

УДК 004.89; 004.67

Ю. В. Николаева, аспирант, ИжГТУ имени М. Т. Калашникова
Е. А. Сучкова, аспирант, ИжГТУ имени М. Т. Калашникова

ПРИМЕНЕНИЕ DATA MINING В НЕФТЕГАЗОВОЙ ОТРАСЛИ

По данным международной консалтинговой компании Molten [www.molten-group.com] крупные нефтегазовые компании тратят от \$1 до \$3 млрд в год на сбор данных, из них расходы на поддержание и обработку накопленной информации зачастую составляют менее 1 % от этой суммы [1].

Molten, клиентами которой являются лидеры отрасли, такие как bp, Lukoil, Gazprom, подчеркивает важность управления данными в слоганах своей компании:

– Управление данными повышает эффективность бизнеса и снижает риски (“Good data management drives business performance and mitigates risk”);

– Недостоверные данные и недостаточная информация могут иметь серьезные последствия (“Bad data and poor information can have severe consequences”) [2].

Нефтегазовая отрасль пронизана насквозь сложными технологическими и финансово-экономическими процессами, каждый из которых характеризуется огромным количеством показателей и критериев. Все эти показатели, критерии, характеристики и ресурсы процессов необходимо учитывать, записывать в специальные хранилища данных, обеспечивать оперативный доступ к ним заинтересованных лиц, многие из показателей необходимо рассчитывать на основании других.

В связи с непрекращающимся ростом фиксируемых данных оптимальные подходы по работе с ними, обработке полученных массивов данных и извлечению из них полезной информации для принятия стратегических решений актуальны для предприятий нефтегазового сектора.

В настоящее время интерес к технологиям работы с большими объемами данных (big data) растет. Технологии big data включают в себя подходы, инструменты и методы для обработки и анализа больших объемов слабоструктурированных и многокритериальных данных (подробнее о big data см. [3, 4]). Исследователи, аналитики и разработчики технологий по работе с большими данными, характеризующимися значительным объемом, сложной структурой, высокими требованиями к скорости их передачи, пытаются найти ответы на два вопроса:

– Как хранить и передавать данные?

– Как обрабатывать и анализировать данные, преобразовывать их в полезную информацию для конечного пользователя?

Среди методов, отвечающих на второй вопрос, выделяют обширную группу методов data mining [5, 6, 7]. Понятие data mining подразумевает извлечение знаний из массивов данных. Данные представ-

ляются как слабоструктурированные, многовариантные, зашумленные массивы, обработка и анализ которых для человека без специальных методов, инструментов, приложений является практически невозможной задачей.

Например, в сложной технологической цепочке по добычи нефти идет сбор показателей различных датчиков большого количества оборудования. Диспетчеру необходимо оперативно отслеживать все эти показания, замечать отклонения от норм, впоследствии анализировать накопленные значения для мониторинга состояния оборудования. Отдельное значение каждого из показателей является единицей данных, однако при анализе важно общее состояние, складывающееся из совокупности значений показателей, из определенных закономерностей и взаимосвязей для выявления причин сбоев оборудования. В приведенном примере знаниями или информацией, полезной пользователю, будут являться состояния оборудования: его исправность или необходимость в ремонте. Эти знания получаются путем анализа исходных значений показателей.

Получение такой «содержательной информационной выжимки» из огромного количества исходных разнотипных данных и является целью методов группы data mining.

Подходы и методы data mining нацелены на решение масштабных аналитических задач следующих классов.

1. Классификация – отнесение отдельных данных к определенному классу, число и характеристики которых заранее известны. Например, разрабатываемые скважины можно разделить на несколько классов по технико-экономической эффективности в зависимости от текущей и накопленной добычи нефти, водонефтяного и газового факторов, коэффициента нефтеотдачи [8].

2. Кластеризация – группировка исходных данных по определенным схожим свойствам в кластеры. Количество образованных кластеров и свойства, по которым определяется принадлежность к кластерам, становятся известными только в ходе анализа. Примером задачи кластеризации в нефтегазовой отрасли может служить выделение объектов разработки (одних или нескольких продуктивных пластов месторождений) по физико-химическим свойствам нефти, воды и газа [9].

3. Регрессия – установление наличия и характера зависимости одной величины от другой или нескольких других величин. В обеспечении функционирования месторождений может быть полезным знание зависимости электропотребления от количества и раз-

мещения эксплуатационных и нагнетательных скважин. Также регрессия используется в целях прогнозирования, то есть по уже выявленной функциональной зависимости определяется значение неизвестной величины. К примеру, по измеренным геолого-физическим свойствам пород, размерам и расположению пластов залежей прогнозируются значения показателей добычи нефти, газа.

4. Ассоциация – определение возможных отношений или связей между величинами. Пусть при геологоразведке нескольких скважин было обнаружено наличие или отсутствие в них газов метана, пропана, бутана, углекислого газа, превышающих минимально заданное значение для каждого. Ставится вопрос, есть ли связь между содержанием газов и объемами нефтяных залежей больше определенного количества? В данном случае наличие или отсутствие превышения минимального уровня содержания газов и объемов нефтяных залежей берется за факт, и отношения устанавливаются именно между фактами [10].

5. Распознавание паттернов – выявление наиболее распространенных шаблонов и отклонений от них. С помощью инструментов и методов распознавания паттернов можно установить последовательности наступления событий. Пусть установлена связь между содержанием газов и объемом залежи нефти в пласте. Можно выявить паттерн уменьшения процентного содержания газов, при котором наступит событие истощения нефтяной залежи.

Таким образом, использование методов data mining помогает решать широкий круг аналитических задач, связанных с технологическими процессами по разработке месторождений и добычи нефти – от кластеризации скважин и классификации месторождений, прогнозирования объемов добычи до прогнозирования нештатных ситуаций на основе показателей работы оборудования. Решение этих задач может принести практическую пользу, упрощая принятие стратегически важных решений.

Помимо технической составляющей отрасли современные методы data mining активно применяются также в финансовой и планово-экономической деятельности предприятий нефтегазового сектора.

В настоящее время крупные предприятия основывают схемы расчета плановых значений KPI (ключевых показателей эффективности), в том числе финансовых, на алгоритмах извлечения данных. Использование в данном вопросе современных методов анализа данных, таких как нейронные сети, нечеткая логика, может стать значительным преимуществом в выполнении стратегических задач и достижении целей предприятия.

На основные финансовые KPI могут влиять множество экономических показателей рынка, таких как цены на сырье, энергоресурсы, закупаемые материалы, к тому же конечные значения показателей эффективности предприятия могут быть отредактированы под воздействием внешнеэкономических факторов, таких как отраслевые индексы, цены на конечную продукцию и даже цены на свои акции и акции конку-

рентов. Все эти зависимости могут быть установлены, спрогнозирована степень их влияния, и выбраны оптимальные варианты действий для увеличения экономической эффективности.

Рассмотрим подробнее несколько примеров применения методов data mining для работы с экономическими показателями предприятий нефтегазового сектора.

Экономическая эффективность предприятия в немалой степени зависит от общего экономического состояния отрасли. В этой связи решение задачи прогнозирования котировок на нефтяные и газовые фьючерсы, отраслевых индексов, акций предприятий-конкурентов нефтегазового сектора становится актуальным. Мощным инструментом data mining для прогнозирования временных рядов является нейросетевая методология, объединяющая широкий круг методов и алгоритмов. Нейронные сети представляют собой математическую модель со структурой, напоминающей по своему строению нервную ткань мозга человека. Нейросетевая методология в отличие от классических статистических методов прогнозирования (регрессионный анализ, модели экспоненциального сглаживания) позволяет корректировать параметры нейронных сетей в процессе обучения, обеспечивая наибольшую точность аппроксимации. Нейросеть состоит из нескольких слоев нейронов, каждый из которых включает в себя некоторое количество нейронов, нейроны слоев сообщаются друг с другом посредством связей – все эти свойства нейронной сети объединяются в понятие «архитектура нейросети» (подробнее о выборе архитектуры нейронной сети см. [11]). Наиболее подходящими архитектурами для задачи прогнозирования являются сети Ворда и многослойный перцептрон.

Также в текущей беспокойной экономической ситуации важно учитывать при планировании финансово-экономических показателей внезапные колебания котировок фьючерсов на нефть и газ, объясняемые зависимостью от кризисных факторов. Анализ временных рядов котировок и выявление в них кризисных паттернов возможны с использованием нейронных сетей классов, направленных на распознавание образов, например, сети Хопфилда (подробнее см. [12]).

Следующей важной задачей повышения экономической эффективности производства является снижение себестоимости продукции. Себестоимость добычи нефти и газа складывается из себестоимости работ на каждом этапе: нефтегазоразведка, геолого-разведочные работы, работы по непосредственному извлечению нефти. Каждый из этапов работ требует использования электроэнергии, транспортно-заготовительных услуг, специального оборудования и материалов (например, при бурении и креплении скважин используются насосы, лебедки, краны, трубы обсадные, цемент, глина, буровые растворы). Для снижения стоимости энергоресурсов, материалов и оборудования необходимо оптимизировать закупочную деятельность, в первую очередь за счет правильного выбора по-

ставщиков, основанного на некотором количестве взвешенных критериев. В процессе выбора оптимального поставщика необходимо учесть различные факторы, такие как цены на услуги и продукты, количество успешных сделок с данным контрагентом, наличие задержек в поставке товаров и т. п. Такой многокритериальный анализ слабоструктурированных данных является одним из направлений data mining. Для проведения такого анализа может быть использовано сочетание метода анализа иерархий с алгоритмами, основанными на fuzzy logic (нечеткая логика) [13]. При принятии решений в закупочной деятельности сложно оперировать понятиями классической математической логики «хорошо/плохо», «да/нет». Необходимую гибкость в процессе обработки критериев выбора поставщика обеспечивает нечеткая логика, которая расширяет спектр доступных вариантов. Например, по критерию «задержки поставок» можно вводить не только варианты «есть», «нет», но и некоторый диапазон промежуточных значений: «задержки были, но кратковременные», «задержки в средних пределах», «частые продолжительные задержки». Еще одним способом эффективной оптимизации закупочной деятельности является управление рисками. Виды закупочных рисков в нефтегазовой сфере и стратегии работы с этими рисками описаны в [14]. Применение методов data mining позволяет получить более точные оценки рисков и последствий их реализации, что позволяет сформировать оптимальную стратегию диверсификации, страхования и формирования резервов с целью минимизации возможных последствий и обеспечения стабильной работы предприятия.

Из приведенных примеров становится ясно, что деятельность предприятий нефтегазовой отрасли является обширным полем для применения методов data mining. Интеллектуальный анализ данных способствуют уменьшению расходов, возникающих при ошибках в производственных вопросах и при выборе стратегий экономического развития. Применение методов и алгоритмов извлечения из данных информации и знаний дает значительное преимущество при принятии стратегических решений.

Таким образом, эффективное управление данными может стать залогом повышения эффективности предприятия.

Библиографические ссылки

1. *Смирнов Е.* Лидерами нефтегаза станут компании, использующие Big Data. – URL: http://www.cnews.ru/news/top/liderami_neftegaza_stanut_kompanii (дата обращения: 25.10.2015).
2. Data Management Assessment Tool. – URL: http://www.molten-group.com/wp-content/uploads/2014/06/Molten_DMAT_information.pdf (дата обращения: 25.10.2015).
3. *Черняк Л.* Большие данные – новая теория и практика // Открытые системы. СУБД – М. : Открытые системы, 2011. – № 10.
4. *Майер-Шенбергер В., Кукьер К.* Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим : пер. И. Гайдюк. – Манн, Иванов и Фербер, 2013. – 240 с.
5. Data Mining – добыча данных. – URL: <http://basegroup.ru/community/articles/data-mining> (дата обращения: 26.10.2015).
6. Методы и модели анализа данных: OLAP и Data Mining / А. А. Барсебян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. – СПб. : ХВ-Петербург, 2004. – 336 с.
7. Введение в современный Data Mining. – URL: <http://statistica.ru/local-portals/data-mining> (дата обращения: 26.10.2015).
8. Понятие о разработке нефтяных месторождений. – URL: <http://www.intuit.ru/studies/courses/4808/1056/lecture/16212> (дата обращения: 27.10.2015).
9. Понятие о разработке нефтяных месторождений. – URL: <http://www.intuit.ru/studies/courses/4808/1056/lecture/16212> (дата обращения: 27.10.2015).
10. Правило ассоциаций в нефтегазодобыче. – URL: http://statsoft.ru/solutions/ExamplesBase/branches/detail.php?ELEMENT_ID=704 (дата обращения: 27.10.2015).
11. *Николаева Ю. В.* Критерии выбора архитектуры нейронной сети для прогнозирования финансовых рынков // Вестник ИжГТУ имени М. Т. Калашникова. – 2015. – № 1(65). – С. 96–97.
12. *Николаева Ю. В.* Распознавание кризисных паттернов в котировках фьючерсов на нефть сорта Brent // Молодые ученые – ускорению научно-технического прогресса в XXI веке : сборник материалов III Всерос. науч.-техн. конф. аспирантов, магистрантов и молодых ученых с международным участием (Ижевск, 22-23 апреля 2015 года) / ИжГТУ имени М. Т. Калашникова.
13. *Сучкова Е. А., Лялин В. Е.* Проблемы выбора поставщика – критерии, инструменты и методы оценки // Математические модели и информационные технологии в организации производства. – 2012. – № 2(25). – С. 39–48.
14. *Сучкова Е. А.* Управление контрактационными рисками при закупке оборудования в нефтегазовой отрасли // Вестник ИжГТУ имени М. Т. Калашникова. – 2015. – № 1(65). – С. 98–99.