

РОЛЬ «БОЛЬШИХ ДАННЫХ» В РАЗВИТИИ БИЗНЕСА В ЭРУ ЦИФРОВОЙ ЭКОНОМИКИ

Актальчиков Р.А.

Оренбургский государственный университет

Как показывает опыт ведущих стран, цифровые технологии могут быть преобразовательными для развития страны, генерируя экономические и социальные выгоды для людей, бизнеса и правительства. Цифровые технологии в настоящее время предоставляют возможности для всестороннего и устойчивого экономического роста во всех секторах экономики.

Важно, чтобы страны предпринимали структурированные усилия по созданию и использованию преимуществ цифровой экономики, с тем чтобы обеспечить более эффективное создание рабочих мест, повысить конкурентоспособность страны, обеспечить большую диверсификацию и стимулировать инновации в сфере предоставления услуг для улучшения жизни граждан.

Российская Федерация добилась значительного прогресса в использовании цифровой революции для решения своих задач в области развития. Страна уже готова к тому, чтобы в полной мере использовать возможности, предлагаемые передовыми цифровыми технологиями. Приспосабливая и определяя цифровую экономику в контексте, специфичном для Российской Федерации, и, используя уже прочные технологические основы, Россия может создать уникальную нишу для себя и позиционировать себя как нового мирового лидера в области цифровой экономики.

Концепция цифровой экономики постоянно развивается из-за ее многогранности и динамичности и из-за трансформационной способности цифровых технологий. Например, австралийское правительство определяет цифровую экономику как «глобальную сеть экономических и социальных мероприятий, которая обеспечивается такими платформами, как Интернет, мобильные и сенсорные сети».

Цифровая экономика определяется Оксфордским словарем как «экономика, которая функционирует в основном с помощью цифровых технологий, особенно электронных транзакций, совершаемых с использованием Интернета». Цифровая экономика иногда называется интернет-экономикой, новой экономикой или веб-экономикой. Он часто воспринимается как ведение бизнеса через рынки, основанные на Интернете и Всемирной паутине.

Европейский союз рассматривает цифровую экономику как «единственный важнейший фактор инноваций, конкурентоспособности и роста в мире».

Недавние исследования подчеркивают распространение цифровой экономики во всей экономике и утверждают, что «она больше не может быть описана как отдельная часть или подмножество основной экономики».

Это выходит за рамки электронной коммерции и электронного бизнеса и включает в себя ведение бизнеса, организацию коммуникаций и предоставление услуг во всех секторах, включая транспорт, финансовые услуги, производство, образование, здравоохранение, сельское хозяйство, розничную торговлю, средства массовой информации, развлечения и бизнес с использованием цифровых технологий.

Цифровая экономика играет важную роль в ускорении глобального экономического развития, повышении производительности существующих отраслей промышленности, культивировании новых рынков и отраслей и достижение всеобъемлющего и устойчивого роста.

Основным двигателем развития цифровой экономики на текущий момент являются так называемые большие данные, или big data, и, как следствие data science, - так в настоящее время называется применение статистики и машинного обучения к бизнес-задачам.

Анализ больших данных и применение результатов этого анализа для принятия решений играют ключевую роль в определении степени конкурентоспособности как отдельного бизнеса, так и страны на международной арене.

Различие больших данных от «обычных данных», которые все мы анализировали ранее, заключается в том, что инструменты, которые используются для сбора, хранения и анализа, должны изменяться, чтобы учесть увеличение размера данных и их сложности. Благодаря новейшим инструментам на рынке больше не нужно полагаться на выборку. Вместо этого возможно полностью обрабатывать наборы данных и получать гораздо более полную картину окружающего нас мира.

Универсальная истина в международном бизнесе сегодня заключается в том, что все дороги ведут к получению необходимых данных. Во все более сложном и взаимосвязанном мире способность организации эффективно собирать, управлять и анализировать огромные объемы данных отделяет её как победителя от занявших второе место или даже проигравших.

На текущий момент в мире стал общепризнанным факт того, что данные, которыми располагают организации, становятся основным компонентом рыночной стоимости этих организаций.

Недавний отчет Capgemini, который опросил более 1000 компаний, показал, что 65% респондентов считают, что их организации рискуют стать неконкурентоспособными из-за высококонкурентного ландшафта данных. При этом оставшиеся 35% компаний указали, что уже представили или находятся в процессе внедрения позиций нового уровня для управления данными.

Дальнейшие выводы из отчета показывают:

- 54% компаний заявили, что их инвестиции в большие данные в течение следующих трех лет обойдут все прошлые инвестиции вместе взятые.
- 61% признают, что большие данные в настоящее время являются движущей силой доходов и становятся столь же ценным для своего бизнеса, как и их существующие продукты и услуги.

- 43% организаций уже провели или в настоящее время проводят реструктуризацию, чтобы использовать новые возможности, связанные с большими данными.
- 36% респондентов заявили, что из-за стратегической важности больших данных им пришлось нанимать целые ИТ-команды для проведения необходимой аналитики данных, которая способствует получению важной бизнес-информации.
- 52% респондентов сообщили о том, что развитию быстрого доступа к данным препятствовали ограничения в процессе разработки ИТ сети в компании.
- Быстрый доступ к большим данным – то, где руководители видят наибольшую ценность, – 77% заявляют, что лица, принимающие решения, все чаще требуют потока данных в режиме реального времени.

Поэтому данные (и статистика) жизненно важны, так как они дают нам четкие, объективные, количественные оценки по всем аспектам нашей жизни и состоянию нашего бизнеса или ситуации в стране.

Наиболее распространенные способы использования больших данных компаниями – отслеживание бизнес-процессов и результатов деятельности, а также построение широкого спектра интеллектуальных предиктивных моделей. Хотя бизнес-аналитика, несомненно, улучшила эффективность многих организаций, однако предиктивное моделирование лежит за множеством поразительных информационных продуктов и услуг, внедренных в последние годы многими компаниями.

Некоторые примеры будут знакомы всем. Рекомендации Amazon и Netflix опираются на прогностические модели того, какую книгу или фильм может захотеть приобрести человек. Результаты поиска и новостей Google основаны на алгоритмах, которые предсказывают релевантность определенных веб-страниц или статей. Функция автозаполнения Apple пытается предсказать остальную часть своего текста или электронной почты на основе прошлых шаблонов использования. Онлайн-реклама и маркетинг в значительной степени зависят от автоматизированных прогнозирующих моделей, ориентированных на людей, которые могут быть особенно склонны реагировать на предложения.

Применение интеллектуальных алгоритмов выходит далеко за пределы онлайн-мира. В сфере здравоохранения страховщикам часто приходится корректировать платежи, основанные на «показателях риска», которые получены на основе прогнозирующих моделей отдельных факторов. Банки, выпускающие кредитные карты, используют прогнозные модели для оценки наступления дефолта или погашения, чтобы правильно проводить андеррайтинг, ценообразование и маркетинговую деятельность. Одна компания в Пало Альто, «Palantir», стала многомиллиардным бизнесом, разработав алгоритмы, которые могут использоваться для идентификации террористических угроз с использованием коммуникаций и других данных, а

также для выявления мошеннических действий в сфере здравоохранения и финансовых услуг.

В применении на практике, эти приложения полагаются на преобразование большого количества неструктурированных данных в прогнозные оценки, часто полностью автоматизированным и масштабируемым способом, а иногда и в режиме реального времени. Затем оценки можно использовать различными способами. Во-первых, они могут ускорить или автоматизировать существующие процессы: Amazon рекомендует товары, которые, по его прогнозам, могут быть релевантны для данного потребителя или ситуации, заменив тем самым рекомендацию, которую ранее можно было получить, например, у онлайн-консультанта. Во-вторых, они могут быть использованы для предоставления нового набора услуг: Apple принимает слово или предложение с высокой балльной оценкой и предлагает его как автоматическое завершение текста или сообщения. Наконец, оценки могут использоваться в области принятия решений. Например, в контексте мошенничества с кредитными картами банки могут осуществлять политику, которая диктует, какие транзакции одобрены, какие отклоняются, и какие требуют дальнейшего расследования, основываясь на балльной оценке транзакций.

Проделана значительная работа по статистическим и машинным методам обучения, которые лежат в основе этих приложений, такие как регрессии Лассо и Ридж и классифицирующие модели. Эти методы в настоящее время распространены в области статистики и информатики, но они редко используются в эмпирической микроэкономике. Хотя подробное описание методов будет далеко не ограничиваться рамками настоящего доклада, краткий концептуальный обзор не окажется лишним.

Задачу прогностического моделирования можно описать, представив N записей, которые связаны с N результатами измерений, а также набор K потенциальных предикторов. Во многих случаях информация о каждой записи велика и неструктурирована, поэтому существует много возможных предикторов, которые могут быть сгенерированы. Действительно, число потенциальных предикторов K может быть больше, чем число наблюдений N . Очевидной проблемой является чрезмерная переоценка: при $K > N$, как правило, можно будет прекрасно объяснить наблюдаемые результаты, но производительность вне выборки может быть низкой.

В этой задаче цель состоит в построении статистической модели, которая максимизирует прогностическую мощность в выборке, но в то же время не «подгоняется» таким образом, чтобы приводить к плохой производительности вне выборки. Различные методы различаются в зависимости от того, как построена модель и как ограничивается чрезмерное использование потенциальных предикторов. Например, регрессия Лассо выбирает коэффициенты для минимизации суммы квадратов отклонений при условии ограничения суммы абсолютных значений коэффициентов. Также принято оценивать компромисс между прогнозируемой мощностью и выборочным

анализом в выборке путем разделения выборки на «образец обучения», используемый для оценки параметров модели, и «тестовый образец», используемый для оценки результативности модели. Такой подход к перекрестной проверке редко используется в эмпирической микроэкономике.

Важнейшим, но часто подразумеваемым предположением в компьютерном обучении является то, что изучаемая среда относительно стабильна в том смысле, что образец оценки (как обучающие, так и тестовые образцы) генерируется одними и теми же независимыми испытаниями, которые позже будут генерировать образец, для которого требуется прогнозирование. Конечно, среда меняется со временем, поэтому это не идеальное предположение. В приложениях, где новые данные становятся доступными на высокой частоте, алгоритм также может «переобучиться» сам по себе, а также корректировать прогнозируемую модель с течением времени по мере изменения среды.

Когда экономисты считают полезность методов машинного обучения, то, что приходит на ум в первую очередь, часто является некой версией «Критики Лукаса». Критика Лукаса, названная в его честь при исследованиях разработки макроэкономической политики, основывается на утверждении о наивности попыток предсказания последствий изменения экономической политики только на основе взаимоотношений показателей в имеющихся исторических данных, особенно сильно агрегированных исторических данных. Если модель прогнозирования используется для принятия решения о вмешательстве в политику, результат может быть не таким, как предсказывает модель, поскольку изменение политики может повлиять на базовое поведение, порождающее отношения в данных. То есть модели являются «прогностическими», а не «структурными». Конечно, это не делает прогностическую модель бесполезной, потому что степень критики во многом зависит от ситуации. Например, возможно, что некоторые покупатели Amazon поняли, как создаются рекомендации, и изменили своё покупательское поведение, чтобы получить разные рекомендации. Но, вероятно, большинство из них не изменили своего поведения. С другой стороны, если Amazon начала предлагать большие таргетированные скидки с использованием подобной модели, они могут вызвать гораздо большее изменение поведения покупателей.

Перед экономистами, желающими воспользоваться большими новыми наборами данных, стоят несколько проблем. К ним относятся получение доступа к данным, разработка возможностей управления данными и программирования, необходимых для работы с крупномасштабными наборами данных, и, наконец, творческое мышление о подходах, выбираемых для обобщения, описания и анализа информации, содержащейся в этих данных.

Доступ к данным. Возьмем в качестве примера США. Исследования по таким темам, как экономика труда, производительность и потребление домашних хозяйств, традиционно основывались на данных правительственных опросов, таких как перепись населения США, групповое исследование динамики доходов (PSID) и национальное продольное исследование молодежи

(NLSY). Для многих из этих данных существуют хорошо установленные протоколы для доступа к ним и их использования. В некоторых случаях, например, в Центрах исследований данных переписей США, эти протоколы являются громоздкими и сложными, и, вероятно, препятствуют набору справедливого числа исследователей, но, по крайней мере, они отражают сознательные усилия по обмену информацией между доступом к исследованиям и проблемами конфиденциальности.

Многие европейские страны, такие как Норвегия, Швеция и Дания, значительно продвинулись вперед в содействии исследованиям. Опыт в этих странах свидетельствует о том, что возможен более широкий доступ и что снижение барьеров на пути доступа к данным может оказать глубокое влияние на объем и качество исследований.

Большая часть новых данных, которые мы обсуждали, принадлежит частным компаниям. Доступ к данным частных компаний создает несколько проблем для исследователей. Во-первых, и, что очевидно, не каждая компания хочет работать с исследователями. Хотя многие считают это потенциально выгодным и полезным способом учиться у посторонних, другие могут рассматривать его как отвлечение фокуса компании от основной деятельности или сосредоточение внимания на рисках для рекламы. Исследователи, которые сотрудничают с компаниями, как правило, должны заключать контракты, чтобы предотвратить раскрытие конфиденциальной информации, и могут столкнуться с некоторыми ограничениями по вопросам, которые они могут изучать. Но эксперты говорят, что их преимущества заключаются в том, что выгоды от работы с данными компании в целом значительно перевешивают затраты, но для обеспечения успешного сотрудничества все же требуются значительные усилия от обеих сторон.

Управление данными и вычисления. Один из способов, которым некоторые исследователи определили категорию больших данных, состоит в том, что для обработки таких данных требуются значительные затраты времени и ресурсов. Практически все успешные интернет-компании и все больше и больше компаний с интенсивными потоками информации в других секторах инвестировали огромные ресурсы в хранение данных и распределенную обработку данных. Они также инвестировали в опытных программистов и инженеров, именуемых сейчас Data Scientists. Интересно, что даже когда эти компании нанимают таких ученых, чья работа состоит в том, чтобы анализировать данные для поиска эмпирических моделей, они обычно ищут людей, обучающихся в области информатики, а не эконометрики. Именно поэтому ожидается, что будущие экономисты, которые хотят работать с большими наборами данных, должны будут владеть по крайней мере некоторыми из современных инструментов исследователей данных, например, SQL и R и стандартные алгоритмы машинного обучения, упомянутые ранее.

Несомненно, в течение следующих десятилетий большие данные изменят ландшафт экономической политики и экономических исследований. Необходимо подчеркнуть, что большие данные не смогут кардинально

повлиять на развитие экономической теории или отменить необходимость тщательных экономических исследований. Скорее, они будут дополнять их. Как именно преобразятся экономические исследования и профессия экономиста в будущем – пока сложно предположить.

Список литературы

- 1. Хасте Т., Тибширани Р., Фридман Дж. Элементы статистического обучения, интеллектуального анализа и прогнозирования данных / Стэнфорд, Калифорния, 2008 г. – 764 с.*
- 2. Эйнав Л., Левин Дж. Революция данных и экономический анализ / Стэнфорд, Калифорния, 2014 г. – 24 с.*