

Методика обучения интеллектуальному анализу образовательных данных студентов педагогического вуза

Цель статьи состоит в обсуждении и обосновании обучения студентов педагогических ВУ-Зов интеллектуальному анализу образовательных данных и описанию методической системы обучения студентов со средним уровнем знаний математических и ИТ-дисциплин интеллектуальному анализу образовательных данных, способствующей развитию научно-исследовательской компетентности студентов. Актуальность исследования обусловлена требованиями к способности выпускников высших учебных заведений проводить анализ информации и осуществлять исследовательскую деятельность с применением современных методов и технологий, которые обозначаются в образовательных стандартах и государственном заказе и связаны с все возрастающим количеством накапливаемых данных в различных областях деятельности и стоимостью добываемых из них знаний.

Материалы и методы. В статье описывается авторская методика обучения интеллектуальному анализу образовательных данных, разработанная на основе: анализа требований и ожиданий к уровню исследовательской компетентности, навыкам анализа данных и современному образованию в целом; сравнении и анализе содержания учебных программ вузов, книг и курсов по интеллектуальному анализу данных и смежным дисциплинам, обобщении педагогического опыта к соответствующей подготовке. Основными аспектами, лежащими в основе методики, являются: форма перевернутого обучения, концентрическая (итеративная) структура содержания, исследовательские методы обучения, комплекс практических

заданий на развитие исследовательских компетенций и программное средство Weka для интеллектуального анализа данных в качестве основного технического средства обучения, на основе которого реализуются практические задания. Эффективность разработанной методической системы проверялась путем наблюдения за учебным процессом, анкетирования студентов и статистической обработки данных анкет.

Результаты. В ходе исследования обосновано обучение интеллектуальному анализу образовательных данных студентов педагогических ВУЗов, обучающихся на математических и информационных направлениях; предложена методика обучения. Экспериментальное внедрение методики показало, что применение разработанной методики для студентов старших курсов педагогического вуза позволяет повысить уровень исследовательской компетентности студентов и существенно развить компетенцию анализа данных.

Заключение. Описанная методическая система может быть частично или полностью использована преподавателями и методистами для обучения анализу данных на современном уровне и развития научно-исследовательской компетентности студентов со средним уровнем знания математических и ИТ-дисциплин.

Ключевые слова: интеллектуальный анализ данных, обучение интеллектуальному анализу данных, интеллектуальный анализ образовательных данных, исследовательская компетентность, перевернутый класс, концентрическое обучение

Ekaterina A. Terbusheva

Herzen State Pedagogical University, Saint-Petersburg, Russia

Methods of teaching educational data mining for pedagogical students

The aim of the article is to discuss and argue teaching educational data mining for pedagogical students and to describe the methodical system of educational data mining teaching for students with a middle level of mathematical and IT disciplines, that contributes to the development of student's research competence. The relevance of the study is determined by the requirements for the ability of higher education graduates to analyze information and perform research using modern methods and technologies that are mentioned in the educational standards and the government order. They are associated with an increasing amount of accumulated data in various fields and the cost of the knowledge extracted from data.

Materials and methods. The article describes the author's methodical system of educational data mining teaching, which was developed rely on: analysis of requirements and expectations to the research competence level, data analysis skills and modern education in general; comparison and analysis of the content of educational programs, books and courses on data mining and related disciplines, generalization of pedagogical experience. The main aspects underlying the methodology: a form of flipped learning, a concentric (iterative) content structure,

research teaching methods, a set of practical tasks for developing research competencies and Weka software for data mining as the main technical training tool for practical tasks implementation. The effectiveness of the developed methodological system was tested by the educational process monitoring, students questioning and statistical processing of questionnaires data.

Results. The study shows the relevance of educational data mining teaching for students of pedagogical universities, studying in mathematical and informational specialization. The use of the described methodic system for senior pedagogical students allows increasing the level of research competence of students and significantly developing the competence of data analysis.

Conclusion. The described methodical system can be used partially or completely by teachers and methodologists for teaching data analysis at the modern level and development of research competence of students with an average level of knowledge in mathematical and IT disciplines.

Keywords: data mining, teaching of data mining, education data mining, research competence, flipped classroom, concentric content model

Введение

Современные тенденции в области цифровизации образования приводят к накоплению больших объемов данных, которые содержат скрытые шаблоны и закономерности и могут быть использованы для принятия решений с целью повышения качества образования. Процесс обнаружения таких скрытых и потенциально полезных закономерностей в данных называется интеллектуальным анализом данных (ИАД). Опыт применения методов ИАД в сфере образования аккумулируется в сравнительно новом направлении, обозначенном в 2008 году и получившем название «Интеллектуальный анализ образовательных данных» (ИАОД, англ. Educational Data Mining, EDM). Также данное направление в некоторых источниках переводится как «Анализ образовательных данных» (АОД). ИАОД представляет собой научное направление, связанное с разработкой методов изучения уникальных данных, получаемых из образовательных учреждений и с использованием этих методов для лучшего понимания студенческого запроса и повышения качества процесса образования.

Примерами исследований по анализу образовательных данных с применением методов ИАД являются: выявление зависимостей между баллами ЕГЭ и дальнейшей успеваемостью студентов [1], оценка качества проведения курсов на основе опроса слушателей [2], оценка качества курсовых работ [3]. В работах [4–6] описываются подходы к созданию систем обработки накапливаемых в вузах данных, в которые встраиваются методы ИАД. В литературе [7–12] рассматриваются перспективы применения методов интеллектуального анализа данных в сфере образования и на основе анализа научных статей выде-

ляются следующие направления применения методов ИАОД: обеспечение обратной связи для поддержки преподавателей, разработка индивидуального образовательного маршрута, прогнозирование успеваемости учащихся, разделение студентов на группы, выявление поведенческих стереотипов студенчества, построение модели студента, моделирование образовательного содержания обучения, разработка концептуальных карт, оценка качества образования и др.

В связи с широкими возможностями ИАОД для решения задач образования актуальным становится вопрос подготовки специалистов, владеющих соответствующими методами обработки и анализа образовательных данных. В настоящий момент целенаправленных программ обучения будущих специалистов интеллектуальному анализу образовательных данных в РФ нет, как нет и соответствующих русскоязычных курсов или учебной литературы.

Дисциплины по ИАД активно вводятся в программы обучения будущих специалистов в области Информационных технологий [13–15], но задачи и проблемы области образования, а также специфика обработки образовательных данных в них не рассматриваются. Вопрос актуальности смежного курса «Машинное обучение» при подготовке учителей математики и информатики поднимался Кузнецовым О.А [16], по мнению которого изучение основ машинного обучения поможет будущим учителям сгладить переход в сознании школьника от восприятия функционирования современного цифрового мира как черного ящика к пониманию того, что это интересная и сложная интеграция математики и информатики. В ряде зарубежных университетов: Пенсильванском университете

(США), педагогическом колледже Колумбийского университета, университете Хельсинки проводятся курсы по ИАОД. Обучение студентов интеллектуальному анализу образовательных данных здесь часто строится на чтении и обсуждении научных статей, связанных с ИАД по этой предметной области и выполнении самостоятельных проектов.

В федеральных образовательных стандартах высшего образования РФ к будущим специалистам предъявляются требования к знанию современных методов обработки экспериментальных данных и компьютерных систем обработки данных, а в государственном запросе отмечается актуальность повышения эффективности исследовательской работы и необходимость в подготовке молодых специалистов к инновационной и исследовательской деятельности, что требует развития новых исследовательских компетенций. С нашей точки зрения, приобретение навыков анализа данных и развитие исследовательской компетентности на современном уровне являются тесно связанными задачами. С одной стороны, компетенции по анализу данных являются необходимой составляющей научно-исследовательской компетентности (НИК). С другой стороны, для проведения качественного интеллектуального анализа данных необходимо обладание исследовательской компетентностью, поскольку любая задача ИАД является по сути исследовательской задачей и имеет с исследованием схожие этапы (сбор данных, преобразование данных, анализ данных, оценка результатов).

Возможность обучения ИАОД студентов педагогических направлений с целью развития профессиональных и научно-исследовательских компетенций была рассмотрена в статье [17]. Здесь выде-

лены основные аспекты, способствующие естественным образом воспитанию НИК при освоении курса: специфика аппарата ИАД, навыки сравнения и аналитики, совершенствование англоязычных навыков, развитие поисковых навыков, практика в решении междисциплинарных задач, практика в решении исследовательских задач.

Таким образом, введение в учебную программу курса по интеллектуальному анализу образовательных данных может способствовать устранению противоречий между необходимостью в специалистах, владеющих современными методами анализа образовательных данных и отсутствием соответствующей подготовки в российских вузах, между требуемым уровнем научно-исследовательской компетентности для успешного проведения исследований и недостаточным уровнем НИК выпускников. Целью данной работы является разработка подходов к введению такого курса и подробное описание методической системы обучения ИАОД.

Подходы к отбору контингента обучаемых ИАОД

Прежде, чем остановиться на подробном описании предлагаемой нами методики обучения интеллектуальному анализу образовательных данных, рассмотрим варианты контингента обучаемых, для которых в педагогических вузах может быть развернута учебная дискуссия (или образовательный процесс) по ИАОД.

Мы выделяем три принципиально различных подхода к отбору специалистов для обучения ИАОД:

1. Студенты ИТ направлений. В настоящее время именно для студентов ИТ направлений активно вводятся дисциплины по интеллектуальному анализу данных, машинному обучению, нейрон-

ным сетям и другим смежным дисциплинам. Также открываются отдельные магистратуры, готовящие специалистов по анализу данных. В данном случае предполагается, что ИТ-специалист сможет работать с любыми данными в любых областях. Отдельно специфика анализа образовательных данных не рассматривается.

Минусом данного подхода является необходимость дополнительных временных затрат ИТ-специалиста для изучения образовательной среды, образовательного запроса, стандартов и специфики образовательных систем и данных. Открытыми вопросами являются готовность технических специалистов к работе в сфере образования, обеспеченность участников рынка труда необходимым числом специалистов по анализу данных, готовность и возможность образовательных учреждений оплачивать работу подобных специалистов.

Плюсом является наличие технических знаний и опыта для качественного решения задачи анализа данных после изучения предметной области.

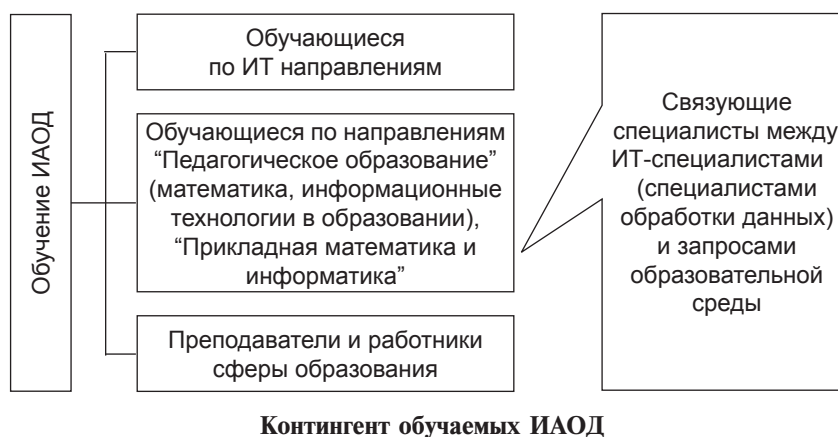
2. Студенты педагогических вузов, обучающихся на математических и информационных направлениях, таких как “Педагогическое образование” (математика, информационные технологии в образовании), “Прикладная математика и информатика” и т.п. В силу педагогической направ-

ленности вуза данная группа будущих специалистов тесно связана с образовательной системой, как правило знакома с тенденциями, существующими проблемами и запросами образования, существующими методиками обучения и образовательными данными. С другой стороны, выбор студентов математических и технических специализаций данного вуза позволит им освоить курс по интеллектуальному анализу образовательных данных, вникая в существующие подходы и алгоритмы.

Минусами данного подхода является более слабая техническая подготовка студентов по сравнению со студентами профильных вузов, необходимость в разработке соответствующего курса и совершенствование программ предшествующих курсов, знания из которых необходимы для освоения ИАД, нехватка преподавателей, готовых поддерживать данную дисциплину.

Плюсами является заинтересованность студентов в работе в области образования и овладении современными технологиями и инструментами для осуществления профессиональной деятельности.

3. Будущие преподаватели и работники сферы образования. Здесь предполагается обучение ИАОД всех будущих (в программе обучения студентов по направлению педагогического образования) и настоящих (в рамках программ повышения



квалификации) специалистов сферы образования.

Минусами является недостаточность знаний дисциплин математического и информационного циклов для освоения методов интеллектуального анализа данных на детальном уровне.

Плюсами — знание о возможных закономерностях в данных и некоторых путях их поиска, современных инструментах ИАД и примерах анализа образовательных данных; возможность применять методы ИАД с помощью соответствующих компьютерных программ для анализа своей деятельности.

Нам представляется оптимальным второй подход, когда интеллектуальному анализу образовательных данных обучаются студенты математических и информационных направлений подготовки педагогических ВУЗов (рисунок). В этом случае образуются специалисты, готовые к анализу данных своей профессиональной деятельности с помощью современных методов и средств, способные передавать в случае работы в учебных заведениях свои знания по ИАД обучаемым, участвовать в организации исследовательской деятельности и готовы выступить при необходимости в качестве посредников между запросами образовательной среды и ИТ-специалистами.

Организация методической системы обучения ИАОД

4.1. Цели обучения. Развитие научно-исследовательских компетенций посредством работы с аппаратом интеллектуального анализа данных, развитие навыков интеллектуального анализа данных, знакомство с интеллектуальным анализом образовательных данных.

Под развитием научно-исследовательских компетенций мы понимаем развитие ка-

честв и умений, являющихся составляющими структурных компонентов НИК. К ним относятся: самостоятельность, саморазвитие, усердность, настойчивость, рефлексивность, целеустремленность, научное общение, способность донести информацию, иноязычное общение, методы исследования, анализ данных, мотивация к исследованиям и др. Детальнее возможности ИАД для развития НИК рассматриваются в работах [17, 18].

4.2. Формы обучения. Развитие информационной цифровой образовательной среды, появление новых форм обучения позволяет сделать более эффективным процесс обучения. В качестве основной формы реализации учебного процесса нами была выбрана форма перевернутого класса. «Перевернутый класс» предполагает самостоятельное изучение традиционного лекционного (теоретического) материала в качестве домашнего задания, в то время как в классе уделяется время активному обучению и практическому закреплению пройденного материала. Учитывая специфику учебного процесса в высших учебных заведениях отойдем от понятий «класс» и «домашнее задание» к понятиям аудиторной и самостоятельной (внеаудиторной) работы.

Мы выделяем следующие преимущества формы обучения «Перевернутый класс» при обучении студентов интеллектуальному анализу данных, способствующие достижению поставленных выше целей:

- Аудиторные часы освобождаются для реализации практической составляющей курса, что является особенно важным при освоении дисциплин информационного цикла.

- В аудитории у преподавателя есть возможность наблюдать процесс выполнения заданий, скорректировать действия студента для достижения требуемого результата, научить

преодолевать трудности в работе.

- Больше возможностей для организации активного обучения.

- Во время аудиторных занятий обучаемые сосредоточены на высших формах когнитивной деятельности в соответствии с перевернутой таксономией Блума, т. е. на применении, анализе, оценке и создании, что способствует развитию научно-исследовательской компетентности под контролем преподавателя, а наиболее низшие уровни познавательной деятельности, такие как запоминание и понимание, практикуются во внеаудиторные часы.

4.3. Методы и средства обучения. При работе студентов с новым материалом мы выделяем несколько последовательных этапов: 1) самостоятельное изучение теоретических материалов, 2) тестирование в аудитории, 3) обсуждение, 4) совместная работа, 5) индивидуальная работа. Таким образом, начинается цикл освоения новой темы с самостоятельного изучения теоретического материала во внеаудиторные часы, что соответствует формату перевернутого обучения. Остальной учебный материал по теме спланирован таким образом, чтобы этапы 2–5 цикла, соответствующие аудиторным занятиям, проходили за 4 академических часа.

Для достижения поставленных целей обучения рекомендуется применять методы обучения, способствующие формированию НИК. В зависимости от этапа текущего цикла нами были использованы следующие методы и средства, показанные в таблице 1.

4.4. Программное обеспечение Weka как средство обучения. Реализация практических заданий на развитие исследовательских компетенций возможна с помощью существующих программных средств

Методы и средства обучения

Этап цикла работы с новым материалом	Методы обучения	Средства обучения
Самостоятельное изучение теоретических материалов	Самостоятельная работа: чтение/изучение электронных материалов.	<ul style="list-style-type: none"> • Разработанные теоретические материалы (включая определения, примеры, алгоритмы) по темам в соответствие с отобранным содержанием. • Среда Moodle для размещения материалов и дистанционной поддержки курса.
Тестирование	Тестирование (в начале каждого аудиторного занятия проводится тестирование, на котором предлагаются вопросы по теоретическому материалу. Вопросы тестов имеют различный характер – на воспроизведение информации, выделение и формулирование основных идей, понимание изученного, сравнение, применение. Тестирование рассматривается нами именно как метод обучения, а не оценки. Вторая функция тестирования – мотивационная к самостоятельному изучению теоретического материала)	<ul style="list-style-type: none"> • Набор тестовых вопросов по каждой теме.
Обсуждение	Обсуждение, объяснение, дискуссия (после этапа тестирования применяются словесные методы обучения, со студентами разбираются вопросы, вызвавшие затруднения)	<ul style="list-style-type: none"> • Набор тестовых вопросов по каждой теме.
Совместная работа	Обсуждение, объяснение, дискуссия. Иллюстрация и демонстрация. Воспроизводящие упражнения (объясняются некоторые теоретические моменты, обсуждаются результаты, полученные при работе в системе при выполнении определенной последовательности действий)	<ul style="list-style-type: none"> • Презентации. Содержат в себе повторение теоретического материал с дополнительными иллюстрациями и примерами, а также инструкции и примеры по работе в программе интеллектуального анализа данных — Weka. Являются промежуточным средством обучения, способствующим переходу от теории к практике. • Weka. Открытое программное обеспечение для интеллектуального анализа данных и машинного обучения.
Индивидуальная работа	Тренировочные упражнения. Исследовательские методы: сравнение, анализ, синтез, эксперимент.	<ul style="list-style-type: none"> • Исследовательские задачи и упражнения на формирование тех или иных исследовательских компетенций. • Методические указания к заданиям. • Weka и другие программы для анализа данных (служат техническими средствами для реализации заданий). • Среда Moodle для размещения материалов и дистанционной поддержки курса.

для интеллектуального анализа данных. Программы ИАД сгруппированы нами следующим образом:

1) программы, включающие в себя возможность ИАД как часть основного или дополнительно устанавливаемого функционала; в настоящее время сюда относятся многие изначально статистические программы (Statistica, Matlab, Deductor, SPSS Statistics и др.);

2) специализированные программы для ИАД, т.е. такие, которые изначально создавались для решения задач именно интеллектуального анализа данных (RapidMiner, Orange, Weka, Knime, и др.);

3) программные среды для программирования ИАД (R, Python и др.).

Для первоначального знакомства с возможностями ИАД было решено использовать программы, разработанные специально для ИАД. Это позволяет сосредоточиться на методах ИАД без перегрузки разнообразным функционалом (в отличие от п. 1) и при этом не погружаться в детали реализации методов и изучение нового языка программирования (в отличие от п. 3).

Среди представителей второй группы программ была выбрана программа Weka – свободно распространяемое ПО разработанное в университете

Вайкато, Новая Зеландия [19]. Эту программу характеризует неперегруженный интуитивно понятный интерфейс и при этом сильный функционал. В Weka реализованы ряд методов для предварительной обработки данных (дискретизация, нормализация, преобразование типа атрибутов, добавление шумов, замещение пропущенных значений, манипуляции с атрибутами и значениями и др.), классификации (ленивые методы – к ближайших соседей и др., на основе деревьев – C4.5, случайный лес и др., на основе правил – таблица решений и др., байесовские методы, методы на основе функций и правил и др.), регрессии

(линейная регрессия, логистическая регрессия, SMO и др.), кластеризации (k -среднее, EM, иерархическая кластеризация и др.), поиска ассоциативных правил (Apriori, FPG). Также предусмотрены современные возможности визуализации результатов, построение потоков работ, работа через командную строку. Отдельно отметим, что в программе, в отличие от известных нам аналогичных средств, есть панель “Экспериментатор”, позволяющая проводить серию экспериментов и сравнивать различные подходы. В мировом научном сообществе Weka часто используется для проведения небольших исследований, как в области образования [20, 21], так и в других областях [22, 23]. Таким образом, в нашей методике программа Weka является ключевым техническим средством обучения.

Студентам в процессе обучения предлагается познакомиться и с другими, отличными от Weka, средствами ИАД. По выбранному программному средству обучаемым необходимо подготовить презентацию, к которой предъявляются следующие требования: в презентации должны быть отражены название программы, электронный адрес и условия доступа, краткое описание функционала, примеры интеллектуального анализа данных, краткий отзыв о работе программного средства. В дальнейшем, после обсуждения презентаций в аудитории, в ряде последующих заданий на выбор предлагается использование данных средств.

4.5. Содержание обучения.

При формировании содержания мы опирались на принципы соответствия цели обучения, профессиональной направленности, фундаментальности, развивающего обучения, доступности содержания, междисциплинарности и практической направленности. В курсе выделяются теорети-

ческая и практическая части. Для практики разработаны задания, направленные на закрепление, понимание и применение изученного материала, освоение программного обеспечения для анализа данных и машинного обучения Weka и развитие научно-исследовательских компетенций. Как правило для коротких курсов в 1–2 семестра используется линейная структура содержания. В разработанной методике для изложения материала выбрана концентрическая структура, которая предполагает возвращение к ранее изученным знаниям, повторение одного и того же вопроса несколько раз с постепенным расширением его содержания, обогащением новыми сведениями, связями и зависимостями [24, с. 65–78]. Это объясняется тем, что студенты педагогических вузов часто имеют по сравнению со студентами технических вузов более слабый уровень подготовки в области прикладной математики и информационных технологий, включающей такие дисциплины как теория алгоритмов, программирование, структуры данных, математическая статистика и др., которые необходимы для

глубокого понимания курса интеллектуального анализа данных. Концентрическая структура изложения содержания призвана облегчить восприятие студентами принципиально нового и сложного для них курса. В содержании курса нами были выделены три концентра, основные цели которых обозначены в табл. 2. В каждом концентре на разном уровне рассматриваются ключевые темы ИАД, как препроцессинг, поиск ассоциативных правил и последовательных шаблонов, регрессия и классификация, кластеризация. Концентрическая модель обучения идейно схожа с успешно себя зарекомендовавшей итерационной моделью разработки программного обеспечения. Каждый концентр представляет собой некоторую итерацию в процессе обучения.

Подробно приведем содержание теоретической части.

Итерация 1: Понятия и основные задачи ИАД. Знакомство с системой Weka. Предварительная обработка данных (примеры задач, преобразование в требуемый формат, дискретизация). Ассоциативные правила (понятие ассоциативного правила, определение

Таблица 2

Цели обучения по итерациям

Номер итерации	Теория	Практика
1 итерация	Знакомство с основными понятиями и задачами ИАД.	Знакомство с программой Weka, выработка навыков поиска новой информации, необходимой для работы в программе.
2 итерация	Углубленное изучение обозначенных на первой итерации задач	Уверенное пользование системой Weka, обзор возможностей системы. Применение изученных методов ИАД для решения задач. Выработка навыков сравнения различных подходов для решения задачи, сопоставления различных форм представления результатов, анализа и критического осмысления результатов.
3 итерация	Задачи интеллектуального анализа данных в образовании. Дополнительный материал	Работа с образовательными данными. Свободное владение изученными методами. Выполнение проектов / исследовательских задач, в т.ч. требующих самостоятельной постановки цели, выбора подхода к решению и анализа результатов.

поддержки и достоверности правила, примеры, алгоритм Априори). Классификация (отличие от задачи регрессии, этапы классификации, понятие ленивых и активных классификаторов, простейшие алгоритмы ZeroR и OneR). Линейная регрессия (постановка задачи, гипотеза и функция стоимости, метод градиентного спуска). Кластеризация (классификация алгоритмов кластеризации: плоские и иерархические, четкие и нечеткие. Алгоритм k -среднее, примеры метрик для определения “похожести” объектов, иерархическая кластеризация и метрики расстояний между кластерами).

Итерация 2: Знакомство с различными программами для ИАД. Предварительная обработка данных (примеры проблем в данных (пропуски, противоречивость, аномальные значения, шумы) и подходы к их устранению, оптимизация данных). Последовательные шаблоны (понятие, алгоритм поиска последовательных шаблонов по R. Agrawal и R.Srikant). Классификация (обобщение формулы Байеса, наивный байесовский классификатор, пример пошаговой работы алгоритма). Кластеризация (классификация алгоритмов кластеризации: методы разбиения, иерархические методы, методы на основе плотности, на основе сетки, на основе моделей. Алгоритм EM/Expectation-Maximization).

Итерация 3: Предварительная обработка данных (выборка атрибутов: методы поиска оптимального подмножества атрибутов, схемо-зависимые и схемо-независимые методы оценки подмножества атрибутов). Классификация текстов. Интеллектуальный анализ данных в образовании (примеры задач и методов ИАД для их решения, разбор научных статей по ИАОД). Материал по выбору (алгоритм классификации C4.5 путем построения

дерева решений, потоки работ, работа в различных системах ИАД и др.)

Таким образом, специфика анализа образовательных данных появляется только в третьем концентре, в то время как на первых двух происходит освоение терминологического аппарата интеллектуального анализа данных, изучение основных задач, методов и алгоритмов. На третьей итерации после знакомства с данными в сфере образования, существующими задачами, интеллектуальным анализом образовательных данных, с изучением возможностей и примеров применения ИАД для образовательных целей, на практике студенты работают в основном с образовательными данными.

Обсуждение результатов исследования

Участниками исследования по апробированию описанной методики выступили бакалавры 4 курса Российского государственного педагогического университета им. А.И. Герцена, обучающиеся по направлению «Прикладная математика и информатика». Курс по интеллектуальному анализу данных в объеме 72 часа проводился в течении трех лет (2016–2018 г.) За три года слушателями курса были 37 человек. Работа по разработанной методике осуществлялась в небольших группах 7–11 человек в связи с низкой наполняемостью групп.

Наше наблюдение в ходе учебного процесса показало, что в начале обучения студенты в целом демонстрируют низкий для 4 курса уровень исследовательской компетентности. Типичными ситуациями являются следующие:

- *Действия студента:* Если при работе с программой появляется системное сообщение об ошибке, то студент сразу обращается за помощью к преподавателю, не пытаясь само-

стоятельно понять данное сообщение, найти информацию о способах ее устранения на тематических форумах в сети Интернет или решить задачу другим способом.

- *Недостаточное развитие следующих составляющих НИК:* самостоятельность, усердность, находчивость, целеустремленность, иноязычное общение (т.к. сообщение на английском языке), поиск информации.

- *Действия студента:* При знакомстве с новым алгоритмом и обнаружении некоторых неизвестных параметров, студентом не осуществляется изучение или поиск описания данных параметров и примеров их задания, а остаются значения, установленные по умолчанию. *Недостаточное развитие следующих составляющих НИК:* критическое мышление, ориентация на результат, поиск информации, алгоритмы и методы анализа данных.

- При выполнении домашнего задания незнакомый термин или мало знакомый теоретический материал пропускается без попыток найти дополнительные источники информации по нему и понять написанное.

- *Недостаточное развитие следующих составляющих НИК:* самостоятельность, саморазвитие, ответственность, рефлексивность, усердность.

На последней третьей итерации в курсе студенты демонстрируют большую заинтересованность, самостоятельность и настойчивость в разрешении проблемных ситуаций, лучше справляются с творческими и исследовательскими заданиями.

После окончания обучения студентам было предложено самостоятельно оценить степень влияния курса на развитие у себя тех или иных качеств, являющихся составляющими научно-исследовательской компетентности. Напротив каждого свойства нужно было проставить балл от 0 до 3, где

0 — никак не повлиял, 1 — незначительно повлиял, 2 — средне повлиял, 3 — сильно повлиял. По мнению студентов, курс значительно поспособствовал повышению уровня знаний в области анализа данных, развитию таких качеств, как самостоятельность, ориентация на результат, эрудиция. Влияние средней степени курса оказал на повышение мотивации к исследованиям, способность донести информацию, усердность, критическое мышление, рефлексивность, целеустремленность, находчивость и способность анализировать.

Другая анкета была направлена на оценку методики обучения и отношения к дисциплине. Результаты показали, что:

- 100% учащихся сочли данный курс полезным;
- 95% учащихся посчитали курс интересным;
- для 95% учащихся форма подачи материала оказалась доступной, для 5% — частично доступной;
- 86% думают, что им могут пригодиться полученные знания, 14% — нет.
- 16% планируют изучать данный предмет далее самостоятельно, 32% — не планируют изучать данный предмет далее самостоятельно и 52% планируют изучать при необходимости в дальнейшей профессиональной деятельности;

В ответах студентов на вопрос «Что Вам понравилось в курсе, показалось особенно интересным?» были выделены практическая направленность курса и его полезность, применение видов анализа на реальных задачах, изучение и разбор алгоритмов, работа в программе. Некоторыми студентами было отмечено, что понравилось самостоятельно изучать новый инструмент ИАД и готовить по нему презентацию, а также слушать выступления одногруппников. Другим примером является ответ: «Проводить мини-исследовательскую работу».

К необходимости самостоятельного изучения теоретического материала, что соответствует формату перевернутого обучения, студенты первоначально относятся сдержанно-негативно. Тем не менее, в процессе курса меняют свое мнение, отмечая, что тестирования носят не оценочный характер, а стимулирующий к предварительному прочтению материала и вспомогательный для определения непонятных моментов и дальнейшего их детального рассмотрения. Приведем мнение из анкеты: «Мне понравилась несколько новая для меня форма работы: тесты, которые поначалу воспринимались негативно, действительно заставляют готовиться и разбираться в материале, что является большим плюсом; основная работа была практической, используя теор. часть, которую разобрали дома, что тоже очень помогает воспринимать материал», «Немного повторяющиеся вопросы в тестовых заданиях помогают лучше отложиться информации в голове». В целом после первых недель напряжение от столкновения с перевернутым обучением спадает, студенты принимают данный формат, спокойно относятся к изучению теории самостоятельно и с готовностью и интересом участвуют в практической работе в классе, что является важным для освоения практически направленного курса.

20% учащихся в течение учебного семестра выразили желание выполнять дипломную работу в области анализа данных. Некоторые студенты применили приобретенные в курсе знания для обработки экспериментальных данных в своих дипломных проектах.

Заключение

В статье для знакомства студентов педагогических вузов с современными возможностями обработки и анализа

экспериментальных данных предлагается ввести курс «Интеллектуальный анализ образовательных данных», в рамках которого студенты изучают теоретические и практические аспекты анализа данных, возможности применения ИАД для решения задач образования. В основе предлагаемой методической системы обучения лежат концентрическая структура содержания, форма перевернутый класс, исследовательские методы обучения. Как показало проведенное исследование, обучение студентов в соответствии с описанной методикой повышает интерес студентов к области анализа данных, позволяет приобрести навыки решения исследовательских задач с помощью современных методов и инструментов анализа данных, способствует повышению мотивации к исследованиям, развивает такие необходимые для научно-исследовательской деятельности качества, как самостоятельность, ориентация на результат, способность донести информацию, усердность, критическое мышление, находчивость, рефлексивность, целеустремленность, способность анализировать.

Дальнейшие перспективы настоящей работы заключаются в создании системы взаимосвязанных курсов, способной подготовить квалифицированного специалиста в области интеллектуального анализа образовательных данных.

Авторы считают актуальными дальнейшие обсуждения в научном сообществе таких вопросов, как подходы и методики обучения студентов анализу данных на современном уровне, целесообразность и возможность обучения студентов гуманитарных направлений методам ИАД, необходимая степень «погружения» непрофильных студентов (не ИТ направлений) в сферу ИАД для выполнения грамотного анализа экспериментальных данных.

Литература

1. Петрова М. В., Ануфриева Д. А. Исследование возможностей методов интеллектуального анализа данных при моделировании образовательного процесса в ВУЗе // Вестник Чувашского университета. 2013. № 3. С. 280–285.
2. Гальченко М.И., Гушинский А.Г. Data Mining и статистические методы при оценивании качества курсов повышения квалификации // Открытое образование. 2014. № 6. С. 15–22.
3. Шамсутдинова Т.М. Оценка качества курсовых работ: компетентностный подход // Образование и наука. 2018. № 2 (20). С. 180–195.
4. Акимов А. А. Проектирование информационно-аналитической системы мониторинга на базе технологий OLAP и Data Mining // Труды Международного симпозиума «Надежность и качество». 2011. Т. 1. С. 235–239.
5. Зорина Н.В., Панченко В.М. Экспериментальный программный комплекс для моделирования и интерпретации процессов анализа образовательных данных // Современные информационные технологии и ИТ-образование. 2017. № 4. С. 207–215.
6. Горлушкина Н.Н., Коцюба И.Ю., Хлопотов М.В. Задачи и методы интеллектуального анализа образовательных данных для поддержки принятия решений // Образовательные технологии и общество. 2015. № 1. С. 472–482.
7. Bakhshinategh B., Zaiane O., Elatia S., Ipperciel D. Educational data mining applications and tasks: A survey of the last 10 years // Education and Information Technologies. 2018. Vol. 23. № 1. P. 537–553. DOI: <https://doi.org/10.1007/s10639-017-9616-z>
8. Romero C., Ventura S. Educational Data Mining: A Review of the State of the Art // IEEE Transactions on Systems Man and Cybernetics. Part C (Applications and Reviews). 2010. V. 40. № 6. P. 601–618. DOI: <https://doi.org/10.1109/TSMCC.2010.2053532>
9. Белоножко П.П., Карпенко А.П., Храмов Д.А. Анализ образовательных данных: направления и перспективы применения [Электрон. ресурс] // Интернет-журнал «Науковедение». 2017. Том 9. № 4. URL: <http://naukovedenie.ru/PDF/15TVN417.pdf> (дата обращения: 18.01.2019)
10. Веряев А.А., Татарникова Г.В. Educational Data Mining и Learning Analytics – направления развития образовательной квалитологии // Преподаватель XXI век. 2016. № 2. С. 150–160.
11. Аbruков В.С., Ефремов Л.Г., Кошчев И.Г. Возможности создания системы поддержки принятия решений и управления вузом с помощью аналитической платформы Deductor // Интеграция образования. 2013. № 1 (70). С. 17–23.
12. Пиотровская К.Р., Тербушева Е.А. Интеллектуальный анализ данных в педагогической аналитике // Техническое творчество молодёжи. 2016. № 2 (96). С. 10–14.
13. Ризаев И.С., Яхина З.Т., Мифтахутдинов Д.И. Компьютерные технологии обучения методам Data Mining обработки данных // Образовательные технологии и общество. 2015. № 2. С. 514–526.
14. Нестеров С. А. Об обучении технологиям интеллектуального анализа данных в рамках курса «Корпоративные системы баз данных» // Преподавание информационных технологий в Российской Федерации. Материалы XIV открытой Всероссийской конференции (г. Санкт-Петербург, 19–20 мая 2016 г.) СПб.: СПбГУ, 2016. С. 78–79.
15. Бодряков В.Ю., Быков А.А. Методические подходы к обучению студентов направления «Прикладная математика и информатика» основам интеллектуальной обработки Больших данных // Педагогическое образование в России. 2016. № 7. С. 145–152.
16. Кузнецов О.А. Курс «Машинное обучение» при подготовке учителей математики и информатики // Преподавание информационных технологий в Российской Федерации. Материалы XVI открытой Всероссийской конференции (14–15 мая 2018 г., Москва) М.: МГТУ им. Н.Э. Баумана, 2018. С. 55–57.
17. Тербушева Е.А. Развитие научно-исследовательской компетентности будущих педагогов для эффективной работы в высокотехнологической образовательной среде // Новые образовательные стратегии в современном информационном пространстве. Сборник научных статей по материалам международной научной конференции (12–26 марта 2018 г., Санкт-Петербург) СПб.: РГПУ им. А. И. Герцена, 2018. С. 178–182.
18. Пиотровская К.Р., Тербушева Е.А. Интеллектуальный анализ данных и развитие научно-исследовательских компетенций бакалавров [Электрон. ресурс] // Современные проблемы науки и образования. 2017. № 1. URL: <http://science-education.ru/ru/article/view?id=26084> (дата обращения: 08.06.2019)
19. Eibe Frank, Mark A. Hall, Ian H. Witten. The WEKA Workbench. Online Appendix for «Data Mining: Practical Machine Learning Tools and Techniques». 4th Ed. Burlington: Morgan Kaufmann. 2016.
20. Hajizadeh N., Ahmadzadeh M., Analysis of factors that affect students' academic performance – Data Mining Approach [Электрон. ресурс] // International Journal of advanced studies in Computer Science and Engineering. 2014. V. 3. I. 8. URL: http://www.ijascse.org/volume-3-issue-8/Data_mining_approach.pdf
21. Sunita B. Aher, Louis Lobo. Applicability of data mining algorithms for recommendation system in e-learning // Proceedings of the International

Conference on Advances in Computing, Communications and Informatics (ICACCI '12). New York: ACM, 2012. P. 1034–1040.

22. Brenda Reyes Ayala, Jiangping Chen. A machine learning approach to evaluating translation quality // Proceedings of the 17th ACM/IEEE Joint Conference on Digital Libraries (JCDL '17). Piscataway: IEEE Press, 2017. 281–282.

References

1. Petrova M. V., Anufriyeva D. A. Study of the possibilities of data mining methods in modeling the educational process in the university. *Vestnik CHuvashskogo universiteta = Bulletin of Chuvash University*. 2013; 3: 280–285. (In Russ.)

2. Gal'chenko M.I., Gushchinskiy A.G. Data Mining and statistical methods in assessing the quality of refresher courses. *Otkrytoye obrazovaniye = Open Education*. 2014; 6: 15–22. (In Russ.)

3. SHamsutdinova T.M. Assessment of the quality of coursework: the competence approach. *Obrazovaniye i nauka = Education and Science*. 2018; 2 (20): 180–195. (In Russ.)

4. Akimov A. A. Designing an Information-Analytical Monitoring System Based on OLAP and Data Mining Technologies. *Trudy Mezhdunarodnogo simpoziuma "Nadezhnost' i kachestvo" = Proceedings of the International Symposium "Reliability and Quality"*. 2011; 1: 235–239. (In Russ.)

5. Zorina N.V., Panchenko V.M. Experimental software for modeling and interpretation of educational data analysis processes. *Sovremennyye informatsionnyye tekhnologii i IT-obrazovaniye = Modern information technologies and IT education*. 2017; 4: 207–215. (In Russ.)

6. Gorlushkina N.N., Kotsyuba I.YU., KHlopotov M.V. Objectives and methods of intellectual analysis of educational data to support decision-making. *Obrazovatel'nyye tekhnologii i obshchestvo = Educational technologies and society*. 2015; 1: 472–482. (In Russ.)

7. Bakhshinategh B., Zaiane O., Elatia S., Ipperciel D. Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*. 2018; 23; .1; 537–553. DOI: <https://doi.org/10.1007/s10639-017-9616-z>

8. Romero C., Ventura: Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems Man and Cybernetics. Part C (Applications and Reviews)*. 2010. V. 40 (6); 601–618. DOI: <https://doi.org/10.1109/TSMCC.2010.2053532>

9. Belonozhko P.P., Karpenko A.P., KHramov D.A. Analysis of educational data: directions and prospects of application [Internet]. *Internet-zhurnal «Naukovedeniye» = Internet magazine «Science»*. 2017; 9 (4). URL: <http://naukovedenie.ru/PDF/15TVN417.pdf> (Cited: 18.01.2019) (In Russ.)

23. Tarek Elghazaly, Amal Mahmoud, and Hesham A. Hefny. Political Sentiment Analysis Using Twitter Data // Proceedings of the International Conference on Internet of things and Cloud Computing (ICC '16). New York: ACM, 2016. 5 p.

24. Подласый И.П. Педагогика. Книга 2. 2-е изд. М.: Владос, 2007. 575 с.

10. Veryayev A.A., Tatarnikova G.V. Educational Data Mining and Learning Analytics - directions for the development of educational quality. *Prepodavatel' KHKHI vek = Teacher of the XXI century*. 2016; 2: 150–160. (In Russ.)

11. Abrukov V.S., Efremov L.G., Koshcheyev I.G. . Opportunities for creating a decision support system and university management using the Deductor analytical platform. *Integratsiya obrazovaniya = Education Integration*. 2013; 1 (70): 17–23. (In Russ.)

12. Piotrovskaya K.R., Terbusheva E.A. Intellectual data analysis in pedagogical analytics. *Tekhnicheskoye tvorchestvo molodëzhi = Technical creativity of young people*. 2016; 2 (96): 10–14. (In Russ.)

13. Rizayev I.S., Yakhina Z.T., Miftakhutdinov D.I. Computer technologies of data mining methods of data processing. *Obrazovatel'nyye tekhnologii i obshchestvo = Educational technologies and society*. 2015; 2: 514–526. (In Russ.)

14. Nesterov: A. About training in data mining technologies in the course «Corporate database systems». *Prepodavaniye informatsionnykh tekhnologiy v Rossiyskoy Federatsii. Materialy XIV otkrytoy Vserossiyskoy konferentsii = Teaching information technologies in the Russian Federation. Proceedings of the XIV Open All-Russian Conference (St. Petersburg, May 19–20; 2016)* Saint Petersburg: SPbSU; 2016: 78–79. (In Russ.)

15. Bodryakov V.YU., Bykov A.A. Methodical approaches to teaching students of the direction “Applied Mathematics and Computer Science” to the basics of intellectual processing of Big Data. *Pedagogicheskoye obrazovaniye v Rossii = Pedagogical education in Russia*. 2016; 7: 145–152. (In Russ.)

16. Kuznetsov O.A. The course “Machine learning” in the preparation of teachers of mathematics and computer science. *Prepodavaniye informatsionnykh tekhnologiy v Rossiyskoy Federatsii. Materialy XVI otkrytoy Vserossiyskoy konferentsii = Teaching information technologies in the Russian Federation. Proceedings of the XVI Open All-Russian Conference (May 14–15; 2018, Moscow)* Moscow: N.E. Bauman MSTU; 2018: 55–57. (In Russ.)

17. Terbusheva E.A. Development of research competence of future teachers for effective work in a high-tech educational environment.

Novyye obrazovatel'nyye strategii v sovremennom informatsionnom prostranstve. Sbornik nauchnykh statey po materialam mezhdunarodnoy nauchnoy konferentsii = New educational strategies in the modern information space. Collection of scientific articles on the materials of the international scientific conference (March 12-26; 2018, St. Petersburg) Saint Petersburg: A. I. Gertsen RSPU; 2018: 178-182. (In Russ.)

18. Piotrovskaya K.R., Terbusheva E.A. Intellectual data analysis and development of research skills of bachelors [Internet]. *Sovremennyye problemy nauki i obrazovaniya*. = Modern problems of science and education. 2017; 1. URL: <http://science-education.ru/ru/article/view?id=26084> (Cited: 08.06.2019) (In Russ.)

19. Eibe Frank, Mark A. Hall, Ian H. Witten. The WEKA Workbench. Online Appendix for «Data Mining: Practical Machine Learning Tools and Techniques». 4th Ed. Burlington: Morgan Kaufmann. 2016.

20. Hajizadeh N., Ahmadzadeh M., Analysis of factors that affect students' academic performance -

Data Mining Approach [Internet]. *International Journal of advanced studies in Computer Science and Engineering*. 2014. V. 3. I. 8. URL: http://www.ijascse.org/volume-3-issue-8/Data_mining_approach.pdf

21. Sunita B. Aher, Louis Lobo. Applicability of data mining algorithms for recommendation system in e-learning. *Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI '12)*. New York: ACM; 2012; 1034-1040.

22. Brenda Reyes Ayala, Jiangping Chen. A machine learning approach to evaluating translation quality. *Proceedings of the 17th ACM/IEEE Joint Conference on Digital Libraries (JCDL '17)*. Piscataway: IEEE Press; 2017. 281-282.

23. Tarek Elghazaly, Amal Mahmoud, and Hesham A. Hefny. Political Sentiment Analysis Using Twitter Data. *Proceedings of the International Conference on Internet of things and Cloud Computing (ICC '16)*. New York: ACM; 2016. 5 p.

24. Podlasy I.P. *Pedagogika = Pedagogy*. Book 2. 2nd ed. Moscow: Vldos; 2007. 575 p. (In Russ.)

Сведения об авторе

Екатерина Александровна Тербушева
ассистент кафедры методики обучения
математике и информатике
Российский государственный педагогический
университет им. А.И. Герцена,
Санкт-Петербург, Россия
Эл. почта: terbushevae@gmail.com

Information about the author

Ekaterina E. Terbusheva
Assistant of the Department of Mathematics
and Informatics Teaching
Herzen State Pedagogical University,
Saint-Petersburg, Russia
E-mail: terbushevae@gmail.com