

Интеллектуальный анализ данных в задачах управления качеством образовательного процесса



Г.Ж. Солтан



С.С. Смаилова



И.М. Увалиева



А.К. Томилин

Восточно-Казахстанский государственный технический университет им. Д. Серикбаева
Г.Ж. Солтан, С.С. Смаилова, И.М. Увалиева
Национальный исследовательский Томский политехнический университет
А.К. Томилин

Предложена модель системы интеллектуального анализа данных (ИАД) применительно к образовательному процессу вуза. Проанализированы возможности ИАД с учетом специфики образовательной сферы, описан опыт его применения.

Ключевые слова: образовательный процесс, система менеджмента качества, анализ данных.

Key words: educational process, quality management system, data analysis.

Для эффективной работы системы менеджмента качества вуза необходимо создание информационной среды, которая позволяет управлять процессом сбора и анализа данных. Информационная среда вуза – это проекция бизнес-процессов вуза на область информационных технологий. Как указано в [1], система менеджмента качества любого вуза не может считаться полной и эффективной, если в ней не применяются инструменты, позволяющие осуществлять мониторинг образовательного процесса с использованием информационных технологий.

Таким образом, целью исследования является применение методов интеллектуального анализа данных

(ИАД) в задачах управления качеством образовательного процесса.

Одним из восьми принципов международного стандарта ИСО 9000 [2] является подход к принятию решения на основе фактов. Реализация данного принципа требует, прежде всего, измерения и сбора достоверных и точных данных, относящихся к поставленной задаче. Сбор и последующий анализ данных предполагает владение знаниями и применение специальных методов. Одним из таких методов является интеллектуальный анализ данных, позволяющий получать более интересные данные, нежели средние показатели [3].

Интеллектуальный анализ данных образовательного процесса позволяет решать следующие задачи:

- выявления кластеров студентов, являющихся группой риска по успеваемости;
- оперативный анализ больших объемов данных (например, результатов текущего и итогового контроля успеваемости) и обнаружение отклонений от нормального хода процесса;
- обнаружение проблемных ситуаций в учебном процессе и выявление их причин;
- анализ накопленных данных с целью совершенствования образовательного процесса.

Специфика анализа данных в сфере образования заключается в том, что большинство подобных задач подразумевают построение моделей, которые раскрывают закономерности в имеющихся данных. Поэтому наиболее востребованными являются описательные модели интеллектуального анализа, которые способствуют глубокому пониманию анализируемых данных. Ключевой момент в таких моделях – легкость и прозрачность результатов для восприятия человеком. Возможно, обнаруженные закономерности являются специфической чертой именно конкретных исследуемых данных и больше нигде не встретятся, но это все равно может быть полезно и потому должно быть известно. К такому виду задач относятся кластеризация и поиск ассоциативных правил [4].

В настоящее время в сфере образования задача прогнозирования имеет меньший вес, чем задача описания. Это связано с тем, что сама система образования изменяется достаточно быстрыми темпами и имеется значительное число косвенных факторов.

Для эффективной реализации интеллектуального анализа данных существует свободно распространяемые программные продукты в виде исходного кода реализации популяр-

ных алгоритмов, самостоятельные приложения (SAS Enterprise Miner, Poly Analyst, Deductor, RapidMiner), а также средства, интегрированные в СУБД (Oracle Data Mining, SQL Server Analysis Services).

Использование готовых алгоритмов, в виде исходного кода, очень трудоемко, а в случае применения самостоятельных приложений информационная система выполняет выборку и очистку данных в оперативном режиме, и экспорт данных осуществляет в формате, который приемлем для внешнего инструментального средства. Далее все манипуляции с данными, такие, как визуализация и применение метода анализа, происходят отдельно от информационной системы. На наш взгляд, наиболее подходящим способом является средство, интегрированное в СУБД. Преимуществами такого подхода являются:

- возможность объединения разнородных данных;
- использование встроенных средств подготовки данных к интеллектуальному анализу;
- выбор и применение различных алгоритмов ИАД;
- сохранение и повторное использование моделей;
- возможность сравнительного анализа эффективности различных моделей.

Для реализации задач управления качеством образовательного процесса разработана информационно-аналитическая подсистема образовательного портала ВКГТУ им. Д. Серикбаева, которая позволяет проводить анализ с привлечением средств Data Mining [6]. В данной подсистеме реализованы следующие процессы и функции:

- сбор статистических данных образовательного процесса;
- создание многомерной базы данных ИАД;
- консолидация данных, поступающих из различных источников в базу данных информационно-

аналитической подсистемы, его ведение и использование для проведения оперативного и интеллектуального анализа;

- реализация ИАД;
- подготовка и предоставление аналитику (или представителю руководства) результатов ИАД;
- администрирование прав доступа пользователей к информационным ресурсам информационно-аналитической подсистемы.

Для экспериментальной обработки данных в модуле «Оперативный и интеллектуальный анализ данных» сформированы 168 066 записей за 2009-2013 годы. Из них «Обучающий

набор» составил 100 839, «Тестовый набор» – 40 336, «Данные для прогноза» – 26 891 записей.

Для анализа использованы следующие инструменты ИАД:

- алгоритм дерева принятия решений;
- алгоритм нейронной сети;
- упрощенный алгоритм Байеса;
- алгоритм кластеризации;
- алгоритм логистической регрессии [5].

Модель ИАД определяет, какие комбинации значений входных переменных приводят к высокой успеваемости, а какие – к низкой, что позволяет выделить группы риска среди студентов.

38

Рис. 1. Окно с фрагментом дерева решений, для каждого узла можно просмотреть вероятность для варианта (сдал/не сдал)

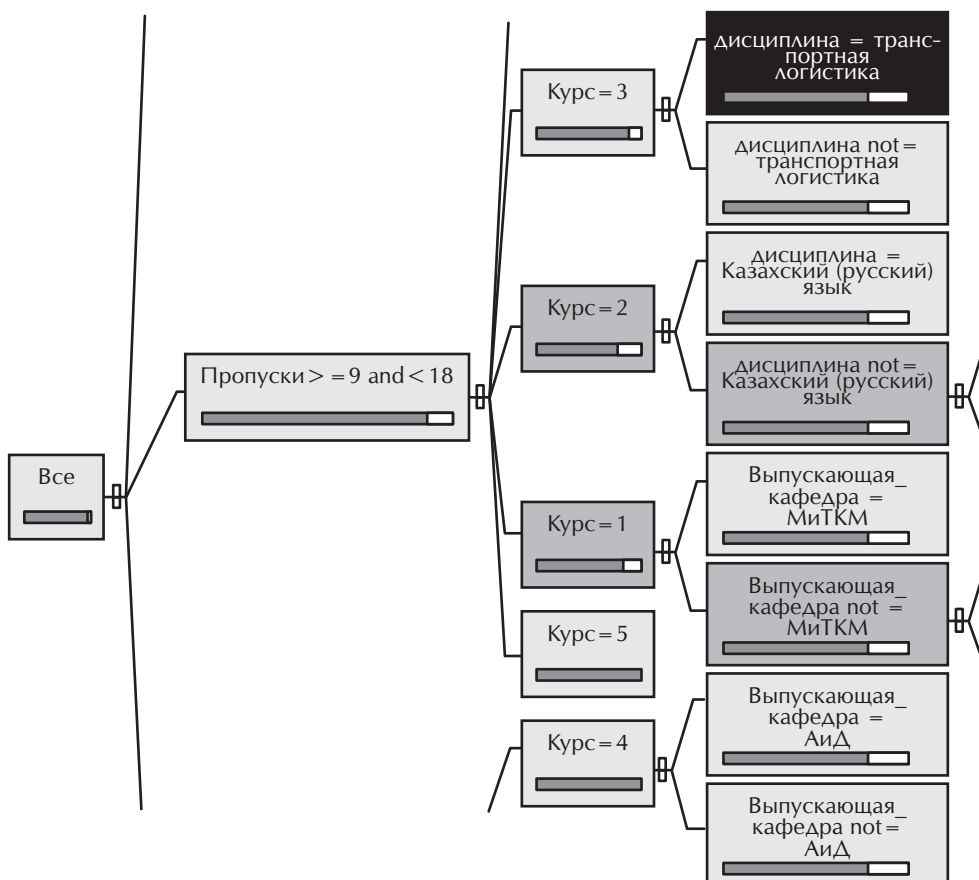


Рис. 2. Результаты алгоритма построения сети зависимостей для дерева решений

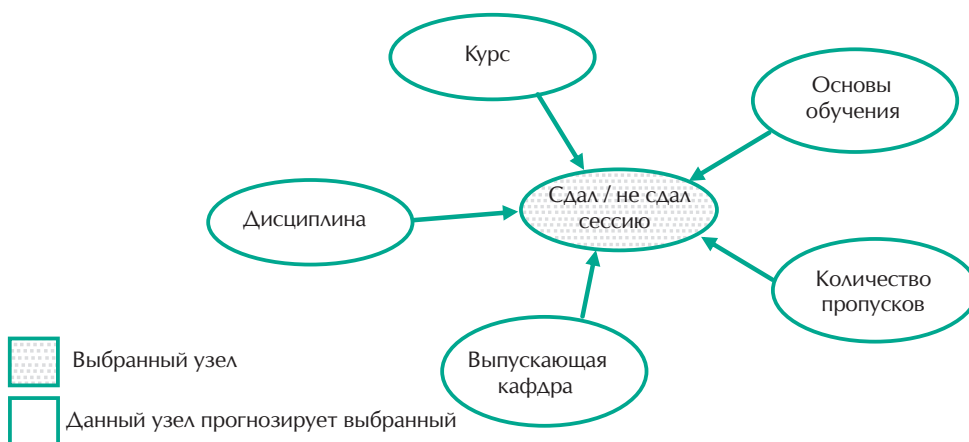


Рис. 3. Результаты классификации с помощью алгоритма нейронной сети

Атрибут	Значение	Не сдал
Дисциплина	История экономических учений	██████████
Специальность	50304	██████████
Дисциплина	Деньги, кредит, банки	██████████
Дисциплина	Статистика	██████████
Дисциплина	Транспортная логистика	██████████
Дисциплина	Компьютерная обработка экономической информации	██████████
Специальность	50724	██████████
Специальность	5В090100	██████████
Специальность	5В050700	██████████
Дисциплина	Основы электротехники	██████████
Выпускающая_кафедра	ЭП	██████████
Специальность	5В050900	██████████
Специальность	5В050600	██████████
Специальность	5В050800	██████████
Специальность	50806	██████████
Факультет	ФМТ	██████████
Специальность	50507	██████████
Специальность	50508	██████████
Дисциплина	Финансовые рынки и посредники	██████████
Дисциплина	Дополнительные главы математики	██████████

Далее представлены результаты моделирования данных образовательного процесса. На рис. 1 представлены результаты классификации на основе алгоритма дерева решений.

На рис. 2 представлено окно с изображением сети зависимостей для дерева решений, где выявлены факторы, влияющие на результаты сессии.

По результатам алгоритма построения сети зависимостей для дерева решений были выявлены следующие основные факторы, влияющие на результаты сессии: дисциплина, выпускающая кафедра, количество пропусков, основы обучения и курс. Самым сильным является фактор «Пропуски».

На рис. 3 представлены результаты ИАД на основе алгоритма нейронной сети.

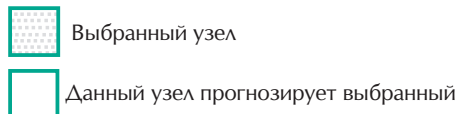
Результаты классификации с помощью алгоритма Байеса представлены на рис. 4. Как видно из отчета сравнения атрибутов на основе алгоритма Байеса, на «не сдачу сессии», в наибольшей степени, влияет такой фактор, как «количество часов пропуска».

Как видно из рис. 5, сеть зависимостей для алгоритма Байеса выявил больше факторов на «сдачу сессии», чем сеть зависимостей, построенная на основе алгоритма построения решений. По результатам построения сети зависимостей для алгоритма Байеса были выявлены следующие факторы, влияющие на результаты сессии: дис-

Рис. 4. Результаты классификации с помощью алгоритма Байеса

Атрибут	Значение	Сдал	Не сдал
Пропуски	0		
Курс	4		
Курс	1		
Курс	2		
Специальность	5B042000		
Основа_обучения	Договор		
Основа_обучения	Грант		
Специальность	5B090100		
Пропуски	20		
Пропуски	23		
Пропуски	27		
Пропуски	18		
Курс	5		
Пропуски	25		
Пропуски	14		
Пропуски	12		
Специальность	50729		
Пропуски	17		
Пропуски	10		
Пропуски	15		

Рис. 5. Результаты построения сети зависимостей для алгоритма Байеса



циплина, количество кредитов, количество пропусков, основы обучения, курс и язык обучения.

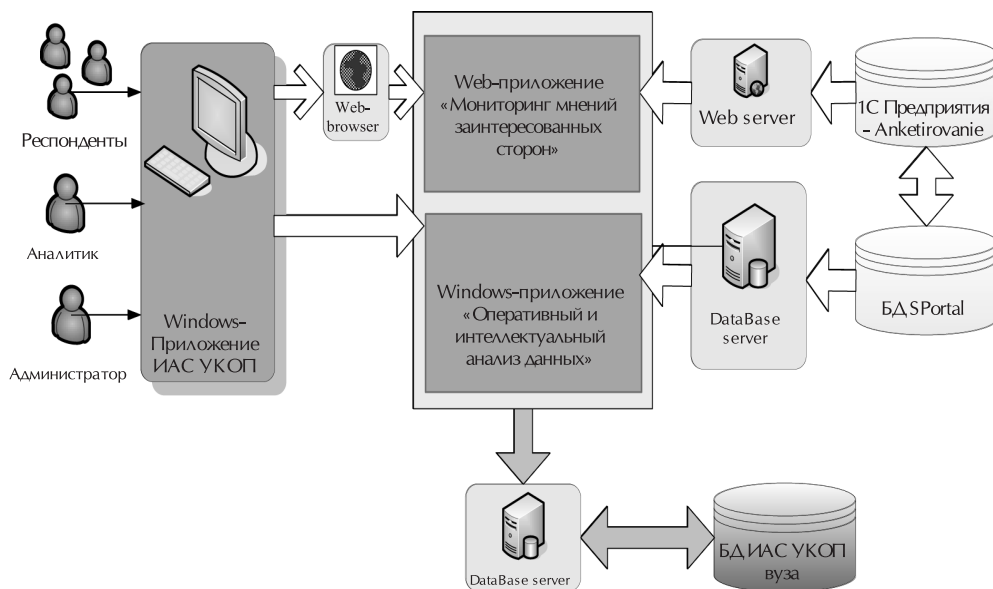
Таким образом, был произведен интеллектуальный анализ данных учебного процесса на основе реализации алгоритма дерева принятия решений, алгоритма нейронной сети и упрощенного алгоритма Байеса.

Результаты вышеуказанных видов анализа полезны не только для принятия управленческих и организационных решений руководящими лицами, но и для обучающихся. Они будут содержать информацию, по которой студент сможет объективно судить о своих способностях и выработать собственные стратегии обучения. Введение оперативного и интеллектуального анализа данных дает возможность замкнуть цикл управления образовательным процессом.

Архитектура информационно-аналитической подсистемы вуза представлена на рис. 6.

Таким образом, информационно-аналитическая система управления качеством образовательного процесса вуза должна не только обеспечивать административное управление вузом информацией о текущем и перспективном состоянии, но и выявлять проблемные зоны образовательного процесса вуза, вырабатывать корректирующие действия для усиления работы в конкретном направлении. Корректирующие действия, представленные в виде результатов оперативной и интеллектуальной обработки, позволяют судить о качестве учебного процесса и выдвигать гипотезы относительно способов управления его совершенствованием.

Рис. 6. Архитектура информационно-аналитической подсистемы



ЛИТЕРАТУРА

1. Управление качеством в высшем учебном заведении / Г.М. Мутанов, А.К. Томилин, Ю.Е. Кукина [и др.]. – Усть-Каменогорск, 2011. – 116 с.
2. СТ РК ИСО 9000-2007. Системы менеджмента качества. Основные положения и словарь. – Взамен СТ РК ИСО 9000-2001; Введ. 2008-01-07. – Астана, 2007. – 76 с.
3. Константиновский Д.Л. От сбора статистических данных – к информационному обеспечению принятия решений / Д.Л. Константиновский, М.Л. Агранович, О.Я. Дымарская. – 2-е изд., доп. и перераб. – М., 2006. – 160 с.
4. Научно-методические и технологические основы информационной системы управления качеством учебного процесса / под ред. Л.И. Григорьева. – М., 2008. – 132 с.
5. Методы и модели анализа данных: OLAP и Data Mining / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – СПб., 2004. – 336 с.
6. Томилин А.К. Внутренний контроль качества образования в вузе // Инж. образование. – 2012. – № 9. – С. 56–61.