УДК 004.043

МЕТОДЫ ВОССТАНОВЛЕНИЯ ОТСУТСТВУЮЩИХ ДАННЫХ В ЗАДАЧАХ ЭКОНОМИЧЕСКОЙ ДИАГНОСТИКИ, ОСНОВАННЫХ НА РАССУЖДЕНИЯХ ПО ПРЕЦЕДЕНТАМ[1]

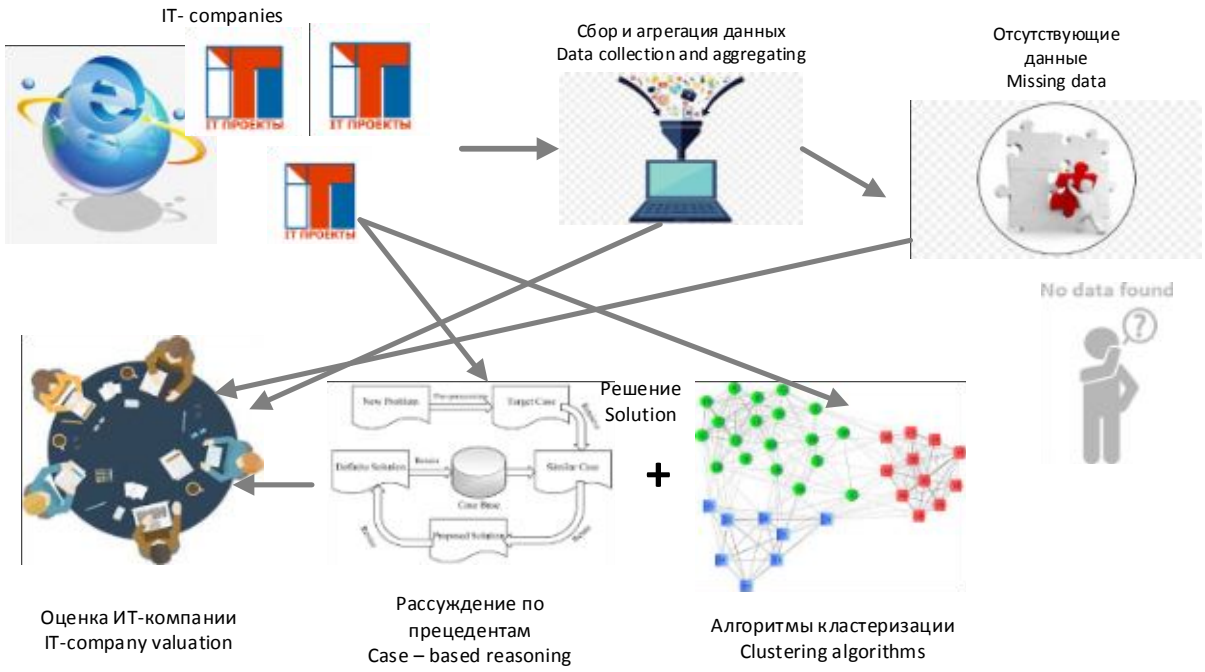
Статья поступила в редакцию 01.02.2019, в окончательном варианте – 12.02.2019.

Квятковская Анастасия Евгеньевна, Астраханский государственный технический университет, 414056, Российская Федерация, г. Астрахань, ул. Татищева, 16,
ассистент, e-mail: zima00@list.ru
Чертина Елена Витальевна, Астраханский государственный технический университет, 414056, Российская Федерация, г. Астрахань, ул. Татищева, 16,
кандидат технических наук, доцент, ORCID 0000-0001-8866-2487, e-mail: saprikinae_1912@mail.ru
Шуршев Валерий Федорович, Астраханский государственный технический университет, 414056, Российская Федерация, г. Астрахань, ул. Татищева, 16,
доктор технических наук, профессор, e-mail: v.shurshev@mail.ru

В исследовании рассматривается проблема оценки экономического состояния IT-компании, основанной на рассуждениях по прецедентам, в условиях недостаточности данных. База прецедентов представлена как база знаний о IT-компаниях, обладающих определенными характеристиками и свойствами. Для экономической диагностики компании-аналога используется мера близости прецедента и аналога, определяемая по имеющимся характеристикам в базе прецедентов. Авторы отмечают, что проблема неполноты имеющихся данных может привести к нарушению структуры массива данных в базе прецедентов и, как следствие, невозможности или недостоверности определения экономического состояния исследуемой IT-компании. Для решения этой проблемы предлагается комплекс методов по обработке пропущенных данных: исключение объектов с пропущенными данными; использование инструментов математической статистики; использование аппарата кластерного анализа и классификации. Разработан алгоритм оценки исследуемого аналога, для которого отсутствуют измеренные значения некоторых данных. Данный алгоритм может быть использован ЛПР при принятии управленческих решений по оценке ITкомпании в условиях информационной неопределённости.

Ключевые слова: IT-компания, экономическая диагностика, рассуждение по прецедентам, компания-аналог, принятие решений, алгоритмы нечеткой кластеризации, пропущенные данные

Графическая аннотация (Graphical annotation)



[1] Работа выполнена при финансовой поддержке РФФИ, проект № 18-37-00130.

MISSING DATA RECOVERY METHODS FOR THE ISSUES OF ECONOMIC DIAGNOSTICS, BASED ON CASE-BASED REASONING

The article was received by editorial board on 01.02.2019, in the final version – 12.02.2019.

Kvyatkovskaya Anastasia E., Astrakhan State Technical University, 16 Tatishchev St., Astrakhan, 414056, Russian Federation,
Assistant, e-mail: zima00@list.ru
Chertina Elena V., Astrakhan State Technical University, 16 Tatishchev St., Astrakhan, 414056, Russian Federation,
Cand. Sci. (Engineering), Associate Professor, e-mail: saprikinae_1912@mail.ru
Shurshev Valeriy F., Astrakhan State Technical University, 16 Tatishchev St., Astrakhan, 414056, Russian Federation,
Doct. Sci. (Engineering), Professor, e-mail: v.shurshev@mail.ru

The study addresses the problem of the economic state valuation of an IT company, based on case-based reasoning in the context of inadequate data. The base of precedents is presented as a knowledge base of IT companies with certain characteristics and features. A proximity measure of a precedent and an analogue is used for the economic diagnostics of a peer company. A proximity measure is determined by the available characteristics in the base of precedents. The authors note that the problem of available data incompleteness can lead to structural imperfection of the data set in the base of precedents and, as a result, the impossibility or unreliability of the economic state valuation of an IT company. To solve this problem, a complex of methods for missing data processing is proposed: the deletion of objects with missing data; using of mathematical statistics tools; using of cluster analysis and classification tools. There has been developed an algorithm for valuation of analyzed peer company which tests the peer company's characteristics with some data on it being missing. This algorithm can be used by the decision - maker for decisions making process on IT company valuation in the context of information uncertainty.

Key words: IT company, case-based reasoning, peer company, decision making, fuzzy clustering algorithms, missing data

Введение. Исследование экономических систем обязательно основано на их анализе и прогнозировании состояний. В то же время это исследование связано с неопределенностью, вследствие отсутствия достоверных источников информации, верифицированных данных, недостаточностью или полным отсутствием информации об исследуемом объекте.

При неполноте выборки, с одной стороны, не исключается применимость методов статистического анализа набора данных – пространственной или временной выборки о деятельности объекта управления. В то же время при исследовании статистических данных необходимо учитывать проблему пропущенных значений при анализе данных, осложняющую эффективность существующих способов обработки информации.

В данной работе предполагается продемонстрировать применение методов восстановления пропущенных данных в сочетании с решением задач экономической диагностики, включающих анализ и прогнозирование экономического состояния предприятия на основе изучения ретроспективной (причем иногда неполной) информации о его деятельности.

Целью данной работы является определение совокупности методов, позволяющих моделировать пропущенные данные, полученные из различных источников, для решения задач экономической диагностики IT-компаний с применением методов рассуждения по прецедентам.

Общая характеристика проблематики статьи, постановка задачи и подходы к ее решению. В настоящее время исследования экономических систем с применением инструментов искусственного интеллекта развиваются в направлении применения методов, использующих рассуждение по прецедентам. Слабая формализуемость ряда предметных областей, в числе которых и экономика, не позволяет построить и верифицировать цепочку правил, рассуждений, приводящих к обоснованному результату. Методы рассуждения по прецедентам предлагают рассмотреть некоторый объект с заранее определенными свойствами, формализовать его описание, внести в базу знаний о прецедентах. Это позволяет в последующих ситуациях сравнивать новые объекты, полученные для анализа (оценки), с данными известных прецедентов. Предполагается, что при последующем мониторинге свойств прецедента тенденции его развития могут быть спроецированы на исследуемые объекты со схожими характеристиками. Накопление базы прецедентов тождественно созданию базы знаний о предметной области, сохраняющей эмпирические данные о поведении прецедента в различных ситуациях [5, 10].

Предлагается применить методы рассуждения по прецедентам к предприятию информационно-коммуникационной сферы, в частности – к молодой IT-компании или стартапу, нуждающихся в экономической диагностике для решения задач эффективного управления. Для решения задачи оценки стоимости бизнеса выбран сравнительный подход. Он основан на сравнении заданного предприятия с ком-

паниями-аналогами по группе заранее заданных характеристик. Эти характеристики могут быть метрическими (выручка, число сотрудников, валовая прибыль, базовая стоимость, задолженность, регулярный месячный доход, сумма затрат на привлечение новых клиентов, уровень возврата инвестиций и др.) и неметрическими (различные виды рисков, оценка команды и др.). Поскольку данная задача решается на основе методов информационного поиска, то один из вариантов решения – поиск сведений о компании-аналоге в интернетt. При этом встает вопрос о недостаточности данных для принятия решения с необходимым уровнем достоверности в отношении сходства тенденций развития исследуемой компании с найденными компаниями-аналогами.

При решении задачи поиска подходящего прецедента возникает ряд *проблем.*

1. Проблема выбора и обоснования метрики (меры) близости прецедента и компании-аналога. Природа характеристик может быть обусловлена как метрическими, так и неметрическими свойствами, что не позволяет использовать для них общую метрику близости.

2. Проблема неполноты информации, отсутствия данных, связанных с ограниченными возможностями информационного поиска при выполнении поисковых запросов, касающихся финансовой или другой информации, которую предприятие считает конфиденциальной.

3. Проблема использования экспертных знаний для оценивания аналогов из-за недостатка информации о предметной области.

Общая постановка задачи выглядит следующим образом.

В целях поиска аналогов элементы базы данных прецедентов, состоящей из n IT-компаний $A = \{a_1, a_2, \ldots a_n\}$, сравниваются с исследуемой компанией a^* по группе из m свойств $S = \{s_1, s_2, \ldots s_m\}$. Каждому свойству IT-компании a_i сопоставлено множество характеристик $S = \{s_j\}$, $j = 1,\ldots, m$, каждая IT-компания представима в виде: $a_i = \{f_1(s_1), f_2(s_2), \ldots f_m(s_m)\}$, где $f_j(s_j)$, – характеристическая функция, определяющая подмножество свойств $S^* \subseteq S$ для i-й IT-компании.

Из множества прецедентов A необходимо выбрать подмножество $A^* \subseteq A$, состоящее из прецедентов, близких к исследуемой компании a^*:

$$a^*_p = \{a \in A \mid \rho(a_i, a^*) \le \varepsilon, p \in \{1, 2, \ldots, n\}\}, \quad (1)$$

где $\rho(a_p, a^*)$ – метрика близости прецедентов и аналога, вычисляемая по m характеристикам аналога и прецедентов; ε – заданный уровень отклонения j-х характеристик аналога и прецедента друг от друга, $A^* = \{a^*_p\}$.

База прецедентов представляет собой набор записей, атрибуты которых соответствуют характеристикам, а поля содержат значения этих характеристик. При мониторинге состояния n прецедентов по m характеристикам в моменты времени $t_\tau = t_1, t_2, \ldots t_T$ мы имеем массив данных $Data(i, j, \tau)$, сохраненный в хранилище данных.

Проблема неполноты или недостоверности информации выражается в том, что при измерении объекта a_i, произведенного в заданные моменты времени t_τ, нарушается заполнение массива $Data$ – вследствие отсутствия информации о значениях характеристики (характеристик) $s_j, j \in \{1, 2, n\}$.

В то же время предполагается дальнейшее решение задач выбора с участием заданного множества прецедентов A:

– при введении на множестве A отношения эквивалентности RE, функция выбора $C^{RE}(A)$ определяется для $a, a^* \in A$:

$$C^{RE}(A) = \{ a \in A \mid a \sim a^* \}, \quad (2)$$

где $C^{RE}(A)$ – множество аналогов компаний, сходных с прецедентом a^*, решается задача поиска аналогов, сходных с прецедентом;

– при введении на множестве А отношения порядка RP, если поставить в соответствие множеству A множество перестановок длины n $\Omega = \{<1, 2,\ldots, n>, <1, 3,\ldots, n, 2>,\ldots,<n, n–1,\ldots, 1>\}$, решается задача ранжирования, упорядочивающая аналоги по убыванию (возрастанию) сходства с прецедентом, при этом $C^{RP}(\Omega) = <num_1, num_2, \ldots, num_n>$, где num_i – номер i-го аналога при указанном упорядочении;

– если классификационные свойства аналогов заранее неизвестны, но задана функция расстояния между аналогами и прецедентом $\rho(a_i, a^*)$, решается з*адача* кластеризации, состоящая в разбиении исходного множества аналогов A на попарно непересекающиеся подмножества A^l, $l = 1, \ldots, L$, где L – количество кластеров, $\Omega = \{1, \ldots, L\}$ – множество меток кластеров. Элементы каждого кластера близки по метрике ρ и существенно отличаются от элементов других кластеров.

Решение последней задачи связано с введением на множестве A отношения эквивалентности для оценки принадлежности компании к какому-либо классу, оцениваемому по номинальной шкале. Внутри каждого класса содержатся прецеденты, близкие по своим свойствам друг другу.

Рассмотрим возможные подходы к обработке пропущенных данных в информации, полученной при мониторинге объекта методами экономической диагностики [3, 11, 12, 13, 14, 15].

1. Исключение объектов с пропущенными данными.

В предположении о пропущенных данных предлагается исключить из хранилища данных строки, содержащие атрибутивные данные о временных или пространственных характеристиках объекта исследования с пропущенными значениями, в случае, если они носят случайный характер. Примером является отсутствие информации по какому-либо признаку при информационном поиске в интернете, когда налицо недостаточность информации. Например, это может относиться к финансовому состоянию объекта или к представлению устаревших, неактуальных или недостоверных данных, данных с большими погрешностями.

К недостаткам данного подхода следует отнести значительные искажения при анализе такой информации, пропуск важнейших тенденций или закономерностей.

2. Использование методов математической статистики, применяющих специальные методики замены пропущенных данных их приблизительными, усредненными значениями. Это могут быть, например, средние значения характеристик, их моды, медианы характеристик.

Данные методы применяются, когда пропущенные данные существуют, но не измерены вследствие единичных искажений, ошибок. Такой подход позволяет прогнозировать или экстраполировать значения этих характеристик методами математической статистики. Возможно, что факт появления пропущенных данных сам по себе является закономерностью, что позволяет описать его возникновение теми же методами. Если пропуски данных имеют систематический характер, то при их удалении может произойти потеря факторов влияния. Следует отметить, что методы математической статистики в большей степени уместны, когда заранее известно, что исследуются однородные объекты с одинаковыми свойствами, тенденциями, закономерностями, имеющими одинаковое пространство признаков [7]. Отсутствие данных в таких случаях обусловлено наличием шумов, небрежностью проведения измерений, погрешностями. Данные методы в меньшей степени используются для восстановления пропущенных данных базы прецедентов вследствие уникальности хранящихся в ней записей.

В нашем случае, когда сканируется интернет-пространство в целях получения данных о раритетном по информационным свойствам объекте, говорить о закономерностях затруднительно. Неопределенность представленной задачи вызвана тем, что для каждого нового прецедента может быть сформировано собственное пространство признаков и собственная мера близости с другими прецедентами. Это исключает использование усредненных данных вместо пропущенных.

3. Использование методов классификации и кластеризации для работы с данными с отсутствующими значениями, основанными на анализе предметной области.

Для разбиения исходного множества прецедентов на классы или кластеры можно воспользоваться различными методами. Например, это могут быть основные алгоритмы нечеткой кластеризации, в том числе по методу с-средних, алгоритм Густафсона-Кесселя [6, 8, 9]. Задача нечетной кластеризации заключается в определении структуры нечетких кластеров, присутствующих в рассматриваемых данных, путем разбиения базы прецедентов в целях последующего определения степени соответствия прецедента нечетким кластерам.

При недостатке данных при описании прецедентов сравниваемая с ними компания может попасть в пересечение классов. При работе с пропущенными значениями используем понятие «локальной контекстно-независимой метрики» [2, 4].

Разработаем следующий *алгоритм* для оценки исследуемого аналога $\tilde{a}$, при измерении свойств которого отсутствуют данные по характеристике $\tilde{s}$ (в данной работе продемонстрируем выполнение процедуры при отсутствии данных только по одной характеристике). Алгоритм включает в себя следующие шаги.

1. Произведем разбиение исходного множества A на L классов эквивалентности $A^1, A^2, \ldots A^L$ по совокупности из m характеристик в условиях полной информации: отсутствие пропусков данных у прецедентов a_i базы знаний.

2. Определим границы каждого класса, поставив в соответствие каждому объекту a из базы прецедентов $a^l_{i,j}$ – значение j-й характеристики i-го прецедента l-го класса: $Lt(A^l) = min\ (a^l_{i,j})$; $Rt(A^l) = max\ (a^l_{i,j})$ – соответственно левая и правая границы класса. Для этого произведем полный перебор базы прецедентов.

3. Оценим принадлежность исследуемого аналога a^* с известным множеством характеристик $\tilde{S} = \{s_1, s_2, \ldots s_m\} \backslash \tilde{s}$ каждому классу A^l. Очевидно, что отсутствие данных по характеристике $\tilde{s}$ допускает соотнесение исследуемого аналога к одному и более классам. Это означает, что часть прецедентов, близких к исследуемому аналогу, попадают в пересечение классов. Для этого будем проверять значение каждой характеристики исследуемого аналога a^* на выполнение условия:

$$Lt(A^l) \le a^*_{i,j} \le Rt(A^l) \text{ для всех } l = 1, 2, \ldots, L. \quad (3)$$

Введем функцию $Pr(l,a)$, определяющую для прецедента a число характеристик, значения которых попадают в границы одного класса:

$$Pr(l,a) = \sum_{j\in\{1,2,\ldots,m\}} (if\,(\mathrm{Lt}(A^l) \le a^l_{i,j} \le \mathrm{Rt}(A^l),\ then\ 1,\ else\ 0)), \quad (4)$$

Будем считать, что прецедент a принадлежит классу A^l, если неравенство (3) выполняется для всех характеристик $\{s_1, s_2, \ldots s_m\}\setminus \widetilde{s}$, т.е. $Pr(l,a) = |\widetilde{S}|$.

4. При подборе аналогов в базе прецедентов для текущего исследуемого аналога a^* возможны случаи, когда у имеющихся прецедентов также имеется отсутствующая информация, но по другим характеристикам. Дадим определение *аналогов* с учетом наличия у них пропущенных данных: прецедент a_i является аналогом a^*, если у обоих имеются пропущенные данные по одинаковым характеристикам и они попадают в одни и те же классы прецедентов.

Определим вектор $K(a)$, определяющий попадание объекта a в классы прецедентов l, элемент которого $K_l(a)$ определяется следующим образом:

$$K_l(a) = \begin{cases} 1, & \text{если } a \in \mathrm{A}^l, \\ 0, & \text{в противном случае.} \end{cases} \quad (5)$$

Найдем $K_l(a^*)$ и зафиксируем lk – метки классов, в которые попадает исследуемый аналог a^*, $lk \in \{1, 2, \ldots, L\}$.

Далее найдем расстояние между векторами $K(a^*)$ и $K(a)$ на основе метрики Хемминга, используя только классы, помеченные метками lk:

$$\rho_{local}(a, a^*) = \sum_{lk\in\{1,2,\ldots,L\}} |K_{lk}(a) - K_{lk}(a^*)|. \quad (6)$$

Для всех элементов известного множества прецедентов $\{a_i\}$ найдем $\rho_{local}(a_i, a^*)$. Наиболее точным аналогом считается тот, у которого $\rho_{local}(a_i, a^*) = 0$. В остальных случаях можно ранжировать аналоги-прецеденты по степени убывания сходства по отношению к исследуемого аналогу, образуя семейство классов, между которыми можно ввести отношение частичного порядка $\Omega^1(a^*)$, $\Omega^2(a^*)$, …, – результат разбиения исходного множества прецедентов A на классы $\Omega^q(a^*)$, где $q = \rho_{local}(a_i, a^*)$.

5. Для каждого класса $\Omega^q(a^*)$ также можно уточнить диапазон границ для пропущенных данных по характеристике $\widetilde{S}$:

$$a_{p,s} \le a^*_{s} \le a_{p,s},\ p = 1, 2, \ldots, |\Omega^q(a^*)|. \quad (7)$$

6. Останов.

Результатом вычислений по данному алгоритму могут быть два варианта:

– пустое множество решений, т.е. не встретился ни один экземпляр, попавший в границы пересечений классов, определяемых исследуемым аналогом;

– избыточность множества решений, когда, например, по известным характеристикам «число сотрудников», «базовая стоимость», «валовая прибыль» обнаруживается значительное количество аналогов.

В обоих случаях необходимо участие лица, принимающего решение, способного либо заполнить дополнительный признак, либо указать его границы.

Заключение. Рассмотрены методы, применение которых позволяет обрабатывать слабоформализуемые данные, полученные из различных источников, с целью осуществления экономической диагностики IT-компаний, основанной на рассуждениях по прецедентам.

Использование предложенного комплекса методов будет способствовать эффективному принятию управленческих решений при определении прогнозных состояний IT-стартапов в условиях неопределённости и недостаточности данных.

Библиографический список

1. Брумштейн Ю. М. Модели оптимизации подбора ресурсов при управлении совокупностью проектов с учетом зависимости качества результатов, рисков, затрат / Ю. М. Брумштейин, И. А. Дюдиков // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. – 2015 – № 1. – С. 78–88.

2. Дюк В. А. Методология поиска логических закономерностей в предметной области с нечеткой системологией: на примере клинико-экспериментальных исследований : дис. ... д-ра техн. наук : 05.13.01. – Санкт-Петербург, 2005. – 309 с. : ил. РГБ ОД, 71 06-5/46.

3. Карлов И. А. Восстановление пропущенных данных при численном моделировании сложных динамических систем / И. А. Карлов // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление. – 2013. – № 6 (186). – С. 137–144.

4. Карпов Л. Е. Методы добычи данных при построении локальной метрики в системах вывода по прецедентам / Л. Е. Карпов, В. Н. Юдин // Препринты Института системного программирования РАН. – Препринт 18. – 2006. – С. 1–42.

5. Квятковская А. Е. Информационная технология поиска компаний-аналогов для оценки стоимости бизнеса, использующая интеллектуальных агентов / А. Е. Квятковская, Е. В. Чертина, А. О. Полумордвинова,

И. Ю. Квятковская // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. – 2019. – № 1. – С. 99–106.

6. Квятковская И. Ю. Разработка алгоритма подбора приоритетных венчурных IT-проектов / И. Ю. Квятковская, Е. В. Чертина // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. – 2015. – № 4. – С. 118–123.

7. Маркин А. В. Метод автоматического восстановления значений в потоках данных на основе взвешенной модели / А. В. Маркин, М. В. Щербаков // Прикаспийский журнал: управление и высокие технологии. – 2013. – № 3 (23). – С. 049-054.

8. Чертина Е. В. Использование алгоритмов Густафсона-Кесселя и FCM для решения задачи многокритериального выбора инновационных ИТ-проектов / Е. В. Чертина // Современные тенденции развития наук и технологии : сборник трудов XVI Международной научно-практической конференции. – Белгород, 2016. – С. 89–93.

9. Чертина Е. В. Принятие решений по инвестированию ИТ-инноваций на основе нечеткой экспертной информации / Е. В. Чертина, Л. Б. Аминул, О. О. Еременко // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. – 2018. – № 1. – С. 103–111.

10. Шуршев В. Ф. Модель системы поддержки принятия решений на основе рассуждений по прецедентам / В. Ф. Шуршев, Г. А. Кочкин, В. Р. Кочкина // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. – 2013. – № 2. – С. 175–183.

11. Юдин В. Н. Неполностью описанные объекты в системах поддержки принятия решений / В. Н. Юдин, Л. Е. Карпов // Программирование. – 2017. – № 5. – С. 24–31.

12. Budylsky A. V. Using coevolution genetic algorithm with Pareto principles to solve project scheduling problem under duration and cost constraints / A. V. Budylsky, I. Yu. Kvyatkovskaya // Journal of Information and Organizational Sciences. – 2014. – Vol. 38, iss. 1. – P. 1–9.

13. Kosmacheva I. Algorithms of Ranking and Classification of Software Systems Elements / I. Kosmacheva, I. Kvyatkovskaya, I. Sibikina, Y. Lezhnina // Communications in Computer and Information Science. – 2014. – Vol. 466 CCIS. – P. 400–409.

14. Kvyatkovskaya I. Yu. Methodology of a support of making management decisions for poorly structured problems / I. Yu. Kvyatkovskaya, V. F. Shurshev, M. B. Frenkel // Communications in Computer and Information Science. – 2015. – Vol. 535. – P. 278–291.

15. Little R. J. A., Rubin D. B. Statistical Analysis with Missing Data / R. J. A. Little, D. B. Rubin. – New York : John Wiley & Sons, 1987.

References

1. Brumshteyn Yu. M., Dyudikov I. A. Modeli optimizatsii podbora resursov pri upravlenii sovokupnostyu proektov s uchetom zavisimosti kachestva rezultatov, riskov, zatrat [Optimization models of resources in the management of the selection of projects according to a set based on the quality of results, risks, costs]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: upravlenie, vychislitelnaya tekhnika i informatika* [Bulletin of the Astrakhan State Technical University. Series: Management, Computer Science and Informatics], 2015, no. 1, pp.78–88.

2. Dyuk V. A. *Metodologiya poiska logicheskih zakonomernostej v predmetnoj oblasti s nechetkoj sistemologiej: na primere kliniko-ehksperimentalnykh issledovaniy* [Methodology of search of logical regularity in a problem domain with fuzzy systemology: On the example of clinic-experimental researches]. St. Petersburg, 2005. 309 p.

3. Karlov I. A. Vosstanovlenie propushchennyh dannykh pri chislennom modelirovanii slozhnykh dinamicheskikh sistem [The missing value estimation in numerical modeling of complex dynamic systems] *Nauchno-tekhnicheskie vedomosti Sankt-Peterburgskogo gosudarstvennogo politekhnicheskogo universiteta. Informatika. Telekommunikatsii. Upravlenie* [St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunication and Control Systems], 2013, no. 6 (186), pp. 137–144.

4. Karpov L. E., Yudin V. N. Metody dobychi dannykh pri postroenie lokalnoy metriki v sistemakh vyvoda po pretsedentam [Data mining methods for composition of local metrics in precedents output systems]. *Preprinty Instituta sistemnogo programmirovaniya RAN* [Preprints of the Institute for System Programming of the Russian Academy of Sciences], 2006, no. 18, pp. 1–42.

5. Kvyatkovskaya A. E., Chertina E. V., Polumordvinova A. O., Kvyatkovskaya I. Yu. Informatsionnaya tekhnologiya poiska kompaniy analogov dlya otsenki stoimosti biznesa ispolzuyushchaya intellektualnykh agentov [Information technology of searching comparative companies for business valuation using intelligent agents] *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitelnaya tekhnika i informatika* [Bulletin of Astrakhan State Technical University. Series: Management, computation and informatics], 2019, no. 1, pp. 99–106.

6. Kvyatkovskaya I. Yu., Chertina E. V. Razrabotka algoritma podbora prioritetnykh venchurnykh IT-proektov [The development of algoritm of selection of the prior IT-projects]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: upravlenie, vychislitelnaya tekhnika i informatika* [Bulletin of Astrakhan State Technical University. Series: Management, Computation and Informatics], 2015, no. 4, pp. 118–123.

7. Markin A. V., Shcherbakov M. V. Metod avtomaticheskogo vosstanovleniya znacheniy v potokakh dannykh na osnove vzveshennoy modeli [A method for data imputation based on weighted backcast models selection] *Prikaspiyskiy zhurnal: upravlenie i vysokie tekhnologii* [Caspian Journal: Control and High Technologies], 2013, no. 3 (23), pp. 49–54.

8. Chertina E. V. Ispolzovanie algoritmov Gustafsona-Kesselya i FCM dlia resheniia zadachi mnogokrite- rialnogo vybora innovatsionnykh IT-proektov [Application of Gustafson-Kessel algorithms for solving a problem of multicriterial choice if innovative IT projects]. *Sovremennye tendentsii razvitiia nauki i tekhnologii : sbornik trudov XVI Mezhdunarodnoy nauchno-prakticheskoy konferentsii* [Modern tendencies of development of science and technologies]. Belgorod, 2016, pp. 89–93.

9. Chertina E. V., Aminul L. B., Eremenko O. O. Prinyatie resheniy po investirovaniyu IT-innovatsiy na osnove nechetkoy ekspertnoy informatsii [Decision-making on investment of IT-innovation based on fuzzy expert information]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitelnaya tekhnika i informatika* [Bulletin of Astrakhan State Technical University. Series: Management, Computation and Informatics], 2018, no. 1, pp. 103–111.

10. Shurshev V. F., Kochkin G. A., Kochkina V. R. Model sistemy podderzhki priniatiya resheniy na osnove rassuzhdeniy po pretsedentam [Model of decision support system using case-based reasoning]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitel'naia tekhnika i informatika* [Bulletin of Astrakhan State Technical University. Series: Management, Computation and Informatics], 2013, no. 2, pp. 175–183.

11. Yudin V. N., Karpov L. E. Nepolnostyu opisannye obekty v sistemakh podderzhki prinyatiya resheniy [Incompletely described objects in decision support]. *Programmirovanie* [Programming and Computer Software], 2017, vol. 43, no. 6, pp. 294–299.

12. Budylsky A. V., Kvyatkovskaya I. Yu. Using coevolution genetic algorithm with Pareto principles to solve project scheduling problem under duration and cost constraints. *Journal of Information and Organizational* Sciences, 2014, vol. 38, iss. 1, pp. 1–9.

13. Kosmacheva I., Kvyatkovskaya I., Sibikina I., Lezhnina Y. Algorithms of Ranking and Classification of Software Systems Elements. *Communications in Computer and Information Science*, 2014, vol. 466 CCIS, pp. 400–409.

14. Kvyatkovskaya, I. Yu., Shurshev V. F., Frenkel M. B. Methodology of a support of making management decisions for poorly structured problems. *Communications in Computer and Information* Science, 2015, vol. 535, pp. 278–291.

15. Little R. J. A., Rubin D. B. *Statistical Analysis with Missing Data*. New York, John Wiley & Sons, 1987.

УДК 004.02:[001.6+002]

АНАЛИЗ РАСПРЕДЕЛЕНИЯ НАУЧНЫХ СПЕЦИАЛЬНОСТЕЙ, ОТНОСЯЩИХСЯ К ОТРАСЛИ «ТЕХНИЧЕСКИЕ НАУКИ», В РОССИЙСКИХ ЖУРНАЛАХ ИЗ СПИСКА ВАК

Статья поступила в редакцию 19.01.2019, в окончательном варианте – 12.02.2019.

Брумштейн Юрий Моисеевич, Астраханский государственный университет, 414056, Российская Федерация, г. Астрахань, ул. Татищева, 20а,

кандидат технических наук, доцент, ORCID http://orcid.org/0000-0002-0016-7295, https://elibrary.ru/author_profile.asp?authorid=280533, e-mail: brum2003@mail.ru

Васильев Никита Вячеславович, Астраханский государственный университет, 414056, Российская Федерация, г. Астрахань, ул. Татищева, 20а,

студент, e-mail: nikivas97@mail.ru

Коновалова Дарья Игоревна, Астраханский государственный университет, 414056, Российская Федерация, г. Астрахань, ул. Татищева, 20а,

студент, e-mail: haizbq@mail.ru

С января 2019 г. для российских научных журналов, включенных в список изданий, рекомендованных ВАК России для публикации результатов кандидатских и докторских диссертаций, вместо групп научных специальностей утверждаются конкретные специальности. В статье обсуждены вероятные последствия такого изменения для аспирантов и докторантов, других авторов научных статей, редакций изданий (включая главных редакторов и ответственных секретарей), издателей. Представлены методики автоматизированного анализа списка журналов, включенных в список ВАК по различным направлениям, представляющим практический интерес для указанных выше категорий физических и юридических лиц. Оценено количество журналов, для которых ВАК утверждены научные специальности, отнесенные к отрасли «Технические науки». Эти журналы разбиты на две категории: (**Т**) «чисто технические», в которых используются только специальности, отнесенные к отрасли «технические науки»; (**С**) издания, в которых дополнительно указываются те же специальности, отнесенные к отрасли «физико-математические науки». Для изданий категории «Т» получено процентное распределение журналов по количествам используемых в них научных специальностей. Для журналов категории «С» – распределение в отношении долей специальностей, отнесенных к отрасли «технические науки», по отношению к общему количеству специальностей, утвержденных для соответствующих изданий. Для всех специальностей, относящихся по классификации ВАК к отрасли «технические науки», представлены таблицы с количествами журналов из списка ВАК, в которых эти специальности встречаются. Для некоторых групп специальностей, относящихся к «техническим наукам», даны оценки совместной встречаемости отдельных специальностей по совокупностям рассмотренных в данной статье журналов. На основании представленного в статье материала сделаны выводы об относительной «востребованности» научных специальностей в изданиях, включенных в список ВАК, опубликованный в начале января 2019 г.

Ключевые слова: журналы из перечня ВАК, тематика журналов, отрасли науки, технические науки, физико-математические науки, научные специальности, группы специальностей, частоты встречаемости специальностей, статистическая обработка, распределения журналов, распределение специальностей