

УДК 378:004.9
ББК 74.4

М.В. ПЕТРОВА, Д.А. АНУФРИЕВА

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ПРИ МОДЕЛИРОВАНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА В ВУЗЕ

Ключевые слова: образовательный процесс, вуз, моделирование, data mining, искусственные нейронные сети, ЕГЭ, успеваемость студентов.

Исследованы возможности методов интеллектуального анализа данных (Data Mining) при моделировании образовательного процесса в вузе. С помощью искусственных нейронных сетей получены вычислительные модели связи баллов, полученных студентами на ЕГЭ и успеваемостью студентов на первых курсах.

M.V. Petrova, D.A. Anufrieva
**A STUDY OF POSSIBILITIES OF DATA MINING FOR MODELING OF EDUCATIONAL
PROCESS IN HIGH SCHOOL**

Key words: educational process, high school, modeling, data mining, artificial neural networks, exam, student performance

Possibilities of Data Mining in the simulation of the educational process at the university are studied. Computational models of connections between scores obtained by students in the exam and academic performance of students in the first courses were obtained.

Введение

В декабре 2012 года Правительством РФ утвержден план мероприятий ("дорожная карта") "Изменения в отраслях социальной сферы, направленные на повышение эффективности образования и науки" (далее План).

По отношению к системе высшего образования План предусматривает изменения, направленные на повышение эффективности и качества услуг в сфере образования, совершенствование структуры и сети государственных образовательных организаций, совершенствование структуры образовательных программ и включает в себя, в частности, следующее.

- проведение ежегодного мониторинга эффективности образовательных организаций высшего образования;
- модернизацию системы лицензирования и аккредитации образовательных программ в системе высшего образования;
- введение прикладного бакалавриата в высшем образовании;
- обеспечение высокого качества программ магистратуры;
- создание новой модели аспирантуры на базе образовательных организаций высшего образования, активно участвующих в научно-исследовательской работе;

- поддержку программ развития сети национальных исследовательских университетов
- реализацию программ стратегического развития образовательных организаций высшего образования;
- создание системы оценки качества подготовки бакалавров;
- разработку и внедрение механизмов эффективного контракта с научно-педагогическими работниками образовательных организаций высшего образования;
- информационное и мониторинговое сопровождение введения эффективного контракта с вузом.

В данной работе была поставлена задача [1] исследования возможностей методов интеллектуального анализа данных – МИАД (Data Mining и, в частности, искусственных нейронных сетей) для анализа выполнения некоторых мероприятий Плана, в частности, анализа состояния системы образования в вузе и прогнозирования ее развития.

Поиск в электронной библиотеке России eLibrary.ru показал, что в России примеров применения МИАД при решении подобных задач нет. За рубежом такие примеры есть [6-8].

Методология исследований

Главным условием применения МИАД, реализующих «информационный подход» к задачам анализа и прогнозирования, является сбор данных об анализируемой системе. Источниками данных могут быть базы данных учетных систем вуза; данные, которые непосредственно или косвенно касаются участников образовательного процесса вуза, но которые отсутствуют в учетных системах и которые можно получить с помощью развернутых анкет-интервью участников образовательного процесса; внешние по отношению к вузу данные (макроэкономические показатели региона вуза, конкурентная среда, демографические и иные статистические данные).

Общие задачи, которые могут решать МИАД – следующие [2-5]:

- классификация – отнесение объектов (наблюдений, событий) к одному из заранее известных классов;
- регрессия, в том числе задачи прогнозирования - установление зависимости выходных параметров (целевых функций) от входных переменных (факторов);
- кластеризация – группировка объектов (наблюдений, событий) на основе данных (свойств), описывающих сущность этих объектов.

Проблемы анализа и моделирования образовательного процесса в вузе формулируются похожим образом, и решение большинства из них сводится к той или иной задаче Data Mining или к их комбинации.

Результаты анализа и моделирования образовательного процесса в вузе

На сайте <http://mfi.chuvsu.ru/opros/> размещены анкеты – интервью, заполняемые в режиме он-лайн и содержащие около 100 вопросов по процессу образования в целом и качеству образовательного процесса. К настоящему времени собрано около 300 анкет (около 30000 ответов).

Схема работы с анкетами следующая. Данные ответов на вопросы анкет представляются в виде таблицы, затем проводится их обработка, анализ и моделирование с помощью МИАД.

В таблице 1 приведены результаты оценки корреляции между результатами ЕГЭ и результатами первой сессии студентов.

Таблица 1 Корреляция между результатами ЕГЭ и результатами первой сессии студентов

Входные поля		Корреляция с выходными полями		
№	Поле	Кол-во пятерок 1 семестре	Кол-во четверок 1 семестре	Кол-во троек 1 семестре
1	Баллы ЕГЭ по математике	0,419	-0,322	-0,459
2	Баллы ЕГЭ по физике	0,247	-0,243	-0,410
3	Баллы ЕГЭ по русскому языку	0,327	0,279	-0,342

В левом столбце «Входные поля» указаны факторы (баллы ЕГЭ по трем дисциплинам), в столбце «Корреляция с выходными полями» в первой строчке - название целевых функций – количество оценок «Отлично», «Хорошо», «Удовлетворительно», ниже – показатели корреляции (цифры в виде десятичной дроби со знаком + или -).

Результаты показывают, что непосредственной линейной связи между баллами ЕГЭ и оценками первой сессии студентов нет и что связь эта существенно нелинейная.

Для выявления этой связи были использованы искусственные нейронные сети, с помощью которых были построены вычислительные модели, позволяющие по баллам ЕГЭ прогнозировать успеваемость студентов в первую сессию. Результаты представлены на рис. 1-3.

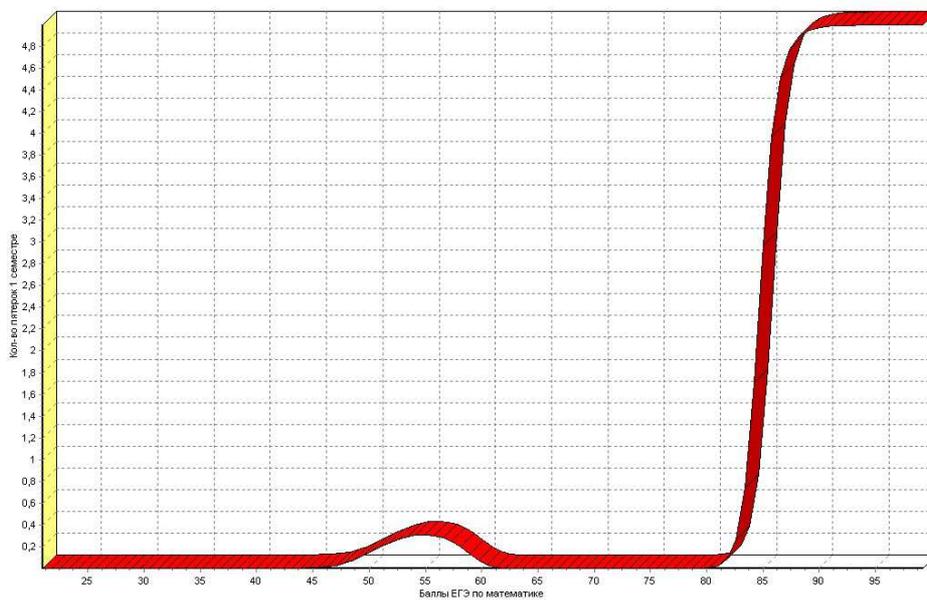


Рис. 1. Зависимость количества «пятерок» в первую сессию от результатов ЕГЭ по математике.

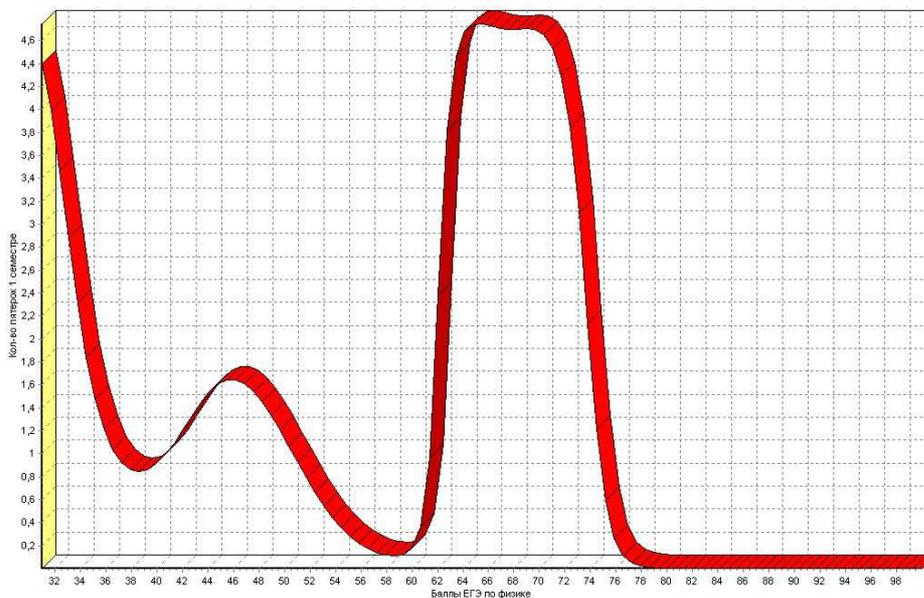


Рис. 2. Зависимость количества «пятерок» в первую сессию от результатов ЕГЭ по физике.

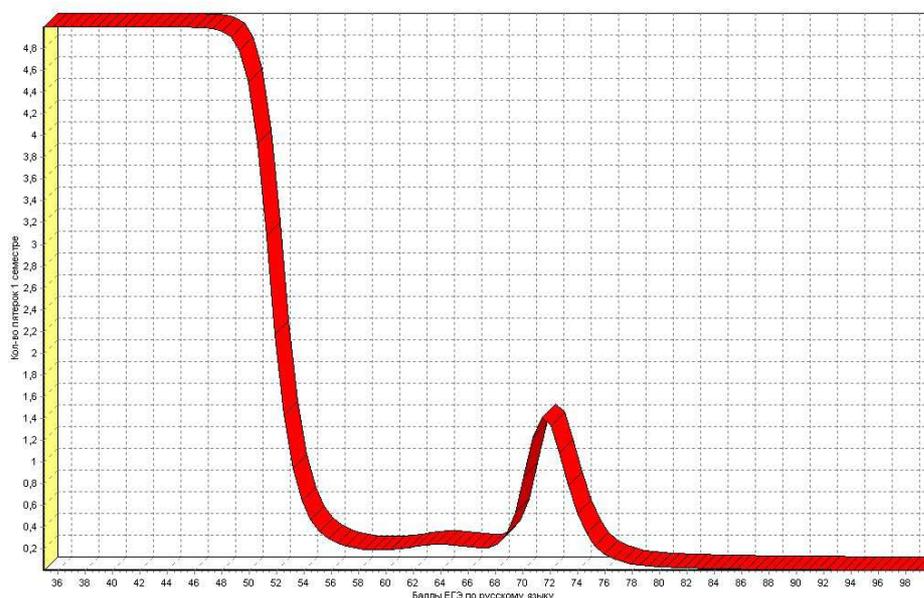


Рис. 3. Зависимость количества «пятерок» в первую сессию от результатов ЕГЭ по русскому языку.

Результаты показывают следующее. Количество «пятерок» в первую сессию не зависит (близко к 0) от баллов ЕГЭ по математике до значения 80 (рис. 1). При баллах ЕГЭ более 80 количество «пятерок», полученных студентами в первую сессию начинает резко расти, достигая максимума при баллах ЕГЭ более 90. Зависимость количества «пятерок» в первую сессию от результатов ЕГЭ по физике показывает, что максимальное количество «пятерок» соответствует диапазону 64 ... 70 (рис. 2). Обращает на себя внимание то, что велико количество «пятерок» для баллов ЕГЭ по физике менее 34. Последнее требует дополнительного изучения. При анализе зависимости количества «пятерок» от баллов ЕГЭ по русскому языку (рис. 3), можно отметить тенденцию то, что меньшие баллы ЕГЭ по русскому языку определяют большее количество «пятерок» в первую сессию, что, возможно, определяет отношение тех, кто больше внимания уделяет математике, к русскому языку.

В целом можно сделать вывод о том, что средние баллы ЕГЭ не могут служить основанием для определения «эффективности» вуза с точки зрения «качества образования» понимаемого как успеваемость студентов.

Были исследованы закономерности «траектории» качества учебной работы студентов в виде связей между оценками, полученными студентами на экзаменационных сессиях разных семестров (с 1-го по 4-й).

На рис.4 представлен скриншот аналитической платформы Deductor. График отражает зависимость количества «пятерок» во 2-ом семестре от количества «троек» в 1-ом семестре. Он показывает, что большое количество

Работа в настоящее время продолжается. В дополнение к данным анкет – интервью планируется собрать данные по финансовому обеспечению образовательного процесса, экспертным оценкам итоговых аттестаций, российскому рейтингу вуза, уровню конкурсного отбора абитуриентов и т.д.

Предполагается решение комплекса задач, например:

- прогнозирование «траектории» учебной и научной работы студентов различных специальностей и направлений подготовки и определение мер способствующих повышению качества учебной и научной работы студентов.
- кластеризация – группировка специальностей и направлений подготовки по качеству приема абитуриентов, качеству учебной и научной работы студентов, районам проживания абитуриентов и студентов с хорошим качеством учебной и научной работы, уровню востребованности выпускников различных специальностей и направлений подготовки.
- решение задач регрессии (получения многофакторных вычислительных моделей) устанавливающих зависимости таких целевых функций, как: качество учебы по курсам и качество выпускных квалификационных работ, востребованность выпускников и уровень их заработной платы после трудоустройства, степень удовлетворенности участников процесса образования (студенты, аспиранты, преподаватели, учебно-вспомогательный персонал, административно-хозяйственный аппарат, потребители – работодатели, родители студентов), и другие от таких факторов, как: потребность (федеральная, региональная) в специалистах; личностная потребность в высшем образовании, ресурсы вуза; бюджетное и внебюджетное финансирование; степень внедрения информационно-телекоммуникационных технологий; уровень требований при конкурсном отборе абитуриентов, степень участия преподавателей в НИР; уровень кадрового обеспечения образовательного процесса в целом; заработная плата преподавателей и учебно-вспомогательного персонала; и т.п.

Мы приглашаем студентов, аспирантов и преподавателей вузов Чувашии к совместной работе.

Литература

1. *Абруков В.С., Ефремов Л.Г., Коцеев И.Г.* Новые подходы к разработке моделей системы поддержки принятия решений и управления вузом // Вестник Чувашского университета. 2013. № 1. С. 224 - 229.
2. *Абруков В.С., Николаева Я.Г.* Количественные и качественные методы: соединяем и властвуем! // СОЦИС. Москва. 2010. N 1. С. 142-145.
3. *Абруков В.С., Николаева Я.Г., Макаров Д.Н., Сергеев А.А., Карлович Е.В.* Применение средств интеллектуального анализа данных (Data Mining) для исследования неполно определенных систем // Вестник Чувашского университета. № 2. 2008. С. 233-241.
4. *Анализ бизнес информации – основные принципы.* [Электронный ресурс] // [сайт]. [1995-2013].URL: <http://www.basegroup.ru/library/methodology/analysisbusinessdata> (дата обращения: 15.04.2013).
5. *Слаутская Е.В., Абруков В.С., Слаутский Л.А.* Интеллектуальный анализ данных психодиагностики школьников предпубертатного возраста // Вестник Чувашского университета. 2012. № 3. С. 226 - 231.

6. *Goyal M.* Applications of Data Mining in Higher Education // International journal of computer science. 2012. 9 (2). P. 113.

7. *Kovacic Z.J.* Predicting student success by mining enrolment data // Journal of research in higher education. 2012. 15. P. 54.

8. *Ranjan J.* Effective educational process: a data-mining approach // Journal of information and knowledge management systems. 2007. 37 (4). P. 502.

ПЕТРОВА МАРИНА ВЕНИАМИНОВНА – магистрант кафедры прикладной физики и нанотехнологий, Чувашский государственный университет, Россия, Чебоксары (vasilkovam@mail.ru).

PETROVA MARINA VENIAMINOVNA - master student of the department of applied physics and nanotechnology, Chuvash State University, Russia, Cheboksary (vasilkovam@mail.ru).

АНУФРИЕВА ДАРЬЯ АЛЕКСАНДРОВНА – магистрант кафедры прикладной физики и нанотехнологий, Чувашский государственный университет, Россия, Чебоксары (dashko-08@mail.ru).

ANUFRIEVA DARYA ALEXANDROVNA - master student of the department of applied physics and nanotechnology, Chuvash State University, Russia, Cheboksary (dashko-08@mail.ru).
