

ИСПОЛЬЗОВАНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ ПОЛУЧЕНИЯ ЗНАНИЙ ОБ ОБРАЗОВАТЕЛЬНОМ ПРОЦЕССЕ

Аннотация.

Исторически было трудно изучить, насколько сильно различия между преподавателями и группами обучающихся влияют на конкретные аспекты учебного опыта; этот вид анализа становится намного проще с использованием методов интеллектуального анализа образовательных данных. В статье отражена важность применения методов интеллектуального анализа данных, возникающих в рамках образовательного процесса, для поддержки принятия решений по управлению образовательным процессом. Изложены основные задачи и методы интеллектуального анализа данных в сфере образования, а также дан тематический обзор работ, посвященных данному направлению.

Ключевые слова: интеллектуальный анализ данных, образовательная аналитика, машинное обучение, адаптивное обучение, educational data mining.

1. Предисловие

Принятие эффективных решений по управлению образовательным процессом – сложная задача, с которой ежедневно сталкиваются административные работники научных школ. Для решения этой задачи требуется собирать и анализировать информацию, поступающую от лиц, участвующих в образовательном процессе, на всех его этапах.

Методы и инструменты сбора, анализа данных, которые когда-то ограничивались исследовательскими лабораториями, были приняты перспективными отраслями промышленности для создания бизнес-аналитики и в целях улучшения процессов принятия решений. Некоторые крупные университеты уже успешно используют данные методы, они позволяют лучше

понимать учащихся, узнавать, какие предметы вызывают наибольшие затруднения, как лучше построить курс, чтобы получить максимально высокие баллы, с какими тестами студенты справляются, какую форму занятий предпочитают, в какой области научных интересов преуспевают.

В октябре 2012 года Министерство образования США опубликовало отчет о необходимости использования методов интеллектуального анализа данных для совершенствования обучения [1]. Краткое описание проблемы призвано помочь руководству научных школ и контролирующим ведомств понять, как аналитика и интеллектуальный анализ данных могут быть применены для улучшения образования.

2. Интеллектуальный анализ образовательных данных

Интеллектуальный анализ данных (от англ. Data mining, далее — ИАД), также называемый «Поиском знаний в базах данных» (от англ. Knowledge Discovery in Databases), это область, занимающаяся обнаружением новой и потенциально полезной информации в больших объемах данных. ИАД применяется во многих областях, включая торговлю на бирже, биоинформатику и борьбу с терроризмом. За последние годы, повысился интерес в использовании ИАД для изучения научных вопросов в рамках образовательных исследований. Интеллектуальный анализ образовательных данных (от англ. Educational data mining, далее — ИАОД) определяется как область научного исследования, сосредоточенная вокруг разработки методов совершения открытий в уникальных типах данных, поступающих из образовательных учреждений, и использования этих методов для лучшего понимания учащихся и окружения, в котором они учатся.

В интеллектуальном анализе образовательных данных применяются методы машинного обучения, статистики, глубинного анализа данных, психопедагогики, поиска информации, когнитивной психологии, рекомендательных систем и их методов к различным наборам образовательных данных для решения задач обучения [2]. Международное объединение

интеллектуального анализа образовательных данных¹ определяет ИАОД как «развивающуюся дисциплину, занимающуюся разработкой методов исследования уникальных типов данных, поступающих из образовательной среды, и использование методов для обнаружения взаимосвязей с целью получения практически полезных знаний об образовательном процессе и его участниках. Хотя глубинный анализ данных является известной областью исследований в области естественных наук, применение его в образовательном контексте ограничено. [3]

ИАОД может быть представлен как комбинация трех основных областей: *компьютерные науки, образование и статистика*. [Рис.1] Кроме того, пересечение данных областей формирует близкие к ИАОД темы, такие как компьютеризированное обучение, интеллектуальный анализ данных и машинное обучение, образовательная аналитика.

ИАОД и образовательная аналитика (*от англ. Learning analytics*) позволяют сделать видимыми данные, которые были незаметны и неинтерпретируемы ранее.

В области ИАОД применены и испытаны различные методы обработки данных и извлечения знаний. Предполагается, что в связи с иерархической структурой и независимым характером данных об образовании, стандартные методы ИАД не могут быть использованы в ИАОД [4].

3. Основные методы ИАОД

В рамках ИАОД широко распространено множество современных методов. Эти методы подразделяются на следующие общие категории: предсказание, кластеризация, исследование отношений, открытие с моделями и перегонка данных для суждения человека [5].

¹ <http://www.educationaldatamining.org/> - International Educational Data Mining Society



Рис.1. Основные области, связанные с ИАОД

Первые три категории в значительной степени признаны универсальными в разных областях ИАД (хотя в некоторых случаях называются они по-разному), однако четвертый и пятый подходы относятся именно к анализу данных в образовательных доменах. (Табл.1).

Таблица 1. Основные методы ИАОД

Метод	Описание метода
Предсказание	Разработка модели, в которой вывод одиночного аспекта данных (предсказываемой переменной) осуществляется на основе комбинации других аспектов (переменных-предсказателей).
Кластеризация	Нахождение точек, которые естественным образом соединены вместе, разделение данных на кластеры. Алгоритм позволяет прогнозировать успешность учащихся, ассоциировать стиль обучения разных типов учеников и их поведение. Применяется, когда ничего не

	известно о структуре данных.
Анализ взаимоотношений	Изучение отношений между переменными в наборе из многих переменных.
Исследование с помощью моделей	Изучение предварительно составленных моделей (разработанных с помощью методов прогнозирования ИАОД, кластеризации, инженерии знаний)
Преобразование данных к виду, понятному человеку	Преобразование сложных данных к виду, понятному для человека, для использования данных в человеческих суждениях

4. Меры и способы развития ИАОД

В информационных системах учебных заведений на данный момент накоплено достаточное количество информации о различных аспектах образовательного процесса: успеваемость студентов, научно-образовательная работа преподавателей, результаты тестирования и анкетирования обучающихся [6]. Этого объема информации достаточно для выявления взаимосвязей в данных, их интерпретации, а также применения методов статистики, машинного обучения, которые будут полезными для всех участников образовательного процесса: обучающихся, преподавателей, разработчиков учебных курсов, административных работников в сфере образования и методистов. [Табл.2] В частности, накопленные данные о разных аспектах работы образовательного учреждения могут быть проанализированы и использованы для поддержки принятия эффективных решений по управлению образовательным процессом.

Таблица 2. Участники образовательного процесса и их интересы.

Участники образовательного	Назначение
----------------------------	------------

процесса	
Обучающиеся	Улучшение успеваемости, получение адаптивной обратной связи и рекомендаций по индивидуализации образовательной траектории.
Преподаватели	Понимание социальных, поведенческих и когнитивных аспектов учебного процесса; применение адекватных методов и приемов обучения
Исследователи	Оценка эффективности и результативности учебного процесса, развития и сравнения между собой методик ИАОД
Администрация	Находить способы наилучшей организации ресурсов на уровне учебного заведения

Чтобы повысить эффективность дальнейших исследований и ценность практического применения ИАОД, преподаватели и административные сотрудники должны разработать и поддерживать культуру использования данных для принятия учебных решений; привлекать ИТ-отделы к процессу накопления и исследования данных.

Примерная схема работы участников образовательного процесса:

1. Ученик работает в электронной обучающей системе, система предлагает персонализированный контент в зависимости от его интересов и способностей.
2. Детализированная информация о работе учащегося в системе сохраняется в базе данных.
3. Эта информация используется для прогнозирования успешности учащегося в будущем.
4. Предсказания и обратная связь отображается в графическом виде на информационной панели учащегося.

5. Ученик получает материалы в соответствии с его уровнем подготовки.

6. Преподаватели и администрация может вмешиваться в процесс обучения для помощи учащимся.

5. Направления исследовательских работ в ИАОД

После анализа статей, опубликованных в научных журналах, можно выделить основные направления работ (Табл. 3), связанных с интеллектуальным анализом образовательных данных. Приведены примеры работ по каждому из направлений.

Таблица 3. Участники образовательного процесса и их интересы.

Направление исследования	Цель работы
Анализ портфолио обучающегося	<ul style="list-style-type: none">- рекомендация лучшей комбинации курсов для учащихся [7].- определение, как выбор собранных экземпляров и атрибутов, которые влияют на точность и понятность прогноза. (114 студентов первого курса) [8].
Групповое обучение	<ul style="list-style-type: none">- нахождение активных и пассивных сотрудников в группе [9].- определение способности обучаться совместно [10]. <p>(116 учащихся задействовано, по результатам эксперимента способности выросли на 55%, в контрольной группе – на 11%)</p>
Моделирование обучающихся	<ul style="list-style-type: none">- моделирование профиля учащегося в онлайн-дискуссионных форумах [11]. (672 обучающихся из 18 групп обучения)- персонализация электронной среды обучения с помощью моделирования личности [12].

<p>Планирование экзаменационной сессии</p>	<p>- измерение поверхностности обучения студентов [13]. (71 студент)</p>
<p>Предсказание успешности сдачи экзамена</p>	<p>- определение успеваемости учащихся на основе их истории обучения (2228 результатов экзаменов в UTN, Малайзия. Результаты показывают точность правильного предсказания 70,17%) [14].</p>
<p>Стиль обучения</p>	<p>- определение ключевых факторов, необходимых для успешного обучения (база данных учебной подготовки персонала Китайской автомобильной корпорации) [15].</p> <p>- выбор места для студентов в классе и его последствия для их оценок (220 студентов Университета Нови-Сад. Результаты эксперимента показали, что студенты, выбирающие просторное расположение набирают в среднем на 10% больше баллов в течение семестра) [16].</p>
<p>Электронное обучение</p>	<p>- получение диаграммы социальной сети в действиях студента по электронному обучению (Moodle LMS Силезийского университета) [17].</p> <p>- выявление и фиксация эффективности поведения студентов при взаимодействии с системой (лог-записи обучающей системы и eye-трекинг) [18].</p>

Таким образом, ИАОД предоставляет мощный инструментарий для извлечения информации и ее преобразования в форму, необходимую для последующего использования. Использование таких методов в

образовательном процессе может решить множества проблем, таких как понимание студентов, улучшение качества курсов, уменьшение затрат на организацию процесса обучения.

В работе было показано, что благодаря большому объему данных в образовательном процессе (базы данных студентов с информацией об успеваемости, предыдущих этапах обучения, программы курсов, данные дистанционные системы обучения и др.), существует нерешенных задач для анализа и исследования в данной области.

СПИСОК ЛИТЕРАТУРЫ

1. Enhancing Teaching and Learning Through *Educational Data Mining* and *Learning Analytics: An Issue Brief* [Электронный ресурс] // Официальный сайт Министерства образования США. URL: <https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf> (дата обращения: 10.04.2017)
2. Romero C., Ventura S. Educational data mining: a review of the state of the art // *Systems, Man, and Cybernetics, Part C: Applications and Reviews*, IEEE Transactions, 2010. Vol. 40. pp. 601-618.
3. Ranjan J., Malik K. Effective educational process: a data-mining approach // *Vine*, 2007. Vol. 37. pp. 502-515.
4. Baker R. S., Yacef K. The state of educational data mining in 2009: A review and future visions // *JEDM-Journal of Educational Data Mining*, 2009.
5. Хлопотов М.В. Модели и алгоритмы интеллектуального анализа образовательных данных для поддержки принятия решений: Дисс. ... канд. истор. наук. Санкт-Петербург, 2014.
6. Веряев А.А., Татарникова Г.В. Educational Data Mining и Learning Analytics – направления развития образовательной квалитологии // *Преподаватель XXI век*, 2016. № 2. с. 150-160.

7. Aher S. B., Lobo L. Applicability of Data Mining Algorithms for Recommendation System in E-Learning // Proc. International Conference on Advances in Computing, Communications and Informatics, 2012.
8. Romero C., López M. I., Luna J. M., Ventura S. Predicting students' final performance from participation in on-line discussion forums // Computers & Education, 2013. Vol. 68. pp. 458-472.
9. Anaya A. R., Boticario J. G. Clustering Learners according to their Collaboration // Proc 13th Int Conf "Computer Supported Cooperative Work in Design, 2009.
10. Chang W. C., Wang T. H. Learning Ability Clustering in Collaborative Learning // Journal of Software, 2010. Vol. 5. pp. 1363-1370.
11. Cobo G., García-Solórzano D., Morán J. A. Using agglomerative hierarchical clustering to model learner participation profiles in online discussion forums // Proc. 2nd International Conference on Learning Analytics and Knowledge, 2012.
12. Feng T., Shibin W., Cheng Z., Qinghua Z. Research on e-learner personality grouping based on fuzzy clustering analysis // Proc. 12th Int. Conf. Computer Supported Cooperative Work in Design, 2008. pp. 1035-1040
13. Baker R. S., Gowda S. M. Towards automatically detecting whether student learning is shallow // Intelligent Tutoring, 2012.
14. Wook M., Yahaya Y. H., Wahab N., Isa M. R. M., Awang N. F., Seong H. Y. Predicting NDUM Student's Academic Performance Using Data Mining Techniques // Proc. 2nd Int. Conf. Computer and Electrical Engineering, 2009. Vol. 2. pp. 357-361.
15. Dharmarajan A., Velmurugan T. Applications of Partition based Clustering Algorithms: A Survey // Proc Int. Conf. Computational Intelligence and Computing Research, 2013.

16. Ghorbani F., Montazer G. A. Learners grouping improvement in e-learning environment using fuzzy inspired PSO method // Proc. 3rd Int. Conf. E-Learning and E-Teaching, 2012.
17. Ivancevic V., Celikovic M., Lukovic I. The Individual Stability of Student Spatial Deployment and its Implications. Proc. Int. Sym. Computers in Education, 2012.
18. Amershi S., Conati C. Combining unsupervised and supervised classification to build user models for exploratory learning environments // Journal of Educational Data Mining, 2009. Vol. 1. No. 1.