

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПРОГРАММ ДЛЯ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

Н.Б. Джамансариев

(г. Юрга, Юргинский технологический институт (филиал)

Национального исследовательского Томского политехнического университета)

E-mail: kochetkovaev@mail.ru

COMPARATIVE ANALYSIS OF PROGRAMS FOR DATA MINING

N.B. Dzhamansariyev

(Yurga, Yurginsky institute of technology (branch)

of National Research Tomsk polytechnic university)

Abstract. The article presents a comparative analysis carried out by several criteria, programs for data mining. For more details and a description of features of the program Deductor Academic Studio.

Keywords: Data Mining, program, comparison analogues.

Введение. Повсеместное использование компьютеров привело к пониманию важности задач, связанных с анализом накопленной информации для извлечения новых знаний. Возникла потребность в создании хранилищ данных и систем поддержки принятия решений, основанных, в том числе, и на методах теории искусственного интеллекта. Действительно, управление предприятием, банком, различными сферами бизнеса, в том числе электронного, немыслимо без процессов накопления, анализа, выявления определенных закономерностей и зависимостей, прогнозирования тенденций и рисков. Именно давний интерес людей к методам, алгоритмическим моделям и средствам их реализации, используемые на этапе анализа данных, явился причиной подготовки моей работы. Анализ данных – это широкое понятие. Сегодня существуют десятки его определений. В самом общем смысле анализ данных – это исследования, связанные с обчетом многомерной системы данных, имеющей множество параметров. В процессе анализа данных исследователь производит совокупность действий с целью формирования определенных представлений о характере явления, описываемого этими данными. Как правило, для анализа данных необходимы различные довольно сложные математические методы, в связи с чем их применение на практике бывает затруднено.

Программы для интеллектуального анализа данных. Для решения указанной проблемы разработано и применяется на практике большое количество специализированных программ для анализа данных. Для сравнения были выбраны некоторые самые популярные из них. Сравнение было проведено по нескольким ключевым критериям, наиболее критичным с точки зрения пользователя. Оценка выставлялась в баллах от 1 до 3. Чем лучше, с точки зрения пользователя, значение критерия, тем выше балл. Результаты проведенного анализа приведены в табл. 1.

Таблица 1

Сравнительный анализ программ для интеллектуального анализа данных

Название программ	Наличие русско-язычной версии	Наличие демо-версии	Функционал демо-версии	Стоимость полной версии программы	Дружественный интерфейс	Системные требования	Наличие справки по работе с программой	Общий балл
OLAP ModelKit 3.0	3	2	2	1	1	3	1	13
STATISTICA Base	3	1	2	1	2	2	3	14
SPSS Statistics	3	2	2	1	2	2	3	15
Deductor Academic	3	3	1	3	3	1	3	17
IBM SPSS Modeler premium	3	1	3	2	1	1	2	13

По результатам проведенного анализа наилучшую оценку получила программа Deductor Academic Studio 5.3 – 17 баллов (из 21). Связано это в первую очередь с тем, что эта программа является бесплатной, имеет понятный и удобный интерфейс и, не смотря на низкую оценку общей функциональности (в сравнении с другими программами), Deductor Academic имеет достаточно обширный набор методов предназначенных именно для интеллектуальной обработки данных (Data Mining) (рис. 1–2).

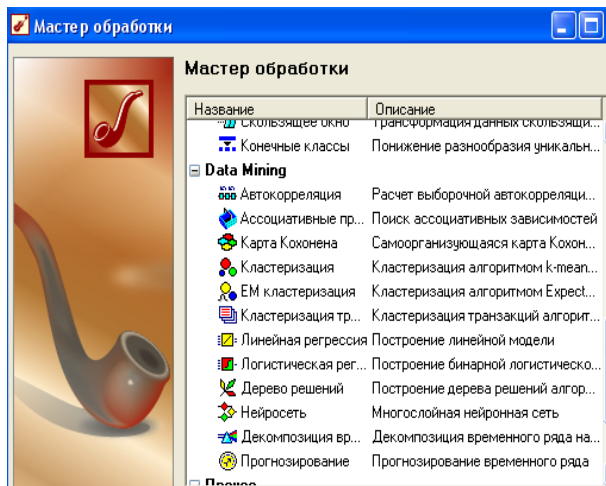


Рис. 1. Содержание функций в «Мастере обработки»

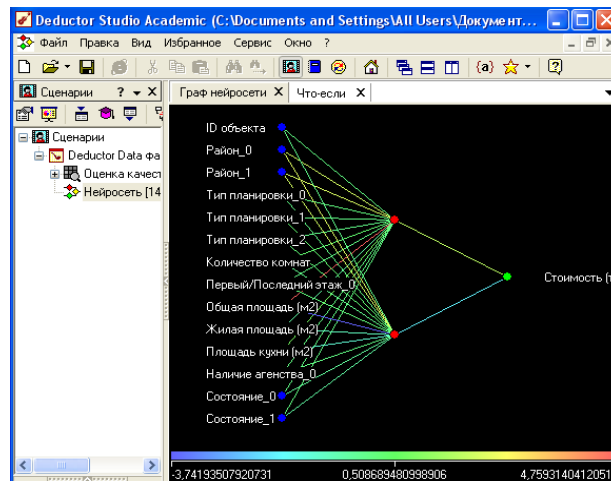


Рис. 2. Граф нейросети

Deductor Academic 5.3 – это платформа для анализа данных. Программа позволяет проводить кластеризацию, визуализировать и прогнозировать данные, применять различные подходы к их обработке и исследованию. Бесплатный вариант платформы включает в себя программу анализа Studio в комплекте с многомерной базой данных Warehouse и предназначается для использования в исключительно образовательных, некоммерческих целях. Программа поддерживает экспорт и импорт данных из текстовых файлов, позволяет строить различные модели (в т. ч. – на основе нейронных сетей, линейной регрессии и др.) и визуализировать полученные результаты в виде таблиц, диаграмм, графиков и т. д.

Deductor содержит все необходимое для решения задач Data Mining. Применение методов Data Mining – фактически единственная возможность извлечь пользу из накопленной информации, в противном случае собранные данные будут лежать «мертвым грузом». Data Mining позволяет извлекать из данных знания и превратить в конкурентные преимущества: качественно прогнозировать, точнее, выявлять целевые аудитории, предсказывать развитие событий, управлять рисками и прочее.

Заключение. В результате проведенного анализа было установлено, что по соотношению «цена-качество» наиболее предпочтительным для интеллектуального анализа данных является программа Deductor Academic Studio 5.3.

Список литературы

1. Сайт программы Deductor Academic 5.3 [Электронный ресурс]. URL: <http://www.basegroup.ru/deductor/> (Дата обращения 08.05.2015 г.).
2. Сайт программы OLAP ModelKit 3.0 [Электронный ресурс]. URL: http://freesoft.ru/olap_modelkit_30 (Дата обращения 08.05.2015 г.).
3. Сайт программы STATISTICA Base [Электронный ресурс]. URL: <http://www.statsoft.ru/> (Дата обращения 08.05.2015 г.).
4. Сайт программы SPSS Statistics [Электронный ресурс]. URL: <http://www-03.ibm.com/software/products/ru/spss-stats-standard> (Дата обращения 08.05.2015 г.).
5. Сайт программы [Электронный ресурс]. URL: <http://www-01.ibm.com/software/analytics/spss/products/modeler/> (Дата обращения 08.05.2015 г.).