

грам ответственности и местам формирования затрат //Бухгалтерский учет. -2000. – № 5. – с. 56–59.

4 Аверчев, И.В МСФО. 1000 примеров применения/ Аверчев И.В. – РидГрупп, 2011. – 992 с.

УДК 004:338

Н.Г. Сияк, доц., канд. экон. наук siniakn@mail.ru

Хабиб А., асп.; Шариф Н., асп.
habib.awada@liu.edu.lb (БГТУ, г. Минск)

МЕТОДИКИ И НЕОБХОДИМОСТЬ ИСПОЛЬЗОВАНИЯ АНАЛИЗА БОЛЬШИХ ДАННЫХ В ОТРАСЛЯХ ЭКОНОМИКИ

Существует множество разнообразных методик анализа массивов данных, в основе которых лежит инструментарий, заимствованный из статистики и информатики. Ниже приведен список методик анализа больших данных [1]. Список не претендует на полноту, однако в нем отражены наиболее востребованные в различных отраслях подходы. Исследователи продолжают работать над созданием новых методик и совершенствованием существующих. Кроме того, некоторые из перечисленных них методик не обязательно применимы исключительно к большим данным и могут с успехом использоваться для меньших по объему массивов (например, А/В-тестирование, регрессионный анализ).

A/B testing. Методика, в которой контрольная выборка поочередно сравнивается с другими. Тем самым удается выявить оптимальную комбинацию показателей для достижения, например, наилучшей ответной реакции потребителей на маркетинговое предложение. Большие данные позволяют провести огромное количество итераций и таким образом получить статистически достоверный результат.

Associationrulelearning. Набор методик для выявления взаимосвязей, т.е. ассоциативных правил, между переменными величинами в больших массивах данных. Используется в **datamining**.

Classification. Набор методик, которые позволяют предсказать поведение потребителей в определенном сегменте рынка (принятие решений о покупке, отток, объем потребления и проч.). Используется в **datamining**.

Clusteranalysis. Статистический метод классификации объектов по группам за счет выявления наперед не известных общих признаков. Используется в **datamining**.

Crowdsourcing. Методика сбора данных из большого количества источников.

Datafusionanddataintegration. Набор методик, который позволяет анализировать комментарии пользователей социальных сетей и сопоставлять с результатами продаж в режиме реального времени.

Datamining. Набор методик, который позволяет определить наиболее восприимчивые для продвигаемого продукта или услуги категории потребителей, выявить особенности наиболее успешных работников, предсказать поведенческую модель потребителей.

Ensemblelearning. В этом методе задействуется множество предикативных моделей за счет чего повышается качество сделанных прогнозов.

Geneticalgorithms. В этой методике возможные решения представляют в виде `хромосом`, которые могут комбинироваться и мутировать. Как и в процессе естественной эволюции, выживает наиболее приспособленная особь.

Machinelearning. Направление в информатике (исторически за ним закрепилось название `искусственный интеллект`), которое преследует цель создания алгоритмов самообучения на основе анализа эмпирических данных.

Naturallanguageprocessing (NLP). Набор заимствованных из информатики и лингвистики методик распознавания естественного языка человека.

Networkanalysis. Набор методик анализа связей между узлами в сетях. Применительно к социальным сетям позволяет анализировать взаимосвязи между отдельными пользователями, компаниями, сообществами и т.п.

Optimization. Набор численных методов для редизайна сложных систем и процессов для улучшения одного или нескольких показателей. Помогает в принятии стратегических решений, например, состава выводимой на рынок продуктовой линейки, проведении инвестиционного анализа и проч.

Patternrecognition. Набор методик с элементами самообучения для предсказания поведенческой модели потребителей.

Predictivemodeling. Набор методик, которые позволяют создать математическую модель наперед заданного вероятного сценария развития событий. Например, анализ базы данных CRM-системы на предмет возможных условий, которые подтолкнут абоненты сменить провайдера.

Regression. Набор статистических методов для выявления закономерности между изменением зависимой переменной и одной или несколькими независимыми. Часто применяется для прогнозирования

и предсказаний. Используется в datamining.

Sentimentanalysis. В основе методик оценки настроений потребителей лежат технологии распознавания естественного языка человека. Они позволяют вычлениить из общего информационного потока сообщения, связанные с интересующим предметом (например, потребительским продуктом). Далее оценить полярность суждения (позитивное или негативное), степень эмоциональности и проч.

Signalprocessing. Заимствованный из радиотехники набор методик, который преследует цель распознавания сигнала на фоне шума и его дальнейшего анализа.

Spatialanalysis. Набор отчасти заимствованных из статистики методик анализа пространственных данных – топологии местности, географических координат, геометрии объектов. Источником больших данных в этом случае часто выступают геоинформационные системы (ГИС).

Statistics. Наука о сборе, организации и интерпретации данных, включая разработку опросников и проведение экспериментов. Статистические методы часто применяются для оценочных суждений о взаимосвязях между теми или иными событиями.

Supervisedlearning. Набор основанных на технологиях машинного обучения методик, которые позволяют выявить функциональные взаимосвязи в анализируемых массивах данных.

Simulation. Моделирование поведения сложных систем часто используется для прогнозирования, предсказания и проработки различных сценариев при планировании.

Timeseriesanalysis. Набор заимствованных из статистики и цифровой обработки сигналов методов анализа повторяющихся с течением времени последовательностей данных. Одни из очевидных применений – отслеживание рынка ценных бумаг или заболеваемости пациентов.

Unsupervisedlearning. Набор основанных на технологиях машинного обучения методик, которые позволяют выявить скрытые функциональные взаимосвязи в анализируемых массивах данных. Имеет общие черты с **ClusterAnalysis**.

Visualization. Методы графического представления результатов анализа больших данных в виде диаграмм или анимированных изображений для упрощения интерпретации облегчения понимания полученных результатов.

Хотя большие данные и бизнес-аналитика имеют одинаковую цель (поиск ответов на вопрос), они отличаются друг от друга. Работа с большими данными не похожа на обычный процесс бизнес-аналитики, где простое сложение известных значений приносит ре-

зультат: например, итог сложения данных об оплаченных счетах становится объемом продаж за год. При работе с большими данными результат получается в процессе их очистки путём последовательного моделирования: сначала выдвигается гипотеза, строится статистическая, визуальная или семантическая модель, на ее основании проверяется верность выдвинутой гипотезы и затем выдвигается следующая [2].

Технологии BigData успешно реализуются в различных отраслях в западных странах: банки, телеком, ритейл, энергетика, медицина, строительство и управление городской инфраструктурой. При всем разнообразии задач решения в сфере BigData пока не приобрели ярко выраженной отраслевой направленности. Рынок находится не просто на стадии активного формирования, а в самом начале этой стадии.

Несмотря на малый срок существования сектора BigData, уже есть оценки эффективного использования этих технологий, основанные на реальных примерах. Один из самых высоких показателей относится к энергетике – по оценкам аналитиков, аналитические технологии BigData способны на 99% повысить точность распределения мощностей генераторов. А здравоохранение США, благодаря BigData, может сэкономить до \$300 млрд.

Готовность к применению технологии BigData складывается из пяти составляющих: желание, накопленные данные, адаптация технологий, отлаженные процессы и персонал.

По оценке IDC, например, коммунальный сектор находится на начальных стадиях освоения технологий BigData. Эти технологии помогают оптимизировать производство энергии, операционную эффективность и работу с клиентами. Аналитика позволяет заблаговременно готовиться к отключениям, а также оценивать энергетический рынок, прогнозировать спрос и вести подсчет финансовых показателей.

«Большие данные» помогают промышленным предприятиям урезать затраты, увеличивать эффективность работы и качество продукции, своевременно обнаруживать проблемы в работе и адаптировать продукцию в соответствии с информацией о непосредственных потребителях.

ЛИТЕРАТУРА

1 Berman J. J. Confidentiality for medical data miners / J. J. Berman // *Art Intell Med*. 2002. Pp. 25–36.

2 A universal legal framework as a prerequisite for database interoperability. / D. Greenbaum, M. Gerstein // *Nature Biotechnol*. 2003. 21:979–82.