



УДК 004.451.53

DOI 10.18413/2411-3808-2018-45-3-566-573

ПРИМЕНЕНИЕ ПРОГРАММНОГО КОМПЛЕКСА ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ ОБРАБОТКИ ЭКОНОМИЧЕСКОЙ ИНФОРМАЦИИ

THE APPLICATION OF SOFTWARE COMPLEX INTELLECTUAL DATA ANALYSIS FOR PROCESSING OF ECONOMIC INFORMATION

Ю.Г. Шаталова

J.G. Shatalova

Севастопольский экономико-гуманитарный институт (филиал) Крымского федерального университета им В.И. Вернадского, Россия, 299028, г. Севастополь, ул. Шелкунова, 1

Sevastopol economic-humanitarian Institute (branch) of the Crimean Federal University,
1 Shelkunova St, Sevastopol, 299028, Russia

E-mail: bastion2417@mail.ru

Аннотация

В статье рассматривается программный комплекс интеллектуального анализа данных, предназначенный для обработки результатов научных исследований в экономике. Разрабатываемый комплекс может быть применен для обучения студентов экономических специальностей. Особый интерес представляет комплекс для удаленного интерактивного обучения. Объединение нескольких методов Data Mining в один программный пакет облегчает процесс анализа данных, позволяет исследователю сравнивать результаты, полученные путем применения различных методов, и систематизировать выводы. Также комплекс предусматривает возможность визуализации данных и результатов.

Abstract

The article discusses a software complex of intellectual analysis of data is intended for processing the results of scientific research in Economics. The developed complex can be used for training students of economic specialties. Of particular interest is the complex for remote interactive learning. Combining multiple Data Mining methods in a single software package facilitates the process of data analysis allows the researcher to compare the results obtained by different methods, and to systematize findings. Also the system provides visualization of data and results implemented methods for rendering two-dimensional and multidimensional data. The software package is based on JAVA platform, it has a multimodular structure. Due to the modular approach, the system can easily scales: it is possible to fill it with new modules that implement additional methods of Data Mining. The complex implements the following methods: regression analysis, methods of mathematical programming, genetic algorithms, neural networks.

Ключевые слова: интеллектуальный анализ данных, методы Data Mining, нейронные сети, математическое программирование, регрессионный анализ, генетические алгоритмы.

Keywords: Data Mining methods, neural networks, mathematical programming, regression analysis, genetic algorithms.

Введение

При проведении экономических исследований часто приходится использовать математические методы обработки информации. Например, при исследовании предприятий с целью прогнозирования их дальнейшего развития необходимо построить математиче-



скую модель и провести ее анализ; при статистическом сборе данных по отрасли данные необходимо обработать и проанализировать и т.д. [Асадуллаев и др., 2017]. В арсенале исследователя – разнообразный математический аппарат [Гаврилова и др., 2016], но обработка большого объема данных предполагает применение компьютерной техники. Современная вычислительная техника позволяет выйти на качественно новый уровень обработки данных. Все чаще используется термин «интеллектуальная обработка информации». Интеллектуальная обработка предполагает не только использование нового поколения инструментальных средств, но и нового поколения математического, алгоритмического и программного обеспечения [Путивцева и др., 2015]. С интеллектуальной обработкой данных связывают понятие Data Mining (поиск ценной информации в большой базе данных (data) и добыча горной руды (mining)) [Марманис, Бабенко, 2011].

Для многих математических и статистических методов разработаны программные инструменты [Дюк и др., 2011]; количество их велико, а для всеобъемлющего исследования, возможно, потребуется не один из них. Поэтому исследователю нужно изучить широкий спектр программ.

К сожалению, сегодня нет универсальной программы, обеспечивающей полномасштабную интеллектуальную обработку данных.

В силу сказанного, создание программного комплекса, реализующего средства Data Mining, является актуальной задачей.

Постановка задачи

Разработать программный комплекс, реализующий методы Data Mining и средства визуализации данных. Комплекс должен иметь удобный, понятный пользовательский интерфейс, модульную структуру и предоставлять пользователю возможность выбора методов анализа.

Описание комплекса

В данной работе рассматривается программный комплекс интеллектуальной обработки данных. Цель создания комплекса – предоставить пользователю возможность проанализировать интересующие его данные с помощью нескольких методов Data Mining, а также визуализировать данные или результаты.

В технологию Data Mining включены методы классификации и кластеризации, моделирования и прогнозирования, генетические и эволюционные алгоритмы, методы прикладной статистики, которые составляют фундамент Data Mining, прежде всего это корреляционный и регрессионный анализ, факторный и дискриминантный анализ и многое другое [Семенкин, Ефимов, 2006]. В этой технологии гармонично объединились строго формализованные методы и методы неформального анализа, т.е. количественный и качественный анализ данных. Поэтому исследователь сможет оперировать как традиционными статистическими и математическими методами, так и методами анализа визуальной информации.

При обработке данных экономических исследований одной из задач является построение математической модели предприятия или системы объектов и ее исследование [Бегунов и др., 2010]. Математическая модель в зависимости от задач исследования может быть использована для разных целей: для предметно-семантического анализа объекта или явления, прогнозирования состояния объекта в разных условиях функционирования, оптимизации отдельных параметров и для решения других специфических задач [Кравченко, 2012]. Конечной целью любой обработки данных является выдвижение гипотез о классе и структуре математической модели исследуемого явления, определение параметров функций, анализ выполнения основных предпосылок, лежащих в основе модели и других.

Для достижения этих целей необходимо решить некоторые частные задачи, среди которых можно выделить следующие:

- 1) выдвижение и проверка гипотезы о функциональной зависимости между исследуемыми факторами;
- 2) оценка параметров и числовых характеристик наблюдаемых случайных величин или процессов;
- 3) группировка исходной информации при большом объеме обрабатываемых данных;
- 4) объединение нескольких групп измерений, полученных, возможно, в различное время или в различных условиях, для совместной обработки [Холод, 2015];
- 5) выявление скрытых связей и взаимозависимости различных измеряемых факторов и результирующих переменных;
- 6) оптимизация функций;
- 7) классификация объектов или явлений на основании полученных результатов [Суркова, Буденков, 2012];
- 8) визуализация исходных данных и/или результатов.

Рассматриваемый программный комплекс базируется на JAWA-платформе, имеет многомодульную структуру. Благодаря модульному подходу комплекс легко масштабируется: есть возможность пополнять его новыми модулями, реализующими дополнительные методы Data Mining.

Одной из проблем при проектировании программного комплекса оказался выбор методов интеллектуальной обработки данных, которые предполагалось реализовать. Перечень методов Data Mining огромен [Холод, 2011]. Поскольку экономические исследования предполагают математическую и статистическую обработку данных, автором были выбраны следующие методы: регрессионный анализ, методы математического программирования [Черняк, 2014], генетические алгоритмы, нейронные сети [Замятин, 2006], визуализация двумерных и многомерных данных [Тиндова, 2008].

Выбор инструментария определяется пользователем. На рисунке 1 представлен главный интерфейс комплекса. При нажатии соответствующей кнопки запускается модуль выбранного метода.

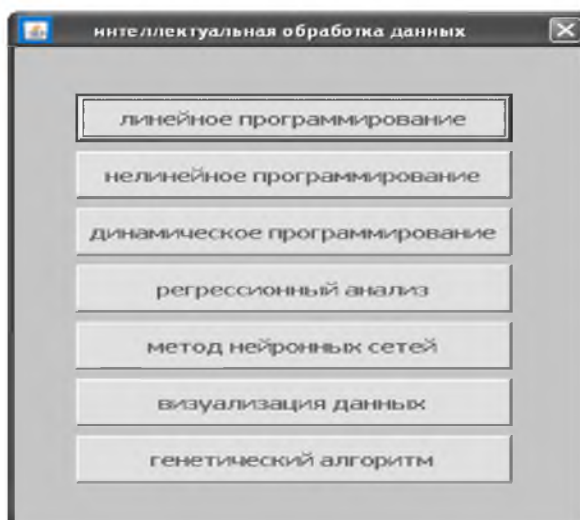


Рис. 1. Интерфейс программного комплекса
Fig. 1. Interface of the software complex

В зависимости от решаемой задачи исследователь может воспользоваться соответствующим набором инструментов. Например, задача, направленная на построение

и проверку адекватности математической модели исследуемому явлению, связана с проведением регрессионного анализа.

Рассмотрим следующий пример: пусть требуется построить математическую модель с целью исследования зависимости прибыли предприятия от вложений капитала в него. Данные наблюдений за восемь лет приведены в таблице.

Таблица
Table

Данные наблюдений за восемь лет
Observational data for eight years

Номер наблюдения	Вложение капитала в предприятие, тыс. руб., x	Прибыль с капитала в течение нескольких лет, тыс. руб., y
1	8	13
2	10	14
3	15	15,4
4	20	16,3
5	30	17,2
6	40	17,8
7	60	18,5
8	80	18,8

Для построения зависимости прибыли от капиталовложений можно использовать методы регрессионного анализа.

В рассматриваемом комплексе реализованы парный линейный и нелинейный регрессионный анализ, а также множественный регрессионный анализ. Пользователь вводит данные, может увидеть диаграмму рассеивания, установить линию тренда, принять решение о соответствии модели выдвинутой гипотезе. В качестве метода определения параметров используется метод наименьших квадратов.

Пример использования парного нелинейного регрессионного анализа для данной задачи приведен на рисунке 2.

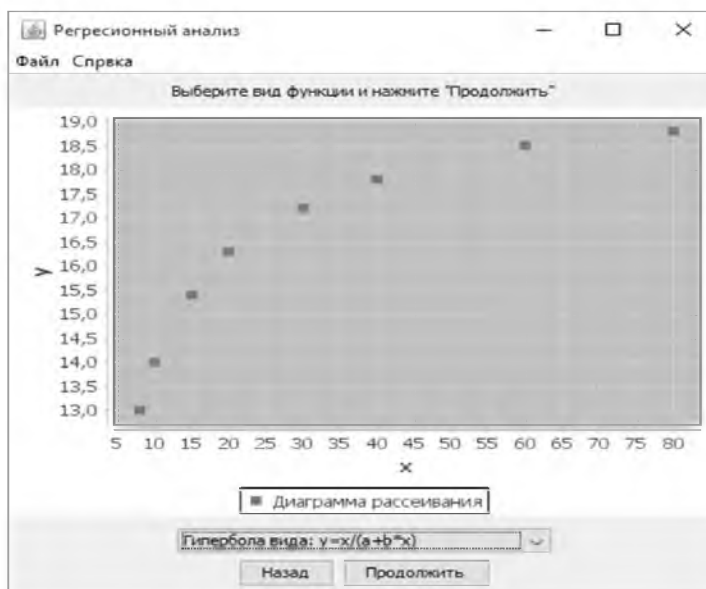


Рис. 2. Пример работы модуля регрессионного анализа
Fig. 2. Example of regression analysis module operation

Проведение регрессионного анализа позволило выявить гиперболическую зависимость между исследуемыми факторами. Дальнейший анализ позволил рассчитать параметры модели. Для рассматриваемого примера модель имеет вид:

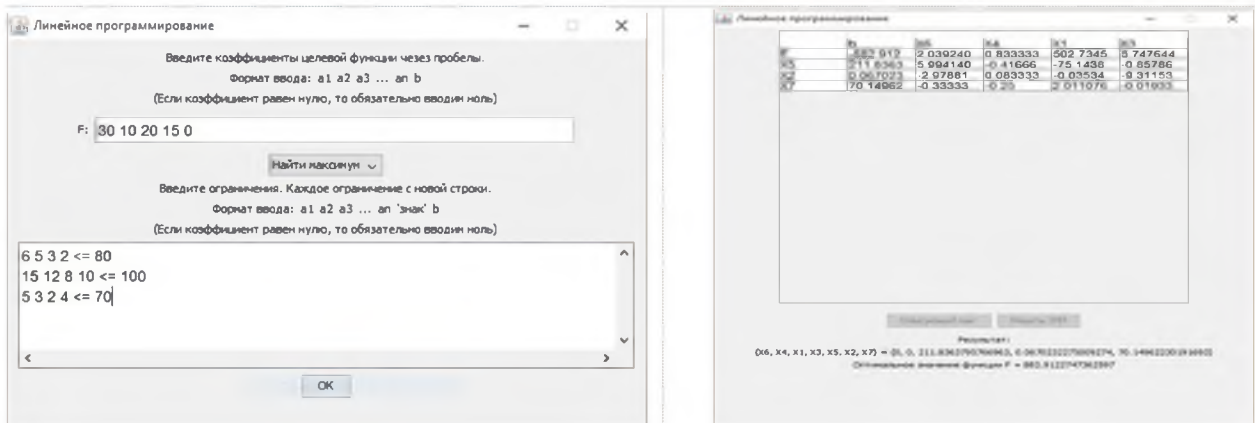
$$y = \frac{x}{13,9 + 0,07 \cdot x}$$

Исследование построенной модели позволило сделать вывод, что вкладывать в предприятие более 50000 руб. нецелесообразно, так как большие вложения приносят меньший эффект.

Решение задач оптимизации функций можно осуществить с помощью методов математического программирования [Акулич, 2011]. Для реализации методов математического программирования были реализованы методы линейного, нелинейного и динамического программирования. В основе реализации метода линейного программирования лежит симплекс-метод, в основе нелинейного анализа – метод множителей Лагранжа.

Для реализации методов динамического программирования был выбран двухэтапный алгоритм последовательных приближений, основанный на определении условных оптимальных решений по функциональному уравнению Беллмана [Гладких, 2010].

Пример использования линейного программирования приведен на рисунке 3, нелинейного – на рисунке 4.



а) ввод исходных данных

б) Симплекс-таблица и результат вычислений

Рис. 3. Пример работы метода линейного программирования

Fig. 3. An example of the method of linear programming

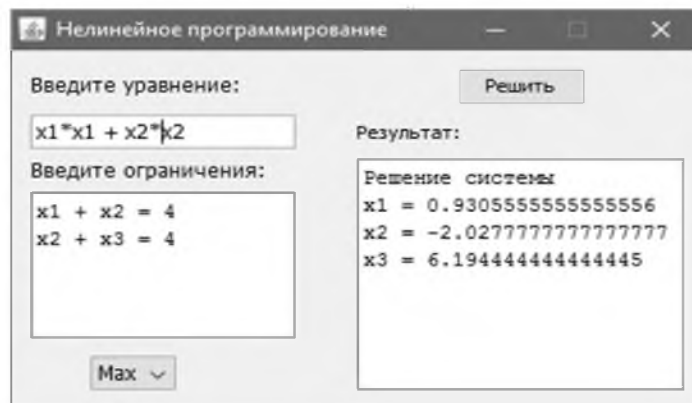


Рис. 4. Пример работы метода нелинейного программирования

Fig. 4. An example of the method of nonlinear programming

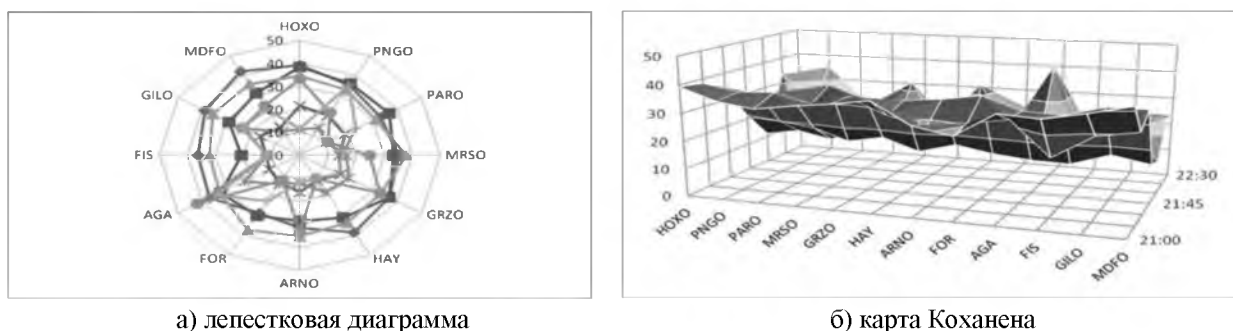
Задачи оптимизации могут быть решены с помощью генетических алгоритмов и нейронных сетей. В качестве генетического алгоритма реализован канонический алгоритм (целочисленное кодирование, битовая мутация, одноточечный оператор кроссинговера). Для реализации метода нейронных сетей [Ефимов, Якимов, 2009] была выбрана многослойная сеть прямого распространения с одним промежуточным слоем.

Помимо применения отдельных методов исследования пользователю предоставляется возможность применить несколько методов к одному и тому же набору данных [Лю, 2005].

Этот подход позволяет рассматривать классификацию данных по различным критериям, выполнять сравнительный анализ результатов и, если надо, визуализировать данные или результаты анализа. Кроме того, у исследователя появляется возможность выбора наиболее подходящего для конкретной задачи метода анализа.

Для наглядного представления полученных результатов можно использовать средства построения двумерных графиков и диаграмм. Выбор вида диаграмм определяется пользователем. Достаточно нажать кнопку «Визуализация данных» в окне главного интерфейса, и пользователь сможет оперировать графическим инструментарием программного комплекса. Также предусмотрена возможность перехода к визуализации данных из каждого выбранного метода.

Особенный интерес представляет многомерная визуализация данных. Подобная визуализация может использоваться не только для наглядного представления результатов, но и для отображения кластеризации и классификации данных. В рассматриваемом комплексе визуализация представлена в виде лепестковых диаграмм и карт Коханена [Манжула, Федяшов, 2011]. Примеры визуализации приведены на рисунке 5.



а) лепестковая диаграмма

б) карта Коханена

Рис. 5. Пример визуализации многомерных данных
 Fig. 5. Example of visualization of multidimensional data

Разработанный комплекс может применяться в процессе обучения студентов экономических специальностей [Пименов, Ипатов, 2009]. Комплекс подходит как для аудиторных занятий, так и для дистанционного обучения по таким дисциплинам, как эконометрия, статистика и т.д. [Щеголькова, 2012]. Студент может выбрать наиболее подходящий метод для решения поставленной задачи, построить математическую модель и провести ее анализ.

Заключение

Таким образом, программный комплекс интеллектуальной обработки данных, отвечающий требованиям поставленной задачи, был разработан. В дальнейшей работе над программным комплексом предполагается пополнить набор реализуемых методов. Планируется добавить методы обработки данных в базах данных и методы нечеткой логики [Вахитов, Силич, 2010]. Также планируется доработать интерфейс системы и дополнить методы визуализации данных. В частности, предполагается добавить построение гиперболических деревьев и кубов Кодда.

Список литературы References

1. Акулич И.Л. 2011. Математическое программирование в примерах и задачах: учебное пособие. СПб., Издательство «Лань», 352.
 Akulich I.L. 2011. Mathematical programming examples and problems: a tutorial. SPb., Lan Publishing House, 352. (in Russian)



2. Асадуллаев Р.Г., Ломакин В.В., Белоконь Ю.Ю., Зайцева Т.В., Резниченко О.С. 2017. Модели и средства поддержки принятия решений в системах переподготовки кадров предприятия. *Научные ведомости Белгородского государственного университета. Экономика. Информатика.* 23 (272), вып. 44: 148–158.

Asadullaev R.G., Lomakin V.V., Belokon Y.Y., Zaitseva T.V., Reznichenko O.S. 2017. Models and methods of decision support in the information systems for requalification of employees. *Belgorod State University Scientific Bulletin. Economics. Information technologies.* 23 (272), Issue 44: 148–158.

3. Бегунов Н.А., Клебанов Б.И., Рапопорт И.А. 2010. Объединение подходов интеллектуального анализа данных и имитационного моделирования для прогнозирования доходов бюджета. *Автоматизация и современные технологии*, 12: 37–40.

Begunov N.A., Klebanov B.I., Rapoport I.A. 2010. Combining the approaches of data mining and simulation modeling to predict future revenues. *Automation and modern technologies*, 12: 37–40. (in Russian)

4. Вахитов А.Р., Силич В.А. 2010. Использование нечеткого логического вывода для интеллектуального анализа данных. *Известия Томского политехнического университета*, 317 (5): 171–174.

Vakhitov A.R., Silich V.A. 2010. The use of fuzzy inference for data mining. *Proceedings of Tomsk polytechnic university*, 317 (5): 171–174. (in Russian)

5. Гаврилова Т.А., Кудрявцев Д.В., Муромцев Д.И. 2016. Инженерия знаний. Модели и методы. М., Издательство «Лань», 324.

Gavrilova T.A., Kudryavtsev, D.V., Muromtsev D.I. 2016. Knowledge Engineering. Models and methods. M., Publishing House "LAN", 324. (in Russian)

6. Гладких Н.А. 2010. Оптимизация систем электронного документооборота на основе интеллектуального анализа данных. *В мире научных открытий*, 4 (11): 122–123.

Gladkich N.A. 2010. Optimization of electronic document management systems based on data mining. *In the world of scientific discoveries*, 4 (11):122–123. (in Russian)

7. Дюк В.А., Флегонтов А.В., Фомина И.К. 2011. Применение технологий интеллектуального анализа данных в естественнонаучных, технических и гуманитарных областях. *Известия российского государственного педагогического университета им. А.И. Герцена*, 138: 77.

Duke V.A., Flegontov A.V., Fomin I.K. 2011. The application of data mining technology in scientific, technical and humanitarian fields. *Bulletin of the Russian State Pedagogical University. A.I. Herzen.* (138): 77. (in Russian)

8. Ефимов С.Н., Якимов Л.С. 2009. Применение технологии нейронных сетей для интеллектуального анализа данных. *Сборник научных трудов по материалам международной научно-практической конференции.* 2 (2): 64–70.

Efimov S.N., Yakimov L.S. 2009. Application of neural network technology for data mining. *Collection of scientific works on the materials of the international scientific-practical conference.* 2(2): 64–70. (in Russian)

9. Замятин Н.В., Максимов В.П., Платонов Н.В., Тарасевич М.Н. 2006. Информационная нейросетевая система для интеллектуального анализа данных. *Известия томского политехнического университета.* 309(8): 107–112.

Zamyatin N.In. Maksimov V.P., Platonov N.V. Tarasevich M.N. 2006. Information neural network system for data mining. *Proceedings of the Tomsk polytechnic university.* 309. (8):107–112. (in Russian)

10. Кравченко Ю.А. 2012. Имитационная модель анализа данных в интеллектуальных информационных системах. *Известия кабардино-балкарского научного центра рн.* (1): 25–31.

Kravchenko Y.A. 2012. A simulation model for data analysis in intelligent information systems. *Proceedings of the kabardino-balkar scientific center of ras.* (1): 25–31. (in Russian)

11. Лю Б. 2005. Теория и практика неопределенного программирования. Пер. с англ. М. Бином, 416.

Liu B. Theory and practice of indefinite programming. Per. With English, Moscow, Binom, 416. (in Russian)

12. Манжула В.Г., Федяшов Д.С. 2011. Нейронные сети кохонена и нечеткие нейронные сети в интеллектуальном анализе данных. *Фундаментальные исследования.* (4): 108–114.

Manzhula V.G., Fedyashov D.S. 2011. Kohonen neural network and fuzzy neural network in data mining. *Fundamental research.* (4): 108–114. (in Russian)



13. Марманис Х., Бабенко Д. 2011. Алгоритмы интеллектуального интернета: пер. с англ. СПб., Символ-плюс, 480.
Marmanis X., Babenko D. 2011. Algoritmy intellektual'nogo interneta: per. S angl. Spb., simvol-pljus, 480. (in Russian)
14. Пименов В.И., Ипатов О.С. 2009. Разработка обучающих систем по дисциплинам технологического цикла на основе методов интеллектуального анализа данных. Вестник компьютерных и информационных технологий, (9): 38–44.
Pimenov, V.I., Ipatov O.S. 2009. Development of training systems in the disciplines of technological cycle-based data mining techniques. Bulletin of computer and information technologies, (9): 38–44. (in Russian)
15. Путивцева Н.П., Зайцева Т.В., Игрунова С.В., Нестерова Е.В., Пусная О.П. 2015. О разработке пакета компьютерной поддержки принятия решений для выбора корреляционно-регрессионных моделей анализа и прогнозирования эмпирических данных. Научные ведомости Белгородского государственного университета. Экономика. Информатика, 19 (216), вып. 36/1: 126–132.
Putivzeva N.P., Zaitseva T.V., Igrunova S.V., Nesterova E.V., Pusnaya O.P. 2015. On the elaboration of a package of computer decision support for the choice of correlation and regression models of analysis and forecast of empirical data. Belgorod State University Scientific Bulletin. Economics. Information technologies. 19 (216), Issue 36/1: 126–132.
16. Семенкин Е.С., Ефимов С.Н. 2006. Модели и алгоритмы автоматизации проектирования специализированных многопроцессорных вычислительных систем интеллектуального анализа данных в режиме реального времени. Вестник Сибирского государственного аэрокосмического университета им. академика М.Ф. Решетнева, 6 (13): 17–21.
Semenkin E.S., Efimov S.N. 2006. Models and algorithms of automated design of specialized multiprocessor computing systems data mining in real-time. Bulletin of the Siberian state aerospace university named after academician M.F. Reshetnev, 6 (13): 17–21. (in Russian)
17. Суркова А.С., Буденков С.С. 2012. Построение модели и алгоритма кластеризации в интеллектуальном анализе данных. Вестник Нижегородского университета им. Н.И. Лобачевского, 2 (1): 198–202.
Surkov A.S., Butenkov S.S. 2012. Model building and clustering algorithm in data mining. Vestnik Nizhegorodskogo universiteta im. N. Lobachevsky, 2 (1): 198–202. (in Russian)
18. Тиндова М.Г. 2008. Предварительная кластеризация многомерных объектов в интеллектуальном анализе данных. Вестник саратовского государственного социально-экономического университета, 4: 137–138.
Tindova M.G. 2008. Preliminary multivariate clustering of objects in data mining. Bulletin of saratov state socio-economic university, 4: 137–138. (in Russian)
19. Холод И.И., Батасова С.Я., Семенченко А.С., Ефимова М.С. 2015. Подготовка распределенных разнородных данных к интеллектуальному анализу данных. В кн.: Международная конференция по мягким вычислениям и измерениям. Изд-во Санкт-Петербургского государственного электротехнического университета им. В.И. Ульянова (Ленина), 117–120.
Kholod I.I., Butasova J.S., Semenchenko A.S., Efimova M.S. 2015. Preparation of distributed heterogeneous data mining. International conference on soft computing and measurements. In: Publisher Saint-Petersburg state electrotechnical university named after V.I. Ulyanov (Lenin), 117–120. (in Russian)
20. Холод И.И. 2011. Архитектура системы разработки алгоритмов интеллектуального анализа распределенных данных. Известия СПбГЭТУ ЛЭТИ, 9: 448–55.
Kholod I.I. 2011. The system architecture is the development of algorithms for mining distributed data. Izvestiya SPBGU "LETI", 9: 448–55. (in Russian)
21. Черняк А.А., Черняк Ж.А., Метельский Ю.М. 2014. Математическое программирование. Алгоритмический подход: учебное пособие. Издательство «Высшая школа», 352.
Chemyak and A.A., Chernyak Zh.A., Metelsky Yu.M. 2014. The Mathematical programming. Algorithmic approach: textbook. Publishing house "Higher school", 352. (in Russian)
22. Щеголькова В.А. 2012. Определение стереотипа обучаемого на основе методов интеллектуального анализа данных. Сборник научных трудов sword, 4 (1): 87–88.
Shchegolkova V.A. 2012. The definition of the stereotype of the student-based data mining techniques. Collection of scientific works of sword, 4 (1): 87–88. (in Russian)