

# Международный журнал информационных технологий и энергоэффективности |



Том 2 Номер 1



2017



## СОДЕРЖАНИЕ / CONTENT

- 
1. **Балашов О.В., Кондратова Н.В.** Комплексное использование экстенсивной и нормальной форм анализа решений в условиях риска **2**

**Balashov O.V., Kondratova N.V.** Complex use of extensive and normal forms of the analysis of decisions in the conditions of risk

- 
2. **Тимощук А.А.** Способ многоуровневой грануляции текста для проведения автоматического определения тональности текста **8**

**Timoshchuk A.A.** Method of multilevel text granulation for text tonality automatic identifying

- 
3. **Чепиков А.М.** Нечеткая модель для управления локальным пожаротушением **16**

**Cherikov A.M.** Fuzzy model for a local fire extinguishing control

- 
4. **Войцицкая А.С.** Способ измерения массы жидких веществ на основе искусственных нейронных сетей **26**

**Voitsitskaya A.S.** The measurement method of mass of liquid substances based on artificial neural networks

- 
5. **Дорофеева Ю.В.** Об аспектах риска в процессе создания IT-проекта **31**

**Dorofeeva Y.V.** About aspects of risk in the process of creating an IT-project

- 
6. **Воевода П.Е., Зернов М.М.** Способ отображения генетических алгоритмов на структуру гетерогенной вычислительной системы **42**

**Voevoda P.E., Zernov M.M.** The mapping of genetic algorithms on the structure of heterogeneous computing system

---



Международный журнал информационных технологий и энергоэффективности

Сайт журнала:

<http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004.83

## КОМПЛЕКСНОЕ ИСПОЛЬЗОВАНИЕ ЭКСТЕНСИВНОЙ И НОРМАЛЬНОЙ ФОРМ АНАЛИЗА РЕШЕНИЙ В УСЛОВИЯХ РИСКА

**Балашов О.В., Кондратова Н.В.**

*Смоленский филиал Российского университета кооперации, Россия, (214018, г. Смоленск, проспект Гагарина, дом 58); e-mail: smolensk@rucoop.ru*

---

**Вводится понятие комбинированного показателя анализа решений в условиях риска. Предлагается комплексное использование экстенсивной и нормальной форм анализа решений в условиях риска.**

---

Ключевые слова: принятие решений, риск, вероятность, ценность, полезность, граф переходов.

## COMPLEX USE OF EXTENSIVE AND NORMAL FORMS OF THE ANALYSIS OF DECISIONS IN THE CONDITIONS OF RISK

**Balashov O.V., Kondratova N.V.**

*Smolensk branch of the Russian university of cooperation, Smolensk, Russia (214018, Smolensk, Gagarin ave., 58); e-mail: smolensk@rucoop.ru*

---

**The concept of the analysis of decisions combined indicators in the conditions of risk is entered. Complex use of extensive and normal forms of the analysis of decisions in the conditions of risk is offered.**

---

Key words: decision-making, risk, probability, value, utility, a transition graph.

При принятии управленческих решений, когда начальной информацией для принятия решения являются вероятностные характеристики реализации множества альтернатив в конечном множестве исходов, то говорят, что решение принимается в условиях риска [1, 2] или статистической неопределенности. Следовательно, в условиях статистической неопределенности, кроме определенного множества исходов и альтернатив, известны ещё и законы распределения вероятностных характеристик связей между альтернативами и исходами.

Вследствие того, что при принятии решения в реальных условиях обстановки всегда существует фактор случайности, а исходная информация часто не полностью определена и получена, то справедливо считать условия риска более близкими к реальным условиям, чем условия определённости, а условия неопределённости – ещё большим приближением к объективной реальности. Непременным условием решения задачи принятия решения является то, что для оценки исходов, альтернатив и существующих между ними связей должны быть выбраны количественные или качественные показатели, а для выбора лучшей альтернативы – критерий (или множество критериев).

В теории принятия решений существуют два подхода к постановке и решению задачи выбора в условиях статистической неопределенности.

Первый подход [3], называемый анализом решений в экстенсивной форме, реализует простую одноиндексную задачу принятия решений (рисунок 1). При анализе решений в экстенсивной форме альтернативы отображаются в исходы в каждой информационной ситуации  $S$ .

В такой постановке задачи выбора множество  $S$  в явном виде в анализе решений не присутствует. Статистическая неопределенность описывается распределениями вероятностей появления исходов, соответствующих альтернативам из множества  $A$ .

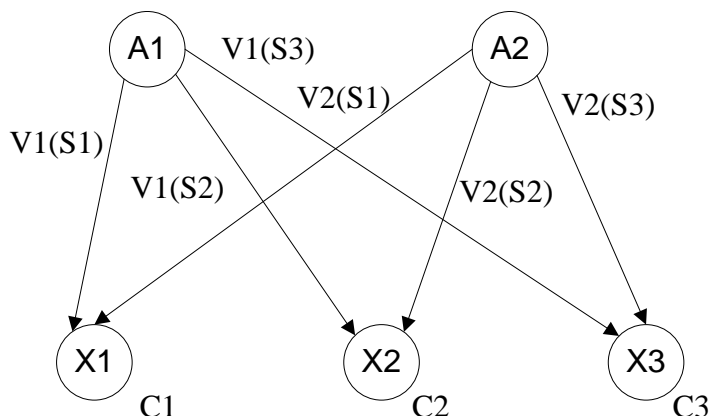


Рисунок 1 – Одноиндексная задача принятия решений

При одноиндексной задаче вероятность реализации  $j$ -й альтернативы в  $i$ -й исход ( $V_{ij}$ ) равна вероятности появления этого исхода в  $k$ -ой ситуации  $V(x_{ik})$ . Лучшая та альтернатива, показатель ожидаемой ценности исходов которой по всем ситуациям максимален (минимален)

$$A^* \rightarrow I^*, A^* \subset A, VC(A^*) = \max (\min) \left\{ \sum_{k=1}^K [V(x_k) \times C(x_k)] \right\}, \quad (1)$$

где  $V(x_k)$  – вероятностная характеристика получения исхода в  $k$ -й ситуации;

$C(x_k)$  – ценность исхода в  $k$ -й ситуации.

Другой подход и принятие решений по схеме Сэвиджа [4], называемый анализом решений в нормальной форме, базируется на предположении, что один и тот же исход может быть получен в результате реализаций нескольких альтернатив в различных информационных ситуациях.

При анализе решений в нормальной форме альтернатива  $A$  определяется как отображение состояний в исходы, а статистическая неопределенность описывается с помощью одного распределения вероятностей на множестве состояний ( $V(S_k)$ ). Это распределение не зависит от выбора той или иной альтернативы. Функция ценности (полезности), по Сэвиджу, строится на множестве пар «альтернатива – состояние»  $\{C(A_i; S_{kj})\}$ . Выбор на множества альтернатив основан на максимизации показателя ожидаемой ценности (полезности) альтернатив по зависимости:

$$A^* \rightarrow I^*, A^* \subset A, VC(A^*) = \max (\min) \left\{ \sum_{k=1}^K [V_{ij}(S_k) \times C(A_i, S_k)] \right\}, \quad (2)$$

где  $V_{ij}(S_k)$  – вероятностная характеристика получения реализаций  $i$ -й альтернативы в  $j$ -й исход в  $k$ -й ситуации;

Подход реализует двухиндексную задачу ПР (рисунок 2).

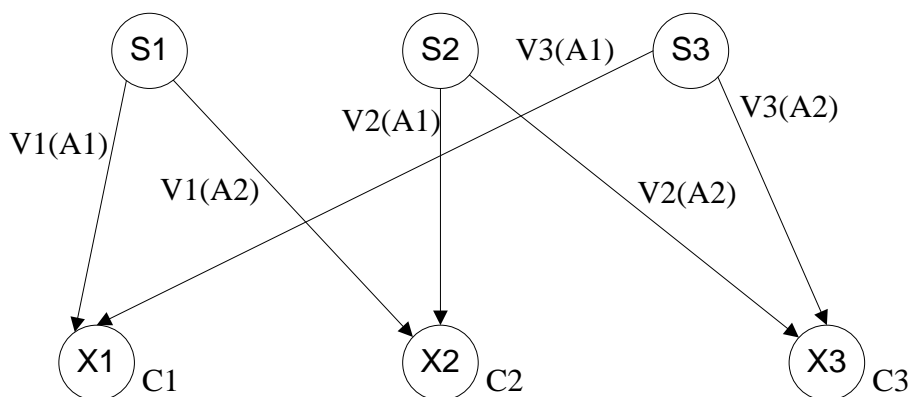


Рисунок 2 – Двухиндексная задача принятия решения

Поскольку при двухиндексной задаче возможно появление одного исхода в результате реализации различных альтернатив, то значения ожидаемой ценности альтернатив будут зависеть как от ценности исходов, так и от вероятности реализации альтернатив в конкретной ситуации.

Экстенсивная и нормальная формы анализа решений требуют получения распределений вероятностей либо на множестве альтернатив, либо на множестве ситуаций. В общем случае каждое из распределений влияет не только на оценку альтернативы, но и определенным образом воздействует на оценку исхода.

В реальных условиях принятия решений полезность исхода в различных ситуациях должна меняться даже при неизменной ценности каждого из исходов. Поэтому направлениями дальнейшего совершенствования моделей принятия решений в условиях, адекватных реальным, должны быть: развитие способов представления и учета показателей  $C$  и  $V$ , соответствующим условиям задачи принятия решения; комплексное использование различных форм анализа решений в моделях принятия решений.

При принятии управленческих решений рассматриваются полученные на основе обработки данных экспертных оценок:  $u_j$  – полезность  $j$ -го исхода и  $S_{ji}$  – соответствующая субъективная вероятность. В зависимости от источников получения количественных значений величин  $V$  и  $C$  возможны четыре варианта определения показателя  $VC$  (таблица 1):

- объективная ожидаемая ценность  $[P_{ji} \times w_i]$
- субъективная ожидаемая ценность  $[S_{ji} \times w_i]$
- объективная ожидаемая полезность  $[P_{ji} \times u_i]$
- субъективная ожидаемая полезность  $[S_{ji} \times u_i]$

Рассматриваемые показатели образуют полную группу объективно-субъективных характеристик для всех возможных вариантов информации в условиях статистической неопределённости и могут использоваться как в одноиндексной, так и в двухиндексной задачах принятия решений.

Показатели принятия решений в условиях риска и полученные на их основе комбинированные показатели применяются для решения сложных задач в соответствии с принятыми схемами выбора.

В том случае, когда решение принимается применительно к нескольким ситуациям, в которых численные значения показателей  $V$  и  $C$  могут изменяться от ситуации к ситуации, схемы выбора могут информационно взаимодействовать между собой, причём в нормальную

форму может входить экстенсивная форма, позволяющая заранее определить значения ожидаемой функции ценности  $C(A_i, S_k)$  или полезности  $u(A_i, S_k)$  для каждого из состояний  $S$ .

Таблица 1 – Показатели оценки альтернатив для условий риска

		Численное значение вероятности реализации $i$ -й альтернативы ( $A_i$ ) получено:	
		объективно $\langle p_i \rangle$	субъективно $\langle s_i \rangle$
Численное значение ценности $j$ -го исхода ( $I_j$ ), полученное по:	результатам опытов (объективно) $\langle w_i \rangle$	Объективная ожидаемая ценность (математическое ожидание ценности) $[P_i \times w_i]$	Субъективная ожидаемая ценность $[s_i \times w_i]$
	оценкам экспертов (субъективно) $\langle u_i \rangle$	Объективная ожидаемая полезность $[P_i \times u_i]$	Субъективная ожидаемая полезность $[s_i \times u_i]$

В целях получения более точных оценок альтернатив при наличии информации о вероятности появления ситуаций, вероятности реализации альтернатив в каждой из ситуаций и численных значений ценности (полезности), получаемых при этой реализации исходов, выражение (2) преобразовывается к виду

$$VC(A) = \max \{ VC(A_i) = \sum_{k=1}^K C(X_j, A_i, S_k) V(A_i, S_k) V(S_k) \}, \quad (3)$$

где  $V(A_i, S_k)$  – вероятность реализации  $i$ -й альтернативы при условии, что имеет место  $k$ -я ситуация;

$C(X_j, A_i, S_k)$  – ценность  $j$ -го исхода, к которому приводит  $i$ -я альтернатива в  $k$ -й ситуации;  
 $V(S_k)$  – вероятность появления  $k$ -й ситуации.

В выражении (3) учитываются оценки исходов в каждой информационной ситуации. Вероятность реализации альтернатив описывает схему отображений множества альтернатив в исходы для каждой ситуации. Ситуации описываются с помощью функции распределения вероятности их появления.

В качестве примера рассматривался задача выбора с использованием критерия (3).  
 Задаются непустые множества:

- исходов  $\{X1, X2, X3, X4\}$ ;
- альтернатив  $\{A1, A2, A3\}$ ;
- ситуаций  $\{S1, S2, S3\}$ .

Значения показателей задачи принятия решения показаны в таблице 2.

Таблица 2 – Значения показателей для рассматриваемого примера

Ситуация S1	Ситуация S2	Ситуация S3
$V(A1, S1) = 0,3$ $V(A2, S1) = 0,2$ $V(A3, S1) = 0,5$	$V(A1, S2) = 0,4$ $V(A2, S2) = 0,5$ $V(A3, S2) = 0,1$	$V(A1, S3) = 0,3$ $V(A2, S3) = 0,3$ $V(A3, S3) = 0,4$
$C(X1, A1, S1) = 5$ $C(X2, A2, S1) = 4$ $C(X3, A3, S1) = 1$	$C(X2, A1, S2) = 4$ $C(X3, A2, S2) = 3$ $C(X4, A3, S2) = 3$	$C(X1, A1, S3) = 5$ $C(X4, A2, S3) = 3$ $C(X3, A3, S3) = 2$

В соответствии с информационными ситуациями строятся графы отображений альтернатив в исходы (рисунок 3) и присваиваются значения показателям:  $C(X_j, A_i, S_k)$ ,  $V(A_i, S_k)$ ,  $V(S_k)$ .

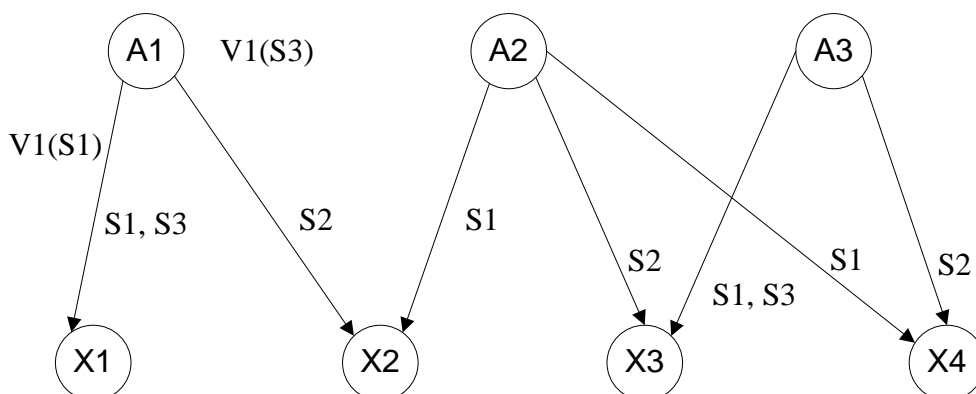


Рисунок 3 – Графы отображений альтернатив в исходы по ситуациям

Выбор наилучшей альтернативы проводится по критерию максимума численных значений ожидаемой ценности альтернатив в соответствии с выражением  $VC(A^*) = \max \{VC(A_i)\}$ ;

$$VC(A^*) = \max \{VC(A1) = 1,51; VC(A2) = 0,92; VC(A3) = 0,63\}.$$

Из расчетов следует, что лучшей альтернативой является A1.

Предложенная формальная модель решения задач в условиях статистической неопределённости может быть использована только при условии получения множества стохастических оценок для полной группы событий. Отсутствие объективно полученных значений показателей может быть скомпенсировано проведением экспертного опроса и обработки субъективных данных [5].

Однако субъективно полученные оценки позволяют решить задачу принятия решения и выбора на множестве альтернатив с меньшей степенью точности, чем объективно полученные оценки.

Переход к более сложным формальным моделям показателей функции полезности, учитывающим физическую природу предпочтений на множестве альтернатив для каждого конкретного случая представления исходной информации и определенных условий принятия решения, может существенно повысить качество принимаемых решений.

### Список литературы

1. Трухаев Р. И. Модели принятия решений в условиях неопределенности. – М.: Наука, 1981.
2. Блюмин С. Л., Шуйкова И.А. Модели и методы принятия решений в условиях неопределенности. – Липецк: Изд-во ЛЭГИ, 2001.
3. Трахтенгерц Э.А. Компьютерные методы реализации экономических и информационных управленческих решений. В 2-х томах. Том 1. Методы и средства. – М.: СИНТЕГ, 2009.
4. Ларичев О.И. Теория и методы принятия решений. – М.: Логос, 2002.
5. Балашов О.В., Кондратова Н.В. Подход к формированию начального множества альтернатив в системах поддержки принятия решений// Материалы международной научно-практической конференции. – Смоленск: СФ РУК, 2012. – С. 188-193.

### References

1. Trukhayev R. I. Decision-making models in the conditions of uncertainty. – М.: Science, 1981. (in Russian)

2. Blumin S. L. Shyikova I. A. Models and methods of decision-making in the conditions of uncertainty. - Lipetsk: publishing house LEGY, 2001. (in Russian)
  3. Trakhtengerts E. A. Computer methods of realisation of economic and information administrative decisions. In 2 volumes. Volume 1. Methods and means. - M: SINTEG, 2009. (in Russian)
  4. Larichev O. I. Theory and methods of decision-making. – M.: Logos, 2002. (in Russian)
  5. Balashov O.V., Kondratova N.V. The approach to formation of initial set of alternatives in systems of support of decision-making//Materials of the international scientifically-practical conference. - Smolensk: publishing house Smolensk branch of the Russian university of cooperation, 2012. – P. 188-193. (in Russian)
-





ОТКРЫТАЯ НАУКА  
ИЗДАТЕЛЬСТВО

Международный журнал информационных технологий и  
энергоэффективности

Сайт журнала:

<http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004.896

## СПОСОБ МНОГОУРОВНЕВОЙ ГРАНУЛЯЦИИ ТЕКСТА ДЛЯ ПРОВЕДЕНИЯ АВТОМАТИЧЕСКОГО ОПРЕДЕЛЕНИЯ ТОНАЛЬНОСТИ ТЕКСТА

Тимощук А.А.

*Филиал федерального государственного бюджетного образовательного учреждения высшего образования «Национальный исследовательский университет МЭИ» в г. Смоленске, Россия (214013, г. Смоленск, Энергетический проезд, дом 1); email: [timoshchuk.anastasia@mail.ru](mailto:timoshchuk.anastasia@mail.ru)*

Статья посвящена проблеме автоматического анализа тональности текста. Предлагается новый способ многоуровневой грануляции текста для автоматического определения тональности, комбинирующий результаты работы классификатора на основе векторного представления текста (Doc2Vec) и метода ключевых слов. Doc2Vec - алгоритм обучения без учителя, учится получать распределенные векторы для частей текстов. Метод ключевых слов основан на подсчете весов входящих в текст признаков. Способ многоуровневой грануляция текста включает две фазы – обучение и распознавание. На первой фазе на основе размеченной коллекции текстов происходит обучение Doc2Vec-классификатора и классификатора на базе ключевых слов. На второй фазе результаты распознавания нового текста обоими классификаторами объединяются и формируется итоговое решение; при этом учитываются степени уверенности классификаторов в своих результатах.

Ключевые слова: анализ тональности текста, текстовая классификация, метод ключевых слов, машинное обучение, Doc2Vec.

## METHOD OF MULTILEVEL TEXT GRANULATION FOR TEXT TONALITY AUTOMATIC IDENTIFYING

Timoshchuk A.A.

*Smolensk Branch of Federal state budgetary educational institution of higher education "National research University Moscow power engineering Institute", Russia (214013, Smolensk, Energeticheski proezd, 1); e-mail: [timoshchuk.anastasia@mail.ru](mailto:timoshchuk.anastasia@mail.ru)*

The article is devoted to the problem of automatic text sentiment analysis. The new method of multilevel text granulation for text tonality automatic identifying which combines the results of work of classifier based on vector text representation (Doc2Vec) and Keywords method is suggested in the article. Doc2Vec is an unsupervised learning algorithm, learning to receive distributed vectors for parts of texts. The Keywords method is based on the counting of weights of features which the text contains.

The method of multilevel text granulation contains two phases – training and recognition. On the first phase Doc2Vec -classifier and Keywords-classifier are trained on the labeled collection of texts. On the second phase the results of recognition by both classifiers for a new text are combined and the final decision is formed; confidence levels of both classifiers are taken into account.

Keywords: text sentiment analysis, text classification, keywords method, machine learning, Doc2Vec.

Мнения занимают основное место почти во всех областях человеческой деятельности. Наши убеждения и представления о реальности и выбор, который мы делаем, в значительной мере зависит от того, как другие видят и оценивают мир. По этой причине, когда нам нужно принять решение, мы часто ищем чужие мнения. Это справедливо не только для людей, но и для организаций. Мнения и связанные с ними понятия, такие как чувства, оценки, отношения и эмоции являются предметом изучения анализа настроений (sentiment analysis). Зарождение и быстрое развитие этой области связано с интересами людей в Интернете, как правило, спрос всегда порождает предложение. В качестве примера можно привести различные отзывы, форумы, обсуждения, блоги, социальные сети. Впервые в человеческой истории имеется огромный объем мнений, записанных в цифровом формате. С начала XXI века сфера анализа тональности данных стала одной из наиболее активно развивающихся и исследуемых направлений в области обработки естественных языков.

Сентимент анализ (анализ тональности текста) – это обработка естественного языка, классифицирующая тексты по эмоциональной окраске. Такой анализ можно рассматривать как метод количественного описания качественных данных, с присвоением оценок настроения. Например, необходимо выявить автора текста – определить субъект, затем то, о чем ведется речь – объект разговора, и, наконец, отношение первого ко второму – определение тональности.

Формализация задачи сентимент анализа выглядит следующим образом. Пусть  $X$  – множество всех документов,  $Y$  – множество меток размера  $N$ ,  $\varphi: X \rightarrow Y$  – целевая функция определения тональности документов, значения которой известны только для обучающего подмножества  $X_{train} = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ . Необходимо найти алгоритм  $\alpha: X \rightarrow Y$ , который способен классифицировать произвольный документ  $x \in X$ , расставить метки.

Задача определения эмоциональной окраски текста является задачей классификации, она может быть бинарной (негативный, позитивный), тернарной (негативный, нейтральный, позитивный) или  $n$ -арной (например, для  $n=5$ : сильно негативный, умеренно негативный, нейтральный, умеренно позитивный, сильно позитивный).

Все способы сентимент анализа можно разделить на три группы:

1. способ, основанный на словарях;
2. способ, основанный на и правилах;
3. машинное обучение.

Для получения более точных результатов классификации многие исследователи предлагают комбинировать результаты различных способов. Так в статье [2] объединяется метод опорных векторов (англ. SVM, support vector machine) и метод ключевых слов (англ. KWM, key words method). Во время экспериментов на тестовой коллекции (набор отзывов пользователей портала Imhonet.ru на различные фильмы) были получены следующие значения точности классификации (процент текстов, по которым классификатор принял правильное решение):

1. SVM-классификатор – 0,869%;
2. KWM-классификатор – 0,875%;
3. Комбинированный метод – 0,888%.

Результаты классификации комбинированным методом превосходят результаты, полученные SVM-классификатором и классификатором на основе ключевых слов. И это

подтверждает целесообразность использования комбинаций различных способов для проведения сентимент анализа текста.

В данной работе предлагается способ многоуровневой грануляции текста автоматического определения тональности текста, комбинирующий результаты работы классификатора на основе векторного представления текста и метода ключевых слов.

В качестве векторного классификатора предлагается использовать алгоритм Doc2Vec, предложенный в 2014 году Томасом Миколовым [7].

### Схема способа

Способ многоуровневой грануляции текста объединяет результаты работы классификатора на основе векторного представления текста (Doc2Vec-классификатор) и классификатора на базе ключевых слов (KWM-классификатор). До начала применения предлагаемого способа необходимо задать дискретную шкалу для измерения тональности, содержащую  $K$  классов. Способ включает в себя две фазы – обучение и распознавание (классификацию) (рисунок 1). Для фазы обучения необходим обучающий текстовый набор данных, каждый текст в котором отнесен к одному из  $K$  классов.



Рисунок 1 – Обобщенная схема способа многоуровневой грануляции документов для проведения автоматического определения тональности текста

Схема фазы обучения (рисунок 2) включает в себя следующие этапы [6]: предобработку текста, формирование словарей признаков (слов), взвешивание признаков методом ключевых слов, формирование векторной модели Doc2Vec, обучение векторного Doc2Vec-классификатора, настройку параметров для объединения результатов работы двух классификаторов.

Схема фазы распознавания (рисунок 3) также состоит из нескольких этапов, первые два из которых совпадают с этапами предобработки и формирования векторной модели в фазе обучения.

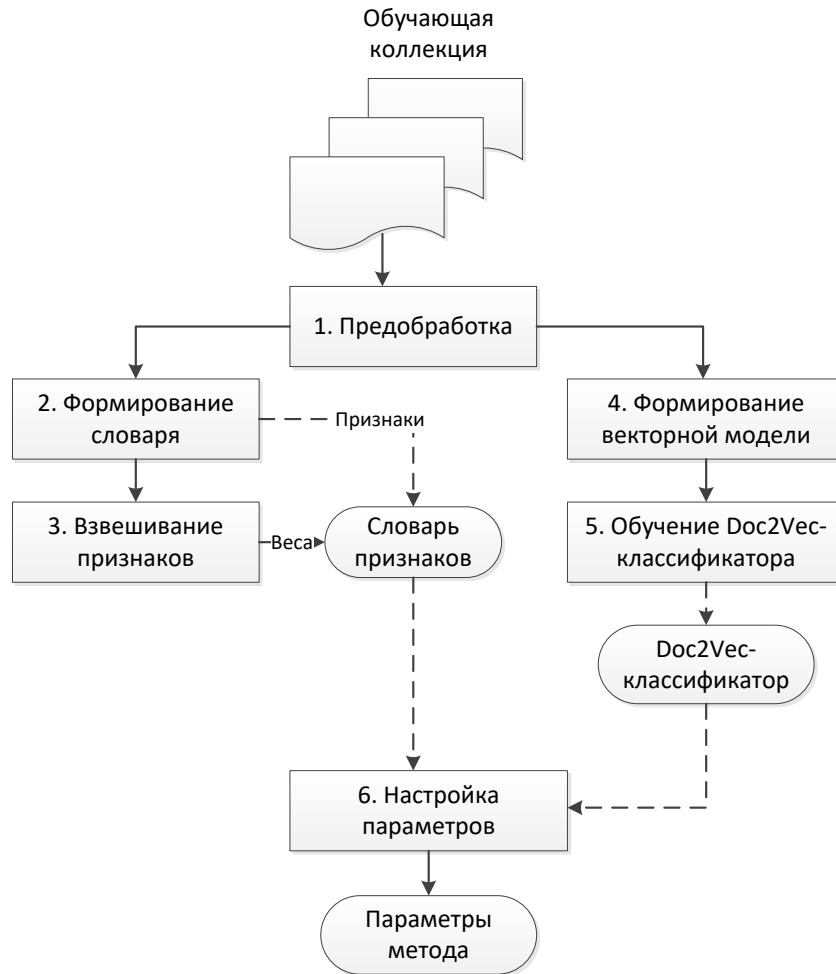


Рисунок 2 – Схема фазы обучения

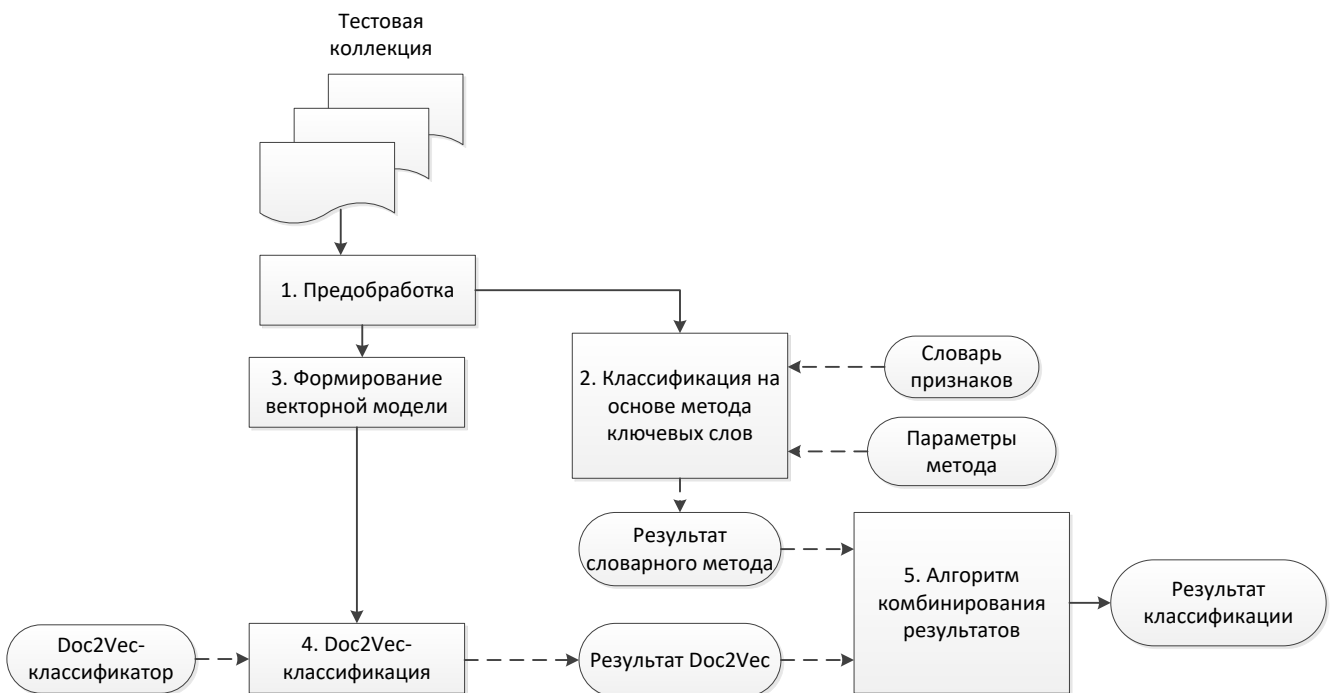


Рисунок 3 – Схема фазы распознавания

Затем для всех текстов из тестовой коллекции применяются классификации на базе метода ключевых слов и Doc2Vec, в результате чего каждый классификатор определяет к какому из  $K$  классов относится текст.

На завершающем этапе результаты классификации второго и третьего этапов объединяются, и полученные значения классификации составляют итоговый результат работы способа многоуровневой грануляции документов.

### **Doc2Vec**

Многие методы машинного обучения требуют на вход данные, представленные в виде вектора признаков фиксированной длины. Когда дело доходит до текстов одним из самых распространенных способов представления векторов фиксированной длины является «мешок слов». Однако эта модель имеет много недостатков. Порядок слов теряется, разные предложения, имеющие одни и те же слова, могут иметь одинаковые представления. Алгоритм Doc2Vec решает эту проблему.

Алгоритм Doc2Vec (первоначальное название Paragraph Vector), алгоритм обучения без учителя, учится получать распределенные векторы для частей текстов [7]. Тексты могут быть переменной длины: от предложения до большого документа. Для простоты все входные тексты в данном разделе будут именоваться документами.

В данной модели векторные представления документов обучаются предсказывать слова в документе, точнее берется вектор документа и объединяется с несколькими векторами слов из него, и модель пытается предсказать следующее слово с учетом контекста. Векторы слов и документов обучаются с использованием метода стохастического градиентного спуска и метода обратного распространения ошибки. Векторы документов являются уникальными, а векторы одинаковых слов в разных документах совпадают.

Существует 2 архитектуры для построения векторных представлений документов:

#### **1. DM**

В данной конструкции каждый документ представлен уникальным вектором в виде столбца в матрице  $D$ , и каждый термин представлен уникальным вектором в виде столбца в матрице  $W$ . Вектор документа и векторы слов в нем объединяются или усредняются для предсказания следующего слова из контекста.

В этой архитектуре можно рассматривать токены как отдельные слова. Они действуют как память, которая помнит, что отсутствует в текущем контексте или теме документа. По этой причине модель называется Distributed Memory.

#### **2. DBOW**

Описанный выше метод рассматривает объединение вектора документа с векторами слов, входящих в него, для предсказания следующего слова в текстовом окне. Другой способ заключается в игнорировании слов из контекста на входе, но при этом модель должна предсказывать случайно отобранные слова для документа на выходе. Это означает, что на каждой итерации стохастического градиентного спуска просматривается текстовое окно, затем просматривается случайное слово в текстовом окне и формируется задача классификации с учетом вектора документа. Эта версия называется Distributed Bag-of-Words [7].

Для определения тональности текста Doc2Vec-классификатор сначала обучается на текстах неизвестной тональности для составления векторной модели, при этом создаются

экземпляры двух моделей Doc2Vec – DM и DBOW. Затем готовой модели передаются тексты с указанной тональностью, и модель обучается предсказывать тональность текста, используя логистическую регрессию. В результате таких действий получается готовая модель, которой на вход подается текст, а на выходе – вероятность принадлежности текста одному из  $K$  классов.

### Метод ключевых слов

Идея метода ключевых слов заключается в том, что тональность текста определяется на основе подсчета весов входящих в него признаков (ключевых слов). Для вычисления весов признаков (слов) в данном методе будет использоваться RF (Relevance Frequency – релевантная частота), предложенная в [1, 3].

Обозначим  $t$  как количество текстов, содержащих  $i$ -й признак и принадлежащий классу  $D$ ;  $a$  – это количество текстов, содержащих  $i$ -й признак и не принадлежащих классу  $D$ . Тогда вес  $i$ -го признака для класса  $D$  будет вычисляться формулой

$$RF_i^D = \log_2 \left( 2 + \frac{t}{a} \right). \quad (1)$$

В процессе использования метода ключевых слов [5] для каждого класса  $D$  вес  $i$ -го признака  $RF_i^D$  вычисляется независимо по формуле (1). Признаки с весами для всех классов сохраняются в словаре признаков.

Для каждого класса  $D_l$  вычисляется вес текста  $V_l$  путем простого суммирования весов входящих в текст признаков:

$$V_l = \sum_{i=1}^K v_i^l, l \in [1, \dots, K], \quad (2)$$

где  $v_i^l$  – вес  $i$ -го признака в тексте  $T$  по отношению к классу  $D_l$ .

Затем среди полученных весов находится максимальный:

$$V_{max} = \max\{V_l\}, l = 1, \dots, K. \quad (3)$$

Текст  $T$  относится к тому классу, для которого получен максимальный вес  $V_{max}$ .

### Алгоритм комбинирования результатов классификации

После применения классификаторов будут получены следующие параметры:

- $K_{Doc2Vec}$  – класс, выданный Doc2Vec-классификатором на основе сформированной векторной модели текстов;
- $K_{KW}$  – класс, выданный KWM-классификатором;
- $Conf(K_{Doc2Vec})$  – степень уверенности Doc2Vec-классификатора;
- $Conf(K_{KW})$  – степень уверенности KWM-классификатора.

Алгоритм, объединяющий результаты двух классификаторов для текста  $T$  состоит из следующих шагов:

1. Текст  $T$  относится к классу  $K$ , если решения классификаторов совпадают, то есть  $K_{Doc2Vec} = K_{KW} = K$ ;

2. Если один из классификаторов сильно уверен в своем решении, а другой слабо, то текст  $T$  относится к классу, который выдал классификатор с сильной уверенностью.

3. Если одновременно оба классификатора слабо или сильно уверены в своих решениях, но при этом решения не совпадают, выбирается решение Doc2Vec-классификатора, так как он выбран приоритетным.

В итоге данный алгоритм позволяет решать конфликты двух классификаторов, используя оптимальные параметры, найденные на фазе обучения.

### Тестирование

Для обоснования целесообразности использования векторного Doc2Vec-классификатора сравним его по точности классификации текстов с векторным SVM-классификатором.

Тестирование и сравнительный анализ выбранных способов автоматического сентимент анализа текста производилось на набор данных IMDB, который часто используется в качестве ориентировочного набора обучающей и тестовой коллекции текстов для проведения сентимент анализа.

Набор данных состоит из 100 000 отзывов на различные фильмы, случайным образом взятых с сайта IMDB (крупнейший в мире веб-сайт о кинематографе). Одним из ключевых аспектов этого набора данных является то, что каждый отзыв на фильм имеет несколько предложений. 100 000 отзывов делятся на три набора данных: 25 000 размеченных текстов по тональности для обучения, 25 000 размеченных текстов по тональности для тестирования и 50 000 неразмеченных текстов по тональности для обучения.

В данном наборе данных задана дискретная шкала измерения тональности, содержащая два класса отзывов: положительный и отрицательный. Отзывы сбалансированы как в наборе для обучения, так и в тестовом наборе.

Для сравнения работы классификаторов будем использовать метрику – точность классификации (процент текстов, по которым классификатор принял правильное решение). Результаты тестирования представлены в таблице 1.

Таблица 1. Результаты работы классификаторов на наборе IMDB

Модель классификатора	Точность, %
SVM	86.95
Doc2Vec	<b>90.24</b>

Из таблицы 1 видно, что Doc2Vec-классификатор является более точным, что подтверждает целесообразность использования в качестве векторного классификатора.

Следовательно, на основе проведенных исследований предполагается, что способ многоуровневой грануляция текста для автоматического определения тональности текста, объединяющий результаты двух классификаторов (Doc2Vec-классификатора и KWM-классификатора), позволит добиться повышения точности классификации.

### Список литературы

1. Котельников Е.В., Клековкина М.В. Автоматический анализ тональности текстов на основе методов машинного обучения // Компьютерная лингвистика и интеллектуальные технологии: по матер. ежегодн. Междунар. конф. «Диалог». 2012. № 11 (18). С. 753–762.
2. Котельников Е.В. Комбинированный метод автоматического определения тональности текста // Журнал Программные продукты и системы. 2012. № 3. С. 189–195.

3. Chetviorkin I., Braslavskiy P., Loukachevitch N. Sentiment Analysis Track at ROMIP 2011 // Computational Linguistics and Intellectual Technologies: Annual International Conf. «Dialogue», CoLing&InTel, 2012, no. 11 (18), pp. 739–746.
4. Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection // Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 1995, no. 2 (12), pp. 1137–1143.
5. Salton G., Buckley C. Term-weighting approaches in automatic text retrieval // Information Processing & Management, 1988, Vol. 24, no. 5. pp. 513–523.
6. Sebastiani F. Machine learning in automated text categorization // ACM Computing Surveys, 2002, Vol. 34, no. 1. pp. 1–47.
7. Tomas Mikolov, Quoc Le. Distributed Representations of Sentences and Documents. // In Proceedings of Workshop at The 31st International Conference on Machine Learning (ICML) – 2014.

### References

1. Kotelnikov E.V., Klekovkina M.V., CoLing&InTel, 2012, no. 11 (18), pp. 753–762.
  2. Kotelnikov E.V. Combined method of text tonality automatic identifying// Journal of Software products and systems, 2012, no. 3, pp. 189–195.
  3. Chetviorkin I., Braslavskiy P., Loukachevitch N. Sentiment Analysis Track at ROMIP 2011 // Computational Linguistics and Intellectual Technologies: Annual International Conf. «Dialogue», CoLing&InTel, 2012, no. 11 (18), pp. 739–746.
  4. Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection // Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 1995, no. 2 (12), pp. 1137–1143.
  5. Salton G., Buckley C. Term-weighting approaches in automatic text retrieval // Information Processing & Management, 1988, Vol. 24, no. 5. pp. 513–523.
  6. Sebastiani F. Machine learning in automated text categorization // ACM Computing Surveys, 2002, Vol. 34, no. 1. pp. 1–47.
  7. Tomas Mikolov, Quoc Le. Distributed Representations of Sentences and Documents. // In Proceedings of Workshop at The 31st International Conference on Machine Learning (ICML) – 2014.
-





ОТКРЫТАЯ НАУКА  
ИЗДАТЕЛЬСТВО

Международный журнал информационных технологий и  
энергоэффективности

Сайт журнала: <http://www.openaccessscience.ru/index.php/ijcse/>



## НЕЧЕТКАЯ МОДЕЛЬ ДЛЯ УПРАВЛЕНИЯ ЛОКАЛЬНЫМ ПОЖАРОТУШЕНИЕМ

**Чепиков А.М.**

*Филиал федерального государственного бюджетного образовательного учреждения высшего образования «Национальный исследовательский университет МЭИ» в г. Смоленске, Россия (214013, г. Смоленск, Энергетический проезд, дом 1); e-mail: a.m.chepikov@yandex.ru*

В статье предложена нечёткая модель поведения интеллектуального агента, управляющего пожаротушением. В отличие от существующих средств пожаротушения, модель отличается локальным применением средств пожаротушения, распределяет имеющиеся ресурсы для пожаротушения пропорционально величине угрозы, имеет возможность «делиться» ресурсами для тушения пожаров со смежными интеллектуальными агентами. Эффективность модели определяется снижением площади помещений, затронутой пожаром или пострадавших от средств пожаротушения (например, затопленных водой).

Ключевые слова: интеллектуальное поведение агента, способ управления огнем, пожаротушение, нечёткая модель.

## FUZZY MODEL FOR A LOCAL FIRE EXTINGUISHING CONTROL

**Chepikov A.M.**

*The Branch of Federal state budgetary educational institution of higher education "National research University Moscow power engineering Institute" in Smolensk, Russia (214013, Smolensk, Energeticheski proezd, 1); e-mail: a.m.chepikov@yandex.ru*

The article considers a fuzzy intelligent agent behavior model that controls fire fighting. In contrast to existing fire extinguishing means the model is notable for local application of fire extinguishing means. It allocates available firefighting resources in proportion to the rate of the threat and is able to "share" firefighting resources with allied intelligent agents. Effectiveness of the model is determined by reduction of the area affected by the fire or damaged by fire extinguishing means (e.g. flooded).

Keywords: intelligent agent behavior, method of fire extinguishing control, fire suppression, fuzzy model.

Современные технологии пожаротушения на промышленных предприятиях представляют собой сложные системы, зачастую, объединяющие и интегрирующие данные от множества систем: контроля над опасными производственными объектами [1, 2], экологического мониторинга, ERP [3] и MES-систем [4]. Однако, такая консолидация, как

правило, заключается на отдельные части единого пульта управления или диспетчерского пульта предприятия.

Зачастую, роль такого диспетчерского пульта, как правило, сводится к непрерывному мониторингу состояния отдельных объектов на предприятии, характеристик, протекающих на нём процессов для своевременного выявления отклонений от нормы, которые способны привести к некоторой аварии или снижению производительности (эффективности) предприятия. Кроме того, диспетчерский пульт может обеспечить контроль развития аварии в случае её наступления [5].

Однако, такая структура, зачастую, не позволяет своевременно идентифицировать, проанализировать, локализовать и устранить аварию или пожар на этапе их зарождения или начальных этапах развития. Основными рисками подобного подхода к построению системы для борьбы с авариями являются: гибель оператора, повреждение пульта управления пожаром, повреждение связей между пультом и датчиками.

Для устранения указанных недостатков, в [5] предложен подход к построению системы пожаротушения на основе интеллектуальных агентов и приведена структура таких агентов. В настоящей статье предлагается модель поведения этих агентов, обеспечивающая возможность построения сложных мультиагентных систем.

#### **Задача адресного тушения пожара**

В настоящее время хорошо изученными являются вопросы развития пожаров, моделирования возгораний и работы средств пожаротушения [6, 7, 8].

Рассмотрим задачу тушения пожара с учетом следующих ограничений.

1. Тушение пожара происходит в помещении, оборудованном водяной системой пожаротушения.

2. Вода к сплинкерам подводится по нескольким независимым трубам с определенной одинаковой пропускной способностью.

3. По помещению распределены температурные датчики.

4. В помещении работает человек, который должен эвакуироваться в случае возникновения пожара.

5. Сплинкеры могут управляться независимо.

6. Управление сплинкерами осуществляется интеллектуальными агентами, в помещении одновременно работает несколько интеллектуальных агентов.

Задача состоит в определении такого набора управляющих воздействий для сплинкеров, которые позволят эвакуировать человека, потушить пожар и нанести минимальный урон (огнем или водой) находящемуся в помещении имуществу. Обобщенно, модель такого помещения приведена на рисунке 1:

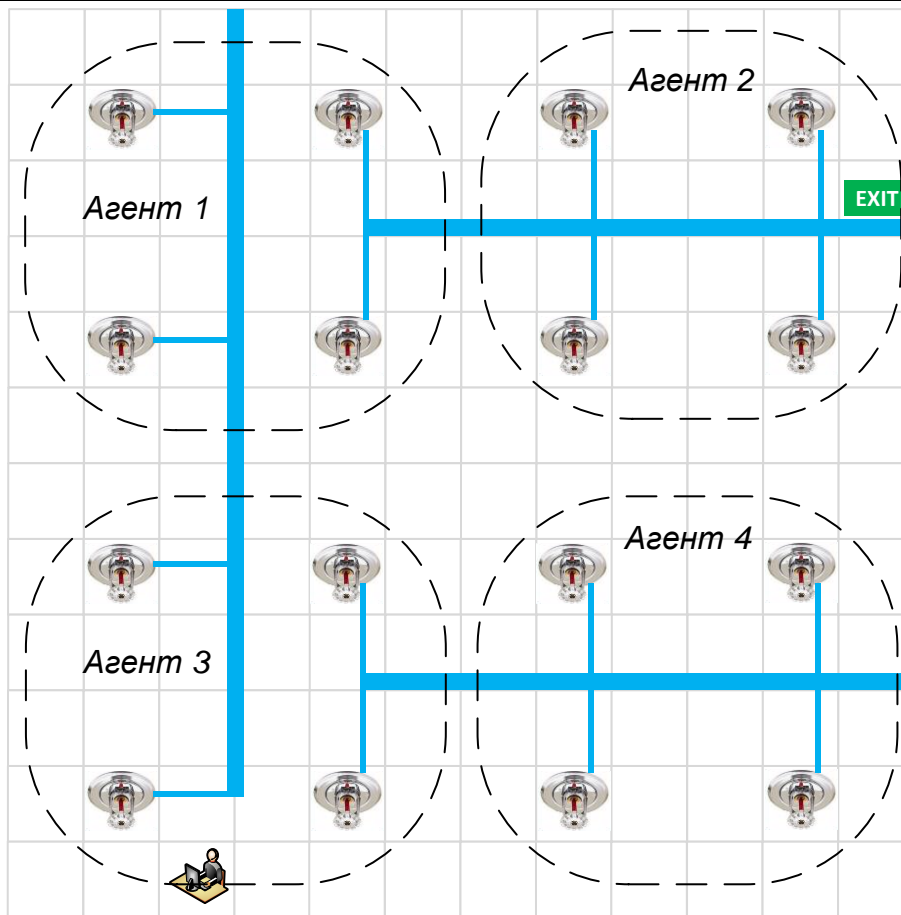


Рисунок 1 – Обобщенная модель помещения

### Структура нечёткой модели поведения интеллектуального агента

Для решения указанной задачи разработана нечеткая модель поведения интеллектуального агента, представленная на рисунке 2. Структура модели напоминает структуру нейро-нечёткого классификатора, однако, построение модели осуществляется не на основе нечётких продукционных правил, а на основе сведений о реальной системе пожаротушения.

Представленная структура модели состоит из 5 слоев.

Слой 1. На вход модели подается матрица нормированных температур  $t$  и матрица путей эвакуации человека  $m$ .

Слой 2. На вход 2 слоя подаются значения матрицы принадлежности температур и матрицы эвакуации человека. Элементы слоя реализуют операцию S-нормы:

$$S(t, m) = \begin{cases} \max(t, m), & \min(t, m) = 1 \\ 1, & \min(t, m) > 0 \end{cases}$$

где  $t$  – значение матрицы нормированных температур,  $m$  – значение матрицы эвакуации человека.

Данный слой предназначен для получения матрицы тушения. Эта матрица имеет размер, совпадающий с размером помещения. Каждый элемент матрицы имеет значение в нечётком диапазоне  $[0, 1]$ , характеризующей уровень тушения рассматриваемого участка помещения.

Слой 3 обеспечивает учет подводящих труб системы пожаротушения. Каждая труба представлена отдельным элементом слоя 3. На вход каждого элемента подаются значения элементов слоя 2 из области, тушение которой обеспечивается рассматриваемой трубой. В элементе слоя выполняется операция среднего взвешивания. Выходом данного слоя является нормированный расход воды с трубы. При выполнении операций среднего взвешивания, желательно использовать операции, ограничивающие рост неопределенности, описанные в [9, 10].

Слой 4 служит для перераспределения ресурсов между агентами в рамках одной трубы. Если от одной трубы поступает вода сразу на несколько интеллектуальных агентов, они должны распределить ресурсы между собой. Для этого, на слой 4 подаются значения нормированного расхода воды из трубы всех агентов, в которые поступает вода от этой трубы. На выходе 4 слоя получается нормированное выделяемое интеллектуальному агенту количество воды, рассчитываемое средним взвешиванием нечетких значений. Количество элементов слоя соответствует количеству труб для подачи воды.

Слой 5. Обеспечивает управление сплинкерами пропорционально силе горения с учётом выделяемого ресурса трубы. Для этого на вход слоя подаются результаты работы 2 и 4 слоев модели. Количество элементов 5 слоя соответствует количеству сплинкеров.

На рисунке 2, представлена, нечёткая модель поведения интеллектуального агента.

### **Способ применения нечёткой модели**

Процедура использования предложенной модели включает в себя следующие этапы.

Этап 1. Получение матрицы температур.

Данная матрица находится путем сбора данных с температурных датчиков, и дальнейшего их нормирования к величинам  $[0, 1]$ . При этом предполагается, что количество датчиков не позволяет контролировать точно температуру на каждом участке помещения, поэтому для получения температуры допускается применение операций аппроксимации. Полученные нормированные значения являются неточными, в связи с чем, их целесообразно представлять, нечёткими числами, например, треугольного вида.

Этап 2. Получение матрицы пути эвакуации человека.

Матрица пути эвакуации человека может быть построена с использованием волнового алгоритма. На первом шаге запускается волна от места нахождения человека к запасному выходу. При этом все препятствия и горящие участки считаются непроходимыми. Полученный путь является оптимальным, однако, в условиях пожара, не всякий человек сможет пройти таким оптимальным маршрутом. В критических ситуациях, человек склонен к панике, может наступить дезориентация. Для того чтобы учесть возможные направления движения человека в условиях психологического давления, полученный на предыдущем шаге четкий путь движения следует «размыть». При этом значение ячейки в матрице движения будет ассоциироваться с возможностью прохождения человеком указанного квадрата. Эти значения также могут быть представлены нечёткими числами треугольной формы.

Этап 3. Функционирование нечёткой модели поведения интеллектуального агента.

Работа нечёткой модели представляет собой следующую последовательность шагов:

Шаг 1. На вход модели подаются полученные на этапах 1 и 2 матрицы.

Шаг 2. Получение матрицы тушения.

Шаг 3. Расчет необходимого расхода воды из трубы.

Шаг 4. Распределение ресурсов трубы между агентами пропорционально их потребности.

Шаг 5. Формирование уровня выходного управляющего воздействия для каждого сплинкера.

Предложенная модель предполагает периодический пересчет управляющих воздействий в зависимости от динамики развития пожара и положения человека.

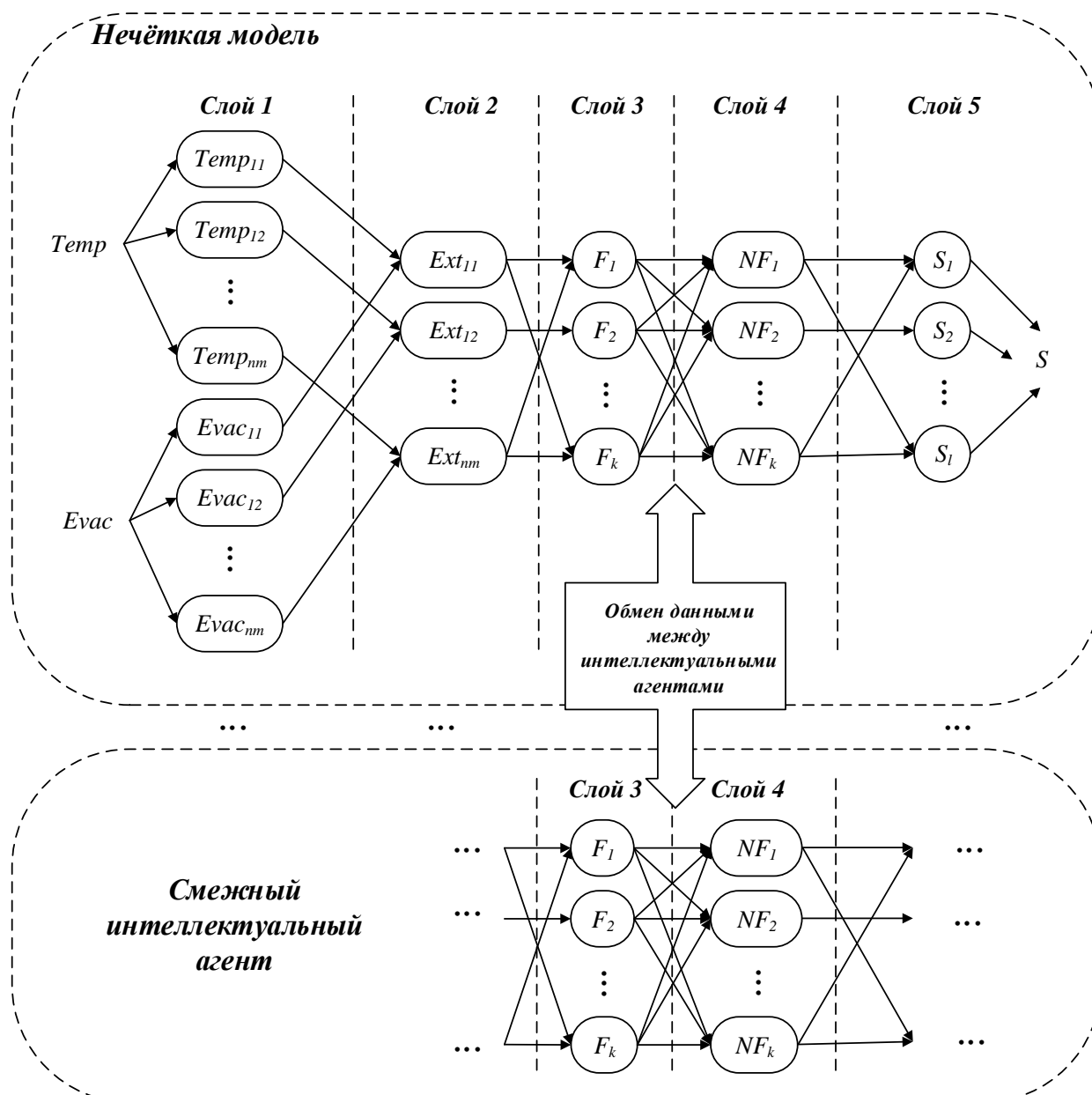


Рисунок 2 – Нечёткая модель поведения интеллектуального агента

### Пример применения нечёткой модели

Для апробации предложенной модели была разработана программа, в рамках которой пользователю предоставляется возможность задать:

1. структуру помещения;

2. положение очагов пожара;
3. положение человека;
4. положение сплинкеров и их характеристики;
5. задать трубы водоснабжения, их характеристики и связать сплинкеры с этими трубами.

В качестве примера рассматривалось помещение (склад) 6 м<sup>2</sup> шириной и 36 м<sup>2</sup> длиной (общая площадь 216 м<sup>2</sup>), в котором хранится линолеум с температурой возгорания 480-500 °С. На складе установлена система пожаротушения, включающая 24 сплинкера, запитанные 3 трубами подачи воды. Сплинкер тушит площадь размером 2\*2 м<sup>2</sup>. Сплинкеры управляются четырьмя интеллектуальными агентами. На складе работает кладовщик, который имеет свое рабочее место. В результате нарушения мер противопожарной безопасности, на складе возникает возгорание. На рисунке 3 отражен момент начала пожара. Отмечены охваченные огнем площади и место положения кладовщика.

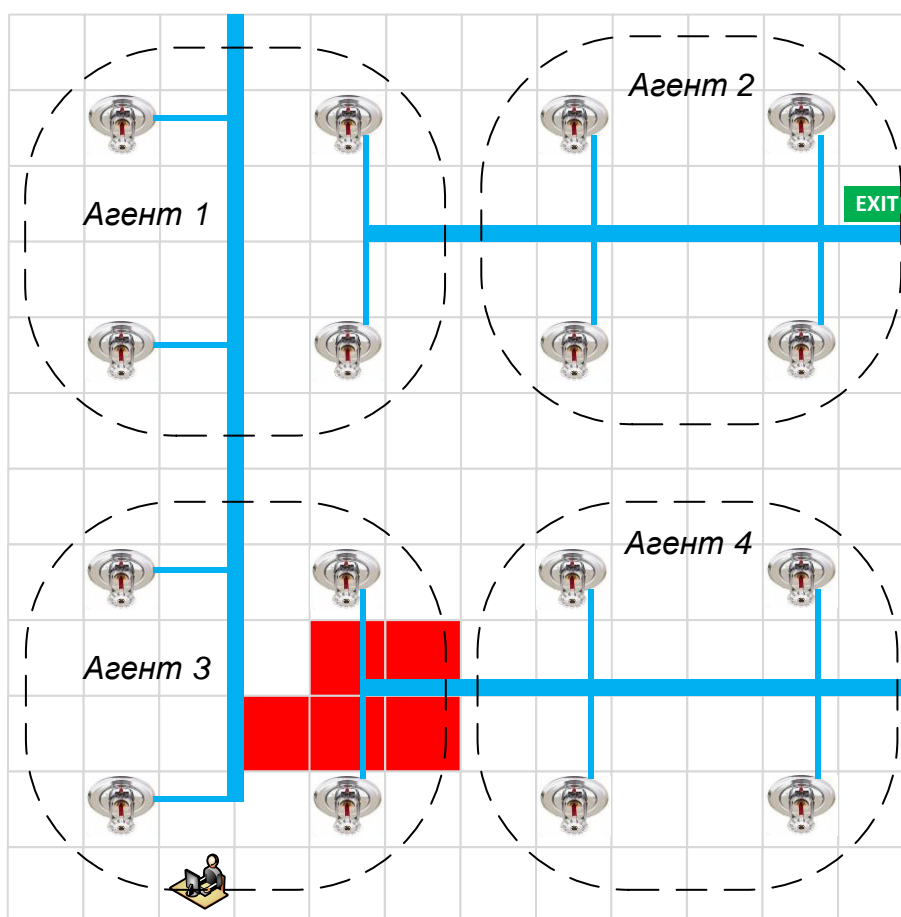


Рисунок 3 – Пример момента начала пожара

В соответствии с [11] для упрощения моделирования предполагаем наличие данных о температуре каждого квадратного метра склада. При наличии в соседней клетке огня, рассматриваемая клетка начинает постепенно нагреваться. Клетка воспламеняется при достижении ей температуры воспламенения линолеума. Также предполагаем, что склад не имеет системы вентиляции и огонь распространяется равномерно по всем направлениям.



слой 4 для трубы 1 подается значение слоя 3, рассчитанное агентом 3, а для трубы 2 – значение слоя 3, рассчитанное агентом 2. В результате получаем требуемый расход воды для интеллектуального агента (таблица 2).

Таблица 2 – Требуемый расход воды для интеллектуального агента

№ Трубы	№ Интеллектуального агента			
	1	2	3	4
1	0	X	0.256	X
2	0	0.7	X	X
3	X	X	1.078	1.911

В результате, мы получаем нормированное количество воды, предназначенное для интеллектуального агента (таблица 3).

Таблица 3 – Нормированное количество воды, выделяемое на интеллектуального агента

№ Трубы	№ Интеллектуального агента			
	1	2	3	4
1	0	X	1	X
2	0	1	X	X
3	X	X	0.361	0.639

На основании полученных нормированных значений выделяемых на интеллектуального агента, рассчитываем количество воды, поступающее на сплинкеры (таблица 4).

Таблица 4 – Количество воды, поступающее на сплинкеры

Номер сплинкера по строке	Номер сплинкера по столбцу			
	1	2	3	4
1	0	0	0	0.144
2	0	0	0	0.556
3	0	0.074	0.227	0.201
4	1	0.287	0.201	0.010

Моделирование проводилось пошагово. За один шаг температура горящих клеток росла на 10%, человек передвигался на 1 клетку. В результате моделирования получены следующие результаты:

Площадь, пострадавшая от пожара: 144 м<sup>2</sup>;

Площадь, пострадавшая от тушения: 81 м<sup>2</sup>;

Площадь, не затронутая ни пожаром, ни тушением: 63 м<sup>2</sup>;

При этом использовании классических систем пожаротушения, вся площадь склада будет зависеть от воды, таким образом, 36.8% склада было сохранено.



Предложенная нечёткая модель поведения интеллектуального агента, управляющего пожаротушением, отличается локальным применением средств пожаротушения, распределяет имеющиеся ресурсы для пожаротушения пропорционально величине угрозы, имеет возможность «делиться» ресурсами для тушения пожаров со смежными интеллектуальными агентами. Эффективность модели определяется снижением площади помещений, затронутой пожаром или пострадавших от средств пожаротушения (например, затопленных водой).

Предложенная модель обеспечивает работу только с одним средством пожаротушения – водяным и не учитывает возможного наличия порошковых огнетушителей и прочих средств. В будущих работах планируется разработать модель интеллектуального агента, комплексно управляющего различными типами средств пожаротушения. Такая система должна учитывать возможность применения тех или иных средств пожаротушения в отдельных зонах помещения. Например: использовать порошковые огнетушители в зоне нахождения человека нельзя, а тушение высокотемпературных пожаров водой запрещено. Кроме того, должны быть развиты механизмы моделирования возгорания, которые позволят прогнозировать движение огня и прокладывать путь эвакуации по безопасным траекториям с учетом динамики пожара.

### Список литературы

1. Acevedo, M.F. Real-Time Environmental Monitoring: Sensors and Systems, CRC Press, 2015.
2. Campbell, M., Sensor Systems for Environmental monitoring, Blackie Academic & Professional, 1997.
3. Bradford, M. Modern ERP: Select, Implement, & Use Today's Advanced Business Systems, 2015.
4. Kletti, J. Manufacturing Execution Systems – MES. Springer, 2007.
5. Сеньков, А.В. Андреева, О.Н. Подход к созданию интеллектуальной системы для контроля пожаров, аварий и инцидентов на промышленных предприятиях на основе теории многоагентных систем// *Фундаментальные исследования*. – 2016. – № 10-3. –с. 560-565; URL: <https://fundamental-research.ru/ru/article/view?id=40895>.
6. Baum, H. R. Analysis of the Pressure Rise in Sealed Magazine Fires. National Bureau of Standards, Gaithersburg, 1987.
7. Levine, R.S. Mathematical Modeling of Fires. National Bureau of Standards, 1980.
8. Rehm, R.G. McDermott, R.J. Mathematical Modeling of Wildland-Urban Interface Fires, *Monografías de la Real Academia de Ciencias de Zaragoza* 34, 105–142.
9. Федулов, А.С. Вид взаимодействия нечётких чисел, ограничивающих рост неопределенности при выполнении операций нечёткой арифметики // *Вестник МЭИ*, № 1, 2006. – с. 101–110.
10. Федулов, А.С. Работа с нечёткими числами. // *Нейрокомпьютеры: разработка, применение*, № 1, 2007. – с. 37–46.
11. Чепиков, А.М. Алгоритм моделирования пожара в помещении. // *Международный журнал информационных технологий и энергоэффективности*. – 2016. – Том 1 №2 с. 39-43 URL: <http://www.openaccessscience.ru/index.php/ijcse/article/view/11/pdf>.

### References

1. Acevedo, M.F. Real-Time Environmental Monitoring: Sensors and Systems, CRC Press, 2015.
2. Campbell, M., Sensor Systems for Environmental monitoring, Blackie Academic & Professional, 1997.

3. Bradford, M. Modern ERP: Select, Implement, & Use Today's Advanced Business Systems, 2015.
  4. Kletti, J. Manufacturing Execution Systems – MES. Springer, 2007.
  5. Senkov, A.V., Andreeva, O.N. Approach to Creation Of Intellectual System For Control Of The Fires, Accidents And Incidents At The Industrial Enterprises On The Basis Of The Theory Of Multi-Agent Systems // Fundamental Research. – 2016. – № 10-3. – pp. 560-565; URL: <https://fundamental-research.ru/ru/article/view?id=40895>.
  6. Baum, H . R . Analysis of the Pressure Rise in Sealed Magazine Fires. National Bureau of Standards, Gaithersburg, 1987.
  7. Levine, R.S. Mathematical Modeling of Fires. National Bureau of Standards, 1980.
  8. Rehm, R.G. McDermott, R.J. Mathematical Modeling of Wildland-Urban Interface Fires, Monografías de la Real Academia de Ciencias de Zaragoza 34, 105–142.
  9. Fedulov, A.S. The type of interaction of fuzzy numbers limiting increase of uncertainty when performing operations of fuzzy arithmetics // MPEI bulletin, № 1, 2006. – pp. 101–110.
  10. Fedulov, A.S. Steady operation of accumulation of fuzzy numbers // Neurocomputers: development, application, № 1, 2007. – pp. 37–46.
  11. Chepikov, A.M. Indoors fire modeling algorithm // International journal on information technologies and energy efficiency. – 2016. – Issue 1 №2 pp. 39-43 URL: <http://www.openaccessscience.ru/index.php/ijcse/article/view/11/pdf>.
-



ОТКРЫТАЯ НАУКА  
издательство

Международный журнал информационных технологий и  
энергоэффективности

Сайт журнала:

<http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004.9

## СПОСОБ ИЗМЕРЕНИЯ МАССЫ ЖИДКИХ ВЕЩЕСТВ НА ОСНОВЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

**Войцицкая А.С.**

*Филиал федерального государственного бюджетного образовательного учреждения высшего образования «Национальный исследовательский университет «МЭИ» в г. Смоленске (Россия, 214013, г. Смоленск, Энергетический проезд, дом 1);  
e-mail: voitsitskay@mail.ru*

---

Статья посвящена описанию предлагаемого способа измерения массы жидких веществ, процесс тарирования которого реализуется на основе искусственных нейронных сетей. Предлагаемый способ может быть реализован для измерения массы разных жидких веществ в различных резервуарах без его перенастройки.

---

Ключевые слова: искусственные нейронные сети, процесс тарировки.

## THE MEASUREMENT METHOD OF MASS OF LIQUID SUBSTANCES BASED ON ARTIFICIAL NEURAL NETWORKS

**Voitsitskaya A.S.**

*Branch of the federal state budget education institution of higher education «National Research University» MEI» in Smolensk (Russia, 214013, Smolensk, Energeticheskiy proezd, house 1);  
e-mail: voitsitskay@mail.ru*

---

The article is devoted to the description of the offered way of measurement of mass of liquid substances which process of a tarirovaniye is implemented on the basis of artificial neural networks. The offered way can be realized for measurement of mass of different liquid substances in various tanks without his change-over.

---

Key words: artificial neural networks, calibration process.

### Введение

В настоящее время известно большое количество систем [1] и способов [2] измерения массы жидких веществ, основными проблемами при этом являются измерения больших объемов при условии сложной геометрической формы резервуаров, и необходимости учета внутренних встроенных элементов. Существующие способы измерения, как правило, не учитывают влияние определенных факторов: температуры, атмосферного давления, химического состава жидких веществ.

Традиционно при измерении массы жидких веществ используются процедуры калибровки и градуировки на основе уровнемеров, поплавков, весов. Однако эти процедуры зависят от свойств и характеристик измеряемых жидкостей и не позволяют применить типовые подходы к измерениям. В свою очередь, это приводит к тому, что не удается полностью выработать из резервуаров измеряемые жидкие вещества [3].

Помимо этого, жидкие вещества, масса которых измеряется, зачастую представляют опасность при непосредственном измерении, они являются вредными/опасными, что, в свою очередь, приводит к необходимости автоматизации процессов измерений.

В большинстве измерительных систем применяется расчетный способ [4], особенностью которого является учет при расчете и оценке параметров жидкости (плотности, объема, давления, температуры). Вместе с тем, точность измерений с использованием этого способа ограничена сложностью учета геометрических параметров резервуара и внешних условий проведения измерений.

В настоящей статье предлагается способ измерения массы жидких веществ, особенностями которого являются: во-первых, измерение столба жидкости гидростатического давления; во-вторых, то, что процесс тарирования осуществляется на основе искусственных нейронных сетей, при этом за счет особенностей используемых при тарировании параметров (гидростатического давления жидкости, высоты столба заливаемой жидкости, учета внешних факторов окружающей среды) удается осуществить настройку информационно-измерительной системы и в дальнейшем осуществить измерение различных жидкостей без перенастройки этой системы.

#### **Обоснование выбора датчика давления для измерения массы жидких веществ**

Выбор датчика давления для измерения массы жидкого вещества позволяет осуществить подход к измерению без учета плотности и объема жидкости, существенно зависящих от параметров внешней среды.

Датчик давления может быть выбран по разным критериям, например, в зависимости от размера резервуара, возможного места размещения, химического состава жидкости (агрессивная, неагрессивная среда, пищевая продукция), требуемого диапазона измерения давления.

Существуют различные датчики давления: оптические, волоконно-оптические, оптоэлектронные, магнитные, емкостные, ртутные, пьезоэлектрические, пьезорезонансные, резистивные, гидростатические. Для решаемой задачи выбраны гидростатические датчики давления, характеризующиеся следующими особенностями:

- предназначены для измерения давления;
- характеризуются большим диапазоном измеряемого давления;
- имеют высокую степень защиты от внешних воздействий;
- наличие термокомпенсации;
- инертность материала, из которого изготавливаются датчики, к воздействию веществ различного химического состава;
- удобство представления выходных сигналов (ток, напряжение).

#### **Описание способа измерения массы жидкого вещества на основе искусственных нейронных сетей**

Предлагаемый способ измерения массы жидкого вещества включает в себя следующие

обобщенные этапы:

### **Этап 1.** Тарирование резервуара

Осуществляется определение минимального значения массы, которое соответствует константному (на границе чувствительности) значению для данного датчика. Далее выполняется оценка измерения давления для соответствующего значения массы вещества. Далее осуществляется заполнение резервуара минимальными порциями и фиксация соответствующего значения датчика давления.

Процесс тарирования по каждому значению осуществляется до полного заполнения резервуара. В итоге формируется т.н. тарировочная таблица. Одновременно с этим фиксируется высота «столба» жидкости, температура и влажность окружающей среды, атмосферное давление.

Полученные данные являются исходной информацией для последующего обучения искусственной нейронной сети.

*Примечание.* В качестве жидкости в процессе тарирования может быть использована как жидкость, масса которой в дальнейшем будет измерена, так и, например, вода. Вследствие особенностей предлагаемого способа, тип жидкости никак не влияет на подготовку исходных данных для обучения искусственной нейронной сети.

### **Этап 2.** Обучение искусственной нейронной сети

Осуществляется обучение искусственной нейронной сети по заполненной на предыдущем этапе тарировочной таблице. Следует отметить, что обучение искусственной нейронной сети может осуществляться как по мере формирования каждого обучающего примера из тарировочной таблицы, так и по результатам полного заполнения этой таблицы для всех значений.

На рисунке 1 приведена структура нейросетевой модели.

Входной слой нейронной сети представлен элементами входного вектора:

$$X = \{P, h, \varphi, t, P_{atm}\},$$

где  $P$  – измеренное значение давления (высота резервуара, в котором осуществляется измерение жидкого вещества);

$h$  – значение высоты залитого вещества;

$\varphi$  – значение влажности окружающей среды;

$t$  – температура окружающей среды;

$P_{atm}$  – атмосферное давление.

Выходной слой нейронной сети представлен элементом выходного вектора:

$$Y = \{M_z\},$$

где  $M_z$  – масса измеряемой жидкости.

В качестве алгоритма обучения искусственной нейронной сети целесообразно использовать алгоритм обратного распространения ошибки [6].

### **Этап 3.** Измерение массы жидкого вещества

На данном этапе осуществляется непосредственно измерение массы жидкого вещества. Резервуар до некоторого уровня заполняется жидким веществом. С датчика в измерительную систему на вход искусственной нейронной сети поступает считываемое значение давления, а также значения дополнительных параметров (высота «столба» залитого вещества; влажность

и температура окружающей среды; атмосферное давление).

На выходе искусственной нейронной сети формируется точное или обобщенное значение выходного параметра (массы жидкого вещества). Результат выводится на устройство индикации информационно-измерительной системы, схема которой приведена на рисунке 2.

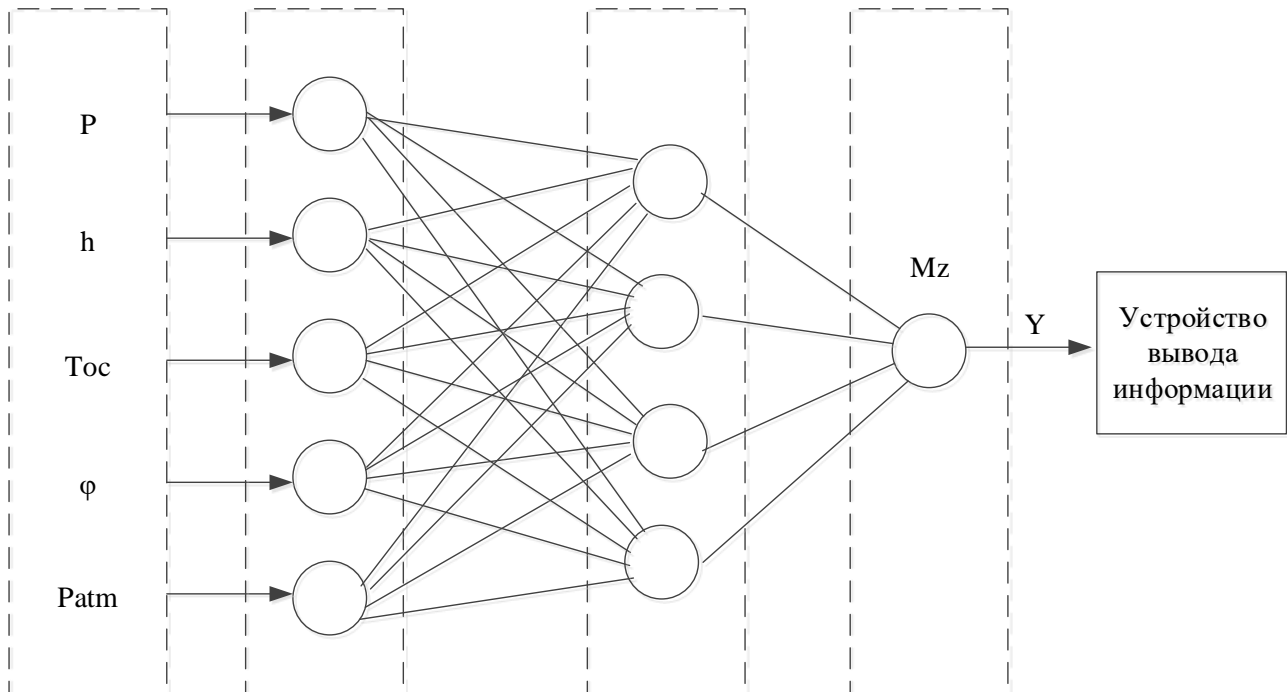


Рисунок 1 – Модель нейронной сети для измерения массы жидкого вещества

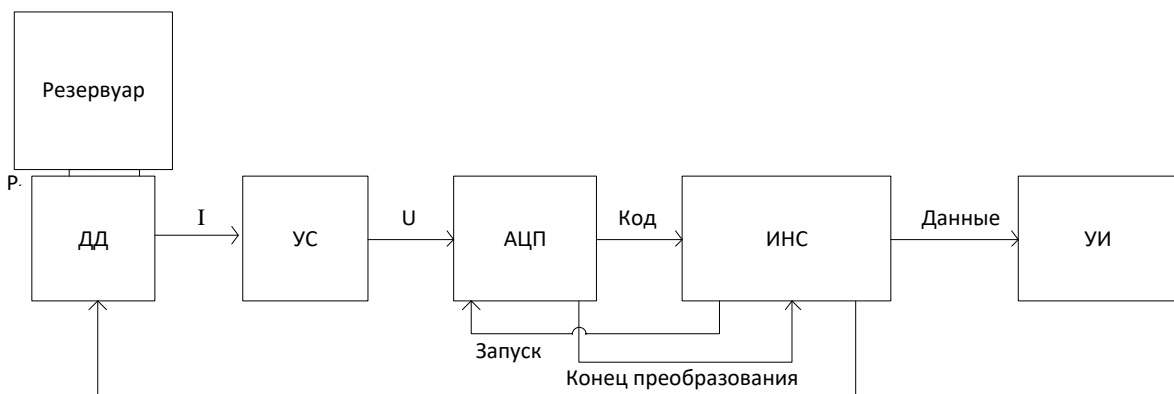


Рисунок 2 – Схема информационно-измерительной системы

На рисунке: ДД – датчик давления; УС – усилитель; АЦП – аналого-цифровой преобразователь; ИНС – контроллер, реализованный на искусственных нейронных сетях; УИ – устройство индикации.

Предлагаемый способ основан на измерении гидростатического давления массы жидких веществ и за счет использования искусственных нейронных сетей позволяет осуществлять измерение различных жидкостей без перенастройки информационно-измерительной системы для выбранного резервуара. Использование этого способа позволяет сократить время настройки системы для другого резервуара, за счет обучения искусственной нейронной сети,

с обеспечением заданной точности измерений и может быть применен для измерения массы разных жидких веществ в различных резервуарах.

### Литература

1. Патент 2 343 426 Российская Федерация. Система для измерения массы жидкого вещества в резервуаре [Текст] / Галкин А.С.; заявитель и патентообладатель Закрытое акционерное общество "НТФ НОВИНТЕХ". - № 2007131221/28; заявл. 16.08.2008; опубл. 10.01.2009. – 7 с.
2. Патент РФ №2497085, 27.01.2010 Способ измерения массы жидкости в резервуаре [текст] / Б.В. Скворцов, С.А. Борминский, А.В. Солнцева, Д.И. Блинов. – Опубл. 27.10.2013 Бюл. №30.
3. Патент 2494353 Российская Федерация. Способы определения массы жидкого вещества в емкости [Текст] / Грузевич Ю.К.; заявитель и патентообладатель Грузевич Ю.К., Пергамент А.М., Тележников В.Н. - № 2012110379/28; заявл. 20.03.2012; опубл. 27.09.2013 Бюл. № 27. – 11 с.
4. Визуальный способ измерения [Электронный ресурс] // Официальный интернет портал [сайт]. URL: <http://www.ngpedia.ru/id148543p2.html>
5. Войцицкая А.С., Борисов В.В. Измерение массы жидкостей различных резервуаров с применением искусственных нейронных сетей [Текст] / XIII Международная научно-техническая конференция студентов и аспирантов: сб. трудов. – Смоленск, 2016. – С. 232-235.
6. Борисов В.В. Искусственные нейронные сети. Теория и практика/ В.В. Круглов, В.В. Борисов – М.: Горячая линия-Телеком, 2002. — 382 с. — ISBN 5-93517-031-0.
7. Войцицкая А.С., Борисов В.В. Измерение массы жидких веществ в различных резервуарах с применением искусственных нейронных сетей [Текст] / XXIII REEPE 2017 Международная научно-техническая конференция студентов и аспирантов: сб. тезисов. – Москва, 2017. – С. 323-324.

### References

1. Patent 2 343 426 Russian Federation. System for measuring weight of the liquid substance in the reservoir [Text] / Galkin, A. S.; applicant and patentee closed joint stock company "NTF NOVINTEKH". No 2007131221/28; Appl. 16.08.2008; publ. 10.01.2009. – 7 S.
  2. RF patent №2497085, 27.01.2010 Method of measuring the mass of liquid in the reservoir [text] / B. V. Skvortsov, S. A. Borminskaya, A. V. Solntsev, D. I. Blinov. – Publ. 27.10.2013 bull. No. 30.
  3. Patent 2494353 Russian Federation. Methods of determining the weight of the liquid substance in the tank [Text] / Gruzevich Y. K.; applicant and patentee of Gruzevich Y. K., A. M. Parchment, Tsyalezchnikau V. N. No 2012110379/28; Appl. 20.03.2012; publ. 27.09.2013 bull. No. 27. – 11 S.
  4. The visual method of measurement [Electronic resource] // Official Internet portal [website]. URL: <http://www.ngpedia.ru/id148543p2.html>
  5. Voitsitska A. S., Borisov V. V. the mass Measurement of liquids in various tanks using artificial neural networks [Text] / XIII international scientific and technical conference of students and postgraduates: collected works. – Smolensk, 2016. – Pp. 232-235.
  6. Borisov V. V. Artificial neural network. Theory and practice/ V. V. Kruglov, V. V. Borisov, M.: Hot line-Telecom, 2002. — 382 p. — ISBN 5-93517-031-0.
  7. Voitsitska A. S., Borisov V. V. the Measurement of the mass of liquid in various tanks using artificial neural networks [Text] / XXIII REEPE 2017 International scientific-technical conference of students and graduate students: collection of abstracts. – Moscow, 2017. – S. 323-324.
-



Международный журнал информационных технологий и энергоэффективности

Сайт журнала:

<http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004.896

## ОБ АСПЕКТАХ РИСКА В ПРОЦЕССЕ СОЗДАНИЯ IT-ПРОЕКТА

**Дорофеева Ю.В.**

Филиал федерального государственного бюджетного образовательного учреждения высшего образования «Национальный исследовательский университет МЭИ» в г. Смоленске, Россия (214013, г. Смоленск, Энергетический проезд, дом 1); e-mail: [ulya.dorofeeva@rambler.ru](mailto:ulya.dorofeeva@rambler.ru)

В статье дается представление о классификации аспектов риска, возникающих в процессе создания IT-проекта, описывается влияние источника риска, риск-события, события в контексте процессной характеристики на структуру бизнес-процесса и на проект. Проводится обзор видов аспектов риска, проявляющихся при разработке проекта, рассматривается бизнес-процесс этапов создания, а также прелиминарные события, предотвращающие или уменьшающие вероятность возникновения риска. Приведены актуальные проблемы, связанные с возникновением рисков при разработке.

Ключевые слова: программное обеспечение, бизнес-процесс, IT-проект, процесс разработки, технические риски, управленческие риски, управление рисками, нотация рисков, источники риска, событие, риск-событие.

## ABOUT ASPECTS OF RISK IN THE PROCESS OF CREATING AN IT PROJECT

**Dorofeeva Y.V.**

The Branch of Federal state budgetary educational institution of higher education "National research University Moscow power engineering Institute" in Smolensk, Russia (214013, Smolensk, Energeticheski proezd, 1); [ulya.dorofeeva@rambler.ru](mailto:ulya.dorofeeva@rambler.ru)

The article gives an idea about the classification of aspects of risk arising in the process of creating an IT project, describes the impact of the source of risk, risk events, in the context of process characteristics on the structure of the business process and the project. Provides an overview of the types of aspects of risk, manifested in the development of the project deals with the business process of creation, as well as preliminary events to prevent or reduce the likelihood of the risk occurring. Given current problems associated with the occurrence of risks during development.

Keywords: software, business process, IT-project, development process, technical risks, management risks, risk management, the notation of risk, sources of risk event, risk event.

Учет и оценка рисков являются необходимой и обязательной частью в успешной деятельности любой компании на рынке, в частности, предприятия, связанного с информационными технологиями. Воспользовавшись определением Ф. Найта установим, что



риск – это ситуация, в которой существует конечное число исходов при известных вероятностях для каждого из них [5]. С точки зрения рыночных отношений, любой риск – это характеристика предполагаемо возможных потерь.

Процесс анализа источников рисков при создании IT-проекта сопровождается использованием нечетких значений в виде оценок экспертов. Для создания эффективной базы знаний о рисках при создании IT-проекта необходимо в достаточной степени классифицировать аспекты потенциально опасной ситуации для однозначной идентификации риска в системе проекта. Входными параметрами на данном этапе исследования являются источники рисков, события, риск-события в формате лингвистических переменных.

#### **Разработка диаграмм согласно графической нотации**

Согласно [6], для определения характера ситуаций, возникающих в условиях риска, будем использовать нотацию рисков. Изучение бизнес-процессов возможно при помощи использования изобразительных средств нотации ARIS [8], BPMN [7], диаграмм деятельности UML [9]. Графическая нотация для представления процесса управления комплексными рисками обладает большей наглядностью, что определяет её использование. Нотация предполагает использование диаграмм шести основных подходов к изучению течения бизнес-процесса и выявления риска в проекте, а также обладает наиболее наглядным способом представления событий в системе проекта:

- 1) Диаграмма процессного аспекта рисков;
- 2) Диаграмма структурного аспекта рисков;
- 3) Диаграмма системного аспекта рисков;
- 4) Диаграмма риск-ситуаций;
- 5) Диаграмма прецедентного управления рисками;
- 6) Диаграмма прелиминарного управления рисками.

Создание проекта в области информационных технологий является совокупностью последовательных процедур с возможностью отсылки к предыдущему этапу. На основании этого утверждения опишем бизнес-процесс этапов IT-проекта, за основу которого возьмем стандарт ГОСТ 34.601-90. Настоящий стандарт ГОСТ 34.601-90 распространяется на автоматизированные системы, используемые в различных видах деятельности – исследование, проектирование, управление, включая их сочетания, создаваемые в организациях, объединениях и на предприятиях [3]. Стандарт предполагает использование основных восьми этапов, включая эксплуатацию. Опираясь на принципы исполнения и сдачи проекта в использование – рассмотрим риски первых семи этапов по отдельности для каждого аспекта. При анализе выяснили, что преобладающим стал процессный подход, в основе которого лежит взаимодействие событий и риск-событий во времени. Для этого сначала выявили источники риска, которые непосредственно влияют на приведение риска в действие, затем изобразили взаимодействие компонентов в пределах одного этапа проектирования.

Диаграммы нотации базируются на основных графических элементах, определяющие компонент системы. [6]

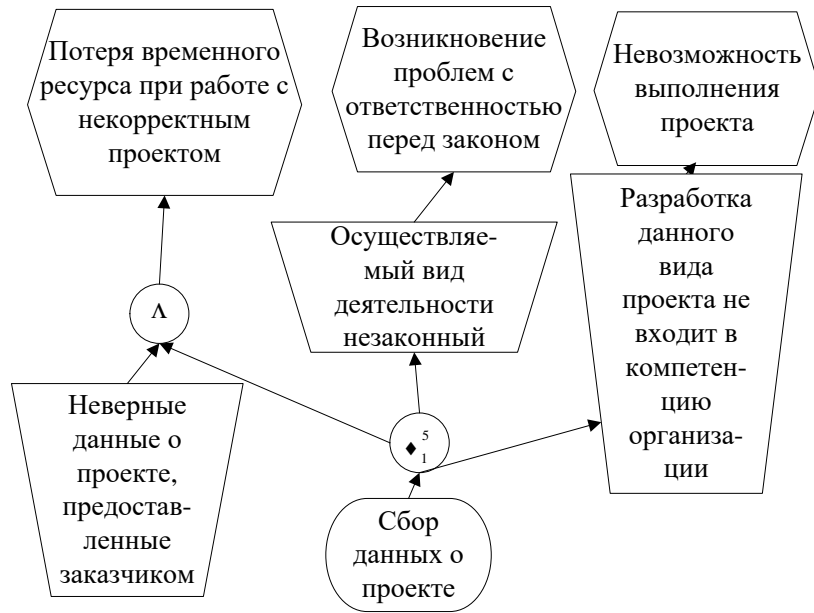


Рисунок 1 – Диаграмма процессного аспекта нотации рисков для этапа 1 согласно стандарту ГОСТ 34.601-90

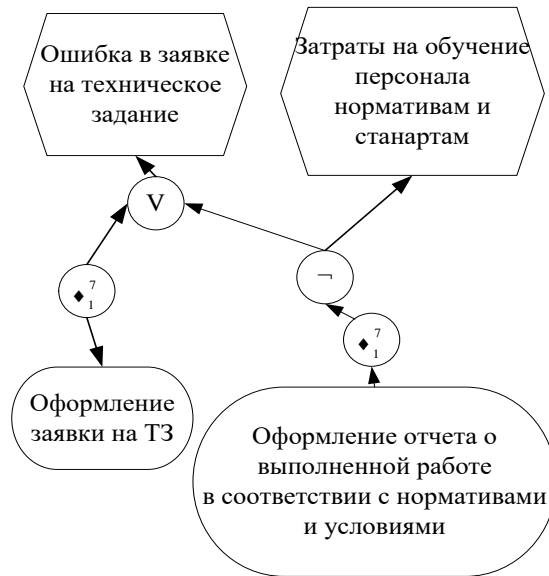


Рисунок 2 – Диаграмма процессного аспекта нотации рисков для этапа 1 согласно стандарту ГОСТ 34.601-90

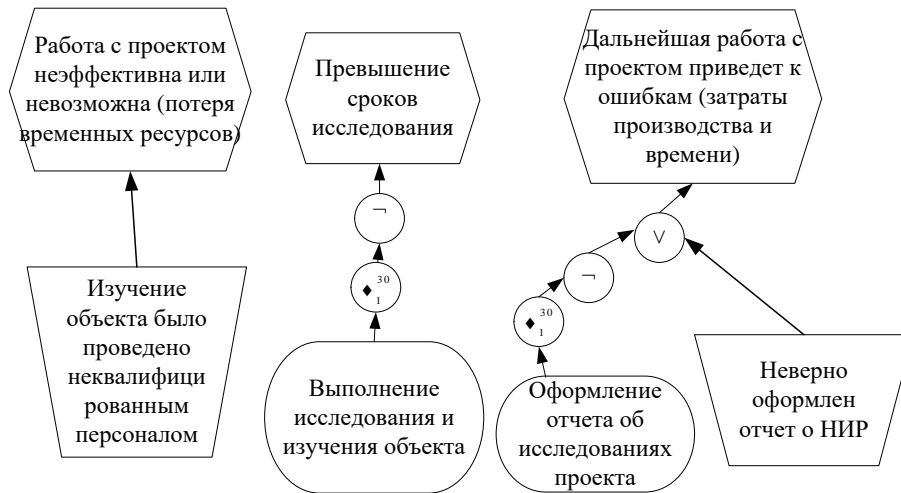


Рисунок 3 – Диаграмма процессного аспекта нотации рисков для этапа 2 согласно стандарту ГОСТ 34.601-90

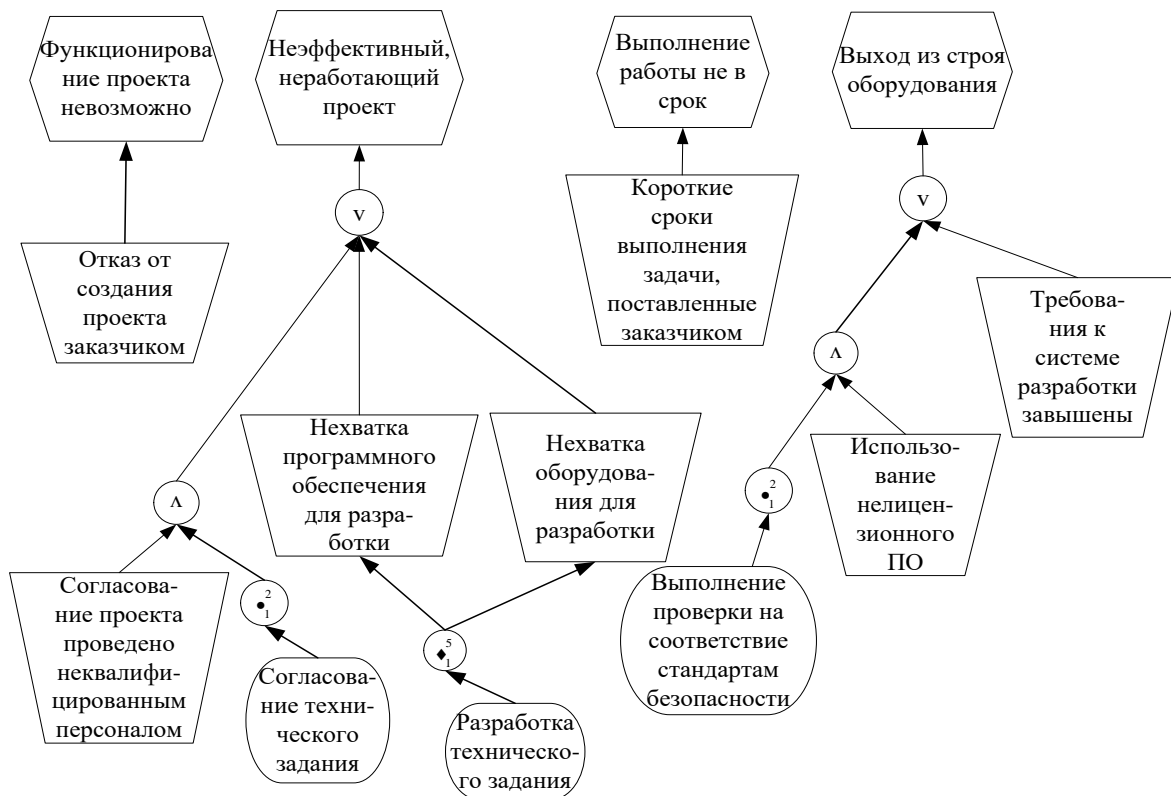


Рисунок 4 – Диаграмма процессного аспекта нотации рисков для этапа 3 согласно стандарту ГОСТ 34.601-90

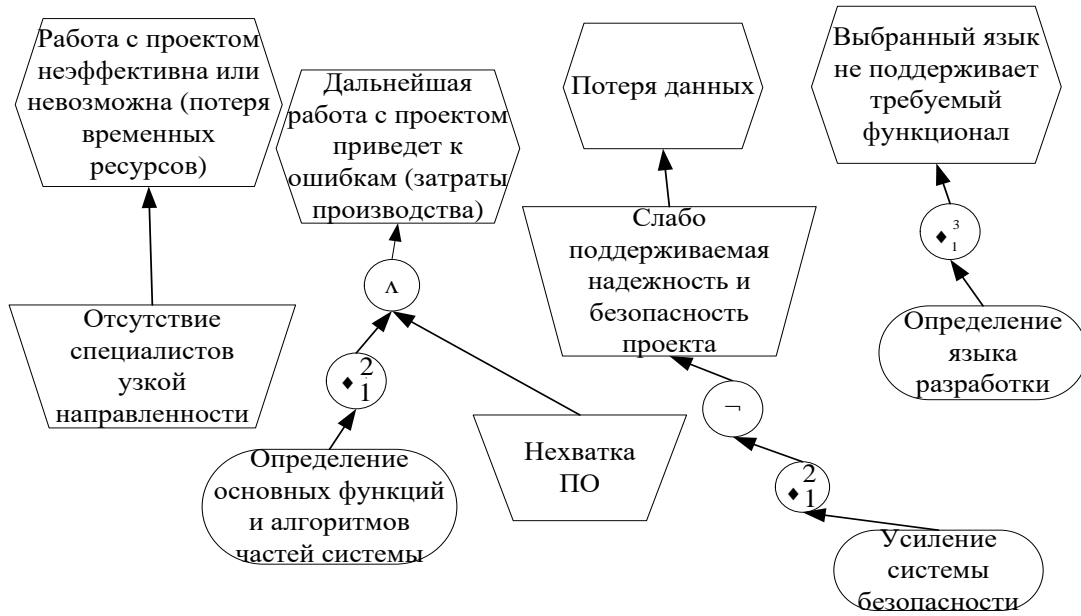


Рисунок 5 – Диаграмма процессного аспекта нотации рисков для этапа 4 согласно стандарту ГОСТ 34.601-90

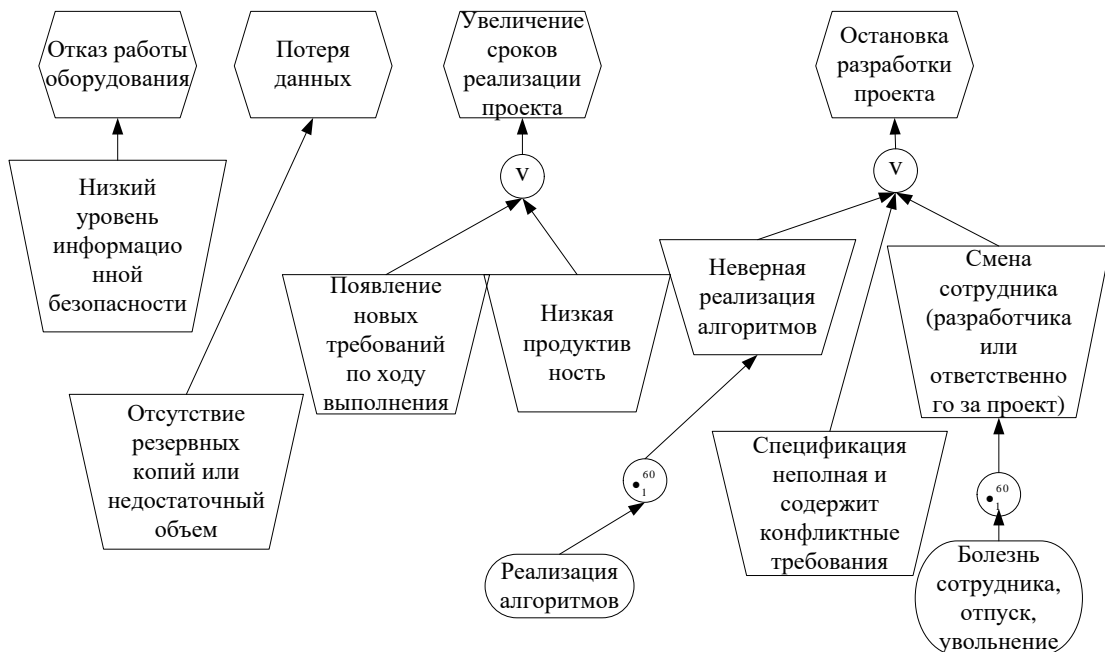


Рисунок 6 – Диаграмма процессного аспекта нотации рисков для этапа 5 согласно стандарту ГОСТ 34.601-90

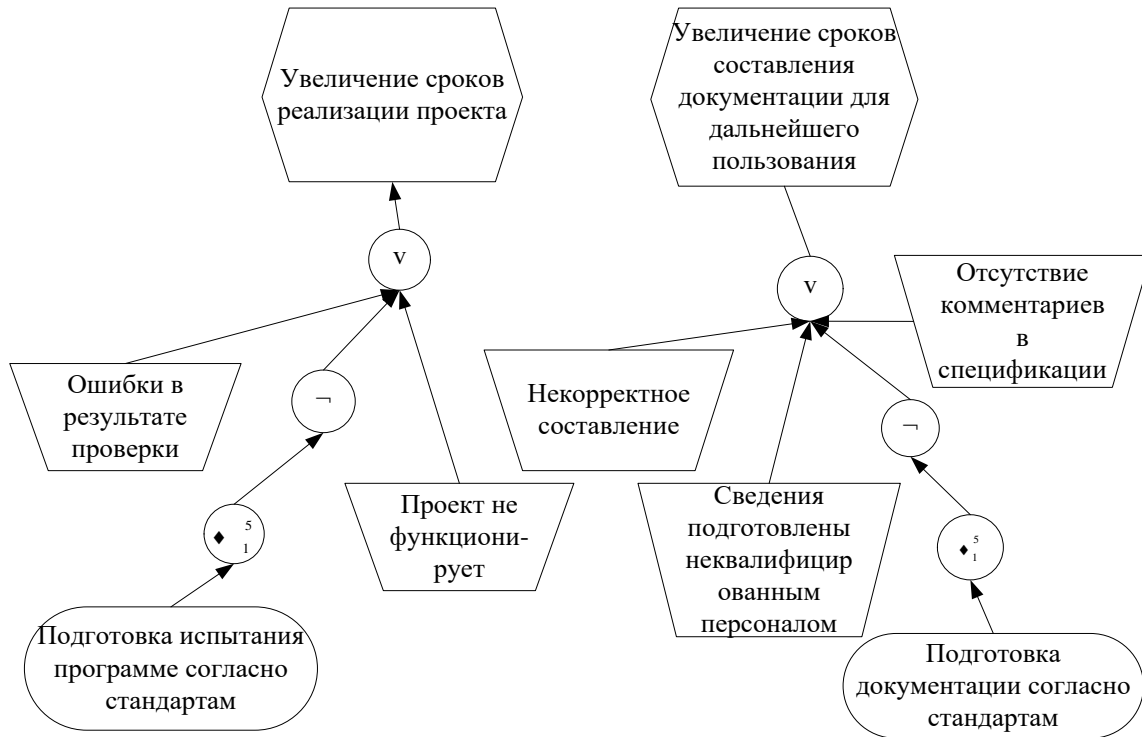


Рисунок 7 – Диаграмма процессного аспекта нотации рисков для этапа 6 согласно стандарту ГОСТ 34.601-90

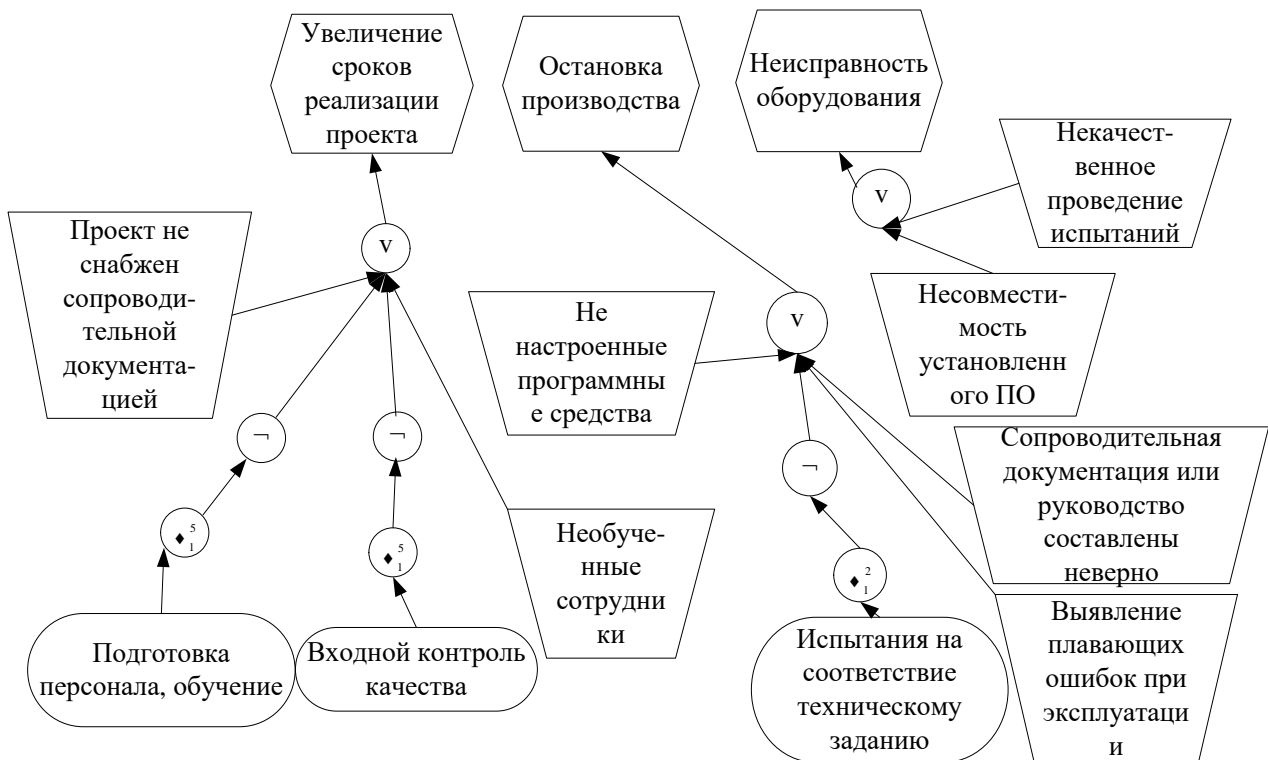


Рисунок 8 – Диаграмма процессного аспекта нотации рисков для этапа 7 согласно стандарту ГОСТ 34.601-90

### Оценка количественной составляющей риска

На основании приведенных графических диаграмм, выясним количественное влияние источников рисков на возникновение риск-событий и составим реестр значений. Оценим подходы к анализу показателя риска.

### Количественная величина вероятности возникновения риска

Величина вероятности возникновения риска связана с осуществлением конкретной угрозы в отношении конкретного события [1], что можно выразить при помощи следующей математической формулы:

$$Q_{sr} = P_s * S_d,$$

где  $Q_{sr}$  - величина риска;

$P_s$  - вероятность риск-события,

$S_d$  – размер ущерба.

В свою очередь,

$$P_s = P_{sor} * P_v,$$

где  $P_{sor}$  - вероятность возникновения источника риска;

$P_v$  - величина уязвимости.

*Вероятность риск-события* - вероятность успешной реализации источника риска с использованием уязвимости и причинения ущерба организации. *Вероятность возникновения источника риска* - вероятность того, что угроза (источник риска) в отношении этапа IT-проекта будет реализовываться (успешность либо неуспешность реализации угрозы определяется величиной уязвимости). *Величина уязвимости* - вероятность того, что в случае реализации угрозы в отношении процесса на этапе IT-проекта, эта угроза будет реализована успешно с использованием данной уязвимости, то есть безопасность в результате будет нарушена и организация понесет определенный ущерб.

Для определения величины риска используются оценочные количественные значения, полученные путем экспертных оценок, прогнозирования, а также на основании статистических данных, исходя из конкретного проекта.

Размер ущерба, как правило, выражается в денежных единицах, величина уязвимости принимает значения в диапазоне от 0 до 1, а вероятность возникновения источника риска является целым положительным числом, определяющим ожидаемое количество попыток реализации угрозы за определенный период времени.

### Статистический метод оценки риска согласно [2]

Одним из наиболее распространенных подходов к количественной оценке риска является использование выражения

$$R = H_{\pi} * P,$$

где  $H_{\pi}$  - величина потерь;

$p$  - вероятность наступления рискового события.

Таким образом, степень риска определяется как произведение ожидаемого ущерба на вероятность того, что такой ущерб будет нанесен. В качестве критерия при количественной оценке риска проектами вложения капитала широко используются два показателя:

- среднее ожидаемое значение ( $\bar{X}$ ) возможного результата (отдачи), которое является средневзвешенным для всех возможных результатов, где вероятность каждого результата используется в качестве частоты или веса соответствующего значения;

- среднее квадратическое отклонение ( $\sigma$ ) как мера изменчивости (колеблемости) возможного результата. В качестве отдачи могут выступать, например, доходы, прибыль, дивиденды.

На основании разработанных процессных диаграмм можно видеть ключевые темпоральные операторы прошлого времени. Это «когда-либо в прошлом» ( $\diamond$ ), «всегда в прошлом» ( $\bullet$ ), «когда-либо в прошлом на интервале  $n, \dots, m$ » ( $\diamond_n^m$ ), «всегда в прошлом на интервале  $n, \dots, m$ » ( $\bullet_n^m$ ) [3]. Неотъемлемой частью диаграмм нотации процессного типа является временной аспект, поэтому, исходя из приведенных выше методик оценки, необходимо ввести дополнительный параметр  $T$  - оценка времени возникновения. Существующая зависимость между временем, которое имеет в распоряжении управленческий персонал, и риском требует от менеджеров проекта держать этот аспект управленческой деятельности под постоянным контролем. И если в процессе управленческих воздействий на объект в ситуации риска удастся прийти к разрешению неопределенности, то опоздание в реализации приведет к нарастанию неопределенности в деятельности фирмы, компании.

Изобразим зависимость вероятности появления риска  $P$  от времени графически для трех ситуаций - «когда-либо в прошлом на интервале  $n, \dots, m$ » ( $\diamond_n^m$ ), «всегда в прошлом на интервале  $n, \dots, m$ » ( $\bullet_n^m$ ), и в случае, если время действия события прошло, событие не было приведено в действие.

1. Когда-либо в прошлом на интервале  $n, \dots, m$  ( $\diamond_n^m$ )

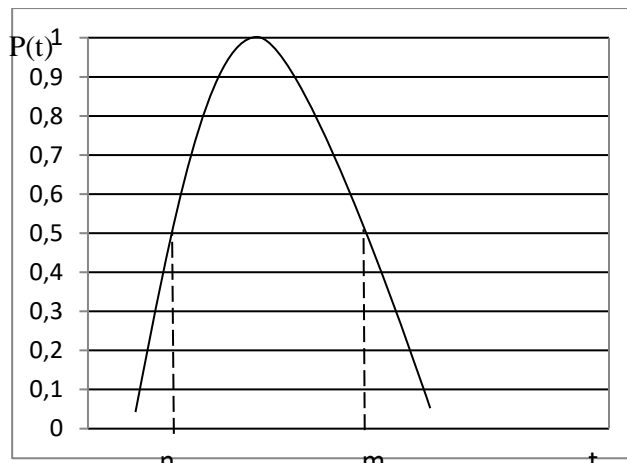


Рисунок 9 – График функции для оператора  $\diamond_n^m$

2. Всегда в прошлом на интервале  $n, \dots, m$  ( $\bullet_n^m$ )

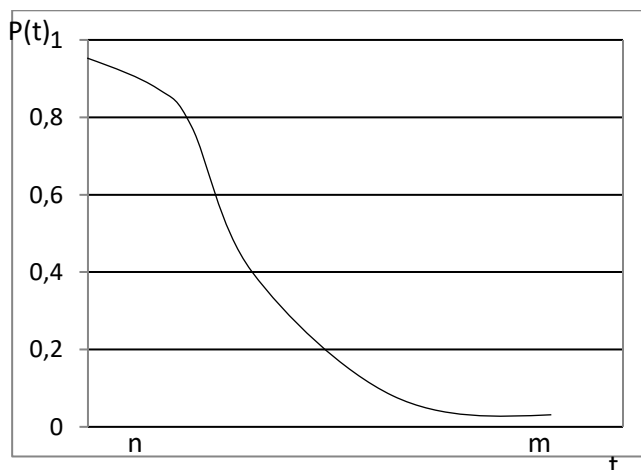


Рисунок 10 – График функции для оператора  $\bullet_n^m$

3. В случае если время действия события прошло, событие не было приведено в действие

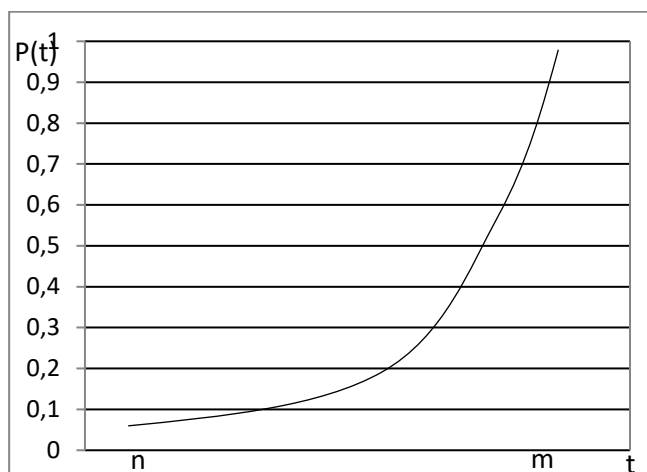


Рисунок 11 - График функции для оператора с отрицанием

Составим таблицу, где опишем те источники риска, которые можно предотвратить или свести к минимуму при введении прелиминарных событий, примем максимальное значение уязвимости, а затем будем вводить события, уменьшающие риск.

Таблица 1 – Источники рисков, напрямую влияющие на исполнение риск-события

Источник риска	Уязвимость	Риск	Последствия	Событие, уменьшающее возникновение риска
Неверные данные о проекте, предоставленные заказчиком	Некомпетентность заказчика	Потеря временного ресурса при работе с некорректным проектом	Финансовые затраты за потраченные трудочасы	Работа с заказчиком, являющимся профессионалом своей области.
Изучение объекта было проведено неквалифицированным персоналом	Некомпетентность персонала	Работа с проектом неэффективна или невозможна (потеря временных ресурсов = трудочасов)	Финансовые затраты за потраченные трудочасы	Обучение персонала
Согласование проведено неквалифицированным персоналом	Некомпетентность персонала	Работа с проектом неэффективна или невозможна (потеря временных ресурсов = трудочасов)	Финансовые затраты за потраченные трудочасы	Обучение персонала
Отказ от проекта/ создания проекта заказчиком	Человеческий фактор со стороны заказчика	Потеря временных ресурсов = трудочасов	Финансовые затраты за потраченные трудочасы	Составление акта-договора страхования от случаев отказа по вине заказчика
Нехватка программного обеспечения для разработки	Ограниченные средства ПО	Отказ оборудования	Финансовые затраты по замене нового оборудования	Закупка необходимого ПО



Окончание таблицы 1

Источник риска	Уязвимость	Риск	Последствия	Событие, уменьшающее возникновение риска
Нехватка оборудования для разработки	Ограниченные средства АО	Отказ оборудования	Финансовые затраты по замене нового оборудования	Закупка необходимого АО
Использование нелегального ПО	Нелегальное ПО	Отказ оборудования	Финансовые затраты по замене нового оборудования	Закупка лицензии
Требования к системе разработки завышены	Завышенные требования к системе	Отказ оборудования	Финансовые затраты по замене нового оборудования	Перейти к этапу согласования требований, согласно системе
Отсутствие специалистов узкой направленности для выполнения задач проекта	Недостаток кадров	Выполнение проекта не в срок,	Санкции, пени – финансовые затраты	Использование услуг сторонних специалистов, обучение персонала требуемым навыкам
Слабо поддерживаемая надежность и безопасность проекта	Отсутствие защиты информации	Потеря данных	Потеря информации	Установка утилит по защите данных
Отсутствие резервных копий или недостаточный объем	Резервные копии не созданы	Потеря данных	Потеря информации	Своевременное копирование, создание резервных данных
Ошибки в результате проверки тестирования	Недостаточное тестирование	Выполнение проекта не в срок со стороны разработчика	Санкции, пени – финансовые затраты	Использование дополнительных ресурсов по привлечению тестировщиков
Некачественное выполнение сопроводительной документации	Некачественная документация	Выполнение проекта не в срок со стороны разработчика, ошибки при эксплуатации	Санкции, пени – финансовые затраты. Затраты при неправильной работе ПО	Обучение сотрудников нормам и стандартам составления документов
Проект не снабжен сопроводительной документацией	Отсутствие документации	Выполнение проекта не в срок со стороны разработчика, ошибки при эксплуатации	Затраты при неправильной работе ПО	Обучение сотрудников нормам и стандартам составления документов
Необученные сотрудники	Некомпетентность сотрудников	Неэффективное использование проекта, потери трудочасов	Финансовые затраты за потраченные трудочасы	Обучение персонала
Несовместимость установленного ПО с проектом	Неустановленное ПО, нелегальное ПО при вводе в эксплуатацию	Отказ оборудования, остановка производства	Финансовые затраты по замене нового оборудования	Закупка и установка совместимого ПО

Таким образом, проведен анализ бизнес-процесса, в ходе которого выявлены источники рисков, риски, а также выявлены предварительные события, при совершении которых вероятность возникновения риска снижается или сводится к нулю. Представленные результаты использованы при разработке системы управления рисками бизнес-процесса IT-проекта.

### Список литературы

1. Астахов А.М. Искусство управления информационными рисками – М.: ДМК Пресс, 2010
2. Гранатуров В.М. Экономический риск: сущность, методы измерения, пути снижения М.: "Дело и Сервис", 2002
3. ГОСТ 34.601-90 Автоматизированные системы. Стадии создания.
4. Еремеев А.П. Логика ветвящегося времени и ее применение в интеллектуальных системах поддержки принятия решений // Сб. тр. 10-й Нац. конф. по искусственному интеллекту с междунар. уч. КИИ-2006. В 3-х т. Т.3. – М.: Физматлит, 2006. – С.746–754.
5. Найт Ф. Риск, неопределенность и прибыль. М. «Дело», 2003
6. Сеньков А.В. Графическая нотация для представления процесса управления комплексными рисками: Современные наукоемкие технологии. – 2016. – № 12 (часть 1) – С. 72-81.
7. Object Management Group. Business Process Model and Notation v.2.0. Режим доступа: <http://www.omg.org/spec/BPMN/2.0/PDF>, accessed 01.10.2016.
8. Шеер А.-В. Моделирование бизнес- процессов. — М.: Весть-МетаТехнология, 2000.
9. Bruce Powel Douglass Real Time UML : Advances in the UML for Real-Time Systems (3rd Edition) (Addison-Wesley Object Technology Series); Издательство деловой и учебной литературы - Москва, 2010. - 836 с.

### References

1. Astahov A.M. Iskusstvo upravleniya informacionnymi riskami – M.: DMK Press, 2010
  2. Granaturov V.M. EHkonomicheskij risk: sushchnost', metody izmereniya, puti snizheniya M.: "Delo i Servis", 2002
  3. GOST 34.601-90 Avtomatizirovannye sistemy. Stadii sozdaniya.
  4. Eremeev A.P. Logika vetvyashchegosya vremeni i ee primeneniye v intellektual'nyh sistemah podderzhki prinyatiya reshenij // Sb. tr. 10-j Nac. konf. po iskusstvennomu intellektu s mezhdunar. uch. KII-2006. V 3-h t. T.3. – M.: Fizmatlit, 2006. – S.746–754.
  5. Najt F. Risk, neopredelennost' i pribyl'. M. «Delo», 2003
  6. Sen'kov A.V. Graficheskaya notaciya dlya predstavleniya processa upravleniya kompleksnymi riskami: Sovremennyye naukoemkie tekhnologii. – 2016. – № 12 (chast' 1) – S. 72-81.
  7. Object Management Group. Business Process Model and Notation v.2.0. Rezhim dostupa: <http://www.omg.org/spec/BPMN/2.0/PDF>, accessed 01.10.2016.
  8. SHeer A.-V. Modelirovanie biznes- processov. — M.: Vest'-MetaTekhnologiya, 2000.
  9. Bruce Powel Douglass Real Time UML : Advances in the UML for Real-Time Systems (3rd Edition) (Addison-Wesley Object Technology Series); Izdatel'stvo delovoj i uchebnoj literatury - Moskva, 2010. - 836 s.
-



Международный журнал информационных технологий и энергоэффективности

Сайт журнала: <http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004.023

## СПОСОБ ОТОБРАЖЕНИЯ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ НА СТРУКТУРУ ГЕТЕРОГЕННОЙ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЫ

**Воевода П.Е., Зернов М.М.**

*Филиал федерального государственного бюджетного образовательного учреждения высшего образования «Национальный исследовательский университет МЭИ» в г. Смоленске, Россия (214013, г. Смоленск, Энергетический проезд, дом 1); pavel-voevoda@mail.ru*

**Рассмотрены основные модели параллельных генетических алгоритмов: «клиент-серверная», «модель островов» и асинхронный генетический алгоритм. Реализован параллельный вариант алгоритма на основе модели «клиент-сервер» в существующей библиотеке классов. Библиотека адаптирована для вычислений на гибридном вычислительном кластере СФ МЭИ.**

Ключевые слова: параллельный генетический алгоритм, параллельные модели генетических алгоритмов, вычислительный кластер.

## THE MAPPING OF GENETIC ALGORITHMS ON THE STRUCTURE OF HETEROGENEOUS COMPUTING SYSTEM

**Voevoda P.E., Zernov M.M.**

*The Smolensk branch of the Moscow Power Engineering Institute (National Research University), Russia (214013, Smolensk, Energetichesky proezd, 1); pavel-voevoda@mail.ru*

**Had been considered main models of parallel GAs, including master-slave, multiple-population coarse-grained GAs and asynchronous GA. Had been implemented parallel version of GA using master-slave model in existing GA library. It may be used for calculating on computer cluster in Smolensk Branch of MPEI.**

Key words: parallel genetic algorithm, models of parallel GAs, computer cluster.

Рассмотрим основные способы параллельной реализации генетических алгоритмов. Из наиболее распространенных можно выделить модель «клиент-сервер», «модель островов» и асинхронные генетические алгоритмы.

Наиболее простой в реализации вариант - это модель «клиент-сервер». В классическом варианте данная модель предполагает, что есть основной цикл построения популяции, который соответствует процессору (поток) «серверу». Все остальные процессоры (поток) являются клиентскими и реализуют вычисления.

Данная модель требует наименьших изменений в существующей версии программного обеспечения, реализующего последовательный ГА и дает неплохие результаты.

При этом затраты по вычислению значений фитнес-функций равномерно распределяются по всем процессорам, для которых используется одна и та же фитнес-функция. Поэтому для  $n$  особей и  $P$  (одинаковых) процессоров мы каждому процессору относим  $\frac{n}{P}$  особей. Значения фитнес-функции вычисляются соответствующими (рабочими) процессорами и посылаются в один процессор (хозяин), который собирает всю информацию, обрабатывает и передает ее снова рабочим процессорам. Процессор «хозяин» имеет информацию о значениях фитнес-функции для всех особей и может генерировать следующее поколение на этой основе.

В конце каждого из этапов помещаются точки синхронизации. Когда процессор-хозяин достигает эти точки, он переходит в режим ожидания, пока все рабочие процессоры не закончат свои задания, что гарантирует глобальную корректность алгоритма. При этом работа между процессором-хозяином и рабочими распределяется следующим образом.

Каждый «рабочий» принимает задание от «хозяина» и определяет значение фитнес-функции для особей, полученный результат посылает хозяину и ожидает следующего задания. Поскольку размер популяции много больше числа процессоров, достигается хороший баланс в загрузке процессоров.

При использовании модели «клиент-сервер» окончательные результаты (поиска решения данной задачи) близки к тем, что получены на однопроцессорной компьютерной системе с использованием аналогичного алгоритма. Качество решения при этом не теряется и в большинстве случаев несколько улучшается, а время его поиска существенно сокращается. В целом данная модель позволяет быстро (с минимальными модификациями) выполнить параллелизацию ГА и дает, прежде всего, ускорение процесса поиска решения. Данную модель не сложно реализовать в локальной сети с использованием технологии сокетов [1].

«Модель островов» используется в распределенных генетических алгоритмах, где каждая подпопуляция развивается на своем "острове". Между островами производится (достаточно редко) обмен лучшими особями. Эта модель может быть реализована в распределенной памяти компьютерной системы, имеющей MIMD-архитектуру. Преимущество распределенных генетических алгоритмов в том, что они работают быстрее даже на однопроцессорных компьютерных системах вследствие лучшей структуризации. Причина заключается в том, что число вычислений сокращается благодаря распределению поиска в различных областях пространства решений [4].

В целом распределенные параллельные генетические алгоритмы более целесообразно использовать для повышения качества решения, в том случае, если его не удастся достичь с помощью обычного (последовательного) генетического алгоритма. При этом мощность каждой подпопуляции должна быть достаточно большой, чтобы не происходило "вырождение особей", которое часто ведет к преждевременной сходимости к локальному экстремуму. Увеличение быстродействия по сравнению с последовательным генетическим алгоритмом также имеет место, но может быть меньше, чем у модели «клиент-сервер». Для повышения быстродействия в этом случае мощность подпопуляций должна быть небольшой, что существенно увеличивает быстродействие (ценой потери качества решения).

Асинхронный генетический алгоритм не имеет явного разбиения эволюции на поколения, а процесс развития реализуется "по событию" в виде непрерывного потока событий отбора, кроссинговера, мутации и т.п. Здесь вновь порожденные потомки сразу замещают (в некотором смысле худшие) особи (не дожидаясь остальных родительских пар своего поколения). Следует

отметить, что такой метод, как правило, проще в реализации (особенно аппаратной) и имеет некоторые преимущества [1].

Как отмечалось, наиболее простой способ создания параллельной модели на основе уже существующей - использование модели «клиент-сервер». Здесь нет необходимости постоянно проводить контроль динамики популяции, а результат параллельной реализации будет наиболее близок к последовательной версии генетического алгоритма. Таким образом, модель будет показывать приблизительно одинаковые результаты независимо от платформы и типа ускорения. В данном случае необходимо создать планировщик задач, который будет распределять вычисления одного этапа по потокам или процессорам. Рассмотрим более подробно реализацию алгоритма на основе данной модели.

### **Отображение структуры генетического алгоритма на вычислительную систему на основе модели «клиент-сервер»**

Рассмотрим наиболее простой способ реализации параллельного генетического алгоритма, при котором параллелизм будет использоваться за счет разделения этапов на несколько процессов. На каждом этапе получения потомков с помощью разделения на  $m/2$  потоков (где  $m$  – число потомков в поколении, причем считаем, что в создании потомка участвует 2 родителя) основного процесса выполнения. Каждый поток проводит скрещивание двух особей, затем при вычислении функции приспособленности каждый узел вычислительной системы отыскивает значения  $F_i$  (функция приспособленности на  $i$ -м этапе) для порождённых на нём особей.

В таком случае можно осуществлять скрещивание и вычисление значений функции приспособленности, следовательно, и целевой функции, одновременно на нескольких потоках (вычислительных узлах). При этом необходимо выделить отдельный поток для управления работой алгоритма и распределением задач и шагов алгоритма.

Схема такого варианта параллельного варианта ГА представлена на рисунке 1.

Выполнение генетических операторов и вычисление целевой функции происходит в клиентских потоках. По достижении условия сходимости или заданного количества итераций в серверной части осуществляется выбор решения с наибольшим значением функции приспособленности, что соответствует наименьшему значению ЦФ.

В качестве основы воспользуемся ранее разработанной библиотекой для работы с генетическими алгоритмами. В ее основе реализована последовательная версия генетического алгоритма. В дальнейшем, была реализована возможность параллельного выполнения команд с использованием OpenMP. Для возможности вычислений с использованием графического процессора были написаны реализации отдельных операторов с использованием OpenCL [3].

При реализации модели «клиент-сервер» и доработке существующей библиотеки необходимо вынести параллельную обработку на уровень выше.

Для проектирования приложения, использующего параллельную обработку данных применяются соответствующие паттерны проектирования. Это обусловлено сложностью распределения команд по потокам и их отделением друг от друга. Для управления всеми потоками должен быть специальный планировщик. Этими задачами занимается паттерн проектирования «активный объект» (ActiveObject), поэтому разумно использовать его базовую версию для адаптации существующей библиотеки под многопоточную работу.

Основная задача - возможность параллельно выполнять генетические операторы. Каждый оператор представлен уже существующим классом, являющимся наследником COperator, причем модификаций этого класса почти не планируется.

Структура данного паттерна после реализации выглядит следующим образом (рисунок 2).

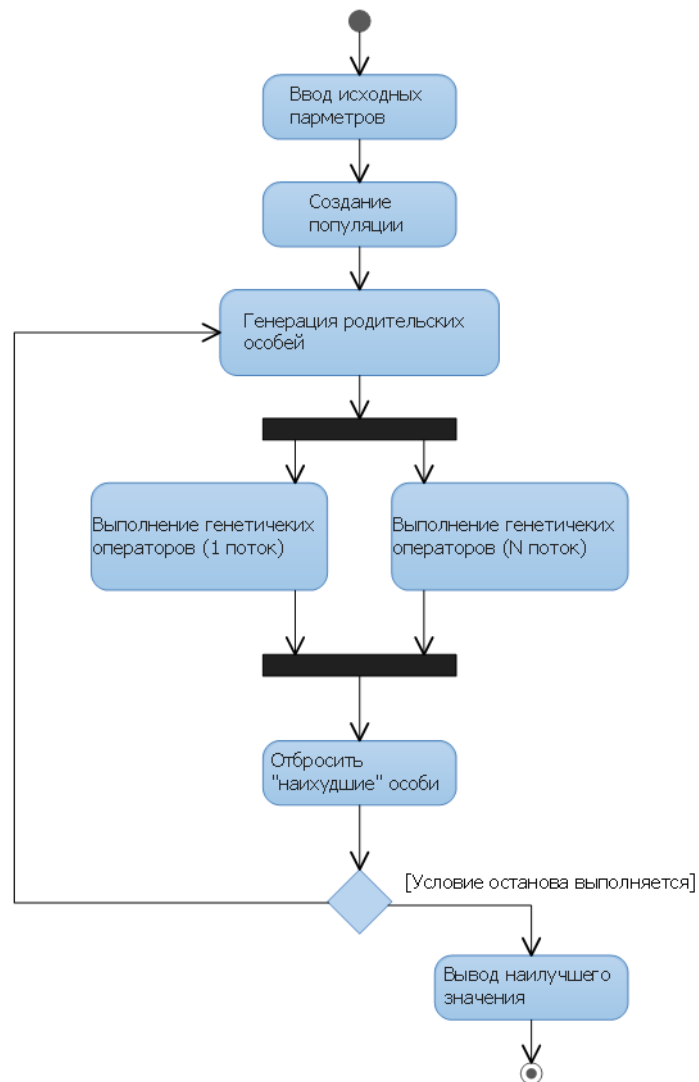


Рисунок 1 - Схема параллельного генетического алгоритма на основе модели «клиент-сервер»

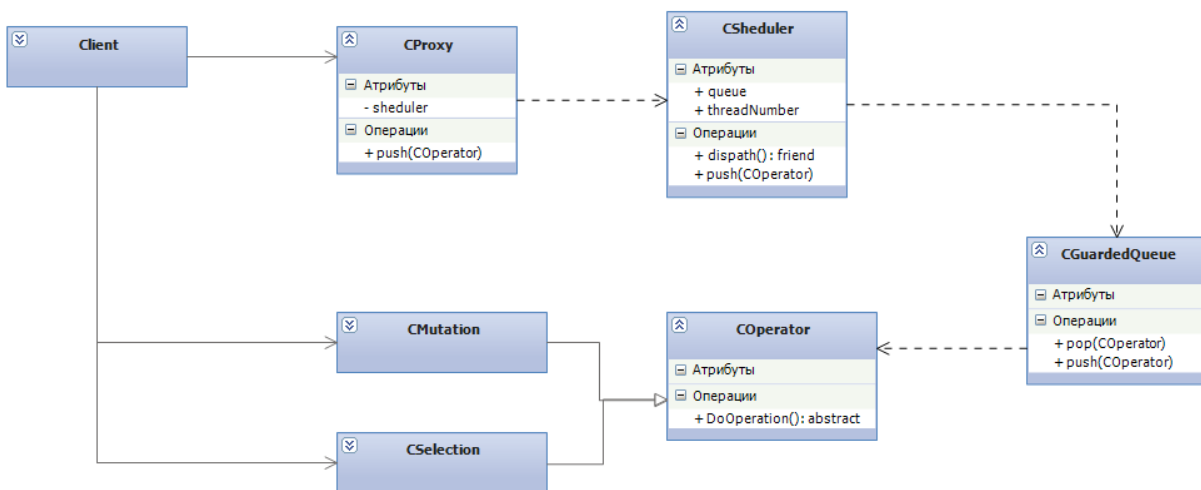


Рисунок 2 - Диаграмма классов для паттерна ActiveObject

Клиенту доступны классы-наследники класса `COperator`, которые переопределяют виртуальный метод `DoOperation()`, а также `CProхy`, с помощью которого будет происходить взаимодействие с создаваемой системой.

Класс `GuardedQueue` – это потоко-защищённая очередь, которая реализует методы `push` и `pop`. Очередь оперирует объектами подклассов класса `COperator`.

Основным элементом паттерна `ActiveObject` является класс `Sheduler` или «планировщик». Именно логика его построения определяет эффективность всей системы.

В конструкторе создается определенное количество потоков, каждый из которых не заканчивает свою работу до завершения работы приложения.

В каждом из созданных потоков запускается функция `dispatch()`, которая в бесконечном цикле вытаскивает верхний элемент из очереди задач (генетических операторов) и запускает его метод `DoOperation ()`.

Эффективность работы паттерна определяется тем, что для каждой задачи не приходится создавать свой поток (процесс создания потока, как известно, занимает продолжительное время), а также тем, что потоки «не простаивают», если есть задачи в очереди.

### **Применение реализации генетических алгоритмов на вычислительном кластере**

Разработка параллельной версии генетического алгоритма производится с учетом особенностей и возможностей гибридного вычислительного кластера филиала НИУ МЭИ в г. Смоленске и адаптирована для использования на нем. Кластер состоит из двух вычислительных узлов, узла управления, необходимых интерфейсных (связных) компонентов.

Каждый вычислительный узел (ВУ) включает в себя два 10-ядерных процессора, оперативную память объемом 64 Гбайт, два твердотельных накопителя SSD с интерфейсом SATA III, организованных в массиве RAID 1, объемом 240 Гбайт, а также по два интерфейсных канала, обеспечивающих подключение вычислительных узлов к внутренней сети кластера.

Кроме того, ВУ 1 содержит две видеокарты на базе GPU NVIDIA, а также два дисковых накопителя HDD объёмом 1 Тбайт с интерфейсом SATA III, организованных в виде массива RAID 1. Вычислительный узел 2 содержит две видеокарты на базе GPU AMD.

Узел управления содержит два 8-ядерных процессора Intel Xeon E5-2620 v4, оперативную память объемом 64 Гбайт стандарта DDR4, девять дисковых накопителей: SSD-накопитель объёмом 256 Гбайт с интерфейсом PCIe 3.0×4 в стандарте M.2; два SSD-накопителя объёмом 250 Гбайт с интерфейсом SATA III в массиве RAID 1; два HDD-накопителя объёмом 2 Тбайт с интерфейсом SATA III в массиве RAID 1; четыре HDD-накопителя объёмом 3 Тбайт с интерфейсом SATA III, размещенных в массиве RAID 10 с объёмом 6 Тбайт. Наличие нескольких дисковых накопителей позволяет, в том числе, создавать на управляющем узле требуемую систему виртуальных машин [2].

Обработка данных может осуществляться посредством как центральных процессоров на основе технологий MPI и Open MP, так и графических на основе технологий OpenCL и CUDA.

После реализации библиотеки для работы с генетическими алгоритмами с применением технологий OpenMP и OpenCL [3] для локального распараллеливания отдельных этапов алгоритма в ней был реализован параллельный генетический алгоритм на основе модели «клиент-сервер». В качестве технологий параллелизма для реализации модели «клиент-сервер» были использованы POSIX threads и Open MPI. Основным этапом перед использованием данной библиотеки в вычислительном эксперименте является ее сборка.

Для сборки библиотеки используется Makefile, представленный в листинге. Данный Makefile адаптирован под сборку проекта с использованием OpenCL и OpenMPI. Файл помещается в папку с исходниками библиотеки, а сборка осуществляется выполнением команды *make*.

#### Листинг1 - Текст Makefile

```
TARGET=$(shell basename `pwd`)
SOURCES=$(wildcard *.cpp)
OBJECTS=$(SOURCES:%.cpp=%.o)

CXXFLAGS=-std=gnu++11
# Check for 32-bit vs 64-bit
PROC_TYPE = $(strip $(shell uname -m | grep 64))

# Linux OS
LIBS= -I/usr/local/cuda-8.0/targets/x86_64-linux/include/ -lOpenCL
CFLAGS+= -I/usr/local/cuda-8.0/targets/x86_64-linux/include
LDFLAGS += -L/usr/local/cuda-8.0/targets/x86_64-linux/lib/x86
LDLIBS += -lOpenCL
ifeq ($(PROC_TYPE),)
    CFLAGS+=-m32
else
    CFLAGS+=-m64
endif

all: $(TARGET)

$(OBJECTS): $(SOURCES)
$(TARGET): $(OBJECTS)
    $(CXX) -o $(TARGET) $(LDFLAGS) $(OBJECTS) $(LOADLIBES) $(LDLIBS) $(CFLAGS)

clean:
    $(RM) $(OBJECTS) $(TARGET)

.PHONY: all clean
```

Данный файл определяет следующую конфигурацию:

- в качестве целевой используется текущая папка;
- исходные файлы в данной папке имеют расширение сpp;
- компилятор использует поддержку стандарта C++11;
- заданы пути для использования и поддержки OpenCL;
- заданы цели для сборки проекта.

В результате выполнения будет собрана библиотека, которую можно подключать к пользовательским проектам в качестве статической библиотеки.

В статье предложен способ отображения генетического алгоритма на структуру параллельной гетерогенной вычислительной системы, отличающийся сочетанием модели параллельной реализации генетического алгоритма «клиент-сервер» с использованием шаблона объектно-ориентированного проектирования «активный объект» и разработанной ранее модели параллельного выполнения генетических операторов с использованием технологии OpenCL. Предложенный способ позволяет одновременно задействовать в ходе работы генетического алгоритма как технологии, задействующие многопроцессорные вычисления (POSIX threads и OpenMPI) и графические вычисления (OpenCL), получая реализацию генетического алгоритма результат которого соответствует последовательной версии с точностью до реализации применяемых случайных последовательностей.

Способ предусматривает наличие планировщика потоков на основе паттерна ActiveObject, управляющего общим ходом вычислений и отслеживающего динамику популяции, управляемых потоков соответствующих отобраным парам родительских особей работающих на основе



технологий POSIX threads и OpenMPI, а также параллельную реализацию самих генетических операторов на основе SIMD-модели на базе технологий OpenMP или OpenCL.

Способ нашёл отражение в разработанной библиотеке классов для реализации генетических алгоритмов, которая была собрана для дальнейших исследований на гибридном вычислительном кластере СФ МЭИ и ориентирована на его структуру.

### Список литературы

1. Скобцов Ю.А., Сперанский Д.В. Эволюционные вычисления: учебное пособие.-М.: Национальный Открытый Университет “ИНТУИТ”, 2015. –331с
2. Борисов В. В., Зернов М. М., Федулов А. С., Якушевский К. А. Исследование характеристик гибридного вычислительного кластера // Системы управления, связи и безопасности. 2016. №4. С. 129-146.
3. Воевода П.Е., Зернов М.М. Применение технологий параллельного программирования в библиотеке для реализации генетических алгоритмов // Информационные технологии, энергетика и экономика XIII международная научно-техническая конференция студентов и аспирантов. - Смоленск: 2016 г., стр. 225-232.
4. Курейчик В.М., Кныш Д.С. Параллельный генетический алгоритм. Модели и проблемы построения // V Международная научно-практическая конференция "Интегрированные модели и мягкие вычисления в искусственном интеллекте" - Коломна: 2009 г.

### References

1. Skobcov YU.A., Speranskij D.V. Evolyucionnye vychisleniya: uchebnoe posobie.-M.: Nacional'nyj Otkrytyj Universitet “INTUIT”, 2015. –331 p.
  2. Borisov V. V., Zernov M. M., Fedulov A. S., YAkushevskij K. A. Issledovanie harakteristik gibridnogo vychislitel'nogo klastera // Sistemy upravleniya, svyazi i bezopasnosti. 2016. №4. pp. 129-146.
  3. Voevoda P.E., Zernov M.M. Primenenie tekhnologij parallel'nogo programmirovaniya v biblioteke dlya realizacii geneticheskikh algoritmov // Informacionnye tekhnologii, ehnergetika i ehkonomika XIII mezhdunarodnaya nauchno-tekhnicheskaya konferenciya studentov i aspirantov. - Smolensk: 2016, pp. 225-232.
  4. Kurejchik V.M., Knysh D.S. Parallelnyj geneticheskij algoritm. Modeli i problemy postroeniya // V Mezhdunarodnaya nauchno-prakticheskaya konferenciya "Integrirovannye modeli i myagkie vychisleniya v iskusstvennom intellekte" - Kolomna: 2009.
-