

# 07, июль 2017

УДК 004.62; 004.04

## Применение интеллектуального анализа данных в сфере высшего образования

*Тихонов И. А., магистрант  
Россия, 105005, г. Москва, МГТУ им. Баумана,  
кафедра «Системы обработки информации и управления»*

*Научный руководитель: Виноградова М. В., к.т.н., доцент  
Россия, 105005, г. Москва, МГТУ им. Баумана,  
кафедра «Системы обработки информации и управления»  
[chernen@bmstu.ru](mailto:chernen@bmstu.ru)*

### Введение

В современном информационном обществе данные с каждым годом увеличивают свою ценность. Однако, если раньше потребность была в хранении данных, в основном это информация о транзакциях, то в последнее время появилась необходимость в анализе накопленных данных с целью поиска решений для задач в некоторой предметной области. В связи с этим были разработаны автоматизированные системы поддержки принятия решений (СППР), типовая структура такой системы приведена на рис. 1.



Рис. 1. Структура СППР

Подсистема ввода вывода реализуется на базе обычных систем управления базами данных (СУБД) и осуществляет транзакционную обработку данных [1], введенных оператором. После фильтрации оперативных источников данные попадают в хранилище.

В зависимости от потребностей аналитика к накопленным данным применяются определенные технологии анализа данных. Одной из технологий является Data Mining.

### **Data Mining**

Понятие интеллектуальный анализ данных (DM) было введено американским ученым в области информатики Григорием Пятецким-Шапиро в 1989 году. Такой анализ данных позволяет выявлять знания, удовлетворяющие следующим критериям [2]:

- знания должны быть новыми – целью любого анализа данных является получение ранее неизвестных знаний;
- знания должны быть нетривиальными – это неожиданные, неочевидные т.н. скрытые знания, которые не могут быть получены более простыми способами, например, путем использования средств визуализации;
- знания должны быть практически полезны – приносить определенную выгоду при их получении или при их применении;
- знания должны быть логически объяснимы – полученные результаты должны быть понятны человеку, в противном случае, может иметь место фактор случайности.

В Data Mining принято выделять четыре основных типа решаемых задач:

- классификация – определение класса объекта по его характеристикам;
- регрессия – определение дискретных значений некоторых параметров объекта по его известным характеристикам;
- ассоциация – нахождение частных зависимостей между объектами;
- кластеризация – поиск характеристик, присущих независимым группам (кластерам), на всем множестве анализируемых данных.

Также к задачам Data Mining можно отнести визуализацию результатов анализа.

Поиск ассоциативных правил и кластеризация относятся к типу описательных задач, направленных на улучшение понимания человеком анализируемых данных. К методам решения таких задач относят итеративные и иерархические методы кластерного анализа, самоорганизующиеся карты Кохонена, методы кросс-табличной визуализации, различные методы визуализации и др.

Классификация и регрессия относятся к предсказательному типу, используемому для прогнозирования появления некоторых событий на основе уже имеющихся данных. Для решения подобного рода задача применяются генетические алгоритмы, нейронные сети, деревья решений, линейная регрессия, а также методы ближайшего соседа, метод опорных векторов и др. [3]

Процесс получения новых знаний состоит из подготовительного этапа, из непосредственного применения средств интеллектуального анализа и анализа новых знаний. Более детально этапы приведены на рис. 2.

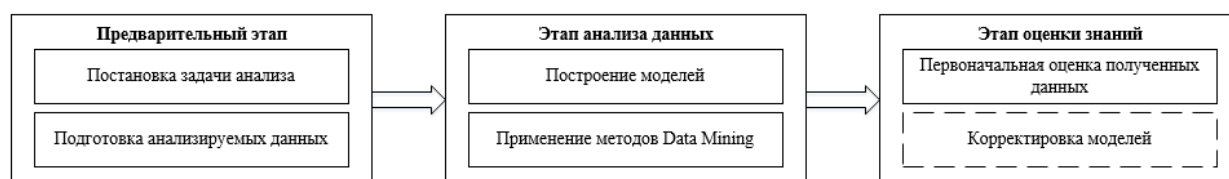


Рис. 2. Этапы получения нового знания

Стоит отметить, что окончательная оценка знаний не может быть получена практически сразу после применения того или иного метода Data Mining. Для этого необходимо через некоторое время проверить модель на адекватность и при необходимости провести ее корректировку.

Сферы применения Data Mining разнообразны и многогранны, их отличительной особенностью является большое количество накопленных данных и высокая значимость полученного результата. В настоящее время интеллектуальный анализ данных применяется во многих отраслях, например, в банковской, финансовой, страховой, телекоммуникационной, в маркетинге, медицине, геологии, сельском хозяйстве, промышленности, в облачных вычислениях, а также в сфере образования.

### **Educational Data Mining**

Основным назначением EDM является получение новых знаний в сфере образования, он активно применяется в обучающих онлайн-системах и в рамках классического процесса обучения. В первом случае EDM решает задачи обучающихся, преподавателей и разработчиков. Сбор данных упрощается использованием файлов-историй в системе курсов, хранящих информацию об активности пользователя, в том числе, включая обратную связь и анализ социальных сетей. На основе этих данных строятся модели студентов, подбираются и оптимизируются курсы, выдаются рекомендации. Однако, в России такое направление только начинает развиваться.

При классическом процессе «интенсивное использование средств вычислительной техники привело к появлению больших массивов данных, накапливаемых в ходе осуществления образовательного процесса - это записи в базах данных, электронные журналы, ведомости. Поэтому начинают осуществляться попытки извлечения из больших массивов данных закономерностей, позволяющих объяснить определенные результаты в

сфере образования, а также выдать рекомендации по организации педагогического процесса» [4]. Таким образом, целью EDM является улучшение и понимание образовательного процесса путем анализа и обработки данных, накопленных в рамках деятельности учебного заведения.

В дополнение к классическим методами Data Mining EDM задействует методы из области психологии преподавания, позволяющие выявлять факторы, влияющие на процесс обучения.

Высшие учебные заведения, используя методы EDM, могут осуществлять всесторонний анализ студенческих показателей, а также использовать оценку для прогнозирования вероятностей различных исходов. На рис. 3. приведены различные категории задач, решаемые в рамках EDM.

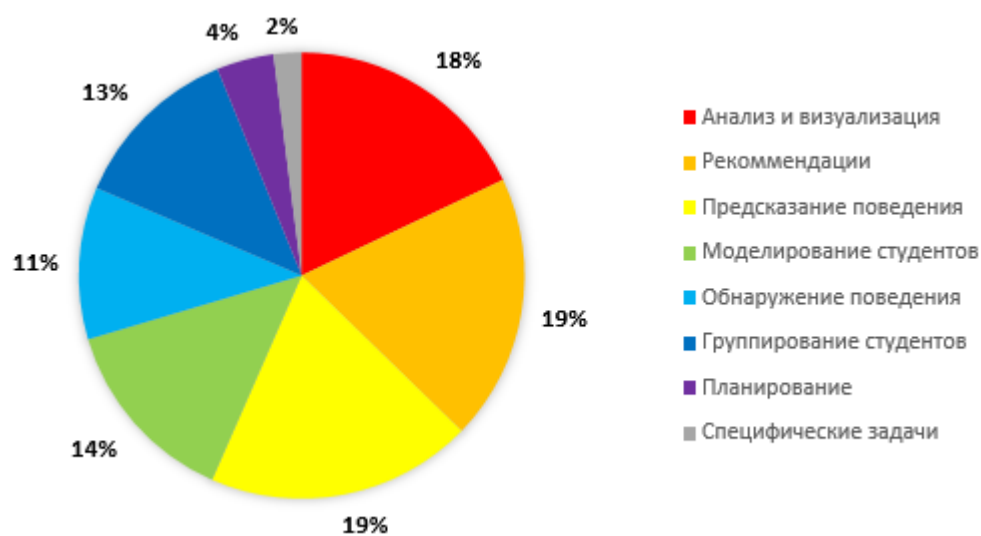


Рис. 3. Возможные спектр применения EDM

На основе этих категорий автором статьи были выделены конкретные виды задач ВУЗов, которые могут решаться с использованием интеллектуального анализа, и самое главное, имеют объективность и практическую ценность [5, 6, 7]:

1. Прогнозирование уровня успеваемости студента/группы студентов по курсу, определение и анализ факторов, влияющих на успеваемость.

2. Выявление и идентификация студентов, которые закончат или не закончат университет. Определение факторов, влияющих на это, и, как следствие, концентрация на таких категориях студентов до момента их отчисления.

3. Анализ и прогноз успеваемости студентов, с целью выдвижения рекомендаций и/или корректировки учебного плана для лучшего усваивания.

4. Изучение типологий студентов. Подход заключается в выявлении и изучении категорий студентов со схожими особенностями (например, быстрота усвояемости материала). Это может быть эффективно использовано при организации учебного процесса

5. Мониторинг данных о студентах с целью помощи им при дальнейшем трудоустройстве. Подход подразумевает классификацию студентов на основе успеваемости, знаний, навыков, предпочтений, профессиональных взглядов и т.п. под имеющиеся вакансии от компаний-партнеров ВУЗа.

Несмотря на очевидные преимущества такого анализ данных, ни один Российский ВУЗ в полной мере не использует EDM (*вывод сделан автором на основе анализа открытых источников*). Основной причиной является:

1. Отсутствие понимания возможностей, преимуществ и направлений применения EDM.

2. Незрелость ИТ-инфраструктуры вузов.

К незрелости ИТ-инфраструктуры можно отнести отсутствие в ВУЗах каких-либо систем и механизмов для осуществления многофакторного анализ, в том числе, хранилищ данных или иных систем, предназначенных для накопления многомерных данных, с целью их агрегации и дальнейшего анализа.

В тоже время, по мнению автора, в МГТУ им. Баумана, имеется базис для внедрения всесторонней системы интеллектуального анализа данных учебного процесса – это успешно внедренная информационная система «Электронный университет» [8].

#### **«Электронный университет» МГТУ им. Баумана - основа для EDM**

Электронный университет (ЭУ) – это автоматизированная система управления вузом разработанная специалистами ВУЗа, которая включает автоматизацию учебной, финансово-экономической, хозяйственной и организационной деятельности вуза [9]. Наибольший интерес в ЭУ для внедрения Educational Data Mining представляют: подсистемы формирования учебных планов и программ учебных дисциплин, подсистемы управления учебным процессом, подсистемы кафедрального уровня и подсистема «Аспирантура». Также важным является то, что система осуществляет всесторонний сбор и хранение информации, а также аналитическую обработку массивов данных. Например, в рамках подсистемы управления учебным процессом решаются задачи контроля и учета посещений занятий, учета и оценки текущей успеваемости, контроля и оценки качества

подготовки студентов, и сопровождение процесса обучения аспирантов в подсистеме «Аспирантура».

Таким образом, можно сделать вывод о развитости ИТ-инфраструктуры ВУЗа и о понимании ценности хранения и преимуществ обработки информации. Однако, в ходе анализа автором был сделан вывод, что с точки зрения анализа, ЭУ (в рамках образовательной деятельности) направлен либо лишь на визуализацию имеющихся данных, либо на выявление тривиальных зависимостей, в то время как выявление нетривиальных зависимостей позволит решать большое число задач, недоступных ранее, и тем самым расширить зоны действия системы.

### **Выводы**

По мнению автора, преимущества интеллектуального анализа данных в сфере образования, особенно высшего, не вызывают сомнений. EDM очень важен для учебных заведений, поскольку становится возможным решение ряда задач, связанных с улучшением организации процесса обучения и с комплексной оценкой студентов. В МГТУ им. Баумана, не смотря на наличие системы «Электронный университет», использование подобного рода анализа позволит решать еще больший спектр задач, направленных на «понимание» студента и улучшение образовательного процесса.

### **Список литературы**

- [1]. Григорьев Ю.А., Ревунков Г.И. Банки данных: учебник для вузов. М.: Изд-во МГТУ им. Н. Э. Баумана, 2002. 318 с.
- [2]. Барсегян А. А. Анализ данных и процессов: учеб. пособие. СПб.: БХВ-Петербург, 2009. 512 с.
- [3]. Методы и стадии Data Mining. Режим доступа <http://www.intuit.ru/studies/courses/6/6/lecture/162> (дата обращения 08.04.2017).
- [4]. Веряев А.А., Татарникова Г.В. Educational Data Mining и Learning Analytics – Направления развития образовательной квалитологии // Наука, образование, культура. 2016. № 2. С. 150-160.
- [5]. Cristobal Romero and Sebastian Ventura. Data mining in education // WIREs Data Mining Knowl Discov. 2013. № 3. Режим доступа: <https://pdfs.semanticscholar.org/c73b/0424e1a4ab2574cfce2e41c505f71f46940e.pdf> (дата обращения 08.04.2017).

- [6]. Naeimeh DELAVARI, Somnuk PHON-AMNUAISUK, Mohammad Reza BEIKZADEH. Data Mining Application in Higher Learning Institutions // Institute of Mathematics and Informatics, Vilnius. Informatics in Education. 2008. № 1. Available at: <https://halshs.archives-ouvertes.fr/hal-00588765/document>, accessed 08.04.2017.
- [7]. Петрова Т.В., Сабинин О.Ю. Исследование механизмов интеллектуального анализа данных (Data Mining) для решения задач извлечения знаний образовательного процесса // Молодежный научный форум: Технические и математические науки: электр. сб. ст. по материалам I студ. междунар. заочной науч.-практ. конф. 2013. № 1(1). Режим доступа: <https://nauchforum.ru/node/325> (дата обращения 08.04.2017).
- [8]. Балдин А. В. Информационные технологии в управлении университетом. МГТУ им. Н. Э. Баумана. М.: Изд-во МГТУ им. Н. Э. Баумана, 2007. 46 с.
- [9]. Агеева Т.И., Балдин А.В., Барышников В.А. Информационная управляющая система МГТУ им. Н.Э. Баумана «Электронный университет»: концепция и реализация / под ред. И.Б. Фёдорова, В.М. Черненко. М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. 376 с.