

И.В. ЛОПАТКИНА, канд.экон.наук, ГБУЗ «Украинская академия
банковского дела», Сумы

АНАЛИЗ БОЛЬШИХ ДАННЫХ – НОВЫЙ ВИТОК В РАЗВИТИИ ИННОВАЦИЙ, КОНКУРЕНЦИИ И ПРОДУКТИВНОСТИ НАЦИОНАЛЬНЫХ ЭКОНОМИК

В статті розглядаються питання інновацій у використанні сучасної інформації, зокрема питання використання великих даних. Сучасні технічні можливості, інноваційні технології стають економічним фактором підвищення конкурентоспроможності і продуктивності. Використання технічних інновацій стає ключом до інновацій в економіці. В статті розглядаються основні переваги та бар'єри аналізу великих даних. Визначені основні сфери, де застосування аналізу великих даних є найбільш вдалим. Робиться висновок про необхідність прийняття прогресивності аналізу великих даних всіма учасниками економічних процесів.

В статье рассматриваются вопросы инноваций в использовании современной информации, в частности вопрос использования больших данных. Современные технические возможности, инновационные технологии становятся экономическим фактором повышения конкурентоспособности и производительности. Использование технических инноваций становится ключом к инновациям в экономике. В статье рассматриваются основные преимущества и барьеры анализа больших данных. Определены основные сферы, где применение анализа больших данных является наиболее удачным. Делается вывод о необходимости восприятия прогрессивности анализа больших данных всеми участниками экономических процессов.

The article considers issues of innovation in use of modern information, specifically the issue of utilizing big data. Modern technological capacity, innovative technology become an economic factor of productivity and competition facilitation. The use of technological innovation becomes a key to innovation in economy. The article considers the main advantages and barriers to big data analysis. The main fields of big data analysis effective application are defined. The core conclusion of the article states that progressive analysis of big data should be embraced by all the participants of economic processes.

Ключевые слова: экономика, инновации, технологии, информатизация, сектора экономики, конкуренция, излишек потребителя.

Вступление. На протяжении всей истории существования человечества одной из наиболее сложных и неразрешимых проблем была и остается проблема хранения и использования информации. Разумная деятельность не одного поколения людей создала массу знаний, аккумулированных в информации, ценной уже потому, что она есть. Это количество информации и информационных источников, различных баз данных настолько огромно и, в то же время, разрозненно, что человек тонет в ней, не имея возможности эффективно использовать наработки предшествующих поколений. Актуальность этой проблемы выросла в разы в связи с информатизацией всех сторон жизнедеятельности человека, развитием компьютерных технологий и

появлением глобальной сети Интернета, что обусловило генерирование больших потоков информации и баз данных. Глобальный финансово-экономический кризис заставил по-новому взглянуть на существующую экономическую систему, выявив ее несовершенство. Человечество находится в состоянии постоянной борьбы за условия существования – экономические, экологические и социальные. При этом нерационально расходует ограниченные и невозобновляемые ресурсы. Использование инновационных методов складирования, хранения, обработки и использования информации позволяет получить конкурентное преимущество в экономическом соревновании стран и человечества в целом. Технические инновации сегодня становятся инновациями в экономике, фактором роста ее производительности, а, следовательно, и повышения конкурентоспособности национальных экономик. Важно не стать аутсайдером в этом процессе.

Постановка задачи: исходя из недостаточной изученности вопроса влияния методов использования больших данных на экономические процессы в Украине и наличия зарубежных исследований в этой области поставлена задача обосновать наличие зависимости между экономическими результатами и методами использования больших данных; изучить состояние и опыт использования больших данных в экономике; осветить инновационные методы использования больших данных.

Методология. В совокупность методов исследования проблемы корреляции вошли концептуальное моделирование, обобщение, анализ, синтез, индукция и дедукция, аналогия, рассуждение, умозаключение. Автор разделяет идеи риторики экономической науки Д. Макклоски ((McCloskey) и берет их за методологическую основу исследования. Социальный конструктивизм, как одно из новых направлений в экономической науке, также нашел свое отражение в данной работе.

Результаты исследования. Вопрос использования больших данных в данный момент рассматривается в основном западными специалистами в силу наличия доступа к эмпирическим методам исследования. Профессора Роджер Бон (Bohn Roger) и Джеймс Шорт (James Short) из Калифорнийского университета Сан Диего в своих работах 2009-2010 годов уделили много внимания накоплению информации среди корпораций и индивидуальных пользователей [6]. В своих работах они раскрыли масштабы накопления информации в глобальном масштабе, а также делают прогнозы по поводу темпов дальнейшего роста объема больших данных. В работе 2011 года Эрик Брынжолфссон (Brynjolfsson Eric) и Хит Лорин (Lorin M. Hitt) подчеркивают увеличение значимости решений, принятых на основании анализа данных в компаниях разных секторов экономики [8]. Особой активностью в вопросе исследования больших данных отличились специалисты McKinsey Global Institute (организация, ассоциированная с международной консалтинговой компанией МакКинси и Ко). В своих работах представители института Жак

Бугин (Bughin, Jaques), Майкл Чиу (Michael Chiu), Джеймс Маниука (James Manyika), Майкл Барбер (Barber, Michael and Lenny Mendoca), Алистер Леви (Alastair Levy), Тимо Кубах (Timo Kubach) и Маркус Лоффлер (Markus Loffler) детально исследуют разнообразные аспекты развития и использования больших данных [2; 4; 8; 9]. В частности, большую роль в своих исследованиях специалисты института уделяют потенциалу использования больших данных в государственном секторе и секторе розничной торговли.

Информация стала большим потоком, врывающимся в каждую область глобальной экономики. Компании генерируют огромные объемы транзакционной информации, охватывающей триллионы байт информации об их клиентах, поставщиках и операциях. Миллионы сетевых сенсоров внедряются в физический мир посредством таких устройств, как мобильные телефоны, разумные энергосберегающие счетчики, автомобили и промышленные машины, которые воспринимают, создают и передают информацию в век Интернета. Действительно, в то время, как компании и организации взаимодействуют с внешним миром, они генерируют огромный «выброс» информации, так называемый побочный продукт их основной деятельности. Социальные сети, смартфоны, и другие потребительские устройства, включая ПК и ноутбуки, позволили миллиардам людей по всему миру сделать свой вклад в доступный объем больших данных. И растущий объем мультимедийного контента сыграл решающую роль в экспоненциальном росте объема больших данных.

Что же мы подразумеваем под «большими данными»: к большим данным относится набор информации, размер которой превышает объемы типичных баз данных, обрабатываемых знакомыми нам программами, анализируемых, управляемых и сохраняемых стандартными процедурами. Эта трактовка намеренно оставляет субъективизм в определении размера больших данных. Большие данные не определяют как такие, которые превышают определенное количество байт, например, терабайт. Предположительно, с развитием технологий будут трансформироваться и критерии определения больших данных. Также следует учитывать, что определение может варьироваться от одного сектора экономики к другому, в зависимости от используемых программных средств.

Многие воспринимают аккумуляцию информации с определенной подозрительностью, видя в бесконечной информатизации своего рода вторжение в личную жизнь. Однако, наряду с этими страхами существует вполне обоснованное мнение о том, что использование больших данных может принести существенный положительный экономический эффект не только на индивидуальном коммерческом уровне, но и на государственном уровне. Многочисленные исследования показывают, что данные могут создать существенную ценность для мировой экономики, увеличивая продуктивность и конкурентоспособность компаний и государственного сектора, а также создавая значительный экономический излишек для потребителей. В одном из исследований [6] был приведен пример положительного эффекта от больших данных в системе здравоохранения США. По предварительным оценкам,

эффект этот может превышать 300 млрд. долл. каждый год, две трети которых были бы в форме уменьшения национальных расходов в сфере здравоохранения на 8%. В сфере розничных продаж этот эффект может быть выражен в форме роста рентабельности на более чем 60%. Использование больших данных в полной мере потенциально может сократить операционные издержки европейских стран на 100 млрд. евро. Дополнительный эффект может быть получен от сокращения количества случаев мошенничества, ошибок и нарушений в сфере налогообложения.

Роль больших данных в системе факторов экономического роста стала достаточно часто обсуждаться в экономических кругах, не смотря на то, что еще несколько лет назад эта тема интересовала, в основном, специалистов сферы IT. Возможности хранить, агрегировать и совмещать информацию и позже использовать результаты этих процессов для дальнейшего анализа значительно увеличились с развитием «облачных» технологий и возросшей доступностью емких средств хранения информации. За менее чем 600 долл. можно приобрести жесткий диск, способный вместить всю мировую музыку. Программные средства, используемые для получения наполненной смыслом информации из данных, также прогрессируют достаточно высокими темпами. Способность генерировать, сообщать, делиться информацией и получать доступ к ней значительно продвинулись в ходе популяризации электронных устройств и сенсоров, связанных в единой сети. В 2010 году более 4 млрд. человек или 60% мирового населения использовали мобильные телефоны и примерно 12% из них имели смартфоны (прирост использования которых составляет в среднем 20% в год) [1].

Многие инновационные компании уже используют большие данные для создания положительного эффекта, таким образом, вынуждая другие компании сектора искать похожие методы для сохранения конкурентоспособности. При этом, многим компаниям придется принять значительные изменения для того, чтобы в полной мере реализовать положительный эффект от использования больших данных. Недостаток аналитиков и кадров в сегменте менеджмента в сфере обработки больших данных создает значительные препятствия для извлечения пользы из доступной информации в краткосрочной и среднесрочной перспективе. Кроме всего прочего, одним из необходимых факторов для интеграции больших данных в систему драйверов роста продуктивности и эффективности в экономике является инфраструктура обработки данных. Обеспечение должной инфраструктуры достигается путем созданием условий для здоровой конкуренции, а также стимулированием инноваций во всех секторах экономики.

Несколько исследовательских команд занимались вопросом изучения объема данных, сгенерированных, хранимых и потребляемых во всем мире. Несмотря на отклонения в результатах исследований разных команд, все экономисты сходятся в одном – размеры данных в мировой экономике растут и будут продолжать расти экспоненциально [10]. По средним оценкам в мировом бизнес-секторе в 2010 году было сгенерировано 7 экзбайт новых данных, в то время как потребители хранят в среднем 6 экзбайт новой информации в год

[6]. Для более понятного осознания, 1 экзбайт более чем в 4000 раз превышает объем информации, аккумулированной Библиотекой Конгресса США. При этом большая часть информации утрачивается вскоре после создания за отсутствием достаточных физических емкостей для ее хранения. Многие компании в Европе и США уже могут позволить себе хранить, обрабатывать и, в конце концов, анализировать большие данные. В этом плане важным является концентрация информации среди корпораций и потребителей. В этом плане у развитых стран существует больше возможностей для аккумуляции достаточного объема данных (что само по себе является первым и основополагающим шагом для реализации последующих процессов) для создания описанного экономического эффекта. В перспективе, однако, существует большой потенциал для мультипликации эффекта от стремительного накопления и интеграции больших данных при благоприятных и правильных условиях. Азия, к примеру, занимает первое место по динамике роста геолокационных данных благодаря растущему использованию мобильных телефонов и смартфонов представителями окрепшего среднего класса. Количество используемых мобильных телефонов в Китае достигло 800 млн. в 2010 году.

Использование больших данных в экономике, как любая инновация, не лишено рисков. Прежде всего, для их предварительной оценки, необходимо ответить на ряд вопросов. Так, в мире, где крупномасштабные эксперименты стали возможными, как будет вести себя корпоративный маркетинг и как будут выстраиваться бизнес-модели? Как бизнесы будут воспринимать перемены и скорость внедрения этих перемен? Сможет ли доступ компании к информации и ее технологический потенциал для обработки этой информации представлять большую ценность, чем продвигаемый годами брэнд? Как будет проходить сегментация рынка и как будут функционировать конкурентные механизмы в среде абсолютной прозрачности данных? Сможет ли это вызвать определенную асимметрию в расстановке ролей, к которой мы привыкли? Например, как большие компании, выстроенные на ортодоксальной бизнес-модели, будут соревноваться с гибкими и мобильными небольшими компаниями нового образца, которые, в свою очередь, способны быстро обработать и воспользоваться детальной информацией о потребителе, которая становится все более и более доступной (например, через средства социальных медиа)? И, в конце концов, как отреагируют рыночные механизмы на смещение экономического излишка от поставщиков (продавцов) к потребителям, по мере того, как последние будут сами получать доступ к данным (как, например, сравнение цен на один и тот же товар одновременно среди тысяч поставщиков)?

Постараемся ответить на вопрос - что дает использование больших данных?

1. Создание прозрачности.

Обеспечение своевременного и легкого доступа к большим данным среди всех вовлеченных сторон может само по себе создать ценность. В государственном секторе, например, создавая доступ к нужной информации, которая была ранее недоступна для разных департаментов, можно значительно

сократить время на поиск и обработку информации. Позитивным примером реализации этой концепции на постсоветском пространстве является создание электронного правительства Эстонии. Электронный переворот власти Эстонии начинали с себя. Зал заседаний Кабинета министров уже более десяти лет самый интерактивный в мире. Доступ к этой программе есть у всех граждан. То есть каждый эстонец может принимать участие в заседании Кабмина.

В производстве интегрирование данных из отделов разработок, инженерского проектирования и самого производства может создать условия для совпадающего проектирования (concurrent engineering), что, в свою очередь, ощутимо сокращает цикл продвижения товара, а также улучшает качество продукции.

2. Создание условий для экспериментов с целью поиска нужд, диверсификации производства и улучшения показателей деятельности.

По мере создания и хранения все большего объема данных организации могут собирать более точные и детальные данные о ключевых производственных показателях, начиная с запасов готовой продукции, и кончая количеством больничных листов. Определенные программные продукты позволяют организациям проводить контролируемые эксперименты и симуляции с полученными данными и показывают разные сценарии взаимодействия этих данных с конечным исходом в реальных и финансовых показателях. Использование больших данных (не выборки) в анализе отклонений в производстве посредством контролируемых экспериментов или анализа реальных статистических данных позволяет определить причины происходящего намного быстрее, чем традиционные методы анализа (например, теория ограничений).

3. Сегментирование групп для индивидуализации действий.

Большие данные позволяют организациям создавать специфические сегментации и подстраивать продукты и сервисы точно под нужды каждого сегмента. Этот подход также известен в маркетинге и риск-менеджменте, но все же может быть революционным в других сферах – например, в государственном секторе, где принято относиться ко всем людям одинаково с макроэкономической точки зрения. Даже компании, производящие потребительские товары и сервисы, которые использовали сегментацию на протяжении множества лет, начинают разворачивать новые более сложные технологические подходы, которые, к примеру, позволяют реализовывать микросегментацию в реальном времени, с целью мгновенной реакции. Логистическая разработка (компьютерная программа) парковки автотранспорта в Сан-Франциско – один из практических примеров успешного использования больших данных с целью сегментирования групп для индивидуализации действий.

4. Замещение/поддержка принятия решений персоналом автоматизированными алгоритмами.

Сложная аналитика может значительно улучшить процесс принятия решений, минимизировать риски и дать путь интересным решениям, которые в ином случае могли бы оказаться потерянными. Такая аналитика могла бы, к

примеру, быть использована налоговыми службами для автоматического отбора объектов более тщательного анализа. Розничные сети могут использовать алгоритмы для оптимизации процесса принятия стандартизированных решений, как, например, адаптация запасов и цен в ответ на анализ продаж в реальном времени. Алгоритм не обязательно должен быть автономным. Напротив, его эффективность будет более высокой в комбинации с человеческим подходом. Однако использование алгоритмов позволит анализировать не просто выборки (которые могли анализировать неподготовленные специалисты), а полноценные массивы данных, таким образом позволяя охватить все значимые изменения и тенденции.

5. Инновации новых бизнес-моделей, продуктов и сервисов.

Большие данные позволяют компаниям создавать новые продукты и сервисы, улучшать существующие и создавать концептуально новые бизнес-модели. Производители используют данные, полученные в ходе использования реальных продуктов, для того, чтобы улучшить развития следующих поколений продуктов и для создания инновационных предложений после продажи. Появление геолокационных данных в реальном времени создало полностью новую сферу сервисов, основанных на местоположении индивидуума: навигация, оценка недвижимости, новые стоимостные модели в сфере страхования и т.д.

6. Использование больших данных станет базой конкуренции и роста индивидуальных фирм.

Использование больших данных становится ключевым способом для больших корпораций получить конкурентное преимущество перед своими соперниками. С другой стороны, неспособность компании адекватно воспринимать большие объемы данных может привести к потере позиций на рынке, где доминирует прозрачная информация.

Большие данные также помогают создать новые возможности для роста инновационных категорий компаний, как, например, компании по агрегированию и анализу данных в конкретной индустрии. Многие из них будут находиться посреди массивных информационных потоков. Участниками нового рынка могут стать компании, которые взаимодействуют с большим числом потребителей, покупающих, в свою очередь, широкий спектр товаров и сервисов, а также компании, являющиеся составными элементами глобальной цепочки поставок.

Некоторые из описанных возможностей создадут новые ресурсы для создания стоимости, другие обусловят существенный сдвиг в модели определения стоимости внутри секторов экономики. В этом процессе роль раннего входа также важна, как и в роль изобретения нового продукта. Те, кто получают доступ к репрезентативной информации, а, главное, к адекватным методам и инструментам ее обработки, получают конкурентное преимущество в своей сфере.

Как измерить ценность больших данных? Когда мы пытаемся определить потенциал больших данных в создании новой стоимости, мы рассматриваем только те действия, которые, по сути, зависят от использования больших

данных, например, действия, в которых использование больших данных необходимо (а не только достаточно) для получения дополнительного мультипликативного эффекта. В данном случае неверно в качестве мультипликативных факторов рассматривать только автоматизацию без использования полного анализа данных (как, например, замена банковских операционистов банкоматами).

Использование больших данных обусловит новую волну роста продуктивности и потребительского излишка.

В пяти основных областях применения, рассматриваемых далее, мы определяем несколько мультипликативных факторов, которые обусловят рост продуктивности. Эти возможности имеют потенциал улучшить показатели эффективности, позволяя организациям производить больше с помощью меньших ресурсов и производить лучше и более качественно. Так, например, компании могут использовать большие данные для настройки дизайна товаров под конкретные узкие нужды потребителей. Данные могут быть использованы для улучшения продуктов уже по мере их использования. Примером такого эффекта может быть мобильный телефон, который постепенно изучает повадки и предпочтения своего владельца и посредством установленных на нем приложений настраивает свой функционал в соответствии с заданными потребностями, и становится, таким образом, более ценным, чем другие устройства с общим набором функций. Для охвата этого потенциала понадобятся инновации во всех операционных процессах.

Стоит обратить внимание также на то, как потребители и граждане страны поглощают экономический излишек, полученный от использования больших данных. Они, по сути, являются прямыми и косвенными бенефициарами инноваций, обусловленных использованием больших данных. Население может ощутить заметный эффект от применения индивидуализированного подхода в предоставлении государственных услуг, прозрачности рыночных цен, а также более тесной связи между нуждами потребителей и продуктами, призванными эти нужды удовлетворять. Во многих развитых странах использование геолокационных данных уже позволило руководству решить множественные проблемы, как, например, заторы на дорогах (что, в свою очередь, тормозит оборот товаров и динамику экономической деятельности в целом), а также позволило значительно улучшить муниципальное транспортное сообщение. В данном случае эффект от использования больших данных вполне оценим количественно в виде экономии на топливе, увеличения скорости товарооборота, снижения выбросов в атмосферу во время пробок, времени, потраченного на путь из одного пункта в другой. Кроме всего прочего по результатам исследований западных ученых позитивный эффект в американской системе здравоохранения может достичь 300 млрд. долл. в год (что составляет 0,7% в годовом росте продуктивности). В сфере розничной торговли этот эффект выражается в потенциальном росте рентабельности продаж на 60% (или рост продуктивности на 0,5-1%). В сфере производства положительный эффект выражается в сокращении расходов на разработку

товаров на 50%. (по данным исследования McKinsey Global Institute) [2; 4; 8; 9].

В то время как все секторы экономики могут выиграть от использования больших данных, некоторые секторы более расположены к поглощению положительного эффекта, чем другие. Очевидно, что индустрии, более зависимые от использования информации, по определению могут получить положительный эффект от инновационных способов ее использования. В то же время сектор финансов и государственный сектор также могут в достаточно высокой степени выиграть от использования больших данных, однако только в том случае, если барьеры для их использования будут полностью преодолены. Последнее далеко не всегда возможно в ряде развивающихся стран, в частности, в Украине. Такие сектора экономики, как сфера развлечений, образовательные услуги и строительство выигрывают от использования больших данных в наименьшей степени (а по результатам некоторых исследований испытывают даже негативных эффект) в силу низкой зависимости их операционной деятельности от больших массивов данных. Сектор транспортных услуг, производственный сектор и также сфера здравоохранения в среднесрочной перспективе будут способны выиграть от использования больших данных, однако без мультипликативного эффекта, по крайней мере, в кратко- и среднесрочной перспективе.

Кроме всего сказанного, следует обратить внимание на барьеры для использования больших данных. В некоторых секторах экономики они выше, чем в других. Государственный сектор (включающий в себя здравоохранение и образовательный сектор) испытывает недостаток данных и должного восприятия данных. Многие государственные службы (например, таможня) обладают доступом к большим данным, однако предпочитают использовать ее с общестатистической точки зрения, обобщая и группируя большие массивы данных, обеспечивая, таким образом, возможность анализировать эти данные без использования дополнительных технических средств. Учет данных в сфере здравоохранения во многих развивающихся странах является невозможным из-за технической отсталости отрасли и отсутствия традиций аккумулирования данных. Такие сектора экономики, как розничные продажи, производство и профессиональные сервисы могут иметь относительно низкие барьеры для использования информации по прямо противоположным причинам (результаты исследования McKinsey Global Institute Analysis).

По мере того, как большие объемы данных переходят в цифровой стандарт и перемещаются сквозь границы организаций, существуют меры, которые будут особенно важными, включая приватность, безопасность, интеллектуальную собственность и ответственность. Очевидно, приватность - это вопрос, важность которого для потребителей растет с каждым днем. Личные данные, такие, как состояние здоровья и финансовая информация могут принести наибольший положительный социальный и экономический эффект, выраженный, например, в более точном прогнозировании медицинского лечения или выборе наиболее подходящего финансового продукта. В тоже время потребители относятся достаточно ревностно к

внешнему доступу к этой информации. В обществе до сих пор существует конфликт в том, какой частью личной информации индивидум готов пожертвовать ради увеличения полезности.

Недавний мировой опыт показал, что в результате недостаточной защищенности данных достоянием общественности могут стать не только личная информация, но и информация государственной важности. В контексте совершенствования технологических средств, способных сделать несанкционированный доступ к информации намного проще, чем прежде, должны совершенствоваться и способы защиты информации. Эта необходимость может породить еще один достаточно динамично развивающийся сектор бизнеса, а, впоследствии, и экономики в целом. Ключевыми вопросами в контексте защиты информации будут: кому принадлежит та или иная информация, и какие права ассоциируются с определенными данными? Что определяем справедливое использование данных? Кто ответственен за негативный эффект обусловленный неправильным использованием информации? Ответы на эти вопросы потребуют глубокой вовлеченности специалистов сектора юриспруденции.

На пути распространения технологий анализа больших данных возникает ряд препятствий разной степени сложности их разрешения.

1. Технологии и техника.

Для того чтобы воспользоваться потенциалом, предоставляемым большими данными, компаниям придется адаптироваться к новым технологиям сбора, анализа и хранения данных. Спектр технологических сложностей и способов их решения будет зависеть во многом от зрелости информации в каждой отдельной организации. Старые архаичные системы и несовместимые стандарты и форматы очень часто тормозят интеграцию данных и более сложную аналитику, которая могла бы принести желаемый положительный эффект. Появление новых задач и растущие мощности современной вычислительной техники дадут серьезный толчок развитию инструментов для анализа больших данных.

2. Организационные изменения и наличие кадров.

Лидеры многих организаций часто лишены правильного понимания больших данных и возможных способов их интерпретации. Несмотря на кажущуюся незначительность этого фактора, имплементация аналитики больших данных во многих организациях может оказаться под вопросом именно из-за отсутствия вовлеченности в управленческих кругах. Кроме этого, даже в случае согласия на использование больших данных не каждая организация сможет наладить правильный процесс аккумуляции данных (правильных, нужных данных) и, тем более, адекватного анализа.

3. Доступ к информации.

Обеспечение трансформационных возможностей потребует от компаний интегрирование информации из множества разных ресурсов. Приобретение доступа к данным станет более популярным способом получения данных. Однако не во всех случаях владельцы информации хотят делиться этой информацией (которая, в свою очередь, может быть ключевой составляющей

совокупного массива больших данных). Поиск путей достижения консенсуса в этом вопросе, формирование взаимовыгодной системы ценностей при обмене данными может стать залогом преодоления этой преграды на пути свободного перемещения данных.

4. Структура индустрии.

Секторы с относительно неконкурентными и непрозрачными механизмами функционирования, а также секторы с высококонцентрированными пулами прибылей в последнюю очередь смогут воспользоваться преимуществами использования больших данных. Например, в государственном секторе не хватает конкурентного давления, которое вынуждало бы организации и их руководство подходить к подбору и обработке данных более эффективно. Высокий уровень конкуренции, вполне логично, является катализатором информационного обмена и более эффективного использования данных.

Стоит отдельно уделить внимание доступным техникам анализа больших данных. Множество способов анализа объединяют в себе статистические методы, компьютерные науки и математические модели. Приведенные ниже методы анализа составляют далеко не полный список доступных ныне методов. Кроме всего прочего, этот список постоянно пополняется за счет стремительного развития IT-индустрии.

A/B тестирование – техника, в которой контрольная группа сравнивается с разнообразными тестовыми группами с целью определения действий, которые приведут к желаемому результату. Этот метод также известен как сплит-тестирование. В случае если во время теста манипулируют более чем одной переменной, то модель называют A/B/N тестирование.

Классификация - набор методов, определяющих категории, к которым принадлежат отдельные элементы данных, основываясь на предварительном распределении других элементов. Одно из применений этого метода заключается в прогнозировании определенного сегментного потребительского поведения с четкой гипотезой объективного исхода.

Кластерный анализ - статистический метод классификации объектов, который разделяет неоднородную группу на более мелкие группы однородных объектов, чьи характеристики схожести не известны заранее. В этом случае анализ данных будет предполагать разделение потребителей на схожие по своему поведению группы для таргетированного маркетинга.

Регрессия – набор статистических методов, определяющих, как значение зависимой переменной изменяется при изменении независимой переменной. Метод чаще всего используется для прогнозирования.

Пространственный анализ – набор методов, в том числе и статистических, которые анализируют топологические, геометрические и/или географические свойства, закодированные в данных. Возможным вариантом применения этого метода в новой эре может быть внедрение пространственного анализа в регрессионный анализ (прогнозирование поведения экономических механизмов в зависимости от географического расположения) или симуляции (сценарии поведения логистических сетей в разных географических условиях).

Другие методы включают также визуализацию, симуляции, знаковый процессинг, анализ настроек, распознавание шаблонов и т.д.

Инструменты для анализа больших данных включают:

- Большие таблицы (разработка Google File System);
- Кассандра (Cassandra) - система управления базами данных, позволяющая оперировать достаточно большими объемами данных в распределенных системах. Изначально эта система была разработана в Фейсбуке (Facebook);
- «Облачные» технологии – парадигма обработки данных, подразумевающая предоставление масштабируемых вычислительных ресурсов через сеть;
- Дупато – собственная разработка компании Amazon;
- ETL (Extract, transform and load – программные инструменты, используемые для выведения данных из внешних ресурсов, трансформирования их в соответствии с операционными нуждами и выгрузка их в базы данных;
- Google File Systems;
- Hadoop;
- SQL – компьютерный язык, разработанный для управления данными в реляционных базах данных. Этот инструмент позволяет вставлять, запрашивать, обновлять и удалять данные, а также управлять структурой баз данных и контролировать доступ к данным.

Другие инструменты включают Mashup (Мэшап), MetaData, HBase, Business Intelligence.

Существует также множество методов визуализации информации, представленной большими данными. Такие методы, как кластерограмма, исторический поток (инструмент визуализации, отображающий вклад каждого отдельного фактора в общий результат с течением времени). Очень удобно иллюстрируется на примере Википедии, где вклад каждого отдельного автора в контент меняется со временем.

Еще одним примером такой визуализации является тэговое облако, которое помогает потребителям и бенефициарам больших данных интуитивно воспринимать значение этих данных.

Базой работы с большими данными является осознание различий между данными, информацией и инсайтом (от англ. Insight). Данные представляют собой набор знаков, цифровых, текстовых и графических, которые сами по себе составляют определенную совокупность, чаще всего объединенную по одному или ряду признаков. Информация, в свою очередь, является результатом интерпретации данных с помощью описанных нами инструментов и методов. Одни и те же данные могут сгенерировать достаточно отдаленную по смысловому наполнению информацию в зависимости от используемых инструментов и методов, а также субъектов, анализирующих данные и потребителей данных. Таким образом, информация является более обширным понятием, чем данные в данном контексте. Последним звеном в этой цепочке является инсайт, то есть способ применения полученной информации для получения положительного социального и экономического эффекта. По сути,

это соотношение информации и реальных процессов, с ней связанных. По нашему мнению достичь правильного инсайта будет сложнее всего, особенно в странах с отсутствующей культурой обработки больших данных. Отсутствие опытных кадров в сфере анализа и инструментов обработки больших данных сведет на нет весь потенциал от накопления самих данных. Подготовка специалистов в этой сфере не является процессом долгосрочным. Продолжительность подготовки может занять в среднем 2-3 года. Однако для старта этого процесса нужен существенный внешний импульс, скорее всего - государственное стимулирование. Специалистов нужно готовить заранее ввиду того, что определенный уровень подготовки требуется уже на этапе сбора и аккумуляции информации, а также ее правильного хранения.

Выводы. Многочисленные исследования говорят о том, что стадия информационной экономики не является последним этапом в экономическом развитии страны. Использование обычных данных имеет ограниченный потенциал в долгосрочной перспективе. Для обеспечения нового витка экономического роста и роста эффективности компаниям и государствам необходимо будет использовать большие данные, накопленные за предыдущие годы. Использование всех данных целиком, а не отдельных статистических выборок позволит кардинально поменять многие процессы в бизнесе и государственном секторе. Однако адаптация к использованию новых моделей обработки и анализа данных потребует от всех участников (в том числе и потребителей информации) больших финансовых и временных инвестиций. Процесс интеграции технологий и больших данных в существующие процессы должен проходить согласованно. Использование новых технологий обработки данных в одной отрасли изолированно существенно сокращает (или сводит на нет) мультипликационный эффект, описанный ранее. В ходе популяризации больших данных, как экономического актива, больше всего смогут выиграть такие секторы как здравоохранение, сектор транспортных услуг, розничная торговля, государственный сектор (в части эффективности административной деятельности и бюджетного планирования), и большие производители. Остается надеяться на то, что потенциал использования больших данных будет в равной степени признан участниками как развитых экономик, так и развивающихся.

Список литературы: 1. *Abele, Eberhard, Tobias, Ulrich Naher, Gernot Stube, and Richard Sykes*, eds. *Global production: A handbook for strategy and implementation*. - Berlin: Springer, 2008. 2. *Akella, Janaki, Timo Kubach, Markus Loffler, and Uwe Schmid*. *Data-driven management: Bringing more science into management*. - McKinsey Technology Initiative Perspective, 2008. 3. *Ayes, Ian*, *Super crunchers: Why thinking-by-numbers is the new way to be smart*. -New York: Bantam Books, 2007. 4. *Barber, Michael, Alastair Levy, and Lenny Mendoca*. *Global trends affecting the public sector*. - McKinsey white paper, February 2008. 5. *Bohn, Roger, and James Short*. *How much information? 2009: Report on American consumers*, University of California, San Diego, Global Information Industry Center, January 2010. 6. *Bohn, Roger, James Short, and Chaitanya Baru*. *How much information? 2010: Report on enterprise server information*, University of California, San Diego, Global Information Industry Center, January 2011. 7. *Bollier, David*. *The*

promise and peril of big data. - Aspen Institute, 2010. **8.** *Brynjolfsson, Eric, Lorin M. Hitt, and Heekyung Hellen Kim.* Strength in numbers: How does data-driven decisionmaking affect firm performance? - McKinsey Quarterly. - April 22, 2011. **9.** *Bughin, Jaques, Michael Chiu, and James Manyika.* “Clouds, big data, and smart assets: Ten tech-enabled trends to watch,” McKinsey Quarterly, August 2010. **10.** *Dunleavy, P.* Innovating out of a recession. – Seminar in London School of Economics. - London, June 22, 2009.

Поступила до редакції 05.04.2012р.