

УДК 004.8

DOI 10.18413/2411-3808-2019-46-3-553-562

**АЛГОРИТМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПО ВЫБОРУ СРЕДСТВ
ОБРАБОТКИ БОЛЬШИХ МАССИВОВ ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ДАННЫХ****THE ALGORITHM OF DECISION SUPPORT IN THE CHOICE OF MEANS
OF PROCESSING LARGE AMOUNTS OF NATURAL LANGUAGE DATA****К.А. Польщиков, О.Н. Польщикова, Е.В. Игитян, М.С. Балакшин
K.A. Polshchikov, O.N. Polshchikova, E.V. Igityan, M.S. Balakshin**Белгородский государственный национальный исследовательский университет,
Россия, 308015, г. Белгород, ул. Победы, 85

Belgorod National Research University, 85 Pobedy St, Belgorod, 308015, Russia

E-mail: polshchikov@bsu.edu.ru, polshchikova@bsu.edu.ru, medvedeva_e@bsu.edu.ru

Аннотация

Статья посвящена обеспечению выбора анализаторов текстов, позволяющих обрабатывать большие естественно-языковые массивы для принятия эффективных управляющих решений. Представлен обзор программных анализаторов текстов и предложен алгоритм поддержки принятия решений по выбору средств обработки массивов больших данных на естественном языке. В процессе работы алгоритм использует программу для оценки эффективности анализаторов текста. Эта программа основана на использовании системы нечеткого выбора, которая служит для расчета интегрального показателя эффективности анализатора текста. При оценке эффективности анализаторов текста учитываются качество и эффективность получения ответов на тестовые вопросы.

Abstract

This article presents an overview of the program text analyzers and proposes an algorithm to support decision making on the choice of solutions for processing large arrays of natural language data. In the process the algorithm uses the program for evaluating the effectiveness of text analyzers. This program is based on the operation of a fuzzy choice system, which serves to calculate the integral indicator of the text analyzer effectiveness. To set up a fuzzy choice system, special software and the participation of experts in a given subject area are required. In the process of setting up this system, the membership functions are approximated and the values of their parameters are determined. When evaluating the effectiveness of text analyzers, the quality and efficiency of getting answers to test questions are taken into account. Text analyzer with the highest integral efficiency indicator value is recommended for further use in the relevant subject area. The application of the algorithm presented in the article allows taking the best version of a management decision based on information obtained as a result of processing big data arrays of natural language.

Ключевые слова: массивы естественно-языковых больших данных, эффективность текстового анализатора, алгоритм поддержки принятия решений, система нечеткого выбора, интегральный показатель, аппроксимация функций принадлежности.

Keywords: arrays of natural language big data, text analyzer efficiency, decision support algorithm, fuzzy selection system, integral index, approximation of membership functions.

Введение

В крупных организациях, учреждениях и ведомствах в течение многих лет накоплены огромные массивы разнообразной информации в виде текстов на естественном языке. Такие информационные ресурсы можно классифицировать как большие данные, содержащие сведения, касающиеся конкретной компании, её целей, задач, структуры,

кадрового состава, проведенных мероприятий, реализованных проектов, финансовых оборотов, отчетов, перспективных планов, партнеров и т. п. Эти большие информационные массивы хранятся в электронном виде в разных форматах (MS Word, MS Excel, txt, pdf, djvu, HTML и др.) во многочисленных корпоративных системах, архивах и порталах, базах данных тех или иных подразделений, системах электронного документооборота, электронной почте, файловых каталогах и т. д. В процессе принятия тех или иных управленческих решений руководителю очень важно учитывать информацию, содержащуюся во всех источниках. Однако ввиду того, что эти данные чрезвычайно велики по объему, разнородны, несистематизированы и распределены по различным хранилищам, руководитель для принятия решений не имеет возможности использовать необходимую для этого полную информацию. По данным исследований корпорации IBM, руководители располагают не более 7 % информации, требуемой для выбора наилучшего варианта решения. В результате страдает качество управления, снижается эффективность функционирования компании, её конкурентоспособность.

Обозначенная выше проблема может быть решена, если из больших и плохо структурированных естественно-языковых массивов выделять нужную информацию, предоставляя возможность руководителю своевременно получать достоверные ответы на конкретные интересующие его вопросы. С этой целью применяются специальные программные средства, позволяющие осуществлять семантический анализ текстов, получая запрос и выдавая ответную информацию на естественном языке. В основе таких средств лежат компьютерные технологии лингвистического процессора, нацеленные на извлечение смысла из больших массивов естественно-языковых данных.

Для решения указанной задачи были разработаны различные варианты программного обеспечения, называемого анализаторами текстов, например, Google Desktop, Yandex.Server. Сейчас могут быть использованы семантические вопросно-ответные поисковые системы AskNet, комплекс программ для анализа русскоязычных текстов Russian Context Optimizer, программные продукты Ontos, информационно-аналитическая система ARION и др.

Цель исследования, изложенного в статье – обеспечить выбор анализаторов текстов, позволяющих обрабатывать большие естественно-языковые массивы для принятия эффективных управляющих решений. Для достижения указанной цели требуется решить следующие задачи:

- 1) предложить аналитические выражения для оценки эффективности программных анализаторов текстов;
- 2) разработать алгоритм поддержки принятия решения по выбору программных средств, обеспечивающих наиболее эффективный анализ текстов в заданной предметной области.

Показатели эффективности программных анализаторов текстов

Пользователь, стремясь разобраться в деталях решаемой задачи, формирует на естественном языке вопрос и отправляет его программному анализатору текстов. Анализатор выполняет семантическую обработку имеющихся текстовых массивов и выдает пользователю ответ на естественном языке.

Эффективность программного анализатора текстов предлагается оценивать на основе вычисления значений двух показателей:

- 1) Q – качество ответов, выдаваемых анализатором;
- 2) D – средняя длительность выдачи ответов.

Показатель Q должен учитывать правдивость и полноту выдаваемого ответа. Значение этого показателя можно вычислить по формуле:

$$Q = \frac{N_Q}{N}, \tag{1}$$

где N_Q – число достаточно качественных ответов, выданных анализатором на вопросы пользователя; N – общее число сформированных вопросно-ответных пар.

Показатель Q может принимать значения от 0 до 1. Чем больше значение Q , тем выше эффективность программного анализатора текстов.

Значение показателя D характеризует способность анализатора оперативно выдавать ответы на вопросы пользователя. Значение этого показателя можно вычислить по формуле:

$$D = \frac{1}{N} \sum_{i=1}^N T_i, \tag{2}$$

где T_i – время, в течение которого анализатором выдан ответ на вопрос номер i .

Величина D показывает, сколько в среднем времени требуется анализатору для формирования и выдачи ответа. Чем больше значение D , тем ниже эффективность программного анализатора текстов.

Величины Q и D являются частными показателями эффективности программного анализатора текстов. На их основе можно предложить интегральный показатель для принятия решений по выбору средств обработки больших массивов естественно-языковых данных.

Уровни качества выдаваемых анализатором ответов и оперативности их выдачи сложно определить строгими численными критериями. В этом случае можно использовать аппарат нечетких множеств. Тогда качество ответов анализатора может оцениваться с помощью нечетких множеств «Высокое качество ответов» и «Низкое качество ответов», а оперативность выдачи ответов на вопросы пользователя может соответствовать нечетким множествам «Высокая оперативность ответов» и «Низкая оперативность ответов». Найти результирующую оценку эффективности программного анализатора текстов можно, вычислив некоторый интегральный показатель E с помощью нечеткого логического вывода [Polshchykov, 2015; Rvachova, 2015; Polshchykov, 2015; Ivaschuk, 2016; Konstantinov, 2017].

Значение интегрального показателя E предлагается вычислять с помощью модели, соответствующей алгоритму нечеткого вывода Сугено нулевого порядка [Takagi, 1985], применение которого позволяет решить многие практические задачи [Усков, 2004; Штовба, 2007; Польщикова, 2014; Польщикова, 2015; Константинов, 2016]. В этом случае используется база нечетких правил:

$$\text{Если } (Q = V_Q) \text{ и } (D = V_D), \text{ то } E = Y_1, \tag{3}$$

$$\text{Если } (Q = V_Q) \text{ и } (D = W_D), \text{ то } E = Y_2, \tag{4}$$

$$\text{Если } (Q = W_Q) \text{ и } (D = V_D), \text{ то } E = Y_3, \tag{5}$$

$$\text{Если } (Q = W_Q) \text{ и } (D = W_D), \text{ то } E = Y_4, \tag{6}$$

где V_Q – нечеткое множество «Высокое качество ответов»;

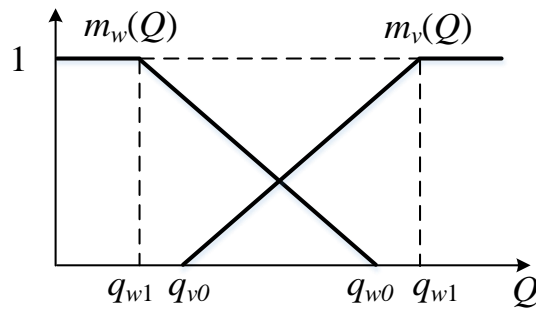
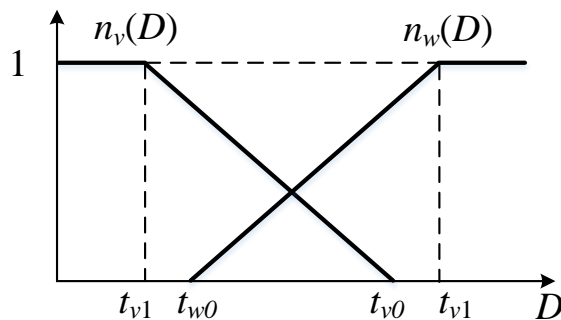
W_Q – нечеткое множество «Низкое качество ответов»;

V_D – нечеткое множество «Высокая оперативность ответов»;

W_D – нечеткое множество «Низкая оперативность ответов»;

$Y_1 = 4$, $Y_2 = 3$, $Y_3 = 2$ и $Y_4 = 1$ – значения индивидуальных выводов нечетких правил.

Степень принадлежности величин Q и D нечетким множествам определяется значениями соответствующих функций принадлежности $m_v(Q)$, $m_w(Q)$, $n_v(D)$ и $n_w(D)$ с параметрами q_{v0} , q_{v1} , q_{w0} , q_{w1} , t_{v0} , t_{v1} , t_{w0} и t_{w1} (рис. 1 и рис. 2).

Рис. 1. Функции принадлежности $m_v(Q)$ и $m_w(Q)$ Fig. 1. Membership function $m_v(Q)$ и $m_w(Q)$ Рис. 2. Функции принадлежности $n_v(D)$ и $n_w(D)$ Fig. 2. Membership function $n_v(D)$ и $n_w(D)$

Алгоритм нечеткого вывода Сугено нулевого порядка включает три основных этапа. На первом этапе выполняется фаззификация:

$$m_v(Q) = \begin{cases} 0, & Q < q_{v0}; \\ \frac{Q - q_{v0}}{q_{v1} - q_{v0}}, & q_{v0} \leq Q < q_{v1}; \\ 1, & Q \geq q_{v1}; \end{cases} \quad (7)$$

$$m_w(Q) = \begin{cases} 1, & Q < q_{w1}; \\ \frac{q_{w0} - Q}{q_{w1} - q_{w0}}, & q_{w1} \leq Q < q_{w0}; \\ 0, & Q \geq q_{w0}; \end{cases} \quad (8)$$

$$n_v(D) = \begin{cases} 1, & D < t_{v1}; \\ \frac{t_{v0} - D}{t_{v0} - t_{v1}}, & t_{v1} \leq D < t_{v0}; \\ 1, & D \geq t_{v1}; \end{cases} \quad (9)$$

$$n_w(D) = \begin{cases} 0, & D < t_{w0}; \\ \frac{D - t_{w0}}{t_{w1} - t_{w0}}, & t_{w0} \leq D < t_{w1}; \\ 1, & D \geq t_{w1}; \end{cases} \quad (10)$$

На втором этапе выполняется агрегирование:

$$G_1 = m_v(Q) \wedge n_v(D); \quad (11)$$

$$G_2 = m_v(Q) \wedge n_w(D); \quad (12)$$

$$G_3 = m_w(Q) \wedge n_v(D); \quad (13)$$

$$G_4 = m_w(Q) \wedge n_w(D). \quad (14)$$

На третьем этапе выполняется дефаззификация:

$$E = \frac{\sum_{r=1}^4 G_r Y_r}{\sum_{r=1}^4 G_r}. \tag{15}$$

Вычисление показателя E является результатом функционирования системы нечеткого вывода, предназначенной для оценки эффективности программного анализатора текстов. Чем выше значение интегрального показателя E , тем большей эффективностью обладает программный анализатор текстов.

Настройка параметров системы нечеткого вывода

Предложенную систему нечеткого выбора необходимо настроить, т. е. определить корректные значения параметров функций принадлежности.

Для этого предлагается провести серию из K экспериментов с участием M_0 экспертов. В каждом эксперименте номер k (где $k = 1, 2, \dots, K$) специально созданная программа для настройки системы нечеткого вывода задает вопросы и выдает на них ответы из предметной области, в которой эксперты имеют глубокие знания и опыт. При этом с помощью указанной программы фиксируется качество ответа и время его выдачи. В результате каждого эксперимента номер k определяются величины N , N_Q и T_i , а затем по формулам (1) и (2) вычисляются значения показателей Q_k и D_k . По окончании эксперимента в интерфейсе программы для настройки системы нечеткого вывода каждый эксперт отмечает, какими (высокими или низкими), по его мнению, были качество и оперативность выдачи ответов тестовой системой.

Далее вычисляются оценки функций принадлежности для каждого значения Q_k и D_k :

$$m_{vk}^* = \frac{M_Q^+(Q_k)}{M_0}; \tag{16}$$

$$m_{vk}^* = \frac{M_Q^-(Q_k)}{M_0}; \tag{17}$$

$$n_{vk}^* = \frac{M_D^+(D_k)}{M_0}; \tag{18}$$

$$n_{vk}^* = \frac{M_D^-(D_k)}{M_0}; \tag{19}$$

где $M_Q^+(Q_k)$ – число экспертов, которые отметили, что при Q_k качество ответов, выдаваемых тестовой системой, было высоким;

$M_Q^-(Q_k)$ – число экспертов, которые отметили, что при Q_k качество ответов, выдаваемых тестовой системой, было низким;

$M_D^+(D_k)$ – число экспертов, которые отметили, что при D_k оперативность выдачи ответов тестовой системой была высокой;

$M_D^-(D_k)$ – число экспертов, которые отметили, что при D_k оперативность выдачи ответов тестовой системой была низкой.

Затем из множества

$$m_v^* = \{m_{v1}^*, m_{v2}^*, \dots, m_{vk}^*, \dots, m_{vK}^*\} \tag{20}$$

путем отбора тех оценок, значения которых удовлетворяют условию $0.1 \leq m_{vk}^* \leq 0.9$, формируется множество

$$\tilde{m}_v = \{\tilde{m}_{v1}, \tilde{m}_{v2}, \dots, \tilde{m}_{vS}, \dots, \tilde{m}_{vS_1}\} \tag{21}$$



где s – номер текущего элемента сформированного множества; S_1 – число элементов в сформированном множестве. Аналогично из множеств

$$m_v^* = \{m_{w1}^*, m_{w2}^*, \dots, m_{wk}^* \dots m_{wK}^*\} \quad (22)$$

$$n_v^* = \{n_{v1}^*, n_{v2}^*, \dots, n_{vk}^* \dots n_{vK}^*\} \quad (23)$$

$$n_w^* = \{n_{w1}^*, n_{w2}^*, \dots, n_{wk}^* \dots n_{wK}^*\} \quad (24)$$

формируются множества

$$\tilde{m}_w = \{\tilde{m}_{w1}, \tilde{m}_{w2}, \dots, \tilde{m}_{ws} \dots \tilde{m}_{wS_2}\} \quad (25)$$

$$\tilde{n}_v = \{\tilde{n}_{v1}, \tilde{n}_{v2}, \dots, \tilde{n}_{vs} \dots \tilde{n}_{vS_3}\} \quad (26)$$

$$\tilde{n}_w = \{\tilde{n}_{w1}, \tilde{n}_{w2}, \dots, \tilde{n}_{ws} \dots \tilde{n}_{wS_4}\} \quad (27)$$

соответственно, элементы которых имеют значения в интервале от 0,1 до 0,9.

Далее требуется получить уравнения прямых:

$$y_1(Q) = a_1 Q + b_1, \quad (28)$$

$$y_2(Q) = a_2 Q + b_2, \quad (29)$$

$$y_3(D) = a_3 D + b_3, \quad (30)$$

$$y_4(D) = a_4 D + b_4, \quad (31)$$

сглаживающих значения элементов множеств \tilde{m}_v , \tilde{m}_w , \tilde{n}_v и \tilde{n}_w соответственно. Аппроксимацию оценок функций принадлежности предлагается выполнить по методу наименьших квадратов, в соответствии с которым коэффициенты прямых вычисляются по формулам:

$$a_1 = \frac{S_1 \sum_{s=1}^{S_1} Q_s \tilde{m}_{vs} - \sum_{s=1}^{S_1} Q_s \sum_{s=1}^{S_1} \tilde{m}_{vs}}{S_1 \sum_{s=1}^{S_1} Q_s^2 - (\sum_{s=1}^{S_1} Q_s)}; \quad (32)$$

$$b_1 = \frac{\sum_{s=1}^{S_1} \tilde{m}_{vs} - a_1 \sum_{s=1}^{S_1} Q_s}{S_1}; \quad (33)$$

$$a_2 = \frac{S_2 \sum_{s=1}^{S_2} Q_s \tilde{m}_{ws} - \sum_{s=1}^{S_2} Q_s \sum_{s=1}^{S_2} \tilde{m}_{ws}}{S_2 \sum_{s=1}^{S_2} Q_s^2 - (\sum_{s=1}^{S_2} Q_s)}; \quad (34)$$

$$b_2 = \frac{\sum_{s=1}^{S_2} \tilde{m}_{ws} - a_2 \sum_{s=1}^{S_2} Q_s}{S_2}; \quad (35)$$

$$a_3 = \frac{S_3 \sum_{s=1}^{S_3} D_s \tilde{n}_{vs} - \sum_{s=1}^{S_3} D_s \sum_{s=1}^{S_3} \tilde{n}_{vs}}{S_3 \sum_{s=1}^{S_3} D_s^2 - (\sum_{s=1}^{S_3} D_s)}; \quad (36)$$

$$b_3 = \frac{\sum_{s=1}^{S_3} \tilde{n}_{vs} - a_3 \sum_{s=1}^{S_3} D_s}{S_3}; \quad (37)$$

$$a_4 = \frac{S_4 \sum_{s=1}^{S_4} D_s \tilde{n}_{ws} - \sum_{s=1}^{S_4} D_s \sum_{s=1}^{S_4} \tilde{n}_{ws}}{S_4 \sum_{s=1}^{S_4} D_s^2 - (\sum_{s=1}^{S_4} D_s)}; \quad (38)$$

$$b_4 = \frac{\sum_{s=1}^{S_4} \tilde{n}_{ws} - a_4 \sum_{s=1}^{S_4} D_s}{S_4}; \tag{39}$$

Получив коэффициенты a_1 и b_1 сглаживающей прямой, можно определить искомые параметры $q_{\alpha 0}$ и $q_{\alpha 1}$ из уравнений

$$a_1 q_{v0} + b_1 = 0; \tag{40}$$

$$a_1 q_{v1} + b_1 = 1. \tag{41}$$

Решив уравнения, получим:

$$q_{v0} = -\frac{b_1}{a_1}; \tag{42}$$

$$q_{v1} = \frac{1 - b_1}{a_1}; \tag{43}$$

Аналогично выводятся значения параметров q_{w0} , q_{w1} , t_{v0} , t_{v1} , t_{w0} и t_{w1} . Полученные формулы для их расчета имеют следующий вид:

$$q_{w0} = -\frac{b_2}{a_2}; \tag{44}$$

$$q_{w1} = \frac{1 - b_2}{a_2}; \tag{45}$$

$$t_{v0} = -\frac{b_3}{a_3}; \tag{46}$$

$$t_{v1} = \frac{1 - b_3}{a_3}; \tag{47}$$

$$t_{w0} = -\frac{b_4}{a_4}; \tag{48}$$

$$t_{w1} = \frac{1 - b_4}{a_4}; \tag{49}$$

Алгоритм поддержки принятия решения по выбору анализаторов текстов

Выполнив настройку параметров синтезированной системы нечеткого вывода, её можно использовать для оценки эффективности программных анализаторов текстов и последующего принятия решения по выбору средств обработки больших массивов естественно-языковых данных. С этой целью необходимо создать программу по оценке эффективности анализаторов текстов, включающую вышеуказанную систему нечеткого вывода, базу данных тестовых вопросов и правильных ответов, а также средств, обеспечивающих вычисление величин Q и D по формулам (1) и (2). Располагая данной программой, при участии человека-тестировщика можно выбрать наиболее эффективный анализатор текстов из совокупности альтернативных вариантов, для чего предлагается следующий алгоритм:

1. Запускается программа по оценке эффективности анализатора текстов.
2. Запускается исследуемый программный анализатор текстов и подключается к большим массивам естественно-языковых данных, содержащих информацию о требуемой предметной области.
3. Программа по оценке эффективности анализатора текстов задает тестовый вопрос.
4. Тестировщик получает вопрос, заданный программой по оценке эффективности анализатора текстов.



5. Тестировщик вводит задаваемый вопрос в интерфейсе исследуемого анализатора текстов и запускает таймер ожидания ответа.

6. Программный анализатор текстов формирует и выдает ответ на заданный вопрос.

7. Тестировщик получает ответ, выданный анализатором текстов, и останавливает таймер ожидания ответа.

8. Тестировщик в интерфейсе программы по оценке эффективности анализатора текстов выбирает вариант, который в наибольшей степени соответствует полученному ответу.

9. Если все тестовые вопросы заданы, то программа по оценке эффективности анализатора текстов вычисляет значение интегрального показателя E , в противном случае выполнение алгоритма возвращается в пункт 3.

10. Если все анализаторы текстов исследованы, то для дальнейшего использования рекомендуется анализатор с наибольшим значением интегрального показателя эффективности, в противном случае выполнение алгоритма возвращается в пункт 2.

Заключение

Таким образом, в статье представлен обзор программных анализаторов текстов и предложен алгоритм поддержки принятия решения по выбору средств обработки больших массивов естественно-языковых данных. При выполнении алгоритма используется программа по оценке эффективности анализаторов текстов. Эта программа основана на работе системы нечеткого выбора, служащей для вычисления интегрального показателя эффективности анализатора текстов. Для настройки системы нечеткого выбора требуются специальные программные средства и участие экспертов в заданной предметной области. В процессе настройки этой системы осуществляется аппроксимация функций принадлежности и определяются значения их параметров. При оценивании эффективности анализаторов текстов учитывается качество и оперативность получения ответов на тестовые вопросы. Анализатор текстов с наибольшим значением интегрального показателя эффективности рекомендуется для дальнейшего использования в соответствующей предметной области. Применение представленного в статье алгоритма позволяет принять наилучший вариант управленческого решения на основе информации, полученной в результате обработки больших массивов естественно-языковых данных.

Дальнейшие исследования по теме статьи будут посвящены разработке программного обеспечения на основе предложенного алгоритма и получению экспериментальных результатов его применения.

Настоящая статья содержит результаты проекта «Разработка инструментальных средств реализации естественно-языковых систем обработки больших данных», выполняемого в рамках реализации Программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по Договору МГУ им. М.В. Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 11.12.2018 № 13/1251/2018.

Список литературы

1. Константинов И.С., Лазарев С.А., Польщиков К.А. 2016. Нечеткая система для оценки эффективности управления информационным обменом сети корпоративных порталов. Вестник компьютерных и информационных технологий, 9: 42–47.

2. Польщиков К.А. 2015. Нейро-нечеткая система выбора межсегментного интервала в телекоммуникационной сети. Вестник НГУ. Серия: Информационные технологии, 13 (4): 33–42.

3. Польщиков К.А. 2014. Нейро-нечеткое прогнозирование длительности ожидания подтверждений в телекоммуникационной сети. Системный анализ и информационные технологии в науках о природе и обществе, 1–2: 57–61.

4. Польщиков К.А. 2015. Система прогнозирования длительности ожидания подтверждений в телекоммуникационной сети. Инфокоммуникационные технологии, 2: 148–153.

5. Польщиков К.А. 2015. Система прогнозирования загрузки маршрутизатора на основе нечеткой нейронной сети. Вестник компьютерных и информационных технологий, 9: 55–60.
6. Польщиков К.А., Здоренко Ю.Н. 2014. Усовершенствованный метод нейро-нечеткого управления отбрасыванием пакетов в транзитных маршрутизаторах телекоммуникационной сети. Проблемы телекоммуникаций, 2 (14): 76–90.
7. Усков А.А., Кузьмин А.В. 2004. Интеллектуальные технологии управления. Искусственные нейронные сети и нечеткая логика. М. Горячая линия-Телеком, 143.
8. Штовба С.Д. 2007. Проектирование нечетких систем средствами MATLAB. М. Горячая линия-Телеком, 288.
9. AskNet. Available at: <http://asknet.ru> (accessed: 26.06.2019).
10. Business Robotics. Available at: <http://ontos.com> (accessed: 26.06.2019).
11. Information and analytical system ARION. URL: <http://sytech.ru/about.php?id=5> (accessed: 26.07.2019).
12. Ivaschuk O., Polschikov K., Lazarev S. et al. 2016. Integral estimate of terrestrial compartment condition in management of Biotechnosphere of Rural and Urban Areas. International Journal of Pharmacy and Technology, 8 (4): 27032–27038.
13. Konstantinov I., Polshchikov K., Lazarev S. 2017. Model of Neuro-Fuzzy Prediction of Confirmation Timeout in a Mobile Ad Hoc Network. CEUR Workshop Proceedings. Mathematical and Information Technologies, 1839: 174–186.
14. Konstantinov I., Polshchikov K., Lazarev S. 2017. The Algorithm for Neuro-Fuzzy Controlling the Intensity of Retransmission in a Mobile Ad-Hoc Network. International Journal of Applied Mathematics and Statistics, 56 (2): 85–90.
15. Polshchikov K., Lazarev S., Zdorovtsov A. 2017. Neuro-Fuzzy Control of Data Sending in a Mobile Ad Hoc Network. Journal of Fundamental and Applied Sciences, 9 (2S): 1494–1501.
16. Polshchikov K., Zdorenko Y., Masesov M. 2015. Neuro-Fuzzy System for Prediction of Telecommunication Channel Load. Problems of Infocommunications Science and Technology. Proceedings of the Second International Scientific-Practical Conference. Kharkiv: 33–34.
17. RCO. Available at: <http://www.rco.ru> (accessed: 26.06.2019).
18. Rvachova N., Sokol G., Polschikov K., Davies J. 2015. Selecting the intersegment interval for TCP in Telecomms networks using fuzzy inference system. Proceedings of the Sixth International Conference 2015 Internet Technologies and Applications (ITA). Wrexham, Glyndwr University: 256–260.
19. SyTech. Available at: <http://www.sytech.ru/about.php?id=60> (accessed: 26.06.2019).
20. Takagi T., Sugeno M. 1985. Fuzzy Identification of Systems and Its Applications to Modeling and Control. IEEE Transactions on Systems, Man, and Cybernetics, 15: 116–132.

References

1. Konstantinov I.S., Lazarev S.A., Pol'shnikov K.A. 2016. Nечetkaja sistema dlja ocenki jeffektivnosti upravlenija informacionnym obmenom seti korporativnyh portalov [A fuzzy system for assessing the effectiveness of managing the information exchange of a network of corporate portals]. Bulletin of Computer and Information Technologies, 9: 42–47.
2. Pol'shnikov K.A. 2015. Nejro-nечetkaja sistema vybora mezhssegmentnogo intervala v telekommunikacionnoj seti [Neuro-fuzzy system of choosing an intersegment interval in a telecommunication network]. Bulletin NGU. Information Technology, 13 (4): 33–42.
3. Pol'shnikov K.A. 2014. Nejro-nечetkoe prognozirovanie dlitel'nosti ozhidanija podtverzhdenij v telekommunikacionnoj seti [Neuro-fuzzy prediction of the waiting time for confirmations in the telecommunications network]. Systems Analysis and Information Technologies in the Sciences of Nature and Society, 1–2: 57–61.
4. Pol'shnikov K.A. 2015. Sistema prognozirovanija dlitel'nosti ozhidanija podtverzhdenij v telekommunikacionnoj seti [The system for predicting the duration of waiting for confirmations in the telecommunications network]. Infocommunication technology, 2: 148–153.
5. Pol'shnikov K.A. 2015. Sistema prognozirovanija zagruzki marshrutizatora na osnove nechetkoj nejronnoj seti [A router load prediction system based on a fuzzy neural network]. Bulletin of Computer and Information Technologies, 9: 55–60.
6. Pol'shnikov K.A., Zdorenko Ju.N. 2014. Uovershenstvovannyj metod nejro-nечetkogo upravlenija otrasyvaniem paketov v tranzitnyh marshrutizatorah telekommunikacionnoj seti [Advanced



method of neuro-fuzzy control of packet drop in transit routers of a telecommunication network]. *Telecommunication problems*, 2 (14): 76–90.

7. Uskov A.A., Kuz'min A.V. 2004. *Intellektual'nye tehnologii upravlenija. Iskusstvennye nejronnye seti i nechetkaja logika [Intelligent Management Technologies]*. M. Gorjachaja linija-Telekom, 143.

8. Shtovba S.D. 2007. *Proektirovanie nechetkih sistem sredstvami MATLAB [Design of fuzzy systems using MATLAB]*. M. Gorjachaja linija-Telekom, 288.

9. AskNet. Available at: <http://asknet.ru> (accessed: 26.06.2019).

10. Business Robotics. Available at: <http://ontos.com> (accessed: 26.06.2019).

11. Information and analytical system ARION. URL: <http://sytech.ru/about.php?id=5> (accessed: 26.07.2019).

12. Ivaschuk O., Polschikov K., Lazarev S. et al. 2016. Integral estimate of terrestrial compartment condition in management of Biotechnosphere of Rural and Urban Areas. *International Journal of Pharmacy and Technology*, 8 (4): 27032–27038.

13. Konstantinov I., Polschikov K., Lazarev S. 2017. Model of Neuro-Fuzzy Prediction of Confirmation Timeout in a Mobile Ad Hoc Network. *CEUR Workshop Proceedings. Mathematical and Information Technologies*, 1839: 174–186.

14. Konstantinov I., Polschikov K., Lazarev S. 2017. The Algorithm for Neuro-Fuzzy Controlling the Intensity of Retransmission in a Mobile Ad-Hoc Network. *International Journal of Applied Mathematics and Statistics*, 56 (2): 85–90.

15. Polschikov K., Lazarev S., Zdorovtsov A. 2017. Neuro-Fuzzy Control of Data Sending in a Mobile Ad Hoc Network. *Journal of Fundamental and Applied Sciences*, 9 (2S): 1494–1501.

16. Polschikov K., Zdorenko Y., Masesov M. 2015. Neuro-Fuzzy System for Prediction of Telecommunication Channel Load. *Problems of Infocommunications Science and Technology. Proceedings of the Second International Scientific-Practical Conference*. Kharkiv: 33–34.

17. RCO. Available at: <http://www.rco.ru> (accessed: 26.06.2019).

18. Rvachova N., Sokol G., Polschikov K., Davies J. 2015. Selecting the intersegment interval for TCP in Telecomms networks using fuzzy inference system. *Proceedings of the Sixth International Conference 2015 Internet Technologies and Applications (ITA)*. Wrexham, Glyndwr University: 256–260.

19. SyTech. Available at: <http://www.sytech.ru/about.php?id=60> (accessed: 26.06.2019).

20. Takagi T., Sugeno M. 1985. Fuzzy Identification of Systems and Its Applications to Modeling and Control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15: 116–132.

Ссылка для цитирования статьи

Reference to article

Польщикова К.А., Польщикова О.Н., Игитян Е.В., Балакшин М.С. 2019. Алгоритм поддержки принятия решений по выбору средств обработки больших массивов естественно-языковых данных. *Научные ведомости Белгородского государственного университета. Серия: Экономика. Информатика*. 46 (3): 553–562. DOI 10.18413/2411-3808-2019-46-3-553-562.

Polshchikov K.A., Polshchikova O.N., Igityan E.V., Balakshin M.S. 2019. The algorithm of decision support in the choice of means of processing large amounts of natural language data. *Belgorod State University Scientific Bulletin. Economics. Information technologies*. 46 (3): 553–562 (in Russian). DOI 10.18413/2411-3808-2019-46-3-553-562.