблиц (LUT). Во второй статье приводится метод для округления на произвольное число K , взаимно простое со всеми модулями выбранного базиса. Метод также основан на базе LUT. За счет использования таблиц и небольшого арифметического блока скорость операции округления значительно увеличивается.

Рассмотрим подробнее метод расширения модулярного базиса. Пусть задан модулярный базис из взаимно простых оснований $\{p_1, p_2, \dots, p_n\}$ и число $X = \{x_1, x_2, \dots, x_n\}$ в рамках этого базиса. Применив китайскую теорему об остатках, можно получить:

$$X = \left\lfloor \sum_{i=1}^n M_i \left\lfloor \frac{x_i}{M_i} \right\rfloor_{p_i} \right\rfloor_M = \sum_{i=1}^n M_i \left\lfloor \frac{x_i}{M_i} \right\rfloor_{p_i} - eM,$$

где $e \in [0; n)$ и $M_i = \frac{M}{p_i}$.

В работе [23] доказано, что для любого модуля p_{n+1} значение $|X|_{p_{n+1}} = x_{n+1}$ может быть рассчитано по следующей формуле:

$$x_{n+1} = |X_E - eM|_{p_{n+1}} = \left| M^{(p)} \left\lfloor \sum_{i=1}^n a_{i,p} \right\rfloor_{\frac{M}{M^{(p)}}} + \sum_{i=1}^p \frac{M^{(p)}}{p_i} \left\lfloor \frac{M}{M^{(p)}} \right\rfloor_{p_i} \left\lfloor \frac{x_i}{M_i} \right\rfloor_{p_i} - eM \right|_{p_{n+1}}. \quad (1)$$

В этой формуле: $M^{(p)} = \prod_{j=1}^p p_j$, $M_i = \frac{M}{p_i}$,

$a_{i,p} = \left\lfloor \frac{M_i \left\lfloor \frac{x_i}{M_i} \right\rfloor_{p_i}}{M^{(p)}} \right\rfloor_{\frac{M}{M^{(p)}}}$ и $e = \{0, 1\}$. Также в этой

формуле p может принимать любое значение на промежутке от 1 до n : $1 \leq p < n$. Значение e равно 1 только при выполнении следующих условий:

$$\begin{cases} \left\lfloor \sum_{i=1}^n a_{i,p} \right\rfloor_{\frac{M}{M^{(p)}}} \geq \frac{M}{M^{(p)}} - p + 1; \\ |X_E|_M \leq (p-1)M^{(p)} - M^{(p)} \sum_{i=1}^p \frac{1}{p_i}. \end{cases} \quad (2)$$

Значения $a_{i,p}$ могут быть предварительно рассчитаны для заданного базиса и затем получены из таблиц. Значения $a_{i,p}$ лежат на интервале: $0 \leq a_{i,p} \leq \frac{M}{M^{(p)}} - \frac{M_i}{M^{(p)}}$. То же самое

касается значений $b_i = \frac{M^{(p)}}{p_i} \left| \frac{M}{M^{(p)}} \left| \frac{x_i}{M_i} \right|_{p_i} \right|_{p_i}$, которые могут быть рассчитаны предварительно. Наибольшую сложность представляет вычислений условий, при которых $e = 1$.

Приняв $T = \left\lfloor \sum_{i=1}^n a_{i,p} \right\rfloor_{\frac{M}{M^{(p)}}}$, формулу (1) можно переписать в виде

$$\begin{aligned} x_{n+1} &= \left\lfloor M^{(p)}T + \sum_{i=1}^p b_i - eM \right\rfloor_{p_{n+1}} = \\ &= \left\lfloor M^{(p)}T \right\rfloor_{p_{n+1}} + \left\lfloor \sum_{i=1}^p b_i \right\rfloor_{p_{n+1}} - e \left\lfloor M \right\rfloor_{p_{n+1}}, \end{aligned} \quad (3)$$

а условие (2) — в виде

$$\begin{cases} T \geq \frac{M}{M^{(p)}} - p + 1; \\ \left\lfloor M^{(p)}T + \sum_{i=1}^p b_i \right\rfloor_M \leq (p-1)M^{(p)} - M^{(p)} \sum_{i=1}^p \frac{1}{p_i}. \end{cases} \quad (4)$$

Предложенный метод позволяет быстро выполнить операцию расширения базиса. Теперь рассмотрим метод для округления на базе расширения модулярного базиса, предложенный в работе [20], т. е. требуется найти такое значение Y , что $Y = \frac{X}{K}$. Данную формулу можно переписать следующим образом: $Y = \frac{X - |X|_K}{K}$ или в модулярном виде:

$$y_i = |Y|_{p_i} = \left| \frac{X - |X|_K}{K} \right|_{p_i} = \left| x_i - |X|_K \right|_{p_i} \left| K^{-1} \right|_{p_i}. \quad (5)$$

Если выбрать K взаимнопростым со всеми базисными модулями, то $|K^{-1}|_{p_i}$ всегда существует. Это хорошо подходит для нашего метода, поскольку мы можем выбрать значение $K = 2^L$, если все базисные модули — это простые числа, большие 2. Если K зафиксировано, то значения $|K^{-1}|_{p_i}$ могут быть рассчитаны заранее и сохранены в таблицах. Для вычисления потребуется только модулярное вычитание и

модулярное умножение на константу, эффективные реализации которых хорошо известны.

Заключение

В статье был предложен метод разработки аппаратной реализации нейронной сети на базе модулярной арифметики. Предложен маршрут реализации для известной мобильной нейронной сети MobileNet. Отмечены узкие места данной реализации, связанные с операциями сравнения и округления в модулярной арифметике. Предложен эффективный алгоритм округления для реализации в аппаратном модуле, а также функция активации, которая не содержит операцию сравнения.

Список литературы

1. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // arXiv preprint arXiv:1409.1556. 2014.
2. He K., Zhang X., Ren S., Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification // Proceedings of the IEEE international conference on computer vision. 2015. P. 1026–1034.
3. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Rabinovich A. Going deeper with convolutions. // Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1–9).
4. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. P. 770–778.
5. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation // International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015. P. 234–241.
6. Abadi M., Barham P., Chen J., Chen Z., Davis A., Dean J., Kudlur M. TensorFlow: A System for Large-Scale Machine Learning // OSDI. 2016. T. 16. C. 265–283.
7. Howard A. G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Adam H. Mobilenets: Efficient convolutional neural networks for mobile vision applications // arXiv preprint arXiv:1704.04861. 2017.
8. Iandola F. N., Han S., Moskewicz M. W., Ashraf K., Dally W. J., Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size // arXiv preprint arXiv:1602.07360. 2016.
9. Han S., Mao H., Dally W. J. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding // arXiv preprint arXiv:1510.00149. 2015.
10. Jouppi N. Google supercharges machine learning tasks with TPU custom chip // Google Blog, May. 2016. Vol. 18.
11. Nakahara H., Sasao T. A deep convolutional neural network based on nested residue number system // Field Programmable Logic and Applications (FPL), 2015 25th International Conference on. IEEE, 2015. P. 1–6.
12. Arrigoni V., Rossi B., Fragneto P., Desoli G. Approximate operations in Convolutional Neural Networks with RNS data representation // European Symposium on Artificial Neural Networks (ESANN 2017).
13. Червяков Н. И., Ляхов П. А., Калига Д. И., Валуева М. В. Архитектура сверточной нейронной сети с вычислениями

в системе остаточных классов с модулями специального вида // Нейрокомпьютеры: разработка, применение. 2017. № 1. С. 3—15.

14. Solovyev R., Kustov A., Telpukhov D., Rukhlov V., Kalinin A. Fixed-Point Convolutional Neural Network for Real-Time Video Processing in FPGA // 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EI-ConRus). 2019. P. 1605—1611.

15. Solovyev R. A., Kalinin A. A., Kustov A. G., Telpukhov D. V., Rukhlov V. S. FPGA implementation of convolutional neural networks with fixed-point calculations. 2018. arXiv preprint arXiv:1808.09945.

16. Cong J., Xiao B. Minimizing computation in convolutional neural networks // International conference on artificial neural networks. Springer, Cham, 2014. P. 281—290.

17. Brönnimann H., Emiris I., Pan V. Y., Pion S. Sign determination in residue number systems // Theoretical Computer Science. 1999. Vol. 210. P. 173—197.

18. Van Vu T. Efficient implementations of the Chinese remainder theorem for sign detection and residue decoding // IEEE Transactions on Computers. 1985. Vol. 100, N. 7. P. 646—651.

19. Hitz M. A., Kaltofen E. Integer division in residue number systems // IEEE transactions on computers. 1995. Vol. 44, N. 8. P. 983—989.

20. Kong Y., Phillips B. Fast scaling in the residue number system // IEEE transactions on very large scale integration (VLSI) systems. 2009. Vol. 17, N. 3. P. 443—447.

21. Low J. Y. S., Chang C. H. A VLSI Efficient Programmable Power-of-Two Scaler for $\{2^n-1, 2^n, 2^n+1\}$ RNS //

IEEE Transactions on Circuits and Systems I: Regular Papers. 2012. Vol. 59, N. 12. P. 2911—2919.

22. Smyk R., Czyżak M., Ulman Z. Scaling of signed residue numbers with mixed-radix conversion in FPGA with extended scaling factor selection // Computer Applications in Electrical Engineering. 2013. Vol. 11.

23. Barsi F., Pinotti M. C. Fast base extension and precise scaling in RNS for look-up table implementations // IEEE transactions on signal processing. 1995. Vol. 43, N. 10. P. 2427—2430.

24. Kuznetsova A., Rom H., Alldrin N., Uijlings J., Krasin I., Pont-Tuset J., Ferrari V. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. 2018. arXiv preprint arXiv:1811.00982.

25. Goncharenko A., Denisov A., Alyamkin S., Terentev E. Fast adjustable threshold for uniform neural network quantization. 2018. arXiv preprint arXiv:1812.07872.

26. Стемниковский А. Л., Амербаев В. М., Соловьев Р. А. Принципы рекурсивных модулярных вычислений // Информационные технологии. 2013. № 2. С. 22—27.

27. Амербаев В. М., Соловьев Р. А., Тельпухов Д. В. Реализация библиотеки модульных арифметических операций на основе алгоритмов минимизации логических функций // Известия Южного федерального университета. Технические науки. 2013. Т. 7, № 144.

28. Амербаев В. М., Соловьев Р. А., Тельпухов Д. В., Балака Е. С. Построение обратных преобразователей модулярной арифметики с коррекцией ошибок на базе полиадического кода // Нейрокомпьютеры: разработка, применение. 2014. № 9. С. 30—35.

R. A. Solovyev, Ph.D., Chief Researcher, e-mail: turbo@ippm.ru,
D. V. Telpukhov, Ph.D., Head of the Department, A. G. Kustov, Engineer,
T. U. Isaeva, Ph.D., Researcher, A. A. Volkov, Engineer,
Institute for Design Problems in Microelectronics

Application of Modular Arithmetic Methods in the Development of Hardware Implementations of Neural Networks

The article proposes to use the methods of modular arithmetic for the implementation of neural networks at the hardware level in VLSI and FPGAs. Widely known mobile neural networks that are highly accurate and suitable for implementation in hardware are considered. The main difficulties in their implementation in the basis of modular arithmetic, namely the rounding operation and the comparison operation, are examined and analyzed. A number of methods have been proposed to solve these problems, such as using convolutions with stride greater than 1 instead of MaxPooling layers, using non-standard activations containing only addition, subtraction and multiplication operations, as well as an efficient algorithm for implementing rounding operations. Additionally, a complete route for designing and transferring a MobileNet neural network to the hardware level in a modular basis is proposed.

Keywords: neural networks, modular arithmetic, residue number system, rounding operation, hardware implementation

DOI: 10.17587/it.25.747-756

References

1. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014.

2. He K., Zhang X., Ren S., Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026—1034.

3. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Rabinovich A. Going deeper with convolutions, *Cvpr*, 2015.

4. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770—778.

5. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation, *International Conference on Medical image computing and computer-assisted intervention*, Springer, Cham, 2015, pp. 234—241.

6. Abadi M., Barham P., Chen J., Chen Z., Davis A., Dean J., Kudlur M. TensorFlow: A System for Large-Scale Machine Learning, *OSDI*, 2016, vol. 16, pp. 265—283.

7. Howard A. G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Adam H. Mobilenets: Efficient convolutional neural

networks for mobile vision applications, arXiv preprint arXiv:1704.04861, 2017.

8. Iandola F. N., Han, S., Moskewicz M. W., Ashraf K., Dally W. J., Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size, arXiv preprint arXiv:1602.07360, 2016.

9. Han S., Mao H., Dally W. J. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding, arXiv preprint arXiv:1510.00149, 2015.

10. Jouppi N. Google supercharges machine learning tasks with TPU custom chip, *Google Blog*, May, 2016, vol. 18.

11. Nakahara H., Sasao T. A deep convolutional neural network based on nested residue number system, *Field Programmable Logic and Applications (FPL)*, 2015 *25th International Conference on*, IEEE, 2015, pp. 1–6.

12. Arrigoni V., Rossi B., Fragneto P., Desoli G. Approximate operations in Convolutional Neural Networks with RNS data representation, *European Symposium on Artificial Neural Networks (ESANN 2017)*.

13. Chervyakov N. I., Lyakhov P. A., Kalita D. I., Valueva M. V. Convolutional neural network architecture using computations in residue number system with specific module set, *Neurocomputery: Razrabotka, Primenenie*, 2017, no. 1, pp. 3–15 (in Russian).

14. Solovyev R., Kustov A., Telpukhov D., Rukhlov V., Kalinin A. Fixed-Point Convolutional Neural Network for Real-Time Video Processing in FPGA. In 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), IEEE, 2019, January, pp. 1605–1611.

15. Solovyev R. A., Kalinin A. A., Kustov A. G., Telpukhov D. V., Rukhlov V. S. FPGA implementation of convolutional neural networks with fixed-point calculations, 2018, arXiv preprint arXiv:1808.09945.

16. Cong J., Xiao B. Minimizing computation in convolutional neural networks, *International conference on artificial neural networks*, Springer, Cham, 2014, pp. 281–290.

17. Brönnimann H., Emiris I., Pan V. Y., Pion S. Sign determination in residue number systems, *Theoretical Computer Science*, 1999, vol.210, pp. 173–197.

18. Van Vu T. Efficient implementations of the Chinese remainder theorem for sign detection and residue decoding, *IEEE Transactions on Computers*, 1985, vol. 100, no. 7, pp. 646–651.

19. Hitz M. A., Kaltofen E. Integer division in residue number systems, *IEEE transactions on computers*, 1995, vol. 44, no. 8, pp. 983–989.

20. Kong Y., Phillips B. Fast scaling in the residue number system, *IEEE transactions on very large scale integration (VLSI) systems*, 2009, vol. 17, no. 3, pp. 443–447.

21. Low J. Y. S., Chang C. H. A VLSI Efficient Programmable Power-of-Two Scaler for $\{2^n-1, 2^n, 2^n+1\}$ RNS, *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2012, vol. 59, no. 12, pp. 2911–2919.

22. Smyk R., Czyżak M., Ulman Z. Scaling of signed residue numbers with mixed-radix conversion in FPGA with extended scaling factor selection, *Computer Applications in Electrical Engineering*, 2013, vol. 11.

23. Barsi F., Pinotti M. C. Fast base extension and precise scaling in RNS for look-up table implementations, *IEEE transactions on signal processing*, 1995, vol. 43, no. 10, pp. 2427–2430.

24. Kuznetsova A., Rom H., Alldrin N., Uijlings J., Krasin I., Pont-Tuset J., Ferrari V. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale, 2018, arXiv preprint arXiv:1811.00982.

25. Goncharenko A., Denisov A., Alyamkin S., Terentev E. Fast adjustable threshold for uniform neural network quantization, 2018, arXiv preprint arXiv:1812.07872.

26. Stempkovsky A. L., Amerbaev V. M., Solovyev R. A. Principles of Recursive Modular Arithmetic, *Informacionnye tekhnologii*, 2013, no. 2, pp. 22–27 (in Russian).

27. Amerbaev V. M., Solovyev R. A., Telpukhov D. V. Library implementation of modular arithmetic operations based on logic functions minimization algorithms, *Izvestiya Yuzhnogo Federalnogo Universiteta. Technicheskie Nauki*, 2013, vol. 7, no. 144 (in Russian).

28. Amerbaev V. M., Solovyev R. A., Telpukhov D. V., Balaka E. S. Construction of residue number system reverse converters with error correction, based on mixed-number system, *Neurocomputery: Razrabotka, Primenenie*, 2014, no. 9, pp. 30–35 (in Russian).

Уважаемые подписчики!

С 2020 г. можно оформить подписку на электронные версии журналов, выпускаемые ООО "Издательство "Новые технологии", через подписное агентство "Урал-Пресс" и ООО «ИВИС». Подписные индексы наших журналов в агентстве "Урал-Пресс":

Индекс	Издание
013308	Безопасность жизнедеятельности Электронная версия (Pdf). Рассылка на e-mail
013309	Информационные технологии Электронная версия (Pdf). Рассылка на e-mail
013310	Мехатроника, автоматизация, управление Электронная версия (Pdf). Рассылка на e-mail
013312	Программная инженерия Электронная версия (Pdf). Рассылка на e-mail
013311	Нано-микросистемная техника Электронная версия (Pdf). Рассылка на e-mail

Контакты ООО "ИВИС": тел. (495)777-65-57, 777-65-58; e-mail: sales@ivis.ru.

В. Н. Костин, канд. тех. наук, доц., e-mail:vladimirkostin5@mail.ru,
Федеральное государственное бюджетное образовательное учреждение высшего образования
"Оренбургский государственный университет"

Оценка эффективности физической защиты информации критически важных объектов на основе марковских цепей

Разработан метод моделирования оценки эффективности физической защиты информации критически важных объектов. Особенностью решения задачи является декомпозиция сложного мультиграфа проникновения нарушителя к информационным ресурсам объекта на множество простых графов, каждый из которых представлен в виде последовательности событий. Иными словами, осуществляется декомпозиция сложной задачи на множество простых подзадач. Результат декомпозиции представлялся в виде матрицы инцидентности: строки — номер пути проникновения, столбцы — ребра графа проникновения нарушителем. Каждый путь нарушителя и реакция физической защиты на проникновение моделируются двумя информационно связанными марковскими цепями. Показателем эффективности физической защиты выбрана вероятность защиты информации. Система оценки защищенности состоит из дизъюнктивно связанных путей проникновения, при этом каждый путь — это набор конъюнктивно связанных ребер проникновения, который определяется вероятностью реализации цели нарушителя. Общая оценка физической защиты информации проводится по самому пессимистическому пути проникновения нарушителя.

Ключевые слова: оценка эффективности физической защиты, мультиграф проникновения нарушителя, матрица инцидентности

Введение

В условиях возрастающих угроз (развития терроризма) и повышения возможностей современных технологий задача оценки эффективности физической защиты (ФЗ) информации становится все более актуальной. Это связано, в первую очередь, с возрастающей сложностью систем защиты. Кроме того, среди последовательных этапов проектирования ФЗ завершающим и важным этапом разработки является оценка ее эффективности.

Вопросам оценки эффективности ФЗ посвящено много статей, в которых рассматриваются несколько подходов к оценке эффективности. Например, А. В. Леус в статье [1] предлагает показатели оценки эффективности ФЗ и математическую модель на основе трехмерного куба. Достоинство метода: при оценке вероятности безопасного состояния используется логарифмическая функция, позволяющая перейти к сложению показателей защищенности объекта. Это удобно при анализе эффектив-

ности ФЗ. Недостатком метода является большая многомерность задачи из-за бесконечного множества точек пространства движения нарушителя. С. И. Корчагин в статье [2] предлагает методику оценки ФЗ с помощью вероятностного подхода, достоинством которой является тот факт, что в модели осуществляется переход к одномерной структуре комплекса, за счет чего упрощаются расчеты оценки эффективности ФЗ. Однако в данной методике нет обоснования значения порога вероятности для перехода к одномерной структуре комплекса. С. С. Звежинский и др. в статье [3] рассматривают эффективность охранной сигнализации для малых объектов, при этом учитывают важность критических элементов путем введения коэффициента важности в целевую функцию оценки рационального расположения технических средств защиты (ТСЗ) на рубежах охраны. Недостатком является решение задачи без ограничений. Кроме того, многие методы оценки эффективности ФЗ основаны исключительно на экспертных оценках, что привно-

сит элемент субъективизма в оценку, а также зависимость от знаний и опыта экспертов.

Оценка эффективности ФЗ с использованием логико-вероятностных методов (ЛВМ) рассматривается в статьях О. А. Панина [4, 5], где делается акцент на большую трудоемкость ЛВМ. Полученные результаты комплексной оценки ФЗ по данной методике [5, 6] при большом числе вариантов проникновения занижают показатель защищенности системы защиты.

Проведенный в монографии [7] анализ показателей и методов оценки эффективности ФЗ показал, что не существует универсальных методик оценки эффективности ФЗ. На сегодняшний день применяется несколько методик оценки эффективности ФЗ: детерминистический подход; логико-вероятностный метод; вероятностно-временной метод. При детерминистическом подходе за эффективность принимается степень выполнения требований по физической защите в соответствии с нормативными документами. Предполагается проведение комплекса проверок требований, содержащихся в руководящих документах. Для каждого фактора состояния разрабатывается аналитическая модель для оценки организационных мероприятий. В логико-вероятностном методе под эффективностью понимается вероятность нахождения системы в безопасном состоянии согласно построенному сценарию развития опасной ситуации. Составленная функция опасности системы в виде логического многочлена заменяется на вероятностный многочлен для определения показателя системы — вероятности наступления опасного события. При вероятностно-временном методе под эффективностью понимается вероятность того, что у сил реагирования резерв времени окажется больше нуля, т. е. ФЗ успешно выполнила функцию. Для этого проводится оценка всех маршрутов движения нарушителя и реагирования группы нейтрализации.

Материал данной статьи является продолжением исследований вопросов оценки эффективности ФЗ на основе методов системного анализа.

Постановка задачи

Для оценки функционирования ФЗ необходимо определить показатель эффективности. По мнению авторов работы [7], наиболее полно характеризующим показателем эффективности функционирования ФЗ является вероятность нахождения информации критически

важного объекта (КВО) в безопасном состоянии. Данный показатель определяется двумя связанными показателями: вероятностью обнаружения нарушителя и вероятностью своевременного прибытия сил охраны для нейтрализации нарушителя:

$$P(V) = P(A)P(R/A)P(V/(R/A)), \quad (1)$$

где $P(A)$ — вероятность получения силами охраны сигнала тревоги; $P(R/A)$ — вероятность развертывания сил охраны в точке перехвата при условии уверенного приема сигнала тревоги; $P(V/(R/A))$ — вероятность нейтрализации нарушителя при условии своевременного развертывания сил охраны. Данный показатель не исследуется — принимается за единицу.

Целью настоящих исследований является разработка метода оценки эффективности ФЗ (оценки варианта размещения ТСЗ) на основе синтеза марковских цепей. Для решения этой задачи необходимо разработать метод формирования логических функций проникновения, на основе которых осуществляется оценка эффективности ФЗ.

Разработанный метод позволяет получить совокупность всех путей проникновения нарушителя к информационным ресурсам объекта, представленных как логические функции проникновения в виде дизъюнкции и конъюнкции логических переменных — ребер графа, которые сформированы в информационную матрицу инцидентности для решения задачи оценки эффективности ФЗ.

Достоинства метода: 1) сложная задача оценки вероятности безопасного состояния информации представлена как оценка надежности системы в виде мультиграфа, который декомпозируется на множество простых графов проникновения нарушителя; 2) входные данные в модель оценки вероятности проникновения нарушителя формируются с помощью проведения натурного эксперимента на объекте защиты, что повышает их точность и, как следствие, достоверность оценки показателя эффективности ФЗ.

Решение задачи рассмотрим на модельном примере. КВО представляет собой сложную систему, состоящую из множества связанных зон охраны различной природы назначения, важности и уровня защищенности (рис. 1). Вся территория объекта имеет двойное ограждение, периметровую охрану и контрольно-пропускной пункт (КПП). На КВО имеется элемент информатизации (ЭИ), подлежащий охране. На рубежах зон охраны располагаются

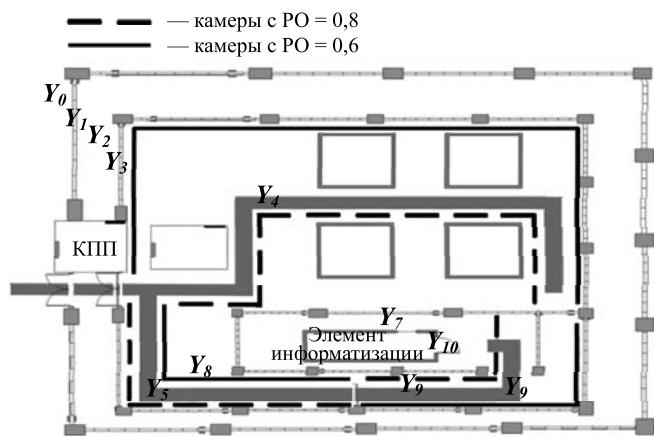


Рис. 1. Расположение ТСЗ (камер видеонаблюдения) на объекте

ТСЗ. Цель нарушителя — проникнуть в ЭИ и совершить хищение информации. Модель проникновения нарушителя представлена в виде разветвленного ориентированного мультиграфа (рис. 2). Мультиграф — это сценарий проникновения нарушителя на охраняемый объект. Вершинами графа являются рубежи зон охраны при достижении нарушителем определенного результата на пути к цели. Ребра графа — это варианты перемещений нарушителя между рубежами охраны. Ребра обозначены X_i , где i — номер ребра (варианта перемещения) в графе. Всего определим n рубежей зон охраны. Следовательно, граф имеет n вершин (событий). Первая вершина — Y_0 , последняя Y_n . Событию Y_0 присваивается единица, т. е. это вероятность нахождения нарушителя в нулевом событии в начальный момент времени. Наступление события Y_n означает факт проникновения нарушителя к информационным ресурсам (нарушитель достиг цели (произошло хищение и т. д.)). Вероятность нахождения события Y_n в безопасном состоянии и будет показателем эффективности ФЗ.

На рис. 2 обозначено: X_1 — преодоление первого ограждения через верх; X_2 — преодоление ограждения путем разрушения ограждения; X_3 — подкоп первого ограждения; X_4 — вариант перемещения зоны бегом; X_5 — преодоление зоны ползком; X_6 — преодоление второго ограждения через верх; X_7 — разрушение второго ограждения; X_8 — подкоп второго ограждения; X_9 — проход

через КПП путем подбора ПИН-кода; X_{10} — проникновение через ворота путем подмены документов; $X_{11}...X_{21}$ — варианты перемещения через зоны между рубежами Y_{3-10} внутри объекта (см. рис. 1).

Для отображения противодействия ФЗ введем еще один ориентированный граф (рис. 3). Вершины графа обозначают рубежи противодействия системы защиты от нарушителя. Ребро в графе ассоциируется с каким-то типом варианта защиты объекта, который будет характеризоваться вероятностью защиты, т. е. ребро — вероятность того, что ФЗ обнаружит нарушителя и окажет противодействие при переходе между рубежами. Ребра также будем обозначать Z_i , где i — номер ребра (вариант противодействия) в графе.

Полученный граф назовем графом противодействия. Всего будем определять n рубежей, как и в предыдущем графе, т. е. граф имеет n вершин. Первая вершина — Y_n , последняя — Y_0 . Событию Y_n присваивается значение 1. Ребра будут направлены от вершины Y_n к Y_0 .

Марковская модель позволяет оценить вероятности состояний двух противоборствующих

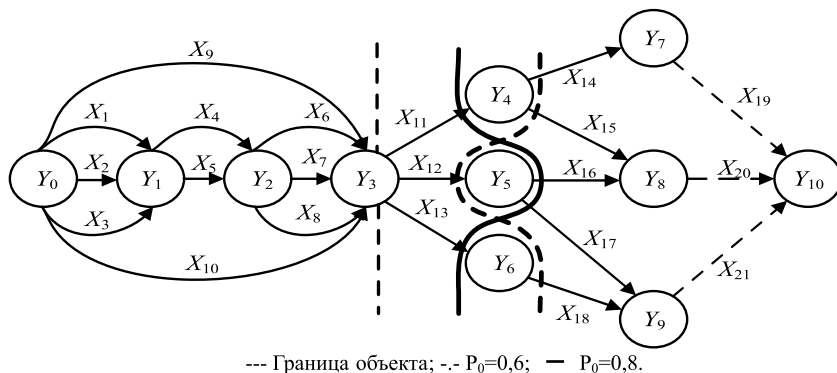


Рис. 2. Граф достижимости нарушителем цели

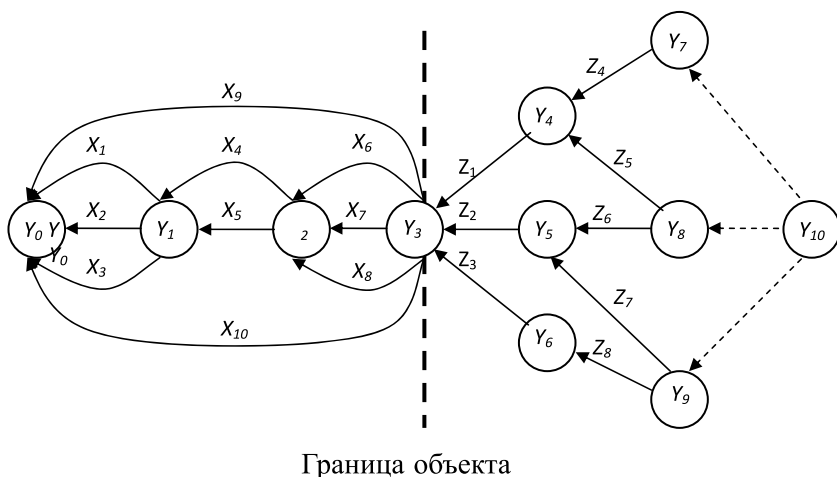


Рис. 3. Граф противодействия ФЗ проникновению нарушителя

Таким образом, имеем марковскую модель с дискретными состояниями, представленную в виде ориентированного взвешенного графа. Переход системы из состояния в состояние будем рассматривать в дискретные моменты времени Δt . Система считается заданной, если заданы два условия: 1 — вероятность начальных состояний системы $P_i(t)$ вектором $P_i^{(0)} = (P_{01}, P_{02}, \dots, P_{0n})$; 2 — условные вероятности переходов $P_{ik}(\Delta t)$ из i -го в k -е состояние за время Δt :

$$P_{ik}(\Delta t) = \begin{vmatrix} P_{11}(\Delta t) & P_{12}(\Delta t) & \dots & P_{1n}(\Delta t) \\ \dots & \dots & \dots & \dots \\ P_{n1}(\Delta t) & P_{n2}(\Delta t) & \dots & P_{nn}(\Delta t) \end{vmatrix}. \quad (2)$$

$$P_k(t + \Delta t) = P_1(t)P_{1k}(\Delta t) + P_2(t)P_{2k}(\Delta t) + \dots \\ \dots + P_k(t)P_{kk}(\Delta t) + \dots + P_n(t)P_{nk}(\Delta t), \quad (3)$$

$$P(t + \Delta t) = P(t)P_{ij}(t + \Delta t), \quad (4)$$

Для однородной цепи Маркова вероятности переходов из i -го в j -е состояние во времени можно записать в виде:

$$P_{ij}(t, t + m\Delta t) = P_{ij}^m(t, t + \Delta t). \quad (5)$$

$$\lim_{m \rightarrow \infty} P_i(t + m\Delta t) = P_i^*, \quad (6)$$

К моменту времени $t + m\Delta t$ вероятности состояний системы двух шагов $m - 1$, m практически равны между собой. Тогда из уравнения Маркова получим вероятность k -го события:

$$P_k = \sum_{i=1}^n P_i P_{ik}. \quad (7)$$

$$\sum_{i=1}^n P_i = 1. \quad (8)$$

Ребра, показанные сплошными линиями, обозначают вероятности переходов нарушителя между рубежами за время Δt , а ребра, показанные штриховыми линиями, — вероятности противодействия ФЗ проникновению. Вектор

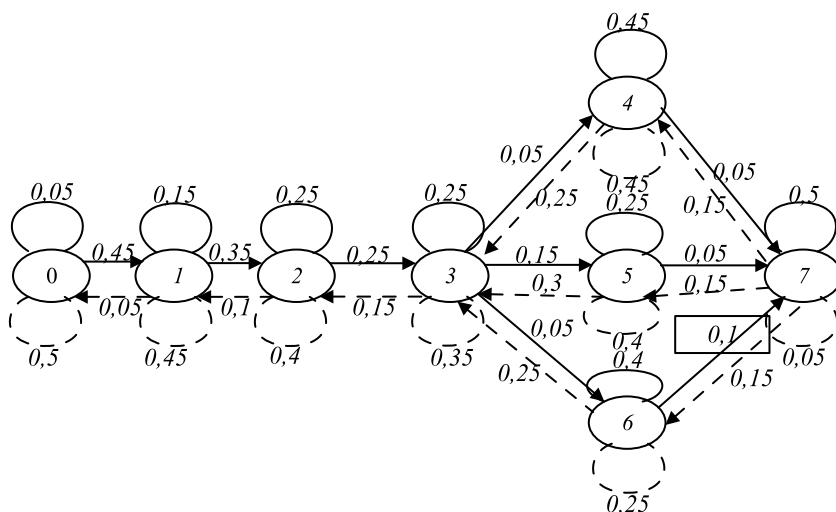


Рис. 4. Граф проникновения нарушителя и противодействия ФЗ

Матрица инцидентности

№ функции проникновения	Номера ребер графа																				
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}	X_{21}
1									1		1			1					1		
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
100			1		1			1					1					1			1

исходного состояния нарушителя представлен в виде вектора-строки. Умножая вектор вероятности состояний на матрицу смежности вероятностей, получаем установившийся результат на определенной итерации. Вероятность конечного n -го события и будет вероятностью безопасного состояния информации. С использованием данных теоретических предположений было разработано программное средство на языке программирования C#, реализующее оценку эффективности ФЗ.

Решение задачи оценки эффективности с помощью марковской модели в такой постановке составило непреодолимую трудность по следующим причинам: большая размерность задачи, сложность проведения нормировки графа при взаимодействии противодействующих сторон и трудности принятия решений, направленных на повышение эффективности ФЗ. По этой причине перешли к синтезу марковских цепей оценки эффективности ФЗ.

Для этого проводили декомпозицию мультиграфа проникновения нарушителя. Определяли все пути перемещений из начальной вершины Y_0 в конечную Y_{10} . Все варианты пути из одной смежной вершины в другую обозначили как дизъюнкцию логических переменных. Например, перемещение из вершины Y_0 в вершину Y_1 будем обозначать $X_1 \vee X_2 \vee X_3$.

Пути из одной вершины в другую определяли с помощью операции композиции матрицы смежности мультиграфа. Чтобы найти пути, состоящие из k ребер, необходимо возвести матрицу смежности в степень k . При этом получим новую матрицу, в которой будут представлены все пути между событиями длиной от одного ребра до k ребер. Таким образом, в полученной логической функции проникновения параллельные маршруты будут представлены как дизъюнкции ребер, а последовательные — как их конъюнкции. Также следует учесть, что умножаемые матрицы смежности содержат логические переменные. Из этого следует, что к результатам умножения ячеек можно применить операции алгебры логики для сокращения

результата умножения [6]: 1) правила для одной переменной $A \vee 1 = 1$; $A \vee 0 = A$; 2) закон тавтологии $A \vee A \vee \dots A = A$; — $A \wedge A \wedge \dots A = A$; 3) распределительный закон — $A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$.

Используя данные теоретические предположения, определим все возможные пути из первого рубежа в последний, представленные как функции алгебры логики в виде конъюнкции весов ребер, по которым проходит путь. Для удобства операцию конъюнкции переменных $X_1 \wedge X_2$ будем обозначать $X_1 X_2$, а операцию дизъюнкции $X_1 \vee X_2$ будем обозначать $X_1 + X_2$. Всего получили 100 логических функций проникновения нарушителя, из которых десять имеют длину в четыре ребра, а девяносто — длиной в шесть ребер. Полученные функции алгебры логики сведены в матрицу инцидентности (табл. 1): строки в матрице — маршруты проникновения (функции), а столбцы — ребра графа между рубежами объекта, на которых расположены ТСЗ с характеристиками (табл. 2).

Элементы матрицы инцидентности в строке связаны конъюнктивно, а сами строки — дизъюнктивно. Полученные логические функции проникновения позволяют оценить вероятность проникновения и противодействия на каждом маршруте, т. е. оценить эффективность ФЗ. В вероятностном смысле эффективность ФЗ будет определяться вероятностью нереализации ни одной из ста функций проникновения.

С точки зрения системного анализа процесс получения всех путей проникновения (функций проникновения) — это декомпозиция сложной задачи на более простые подзадачи.

Таблица 2

Характеристики технических средств защиты

Тип ТСЗ	Расположение на графе проникновения	P об-наружения	Протяженность, м	Удаление от ЭИ, м
CNB-WFL-2IS	$X_{11}, X_{16}, X_{17}, X_{13}$	0,60	880	450
SCANALL	$X_{12}, X_{14}, X_{15}, X_{18}$	0,80	670	300

После этого согласно теории системного анализа решается задача синтеза и оценки эффективности ФЗ.

Вероятности обнаружения и перехвата (своевременного прибытия для нейтрализации) нарушителя обоснованы и заданы на предыдущих этапах проектирования ФЗ [9]. Определим требования к ФЗ: вероятность обнаружения нарушителя на каждом пути проникновения не менее 0,9; вероятность своевременного прибытия в точку пресечения группы реагирования на каждом пути не менее 0,8, т. е. вероятность безопасного состояния объекта P_{6c} не меньше 0,72.

На основе длин ребер на местности определим протяженность пути проникновения как их сумма в соответствии с табл. 1: $X_1 - 25$ м; $X_2 - 15$ м; $X_3 - 35$ м; $X_4 - 14$ м; $X_5 - 24$ м; $X_6 - 20$ м; $X_7 - 15$ м; $X_8 - 24$ м; X_9 и $X_{10} - 13$ м; X_{11} — перемещение между рубежами $Y_3 - Y_4 - 450$ м; X_{12} — перемещение между рубежами $Y_3 - Y_5 - 320$ м; X_{13} — между рубежами $Y_3 - Y_6 - 220$ м; X_{14} — между рубежами $Y_4 - Y_7 - 180$ м; X_{15} — между рубежами $Y_4 - Y_8 - 50$ м; X_{16} — между рубежами $Y_5 - Y_8 - 160$ м; X_{17} — между рубежами $Y_5 - Y_9 - 50$ м; X_{18} — между рубежами $Y_6 - Y_9 - 120$ м; X_{19} — между рубежами $Y_7 - Y_{10} - 60$ м; X_{20} — между рубежами $Y_8 - Y_{10} - 70$ м; X_{21} — между рубежами $Y_9 - Y_{10} - 50$ м. Расстояние от караула до ЭИ — 310 м.

Для решения задачи оценки вероятности проникновения нарушителя по порядку моделировали все пути проникновения из матрицы инцидентности (см. табл. 1) в виде последовательных событий перемещения нарушителя между рубежами охраны и противодействия проникновению сил охраны. Данная последовательность описывалась двумя простыми связанными моделями: первая модель — модель перемещения нарушителя, вторая — модель перемещения группы нейтрализации нарушителя после его обнаружения. Обе модели опи-

сываются с помощью последовательного графа переходов событий. Модели взаимосвязаны и работают синхронно. Процесс движения $x(t)$ считаем как поток независимых приращений расстояний с интенсивностью (скоростью) $\lambda(t)$. Так как приращения на любом участке времени $t, t + \Delta t$ независимы, то процесс приращения расстояния при движении $x(t)$ является пуассоновским процессом с ограниченным числом состояний $x(t) \leq L$, где L — расстояние между событиями [9]. Вероятности переходов между событиями за время Δt определяются по экспоненциальному закону исходя из значения скорости перемещения между рубежами охраны. Графы проникновения и противодействия представлены на рис. 5.

Вершины графов — рубежи перемещений через зоны охраны нарушителя и группы нейтрализации. Особенностью второго графа переходов является то, что он информационно связан с первым графом (на рис. 5 это показано штриховыми ребрами). Матрица вероятностей переходов изменяется динамически, в зависимости от вероятности состояния событий (1) и (2) модели движения нарушителя. Таким образом, марковская модель реакции группы нейтрализации на проникновение нарушителя не стационарна. По мере перемещения нарушителя вероятность состояний (1) и (2) увеличивается и, следовательно, увеличивается вероятность обнаружения. Из состояний (1) и (2), где находятся средства технического контроля, информация передается в ребро 0 — 1 на второй марковский граф в виде вероятности обнаружения, т. е. содержание матрицы переходов вероятностей динамически изменяется. Эта информация в виде возрастающей вероятности передается во второй граф (матрицу переходов), и, следовательно, вероятность начала перемещения группы нейтрализации увеличивается. Во второй модели графа группы нейтрализации переход из нулевого состояния в первое состояние — это и есть вероятность обнаружения нарушителя ТСЗ в виде логической связи. Таким образом, движение группы нейтрализации начинается с первого события при получении информации (обнаружении нарушителя) с двух рубежей расположения технических средств обнаружения. Последние вершины обоих графов имеют одно и то же физическое положение на местности — ЭИ.

Вероятность перехода из нулевого события в первое (вероятность обнаружения) определяется по формуле умножения вероятностей последовательных событий, так как наруши-

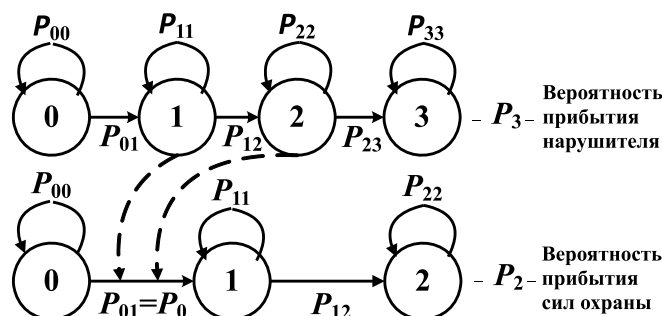


Рис. 5. Синхронизация графов проникновения и противодействия ФЗ

тель последовательно пересекает рубежи, на которых расположены ТСЗ. Результирующее действие оценивает вероятность перехода на графе из нулевого события в первое, как произведение вероятностей не обнаружения на каждом рубеже:

$$P_{01} = 1 - [(1 - P_{ТСЗ}^1 P_1)(1 - P_{ТСЗ}^2 P_2)], \quad (9)$$

где $P_{ТСЗ}^1$, $P_{ТСЗ}^2$ — вероятности обнаружения ТСЗ, расположенных на графе в вершинах событий 1 и 2 соответственно;

P_1 , P_2 — вероятности нахождения нарушителя в зоне обнаружения ТСЗ на рубежах событий вершин графа проникновения 1 и 2 соответственно.

Функционирование графов синхронизировано по времени с шагом продолжительности $\Delta t = 1$ с. В результате моделирования определяли значения вероятности обнаружения, которые передавали во вторую модель противодействия. Во второй модели на графе с помощью марковской модели графа определяется вероятность своевременного прибытия группы нейтрализации, т. е. вероятность наступления события (2).

Выходными значениями марковских моделей являются вероятности наступления конечных событий, т. е. реализация целей нарушителя и группы нейтрализации. Полученные графики изменения вероятностей конечных событий сравниваются, и проводится анализ выходной информации. По соотношению вероятностей конечных событий можно сделать вывод о безопасном состоянии объекта [10].

Процесс моделирования автоматизирован с помощью программы на языке программирования С#. Исходные данные представлены в виде характеристик: расстояния и скорости движения противоборствующих сторон в соответствующих зонах охраны. Результаты решения и динамика изменения вероятности состояний конечных событий представлены на рис. 6, 7.

Достоверность результатов определяется корректностью задаваемых входных матриц вероятности переходов. Данные вероятности определяются расчетным путем и согласовываются с экспериментальными данными [12] как элемент вероятности: $\Delta P_{ij} = \lambda_{ij} \Delta t$, где λ_{ij} — интенсивность (скорость) движения нарушителя между рубежами и сил охраны к месту расположения ЭИ.

Оценка вероятности нахождения объекта в безопасном состоянии по вероятностям двух конечных событий определяется по фор-

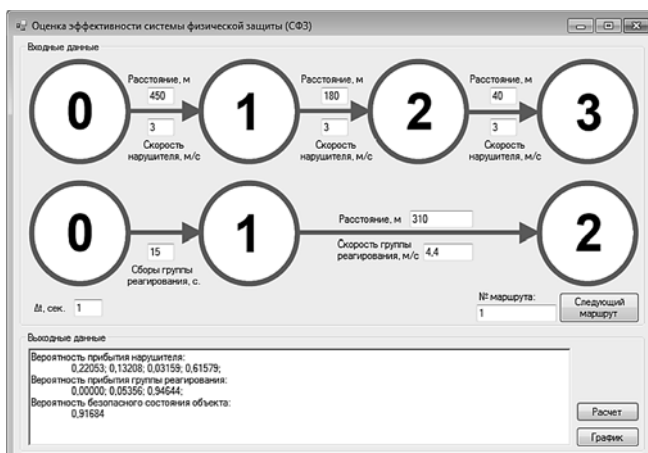


Рис. 6. Входные данные и результаты вычислений

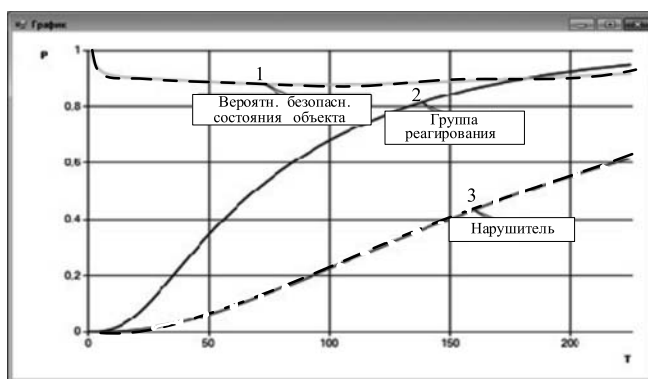


Рис. 7. Динамика изменения вероятности конечных событий

муле гипотез. Оценивается первая гипотеза, состоящая в том, что группа нейтрализации опередит нарушителя, и вторая гипотеза, что нарушитель опередит группу нейтрализации. Условная вероятность появления первой гипотезы определялась по формуле [12, стр. 64]

$$P(H_1/A) = PH_1 / (PH_1 + PH_2), \quad (10)$$

где $PH_1 = P_1(1 - P_2)$ — вероятность первой гипотезы; $PH_2 = (1 - P_1)P_2$ — вероятность второй гипотезы; P_1 , P_2 — вероятности прибытия к месту разворачивания группы нейтрализации и нарушителя соответственно.

Данная задача решается для каждого из ста маршрутов проникновения матрицы инцидентности (см. табл. 1). В результате анализа всех путей проникновения с учетом расположения ТСЗ получили матрицу инцидентности, содержащую всего пять уникальных функций проникновения до границы объекта (табл. 3).

Результаты вычислений вероятностей безопасного состояния сведены в табл. 4.

Анализ табл. 4 показывает, что требования к эффективности ФЗ выполнены [9]. Про-

Инцидентности уникальных маршрутов в границах объекта

№ функции проникновения	Номер дуги графа										
	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}	X_{21}
1	1			1					1		
2	1				1					1	
3		1				1				1	
4		1					1				1
5			1					1			1

Таблица 4

Результаты моделирования марковских процессов

№ функции проникновения	Вероятность прибытия сил охраны PH_1	Вероятность прибытия нарушителя PH_2	Вероятность безопасного состояния объекта $P(H_1/A)$
1	0,951	0,615	0,924
2	0,912	0,633	0,857
3	0,912	0,608	0,870
4	0,837	0,635	0,747
5	0,815	0,612	0,736

граммное средство адекватно реагирует на вводимые изменения в структуру модели, что говорит о чувствительности модели к входным данным.

Заключение

Разработанный метод на основе синтеза марковских цепей позволяет оценить эффективность функционирования ФЗ. Применение разработанного математического аппарата автоматизировано. Программное средство адекватно реагирует на вводимые в структуру модели изменения, что говорит о чувствительности модели к входным данным. Результатами моделирования являются вероятности безопасного состояния ЭИ при попытке нарушителя проникнуть на объект по каждому маршруту. На каждом пути проникновения обеспечивается безопасное состояние объекта на уровне не меньше заданных требований.

Достоинством разработанного метода моделирования является тот факт, что повышается достоверность результатов оценки из-за возможности детализировать маршрут проникно-

вения нарушителя и, получив множество разнородных по сложности участков, корректно задавать входные данные по скорости преодоления этих участков, определяя их экспертным путем и согласовывая с экспериментальными данными натурного эксперимента на реальном объекте.

Список литературы

1. Леус А. В. Математическая модель оценки эффективности систем физической защиты // Т-Comm — Телекоммуникации и Транспорт. 2010. № 6. С. 46—49.
2. Корчагин С. И. Оценка эффективности ИК ФЗ в рамках вероятностного подхода // Т-Comm — Телекоммуникации и Транспорт. 2010. № 4. С. 46—47.
3. Звездинский С. С., Голубков Г. В., Иванов В. А. Оценка функциональной эффективности охранной сигнализации малых объектов // Спецтехника и связь. 2008. № 3. С. 13—20.
4. Панин О. А. Проблемы оценки эффективности функционирования систем физической защиты объектов // Безопасность, достоверность информация (БДИ). 2005. № 3. С. 22—27.
5. Панин О. А. Как измерить эффективность // БДИ. 2008. С. 20—24.
6. Рябинин И. А. Надежность и безопасность структурно-сложных систем. СПб.: Изд-во СПбГУ, 2007. 276 с.
7. Боровский А. С., Тарасов А. Д. Автоматизированное проектирование и оценка систем физической защиты потенциально — опасных (структурно-сложных) объектов. Часть 1: Системный анализ проблемы проектирования и оценки систем физической защиты. Самара; Оренбург: Сам ГУПС, ОриПС — филиал Сам ГУПС, 2012. 163 с.
8. Рекомендации по обеспечению безопасности информации в ключевых системах информационной инфраструктуры. ФСТЭК России, 2007 г.
9. Мишин Е. Т., Соколов Е. Е. Построение систем физической защиты потенциально опасных объектов. М.: Радио и связь, 2005. 200 с.
10. Вентцель Е. С., Овчаров Л. А. Теория случайных процессов и ее инженерные приложения. М.: Высшая школа, 2003. 432 с.
11. Вентцель Е. С. Теория вероятностей. М: Наука, 2013. 368 с.
12. Никитин В. В., Цицулин А. К. Телевидение в системах физической защиты. URL: <https://www.security-bridge.com/> (дата обращения 13.10.2005).

Assessment of Efficiency of Physical Security of Information of Crucial Objects on the Basis of Markov Chains

The method of modeling of assessment of efficiency of physical security of information of crucial objects is developed. Feature of the solution of a task is decomposition of a complicated multigraph of penetration of the violator to information resources of an object on a great number of ordinary graphs, each of which is presented in the form of the sequence of events. This is decomposition of a difficult task on a set of simple subtasks. The result of decomposition was presented in the incidence matrix form: lines — number of a way of penetration, columns — edges of the graph of penetration by the violator. Every way of the violator and reaction of physical protection to penetration is modelled by two information-related Markov chains. Probability of information security is chosen as an indicator of efficiency of physical protection. The system of assessment of security consists of disjunctive-connected ways of penetration, at the same time every way it is a set conjunctive-connected edges of penetration which is defined by the probability of realization of the purpose of the violator. The general assessment of physical security of information is made on the most pessimistic way of penetration of the violator.

Keywords: assessment of efficiency of physical protection, multigraph of penetration of the violator, incidence matrix

DOI: 10.17587/it.25.757-765

References

1. Leus A. V. *T-Comm — Telekommunikacii i Transport*, 2010, no. 6, pp. 46—49 (in Russian).
2. Korchagin S. I. *T-Comm — Telekommunikacii i Transport*, 2010, no.10, pp. 46—47 (in Russian).
3. Zvezhinskij S. S., Golubkov G. V., Ivanov V. A. *Spektsh-nika i Svyaz'*, 2008, no. 3, pp. 13—20 (in Russian).
4. Panin O. A. *Bezopasnost', dostovernost' informaciya (BDI)*, 2005, no. 3, pp. 22—27 (in Russian).
5. Panin O. A. *Bezopasnost', dostovernost' informaciya (BDI)*, 2008, no. 2, pp. 20—24 (in Russian).
6. Ryabinin I. A. Reliability and safety of structurally complex systems, SPb, Publishing house of SPbGU, 2007, 276 p. (in Russian).
7. Borovskij A. S., Tarasov A. D. Computer-aided design and assessment of physical protection systems for potentially dangerous (structurally complex) objects. Part 1: System analysis of the

problem of designing and evaluating physical protection systems, Samara; Orenburg, Sam GUPS, OrIPS—filial Sam GUPS, 2012, 163 p. (in Russian).

8. **Recommendations** for ensuring information security in key systems of information infrastructure, FSTEK Rossii, 2007 g. (in Russian).

9. Mishin E. T., Sokolov E. E. Construction of physical protection systems for potentially dangerous objects, Moscow, Radio i svyaz', 2005, 200 p. (in Russian).

10. Ventcel' E. S., Ovcharov L. A. The theory of random processes and its engineering applications, Moscow, 2003, 432 p. (in Russian).

11. Ventcel' E. S. Probability theory, Moscow, Nauka, 2013, 368 p. (in Russian).

12. Nikitin V. V., Ciculin A. K. Television in physical protection systems, available at: <https://www.security-bridge.com/> (date of access 13.10.2005) (in Russian).

17—18 апреля 2020 г. в Барнауле на базе Алтайского государственного университета состоится



X Международная научно-практическая конференция
«ВЫСОКОПРОИЗВОДИТЕЛЬНЫЕ
ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ
В НАУЧНЫХ ИССЛЕДОВАНИЯХ, АВТОМАТИЗАЦИИ
УПРАВЛЕНИЯ И ПРОИЗВОДСТВА
(ВВСТ-2020)»



Секции работы конференции:

Секция 1

Многопроцессорные вычислительные системы и сети, многоядерные процессоры и программируемые логические структуры, цифровая обработка сигналов

Секция 2

Параллельное программирование и компьютерное моделирование явлений и процессов в естественных областях с использованием параллельных вычислений

Секция 3

Робототехника, автоматизация управления, автоматизация производства и научного эксперимента

Подробная информация на сайте конференции: <http://konf.asu.ru/hpcst/>

Указатель статей, опубликованных в журнале "Информационные технологии" в 2019 г.

МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ

- Антониади Г. Д., Архипов В. О., Цуприков А. А.** Математическая модель задержки автотранспорта на регулируемом перекрестке № 4
- Агасиев Т. А.** Метод секторизации полной карты вариативности целевой функции для оценки характерных признаков задачи оптимизации № 12
- Гуракий М. М.** Эффективный алгоритм моделирования кольцевых генераторов методом гармонического баланса. № 10
- Деменков Н. П., Микрин Е. А., Мочалов И. А.** Нечеткое оптимальное управление линейными системами. Часть 1. Позиционное управление № 5
- Деменков Н. П., Микрин Е. А., Мочалов И. А.** Нечеткое оптимальное управление линейными системами. Часть 2. Программное управление № 6
- Ершов М. Д.** Методы оптимизации первого порядка в машинном обучении № 11
- Зак Ю. А.** Алгоритмы оперативного перераспределения ресурсов в системах сетевого планирования и управления выполнением проектов № 8
- Иванников А. Д.** Теоретические основы выбора множества отладочных тестов проектов цифровых систем на основе алфавита выполняемых функций № 11
- Лебедев Б. К., Лебедев О. Б., Пурчина О. А.** Модифицированный роевой алгоритм «муравьиное дерево» в задаче диверсификации трассировочных ресурсов № 11
- Левин В. И.** Непрерывная логика и анализ надежности сложных систем. Математический аппарат. № 4
- Левин В. И.** Непрерывная логика и анализ надежности сложных систем. Исследование некоторых классов систем № 6
- Медведев А. В., Яременко Д. И.** Непараметрическое моделирование Т-процессов в условиях неполной информации. № 10
- Нефедов А. С., Шакиров В. А.** Автоматизация процедуры заполнения матриц парных сравнений альтернатив по критериям при использовании метода анализа иерархий № 6
- Родионов А. Н.** Абстрактные роли и примитивы ролевого моделирования сущностей в системе «концептуальная-логическая-физическая модели данных» № 8
- Салахов А. З.** Исследование рассеяния электромагнитных полей в биообъектах методом компьютерного моделирования № 12
- Тарасов В. Н., Бахарева Н. Ф., Када Отхмане.** Математическая модель телетрафика на основе системы $NE_2/M/1$ № 4
- Ульянов М. В., Фомичев М. И.** Сравнительный анализ комбинаций метода ветвей и границ с метаэвристическими алгоритмами для решения асимметричной задачи коммивояжера № 10
- Цвиркун А. Д., Топка В. В.** Равномерное распределение невозобновимого ресурса в крупномасштабных инновационных проектах № 11

СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ

- Гридин В. Н., Анисимов В. И., Абухазим М. М.** Индексно-численная технология сжатия данных в системах автоматизации схемотехнического проектирования № 6

- Гридин В. Н., Анисимов В. И.** Организация параллельных вычислительных процессов в распределенных системах автоматизации схемотехнического проектирования № 11
- Щекин А. В.** Конструкторско-технологическая параметризация в составе интегрированной САМ-системы № 7

НЕЙРОСЕТЕВЫЕ ТЕХНОЛОГИИ

- Абдуллаев Н. Т., Исмаилова К. Ш.** Выбор шага итерации в процессе обучения нейронной сети при использовании релаксационных алгоритмов № 3
- Амосов О. С., Амосова С. Г., Иванов Ю. С., Жиганов С. В.** Моделирование интеллектуальной системы контроля и управления доступом транспортных средств с использованием глубоких нейронных сетей № 2
- Барский А. Б., Мельник Д. И., Решетников А. В.** Нейросетевые методы управления качеством при модернизации и развитии сложных систем в условиях финансовых и технологических ограничений № 2
- Барский А. Б., Мельник Д. И.** Нейросетевая модель целераспределения для вычислительной системы архитектуры dataflow № 7
- Савченко Л. В.** Система постановки произношения на основе сверточных нейронных сетей и информационной теории восприятия речи № 5
- Соловьев Р. А., Тельпухов Д. В., Кустов А. Г., Исаева Т. Ю., Волков А. А.** Применение методов модулярной арифметики при разработке аппаратных реализаций нейронных сетей № 12

ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ И СЕТИ

- Абрамов А. Г., Евсеев А. В.** Концептуальные аспекты создания в Российской Федерации национальной исследовательской компьютерной сети нового поколения № 12
- Барский А. Б., Мельник Д. И., Пирожник В. В.** Модель управления качеством вычислительных средств распределенной многоканальной системы массового обслуживания № 6
- Ефанов Д. В.** Условия построения полностью самопроверяемых дискретных устройств на основе метода логического дополнения до равновесного кода «2 из 4» № 3
- Ефанов Д. В.** Троичный код паритета в системах рабочего диагностирования устройств автоматики и вычислительной техники № 7
- Инютин С. А.** Дробно-рациональные конструкции в компьютерной модулярной арифметике № 9
- Карганов В. В., Шевченко А. А., Малышев Б. Ю.** Способ повышения работоспособности информационно-вычислительной сети за счет адаптивного управления защитой № 1
- Кулагин В. П., Малых Е. С.** Проектирование матричных вычислительных структур с использованием сетей Петри № 5
- Петров И. С.** Системы обнаружения скомпрометированных коммутаторов в программно-конфигурируемых сетях № 3
- Попов А. Ю.** Применение гетерогенной вычислительной системы с набором команд дискретной математики для решения задач на графах № 11
- Пуртов А. М.** Использование образцов для выбора маршрутов в сетях передачи данных № 4
- Романов А. Ю., Ведмидь Е. А., Монахова Э. А.** Проектирование сетей на кристалле с топологией кольцевой циркулянт с тремя образующими: разработка алгоритмов маршрутизации № 9
- Тарасов В. Н., Бахарева Н. Ф., Отхмане Када.** Моделирование телетрафика на основе системы $NE_2/H_2/1$ № 9

- Филиппов И. В., Мелик-Адамян А. Ф., Сухомлинов В. А.** Размещение сетевых функций в облачных инфраструктурах. № 4
- Холопов Ю. А., Нгуен Ван Хиеу.** Реализация контроллера активной синхронной измерительной подсистемы цифровой системы управления. № 12
- Юсупов Р. М., Воробьев В. И., Петров М. Ю.** Представление эволюции средств вычислительной техники в экспозиции Музея истории СПИИРАН. № 5

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

- Грибова В. В., Шалфеева Е. А.** Обеспечение жизнеспособности систем, основанных на знаниях. № 12
- Грушо Н. А., Тимонина Е. Е.** Сравнение архитектур много-агентных систем. № 5

БАЗЫ ДАННЫХ

- Гвоздев В. Е., Блинова Д. В., Кириллов К. В.** Информационная поддержка принятия решений по устранению дефектов, выявленных при эксплуатации систем обработки данных и управления. № 3
- Мезенцев Ю. А., Разумникова О. М., Тарасова И. В., Трубинова О. А.** О некоторых задачах кластеризации больших данных по минимаксным и аддитивным критериям, применение в медицине и нейрофизиологии. № 10
- Потапов Д. Р.** Исследование эффективности применения кэша для использования в самоадаптирующихся контейнерах данных. № 4

БЕЗОПАСНОСТЬ ИНФОРМАЦИИ

- Асратян Р. Э.** Организация обработки защищенных сообщений в распределенных системах на основе Cryptographic Message Syntax. № 7
- Коляда А. А., Кучинский П. В., Червяков Н. И.** Пороговый метод разделения секрета на базе избыточных модулярных вычислительных структур. № 9
- Менисов А. Б., Шастун И. А., Капицын С. Ю.** Подход к выявлению вредоносных сайтов сети Интернет на основе обработки лексических признаков адресов (URL) и усредненного ансамбля моделей. № 11
- Салахутдинова К. И.** Повышение точности идентификации программного обеспечения путем использования аддитивного критерия Фишберна. № 10

ПРОГРАММНАЯ ИНЖЕНЕРИЯ

- Pashayev F. H., Pashayeva S. E., Najafov H. T., Suleymanli B. A.** Operating Algorithms for Folders and Files. № 3
- Бершадский А. М., Бождай А. С., Евсеева Ю. И., Гудков А. А.** Концепция рефлексивной самоадаптации прикладных программных систем. № 1
- Сарамуд М. В., Ковалев И. В., Лосев В. В., Петросян М. О., Калинин А. О.** Сравнение отказоустойчивых моделей программного обеспечения в имитационной среде исполнения. № 1
- Юлюгин Е. А.** Корректное и быстрое моделирование инструкций архитектуры INTEL® 64, нарушающих условия обратной совместимости. № 3

WEB-ТЕХНОЛОГИИ

- Бова В. В., Кравченко Ю. А., Кулиев Э. В., Курейчик В. В.** Моделирование поведения субъекта в интернет-сервисах на основе модифицированного алгоритма бактериальной оптимизации. № 7

ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ И ИЗОБРАЖЕНИЙ

- Абдуллаев Н. Т., Дышин О. А., Ибрагимова И. Д., Ахмедова Х. Р.** Сегментирование нестационарных физиологических сигналов с фрактальными свойствами. № 5
- Гречихин И. С., Савченко А. В.** Метод анализа предпочтений пользователя по фото и видео на мобильном устройстве на основе нейросетевых детекторов объектов на изображениях. № 9
- Жарков А. В.** Формирование классификационных признаков распознавания объектов при использовании лазерно-локационных средств их сканирования. № 6
- Игнатьев Н. А.** Компактность объектов классов и селекция обучающих выборок. № 9
- Касаткин С. С., Калайда В. Т.** Метод и реализация системы формализации описания класса изображений методами непараметрической статистики. № 2
- Нагорнов Н. Н., Ляхов П. А., Червяков Н. И.** Исследование шума квантования фильтров дискретного вейвлет-преобразования для обработки трехмерных изображений в медицине. № 7
- Фахми Ш. С., Рыжов Н. Г., Хасан Х. А.** Адаптивный алгоритм кодирования и декодирования изображений в транспортных системах наблюдения. № 8
- Федорова Т. О., Соснина О. А.** Оптимизация продолжительности визуализации сцен с Mental ray. № 10
- Шниперов А. Н., Сосновский М. С., Шипулин П. М.** Робастный метод маркирования изображений цифровым водяным знаком, основанный на ортогональных моментах Цернике. № 7

КОДИРОВАНИЕ И ОБРАБОТКА СИГНАЛОВ

- Ефанов Д. В.** Двухмодульные коды с суммированием — эффективный класс разделимых кодов для синтеза диагностического обеспечения устройств и систем управления. № 1
- Дворников С. В., Пшеничников А. В., Манаенко С. С.** Статистические характеристики помехозащищенных радиолиний с управлением частотным ресурсом. № 1

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

- Гвоздев В. Е., Христовуло О. И.** Информационная поддержка принятия решений в задачах управления отходами на основе математико-геоинформационного моделирования. № 8
- Гвоздкова Ю. Д., Гвоздкова И. А., Курочкин А. В., Черняев А. В.** Информационная система оценки экологической безопасности авиационных материалов и технологий методом анализа иерархий. № 3
- Костин В. Н.** Оценка эффективности физической защиты информации критически важных объектов на основе марковских цепей. № 12
- Кромина Л. А., Ярцев Р. А.** Об учете авторской значимости литературных ссылок при оптимизации заказа книжных изданий для библиотеки вуза. № 6

- Перемитина Т. О., Яценко И. Г.** Алгоритм комплексного анализа многомерных данных об экологическом состоянии окружающей среды. № 1
- Полесский С. Н., Востриков А. В., Прокофьева Е. Н., Романова И. И.** Методика оценки надежности вычислительной техники с учетом влияния отказов вспомогательного оборудования. № 4
- Полищук Ю. В.** Способ хранения электронных документов с квазиструктурированным контентом. № 1
- Полищук Ю. В.** Взаимосвязь ценности и старения информации при управлении большими техническими системами. № 6
- Торчинский В. Е., Торчинская И. В., Файнштейн А. С., Файнштейн С. И.** «Гибридный» алгоритм планирования государственных закупок товарно-материальных ценностей. № 4

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ И ПРОИЗВОДСТВЕ

- Авдошин С. М., Цветков Д. Д.** Методика проведения ICO. . . № 10
- Грибова В. В., Величко А. С., Колмогоров А. В.** Концепция цифровой платформы для производственной деятельности. № 8
- Кравченко Т. К., Брускин С. Н., Гоменюк К. С., Огуречников Е. В.** Разработка архитектуры требований при реализации ИТ-проекта по повышению эффективности управления складской деятельностью торговой компании сферы электронной коммерции. № 10
- Кузнецова Е. В.** Автоматизация проектной деятельности в организациях, выполняющих контрактные ИТ-проекты. № 9
- Левин С. Е., Окренг Я. Н., Нагибин С. Я., Балакирев Н. Е.** Математическая модель технологического процесса производства стирала. № 9

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

- Алдунин Д. А., Федин Г. Г.** Математическая модель для построения оптимальной индивидуальной образовательной траектории обучающегося при изучении массовых открытых онлайн-курсов
- Вишнеков А. В., Иванова Е. М., Басова К. Э., Ветелина Е. О.** Интерактивный учебно-исследовательский комплекс для моделирования процессов в вычислительных системах. . . . № 8

- Лобачев С. Л.** Дистанционные образовательные технологии в очном обучении: опыт, анализ и некоторые выводы. № 11

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В БИОМЕДИЦИНСКИХ СИСТЕМАХ

- Гельцер Б. И.** Концепция гетерогенного хранилища биомедицинской информации. № 2
- Грибова В. В., Москаленко Ф. М., Шахгельян К. И., Гмарь Д. В., Артюшина Е. В., Ершов Д. Е., Прохода Ю. А.** Роль интернет-технологий в коммуникационной деятельности медицинской организации. № 3
- Исмаилова К. Ш.** Применение различных методов оптимизации при расчете погрешности нейронной сети для диагностирования нервно-мышечных заболеваний. . . . № 1

КОМПЬЮТЕРНАЯ ГРАФИКА И ГЕОМЕТРИЧЕСКОЕ МОДЕЛИРОВАНИЕ

- Конопатский Е. В.** Аппроксимация геометрических объектов с помощью дуг кривых, проходящих через наперед заданные точки. № 1
- Соснина О. А., Филинских А. Д., Ложкина Н. А.** Анализ методов создания виртуальных моделей нетривиальных форм. № 11

КОДИРОВАНИЕ И ОБРАБОТКА СИГНАЛОВ

- Ефанов Д. В.** Двухмодульные коды с суммированием — эффективный класс разделимых кодов для синтеза диагностического обеспечения устройств и систем управления. № 1
- Дворников С. В., Пшеничников А. В., Манаенко С. С.** Статистические характеристики помехозащищенных радиолиний с управлением частотным ресурсом. № 1

НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

- Жевнерчук Д. В.** Обобщенный метод синтеза многокомпонентных интероперабельных структур на основе онтологии и недетерминированного конечного автомата. . . № 2
- Трубочкина Н. К., Поляков С. К.** Система электронного голосования на основе технологии блокчейн с использованием смарт-контракта. № 2

Адрес редакции:

107076, Москва, Стромьинский пер., 4

Телефон редакции журнала (499) 269-5510

E-mail: it@novtex.ru

Технический редактор *Е. В. Конова*.

Корректор *Н. В. Яшина*.

Сдано в набор 10.10.2019. Подписано в печать 26.11.2019. Формат 60×88 1/8. Бумага офсетная.

Усл. печ. л. 8,86. Заказ IT1219. Цена договорная.

Журнал зарегистрирован в Министерстве Российской Федерации по делам печати, телерадиовещания и средств массовых коммуникаций.

Свидетельство о регистрации ПИ № 77-15565 от 02 июня 2003 г.

Оригинал-макет ООО "Авансд солюшнз". Отпечатано в ООО "Авансд солюшнз".

119071, г. Москва, Ленинский пр-т, д. 19, стр. 1. Сайт: www.aov.ru

«МЕТОД СЕКТОРИЗАЦИИ ПОЛНОЙ КАРТЫ ВАРИАТИВНОСТИ ЦЕЛЕВОЙ ФУНКЦИИ ДЛЯ ОЦЕНКИ ХАРАКТЕРНЫХ ПРИЗНАКОВ ЗАДАЧИ ОПТИМИЗАЦИИ»

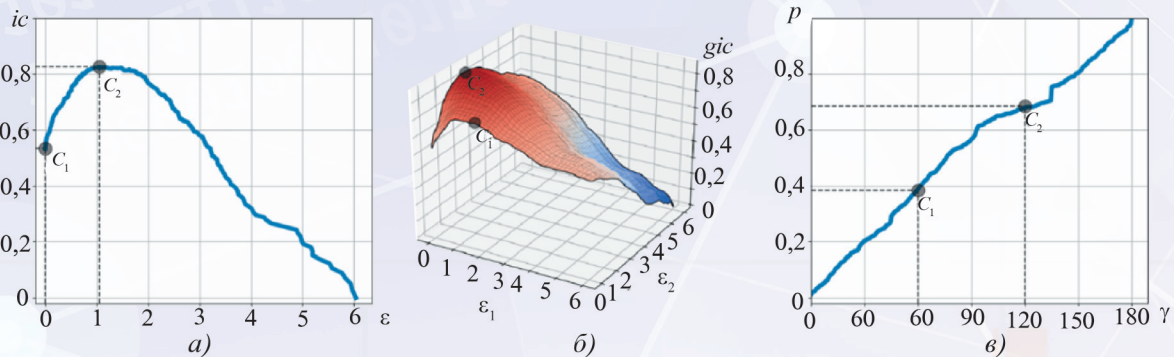


Рис. 9. Примеры точек характеризующих функций, координаты которых используем в качестве элементов вектора ХП в соответствии с методами информационного содержания (а), обобщенного информационного содержания (б) и секторизации карты вариативности (в)

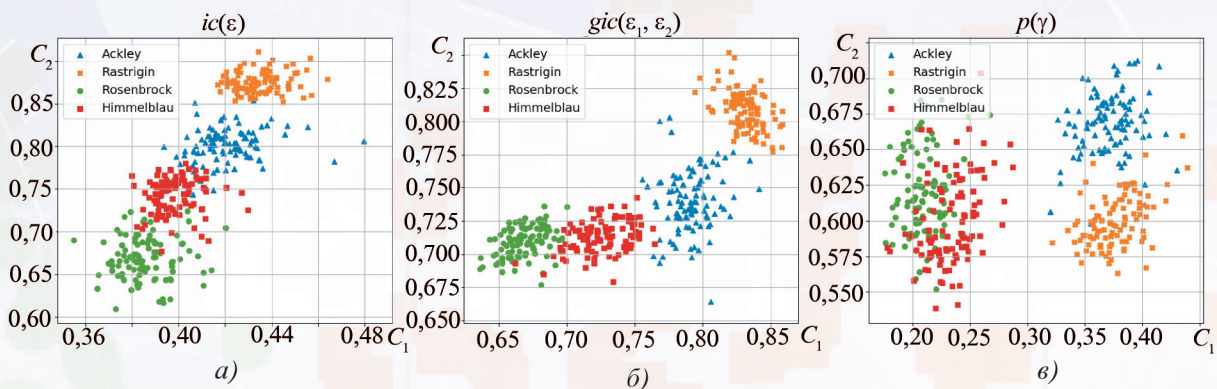


Рис. 10. Множества векторов $C_r, r \in [1:R]$, составленных на базе значений характеризующих функций методов информационного содержания (а), обобщенного информационного содержания (б) и секторизации карты вариативности (в); цветами обозначены соответствующие им метки

Рисунок к статье Р. А. Соловьева, Д. В. Тельпухова, А. Г. Кустова,
Т. Ю. Исаевой, А. А. Волкова

«ПРИМЕНЕНИЕ МЕТОДОВ МОДУЛЯРНОЙ АРИФМЕТИКИ ПРИ РАЗРАБОТКЕ АППАРАТНЫХ РЕАЛИЗАЦИЙ НЕЙРОННЫХ СЕТЕЙ»

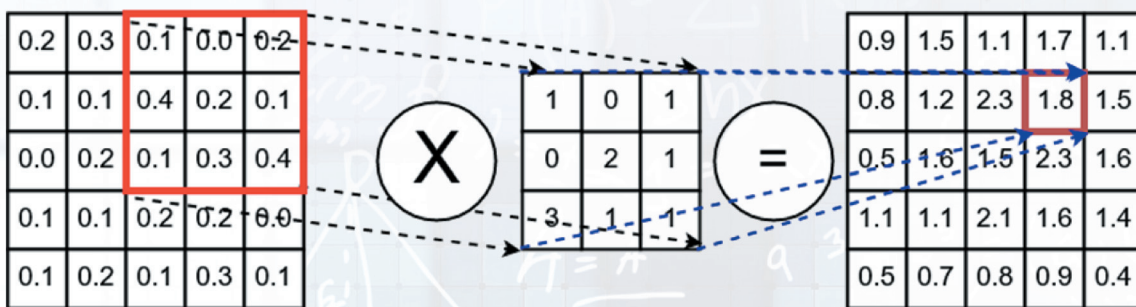
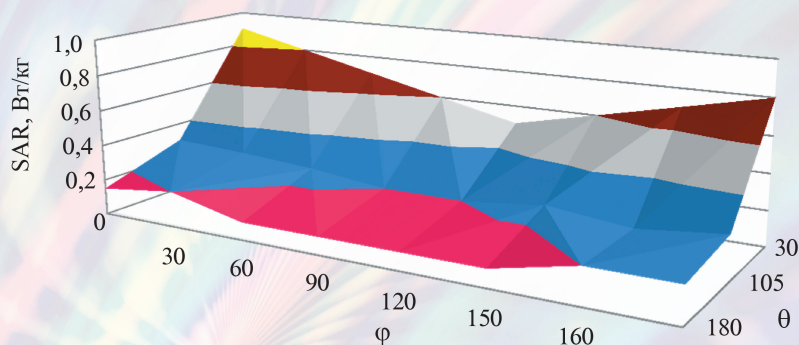
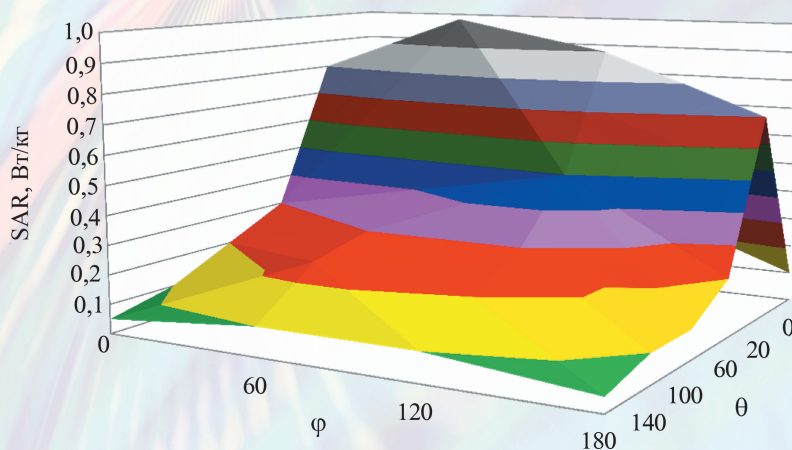


Рис. 3. Схематическое изображение операции Convolution2D

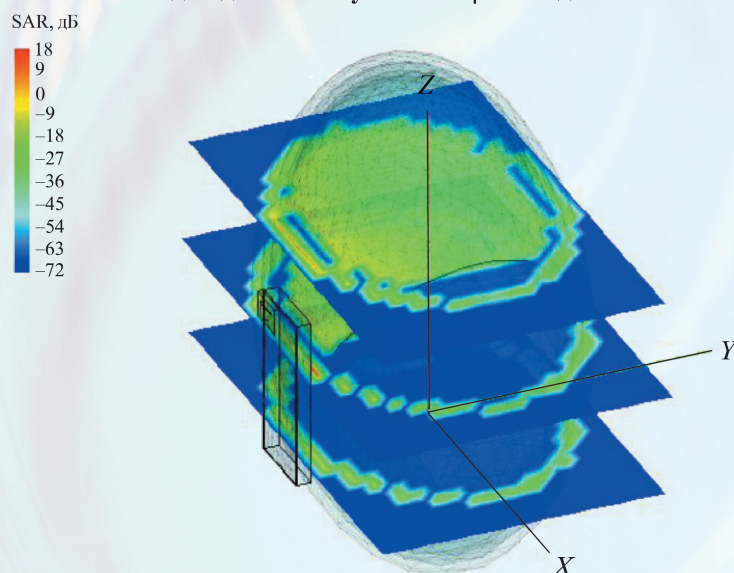
**«ИССЛЕДОВАНИЕ РАССЕЯНИЯ ЭЛЕКТРОМАГНИТНЫХ ПОЛЕЙ
В БИООБЪЕКТАХ МЕТОДОМ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ»**



**Рис. 6. Распределение SAR в первом слое сферического образования (кожа)
для диапазона углов θ и φ от 0° до 180°**



**Рис. 8. Распределение SAR во втором слое сферического образования (жировая ткань)
для диапазона углов θ и φ от 0° до 180°**



**Рис. 10. Результаты вычисления SAR в различных сечениях модели
с помощью компьютерного моделирования**