

МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В ОБРАЗОВАНИИ

© Э. Д. Алисултанова, Л. К. Хаджиева, З. А. Шудуева

ГГНТУ им. акад. М. Д. Миллионщикова, Грозный, Россия

В данной статье проводится обзор и обосновывается актуальность применения методов интеллектуального анализа данных в образовании. Изложены особенности и основные методы, применяемые для анализа данных в исследуемой области. Описанные методы наиболее актуальны и употребимы в системах поддержки принятия решений. На современном этапе развития информационного объема данных рынок труда требует новых инструментов и методов для поддержки больших хранилищ данных для оптимальной выборки и получения необходимой информации. Интеллектуальный анализ данных (Data mining) направлен на выявление и обработку информации из большого массива, требуемой для принятия решений в определенных сферах деятельности человека. На сегодняшний день области применения Data mining включают такие сферы, как бизнес, образование, сельское хозяйство, медицину и другие. В данном аспекте использование искусственного интеллекта, машинного обучения и методов визуализации данных имеют колоссальное значение для цифровой экономики РФ.

Ключевые слова: интеллектуальный анализ данных, поддержка принятия решений, искусственный интеллект, машинное обучение, нейронные сети, методы, классификация, обучающие системы.

В современном мире образование является самым важным компонентом экономического роста и развития страны. Его роль для человека и общества в целом только растет. Мы живем в таких условиях, что требуют высокого уровня профессионализма и интеллектуальных знаний для принятия правильных решений в различных ситуациях. Пожалуй, принятие решений в сфере образования является самым сложным процессом, в котором участвует большой круг лиц. Важным критерием является и анализ данных, поступающих от участников образовательного процесса [1].

Если рассматривать аспекты применения интеллектуального анализа данных непосредственно в образовании – это процесс преобразования необработанных данных, полученных из образовательных систем, в полезные данные, которые можно использовать для принятия решений. Процесс интеллектуального анализа данных состоит из девяти основных

этапов. Исходная позиция фиксируется этапом определения целей интеллектуального анализа данных и завершается этапом реализации обнаруженных знаний [2].

Определение предметной области – дает понимание того, что следует делать с различными решениями, такими как преобразование, алгоритмы, представление и т. п. Специалисты, отвечающие за этот этап, должны понимать и характеризовать цели конечного пользователя и среду, в которой будет происходить процесс обнаружения знаний.

Выбор и создание набора данных, на котором будет происходить обнаружение знаний. После определения целей следует определить данные, которые будут использоваться для процесса обнаружения знаний. Это включает в себя обнаружение доступных данных, получение важных данных и последующую интеграцию всех данных для обнаружения знаний в один набор, включающий качества, которые

будут учитываться в процессе. Этот процесс важен, потому что интеллектуальный анализ данных делает открытия из доступных данных.

Предварительная обработка. На этом этапе повышается надежность данных. Он включает в себя очистку данных, например, обработку отсутствующих количеств и удаление шума. В этом контексте он может включать сложные статистические методы или использовать алгоритм интеллектуального анализа данных. Например, когда кто-то подозревает, что определенный атрибут недостаточно надежен или имеет много отсутствующих данных, в этот момент этот атрибут может превратиться в цель контролируемого алгоритма интеллектуального анализа данных. Будет создана модель прогнозирования для этих атрибутов, после чего можно будет прогнозировать отсутствующие данные. Расширение, на которое обращают внимание на этом уровне, зависит от множества факторов. Тем не менее, изучение аспектов важно и регулярно раскрывается само по себе для корпоративных сред данных.

Преобразование данных. На этом этапе подготавливается и разрабатывается создание соответствующих данных для интеллектуального анализа данных. Методы здесь включают уменьшение размерности (например, выбор и извлечение признаков и выборку записей), а также преобразование атрибутов (например, дискретизацию числовых атрибутов и функциональное преобразование). Этот шаг может иметь важное значение для успеха всего процесса интеллектуального анализа данных, и обычно он очень специфичен для проекта. Например, в медицинских оценках соотношение атрибутов часто может быть наиболее значимым фактором, а не каждый из них сам по себе [3].

Прогнозирование и описание. На этом этапе нужно решить, какой тип интеллектуального анализа данных использовать, например, классификацию, регрессию, кластеризацию и т.д. Это в основном зависит от целей интеллектуального анализа данных, а также от предыдущих шагов. В интеллектуальном анализе данных есть две важные цели: первая – это прогноз, а вторая – описание. Прогнозирование обычно называют контролируемым интеллектуальным анализом данных, в то время

как описательный интеллектуальный анализ данных включает в себя аспекты неконтролируемого интеллектуального анализа данных и визуализации. Большинство методов интеллектуального анализа данных зависят от индуктивного обучения, когда модель строится явно или неявно путем обобщения достаточного количества подготовленных моделей.

Выбор алгоритма интеллектуального анализа данных. Этот этап включает в себя выбор конкретного метода, который будет использоваться для поиска паттернов, включающих несколько индукторов. Например, при рассмотрении точности и понятности предыдущий лучше с нейронными сетями, а последний лучше с деревьями решений. Для каждой системы метаобучения есть несколько возможностей того, как ее можно преуспеть. Мета-обучение направлено на выяснение того, что делает алгоритм интеллектуального анализа данных плодотворным или нет в конкретной проблеме. Таким образом, эта методология пытается понять ситуацию, в которой алгоритм интеллектуального анализа данных наиболее подходит. У каждого алгоритма есть параметры и стратегии обучения, такие как десятикратная перекрестная проверка или другое разделение для обучения и тестирования [4].

Использование алгоритма интеллектуального анализа данных. Наконец, реализован алгоритм. На этом этапе может потребоваться использовать алгоритм несколько раз, пока не будет получен удовлетворительный результат. Например, изменяя параметры управления алгоритмами, такие как минимальное количество экземпляров в одном листе дерева решений.

Оценка. На этом этапе оцениваются и интерпретируются найденные шаблоны, правила и надежность в соответствии с целью, описанной на первом этапе. Здесь рассматриваются этапы предварительной обработки с точки зрения их влияния на результаты алгоритма интеллектуального анализа данных. Например, включение функции на четвертом этапе и повтор. Этот шаг фокусируется на понятности и полезности индуцированной модели. На этом этапе выявленные знания также записываются для дальнейшего использования. Последним шагом является использование, а общая об-

ратная связь и результаты обнаружения получаются с помощью интеллектуального анализа данных.

Использование обнаруженных знаний. На этом этапе можно включить знания в другую систему для дальнейшей деятельности. Знание становится эффективным в том смысле, что можно вносить изменения в систему и измерять воздействие. Выполнение этого этапа определяет эффективность всего процесса интеллектуального анализа данных. На этом этапе есть множество проблем. Например, знание было получено из некоего статического изображения, обычно это набор данных, но теперь данные становятся динамическими. Структуры данных могут изменять определенные величины, которые становятся недоступными, и предметная область данных может быть изменена, например, атрибут, который может иметь непредвиденное ранее значение [5].

Хоть и развитие интеллектуального анализа данных в области образования было позднее, чем в других областях, оно оказалось весьма плодотворным. Прогресс этих методов изменил, практически, систему образования. Достаточно долгое время в информационных системах образовательных учреждений собиралась информация о всех студентах и их успеваемости, и преподавателях с их научно-образовательной работой. Создавались различные компьютерные и дистанционные курсы, системы мониторинга и оценки качества образования, образовательные форумы и многое другое. Следовательно, накопился большой объем данных, относящийся к определенным образовательным процессам. И, если не подвергать эти накопленные данные дальнейшему анализу, они остаются только огромными объемами данных [6].

Накопленные данные могут быть использованы для принятия решений по контролю образовательным процессом. Проведенные исследования показывают, что наибольшее воздействие на процессы принятия решений в образовательной сфере оказывают экспертные оценки преподавателей и их опыт. Из этого следует, что лица, принимающие решения, несут огромную ответственность за их последствия.

Понимание поведения обучающихся и того, как они учатся, может помочь управле-

нию образованием улучшить текущие учебные программы и образовательную практику в целом. Анализируя образовательные данные, а также анализируя важность влияния отдельных переменных, можно использовать различные модели интеллектуального анализа данных в качестве поддержки для принятия решений в сфере образования, способствуя тем самым более успешному обучению и повышению качества образования. Результаты интеллектуального анализа образовательных данных могут помочь университетам более эффективно распределять ресурсы.

Инструменты, методы и приемы интеллектуального анализа данных применяются для работы с огромными наборами данных для поиска скрытых закономерностей и взаимосвязей, что помогают многим учреждениям и организациям принимать решения на основе этих данных. Для обнаружения и поиска скрытых данных из базы данных применяются различные методы и алгоритмы, такие как:

- кластеризация и классификация;
- нейронные сети;
- искусственный интеллект.

Кластеризация относится к процессу идентификации и классификации объектов в разные группы, деление данных на кластеры, чтобы данные в каждом разделении имели некоторые общие характеристики аналогичных классов объектов. Кластеризация особенно полезна, когда наиболее распространенные категории в наборе данных заранее неизвестны. Если набор кластеров оптимален, то внутри категории каждая точка данных будет похожа на другие точки данных в этом кластере, чем данные точки в других кластерах.

Классификация относится к процессу описания связей между данными и описывает их для будущих наблюдений. Она сопоставляет данные из одного класса с данными из другого. Существуют различные методы классификации, такие как: методы на основе правил и дерева решений; рассуждения на основе памяти; нейронные сети. В классификации тестовые данные используются для оценки определенности правил классификации. Если определенность приемлема, правила могут быть применены к новым данным. Алгоритм обучения

классификатора использует предварительно классифицированные примеры для определения расположения параметров, необходимых для правильного различения.

Что же касается нейронных сетей, то они рассматриваются как системы, способные самостоятельно извлекать данные из поступающей информации. Их работа напоминает работу нашего мозга. Мы получаем знания в процессе обучения, которые, в свою очередь, хранятся не в одном сегменте, а распределены по всей сети [7].

Под искусственным интеллектом понимается система, способная выполнять те же функции и задачи, что и человек. Главные задачи искусственного интеллекта состоят в том, чтобы понять принципы интеллекта и сделать компьютеры полезными для человека. Для этого разработчикам необходимо узнать, как искусственный интеллект поможет разрешать трудные задачи. Области его применения, можно сказать, безграничны. В образовательных учреждениях он должен обеспечивать обучающихся необходимым пакетом информации, которые помогут для поиска решений задач и другой информации. Использование компьютеров намного упрощает работу. Ручные методы обучения давно исчерпали свои возможности, сейчас все строится на компьютерных системах.

Интеллектуальные системы являются своего рода системами обработки знаний, т. к. знания являются информацией. Исследования в области искусственного интеллекта выявили следующие направления:

Представление знаний – решение задач, связанных с представлением информации (знаний) в памяти искусственного интеллекта. Для этого разрабатывают модели представления знаний и языка их описания.

Манипулирование знаниями – обучение искусственного интеллекта для использования знаний при решении задач. Для этого необходима разработка способов пополнения знаний и методов достоверного вывода имеющихся. Оба направления тесно связаны между собой, их разделяют только условно.

Общение – предполагает формирование методов построения лингвистических процес-

сов. Вопросно-ответных и диалоговых систем, целью которых является обеспечение комфортного общения с человеком.

Восприятие – представление информации в виде зрительных образов; разработка методов перехода от зрительных сцен к их текстовому написанию и обратный переход.

Обучение – разработка методов формирования у систем искусственного интеллекта способности решать задачи, с которыми они ранее не сталкивались; создание методов разбивки общей задачи на более известные искусственному интеллекту.

Поведение – формирование поведенческой процедуры для адекватного взаимодействия искусственного интеллекта с окружающим миром и с другими системами [8].

Специалисты делают акцент на том, что совершенствование интеллектуальных систем во многом определяется успешным решением проблем представления знаний. Последовательное развитие структур данных привело к переходу от представления данных к представлению знаний. Уровень представления знаний отличается более сложной структурой и наличием специальных процедур пополнения знаний.

Существует много задач, в которых машинное обучение и извлечение знаний полезны для участников образовательных отношений. Исследования в данной области предполагают использование интеллектуальных обучающих систем (Intelligent Tutoring Systems) в образовании. Однако в России это направление не так широко развито. В большей части используют методы интеллектуального анализа данных. Отличительной особенностью таких методов является применение специальных методов для описания особенностей среды обучения.

Интеллектуальный анализ показывает, что большое внимание уделяется факторам, влияющим на успеваемость обучающихся. К примеру, информацию выкладывают на сайт (сервер), учащиеся должны изучить ее, а через 2-3 недели написать контрольную работу. Таким образом, они изучают требуемый объем данных отдельными блоками и в конце уже сдают экзамен. Конечно, такой формат обучения требует времени на подготовку.

Изучая все эти данные с применением различных интеллектуальных методов, можно выявить определенные закономерности, которые будут помогать учащимся изучать, анализировать и улучшать свою академическую успеваемость [9].

Интеллектуальный анализ образовательных данных помогает создавать методы разработки для извлечения интересной, интерпретируемой, полезной и новой информации, которая может привести к лучшему пониманию студентов и условий, в которых они учатся.

Стоит отметить и недостатки данного метода, а именно конфиденциальность и безопасность пользователей. Должно быть определено, как и с кем будет использоваться и передаваться информация. Инструменты и методы интеллектуального анализа данных работают с очень большими объемами данных, поэтому на этапе внедрения возникают большие затраты. Требуются квалифицированные ИТ-специалисты для предварительной обработки данных и поиска подходящей модели и метода анализа. Методы интеллектуального анализа данных не являются на 100% точными, что может привести к нежелательным последствиям. Конечно же, при правильной постановке задач и подходов к внедрению можно избежать этих проблем [10].

Образовательные методы интеллектуального анализа данных оказались плодотворными и привели к ряду явлений, касающихся обучения студентов на онлайн-платформе, а также к постоянному повышению точности. Есть важные аспекты, которые необходимо изучить, чтобы оправдать выдающийся прогресс в области образовательных данных, который приводит к признанию того, что не все ключевые данные хранятся в одном потоке данных. Исследования в области образования привели к нескольким новым образовательным усовершенствованиям. Компьютерные инновации изменили то, как мы живем и учимся. Сегодня использование информации, собранной с помощью этих достижений, поддерживает второй этап трансформации во всех областях обу-

чения с различными достижениями. Интеллектуальный анализ данных – это самый мощный метод с невероятным потенциалом, который помогает школам и университетам сосредоточиться на наиболее важной информации из собранных ими наборов данных о поведении обучающихся и потенциальных учениках и студентах [11, 12].

Новая развивающаяся популярная область, называемая интеллектуальным анализом данных, связана с созданием стратегий, которые находят полезные данные, полученные из образовательной среды. Интеллектуальный анализ данных – чрезвычайно обширная область, которая включает в себя использование различных методов и алгоритмов для поиска закономерностей.

Конечной целью интеллектуального анализа данных является улучшение образовательного процесса и объяснение образовательных стратегий для лучшего принятия решений.

Интеллектуальный анализ данных в образовании – молодая дисциплина с высоким потенциалом для каждого участника образовательного процесса. Методы интеллектуального анализа данных были разработаны для автоматического обнаружения скрытых знаний и распознавания закономерностей в данных.

Интеллектуальный анализ данных может использоваться для классификации и прогнозирования успеваемости обучающихся. Это может помочь преподавателям отслеживать академический прогресс, чтобы улучшить учебный процесс, это может помочь студентам в выборе курсов и управлении образованием, чтобы быть более эффективными и действенными. Анализ данных обучающихся имеет решающее значение для понимания того, какие методы обучения эффективны. В этой статье был проведен обзор методов и процесса интеллектуального анализа данных и их роли в образовании. А также были рассмотрены наиболее актуальные исследования, проведенные в области интеллектуального анализа данных, включая данные, использованные в некоторых исследованиях, и используемые методологии.

ЛИТЕРАТУРА

1. *Андрейчиков А. В.* Интеллектуальные информационные системы и методы искусственного интеллекта. М.: Инфра-М, 2021. С. 530.
2. *Богданов Е. П.* Интеллектуальный анализ данных. Волгоград: ФГБОУ ВО ВГАУ, 2019. 112 с.
3. *Нестеров С. А., Смолина Е. М.* Методы интеллектуального анализа данных в задачах оценки результатов дистанционного обучения. Текст научной статьи по специальности «Компьютерные и информационные науки». М.: Системный анализ в проектировании и управлении, 2019. С. 406-412.
4. *Taha S., Al-Janabi M.* Studying of education data mining techniques. Iraq: University of technology, 2018. Pp. 5742-5750.
5. *Минцаев М. Ш., Хакимов З. Л., Лабазанов М. А.* Программирование логических контроллеров фирмы SIEMENS // Лабораторный практикум. Грозный, 2021.
6. *Алисултанова Э. Д., Усамов И. Р., Мамуев А. М.* Роль виртуальной и дополненной реальности в учебном процессе // Актуальные вопросы современной науки: теория, технология, методология и практика: Материалы Международной научно-практической онлайн-конференции, приуроченной к 60-летию член-корреспондента Академии наук ЧР, доктора технических наук, профессора Сайд-Альви Юсуповича Муртазаева. Грозный, 2021.
7. *Алисултанова Э. Д., Моисеенко Н. А., Тасуев У. Р., Юсупова Р. В.* Современные наукоемкие технологии. 2021. № 12-1. С. 15-20.
8. *Алисултанова Э. Д., Хаджиева Л. К., Исаева М. З.* Профориентационная школьная лаборатория по основам кибербезопасности в сфере интернет вещей (ИОТ) // Вестник ГГНТУ. Гуманитарные и социально-экономические науки. 2019. Т. 15. № 4 (18). С. 51-58.
9. *Алисултанова Э. Д., Тасуева Х. Х.* Особенности и формы в образовании // Вестник ГГНТУ. Гуманитарные и социально-экономические науки. 2019. Т. 15. № 1 (15). С. 68.
10. *Моисеенко Н. А., Темирова А. Б.* Научное образование как основа формирования инновационной компетентности в условиях цифровой трансформации общества // Вестник ГГНТУ. Гуманитарные и социально-экономические науки. 2022. Т. 18. № 1 (27). С. 60-66.
11. *Моисеенко Н. А., Алисултанова Э. Д., Албакова А. А.* Использование информационных систем для оценки научной деятельности преподавателей вуза // Управление образованием: теория и практика. 2021. № 6 (46). С. 19-27.
12. *Хасухаджиев А. С. А., Магомаева Л. Р., Бикаев И. М.* Автоматизированная балльно-рейтинговая система оценки учебной деятельности студентов с модулем автоматизированного рабочего места преподавателя. Свидетельство о регистрации программы для ЭВМ RU 2017617180, 30.06.2017. Заявка № 2017614118 от 03.05.2017.

DATA MINING TECHNIQUES IN EDUCATION

© E. D. Alisultanova, L. K. Khadzhieva, Z. A. Shudueva

GSTOU named after acad. M. D. Millionshchikov, Grozny, Russia

This article reviews and substantiates the relevance of the application of data mining methods in education. The features and main methods used to analyze data in the study area are outlined. The described methods are the most relevant and usable in decision support systems. At the present stage of development of the information volume of data, the labor market requires new tools and methods to support large data warehouses for optimal sampling and obtaining the necessary information. Data mining is aimed at identifying and processing information from a large array required for decision making in certain areas of human activity. To date, the areas of application of Data mining include such areas as business, education, agriculture, medicine and others. In this aspect, the use of artificial intelligence, machine learning and data visualization methods are of great importance for the digital economy of the Russian Federation.

Keywords: data mining, decision support, artificial intelligence, machine learning, neural networks, methods, classification, learning systems.

REFERENCES

1. Andreichikov, A. V. (2021) *Intellectual'nye informatsionnye sistemy i metody iskusstvennogo intellekta*. [Intelligent information systems and methods of artificial intelligence]. Infra-M, Moscow, p. 530.
2. Bogdanov, E. P. (2019) *Intellectual'nyi analiz dannykh*. [Data mining]. FGBOU VO VGU, Volgograd, p. 112.
3. Nesterov, S. A. and Smolina, E. M. (2019) *Metody intellektual'nogo analiza dannykh v zadachakh otsenki rezul'tatov distantsionnogo obucheniya. Tekst nauchnoi stat'i po spetsial'nosti «Komp'yuternye i informatsionnye nauki»*. [Methods of data mining in evaluating the results of long-term results. Research paper on the specialty "Computer Science"]. *Sistemnyi analiz v proektirovanii i upravlenii*. Moscow, pp. 406-412.
4. Taha, S. and Al-Janabi, M. (2018) *Studying of education data mining techniques*. University of technology. Iraq, Pp. 5742-5750.
5. Mintsae, M. Sh., Khakimov, Z. L. and Labazanov, M. A. (2021) *Programmirovaniye logicheskikh kontrollerov firmy SIEMENS. Laboratornyi praktikum*. Grozny. [Programming the functions of SIEMENS controllers. Laboratory workshop]. Grozny.
6. Alisultanova, E. D., Usamov, I. R. and Mamuev, A. M. (2021) *'Rol' virtual'noi i dopolnennoi real'nosti v uchebnom protsesse'*. *Aktual'nye voprosy sovremennoi nauki: teoriya, tekhnologiya, metodologiya i praktika. Materialy Mezhdunarodnoi nauchno-prakticheskoi onlain-konferentsii, priurochennoi k 60-ti letiyu chlen-korrespondenta Akademii nauk ChR, doktora tekhnicheskikh nauk, professora Said-Al'vi Yusupovicha Murtazaeva*. [Topical issues of modern science: theory, technology, methodology and practice. Materials of the international scientific and practical online conference dedicated to the 60th anniversary of Corresponding Member of the Academy of Sciences of the Chechen Republic, Doctor of Technical Sciences, Professor Sayd-Alvi Yusupovich Murtazaev]. Grozny.
7. Alisultanova, E. D., Moiseenko, N. A., Tasuev, U. R. and Yusupova, R. V. (2021) *'Sovremennye naukoemkie tekhnologii'*. [Modern science-intensive technologies]. № 12-1. pp. 15-20.
8. Alisultanova, E. D., Khadzhieva, L. K. and Isaeva, M. Z. (2019) *'Proforientatsionnaya shkol'naya laboratoriya po osnovam kiberbezopasnosti v sfere internet veshchei (IOT)'*. *Vestnik GGNTU. Gumanitarnye i sotsial'no-ekonomicheskie nauki*. [Vocational guidance school laboratory on the basics of cybersecurity in the field of the Internet of Things (IOT). Herald of GSTOU. Humanitarian, social and economical sciences]. V. 15. № 4 (18). pp. 51-58.

9. Alisultanova, E. D. and Tasueva, Kh. Kh. (2019) 'Osobennosti i formy v obrazovanii'. *Vestnik GGNTU. Gumanitarnye i sotsial'no-ekonomicheskie nauki*. [Features and forms in education. Herald of GSTOU. Humanitarian, social and economical sciences]. V. 15. № 1 (15). p. 68.
10. Moiseenko, N. A. and Temirova, A. B. (2022) 'Nauchnoe obrazovanie kak osnova formirovaniya innovatsionnoi kompetentnosti v usloviyakh tsifrovoi transformatsii obshchestva' *Vestnik GGNTU. Gumanitarnye i sotsial'no-ekonomicheskie nauki*. [Scientific education as the basis for the formation of innovative competence in the context of the digital transformation of society. Herald of GSTOU. Humanitarian, social and economical sciences]. V. 18. № 1 (27). pp. 60-66.
11. Moiseenko, N. A., Alisultanova, E. D. and Albakova, A. A. (2021) 'Ispol'zovanie informatsionnykh sistem dlya otsenki nauchnoi deyatel'nosti prepodavatelei vuza'. *Upravlenie obrazovaniem: teoriya i praktika*. [The use of information systems to assess the scientific activities of university representatives. Education Management: Theory and Practice]. № 6 (46). pp. 19-27.
12. Khasukhadzhiev, A.S. A., Magomaeva, L. R., Bikaev, I.M. Avtomatizirovannaja ball'no-rejtingovaja sistema ocenki uchebnoj dejatel'nosti studentov s modulem avtomatizirovannogo rabochego mesta prepodavatelja. [Automated score-rating system for evaluating students' educational activities with a module of the teacher's automated workplace. Certificate of registration of the computer program RU 2017617180, 06/30/2017. Application № 2017614118 dated 05/03/2017]. Svidetel'stvo o registracii programmy dlja EVM RU 2017617180, 30.06.2017. Zajavka № 2017614118 ot 03.05.2017.