

# ГЛУБОКОЕ ПОГРУЖЕНИЕ

САМЫЕ ЦЕННЫЕ ПОЛЕЗНЫЕ ИСКОПАЕМЫЕ СЕГОДНЯШНЕГО ВРЕМЕНИ ОТНЮДО НЕ УГЛЕВОДОРОДЫ, А ИНФОРМАЦИЯ. НАБЛЮДАТЬ ЗА КЛИЕНТАМИ И ХРАНИТЬ ДАННЫЕ ОБ ИХ ПРЕДПОЧТЕНИЯХ, ПОВЕДЕНИИ И ПОКУПКАХ НАУЧИЛИСЬ ДАВНО. НО ЛИШЬ НЕДАВНО КОМПЬЮТЕРЫ СТАЛИ ДОСТАТОЧНО СЛОЖНЫМИ ДЛЯ ТОГО, ЧТОБЫ АНАЛИЗИРОВАТЬ ЭТУ ИНФОРМАЦИЮ. РЫНОК ОБРАБОТКИ БОЛЬШИХ ДАННЫХ СТОИТ НА ПОРОГЕ БОЛЬШОГО РОСТА: СОГЛАСНО ПРОГНОЗУ АНАЛИТИКОВ PIERRE AUDOIN CONSULTANTS, В 2016 ГОДУ ЕГО ОБЪЕМ СОСТАВИЛ €20 МЛРД — ПОЧТИ В СЕМЬ РАЗ БОЛЬШЕ, ЧЕМ В 2010 ГОДУ. СВЕТЛАНА РАГИМОВА

**УМНЫЕ ПРОДАЖИ** В 2008 году в Швеции появился сервис по мониторингу будущих событий под названием Recorded Future. Всего за три года работы компания достигла оборота более чем в \$100 млн в год, а среди ее клиентов оказались участники рейтинга Fortune 500, государственные структуры и крупные финансовые корпорации. По сути, команда сервиса примерно из 20 человек занимается тем, что поддерживает и дорабатывает ПО, собирающее данные из открытых источников о событиях, которые пока только планируются и должны произойти в будущем. Это может быть что угодно: от расписания встреч Барака Обамы до анонсов новых смартфонов. Все зависит от задач клиента. Вариантов использования сервиса — масса. Самый очевидный — для увеличения заработков в игре на бирже.

При этом команда Recorded Future состоит не из экономистов или политиков, а в основном обладателей ученых степеней в области компьютерных наук. Сервис, который они создали, и есть замечательный пример использования технологий по добыче информации (data mining).

Директор отделения аналитических технологий компании ФОРС Евгений Курилович объясняет: «Термин data mining обозначает не столько конкретную технологию, сколько сам процесс поиска корреляций, тенденций, взаимосвязей и закономерностей посредством различных математических и статистических алгоритмов: кластеризации, создания субвыборок, регрессионного и корреляционного анализа. Цель этого поиска — представить данные в виде, четко отражающем бизнес-процессы, а также построить модель, при помощи которой можно прогнозировать процессы, критичные для планирования бизнеса (например, динамику спроса на те или иные товары или услуги либо зависимость их приобретения от каких-то факторов)». Подобного рода информация обычно используется при прогнозировании, стратегическом планировании, анализе рисков, и ценность ее для предприятия очень высока. Подобным образом могут быть выявлены самые необычные и даже парадоксальные взаимосвязи.

Один из примеров использования этого класса технологий касается ритейлера Walmart. Компания считается пионером в практическом применении данного метода. Анализ огромных массивов данных производится ритейлером для налаживания взаимоотно-

шений с поставщиками. Еще в 1995 году этот розничный гигант начал собирать информацию обо всех состоявшихся покупках в 2900 своих магазинах и складировать ее в систему хранения. Доступ к данным и аналитике по продажам конкретных продуктов получили все 3500 поставщиков ритейлера, которые до сих пор используют эту возможность для выявления привычек и предпочтений клиентов на уровне конкретных магазинов. Это позволяет составлять поставки так, чтобы максимально точно соответствовать вкусам покупателей, а также правильно выставлять товары на полках, проводить локальные промо-акции и т. д.

Помимо очевидного анализа покупательской корзины, который помогает выработать общую стратегию создания запасов товаров и их раскладки в торговых залах, дата-майнинг также позволяет исследовать временные шаблоны, благодаря которым сети розничной торговли могут, к примеру, определить недельную сезонность товаров или подсчитать, через какое время человек, купивший фотоаппарат, придет за новыми батарейками. С помощью дата-майнинга также можно создавать прогнозирующие модели, которые позволяют на базе полученных данных разрабатывать точно направленные мероприятия по продвижению товара.

Тот же Walmart сейчас использует этот метод для проведения так называемого market-basket анализа, то есть выявления тех продуктов, которые покупатели приобретают совместно, за одну покупку, и впоследствии расставляет эти товары самым удобным для покупателя образом.

Подобного рода интеллектуальным анализом данных занимается и крупнейший в мире интернет-магазин Amazon, который использует полученную информацию для составления персональных и привлекательных для пользователей предложений.

**ВЕРШИНА АЙСБЕРГА** Потенциал использования средств data mining в бизнесе огромен. Евгений Олейник, руководитель департамента HP Autonomy/IM в странах Центральной и Восточной Европы, рассказывает, что на сегодняшний день 85% всех данных в мире составляют неструктурированный поток информации и только 15% данных — структурированный. С каждым годом доля структурированных данных в этом потоке уменьшается: все больше людей используют социальные сети, в городах все больше появляется камер видеонаблюдения. Люди используют одновременно множество разных каналов коммуникации, и возникает необходимость определить взаимосвязь между отдельными информационными событиями, получаемыми из разных источников. «Сегодня организации могут принимать бизнес-решения только за счет анализа 15% структурированных данных. Если организация могла бы использовать и анализировать все 100% имеющихся данных, это позволило бы ей принимать более грамотные бизнес-решения на основе всей су-

ществующей информации», — говорит господин Олейник. К примеру, продукт HP Autonomy позволяет получать представление о том, что содержится в 100% информации, и определять взаимосвязи между разными источниками.

Дмитрий Шепелявый, директор департамента технологических решений SAP СНГ, рассказывает, что дата-майнинг возник и развивается на базе статистики, распознавания образов, методов искусственного интеллекта и теории баз данных. Используемые сырые данные могут быть совсем разными, но применять методы глубинного анализа имеет смысл только к достаточно большим массивам информации. Это могут быть компании, работающие в сфере розничной торговли, банковской отрасли, телекоммуникационной, страховой индустрии и т. д. Безусловно, эти же методы анализа данных могут быть полезны госсектору. Государственные органы могут, к примеру, таким способом выявлять лиц, уклоняющихся от уплаты налогов, или проверять таможенные грузы и т. п. Та же SAP предлагает несколько решений для дата-майнинга — как визуальных данных (SAP BusinessObjects Explorer), так и статистических (BusinessObjects Predictive Workbench).

**СХВАТИТЬ ЗА РУКУ** Сегодня корпорации хотят получать аналитику, влияющую на их бизнес, очень быстро, в идеале — в режиме реального времени. Технологии Real-time Data Management типа SAP HANA способны обеспечить анализ громадных объемов данных практически в онлайн-режиме. Oracle Advanced Analytics, составной частью которого является Oracle Data Mining, также позволяет создавать приложения для анализа в реальном времени и стратегического планирования в таких областях, как потеря и приобретение клиентов, подготовка рекомендаций по продукции и предупреждение мошенничества. Этим инструментарием уже активно пользуются многие крупнейшие глобальные корпорации — BMW, Philips IP&S and Trepapel, Xerox, Walter Reed Medical Center, NTT Comware и другие.

Не так давно крупный европейский банк, действующий в России и СНГ, столкнулся с кражей 150 тыс. руб. со счета клиента за 15 минут из 12 разных банкоматов. Если бы компания использовала решения дата-майнинга, такой ситуации удалось бы избежать. Данная технология активно применяется финансовыми организациями для выявления и предотвращения случаев мошенничества. В страховых компаниях это дает возможность определять, является ли заявка на возмещение ущерба мошеннической или нет. Считается, что от 10% до 20% всех требований на выплату страховок фабрикуются мошенническим путем. Например, если машина куплена в другом регионе и попала в аварию практически сразу после того, как была застрахована, случай с высокой долей вероятности будет включен в выборку для дальнейшего анализа и проверки. Генеральный директор SAS Россия/СНГ Валерий Панкратов рассказывает, что в Бельгии государственная служба по борьбе

с мошенничеством с помощью аналитических решений выявляет целые преступные группы физических и юридических лиц, которые совершают махинации с возмещением экспортного НДС, в результате ежегодно предотвращаются необоснованные выплаты на сумму около €80 млн. На основе уже выявленных случаев мошенничества строятся профили, или набор правил, по которым аналитическая система автоматически определяет, является ли новый случай мошенническим.

Олег Данильченко, руководитель направления data mining, IBM в России и СНГ, говорит, что методы, предполагающие применение средств data mining, разительно отличаются от традиционных подходов к обнаружению мошенничества тем, что выходят далеко за рамки простых отчетов об исключительных ситуациях. Эти средства выявляют подозрительные случаи на основе шаблонов данных, позволяющих сделать предположение о мошенничестве.

Также дата-майнинг актуален при принятии решения о выдаче кредита, в том числе и на основе анализа информации о клиенте в социальных сетях и других открытых источниках. Классическое применение — это кредитный скоринг, то есть оценка платежеспособности заемщика и расчет вероятности невозврата кредита. Для этого анализируются данные о клиентах, которым ранее выдавали кредиты и у которых уже есть кредитные истории; эта информация обрабатывается алгоритмами data mining, в них выявляются скрытые неочевидные закономерности, и на их основе строится математическая модель, которая помогает определять, кому из потенциальных заемщиков выдавать кредит, а кому отказать. Таким образом, сбор данных помогает прогнозировать кредитные риски и автоматически принимать решения о выдаче кредитов. «Эту задачу решают банки по всему миру, в том числе и российские, — говорит господин Панкратов. — У нас, например, были проекты внедрения решений для кредитного скоринга в Сбербанке, ВТБ 24, Банке Москвы, Росбанке, Юникредит-банке».

Господин Шепелявый рассказывает об успешном внедрении дата-майнинга в Bank of America, девиз которого — принесение как можно больше пользы каждому клиенту. С помощью специальных инструментов глубинного анализа данных банк проанализировал профили клиентов, и теперь вместо прослушивания стандартного маркетингового предложения от специалистов колл-центров клиенты получают индивидуальные предложения.

**СНЕЖНЫЙ ТЕЛЕКОМ** В России продукты дата-майнинга пока не слишком распространены. Тем не менее это не повод для пессимизма. Максим Андреев, заместитель директора департамента информационных технологий и руководитель направления бизнес-приложений компании КРОК, говорит: «Принято считать, что в России распространение технологий лет на пять отстает от Запада. Но именно в случае с data mining

**СЕГОДНЯ КОРПОРАЦИИ ХОТЯТ ПОЛУЧАТЬ АНАЛИТИКУ, ВЛИЯЮЩУЮ НА ИХ БИЗНЕС, ОЧЕНЬ БЫСТРО, В ИДЕАЛЕ — В РЕЖИМЕ РЕАЛЬНОГО ВРЕМЕНИ**

↑  
ТRENДЫ