

Интеллектуальный анализ образовательных данных студентов для создания персонализированной среды обучения

Токтарова В.И.

Марийский государственный университет (ФГБОУ ВО МарГУ)
г. Йошкар-Ола, Российская Федерация
ORCID: <https://orcid.org/0000-0002-3590-3053>
e-mail: toktarova@yandex.ru

Попова О.Г.

Марийский государственный университет (ФГБОУ ВО МарГУ)
г. Йошкар-Ола, Российская Федерация
ORCID: <https://orcid.org/0000-0002-2666-1005>
e-mail: olesya_popova10@mail.ru

Цифровизация является одним из важнейших направлений развития современной системы образования. Внедрение в образовательный процесс цифровых технологий приводит к появлению большого объема информации, исследованием которой занимается анализ образовательных данных. В статье рассмотрены основные возможности применения интеллектуального анализа образовательных данных в разработке персонализированной среды обучения. Рассмотрены методы анализа данных в сфере образования и описаны типы задач, которые могут быть решены с их помощью. Предложена модель создания персонализированной среды обучения на основе анализа образовательных данных, целью которой является прогнозирование успешности студента в процессе обучения на основе его индивидуальных характеристик и образовательных результатов.

Ключевые слова: интеллектуальный анализ данных, персонализация, персонализированная среда обучения, образовательные данные, студент, вуз.

Для цитаты:

Токтарова В.И., Попова О.Г. Интеллектуальный анализ образовательных данных студентов для создания персонализированной среды обучения // Цифровая гуманитаристика и технологии в образовании (DHTE 2021): сб. статей II Всероссийской научно-практической конференции с международным участием. 11–12 ноября 2021 г. / Под ред. В.В. Рубцова, М.Г. Сороковой, Н.П. Радчиковой. М.: Издательство ФГБОУ ВО МГППУ, 2021. 343–352 с.

Введение

Обеспечение условий персонализации образовательного процесса является важным элементом развития современного образования. Одной из задач федерального проекта «Кадры для цифровой

экономики» [5] Национальной программы «Цифровая экономика Российской Федерации» [4] является разработка модели «Цифровой университет», основной целью которой является трансформация процессов университета с помощью цифровых технологий. В рамках модели предполагается внедрение в образовательный процесс современных цифровых сервисов и платформ, использующих алгоритмы анализа больших данных и машинного обучения.

Применение в процессе обучения онлайн-платформ, широкое распространение электронного обучения приводит к накоплению большого количества данных о пользовательском поведении и результатах обучения. Одним из направлений использования данного материала в учебном процессе является анализ образовательных данных – направление исследований, которое связано с применением методов интеллектуального анализа данных, машинного обучения и статистики к информации, производимой образовательными учреждениями [4].

Образовательные данные – это данные об образовательных результатах обучающихся, развития образовательных систем, условиях образования (образовательных программах, особенностях образовательной среды, образовательных методиках и др.) [8].

Исследованиям в данной области проводятся российскими и зарубежными учеными. Так, интеллектуальному анализу данных посвящены работы С.И. Барцева, П. Дж. Вербоса., Н. Вапника, А.И. Галушкина, А.Н. Горбань, Ю.И. Журавлева, М. Мински, А.Б. Новикова, К.В. Рудакова, В.А. Охонина, С. Пайперт, Б. Уидроу, Дж. Хопфилд и др.; методы моделирования обучающихся рассматривались в исследованиях Р. Алмонда, Г.А. Атанова, М. Десмаре, Е. Миллан, Д. Рей и др. Большой вклад в разработку подходов к решению проблем анализа образовательных данных внесли Р.Бейкер, П.Л. Брусиловский, С. Вентура, Л.С. Лисицына, Х. Ромеро, И.Д. Рудинский и др.

Целью нашего исследования является использование методов интеллектуального анализа образовательных данных для проектирования и разработки модели персонализированной среды обучения студентов в высших учебных заведениях.

Методы

Методологическую базу исследования составляет комплекс методов: теоретические (анализ нормативных источников, психолого-педагогической и специальной литературы; системный, структурно-функциональный и сравнительно-сопоставительный анализ; контент-анализ, синтез, систематизация, дедукция, индукция, прогно-

зирование); эмпирические (наблюдение, опрос, интервьюирование, тестирование, экспертная оценка); математические (корреляционный и дисперсионный анализ, методы математической статистики, математическое моделирование, Data Mining).

Результаты

Сегодня российские университеты активно внедряют технологии анализа данных в управление образовательным процессом, разрабатывая собственные технические решения. К примеру, информационная система вуза [3] прогнозирует успешность студента в следующем семестре с точностью 71 % на основе данных о баллах ЕГЭ, результатах сессий, участия студента в общественной жизни вуза, его активности в электронных библиотечных системах. В университете предполагают, что данная система будет способствовать снижению процента отчислений из-за академической неуспеваемости.

К основным задачам интеллектуального анализа образовательных данных относят прогнозирование поведения обучающегося в процессе обучения; мониторинг сформированности профессиональных компетенций; разработку новых способов представления учебной информации, которые соответствуют выявленным познавательным способностям студентов; анализ и прогнозирование конкурентоспособности студентов на рынке труда, а также диагностика уровня качества образования [1, 9].

В соответствии с приведенными задачами методы интеллектуального анализа образовательных данных подразделяются на следующие группы [7]:

- методы прогнозирования (построение модели, предсказывающей значение зависимой переменной по значениям независимых переменных. К таким методам относятся алгоритмы классификации и регрессии);
- методы обнаружения структуры (разбиение выборки на группы по каким-либо схожим признакам, к ним относятся алгоритмы кластеризации);
- методы выявления взаимосвязей (установление взаимосвязи между переменными. К данной группе методов относятся корреляционный анализ, поиск связующих правил и поиск последовательных шаблонов).

Процесс интеллектуального анализа образовательных данных состоит из пяти этапов. На первом этапе уточняются цели и задачи анализа. На втором этапе происходит предобработка данных, используемых для анализа. Характер преобразования данных зависит от используемых методов. На третьем этапе происходит построение

моделей, наиболее часто для этого используются методы машинного обучения. На следующем этапе построенные модели проходят проверку, в том числе на контрольном наборе данных, который не участвует в обучении. На последнем этапе проводится интерпретация построенных моделей и полученных результатов с целью их использования в дальнейшем для принятия решений [10].

В совокупности все методы интеллектуального анализа данных могут служить инструментом для создания персонализированной среды обучения. Анализ позволит спрогнозировать успешность в процессе обучения и дать рекомендации каждому студенту индивидуально, т.е. создать персонализированную образовательную среду.

Персонализация в электронном обучении – это использование технологий и информации об обучающихся для адаптации взаимодействия между онлайн-системой и отдельными обучающимися таким образом, чтобы последние добивались лучших результатов обучения. Каждый из студентов имеет собственные предпочтения относительно комфортной организации процесса своего обучения. Кому-то нужно больше времени, чтобы понять учебный материал, кто-то усваивает информацию быстрее. Замечено, что использование различных элементов курса по-разному влияет на успеваемость обучающихся. Персонализированные среды обучения могут изменять свою реакцию на входные данные в зависимости от контекста и обстоятельств. Эффективность измеряется ее способностью оптимизировать результаты, воздействуя на эти изменения. Подобные среды создаются с целью сбора данных, анализа и предоставления рекомендаций и решений для оптимизации определенных критериев.

К примеру, целью разрабатываемой авторской модели создания персонализированной среды на основе алгоритмов интеллектуального анализа образовательных данных является прогнозирование успешности обучения студента на основе его индивидуальных особенностей и образовательных результатов. Ставится задача выявить, какие студенты имеют большую вероятность успешно пройти курс, а какие студенты окажутся в ситуации, когда есть риск получить неудовлетворительную оценку.

Структура модели включает в себя функционально-целевой, содержательный, технологический и результативно-критериальный блоки, и отражает этапы построения персонализированной среды обучения на основе анализа образовательных данных. Функционально-целевой блок модели формируется под воздействием внешней среды и акцентирует внимание на основной цели функционирования, предопределяя приоритетное направление развития электронного обучения студентов, обусловленное социальным заказом

и модернизацией российского образования. Содержательный блок определяется предметным контентом определённой дисциплины. Структура и содержание дисциплин варьируются в зависимости от направления подготовки / специальности студентов. Идея обучения в персонализированной среде строится на предположении о том, что способности каждого конкретного студента раскрываются при оптимально подобранных для него условиях. Технологический блок модели реализуется за счет определения конкретных методов, форм и средств обучения. Процесс обучения в персонализированной среде представляется как дидактическая система с определенной совокупностью взаимодействующих компонентов и функциями управления, которое предполагает непрерывное измерение успешности усвоения учебного материала каждым обучающимся; оперативную обратную связь обучающего и обучаемого; корректировку обучения (в случае необходимости). Результативно-критериальный блок направлен на реализацию функций диагностического, корректирующего и рефлексивного характера.

Проектирование персонализированной среды обучения предполагает следующие шаги:

- 1) диагностику индивидуальных характеристик студентов;
- 2) группировку контингента по кластерам в зависимости от индивидуальных характеристик;
- 3) конструирование модели обучаемого;
- 4) разработку содержания и дифференциацию образовательных технологий, средств и методов обучения по различным параметрам;
- 5) построение персонализированных траекторий обучения в электронной среде.

Данные о студентах можно получить из системы Moodle [6], которую используют многие российские вузы. В Moodle хранятся данные о процессе обучения, выполнении заданий и прогрессе обучаемого. Можно проводить анализ и визуализацию отчётных данных об активности конкретного студента, отдельной группы или всех студентов курса. Объединив все данные о результатах прохождения курса, можно получить более конкретную информацию об успеваемости того или иного студента. Для проектирования модели были использованы следующие данные о студенте: идентификатор пользователя; пол; возраст; дата и время доступа к элементам курса; количество обращений к ресурсам каждого вида (теоретические материалы, практические задания, тесты) и их продолжительность; количество сообщений на форумах и чатах курса; общее количество обращений к курсу за определённые календарные периоды; оценки студента (промежуточные и итоговая); соблюдение сроков графика

обучения (количество дней, прошедших после истечения срока сдачи задания). Данный список переменных представляет собой информацию о студенте и о его активности на курсе. Прогноз основывается на взаимодействии студента с электронной образовательной средой.

Состав набора психолого-педагогических характеристик для диагностики индивидуальных особенностей и способностей обучающихся в нашей работе основан на анализе стилевых характеристик обучающихся, выявленных на основе моделей Колба, Хони–Мамфорда, Грегорка, VARK, Фельдер-Сильверман, целостного мозга Геррманна [11]. На базе физико-математического и психолого-педагогического факультетов Марийского государственного университета был проведен эксперимент по выявлению значений характеристик студентов 3–4-го курсов (165 человек). Группировка студентов по кластерам производилась на основе методов кластерного и дискриминантного анализа в зависимости от выявленных индивидуальных характеристик с использованием нейронной сети Хемминга.

Для создания индивидуального портрета конкретного студента была сформирована модель обучающегося, включающая в себя индивидуальные характеристики студента и данные о его образовательных результатах, описанных выше.

Проектирование и разработка содержания учебного курса включала в себя следующую дифференциацию образовательных технологий, учитывающих: вид представления учебного материала (текстовое описание; графическое представление (иллюстрации, схемы, модели); видео (видеолекции, видеопрактикумы, вебинары); аудио (аудиолекции, аудиословари, аудиосправочники)); объем учебного материала (краткое представление; подробное представление); уровень сложности (начальный; средний; высокий); стратегию подачи учебного материала (последовательное представление небольших фрагментов; полное изложение учебного элемента); формы организации учебной деятельности (теоретическое обучение; выполнение лабораторно-практических работ; подготовка к зачету, экзамену; проведение исследований; комплексное изучение курса); темп обучения (медленный; обычный; ускоренный); педагогические приемы (предоставление методических рекомендаций и инструкций; создание проблемных ситуаций; экспериментирование (в т. ч. симуляции, игры, др.); организация групповой / индивидуальной работы).

На основе значений параметров модели обучающегося была сформирована и предоставлена студенту персонализированная траектория обучения, адаптированная под конкретного студента, где каждый учебный элемент был оформлен в разных вариантах.

Заключение

Таким образом, основное назначение применения интеллектуального анализа образовательных данных – лучше понять обучающегося, его намерения и предпочтения, и помочь каждому индивидуально достичь его целей, т.е. создать персонализированную среду обучения. Предложенная в статье модель персонализированного обучения позволит руководителю курса своевременно выявлять студентов из группы риска и предпринять необходимые меры. Использование анализа образовательных данных при проектировании индивидуальных траекторий обучения позволит обеспечить многообразие вариантов развития обучаемых с учетом их личностных качеств и образовательных результатов, осуществить постоянный мониторинг действий и уровня компетенций студентов, что будет способствовать повышению качества электронного обучения в вузе.

Литература

1. *Белоножко П.П., Карпенко А.П., Храмов Д.А.* Анализ образовательных данных: направления и перспективы применения. Вестник евразийской науки. 2017. № 4 (41).
2. *Мантуленко В.В.* Перспективы использования цифрового следа в высшем образовании [Электронный ресурс]. Преподаватель XXI века. 2020. № 3-1. С. 32-42. DOI: 10.13862/2073-9613-2020-3-32-42
3. Московский государственный педагогический университет [официальный сайт]. URL: <https://www.mgpu.ru/> (дата обращения: 10.09.2021)
4. Паспорт национальной программы «Цифровая экономика Российской Федерации» [Электронный ресурс]. Правительство России. – URL: <http://static.government.ru/> (дата обращения: 10.09.2021)
5. *Паспорт федерального проекта «Кадры для цифровой экономики»* [Электронный ресурс]. URL: <https://digital.ac.gov.ru/poleznaya-informaciya/material/Pasport-federal'nogo-proekta-Kadry-dlya-tsifrovoy-ekonomiki.pdf> (дата обращения: 10.09.2021)
6. *Система управления курсами Moodle* [официальный сайт]. URL: <https://moodle.org/> (дата обращения: 10.09.2021)
7. *Фиофанова О.А.* Анализ больших данных в сфере образования: методология и технологии: монография. – М.: Издательский дом «Дело» РАНХиГС, 2020. – 200 с.
8. *Фиофанова О.А.* Smart Big Data в публичных докладах [Электронный ресурс] // Образовательная политика. 2020. № 4 (84). С. 70–77. DOI: 10.22394/2078–838X-2020–4-70–77
9. *Хлопотов М.В., Коцюба И.Ю.* Методы интеллектуального анализа данных для мониторинга и диагностики качества образования // Дистанционное и виртуальное образование. – Москва, 2014. № 5. С. 18–25.

10. *Чмыхова Е.В., Давыдов Д.Г., Строкопыхтова С.А.* Интеллектуальный анализ данных электронной образовательной среды для оценки личностных особенностей обучаемых [Электронный ресурс] // *Инновации в образовании*. 2020. № 2. С.83–99.
11. *Toktarova V.I.* Pedagogical Management of Learning Activities of Students in the Electronic Educational Environment of the University: a Differentiated Approach. *International Education Studies*. 2015. Vol. 8. № 5. P. 205–212.

Информация об авторах

Токтарова Вера Ивановна, доктор педагогических наук, Марийский государственный университет (ФГБОУ ВО МарГУ), г. Йошкар-Ола, Российская Федерация, ORCID: <https://orcid.org/0000-0002-3590-3053>, e-mail: toktarova@yandex.ru

Попова Олеся Геннадьевна, магистрант института цифровых технологий, Марийский государственный университет (ФГБОУ ВО МарГУ), г. Йошкар-Ола, Российская Федерация, ORCID: <https://orcid.org/0000-0002-2666-1005>, e-mail: olesya_pорова10@mail.ru

Intelligent Educational Data Analysis for Creation a Personalized eLearning Environment

Vera I. Toktarova

Mari State University, Yoshkar-Ola, Russia
ORCID: <https://orcid.org/0000-0002-3590-3053>
e-mail: toktarova@yandex.ru

Olesya G. Popova

Mari State University, Yoshkar-Ola, Russia
ORCID: <https://orcid.org/0000-0002-2666-1005>
e-mail: olesya_popova10@mail.ru

Digitalization is one of the most important directions in the development of the modern education system. The introduction of digital technologies into the educational process leads to the emergence of a large amount of information. The article discusses the main possibilities of using educational data mining in the development of a personalized learning environment. Methods of data analysis in the field of education are considered and the types of problems that can be solved with their help are described. A model for creation a personalized eLearning environment based on the analysis of educational data is proposed, the purpose of which is to predict student success in the learning process based on his individual characteristics and educational results.

Keywords: data mining, personalization, personalized learning environment, educational data, student, university.

For citation:

Toktarova V.I., Popova O.G. Intelligent Educational Data Analysis for Creation a Personalized eLearning Environment // Digital Humanities and Technology in Education (DHTE 2021): Collection of Articles of the II All-Russian Scientific and Practical Conference with International Participation. November 11–12, 2021 / V.V. Rubtsov, M.G. Sorokova, N.P. Radchikova (Eds). Moscow: Publishing house MSUPE, 2021. 343–352 p.

References

1. Belonozhko P.P., Karpenko A.P., Khramov D.A. Analiz obrazovatel'nykh dannykh: napravleniya i perspektivy primeneniya [Analysis of educational data: directions and prospects of application]. *Vestnik evraziiskoi nauki = Bulletin of Eurasian Science*, 2017, no. (41).
2. Mantulenko V.V. Perspektivy ispol'zovaniya tsifrovogo sleda v vysshem obrazovanii [Prospects for using the digital footprint in higher education]. *Prepodavatel' KhKhI veka = Teacher of the XXI century*, 2020, no. 3–1, pp. 32–42. DOI: 10.13862/2073-9613-2020-3-32-42

3. Moskovskii gosudarstvennyi pedagogicheskii universitet [Moscow State Pedagogical University]. Available at: <https://www.mgpu.ru/> (Accessed 10.09.2021)
4. Pasport natsional'noi programmy «Tsfirovaya ekonomika Rossiiskoi Federatsii» [Passport of the national program “Digital Economy of the Russian Federation”]. Available at: <http://static.government.ru/> (Accessed 09/10/2021)
5. Pasport federal'nogo proekta «Kadry dlya tsifrovoi ekonomiki» [Passport of the federal project “Personnel for the digital economy”]. Available at: <https://digital.ac.gov.ru/poleznaya-informaciya/material/Pasport-federal'nogo-proekta-Kadry-dlya-tsfirovoi-ekonomiki.pdf> (Accessed 10.09.2021)
6. Sistema upravleniya kursami Moodle [Moodle course management system]. Available at: <https://moodle.org/> (Accessed 10.09.2021)
7. Fiofanova O.A. Analiz bol'shikh dannykh v sfere obrazovaniya: metodologiya i tekhnologii: monografiya [Big Data Analysis in Education: Methodology and Technologies]. – M.: *Izdatel'skii dom «Delo» = Publishing house “Delo”, RANKhiGS, 2020, 200 p.*
8. Fiofanova O.A. Smart Big Data v publicnykh dokladakh [Smart Big Data in public reports] // *Obrazovatel'naya politika = Educational policy*, 2020, no. 4 (84), pp. 70–77. DOI: 10.22394/2078–838Kh-2020–4-70–77
9. Khlopotov M.V., Kotsyuba I.Yu. Metody intellektual'nogo analiza dannykh dlya monitoringa i diagnostiki kachestva obrazovaniya [Methods of data mining for monitoring and diagnosing the quality of education]. *Distantsionnoe i virtual'noe obrazovanie = Distance and virtual education*, 2014, no. 5, pp. 18–25.
10. Chmykhova E.V., Davydov D.G., Strokopytova S.A. Intellektual'nyi analiz dannykh elektronnoi obrazovatel'noi sredy dlya otsenki lichnostnykh osobennosti obuchaemykh [Intelligent data analysis of the electronic educational environment for assessing the personal characteristics of students]. *Innovatsii v obrazovanii. = Innovations in education*, 2020, no. 2, pp. 83–99.
11. Toktarova V.I. Pedagogical Management of Learning Activities of Students in the Electronic Educational Environment of the University: a Differentiated Approach. *International Education Studies*. 2015. Vol. 8, no 5, pp. 205–212.

Information about the authors

Vera I. Toktarova, Dr. of Pedagogical Sciences, Mari State University, Yoshkar-Ola, Russia, ORCID: <https://orcid.org/0000-0002-3590-3053>, e-mail: toktarova@yandex.ru

Olesya G. Popova, Master's student at the Institute of Digital Technologies, Mari State University, Yoshkar-Ola, Russia, ORCID: <https://orcid.org/0000-0002-2666-1005>, e-mail: olesya_popova10@mail.ru