

# **ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ**

УДК 378+004.896

Л. Ю. Овсяницкая

## **ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ КАК СОСТАВЛЯЮЩАЯ ПЕДАГОГИЧЕСКОГО УПРАВЛЕНИЯ**

*Аннотация.* В работе рассмотрены возможности существующих средств интеллектуальной обработки данных, которые в настоящее время активно используются экономистами, маркетологами и бизнес-аналитиками и позволяют на основе анализа количественных и качественных экономических показателей решать вопросы синтеза и прогнозирования финансово-экономических процессов. Адаптация данных методов для обработки педагогической информации и широкое их внедрение в образовательный процесс могут резко повысить качество управления обучением и контроль над ним.

В статье обсуждаются теоретические и практические аспекты применения интеллектуального анализа данных в педагогике, обоснована актуальность данной темы. Проанализированы этапы возникновения и становления направления интеллектуального анализа данных. Продемонстрированы примеры оценивания качественных характеристик учения и личностных характеристик студентов с помощью средств Data Mining.

Автором предложены подходы к составлению шаблонов, лежащих в основе методов Data Mining, которые в совокупности с применением современных программных средств дают возможность производить обработку материалов педагогического мониторинга, выявлять факторы, оказывающие влияние на результат обучения, своевременно корректировать дальнейшую работу учебного заведения, т. е. осуществлять педагогическое управление.

*Ключевые слова:* интеллектуальный анализ данных, Data Mining, компетенция, компетентность, педагогическое управление.

*Abstract.* The paper investigates the potential of the intelligent data processing tools presently implemented by economists, marketers and business analytics for solving the economic problems of synthesis and prognostication of financial processes based on quantitative and qualitative indicators. The adaptation of the above methods for pedagogical data processing and implementation in educational process can improve the education management quality and control.

The paper presents the theoretical and practical aspects of the intelligent data analysis in pedagogy, the relevance of the problem being substantiated. The author looks at the origination and formation of the intelligent data analysis and gives the estimation examples of qualitative characteristics of teaching process, as well as the students' personal characteristics based on the Data mining resources.

The author recommends the ways for developing the patterns underlying the Data mining methods. The given approach in combination with the modern software tools can provide the data processing related to education management, i.e. pedagogical monitoring, identification of aspects affecting educational outcomes, early correction of educational department activities, etc.

*Keywords:* intelligent data analysis, Data mining, competence, expertise, pedagogical control.

## **1. Теоретические аспекты применения интеллектуального анализа данных для оценивания уровня компетенции и компетентности**

Существует множество различных определений понятий компетенции и компетентности. Несмотря на различные вариации, под компетенцией обычно понимается объем профессиональных знаний и навыков, а также личностные характеристики и установки, проявляемые в поведении человека и требуемые для выполнения непосредственных должностных обязанностей. Компетентность же предполагает наличие у человека определенных компетенций в совокупности с умением проявлять инициативу, способностью организовывать других людей, обладающих соответст-

вующими компетенциями, для выполнения поставленных целей, готовность оценивать, анализировать и предвидеть социальные последствия своих действий.

Иными словами, любая компетенция и компетентность включают в себя несколько характеризующих квалификационный уровень специалиста составляющих; некоторые из них легко поддаются оценке, тогда как другие определить с большой долей точности весьма проблематично. Так, говоря о компетентности в целом, мы, как правило, используем словосочетания «обладает высоким уровнем компетенции», «компетентен», «имеет низкую компетентность».

Д. И. Менделеев утверждал: «Наука начинается там, где начинаются измерения». Мы можем измерить количество правильных ответов в тесте и сумму теоретических знаний и практических навыков обучаемого, так как они представляют собой значения, поддающиеся исчислению. А вот способности, личностные качества, темперамент, инициативность, обязательность и многие другие качества человека, входящие в понятия компетенции и компетентности, не могут быть оценены подобными средствами, так как являются качественными характеристиками, в отношении которых могут использоваться лишь словесные описания. То есть применительно к указанным категориям необходимы иные инструментарий и методы измерения.

Например, если при проведении тестирования мы будем рассматривать каждый параметр, полученный отдельно, скажем – количество правильных ответов на вопросы теста по конкретной теме, то мы никогда не получим объективного представления о компетенции или компетентности ученика. Высокий результат, показанный при прохождении тестирования, не может гарантировать полноты полученных или имеющихся знаний, умений, навыков, личностных характеристик, необходимых для успешной работы. При «выпадении» хотя бы одного из перечисленных пунктов, несмотря на высокую оценку, человек не будет компетентен в своей области, и данный «провал» обязательно проявится в будущей практической деятельности.

В случае, когда мы имеем дело с качественными характеристиками, применение стандартных статистических методов оценивания усложняется еще и тем, что они построены на основе выполнения операций усреднения. Допустим, в группе два человека имеют высокую мотивацию, а три – низкую, тогда определить средний уровень мотивации группы становится невозможно.

Кроме того, существует еще одна проблема. Точность проведения измерений, безусловно, важна, но не менее значима процедура обработки результатов, итогом которой может быть как приведение полученных данных к информативному виду, удобному для восприятия основных статистических показателей, так и полная дискредитация полученных данных, вызванная выбором неверной модели и, как следствие, «выявленными» ложными закономерностями. Более того, материалы комплексного педагогического мониторинга, которые могут быть представлены текстовыми, числовыми, графическими, фото- и видеодокументами, всегда являются разнородными, а при высокой периодичности проведения контрольных мероприятий могут стать избыточными и, значит, неинформативными.

Перечисленные и многие другие вопросы оценивания уровня компетенции и компетентности учащихся пока остаются открытыми.

В настоящее время разработаны и активно используются экономистами, маркетологами и бизнес-аналитиками средства интеллектуального анализа данных, позволяющие на основе как количественных, так и качественных показателей решать задачи синтеза, анализа и прогнозирования финансово-экономических процессов.

На наш взгляд, для педагогической науки изучение опыта применения в сфере экономики средств интеллектуального анализа с целью их адаптации и широкого внедрения в процесс контроля и управления обучением весьма актуально и полезно.

Под интеллектуальным анализом данных – английский синоним Data Mining – понимается собирательное название целой совокупности методов. В основе подходов Data Mining лежат методы

классификации, кластеризации, моделирования и прогнозирования, построения деревьев решений, эволюционного программирования и нечеткой логики [3].

Data Mining – метафора, состоящая из двух понятий: данные (data) и добыча горной руды (mining), поэтому часто переводится как «добыча данных». В литературе можно встретить также варианты «извлечение информации», «раскопка данных», «интеллектуальный анализ данных», «извлечение знаний», «анализ шаблонов», «обнаружение знаний в базах данных». Можно сказать, что Data Mining – это мультидисциплинарная область, возникшая и развивающаяся на базе прикладной статистики, распознавания образов и искусственного интеллекта.

Способ анализа информации, рассматриваемый в данной работе, называется «интеллектуальным». Для понимания характерных особенностей и семантического поля этого понятия проанализируем этапы появления выражений «искусственный интеллект» и «интеллектуальный анализ данных» в русском языке.

Термин «искусственный интеллект» (англ. artificial intelligence) впервые был предложен в 1956 г. в США. Слово «intelligence» полисемично и имеет значения «интеллект; рассудок, разум; умственные способности» [1]. Применительно к информатике и способам обработки данных оно традиционно переводится на русский язык созвучно английскому произношению и согласно первому значению – «интеллект». Но под интеллектом мы понимаем высшую способность человека к мышлению и рациональному познанию [4]. Именно интеллект позволяет индивидууму мыслить, предвидеть, проводить сравнения, абстракции, выводить понятия, суждения, заключения, превращать восприятие в знание или критически пересматривать и анализировать уже имеющиеся знания.

На сегодняшний день под искусственным интеллектом, интеллектуальными средствами обработки информации понимают раздел информатики, включающий разработку методов моделирования и воспроизведения с помощью компьютера отдельных функций творческой деятельности человека, таких как решение

проблемы представления знаний и построение баз знаний, создание экспертных систем, распознавание зрительных и звуковых образов, речи, рукописных текстов, автоматический перевод с одного языка на другой. Эти задачи выходят за рамки ординарных вычислительных задач, но электронная техника не способна к высшему познанию, присущему исключительно человеку, обладающему интеллектом. Поэтому, с нашей точки зрения, наиболее точным переводом «intelligence» стало бы слово «разумность». Ведь, действительно, существующие сегодня интеллектуальные системы демонстрируют лишь элементы разумности. Именно с этих позиций мы и будем в дальнейшем рассматривать применение интеллектуального анализа педагогических данных.

Технологии Data Mining получили развитие в 90-е гг. XX в. В это время резко возрастает количество обрабатываемой информации, увеличиваются объемы устройств хранения данных и совершенствуются алгоритмы обработки информации. Григорий Пиатецкий-Шапиро (Gregory Piatetsky-Shapiro), один из основателей направления интеллектуального анализа данных, предложил следующее определение: «Data Mining – это процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности» [2].

В основу технологии Data Mining положена концепция шаблонов (patterns), которые представляют собой закономерности, выраженные в понятных человеку формах. В таблице буквально на нескольких возможных формулировках задач, касающихся трудоустройства выпускников, продемонстрировано отличие формирования шаблонов интеллектуального анализа данных для управления образованием от стандартных статистических подходов к нему.

Очевидно, что если будут существовать четкие шаблоны результатов поведения или личностных качеств студентов, адекватно отражающие динамику целевых показателей, то с большой до-

лей вероятности можно будет говорить о более качественной организации и возможности при необходимости оперативной корректировки деятельности педагогической системы в целом и отдельных образовательных учреждений в частности.

#### Сравнение содержания задач для получения данных на основе статистического и Data Mining методов

Статистический подход	Data Mining
Определить средний балл выпускников, которые смогли в течение двух месяцев после окончания вуза найти работу, удовлетворяющую их требованиям	Есть ли примеры (шаблоны) итоговых оценок выпускников, которые смогли в течение двух месяцев после окончания вуза найти работу, удовлетворяющую их требованиям
Принимали ли участие в творческих коллективах, студенческих научных обществах, студенческих конференциях выпускники, которые смогли в течение двух месяцев после окончания вуза найти работу, удовлетворяющую их требованиям	Имеются ли шаблоны поведения (участия в творческих коллективах, студенческих научных обществах, студенческих конференциях) выпускников, которые смогли в течение двух месяцев после окончания вуза найти работу, удовлетворяющую их требованиям
Определить средний балл, полученный во время прохождения производственных практик и стажировок студентами, которые были приглашены в дальнейшем на работу на то же предприятие, где практиковались	Встречаются ли шаблоны поведения студентов, которые проходили производственные практики и стажировки на определенном предприятии и впоследствии были приглашены туда на работу

Инструменты Data Mining позволяют выявить факторы, которые оказывают влияние на результат. Поэтому, имея большое количество информации, полученной в процессе мониторинга педагогических показателей, и заложив их в систему Data Mining, мы можем определить, от каких причин и движущих сил, мер, принятых действий в большей степени зависит, например, уровень компетенции выпускников вузов или их возможность в течение определенного времени после завершения обучения найти хорошую работу по специальности. Соответственно, можно определить и факторы, которые или не оказывают статистически значимого воздействия, или отрицательно влияют на результат. Применение интеллектуального анализа статистических данных в обра-

зовательных учреждениях [5, 6] в процессе внедрения сетевых образовательных систем и электронного обучения способствует качественным изменениям традиционной системы образования.

## **2. Практические аспекты вопроса применения интеллектуального анализа данных для оценивания уровня компетенции и компетентности**

Имея достаточное количество результатов педагогических измерений, включающих в себя как количественные, так и качественные величины, мы можем выявить шаблоны поведения, обучения, личностных качеств, которые приводят к достижению высших уровней компетенции и компетентности.

Сегодня уже разработано большое количество программных комплексов, на базе которых проводится аналитическая обработка полученных результатов. Из наиболее популярных отметим:

- Clementine – созданный фирмой IBM один из наиболее популярных продуктов, позволяющих проводить преобразования, моделирование и оценивание и внедрение методов Data Mining [6];
- DBMiner 2.0 Enterprise<sup>1</sup> – программное приложение для исследования больших объемов данных;
- Polyanalyst<sup>2</sup> – ресурс, который помимо полного анализа данных методами Data Mining включает востребованный инструмент анализа текстов Text Mining;
- IBM Intelligent Miner for Data<sup>3</sup> – инструмент, поддерживающий полный процесс Data Mining от подготовки данных до презентации результатов.

Самым доступным средством интеллектуального анализа является программа MS Excel (версии 2007 и выше) с установленной надстройкой «Интеллектуальный анализ данных», которая позволяет использовать преимущества функций прогнозирующего анализа SQL Server. Применение надстройки, содержащей специализированные инструменты, дает возможность на основе табличных или

---

<sup>1</sup> <http://www.spss.com/clementine>.

<sup>2</sup> <http://www.dbminer.com>.

<sup>3</sup> <http://www.megaputer.com>.



внешних источников информации создавать и тестировать модели интеллектуального анализа данных, а также управлять ими.

С помощью инструмента «Анализ ключевых факторов влияния» определяется связь интересующего нас параметра с другими показателями и зависимость его от них. Например, если в качестве результирующего фактора принять нахождение выпускниками работы в течение определенного времени после окончания вуза, т. е. подтверждение того, что компетенции выпускников удовлетворили работодателей, а в качестве влияющих факторов – оценки, полученные студентами на сессиях, государственном экзамене, при защите дипломного проекта; участие в студенческих научных конференциях, научных кружках, творческих коллективах; личные характеристики и др., то можно понять, какие факторы или их комбинации оказали решающее воздействие на результат.

Решение задачи кластеризации, или разделения всего множества вариантов на части, составляющие которых наиболее близки друг к другу по ряду признаков, позволяет обнаружить наличие студентов со сходными психологическими, физиологическими, поведенческими и интеллектуальными характеристиками. А формирование на основе этого и в соответствии с полученными рекомендациями групп для дополнительного обучения, факультативных занятий, тренингов и т. д., предоставляет широкие возможности для индивидуального подхода к обучению.

Используя блок «Прогноз», основанный на заранее выявленных ключевых факторах, оказывающих максимальное воздействие на результат, можно планировать выполнение поставленных целей и, в случае неблагоприятной версии прогноза, своевременно принимать меры воздействия на систему.

Инструмент «Анализ сценариев» позволяет моделировать влияние, которое оказывают изменения одного из параметров на другой, связанный с первым. Благодаря этому можно получить ответ на вопрос, реально ли достичь желаемого результата при модификации значения выбранного параметра. Например: действительно ли существует прямая зависимость между количеством сту-

денческих научных конференций, олимпиад и более высокой оценкой при защите дипломного проекта или обретением желаемой работы?

Поскольку интеллектуальный анализ данных изначально был ориентирован на финансовых и торговых аналитиков, то достаточно популярной операцией в нем является «Анализ покупательской корзины». По аналогии, как на основе данных о покупках выделяются товары, встречающиеся чаще всего в заказах, и формируются рекомендации относительно совместных продаж, с помощью этой операции можно выделять мероприятия, которые охотнее всего посещают студенты различных курсов и направлений, и сформировать рекомендации относительно привлечения учащихся к участию в наиболее значимых видах деятельности.

Инструмент «Деревья решений» сформирует правила классификации в виде иерархической (древовидной) структуры. Описанные с помощью дерева зависимости представляют собой визуализацию правил «если... – то...». Например: «Если студент посещает дополнительные углубленные занятия по предмету *X* и имеет отличные оценки по предмету *Y*, то с большой вероятностью будет приглашен на работу после окончания вуза работодателем *Z*».

### **Выводы**

Применение метода интеллектуального анализа Data Mining в совокупности с современными программными средствами позволяет проводить обработку данных педагогического мониторинга, выявлять факторы, влияющие на результат, учитывать полученную информацию и корректировать дальнейшую работу учебного заведения, т. е. осуществлять педагогическое управление.

### **Литература**

1. Англо-русский словарь общей лексики «Lingvo Universal» (к версии АВВУ Lingvo x3). 9-е изд., испр. и доп. М.: АВВУ. 2011.
2. Дюк В., Самойленко А. Data Mining: учебный курс. СПб.: Питер, 2001.

3. Кузнецова А. В., Сенько О. В. Возможности использования методов Data Mining при медико-лабораторных исследованиях для выявления закономерностей в массивах данных // *Врач и информационные технологии*. 2005. № 2. С. 38–46.

4. *Философский энциклопедический словарь*. М.: Советская энциклопедия, 1983.

5. Sachin R. B., Vijay M. S. A Survey and Future Vision of Data Mining in Educational Field // *Second International Conference on Advanced Computing & Communication Technologies*, Rohtak, Haryana, India, 2012. P. 96–100.

6. Fanqjun Wu. Discussion on experimental teaching of data warehouse & data mining course for undergraduate education // *The 7-th International Conference on Computer Science & Education*, Melbourne, Australia, 2012. P. 1425–1429.

### References

1. *LingvoUniversal (En-Ru)*. М.: FBBYY? 2011. (In Russian)

2. Djuk V., Samojlenko A. *Data Mining*. SPb: Piter, 2001. (In Russian)

3. Kuznecova A. V., Sen'ko O. V. The possibility of using Data Mining methods for medical and laboratory studies to identify patterns in the data. *Vrach i informacionnye tehnologii*. 2005. № 2. P. 38–46. (In Russian)

4. *Encyclopedic Dictionary of Philosophy*. М.: Sovetskaja jenciklopedija, 1983. (In Russian)

5. Sachin R. B., Vijay M. S. A Survey and Future Vision of Data Mining in Educational Field // *Second International Conference on Advanced Computing & Communication Technologies*, Rohtak, Haryana, India, 2012. P. 96–100. (Translated from English)

6. Fanqjun Wu. Discussion on experimental teaching of data warehouse & data mining course for undergraduate education // *The 7th International Conference on Computer Science & Education*, Melbourne, Australia, 2012. P. 1425–1429. (Translated from English)