2. ГОСТ Р 51318.22-99. Совместимость технических средств электромагнитная. Радиопомехи индустриальные от оборудования информационных технологий. Нормы и методы испытаний. – М.: Изд-во стандартов, 2001. – 36 с.

3. Thomas D. W. P., Christopoulos C., Pereira E. T. Calculation of Radiated Electromanetic Fields from Cables Using Time-Domain Simulation // IEEE Transactions on Electromagnetic Compability, 1994. – №3. – P. 201–205.

4. Гизатуллин З.М., Нуриев М.Г., Шкиндеров М.С., Назметдинов Ф.Р. Простая методика исследования электромагнитного излучения от электронных средств // Журнал радиоэлектроники. 2016. №9. С.7.

5. Гизатуллин З.М. Сквозное прогнозирование помехоустойчивости электронно-вычислительных средств внутри зданий при внешних электромагнитных воздействиях // Вестник Казанского государственного технического университета им. А.Н. Туполева . – 2011. – №2. – С. 123-128.

6. Гизатуллин З.М., Гизатуллин Р.М. Исследование электромагнитной совместимости локальных вычислительных сетей при наносекундных электромагнитных воздействиях // Радиотехника и электроника. – 2014. – №5. – С. 463–467.

7. Гизатуллин З.М. Повышение эффективности экранирования корпуса электронных средств // Технологии электромагнитной совместимости. – 2010. – №3. – С. 37-43.

8. Гизатуллин З.М. Снижение электромагнитных помех в межсоединениях многослойных печатных плат // Вестник Казанского государственного технического университета им. А.Н. Туполева. – 2012. – №2 – С. 199-205.

9. Гизатуллин З.М., Гизатуллин Р.М., Назметдинов Ф.Р., Набиев И.И. Повышение помехоустойчивости электронных средств при электромагнитных воздействиях по сети электропитания // Журнал радиоэлектроники: электронный журнал. – 2015. – №6.- С. 2.

А.Н. Назарова, Я.В. Соловьева

МЕТОДЫ ПОСТРОЕНИЯ ДЕРЕВЬЕВ РЕШЕНИЙ В ЗАДАЧАХ КЛАССИФИКАЦИИ МЕТОДАМИ DATA MINING

(Самарский национальный исследовательский университет
имени академика С.П. Королева)

Результатом развития информационных технологий является колоссальный объем данных, накопленных в электронном виде, растущий быстрыми темпами. Собранные за длительный срок данные могут содержать в себе закономерности, тенденции и взаимосвязи, являющиеся ценной информацией при планировании, прогнозировании, принятии решений, контроле за процессами. При этом данные, как правило, обладают разнородной структурой (тексты, изображения, реляционные базы данных и т.д.). В настоящее время анализ неоднородных данных является актуальной проблемой, так как человек физически не

способен эффективно анализировать такие объемы информации. Для решения данной проблемы все чаще используют технологию DataMining, которая призвана минимизировать усилия лица, принимающего решения в процессе анализа данных.

Технология DataMining позволяет выявить среди больших объемов данных закономерности, которые не могут быть обнаружены стандартными способами обработки сведений, но являются объективными и практически полезными. Методы DataMining основываются на базе различных научных дисциплин: статистки, теории баз данных, искусственного интеллекта, алгоритмизации, визуализации и других наук (рисунок 1) [1].

Рисунок 1 – Data Mining - мультидисциплинарная область

Задачи, решаемые методами DataMining [2]:

- классификация;
- регрессия;
- кластеризация;
- ассоциация;
- последовательные шаблоны;
- анализ отклонений.

Деревья решения являются одним из наиболее популярных методов к решению задач DataMining. Они создают иерархическую структуру классифицирующих правил типа "ЕСЛИ..., ТО..." (if-then), имеющую вид дерева. Для принятия решения, к какому классу отнести некоторый объект или ситуацию, требуется ответить на вопросы, стоящие в узлах этого дерева, начиная с его корня. Основа такой структуры - ответы "Да" или "Нет" на ряд вопросов.

На сегодняшний день существует большое количество алгоритмов, реализующих деревья решений: CART, C4.5, CHAID, CN2, NewId, ITrule и другие.

В данной работе рассматриваются два наиболее известных алгоритма построения деревьев решений CART и C4.5 для задачи классификации. Основным отличием данных алгоритмов является устойчивость к шумам и выбросам данных.

Алгоритм CART решает задачи классификации и является самым распространенным способом выявления, структурирования и графического представления логических закономерностей в данных. Его преимущества заключаются в следующем[3]:

- быстрый процесс обнаружения знаний;
- генерация правил в предметных областях, в которых трудно формализуются знания;
- извлечение правил на естественном языке;
- создание интуитивно понятной классификационной модели предметной области;
- прогноз с высокой точностью, сопоставимой с другими методами (статистическими и нейросетевыми);
- построение непараметрических моделей.

В данном алгоритме можно выделить три основные операции: сортировка источника данных при формировании множества условий для атрибутов числового типа, вычисление критерия *Gini* при разбиении узлов бинарного дерева, перемещение в таблице значительных объёмов информации при делении узла.

Отбор наилучшего варианта разбиения узла дерева проводится по наибольшей классифицирующей силе, вычисляемой по критерию *Gini* [3]:

$$GINI = \frac{1}{|L|} \cdot \sum_{i=1}^{Ncp} l_i^2 + \frac{1}{|R|} \cdot \sum_{i=1}^{Ncp} r_i^2$$

Алгоритм C4.5 строит дерево решений с неограниченным количеством ветвей у узла. Данный алгоритм может работать только с дискретным зависимым атрибутом и поэтому может решать только задачи классификации.

Его преимущества заключаются в следующем:

- C4.5 использует относительную энтропию при генерировании деревьев решений;
- C4.5 использует однопроходное отсечение ветвей, чтобы упростить переобучение. Отсечение ветвей существенно улучшает работу алгоритма;
- C4.5 может работать и с непрерывными, и с дискретными данными.

Для работы этого алгоритма были соблюдены следующие требования [4]:

- каждая запись набора данных должна быть ассоциирована с одним из предопределенных классов, т.е. один из атрибутов набора данных должен являться меткой класса;
- классы должны быть дискретными. Каждый пример должен однозначно относиться к одному из классов;
- количество классов должно быть значительно меньше количества записей в исследуемом наборе данных.

Основными различиями изученных методов являются следующие характеристики:

- вид расщепления - бинарное (binary), множественное (multi-way);
- критерии расщепления - энтропия, Gini, другие;
- обработка пропущенных значений;

- процедура сокращения ветвей или отсечения;
- возможности извлечения правил из деревьев.

Исследование выше описанных методов показало, что оба алгоритма имеют довольно высокую скорость работы, а выходные данные просты для восприятия. В ходе работы оба алгоритма были обучены, поскольку для построения дерева классификаций необходим размеченный набор данных. Алгоритм C4.5 по сравнению с CART более прост в обучении благодаря однопроходному отсечению ветвей. Однако оба алгоритма имеют свои недостатки, которые возможно устранить, используя предварительную обработку данных и объединив их с другими алгоритмами классификации, что будет рассмотрено в ходе дальнейшего изучения задач DataMining.

Литература

1 Datamining [Электронный ресурс]. – http://rtb-media.ru/wiki-data-mining/ (дата обращения 15.01.2017 г.);

2 DataMining – добыча данных [Электронный ресурс]. – https://basegroup.ru/community/articles/data-mining (дата обращения 15.01.2017 г.);

3 Деревья решений – CART математический аппарат. [Электронный ресурс]. – https://basegroup.ru/community/articles/math-cart-part1 (дата обращения 15.01.2017 г.);

4 Реализация и распараллеливание алгоритма интеллектуального анализа данных, основанного на деревьях решений. [Электронный ресурс]. – http://intellect-tver.ru/?p=209 (дата обращения 15.01.2017 г.);

А.Л. Никишина, Я.В. Соловьева

ИНТЕРАКТИВНОЕ ПРИЛОЖЕНИЕ ДЛЯ БИЗНЕС-АНАЛИЗА И УПРАВЛЕНИЯ ПРОЕКТАМИ «ANALYSIS AND PROJECT MANAGEMENT»

(Самарский национальный исследовательский университет
имени академика С.П. Королева)

Эффективность и оптимизация являются одними из главных слов в современной реализации бизнес-процессов. Но как сделать процесс более эффективным и оптимизированным и какие метрики можно использовать для оценки данных характеристик? Эти вопросы наиболее часто возникают у бизнес-аналитиков и руководителей проектов в начале карьеры. Если рассматривать ситуацию в городе Самара, то в настоящий момент существует единственная программа, по которой готовят бакалавров по направлению бизнес-аналитики, открывшаяся сравнительно недавно – в 2015 году [1]. Чуть лучше ситуация с современным подходом к управлению бизнес проектов: таких специалистов, бакалавров или магистров готовят в нескольких высших учебных заведениях