

СОВРЕМЕННЫЕ АСПЕКТЫ РАЗВИТИЯ ОТРАСЛЕВЫХ ПРИЛОЖЕНИЙ НА ОСНОВЕ ТЕХНОЛОГИЙ БОЛЬШИХ ДАННЫХ: ВОЗМОЖНОСТИ, ПЕРСПЕКТИВЫ И ОГРАНИЧЕНИЯ

М.В. Рамзаев

Международный институт рынка

Технологии хранения, обработки и интеллектуального анализа Больших данных получают все большее распространение во всем мире. Области применения также расширяются, открывая новые возможности поиска решений самых разнообразных задач. Вместе с тем, важно учитывать риски и ограничения применения данных технологий, проводя глубокие исследования по их выявлению. Таким образом, необходимо формировать новые алгоритмы анализа Больших данных с точки зрения сочетания их широких возможностей и границ в быстроменяющихся условиях.

В современном мире все больше сфер жизнедеятельности человека охватываются теми или иными видами информационных технологий и их всевозможными отраслевыми приложениями. Несмотря на то, что существующие механизмы хранения и обработки данных активно совершенствуются, актуальность разработок и внедрения инноваций ИТ-сектора является одной из наиболее важных задач для экономик стран всего мира. Наиболее передовые, обсуждаемые и, по мнению ученых, перспективные в этой области направления – это обработка, хранение и интеллектуальный анализ Больших данных (BIG DATA) с использованием облачных технологий.

Данные технологии уже активно используются во многих областях науки и жизнедеятельности. В качестве примера можно привести маркетинг, энергетику, промышленность, нефтегазовый сектор, государственное и муниципальное управление, жилищно-коммунальное хозяйство («умные города»), банковскую и страховую деятельность, био- и нанотехнологии, геоинформационные технологии, авиакосмическую отрасль, безопасность различных сфер жизнедеятельности, здравоохранение и др.

Применение технологий Больших данных в России по данным аналитического агентства CNews Analytics и Oracle постепенно растет [3]. Вместе с тем, на данный момент существует и ряд скептических аспектов таких тенденций, которые могут существенно ограничить широкое применение указанных технологий. По мнению мирового экспертного сообщества, одним из наиболее проблемных вопросов внедрения и распространения BIG DATA аналитики в настоящее время является конфиденциальность персональной информации.

С каждым годом число компаний предоставляющих доступ к самым разнообразным базам данных, а также оказывающих услуги по их анализу возрастает. К примеру, по данным аналитического агентства Mind Commerce объем рынка BIG DATA в США в 2013 году составил 20 млрд. долларов, а в 2014 уже 29 млрд. То есть, рост составил 45%. Это существенный показатель формирующихся тенденций и трендов в секторе ИТ.

В этой связи, ввиду отсутствия соответствующей нормативной базы, регулирующей деятельность в области хранения, передачи третьим лицам и аналитику Больших данных, можно закономерно предположить, что будет возрастать и риск использования массивов информации в недобросовестных целях. Вместе с тем, существующий в России закон о конфиденциальности персональных данных [1] накладывает и целый ряд ограничений и не определяет новые возможности использования их в процессе интеллектуального анализа BIG DATA. Хотя уже и сейчас при определенном качестве запроса из любых массивов данных можно получать «персональные портреты» любого конкретного человека на основе сопоставления IP

адресов. А это не только банковские счета и реестр имущества, но и возможность вычислять алгоритмы кодов, необходимых для управления ими.

Вместе с тем, еще одним из наиболее уязвимых элементов персональных данных является личная медицинская информация. Современные тенденции к облачному хранению данных также способствуют возрастанию рисков их утечки. С одной стороны, организация процесса сбора и анализа всей имеющейся и поступающей в реальном времени информации о медицинских данных миллионов пациентов во всем мире может оказать существенное положительное влияние на сферу здравоохранения в целом. Это касается и выявления наиболее эффективных лекарственных средств, схем и методик лечения, видов диагностик и других медицинских мероприятий. С другой стороны, пациент – это один из самых незащищенных элементов всей этой процедуры. Он не имеет влияния на обработку информации о своем здоровье или лечении. Он имеет право только не персонализировать эти данные и остаться анонимом. Однако, анонимность в данном случае – понятие относительное, гарантию которого обеспечить в полной мере невозможно.

Существуют и иные опасения экспертов. Связаны они, прежде всего, с тем, что практически любое аналитическое агентство будет способно в реальном времени сегментировать общество по тем или иным наборам критериев, что, с одной стороны, позволит прогнозировать общественные реакции внутри каждого сегмента, с другой стороны упростит процесс манипуляции ими в корыстных целях. Более того, цены на все эти аналитические услуги в ближайшем будущем будут становиться только доступнее из-за роста предложений. К примеру, по данным агентства Economist Intelligence Unit survey, которое с помощью опросов выявляло приоритеты по развитию BIG DATA на несколько лет вперед, одним из самых резко развивающихся является такое направление, как управление человеческими ресурсами [2]. При этом, полагаю, что механизмы пассионарного управления в обществе давно используются, например, в США. Экспертное сообщество считает, что интерес коммерческих компаний к этой тематике за последующие три года возрастет практически в 2 раза, по сравнению с 2014 годом.

Еще один аспект в сложностях использования Больших данных, который отмечает профессор университета Беркли Майкл Джордан, связан с качественной сегментацией информации и ее анализа. Рост числа данных, рост числа массивов данных и рост уровня доступности к ним формирует у специалистов тенденции завышенных ожиданий, которые, как показывает история развития научных дисциплин, зачастую приводит к противоречиям и недостоверности результатов впоследствии. В этой связи важно уделять внимание изучению не только изучению и развитию широких возможностей BIG DATA, но и исследованиям рисков, негативных последствий, ограничений и соответствующей адаптации. К примеру, если пользователь несколько месяцев изучал информацию о телевизоре, а потом через интернет магазин купил его, то он уже не будет эффективным потребителем таргетированного маркетинга, который будет его преследовать еще несколько месяцев и более в сети интернет и мобильных устройствах. Вряд ли кто-то будет покупать телевизор каждый раз, увидев рекламный баннер в углу своего монитора.

В контексте возможных областей применения технологий BIG DATA все чаще рассматривается государственное управление, включающее и социальную сферу, и ЖКХ, и безопасность, и экономику, и много другое.

В сфере безопасности очень интересен опыт Пакистана, где была создана одна из крупнейших в мире база хранения и обработки данных граждан, включающая их биометрические показатели NADRA (National Database & Registration Authority) - национальное агентство по управлению базами данных и регистрациями граждан [4]. Основной стратегической целью создания такой системы и организации формирования информационной базы является борьба с коррупцией, противодействие терроризму, оптимизация процесса проведения реформ, а также повышение эффективности работы с последствиями стихийных бедствий. В целом экспертное сообщество сходится во мнении,

что данная инициатива реализована успешно и на сегодняшний день является уникальной моделью внедрения IT инноваций и BIG DATA технологий в государственное управление. В то же время, NADRA сегодня – абсолютно самостоятельная и финансово независимая организация, увеличившая свои доходы в 3 раза за последние 5 лет.

Однако проведенный анализ показывает, что повсеместное внедрение и применение BIG DATA аналитики в государственном управлении (в т.ч. real time администрирование) имеет целый ряд ограничений и проблем. Во-первых, это сложность в формировании самих баз или массивов данных о социально-экономических показателях тех или иных территорий. Во-вторых, трудноразрешима задача формирования потоковых данных: организация сбора и анализа информации в реальном времени, что практически сводит на нет любые алгоритмы выявления рисков и оптимальных решений в конкретных ситуациях. В-третьих, неотъемлемой частью применения IT инновация является компетенция кадров, где необходимо не только обучить государственных служащих, но и сформировать саму научную школу, специализирующуюся конкретно на данном направлении, что представляется крайне сложным для стран с низким и средним уровнем развития, а также имеющих обширные территории. В четвертых, необходимо учитывать, что технологии Больших данных как правило не дают точных рекомендаций, которые можно адаптировать в логический алгоритм. Скорее они дают набор результатов, связи которых с запросом не всегда просматриваются. Это значит, что в реальности можно говорить лишь о расчетах вероятностей возникновения тех или иных ситуаций или критических значений определенных показателей, что, в свою очередь означает, что в прогнозировании на основе BIG DATA аналитики речь идет о прогнозе вероятностей. Впрочем, это не говорит о том, что такого рода подходы в государственном управлении не эффективны. Напротив, это свидетельствует о том, что необходимо более взвешенно подходить к средствам хранения, обработки и анализа Больших данных в целом, учитывая и вопросы безопасности данных, и вероятности их неточности, а также подходить к анализу с точки зрения поиска логических взаимосвязей.

На сегодняшний день Data scientists (специалисты в области анализа данных) не приемлют процесс поиска логики, утверждая, что с появлением новых технологий он теряет свою эффективность и значимость. Впрочем, этот вопрос еще долгое время будет оставаться весьма дискуссионным, и баланс между возможностями и ограничениями будет смещаться в ту или другую сторону.

Литература

1. Федеральный закон от 27.07.2006 N 152-ФЗ «О персональных данных» (ред. от 21.07.2014).
2. Аналитический обзор рынка Big Data /<http://www.eiu.com/home.aspx/>
3. Исследование Oracle и CNews Analytics: Большие данные пришли в Россию /<https://www.oracle.com/ru/corporate/pressrelease/study-of-oracle-and-cnews-analytics-20150226.html/>
4. National Database & Registration Authority /<https://www.nadra.gov.pk/>