

КОМПЬЮТЕРНЫЙ АНАЛИЗ ДАННЫХ КАК ВАЖНАЯ СОСТАВЛЯЮЩАЯ ПОДГОТОВКИ СТУДЕНТОВ ФАКУЛЬТЕТА МЕЖДУНАРОДНЫХ ОТНОШЕНИЙ БГУ К ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ

Т. С. Петрушина, Т. И. Рабцевич, В. И. Садовничий

Белорусский государственный университет,
кафедра общей математики и информатики,
г. Минск, Беларусь
E-mail: rbcti@mail.ru

Рассмотрены состояние и основные направления развития современного компьютерного анализа данных, показаны их роль и место в обучающих технологиях для студентов ФМО БГУ. Приведены основные компьютерные инструментарии для решения этих задач: таблицы MS Excel, СУБД, модели интеллектуального анализа данных Data Mining и OLAP. Показан результат работы одного из алгоритмов кластеризации, используемых Data Mining

Интеллектуальный анализ данных, СУБД, запросы, кластерный анализ

В настоящее время современные вычислительные системы и компьютерные сети позволяют накапливать огромные массивы данных для решения задач обработки, анализа данных и распознавания образов. К сожалению, сама по себе машинная форма представления информации того или иного вида, которая необходима человеку, находится в скрытом виде и для её извлечения необходимо использовать специальные методы и программы анализа данных. Большой объём информации, с одной стороны, позволяет получить более точные расчёты и анализ, с другой — превращает задачи поиска решений и распознавания образов в сложную задачу. Неудивительно, что первичный анализ информации в современных условиях переложен на компьютер. В результате появился целый класс программных систем, таких как реляционные СУБД, электронные таблицы, системы обнаружения и интеллектуального анализа данных Data Mining и OLAP, программы, способные распознавать образы. Например, невозможно представить даже само существование сети Интернет без Data Mining, поскольку эта система программ используется каждодневно тысячи раз в секунду — каждый раз, когда кто-то использует Google или другие поисковые системы (search engines) на просторах Интернета. Все эти программные системы достаточно широко применяются и в современном образовательном процессе, каждая в большей или меньшей степени, в зависимости от специфики получаемой студентами специальности.

Поскольку сфера применения программ распознавания образов в образовательном процессе ещё достаточно ограничена, то авторы настоящей статьи более подробно остановятся на анализе данных в СУБД, который является одной из наиболее важных частей курса «Основы информатики и информационных технологий» для всех специальностей факультета международных отношений БГУ.

На современном этапе, по мнению авторов, по-прежнему актуален анализ данных с помощью электронных таблиц MS Excel. При преподавании MS Excel авторами ставится задача ознакомить студентов не только лишь с инструментальными возможностями этой программы, но и показать, как эти возможности можно использовать при решении конкретных задач. На примерах, соответствующих их специализации, студенты учатся решать поставленные задачи. Приведём несколько, с нашей точки зрения важных и интересных, тем заданий, предлагающихся для выполнения студентам: представление и анализ данных [1]; выработка стратегий и создание сценариев различных проектов; построение баз данных и работа с ними; решение различных задач планирования (капиталовложений и сбыта, рекламных компаний, управление персоналом); финансовый анализ; разработка диалоговых окон.

Другой важной составляющей компьютерного образования студентов ФМО БГУ являются проектирование и анализ данных в реляционных базах данных на примере MS Access. В этой части курса студенты занимаются построением сложных форм и отчётов, с внедрением в них динамических графических и мультимедийных объектов. Анализ данных осуществляется с применением перекрёстных запросов, запросов с вычисляемыми полями с помощью статистических функций, запросов с параметрами и других сложных запросов на универсальном для всех реляционных баз данных языке SQL.

Студентам-культурологам в курсе лекций и на семинарских занятиях для анализа произведений искусства предлагаются для рассмотрения диаграммы энтропии, метод семиотического дифференциала и факторный анализ. С применением элементов компьютерной лингвист-

тики [2, с. 7-9] производится анализ художественных текстов, сравнительный анализ языковых групп и отдельных языков, анализируется машинный перевод.

Возвращаясь к анализу данных в БД, отметим, что реляционные СУБД, вероятнее всего были, есть и будут наиболее подходящей технологией для хранения данных. Необходимость существует не в новой технологии БД, а скорее в средствах анализа, дополняющих функции существующих СУБД, и достаточно гибких, чтобы автоматизировать разные виды интеллектуального анализа, таких как Data Mining и OLAP.

На рис.1 с помощью диаграмм Эйлера-Венна представлен результат работы одного из неиерархических алгоритмов кластерного анализа, широко применяемых в вышеназванных технологиях. Данные диаграммы показывают принадлежность каждого из объектов анализа (точки на диаграмме) к одному из кластеров, определяющих две условных характеристики объекта – длину и ширину, где один из объектов принадлежит сразу двум кластерам.

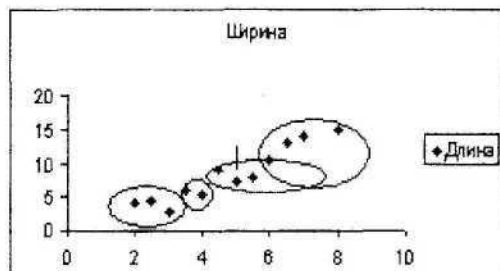


Рис. 1. Результат работы одного из алгоритмов кластерного анализа

В заключении отметим, что современный подход к работе с данными немаловажен без использования компьютерных технологий. Умение использовать эти технологии при решении конкретных задач открывает будущим специалистам новые возможности.

Литература

- [1] Эддоус М. Методы принятия решения / М Эддоус, Р. Стэнсфилд. – Москва Аудит, - 1997. – 590 с.
- [2] Еровенко В. А. Основы высшей математики для физиков: методические замечания и примеры: курс лекций / В. А. Еровенко. – Минск: БГУ, – 2006. 175 с.
- [3] Барсегян А. А. Методы и модели анализа данных: OLAP и Data Mining: учебное пособие/ А. А Барсегян, М.С. Куприянов, В. В. Степаненко, И. И. Холод. – СПб.: БХВ-Санкт-Петербург, - 2004. – 336 с.