

ум – хорошо, а искусственный ум – лучше!

*интеллектуальные методы анализа данных
как инновационный элемент
в деятельности современных фирм*

Переход российской экономики к инновационной модели развития является одной из наиболее важных макроэкономических тем, широко обсуждаемых в научной, специальной и публицистической литературе. Спектр проблем, порождаемых таким переходом и связанных с ним, также весьма обширен.

В частности, актуальным становится привлечение развитых аналитических, программных и вычислительных методов для решения широкого круга экономических задач, возникающих в текущей деятельности предприятий. Именно вопросы применения таких методов и будут предметом рассмотрения в настоящей публикации.

Деятельность любого производственного предприятия, любой коммерческой фирмы сопровождается регистрацией значительных объемов первичной информации. Значимым фактором для развития становится обработка этих данных и получение новых знаний о возможностях и потенциале предприятия.

В современной науке, в разных прикладных областях весьма часто звучит упоминание об интеллектуальных методах анализа данных или о так называемой «добыче данных» (Data Mining). Стандартное определение этого понятия таково:

Data Mining – это процесс обнаружения в исходных данных ранее неизвестных, нетри-

Меркулина И.А.

*канд. экон. наук,
доцент,
декан факультета
дополнительного
образования,
профессор кафедры
прикладной
информатики в
экономике*

Никитин А.П.

*канд. тех. наук,
доцент,
профессор кафедры
прикладной
информатики
в экономике*

*Всероссийская
государственная
налоговая академия
Минфина России*

**цель анализа
данных состоит
в поиске
неочевидных,
неожиданных
закономерностей,
составляющих
так называемые
скрытые знания**

визуальных, практически полезных и доступных для интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности [1].

Принципиальный момент в приведенном определении – нетривиальность разыскиваемых знаний. Акцентируется внимание на том, что цель такого анализа состоит в поиске неочевидных, неожиданных закономерностей и закономерностей, составляющих так называемые скрытые знания. «Неочевидных» означает, что такие закономерности имеют мало шансов быть обнаруженными стандартными статистическими методами обработки и обобщения информации, а также экспертами, если эти закономерности не укладываются в их представления и не вытекают напрямую из их опыта. К методам интеллектуального анализа данных относят различные технологии и алгоритмы кластеризации, классификации, поиска ассоциаций и других логических закономерностей в данных. Мы опирались на обзор состояния этой области на начало 2008 г. [2].

Применительно к экономической сфере практическая полезность обнаруженных знаний означает, что они имеют конкретное бизнес-значение, т.е. могут быть использованы для повышения прибыльности бизнеса, для снижения издержек, для завоевания новых рыночных ниш, для предоставления клиентам новых видов услуг, типов продукции и др. Интеграция таких методов с другими бизнес-приложениями представляет немалую ценность как дополнительный фактор, позволяющий получить ощутимые преимущества в конкурентной борьбе.

Вместе с тем публикаций, посвященных экономическим приложениям интеллектуального анализа данных и знакомящих с ними специалистов-практиков, не так много. Также следует отметить, что для ряда изданий характерно увлечение формально-математической стороной, перегруженность формулами, определениями и понятиями, затрудняющими восприятие этих технологий.

С позиции бизнес-приложений выделяется целый ряд технологических проблем, связан-

ных с применением Data Mining технологий [3]. Среди них особо подчеркнем необходимость специальной подготовки исходной информации (очистки данных от недостоверных записей) перед применением средств и процедур ее обработки.

Обсуждаются в научной литературе и другие вопросы, связанные с обработкой персональной информации, оптимизацией затрат на внедрение программных средств интеллектуального анализа, интерпретации полученных результатов в терминах предметной области, восприятием новых технологий экспертами, менеджерами, практиками [4].

В предыдущих публикациях обсуждался вопрос о месте таких информационных технологий в структуре учебных планов подготовки специалистов в области экономической информатики [5].

Вместе с тем, иногда такие программные технологии рассматриваются как универсальные для фактически любых ситуаций, требующих анализа больших массивов информации (в том числе, экономической). Делается неявный вывод о том, что сами по себе новые информационные технологии, новые технологии обработки данных обеспечивают качественный прорыв в эффективности решения пользовательских задач. Такое обобщение следует признать некорректным.

Для иллюстрации авторы хотели бы остановиться на практических приложениях, в которых действительно элементы технологий интеллектуального анализа данных позволили обеспечить существенное продвижение к поставленной цели.

1. Весьма популярной в настоящее время является концепция CRM-систем (*customer relationship management*), т.е. систем анализа и управления взаимоотношениями с клиентами. О степени востребованности такого подхода можно с уверенностью судить и по спектру предлагаемых программных продуктов, и по многочисленным интернет-публикациям, и по интересу к CRM со стороны топ-менеджмента компаний. Стандартным методическим под-

вывод о том, что сами по себе новые технологии обработки данных обеспечивают качественный прорыв в эффективности решения пользовательских задач является некорректным

ходом является RFM-модель (сокращение от Recency, Frequency, Monetary), которая оперирует показателями, отражающими время, прошедшее от последней покупки клиента, частоту покупок и суммарную стоимость покупок [6].

Сегментация клиентской базы также может производиться по прибыльности, стереотипным действиям, демографическим, социальным и иным характеристикам. Дополнительно рассматривается структура покупок клиента по товарным группам, ценовым группам, оценивается приверженность клиента конкретным магазинам или конкретным маркам товаров.

Выявленные профили клиентов, заинтересованных в товарах или услугах, которые предлагает фирма, дают шанс разработать эффективные маркетинговые программы. Декларируемым элементом CRM-систем является способность предупреждать о потере лояльности клиентом и его возможном уходе.

При анализе клиентских баз весьма различного масштаба (розничная торговля – сотни тысяч клиентов, крупная медицинская компания – десятки тысяч пациентов, инновационные компании, работающие в наукоемкой сфере – сотни клиентов – юридических и физических лиц) удастся, тем не менее, опереться на общность методологии их решения. В эту методологию входят процедуры сегментации, поиска ассоциативных правил в массивах данных, использование схем перебора вариантов для установления характеристик клиентов. Так при обработке данных продуктивна трехзвенная структура производных показателей (товар → покупка → сводная статистика по клиенту, и, аналогично лечебная услуга → визит → сводная статистика по пациенту).

2. Вторая прикладная задача связана с прогнозированием продаж в торгово-розничной компании, структурно представляющей собой комплекс специализированных сетей, филиалов и отдельных магазинов. В целом, кратко- и долгосрочное прогнозирование будущих продаж является одним из наиболее востребованных результатов анализа экономической информации. Для целей прогнозирования может при-

**декларируемым
элементом
CRM-систем
является
способность
предупреждать
о потере
лояльности
клиентом
и его возможном
уходе**

меняться обширный список статистических методов: классический регрессионный анализ, методики скользящего среднего, экспоненциального сглаживания, различные варианты ARIMA и др. [7]. Интеллектуальные программные средства позволяют сделать эти предсказания более точными за счет учета ряда дополнительных факторов. Например, при необходимости прогнозирования не только валовой выручки, но и выручки по отдельным структурным подразделениям предварительная кластеризация магазинов на группы улучшает качество предсказаний (фиксировалось снижение относительной ошибки прогноза примерно на треть). Такая кластеризация основывалась на комплексе показателей, включающих как ретроспективную информацию о деятельности в предыдущие интервалы времени (темпы роста, профиль сезонности), количественные показатели (арендная плата, численность персонала, площадь и т.д.), формализованные характеристики размещения, окружения, транспортной доступности.

3. *Оптимизация товарных потоков в торговой компании* предполагает, в частности, алгоритмизацию задач перераспределения товаров внутри сети, их ротацию, агрегирование, своевременную коррекцию цен и выставление товаров на распродажу. Основной целью в реинжиниринге подобных бизнес-процессов оказывается формализация экспертных знаний. Но и здесь как вспомогательный элемент анализа успешно привлекаются элементы Data Mining, в частности, переборные схемы для анализа цепочек товарного движения (поступление на склад → приход в магазин → продажа с учетом возможных перемещений между магазинами, возвратами и т.д.). Подобная аналитика позволяет получать формализованные характеристики исследуемых бизнес-процессов и, нередко, выявлять расхождения в представлениях экспертов-менеджеров с реальными фактами, подтверждаемыми накопленной информацией.

Подводя итог, отметим, что успешность решения различных аналитических задач опреде-

***вторая
прикладная задача
CRM-систем
связана
с прогнозированием
продаж***

**успешность
решения
аналитических
задач определяется
во многом
квалификацией
самого
пользователя**

ляется не только специфическими особенностями применяемых вычислительных методов, но и во многом квалификацией пользователя – насколько хорошо он разбирается в том, что пытается применить. Полагаем поэтому, что базовые представления и знания о ключевых возможностях методов интеллектуального анализа необходимы многим специалистам, деятельность которых связана с анализом больших массивов экономической информации.

Литература

1. Fayyad U. From data mining to knowledge discovery: an overview / U. Fayyad, G. Piatetsky-Shapiro, P. Smyth. – In Advances in Knowledge Discovery and Data Mining. Menlo Park, CA: AAAI Press/ The MIT Press, 1996. – pp.495-515.
2. Меркулина И.А. Применение методов обнаружения логических закономерностей в экономических задачах / И.А. Меркулина, А.П. Никитин. – М.: ИВЦ «Маркетинг», 2007. – 354 с. – ISBN 978-5-7856-0587-9.
3. Sumathi S. Data Mining in business / Sumathi S., Sivanandam S.N. // Studies in Computational Intelligence. – 2006. – v. 29. – p.387-409.
4. Lavrac N. Introduction: Lessons learned from data mining applications and collaborative problem solving/ Lavrac N., Motoda H., Fawcett T., Holte R., Langley P., Adriaans P. // Machine Learning. – 2004. – v. 57. – p. 13-34.
5. Меркулина И.А. Дипломное проектирование и выпуск конкурентноспособных информатиков-экономистов / Меркулина И.А., Никитин А.П., Каширская Е.Н. // Прикладная Информатика. 2007. – №3. – с. 60-72. – ISSN – 1993-8314.
6. Полежаев И.Е. Метод сегментации клиентских баз данных на основе жизненного цикла клиента [Электронный ресурс]. – Электронный научный журнал «Исследовано в России» 2006. – с.1875-1902. – zhurnal.ape.relarn.ru/articles/2006/200.pdf
7. Makridakis S.G. Forecasting: methods and applications / S.G. Makridakis, S.C. Wheelwright, R.J. Hyndman. – 3rd edition. – Wiley, 1998. – 656 p.

КЭ