

МЕТОДОЛОГИЯ АНАЛИЗА БОЛЬШИХ ДАННЫХ В ОБРАЗОВАНИИ¹

Системно-методологический подход, основанный на анализе образовательных данных, поиска стратегии принятия управленческих и организационно-педагогических решений в образовании

Кирилл Анатольевич Баранников,

проректор по развитию Московского городского педагогического университета, кандидат педагогических наук, Москва, e-mail: barannikovka@mgpu.ru

Сергей Михайлович Лесин,

ведущий научный сотрудник лаборатории исследования образовательной политики управления стратегического развития Московского городского педагогического университета, кандидат педагогических наук, Москва, e-mail: lesinsm@mgpu.ru

В статье выявляются ключевые тенденции анализа образовательных данных (Educational Data Mining), применяемых при обработке производимой образовательными организациями информации, и применение его результатов для принятия организационно-педагогических и управленческих решений в образовании. Даются анализ и обобщение зарубежного и российского опыта при организации и использовании анализа образовательных данных (EDM) в системе образования. Рассматриваются ключевые понятия и методы анализа образовательных данных (EDM), его отличие и специфика по сравнению с традиционным интеллектуальным анализом данных (DM) с учетом особенности организации образовательной деятельности в системе образования. Авторы выделяют стратегии и направления использования результатов анализа образовательных данных в зависимости от деятельности основных участников образовательного процесса. Сделанные выводы подтверждают ключевые возможности применения анализа образовательных данных для систем управления обучением и способов их систематизации для принятия организационно-педагогических и управленческих решений в системе образования.

- анализ образовательных данных • Educational Data Mining (EDM)
- интеллектуальный анализ данных • Data Mining (DM) • Московская электронная школа (МЭШ)

¹ Статья публикуется в рамках гранта РФФИ на реализацию научного проекта № 19-29-14016-мк «Методология анализа больших данных в образовании и ее интеграция в программы профессиональной подготовки педагогов и руководителей общеобразовательных организаций в логике “Педагогика, основанная на данных”, “Управление образованием на основании данных”» (договор № 19-29-14016/19 от 18.11.2019).

Введение

В последние годы наблюдается тенденция повышения интереса к использованию анализа образовательных данных или интеллектуального анализа образовательных данных при принятии управленческих решений в рамках функционирования различных информационных автоматизированных систем, применяемых в системе образования. Появление современных компьютерных обучающих сред, анализ и отслеживание цифровых следов обучающихся, а также их поведения в этих средах позволяют получить новые данные, которые необходимо агрегировать, систематизировать, обработать, чтобы сформировать систему персонализированного обучения.

Метод интеллектуального анализа образовательных данных позволяет специалистам отслеживать учебные события, цифровые следы основных субъектов образовательной деятельности в нелинейной среде, не нарушая сам процесс обучения учащегося или его навигации по образовательному контенту. Данные, полученные в режиме реального времени, позволяют виртуально воссоздать или смоделировать прогноз действий учащихся, определить, как они конструируют знания и отслеживают свой фактический выбор, а также выявить методы, которые они используют для выражения данного выбора посредством саморегулируемого обучения в конкретном контексте.

Следует немного разобраться с терминологией, имеющей отношение к анализу образовательных данных или Educational Data Mining (далее — EDM), который напрямую связан с самим понятием «интеллектуальный анализ данных» или Data Mining, то есть анализом или «добычей» данных. Так, термин Data Mining часто переводится как добыча данных, извлечение информации, раскопка данных, интеллектуальный анализ данных, средства поиска закономерностей, извлечение знаний, анализ шаблонов, извлечение зёрен знаний из гор данных, раскопка знаний в базах данных, информационная проходка

данных, промывание данных. Понятие «обнаружение знаний в базах данных» (Knowledge Discovery in Databases, KDD) можно считать синонимом DM².

Тем не менее сам термин «анализ образовательных данных» связан с обработкой больших данных, где целью поиска становится набор паттернов как разновидности различных данных в виде образцов, схем, закономерностей действий и т.п. в образовательной деятельности. Утверждается, что большие данные (Big Data) позволяют осуществлять влияние на успеваемость учащихся и понимать, как лучше организовать учебный процесс, что технология интеллектуального анализа данных (Data Mining, DM) разнообразит методики обучения, добавляет в учебный процесс больше нюансов. Иногда интеллектуальный анализ данных определяют как процесс поддержки принятия решений, основанный на поиске в данных скрытых закономерностей. Если данные имеют отношение к образованию, управленческие решения направлены на повышение эффективности учебно-воспитательного процесса, на достижение целей, которые поставлены перед современной школой³.

Метод извлечения последовательных паттернов в электронном обучении также вносит свой вклад в образовательные исследования, посвящённые анализу образовательных данных для принятия управленческих решений. Он помогает проанализировать деятельность учащегося и, соответственно, спрогнозировать адаптацию и настройку предоставления ресурсов в рамках электронной образовательной среды, сравнить с ожидаемыми поведенческими паттернами и теоретически приемлемыми траекториями обучения в этой связи. Всё это необходимо, чтобы лучше организовать образовательные интерфейсы

² Веряев А.А., Татарникова Г.В. Educational data mining и learning analytics — направления развития образовательной квалитологии // Преподаватель XXI век. — 2016. — Т. 1. — № 2. — С. 150–160.

³ Там же.

и иметь возможность предлагать учащимся, имеющим сходные образовательные характеристики, генерировать персонализированные действия для различных групп учащихся, идентифицировать последовательности взаимодействия, указывающие на проблемы и паттерны, которые являются маркерами успеха, рекомендовать учащимся наиболее подходящие ссылки на ресурсы или конкретные веб-страницы в рамках адаптивного веб-курса для последующего посещения, совершенствовать системы электронного обучения в целом и выявлять предикторы успеха путём анализа всех записей учащихся в электронной обучающей среде. Все эти примеры предполагают, что интеллектуальный анализ образовательных данных имеет большой потенциал для расширения гораздо более широкого набора инструментов для анализа важных вопросов в системе образования.

Следует учесть, что понятие «анализ образовательных данных» (EDM) является приложением к методам интеллектуального анализа данных (DM), применяемых к образовательным данным. Его цель состоит в том, чтобы проанализировать эти типы данных для решения образовательных исследовательских задач и принятия в дальнейшем организационно-педагогических и управленческих решений.

Тем не менее известным является факт, что сам термин EDM в широком смысле определяется веб-сайтом образовательного сообщества интеллектуального анализа данных (www.educationaldatamining.org) как развивающаяся дисциплина, занимающаяся разработкой методов исследования уникальных типов данных, поступающих из образовательной среды, и использованием этих методов для лучшего понимания студентов и условий, в которых они учатся. EDM часто подчёркивает улучшение студенческих моделей, которые обозначают текущие знания студента, мотивацию и отношения³.

Исследование и методология

Анализ образовательных данных строится на основе концепции шаблонов (patterns) пове-

³ Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.

дения и личностных качеств учащихся⁴. Среди основных паттернов могут быть следующие: информация о предметах, вызывающих большие затруднения (набор различных параметров); информация об эффективных тестах, с которыми учащиеся лучше всего справляются; предпочтения в выборе форм, темпа и времени обучения; информация о предпочтениях в изучаемых темах и способе (последовательности) их изучения учащимися; паттерн компетенций, которые учащемуся необходимы в сфере его будущей профессиональной деятельности.

Примером применения анализа образовательных данных может быть следующая задача. Имеются ли примеры (шаблоны) полученных оценок выпускниками, которые затем смогли бы в течение короткого времени после окончания вуза найти работу, удовлетворяющую их требованиям⁵.

Среди основных методов, которые используются при осуществлении анализа образовательных данных, можно выделить следующие⁶.

1. Классификация — отнесение объектов (наблюдений, событий) к одному из заранее известных классов. Для классификации в Data Mining используется множество различных моделей: нейронные сети, деревья решений, метод k-ближайших соседей.
2. Регрессия, в том числе метод прогнозирования. Применение регрессионных методов позволит смоделировать влияние, которое оказывает изменение одного из параметров на другой,

⁴ Bishop Ch. Pattern Recognition and Machine Learning. Series: Information Science and Statistics. 2006. Т. XX. 740 p.

⁵ Мамедова Г.А., Зейналова Л.А., Меликова Р.Т. Технологии больших данных в электронном образовании. *Открытое образование*. 2017; (6):41-48. <https://doi.org/10.21686/1818-4243-2017-6-41-48>

⁶ Там же.

установить зависимость выходных параметров (целевых функций) от входных переменных (факторов).

3. Кластеризация — разделение всего множества объектов (наблюдений, событий), которые наиболее близки друг к другу по ряду признаков или свойств, на кластеры. Внутри каждого кластера должны оказаться похожие объекты, а в разных кластерах — объекты должны отличаться. Алгоритм действий при применении кластерного анализа включает несколько основных этапов: выбрать объекты для кластеризации; определить множество признаков, по которым будут оцениваться объекты в выборке; применить один из методов кластерного анализа для создания групп сходных объектов (кластеров); визуализировать результаты анализа.

4. Ещё одним важным направлением анализа образовательных данных может стать метод анализа данных социальных сетей. Вполне очевидно, что современные учащиеся большую часть времени проводят вне учебных заведений, тратя его в социальных сетях. Они общаются со сверстниками, обмениваются информацией, сотрудничают, создают совместный сетевой контент, который в дальнейшем может быть использован в образовательной деятельности. Особенность социальных сетей такова, что учащиеся находятся в более комфортных условиях, чем в реальности (не находятся под контролем преподавателя), поэтому анализ информации, извлечённой из социальных сетей, в отличие от других источников, может быть более объективным.

5. Частными методами анализа образовательных данных может служить целый ряд методов статистики, применяемых к информации, производимой образовательными организациями и образовательными электронными платформами. То есть анализ строится на основе цифровых следов, оставляемых участниками образовательной деятельности, например, в таких электронных платформах, как Московская электронная школа

(МЭШ)⁷ или электронный дневник, а также сайт образовательной организации или единый сетевой ресурс, используемый в образовательной организации для управления процессом обучения, включая дистанционное.

Результаты и их обсуждение

Говоря о методике подхода, основанном на анализе образовательных данных (АОД), можно заключить, что он использует большие блоки (паттерны) наборов образовательных данных для качественного понимания процесса обучения в целом и предоставления целостной информации о нём. Отличие метода АОД от любых других, которые также анализируют и обобщают большие данные, заключается в том, что он обычно использует данные от учащихся, изучающих школьные предметы, их цифровые следы, часто в течение всего учебного года. Например, можно наблюдать, как учащиеся формируют навык в течение длительного интервала времени в процессе обучения, и делать выводы о том, какие виды деятельности приводят к лучшему долгосрочному обучению, узнать о влиянии времени, с которым они начинают выполнять домашнее задание, на успеваемость в классе, или понять, как продолжительность времени, которое они тратят на чтение отзывов о своей работе, влияет на качество их последующих усилий.

Для осуществления анализа образовательных данных используются данные из таких источников, как: интеллектуальные компьютерные тьюторские системы (Intelligent Tutors Systems, ITS); традиционные электронные образовательные системы, построенные на использовании компьютерных автоматизированных информационных

⁷ Заславская О.Ю., Кац С.В., Махотин Д.А. Подходы к описанию модели проектирования сценариев уроков по технологии на портале «Московской электронной школы» / О.Ю. Заславская, С.В. Кац, Д.А. Махотин // Вестник Московского городского педагогического университета. Серия: Информатика и информатизация образования. 2019. — № 4 (50). — С. 64–72.

систем⁸; форумы для обсуждения в рамках онлайн-класса, электронные учебники для учителей; образовательные данные о зачислении учащихся в школу и стандартизированных тестов о поступлении и др.⁹ Несмотря на то что многие из этих источников данных существуют уже много лет, благодаря современным информационным технологиям, сбор и сопоставление больших наборов данных по этим источникам значительно упростился и улучшился. Что тем самым произвело революцию в изучении особенностей процесса обучения, а объём и рост образовательных данных позволяет совершенно новому прогнозировать управленческие решения в системе образования.

Подобная тенденция наблюдается во многих сферах научного знания. Аналогично анализ образовательных данных использует доступные источники данных, так же как более 20 лет назад в биоинформатике. Тогда всплеск доступных данных произвёл революцию таким образом, что было проведено много исследований, причём были применены к обработке данных современные методы информатики, такие как интеллектуальный анализ данных и распознавание образов. Всё больше школ сегодня используют компьютерное программное обеспечение, которое способно записывать для последующего анализа каждое действие учащегося и результаты его взаимодействия с электронной обучающей средой или компьютером. В системе российского образования всё чаще можно получить данные по успеваемости учащегося, или журнал его действий на уроке, или действий при выполнении им заданий, что фиксируется в единой информационной электронной системе, например, такой как МЭШ (Московская электронная школа) или электронный дневник. Эти данные агрегируются и аккумулируются, их можно в дальнейшем обрабатывать, чтобы организовать систему

⁸ Инструментальная дидактика: перспективные средства, среды, технологии обучения // Назарова Т.С., Тихомирова К.М., Кудина И.Ю., Кожевников Д.Н., Аверьянов Ю.И., Емельянова М.Л., Бондаренко Е.А., Кожевникова В.В., Пименова В.Н., Франко Г.Ю., Мерзликина И.В., Полторак Д.И., Дударев М.И., Волкова С.А., Васканий А.Г., Куприянова Н.С., Заславская О.Ю., Смелова В.Г., Якушина Е.В., Иванов Ю.Н. и др. / Москва; Санкт-Петербург, 2012.

⁹ Romero, C., Ventura, S., Pechenizkiy, M., & Baker, de, R. S. J. (Eds.) (2011). Handbook of educational data mining. (Chapman and Hall/CRC data mining and knowledge discovery series). Boca Raton: CRC Press.

персонализированного обучения. В дополнение к этим данным нужны соответствующие вычислительные и статистические структуры и методы, чтобы понять смысл данных, а также специалисты, чтобы сформулировать правильные вопросы о данных, причём это будут специалисты смежных или междисциплинарных дисциплин, так как анализ самих данных не представляется простым и линейным.

Информатика и в целом компьютерные науки предоставляют экспертные знания в работе с большими объёмами данных, как с точки зрения машинного обучения, так и с точки зрения методов интеллектуального анализа данных. Всё это хорошо масштабируется в наборы данных с миллионами записей, в решение проблем реального мира, таких как фильтрация данных с гарантией, что систематические ошибки в исходных данных не приведут к ошибочным результатам.

Также необходимы при организации методики анализа образовательных данных специалисты в области статистики и психодиагностики. Они обладают опытом в понимании того, как правильно анализировать сложные исследовательские работы и проекты и должным образом учитывать тот факт, что большинство образовательных данных не являются традиционными рандомизированными управляемыми исследованиями, которые однозначно понятны и чётко спроектированы. Эти специалисты из двух сфер научного знания сильны в статистических и вычислительных методах, но их методов и данных недостаточно для развития в целом научной области анализа образовательных данных — требуются также специалисты с базовым пониманием организации процесса преподавания и построения обучения. В итоге специалисты в сфере образования и психологи являются ключевыми участниками сообщества исследователей в сфере анализа образовательных данных¹⁰.

¹⁰ Там же.

В целом сам процесс анализа образовательных данных преобразует необработанные данные, поступающие из образовательных систем, в полезную информацию, которая могла бы потенциально оказывать большое влияние на образовательные исследования и практику. Этот процесс мало чем отличается от других областей применения интеллектуального анализа данных, таких как бизнес, генетика, медицина и т.д., потому что он следует тем же шагам, что и основной процесс интеллектуального анализа данных: предварительная обработка, интеллектуальный анализ данных и постобработка. Однако важно заметить, что в данном случае термин «интеллектуальный анализ данных» используется в более широком смысле, чем оригинальное/традиционное определение Data Mining¹¹.

Методологически процесс, построенный на анализе образовательных данных, можно представить следующим образом. Вначале, на первом этапе, выполняется осмысление поставленной задачи и уточнение целей, которые должны быть достигнуты методами анализа образовательных данных. Второй этап состоит в приведении данных к форме, пригодной для применения конкретных методов анализа образовательных данных. Характер преобразований, совершаемых над данными, зависит от используемых методов, выбранных на предыдущем этапе. Третий этап — это применение методов анализа образовательных данных. Следующий этап — проверка построенных моделей. На большей части данных происходит обучение компьютерных систем, на меньшей части данных осуществляется проверка моделей. Последний этап — интерпретация полученных моделей человеком в целях их использования для принятия решений¹².

¹¹ C. Romero and S. Ventura, «Educational Data Mining: A Review of the State of the Art» in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 6, pp. 601–618, Nov. 2010.

¹² Веряев А.А., Татарникова Г.В. Educational data mining и learning analytics — направления развития образовательной квалитологии // Преподаватель XXI век. — 2016. — Т. 1. — № 2. — С. 150–160.

В рамках использования анализа образовательных данных применяются не только типичные методы интеллектуального анализа данных, таких как классификация, кластеризация, поиск ассоциативных правил, последовательная добыча, анализ текста и т.д., но и другие, например, регрессия, корреляция, визуализация и т.д. Учёные отмечают, что отдельного рассмотрения требует обзор используемых методов в компьютерных системах EDM/LA. Здесь только кратко перечислим их: классификация, кластеризация, регрессия, анализ текстов, ассоциативные правила, анализ социальных сетей, построение моделей учащихся, визуализация данных. Из всех методов отметим последние, связанные с визуализацией учебных достижений. Представляется, что подобного рода работа полезна школьным учителям и может положительно повлиять на мотивацию учащихся¹³.

Рассмотрим детально отличие анализа образовательных данных (АОД) от типичного интеллектуального анализа данных или просто анализа больших данных (Data Mining, DM).

Первое положение, которое чётко отличает методику, построенную на анализе образовательных данных, заключается в построении целеполагания. Основная цель анализа образовательных данных — улучшение или совершенствование образовательной деятельности, основываясь на тех данных, которые присутствуют в свободном доступе в системе образования. Фактически изучается сам процесс построения, организации и управления образованием на основе цифровых следов, оставляемых его участниками. Статистически и при помощи современных информационных технологий обобщаются и обрабатываются образовательные данные, вырабатываются закономерности и даются рекомендации как преподавателям, так и учащимся, учитывая их уровни усвоения знаний, сформированности умений,

¹³ Там же.

результатов обучения, а также непосредственный уровень успеваемости. На основе анализа образовательных данных формируется пул управленческих решений или генерируются рекомендации по улучшению образовательной деятельности.

Ещё одно принципиальное отличие — это сами данные, которые необходимо анализировать. Их особенность заключается в том, что они имеют сложную иерархичную структуру, которая в обязательном порядке состоит как минимум из 2–3 значимых уровней и взаимодействует с несколькими типами данных одновременно. Например, в онлайн-системах для EDM чаще всего используют лог-файлы, в которых содержится вся информация о работе пользователя на сайте (клики, переходы, оценки и многое другое). В то время как в онлайн-курсе всё уже в едином электронном формате, и масштаб аудитории на таких ресурсах, например Coursera, во много раз превышает стандартный класс. Именно для данных таких размерностей и имеет смысл применять распространённые в DM алгоритмы машинного обучения.

Следующее принципиально важное отличие анализа образовательных данных от других методов анализа — это использование помимо типичных методов интеллектуального анализа данных (кластеризации, классификации, регрессии, корреляции, визуализации, статистики, поиска правил связей (англ. association rule mining)) некоторых специфичных, например, из области психометрики. Так, методы из этой дисциплины помогают разбивать учащихся на группы по восприятию информации (типология MBTI), что, в свою очередь, позволяет адаптировать образовательную деятельность под учащегося: подбирать соответствующий тип контента и определённым образом его организовывать.

Чтобы понять, каким образом принимаются управленческие решения в системе образования с учётом полученных результатов анализа образовательных данных, приведём таблицу с описанием стратегий и направлений использования результатов этого анализа с учётом основных стейкхолдеров образовательного процесса (табл. 1).

На основе анализа стратегических направлений при использовании результатов анализа и алгоритмов образовательных данных можно выявить ключевые преимущества при проектировании и принятии управленческих решений в системе образования. Следует отметить, что это целый комплекс мер, связанных с организацией поддержки педагогов при управлении обучением класса и процессов самого обучения учащихся; с анализом и оценкой эффективности методики и стратегии обучения, а также всесторонней поддержкой и обеспечением обратной связи всех участников образовательной деятельности.

Заключение

На основе проведенного исследования можно сделать ключевые выводы о возможностях применения анализа образовательных данных для систем управления обучением и способов систематизации образовательных данных для принятия организационно-педагогических и управленческих решений в системе образования.

Стратегия и направления принятия управленческих решений с учётом методов анализа образовательных данных может быть следующей.

1. С учётом методов кластеризации образовательных ресурсов (по функциональному признаку, определяющему значение и место образовательного ресурса в образовательной деятельности, по организации текста ресурса, по характеру представляемой информации, по форме изложения, по целевому назначению и т.д.); тестовых заданий (по разным уровням сложности, с учётом индивидуального темпа работы, с учётом индивидуальных возможностей учащегося) возможна гибкая регулировка и управление количеством предъявляемых заданий в зависимости от уровня развития учащегося.

Таблица 1

Стратегии и направления использования результатов анализа образовательных данных¹⁴

Участники/пользователи	Стратегии использования результатов анализа данных
Учащиеся/студенты/ ученики	Персонализация электронного обучения; Рекомендации видов деятельности (мероприятий), ресурсов, учебных задач для учащихся, которые могли бы ещё больше улучшить их обучение; Предложение по изучению интересного практического опыта для студентов; Предложение по сокращению и оптимизации образовательной траектории в рамках всего обучения или упрощение навигации в виде ссылок по ресурсам, генерации адаптивных подсказок по ним, рекомендации актуальных курсов, включая электронных, актуальных обсуждений и форумов, книг и т.п.
Педагоги/учителя/ наставники/тьюторы	Получение объективной обратной связи об обучении; Анализ обучения учащихся и их поведения (прогресс и профиль); Определение учащихся, которым нужна поддержка; Прогнозирование успешность учащихся (успеваемость); Классификация учащихся по тем или иным группам (параметрическая кластеризация и группировка, фильтрация); Поиск регулярных, а также нерегулярных паттернов учащихся (наборов данных, необходимых в процессе обучения); Поиск наиболее часто совершаемых ошибок учащихся; Определение наиболее эффективных видов деятельности в процессе обучения; Улучшение (повышение эффективности) и адаптация (кастомизация), а также персонализации курсов обучения, включая электронные и т.п.
Разработчики курсов/ Исследователи в области образования	Оценка и сопровождение учебного материала курса; Улучшение качества обучения учащихся; Оценка структуры контента курса и его эффективности в процессе обучения; Автоматическое построение модели профиля учащегося и профиля преподавателя (тьютора); Сравнение инструментов и методов для анализа данных, чтобы иметь возможность рекомендовать наиболее полезный метод для каждой задачи; Разработка специфических инструментов анализа данных для обеспечения образовательных запросов и т.п.
Организации/ Провайдеры обучения/ Университеты/ Частные образовательные организации и компании	Совершенствование процессов принятия управленческих и организационных решений в высших учебных заведениях; Повышение и рационализация эффективности процесса принятия решений; Достижение конкретных целей обучения; Предложение определённых курсов, которые могут быть полезны для каждой категории учащихся; Поиск наиболее экономически эффективного способа повышения успеваемости и оценок; Отбор наиболее квалифицированных абитуриентов для получения диплома; Помощь в зачислении студентов, которые будут хорошо учиться в университете и т.д.
Администраторы/ Школьные администраторы и дирекция/ Сетевые администраторы/ Системные администраторы	Разработка оптимального способа организации институциональных ресурсов (человеческих и материальных) и образовательной потребности в них; Более эффективное использование имеющихся ресурсов; Повышение эффективности спроса образовательных программ и определение эффективности подхода к дистанционному обучению; Система оценки преподавателя и учебных программ; Разработка системы параметров повышения эффективности веб-сайта и адаптации его к пользователям (оптимальный размер сервера, распределение сетевого трафика и др.)

¹⁴ C. Romero and S. Ventura, «Educational Data Mining: A Review of the State of the Art» in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 6, pp. 601–618, Nov. 2010.

2. С учётом методов регрессии и прогнозирования можно выяснить, реально ли достичь желаемого результата, если изменить значения выбранного параметра. Управленческое решение в данном случае: прогноз результатов итоговых экзаменов, уровня компетенций выпускника, востребованности на рынке труда и уровня их заработной платы после трудоустройства. Также можно выявить степень влияния на процесс образования таких факторов, как потребность в специалистах, ресурсы вуза (в том числе и финансирование), степень внедрения информационно-телекоммуникационных технологий в образовательный процесс, уровень кадрового обеспечения вуза, заработная плата преподавателей и т.п.

3. Методы кластеризации позволят выявить группу учащихся со сходными психологическими, физиологическими, поведенческими и интеллектуальными характеристиками, что позволит определить, как эти поведенческие паттерны (шаблоны) влияют на успешность в различных видах деятельности, какие методы обучения эффективны по отношению к учащимся с различными стереотипами мышления и психики. Также можно на основе данных кластеризации разрабатывать учебные программы с учётом индивидуальных особенностей для отдельных групп учащихся, с учётом длительности обучения, траектории изложения материала, степени сложности заданий и другими характеристиками изучаемой дисциплины.

4. Анализируя данные, полученные через социальные сети, можно получить информацию о связях учащегося со своими сверстниками, о его интересах, перемещениях по сети, активности (частоты вхождения в сеть и времени

его нахождения в сети) и др. Объективная информация позволит составить профиль учащегося и учесть его реальные интересы в получении образования, настроить систему управления его обучения с учётом его как интеллектуальных, так и коммуникационных особенностей, а для преподавателей даст рекомендации по эффективному проектированию методики преподавания.

5. Анализ цифровых следов участников образовательной деятельности, например, на основе их деятельности в таких электронных информационных системах и платформах, как электронный дневник и МЭШ, позволит образовательной организации спроектировать и наладить персонализированное обучение в рамках построения модели цифровой школы конкретного региона.

Следует отметить, что использование методов анализа образовательных данных и полученных результатов позволит сформировать стратегию принятия организационно-педагогических и управленческих решений в образовании на основании данных и способов их систематизации (big data management in education). Особенностью такой стратегии является то, что она будет основана на математически и статистически измеримых количественных показателях, то есть самих обработанных образовательных данных, что позволит автоматизировать процесс управления обучением и сделать его персонализированным. **НО**

Analysis Methodology Big Data In Education

System-based methodological approach based on educational data analysis, strategy search making managerial and organizational-pedagogical decisions solutions in education

Kirill A. Barannikov, Vice-rector for development of Moscow city pedagogical University, candidate of pedagogical Sciences, Moscow, e-mail: barannikovka@mgpu.ru

Sergey M. Lesin, leading researcher of the educational research laboratory policies of the Moscow city Department of strategic development pedagogical University, candidate of pedagogical Sciences, Moscow, e-mail: lesinsm@mgpu.ru

Abstract. *The article deals with the implementation of a system-methodological approach based on the analysis of educational data (Educational Data Mining) and data mining methods used in the analysis and processing of information produced by educational organizations. We study the application of educational data analysis for learning management systems and ways to systematize educational data for making organizational, pedagogical and managerial decisions in education.*

Keywords: *educational data analysis, Educational Data Mining (EDM), data mining (DM), Moscow electronic school (MES).*

Список использованных источников:

1. *Веряев А.А., Татарникова Г.В.* Educational data mining и learning analytics — направления развития образовательной квалитологии // Преподаватель XXI век. — 2016. — Т. 1. — № 2. — С. 150–160.
2. *Baker, R. S. J. D., & Yacef, K.* (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.
3. *Bishop Ch.* Pattern Recognition and Machine Learning. Series: Information Science and Statistics. 2006. Т. XX. 740 p.
4. *Мамедова Г.А., Зейналова Л.А., Меликова Р.Т.* Технологии больших данных в электронном образовании. Открытое образование. 2017;(6):41–48. <https://doi.org/10.21686/1818-4243-2017-6-41-48>
5. *Заславская О.Ю., Кац С.В., Махотин Д.А.* Подходы к описанию модели проектирования сценариев уроков по технологии на портале «Московской электронной школы» / О.Ю. Заславская, С.В. Кац, Д.А. Махотин / Вестник Московского городского педагогического университета. Серия: Информатика и информатизация образования. 2019. — № 4 (50). — С. 64–72
6. Инструментальная дидактика: перспективные средства, среды, технологии обучения. // Назарова Т.С., Тихомирова К.М., Кудина И.Ю., Кожевников Д.Н., Аверьянов Ю.И., Емельянова М.Л., Бондаренко Е.А., Кожевникова В.В., Пименова В.Н., Франко Г.Ю., Мерзликина И.В., Полторак Д.И., Дударев М.И., Волкова С.А., Васканын А.Г., Куприянова Н.С., Заславская О.Ю., Смелова В.Г., Якушина Е.В., Иванов Ю.Н. и др. / Москва; Санкт-Петербург, 2012.
7. *Romero, C., Ventura, S., Pechenizkiy, M., & Baker, de, R. S. J.* (Eds.) (2011). Handbook of educational data mining. (Chapman and Hall/CRC data mining and knowledge discovery series). Boca Raton: CRC Press.
8. C. Romero and S. Ventura, «Educational Data Mining: A Review of the State of the Art,» in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601–618, Nov. 2010.

Spisok ispol'zovannyh istochnikov:

1. *Veryaev A.A., Tatarnikova G.V.* Educational data mining i learning analytics — napravleniya razvitiya obrazovatel'noj kvalitolgii // Prepodavatel' XXI vek. — 2016. — Т. 1. — № 2. — S. 150–160.
2. *Baker, R. S. J. D., & Yacef, K.* (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.
3. *Bishop Ch.* Pattern Recognition and Machine Learning. Series: Information Science and Statistics. 2006. Т. XX. 740 p.
4. *Mamedova G.A., Zejnalova L.A., Melikova R.T.* Tekhnologii bol'shih dannyh v elektronnom obrazovanii. Otkrytoe obrazovanie. 2017;(6):41–48. <https://doi.org/10.21686/1818-4243-2017-6-41-48>
5. *Zaslavskaya O.Yu., Kac S.V., Mahotin D.A.* Podhody k opisaniyu modeli proektirovaniya scenariyev urokov po tekhnologii na portale «Moskovskoj elektronnej shkoly» / O.Yu. Zaslavskaya, S.V. Kac, D.A. Mahotin / Vestnik Moskovskogo gorodskogo pedagogicheskogo universiteta. Seriya: Informatika i informatizaciya obrazovaniya. 2019. — № 4 (50). — S. 64–72.
6. Instrumental'naya didaktika: perspektivnye sredstva, sredy, tekhnologii obucheniya. // Nazarova T.S., Tihomirova K.M., Kudina I.Yu., Kozhevnikov D.N., Aver'yanov Yu.I., Emel'yanova M.L., Bondarenko E.A., Kozhevnikova V.V., Pimenova V.N., Franko G.Yu., Merzlikina I.V., Poltorak D.I., Dudarev M.I., Volkova S.A., Vaskanyan A.G., Kupriyanova N.S., Zaslavskaya O.Yu., Smelova V.G., Yakushina E.V., Ivanov YU.N. i dr. / Moskva; Sankt-Peterburg, 2012.
7. *Romero, C., Ventura, S., Pechenizkiy, M., & Baker, de, R. S. J.* (Eds.) (2011). Handbook of educational data mining. (Chapman and Hall/CRC data mining and knowledge discovery series). Boca Raton: CRC Press.
8. C. Romero and S. Ventura, «Educational Data Mining: A Review of the State of the Art,» in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601–618, Nov. 2010.