

ИНФОРМАЦИОННАЯ ПЕРЕГРУЗКА: BIG DATA, DATA MINING И МАРКЕТИНГ-МЕНЕДЖМЕНТ



Д. В. Ермолович
Доцент кафедры философии БГУИР, кандидат философских наук, доцент

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь
E-mail: ermolovich@bsuir.by

Abstract. The emergence and development of the information environment made safe storage, processing, transfer information, and obtaining structured information in the fastest time possible. But since the introduction of informatization occurs very rapidly, creating problems of information overload, as well as control over the dynamism and expansion of the information environment through Big Data, Data Mining and marketing-management technologies.

История «Больших данных» (БД) предельно молода, не более 10-ти лет, и повсеместную увлеченность БД можно назвать модой. Как и всякая мода, мода на БД непременно смениться. Уже в 2013 году БД-технология в маркетинге уступила пальму первенства не только мобильному маркетингу (mobile marketing), но и маркетингу контентному (content marketing) [1]. Однако мода проходит, а проблемы «Больших данных» и технология остаются. Быстро эволюционирующая БД-технология «доросла» уже до «четырех V»: «объем» (volume, постоянно расширяющийся), «скорость обработки» (velocity, допускающую оценку действительности в реальном времени), «достоверность» (veracity, максимально возможная степень совпадения с реальностью) и «разнообразие» (variety, возможность одновременной обработки разного уровня структурированности данных) [2], что предполагает более внимательное отношение к качеству получаемой и обрабатываемой информации, предполагает дальнейшую разработку новой технологии анализа данных – Data Mining (DM, интеллектуальный анализ данных).

Обратимся к предыстории возникновения информационного бума и связанными с ним «Большими Данными», а также вытекающим отсюда последствиям: необходимость совершенствования Data Mining и маркетинг-менеджмента в деловой активности общества [3]. В классическом представлении информационная среда – это совокупность технических и программных средств хранения, обработки и передачи информации (данных) для эффективного использования знаний в виде информационного ресурса. Активное внедрение информационных

технологий прослеживается во всех сферах общественной жизни. Появление и развитие информационной среды сделало возможным хранение, обработку, передачу и получение структурированной и неструктурируемой информации в быстрые сроки (это и есть стратегия БД). Если раньше мы имели дело с проблемой использования исчерпаемых природных ресурсов и в принципе демографически неограниченных трудовых ресурсов, то с развитием информационной среды мы сталкиваемся с появлением нового ценного ресурса – информации. Но поскольку внедрение информационной среды происходит очень быстрыми темпами, общий объем информационного потока (и данных) увеличивается в геометрической прогрессии, то возникает проблема сложности контроля над динамичностью и расширением информационной среды. Поскольку человеческое развитие не идет в ногу с развитием информационным, индивид сталкивается с понятием *информационной перегрузки*. Отныне человек не справляется с поиском нужного алгоритма поведения, не успевает перерабатывать поступающие данные, принимать верные решения и вынужден искать поддержку в информационных технологиях (Data Mining одна из них). Возникает проблема информационного стресса для человеческой психики и человеческой экологии в целом, как следствие. Внедрение всякого «новшества» без инструкций по его применению приводит к неправильной интерпретации информационной среды как таковой, поэтому кажется целесообразным поставить задачи поиска новых определенных характеристик информационной среды, анализа проблемы информационной перегрузки (в том числе и в связи с использованием БД-технологий), ее причин и следствий.

Для того чтобы информационная проблема приобрела некоторые очертания, необходимо обратиться к концепции технологических укладов, которая в той или иной мере способна объяснить наличие информационной перегрузки и обращения к БД- и DM-технологиям. Эта концепция, сейчас развиваемая российским экономистом Сергеем Глазьевым [4], является продолжением теории длинных волн Н. Д. Кондратьева. Известно, что Кондратьев, благодаря обработке экономических показателей наиболее развитых капиталистических стран (США, Великобритании, Франции и Германии) за продолжительный промежуток времени – с конца XVIII в. по 20-е годы XX века, эмпирически установил существование циклов капиталистического производства. Основной причиной таких циклов является появление новых технологий и производственных отраслей. В 90-е гг. XX в. смену волн Кондратьева объяснила концепция технологического уклада. Технологический уклад – совокупность связанных между собой технологий, характерных для определенного уровня развития производительных сил и качественно влияющих на производственные отношения. В связи с научным и технико-технологическим прогрессом происходит переход от более низких укладов к более высоким, прогрессивным (причем высокие включают в себя и более низкие уклады).

Сокращение периодичности кондратьевских циклов влечет за собой сокращение сроков технологических укладов. Наступает момент, когда сроки некоторых укладов становятся слишком короткими для возможной и быстрой общественной адаптации. Таким образом, основная причина «межукладных» кризисов, по этой концепции, в том, что общество с запаздыванием реагирует на объективно необходимую смену укладов. Проектируя образовавшуюся информационную среду на концепцию С. Ю. Глазьева, можно заметить, что ее зарождение совпадает с пятым технологическим укладом, где ведущие отрасли составляют программное обеспечение, вычислительная техника, телекоммуникации, информационные услуги. Проникновение продуктов информатизации в общественную жизнь в этом случае осуществляется настолько быстрыми темпами, что информационное развитие среды значительно превышает возможности человеческого развития. Пятый уклад не способен в полной мере реализоваться в обществе в качественно новых обстоятельствах образовавшейся информационной среды, а увеличение информационных потоков в такой среде не гарантирует прогрессивного функционирования общества. Поколениям, выросшим в одних условиях, трудно принять изменения и адаптироваться к иной среде, а невозможность адекватного понимания и адаптации общества к новшеству, безусловно, ведет к человеческому регрессу. Современный человек слишком буквально (утилитарно и прагматически) воспринял способ применения информационных технологий. Возможность получать готовые ответы, идти по пути наименьшего сопротивления «убивает» в человеке деятеля, что приводит к ограничению его интеллектуального потенциала. В результате конкретный индивид организован в гораздо менее разумной степени, чем это необходимо в действительности, и отныне не способен осознавать свои перспективы. Таким образом, первые поколения информационного общества обречены на социо-культурный кризис. Но общество должно пройти через разрушительную школу информационных технологий, с которыми ему надо научиться работать: человечество «обречено» приспособиться к нарастающим информационным потокам (возрождая с одной стороны проблематику искусственного интеллекта, а с другой – искусство результативного анализа данных и системного мышления), «обречено» на маркетинг-менеджмент своей деловой активности.

Процесс информатизации жизненной среды человека настолько быстр и масштабен, что проблема информационной перегрузки принимает глобальный характер. Но, несмотря на то, что информации много, только некоторая ее часть удовлетворяет человеческой потребности. Все остальное, для сфер производства и экономики – информационный мусор («*нон-информация*»). Таким образом, новая среда предоставляет для нас не только информационный ресурс, но и своего рода «отходы» информационной индустрии. «Засорение» информационной среды, своеобразный «информационный шум» требует разработки современных ДМ-технологий, способных успешно работать в условиях информационной перегрузки.

Слово «информация» так быстро внедрилось в бытовую и профессиональную лексику, что научное сообщество все еще не смогло обратить должного внимания на терминологическое определение понятия «информация». Проблема состоит, в том числе, и в определении ключевых различий между качественными информационными ресурсами и ресурсами «нон-информации». Не каждый имеющийся информационный поток предполагает непосредственный контакт с субъектом (пользователем). Со стороны пользователя необходима избирательность и сосредоточенность на определенном типе информационного ресурса, нужно уметь проводить своеобразную фильтрацию информационного потока.

Кроме того, качественная информация объективна (существует независимо от человеческого сознания), достоверна (отражает истинное положение дел), адекватна и актуальна (соответствует текущему моменту времени). Если рассматривать информацию как инновационный продукт, то новизна также является весомой качественной характеристикой, а все, что не привносит новизны в наше понимание сущности объекта, информацией, в заданном значении, уже не является. Только качественный информационный ресурс (соответствующий вышеприведенным и иным качественным критериям) может, поддаваясь обработке, хранению, передаче, послужить причиной синтеза новых информационных продуктов. Информация в таком понимании не поддается старению (не теряет своего качества), иначе она переходит в статус «нон-информации». Характеризуя современные информационные потоки, становится ясно, что с созданием новой среды жизни человека информации не стало больше. Вполне вероятно, что общий объем качественной информации на данный момент меньше, чем это необходимо для поколения (часть информации уже подверглась «старению», новая информация в связи с социо-культурным кризисом синтезируется медленно).

Неумение классифицировать составляющие информационного потока вынуждает человека хранить, обрабатывать и передавать «нон-информацию». В результате из некачественного «нон-ресурса» синтезируется новая «нон-информация». Таким образом, проанализировав проблему информационной перегрузки, мы обнаруживаем преимущественную проблему перегрузки «нон-информацией». Не связаны ли проблемы БД, спровоцированные технологическим прогрессом, с «нон-информацией»?

Если при более ранних технологических укладах общество сталкивалось с физической перегрузкой бесполезным трудом, то в новой среде мы имеем дело с психологической перегрузкой бесполезной информацией. Внедрение «нон-информации» в человеческое сознание приводит к декадансу ценностных ориентиров и неадекватному восприятию реальности. Экология человеческого существования находится под угрозой, так как разрыв между физическим, интеллектуальным и моральным на данный момент беспрецедентен. Информационная среда повела себя хитрее, чем мы ожидали. Удовлетворив наши потребности, она лишила нас интеллектуальной автономности. Но лояльность к происходящему и нежелание осознавать своих перспектив может привести к необратимым последствиям. Построенная на базе БД-технологий информационная система слишком

хорошо знает человека, обнаруживается проблема приватности, ибо большая часть информации сейчас уже генерируется персональными мобильными устройствами [5]. Человек постепенно становится рабом технологий и может утратить способность к самоуправлению. Становление новых технологических укладов при этом покажется немыслимой перспективой. Пессимистическая же перспектива известна: хронический западноевропейский индивидуализм и потеря автономности могут повлечь за собой крах всего человечества.

На данный момент «существует множество комбинаций программного и аппаратного обеспечения, которые позволяют создавать эффективные решения Big Data для различных бизнес дисциплин: от социальных медиа и мобильных приложений, до интеллектуального анализа и визуализации коммерческих данных». Совмещая новые инструменты с широко используемыми в бизнесе и государственном управлении базами данных, БД- и DM-технологии в союзе с *электронной наукой* (e-science) позволяют поддерживать муниципальную жизнь в крупных городах, вести контроль за эпидемиологической ситуацией и последствиями стихийных бедствий, отслеживать правоохранительную обстановку в регионе и многое другое [1; 5].

На какие трудности (здесь остановимся только на общенаучных и формально-логических) следует обратить внимание при разработке приемлемых технологий управления и анализа «Большими данными».

1. Так синергетический подход к анализу действительности требует внимательного обращения к «точечным» данным (а это пока еще сфера естественного, человеческого интеллекта): аттракторный, бифуркационный анализ «в точке» только и снимает неопределенность прогнозных показателей. «Техническое» решение видится в параллельном моделировании причинно-следственных связей «входа» и «выхода» в заданной точке, возможным ответом мог бы стать обнаруженный «резонанс» предварительных результатов анализа.

2. В «*нон-информационной*» модели анализа (а в «*Больших данных*»), конечно, преобладает «*нон-информация*») следует избегать формально-логических ошибок: экстраполяции неизвестного через неизвестное, нарушения закона тождества (особенно часто это делается в процессе интерпретации результатов), подмены оснований кластеризации и классификации и т. п.

3. Увеличение количества итераций и объема данных резко снижают качество получаемой информации, в этом случае на первый план выходят оптимизационные модели анализа. С другой стороны, уменьшение шага итерации приведет к проблеме гештальт-предела как минимального паттерна действительности, за которым целостность, качественность информации разрушается, а сама информация становится принципиально неструктурируемой.

4. В теории вероятностей закон больших чисел, при условии фиксированности распределения, предполагает наличие конечной выборки для получения удовлетворительных характеристик этого распределения. Однако «*нон-информация*» и соответствующие ей «*Большие данные*» ни ментально, ни вещественно

не фиксированы, т. е. в общем виде вероятностное пространство «Больших данных» не имеет известного нам вероятностного распределения, именно поэтому для снижения неопределенности результатов анализа используются DM-технологии: приходится создавать фиксированные «хранилища», проводить пилотную кластеризацию и ориентировочную классификацию данных.

5. Собственник информации и ее заказчик всегда опосредованы, то ли объективно-технологически (через облачные, сетевые и иные технологии), то ли субъективно-психосоциально (через потребности, возможности и долженствование). Коммуникация названных сторон требует согласования, посредничества, что предполагает иметь достаточное количество подготовленных специалистов – маркетологов-менеджеров, чтобы обеспечить необходимое качество такой коммуникации [3].

Литература

- [1]. Работа с Big Data: основные области и возможности [Электронный ресурс]. – Режим доступа: http://www.marketing.spb.ru/lib-research/methods/Big_Data.htm – Дата доступа: 23.04.2016.
- [2]. Черняк, Л. Seriously о технологиях для Больших Данных // Открытые системы. СУБД. – М.: Открытые системы, 2014. – № 1. – ISSN 1028-7493.
- [3]. Ермолович, Д. В. Маркетинг-менеджмент и философия бизнеса // Электронная экономика: теория, модели, технологии / Т.Н. Беляцкая [и др.]; под общ. ред. Т.Н. Беляцкой, Л.П. Князевой. – Минск: БГУИР, 2016. – С. 222-226.
- [4]. Глазьев, С. Ю. Современная теория длинных волн в развитии экономики // Экономическая наука современной России. – 2012. – № 2 (57). – С.8-27.
- [5]. Мусолеси, М. Большие Мобильные Данные: добро или зло? // Открытые системы. СУБД. – М.: Открытые системы, 2014. – № 4. – ISSN 1028-7493.