

Раздел II. Анализ данных и управление знаниями

УДК 681.3.06:378.1

DOI 10.23683/2311-3103-2019-4-81-89

Э.В. Кулиев, Э.С. Цырульникова, Н.В. Кулиева, В.В. Марков

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА УПРАВЛЕНИЯ ЗНАНИЯМИ В ПРОЦЕССАХ ОБУЧЕНИЯ*

Предложена интеллектуальная информационная система с целью поддержки качества образовательной деятельности (ИИС ПКОД). ИИС ПКОД включает в себя модуль «Экспертной оценки качества» и модуль «Анализа и прогнозирования результатов процесса обучения». ИИС ПКОД решает задачи анализа и прогнозирования результатов процесса обучения, используя расширенный набор признаков: данные о посещении учебных и тестовых ресурсов, данные об участии в дискуссионных форумах, данные о просмотре анонсов, данные об учебном курсе, данные об уровне подготовки обучаемого и данные экспертных оценок качества учебно-методических и тестовых материалов. Набор информативных признаков определен с помощью фильтровочного алгоритма и считается оптимальным для построения модели прогноза результатов процесса обучения. Экспертная оценка учебно-методического обеспечения осуществляется по критериям полноты, доступности и актуальности. Экспертная оценка тестовых материалов осуществляется по показателям: трудности, валидности и надежности теста. Экспертные оценки вносятся в специальные формы. Структура форм устроена таким образом, что при заполнении форм на третьем уровне, формы второго и первого уровня могут заполняться автоматически, что облегчает задачу ввода данных. Для построения модели прогноза используется алгоритм классификации: бустинг над решающими деревьями. В ходе вычислительных экспериментов применение алгоритма бустинг над решающими деревьями и расширенного набора признаков показали хорошее качество прогноза по сравнению с аналогичными разработками. Разработанная ИИС ПКОД дает качественный прогноз результатов процесса обучения и возможность эффективно управлять качеством учебного процесса, выявлять проблемные ситуации и совершенствовать информационно-методические материалы. Проведены экспериментальные исследования.

Интеллектуальный анализ данных; прогнозирование; алгоритмы классификации; бустинг; решающие деревья; отбор информативных признаков; результаты процесса обучения; экспертная оценка качества; учебно-методические материалы; тестовые материалы.

E.V. Kuliev, E.S. Tsyrunnikova, N.V. Kulieva, V.V. Markov

INTELLECTUAL KNOWLEDGE MANAGEMENT SYSTEM IN THE PROCESS OF TRAINING

In this work, an intelligent information system of supporting the quality of educational activities (IIS SQEA) is proposed. IIS SQEA includes a module "Expert quality assessment" and a module "Analysis and prediction of the results of the learning process." IIS SQEA solves the problem of analyzing and forecasting the results of the learning process using an extended set of features: data on attendance of training and test resources, data on participation in discussion forums, data on viewing announcements, data on a training course, data on the level of student training and expert evaluations quality teaching materials and test materials. The set of informative features is

* Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов №№ 19-07-00099, 18-07-00055.

determined using a filtering algorithm and is considered optimal for building a model for predicting the results of the learning process. Expert evaluation of educational and methodological support is carried out according to the criteria of completeness, availability and relevance. Expert assessment of test materials is carried out on indicators: difficulties, validity and reliability of the test. Expert assessments are made in special forms. The structure of the forms is arranged in such a way that when filling out forms on the third level, forms of the second and first levels can be filled in automatically, which simplifies the task of data entry. To build a forecast model, a classification algorithm is used: boosting over decision trees. In the course of computational experiments, the application of the algorithm for boosting over decisive trees and an extended set of features showed a good forecast quality compared with similar developments. The developed IIS SQEA provides a qualitative forecast of the results of the learning process and the ability to effectively manage the quality of the educational process, identify problem situations and improve information and teaching materials. Experimental studies have been carried out.

Data mining; forecasting; classification algorithms; boosting; decision trees; selection of informative features; results of the learning process; expert assessment of quality; teaching materials; test materials.

Введение. Современный этап развития общества характеризуется высоким уровнем развития информационных технологий. Особенно интенсивно развивается сфера образовательной деятельности [1]. Разработка интеллектуальных информационных систем обучения является актуальной задачей, предоставляющей широкий спектр возможностей для научной деятельности. Исследования в данной области ведутся в различных направлениях, в том числе в совершенствовании процедур управления качеством учебного процесса с помощью интеллектуальных информационных систем (ИИС). Интеллектуализация систем обучения может быть достигнута путем использования различных современных методов и технологий искусственного интеллекта, например, интеллектуального анализа данных [3], нечетких математических моделей [4–5], технологий адаптации [6] и др.

В данной работе предложена ИИС поддержки качества образовательной деятельности (ИИС ПКОД), имеющая нестандартный набор функций и дополняющая возможности многих систем дистанционного обучения (СДО). ИИС ПКОД решает задачи управления качеством образовательного процесса, основываясь на результатах интеллектуального анализа данных.

1. ИИС поддержки качества образовательной деятельности. Одним из наиболее продуктивных способов повышения качества образовательной деятельности является прогнозирование результатов процесса обучения и принятие на его основе превентивных мер. Прогнозирование результатов процесса обучения считается одним из приоритетных направлений в разработке интеллектуальных систем обучения [7]. Сфера таких исследований весьма разнообразна и получила освещение в множестве научных работ [2–9].

Согласно результатам ряда научных работ [7, 10] – существует тесная связь между востребованностью образовательных ресурсов и успеваемостью обучающихся. Востребованность образовательных ресурсов определяется продолжительностью посещения страниц с учебными материалами, новостных страниц и дискуссионных форумов. Повышение точности модели прогноза с использованием признаков, характеризующих востребованность образовательных ресурсов, составляет более 22 % [10]. Вместе с тем востребованность образовательных ресурсов связана с их качеством. Таким образом, качество образовательных ресурсов влияет на прогноз результатов учебной деятельности. В основе данной работы предложена модель прогноза с расширенным набором признаков, включающим экспертные оценки качества учебно-методических и тестовых материалов. Использование расширенного набора признаков повысит информативность и точность прогноза, и позволит предупреждать проблемные ситуации и управлять качеством образовательного процесса.

На основании проведенного анализа был определен состав, решаемых ИИС ПКОД задач. ИИС ПКОД включает в себя модуль «Экспертной оценки качества» и модуль «Анализа и прогнозирования результатов процесса обучения». Ввиду интероперабельности разработанной системы, возможно совместное использование со многими существующими системами дистанционного обучения. На рис. 1. представлена структурная схема ИИС ПКОД.

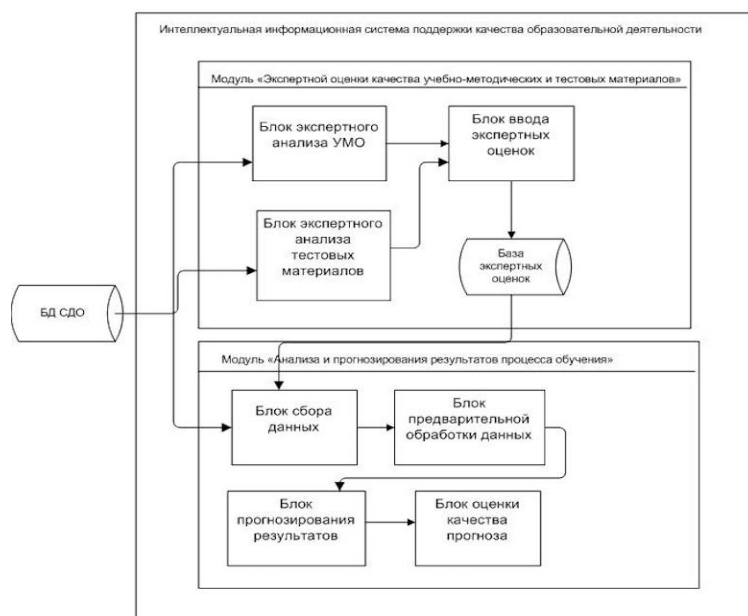


Рис. 1. Функциональная модель ИС ПОД

2. Модуль «Экспертной оценки качества» решает задачи оценки качества учебно-методических и тестовых материалов. Экспертный анализ качества учебно-методического обеспечения осуществляется по критериям полноты, доступности и актуальности.

Мера полноты УМО определяется следующим образом:

$$P = V_{\text{вост}} / V_{\text{общ}}, \quad (1)$$

где P – полнота; $V_{\text{вост}}$ – объем рассмотренной информации; $V_{\text{общ}}$ – общий объем информации в курсе.

Доступность УМО определяет оценку его адекватной интерпретации и возможности получения необходимых материалов. Доступность УМО определяется следующим образом:

$$D = D_{\text{дост}} * D_{\text{инт}}, \quad (2)$$

где D – доступность; $D_{\text{дост}}$ – возможность доступа; $D_{\text{инт}}$ – возможность адекватной интерпретации.

Актуальность УМО отражает соответствие текущему моменту времени. Актуальность можно оценить следующим образом:

$$A = V_{\text{соотв}} / V_{\text{общ}}, \quad (3)$$

где A – актуальность образовательных ресурсов; $V_{\text{соотв}}$ – объем информации, соответствующий текущему моменту времени; $V_{\text{общ}}$ – общий объем информации. На основе приведенных формул рассчитывается обобщенная оценка качества ИМО.

Следующий этап осуществляет оценку тестовых материалов по различным показателям: валидность и надежность теста, показатели качества ТЗ в соответствии с СТТ и ИРТ. В качестве исходных данных используется матрица тестирования $\{x_{ij}\}$, имеющая n строк и m столбцов, где n – количество испытуемых, m – количество тестовых заданий, x_{ij} – оценка успешности выполнения i -м испытуемым j -го задания, представленная в виде числа.

Основной статистической характеристикой тестового задания (ТЗ) является его трудность (p_j), которая, в соответствии с СТТ, рассчитывается по формуле 4.

$$p_j = \sum_{i=1,n} x_{ij} / n. \quad (4)$$

Кроме трудности ТЗ СТТ предоставляет формулы для расчета дисперсии баллов испытуемых, коэффициентов корреляции ТЗ с тестом и между собой, а также различающую (дифференцирующую) способность [14].

Основным отличием ИРТ является методология латентно-структурного анализа, в соответствии с которой способность к обучению невозможно измерить непосредственно. Для расчета начального уровня трудности ТЗ (β_j^0) ИРТ использует формулу 5.

$$\beta_j^0 = \ln (q_j / p_j), \quad (5)$$

где q_j – доля неверных ответов на ТЗ, вычисляемая как $(1 - p_j)$; p_j – доля верных ответов на j -ое ТЗ, равная отношению числа испытуемых, правильно ответивших на данное ТЗ, к общему числу испытуемых.

Далее рассчитывается исходный уровень подготовленности испытуемого (θ_i^0) по формуле 6.

$$\theta_i^0 = \ln (p_i / q_i), \quad (6)$$

где p_i – доля верных ответов i -го испытуемого, равная отношению индивидуально-го балла данного испытуемого к числу заданий в тесте; q_i – доля неверных ответов i -го испытуемого, вычисляемая как $(1 - p_i)$.

Далее формулы (5) и (6) приводят к единой интервальной шкале по формулам 7 и 8.

$$\theta_i = a_\theta \times \theta_i^0 + \beta_{cp}, \quad (7)$$

$$\beta_j = a_\beta \times \beta_j^0 + \theta_{cp}, \quad (8)$$

где a_θ и a_β – угловые коэффициенты необходимые для перехода к единой шкале измерений, вычисляемые на основе стандартных отклонений величин θ_i и β_j ; θ_{cp} и β_{cp} – средние значения величин θ_i и β_j соответственно, β_j – итоговый уровень трудности j -го задания [14].

ИРТ включает следующие модели педагогического измерения [15]: однопараметрическая (модель Раша), двухпараметрическая, трёхпараметрическая.

Валидность теста (Phi) рассчитывается на основе метода «известных групп» по формуле 9.

$$Phi = (a \times d - b \times c) / ((a + b)(c + d)(a + c)(b + d))^{1/2}, \quad (9)$$

где a – число испытуемых из «высокой» группы, хорошо справившихся с тестом; b – число испытуемых из «низкой» группы, хорошо справившихся с тестом; c – число испытуемых из «высокой» группы, плохо справившихся с тестом; d – число испытуемых из «низкой» группы, плохо справившихся с тестом.

Надежность теста можно рассчитать на основании результатов первой и второй процедуры тестирования – X и Y соответственно, по формуле 10.

$$Rt = Cov(X, Y) / (S_x \times S_y), \quad (10)$$

где $Cov(X, Y)$ – ковариация двух переменных X и Y ; S_x и S_y – стандартные отклонения X и Y соответственно. $Cov(X, Y)$ можно рассчитать по формуле 11.

$$Cov(X,Y) = (\sum_{i=1,n} (X_i - X_{cp})(Y_i - Y_{cp})) / (n - 1), \quad (11)$$

где n – количество значений переменных X и Y ; X_{cp} и Y_{cp} – средние значения переменных X и Y соответственно.

В зависимости от значений X и Y можно рассчитать надежность различных видов (рестовая, параллельных форм и др.)

Экспертные оценки вносятся в специальные специальные формы [13]. Структура форм устроена таким образом, что при заполнении форм на третьем уровне, формы второго и первого уровня могут заполняться автоматически, что облегчает задачу ввода данных. Экспертные оценки качества учебно-методических и тестовых материалов являются исходными данными для прогнозирования результатов процесса обучения.

3. Модуль «Анализа и прогнозирования результатов процесса обучения» решает задачи сбора данных СДО, предварительной обработки данных, прогнозирования результатов процесса обучения, оценка качества прогноза.

Процесс сбора данных подразумевает экспорт данных из БД СДО и объединение с данными экспертных оценок. Первым шагом отбираются записи, со следующими информативными признаками: данные о посещении учебных и тестовых ресурсов, данные об участии в дискуссионных форумах, данные о просмотре анонсов, данные о курсе, данные об уровне подготовки обучаемого. Перечисленный набор информативных признаков определен с помощью фильтровочного алгоритма в работах [11] и считается оптимальным для построения модели прогноза результатов обучения. Затем отобранные записи объединяются с экспертными оценками качества.

Предварительная обработка данных – это шаг перед применением алгоритма интеллектуального анализа данных. Предварительная обработка данных включает в себя различные задачи, такие как очистка данных и преобразование данных [23]. Очистка данных является одной из основных задач предварительной обработки и применяется к образовательному набору данных для удаления ненужных элементов и пропущенных значений.

Прогнозирование результатов процесса обучения относится к задачам классификации. Наиболее точные прогнозы дает следующий алгоритм классификации: бустинг над решающими деревьями [17]. В рамках данного алгоритма модель прогноза имеет следующий вид:

$$y_t = \sum_{k=1}^K f_k(x_i), \quad f_k \in F, \quad (12)$$

где x_i – объекты обучающей выборки, а y_i – целевая переменная для объекта с индексом i , k – это количество деревьев в модели, f – функция в функциональном пространстве F , которое представляет собой набор всех возможных деревьев. Функционал качества состоит из функции потерь и параметра регуляризации:

$$O b_j(\theta) = L(\theta) + \mu(\theta), \quad (13)$$

Таким образом, функционал качества, для данной задачи имеет вид:

$$O b_j(\theta) = \sum_l^n l(y_i, y_l) + \sum_{k=1}^K \mu(f_k), \quad (14)$$

В общем случае применяется разложение в ряд Тейлора функционала качества до второго порядка. Полученная функция потерь минимизируется для каждого нового дерева, которое добавляется в модель на шаге t [22]. Примеры использования данного алгоритма для прогнозирования результатов учебной деятельности приведены в работе [10].

4. Вычислительные эксперименты. В наших экспериментах использовались четыре общих измерения для оценки качества алгоритма классификации: Accuracy, Precision, Recall and F-Measure.

В следующей таблице представлены результаты экспериментов и отражено влияние расширенного набора признаков на качество прогноза результатов учебной деятельности. В качестве алгоритма классификации был использован алгоритм бустинг над решающими деревьями (Boosting DT).

Таблица 1

Прогнозирование результатов учебной деятельности с учетом (AddFeatures) и без учета расширенного набора признаков (WTAddFeatures)

Алгоритм классификации	Boosting DT	
	AddFeatures	WTAddFeatures
Расширенный набор признаков		
Accuracy	77.8	79.7
Precision	77.8	79.7
Recall	78.0	79.8
F-Measure	77.9	79.7

Согласно приведенным данным применение алгоритма бустинг над решающими деревьями и расширенного набора признаков показали хорошее качество прогноза. Точность результатов с учетом использования расширенного набора признаков превосходит результаты без учета расширенного набора признаков.

Заключение. Представленная ИИС ПКОД не дублирует функциональность большинства аналогичных разработок, а расширяет ее путем возможности совместного использования разработанных подсистем со сторонними программными продуктами. Представленная ИИС ПКОД основана на использовании интеллектуального анализа данных с использованием расширенного набора признаков и дополнена разработанными алгоритмами и необходимыми формулами. Использование ИИС ПОД в процессе обучения позволит внести элементы интеллектуальности в существующую образовательную среду и дополнить ее функциональность путем использования предложенных подсистем.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Бова В.В., Лежебоков А.А., Нужнов Е.В. Образовательные информационные системы на основе мобильных приложений с дополненной реальностью // Известия ЮФУ. Технические науки. – 2015. – № 6 (167). – С. 200-210.
2. Бова В.В. Представление динамических моделей сложных объектов в интеллектуальных обучающих системах // Известия ЮФУ. Технические науки. – 2010. – № 12 (113). – С. 219-226.
3. Кравченко Ю.А., Лежебоков А.А., Запорожец Д.Ю. Способы интеллектуального анализа данных в сложных системах // Известия Кабардино-Балкарского научного центра РАН. – 2012. – № 3. – С. 52-57.
4. Кравченко Ю.А. Нечеткие модели анализа уровня знаний и поддержки принятия решений в процессе обучения // Известия ЮФУ. Технические науки. – 2014. – № 7 (156). – С. 134-142.
5. Кравченко Ю.А., Бова В.В. Нечеткое моделирование разнородных знаний в интеллектуальных обучающих системах // Открытое образование. – 2013. – № 4 (99). – С. 70-74.
6. Нужнов Е.В. Возможности и средства повышения эффективности адаптивной среды компьютерного обучения // Известия ЮФУ. Технические науки. – 2011. – № 7 (120). – С. 237-244.
7. T. Devasia, V.T.P, and V. Hegde. Prediction of student's performance using educational data mining // In Data Mining and Advanced Computing (SAPIENCE), International Conference on, 2016.
8. Halde R.R., Deshpande A., and Mahajan A. Psychology assisted prediction of academic performance using machine learning // In IEEE International Conference On Recent Trends In Electronics Information Communication Technology, 2016.

9. *Kassarnig V., Bjerre-Nielsen A., Mones E., Lehmann S., and Dreyer D.* Lassen Academic performance and behavioral patterns // Website, 2017. – <https://arxiv.org/abs/1706.09245>.
10. *Marbouti F., Diefes-Dux H.A., and Madhavan K.* Models for early prediction of at-risk students in a course using standards-based grading // *Computers & Education*. – 2016. – Vol. 103. – P. 1-15.
11. *Pardo A., Han F., and Ellis R.A.* Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance // *IEEE Transactions on Learning Technologies*. – 2016. – No. 10 (1). – P. 82-92,
12. *Кулиев Э.В., Запорожец Д.Ю., Курейчик Вл.Вл.* Комбинированный подход адаптации и самоорганизации к обработке проблемно-ориентированных знаний // *Известия ЮФУ. Технические науки*. – 2015. – № 6 (167). – С. 191-200.
13. *C. Romero, M.L.* Predicting students' final performance from participation in on-line discussion forums // *Computers & Education*. – 2013. – P. 458-472.
14. *Бранченко Д.С., Кравченко Ю.А., Кулиев Э.В.* Актуальные задачи сетевых образовательных технологий для оптимизации процесса получения знаний // *Информатика, вычислительная техника и инженерное образование*. – 2016. – № 1 (25).
15. *Wang R., Harari G., Hao P., Zhou X., and Smartgpa C.T.* How smartphones can assess and predict academic performance of college students // In *UBICOMP '15, OSAKA, JAPAN, 2015*.
16. *Новиков А.А., Кулиев Э.В., Самойлов А.Н.* Когнитивная архитектура агентов мультиагентной системы // *Информатизация и связь*. – 2016. – № 2. – С. 127-131.
17. *Казмина И.И.* Организация контроля информационно-методического обеспечения учебного процесса // *Информатика, вычислительная техника и инженерное образование*. – 2013. – № 3 (13).
18. *Ким В.С.* Тестирование учебных достижений: монография. – Уссурийск: Изд-во УГПИ. 2007. – 214 с.
19. *Козлов О.А., Довгань В.В.* Информационно-методическое обеспечение специальных дисциплин учреждений среднего профессионального образования строительного профиля // *Информационная среда образования и науки: Электронное периодическое издание*. – 2012. – № 10.
20. *Ильясов Э.С.* Применение модели градиентного бустинга для прогнозирования развития диабета // *Молодой ученый*. – 2016. – № 27. – С. 1-5.

REFERENCES

1. *Bova V.V., Lezhebokov A.A., Nuzhnov E.V.* Obrazovatel'nye informatsionnye sistemy na osnove mobil'nykh prilozheniy s dopolnennoy real'nost'yu [Educational information systems based on mobile applications with augmented reality], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2015, No. 6 (167), pp. 200-210.
2. *Bova V.V.* Predstavlenie dinamicheskikh modeley slozhnykh ob'ektov v intellektual'nykh obuchayushchikh sistemakh [Presentation of dynamic models of complex objects in intelligent learning systems], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2010, No. 12 (113), pp. 219-226.
3. *Kravchenko Yu.A., Lezhebokov A.A., Zaporozhets D.Yu.* Sposoby intellektual'nogo analiza dannykh v slozhnykh sistemakh [Methods of data mining in complex systems], *Izvestiya Kabardino-Balkarskogo nauchnogo tsentra RAN* [Izvestiya of Kabardino-Balkar scientific center of RAS], 2012, No. 3, pp. 52-57.
4. *Kravchenko Yu.A.* Nechetkie modeli analiza urovnya znaniy i podderzhki prinyatiya resheniy v protsesse obucheniya [Fuzzy models of knowledge analysis and decision support in the learning process], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2014, No. 7 (156), pp. 134-142.
5. *Kravchenko Yu.A., Bova V.V.* Nechetkoe modelirovanie raznorodnykh znaniy v intellektual'nykh obuchayushchikh sistemakh [Fuzzy modeling of heterogeneous knowledge in intelligent learning systems], *Otkrytoe obrazovanie* [Open education], 2013, No. 4 (99), pp. 70-74.
6. *Nuzhnov E.V.* Vozmozhnosti i sredstva povysheniya effektivnosti adaptivnoy sredy komp'yuternogo obucheniya [Opportunities and means to improve the effectiveness of the adaptive environment for e-learning], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2011, No. 7 (120), pp. 237-244.

7. T. Devasia, V.T.P, and V. Hegde. Prediction of student's performance using educational data mining, *In Data Mining and Advanced Computing (SAPIENCE), International Conference on, 2016.*
8. Halde R.R., Deshpande A., and Mahajan A. Psychology assisted prediction of academic performance using machine learning, *In IEEE International Conference On Recent Trends In Electronics Information Communication Technology, 2016.*
9. Kassarnig V., Bjerre-Nielsen A., Mones E., Lehmann S., and Dreyer D. Lassen Academic performance and behavioral patterns, *Website, 2017.* Available at: <https://arxiv.org/abs/1706.09245>.
10. Marbouti F., Diefes-Dux H.A., and Madhavan K. Models for early prediction of at-risk students in a course using standards-based grading, *Computers & Education, 2016, Vol. 103, pp. 1-15.*
11. Pardo A., Han F., and Ellis R.A. Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance, *IEEE Transactions on Learning Technologies, 2016, No. 10 (1), pp. 82-92,*
12. Kuliev E.V., Zaporozhets D.Yu., Kureychik V.I. Kombinirovannyi podkhod adaptatsii i samoorganizatsii k obrabotke problemno-orientirovannykh znaniy [Combined approach of adaptation and self-organization to the processing of problem-oriented knowledge], *Izvestiya YuFU. Tekhnicheskie nauki [Izvestiya SFedU. Engineering Sciences], 2015, No. 6 (167), pp. 191-200.*
13. C. Romero, M.L. Predicting students' final performance from participation in on-line discussion forums, *Computers & Education, 2013, pp. 458-472.*
14. Branchenko D.S., Kravchenko Yu.A., Kuliev E.V. Aktual'nye zadachi setevykh obrazovatel'nykh tekhnologiy dlya optimizatsii protsessa polucheniya znaniy [Actual problems of network educational technologies to optimize the process of obtaining knowledge], *Informatika, vychislitel'naya tekhnika i inzhenernoe obrazovanie [Computer science, computer engineering and engineering education], 2016, No. 1 (25).*
15. Wang R., Harari G., Hao P., Zhou X., and Smartgpa C.T. How smartphones can assess and predict academic performance of college students, *In UBIComp '15, OSAKA, JAPAN, 2015.*
16. Novikov A.A., Kuliev E.V., Samoylov A.N. Kognitivnaya arkhitektura agentov mul'tiagentnoy sistemy [Cognitive architecture of agents of multi-agent system], *Informatizatsiya i svyaz' [Informatization and communication], 2016, No. 2, pp. 127-131.*
17. Kazmina I.I. Organizatsiya kontrolya informatsionno-metodicheskogo obespecheniya uchebnogo protsessa [Organization of control of information and methodological support of the educational process], *Informatika, vychislitel'naya tekhnika i inzhenernoe obrazovanie [Computer science, computer engineering and engineering education], 2013, No. 3 (13).*
18. Kim V.S. Testirovanie uchebnykh dostizheniy: monografiya [Testing of educational achievements: monograph]. Ussuriysk: Izd-vo UGPI. 2007, 214 p.
19. Kozlov O.A., Dovgan' V.V. Informatsionno-metodicheskoe obespechenie spetsial'nykh distsiplin uchrezhdeniy srednego professional'nogo obrazovaniya stroitel'nogo profilya [Information and methodological support of special disciplines of secondary vocational education institutions of construction profile], *Informatsionnaya sreda obrazovaniya i nauki: Elektronnoe periodicheskoe izdanie [Information environment of education and science: Electronic periodical], 2012, No. 10.*
20. Ilyasov E.S. Primenenie modeli gradientnogo bustinga dlya prognozirovaniya razvitiya diabeta [Application of gradient boosting model to predict diabetes development], *Molodoy uchenny [Young scientist.], 2016, No. 27, pp. 1-5.*

Статью рекомендовал к опубликованию д.т.н., профессор Н.И. Витиска.

Кулиев Эльмар Валерьевич – Южный федеральный университет; e-mail: ekuliev@sfedu.ru; 347928, г. Таганрог, пер. Некрасовский, 44; тел.: 88634371651; кафедра систем автоматизированного проектирования; доцент.

Цырульникова Эльвира Сергеевна – e-mail: bolshova@sfedu.ru; кафедра систем автоматизированного проектирования; аспирант.

Кулиева Нина Владимировна – e-mail: holopova@sfedu.ru; кафедра систем автоматизированного проектирования; аспирант.

Марков Владимир Васильевич – e-mail: vmarkov@sfnedu.ru; кафедра систем автоматизированного проектирования; доцент.

Kuliev Elmar Valerievich – Southern Federal University; e-mail: ekuliev@sfnedu.ru; 44, Nekrasovski lane, Taganrog, 347928, Russia; phone: +78634371651; the department of computer aided design; associate professor.

Tsyruhnikova Elvira Sergeevna – e-mail: bolshova@sfnedu.ru; the department of computer aided design; graduate student.

Kulieva Nina Vladimirovna – e-mail: holopova@sfnedu.ru; the department of computer aided design; graduate student.

Markov Vladimir Vasilievich – e-mail: vmarkov@sfnedu.ru; the department of computer aided design; associate professor.

УДК 004.822

DOI 10.23683/2311-3103-2019-4-89-102

В.В. Бова, Ю.А. Кравченко

**БИОИНСПИРИРОВАННЫЙ ПОДХОД К РЕШЕНИЮ ЗАДАЧИ
КЛАССИФИКАЦИИ ПРОФИЛЕЙ ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЕЙ
В ИНТЕЛЛЕКТУАЛЬНЫХ ИНТЕРНЕТ-СЕРВИСАХ***

Рассматриваются проблемы повышения эффективности организации личностно-ориентированного взаимодействия пользователя в интеллектуальных Интернет-сервисах для формирования культуры безопасного поведения в Интернет-пространстве. Для их решения предлагается метод построения профилей поведения пользователей на основании анализа их информационных потребностей, максимально соответствующих их предпочтениям, в том числе и неявным. Актуальность работы определяется растущей популярностью идеи персонализации контента и информационных услуг в Интернет-пространстве. Профиль поведения пользователя рассматривается авторами как слабоформализованный объект в пространстве признаков внутренних и внешних характеристик, описывающих его взаимодействие с Интернет-ресурсом. Предлагаемый в работе метод основан на вероятностном алгоритме EM-кластеризации исследуемых данных о характеристиках пользователя и распределенных Интернет-ресурсов для генерации структуры входных параметров классификаторов модели формирования профиля пользователя. Оптимизация структуры реализуется механизмом отбора информативных признаков профиля, основанного на идее выявления скрытых интересов и предпочтений пользователей с одной стороны, и способностью ресурса удовлетворять заинтересованных пользователей этому набору признаков – с другой. Для снижения размерности исходных данных признакового пространства в задаче классификации предлагается метаэвристический алгоритм оптимизации «кукушкин поиск», отличающийся масштабируемостью и высокой интерпретируемостью выходных данных. Оптимизация параметров классификаторов заключается в подборе параметров функции принадлежности и меток классов обобщенного профиля таким образом, чтобы численный критерий точности классификации признаков ресурсов и предпочтений пользователей сводился к максимуму на реальных данных. Для оценки эффективности предложенного алгоритма проведен вычислительный эксперимент на тестовых наборах данных из открытого репозитория UCI Machine Learning Repository. Результаты которого показали, что построенный на тестовых данных классификатор обладает более высоким уровнем интерпретируемости полученных результатов формирования профилей, сохраняя точность классификации.

Классификация; интеллектуальные Интернет-сервисы; профиль поведения пользователя; EM-алгоритм; биоинспирированные методы; метаэвристика «кукушкин поиск».

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18–29–22019.