

BIG DATA В ПРОГНОЗНОЙ АНАЛИТИКЕ БАНКОВ

Олег Игоревич КЛЮЧНИКОВ¹, к.э.н, доцент

¹Кафедра банковского бизнеса и инновационных финансовых технологий,
Автономная некоммерческая организация высшего образования «Международный
банковский институт имени Анатолия Собчака», Санкт-Петербург, Россия
Адрес для корреспонденции: О.И. Ключников, 191011, Невский пр., 60,
Санкт-Петербург, Россия
Тел.: +79219549889. E-mail: okey003@mail.ru

Аннотация

В статье рассматриваются перспективы, связанные с внедрением в банковское дело больших данных. В концептуальном и хозяйственном плана феномен «большие данные» изменил структуру банковской отрасли, привел к появлению новых свойств и раздвинул границы банков.

Любой информации, включая цифры, слова, графики, таблицы, можно доверять только после тщательного ее изучения. Большие данные можно представить как серию подходов, инструментов и методов для обработки структурированных и неструктурированных данных большого объема в их разнообразии для получения результатов, которые необходимы для принятия решений. Большие данные являются альтернативой традиционным базам данных, на основе которых принимались решения. Технология больших данных играет важную роль в развитии банковского дела. Информация стала ценным активом, своего рода новой нефтью, которая движет информационным обществом точно так же, как традиционная нефть была основным ресурсом в эпоху индустриального развития. Технологии больших данных жизненно важны для управления активами, оценки рисков, удержания и расширения клиентской базы.

Ключевые слова

Большие данные, интеллектуальный анализ данных, банковское дело, прогнозное моделирование.

BIG DATA IN PREDICTIVE ANALYTICS OF BANKS

Oleg Igorevich KLIUCHNIKOV¹, PhD, associate professor

¹Department of Banking and innovative financial technologies,
Autonomous non-profit organization
of higher education «International banking Institute named after Anatoliy Sobchak»,
Saint Petersburg, Russia
Address for correspondence: O. I. Kliuchnikov, 60 Nevsky Ave., 191011,
Saint Petersburg, Russia
Tel.: +79219549889. E-mail: okey003@mail.ru

Abstract

The article discusses the prospects associated with the implementation of big data in banking. Conceptually, the phenomenon of «big data» has changed the structure of the banking industry, led to the emergence of new properties, and pushed the boundaries of banks.

Any information, including numbers, words, graphs, tables, can be trusted only after careful study. Big data can be thought of as a series of approaches, tools, and techniques for processing large-scale structured and unstructured data in all its varieties in order to obtain the results necessary for decision-making. Big data is an alternative to traditional databases for decision-making. Big data technology plays an important role in the development of banking. Information has become a valuable asset, a kind of new oil that drives the information society, just as traditional oil was the main resource in the era of industrial development. Big data technologies are vital for asset management, risk assessment, and customer retention and expansion.

Keywords

Big data, data mining, banking, predictive modelling.

Введение

В банковском деле масштабы и величина всегда определяли перспективы развития. Появление «Больших данных» формально укрепило данную тенденцию. Тем не менее переход на большие данные и соответствующие им технологии (в том числе, облачные базы данных, высокоскоростная алгоритмическая торговля и технологии распределенных реестров) нарушили традиционную бизнес-модель развития и изменили подходы к масштабированию деятельности. Многие новации, в основе которых большие данные, позволили перейти к децентрализации финансовых участников, подключить к банковскому обслуживанию более широкие слои населения и усилили тенденции к демополизации финансового рынка. Банковская отрасль является ярким примером того, как технологии произвели революцию в сфере обслуживания клиентов и во многом изменили образ жизни и отношения населения к банкам.

Большие данные оказали важное воздействие на продуктовую линейку и услуги банков. С одной стороны, они существенно расширили предложение и оказали воздействие на его формы (в основном в направлении онлайн-цифрового взаимодействия с клиентами), а с другой стороны, открыли банки для широкой аудитории и новых участников. В целом большие данные изменили общий банковский ландшафт, а также условия стабильности и устойчивости банковской системы.

Данные служат источником для принятия решений и «сырьем» для отчетности. Без высококачественных данных, дающих достоверную информацию о процессах и явлениях в нужное время, разработка моделей поведения, мониторинг и оценка эффективности деятельности в различных

сферах становятся практически невозможными. Банковская сфера является одним из крупнейших потребителей и одновременно источников больших данных. Оцифровка ручных отчетов и процессов вместе с использованием внешних данных, поступающих из разных источников, идет в контексте «знай своего клиента» [1]. Однако потенциал больших данных в банковской сфере значительно шире.

По мере создания новых наборов данных появляются дополнительные возможности. Огромные объемы финансовых данных могут помочь выявить возникающие риски, позволить участникам рынка и регулирующим органам увидеть и лучше понять финансовые сети и взаимосвязи, повысить финансовую стабильность, усилить защиту потребителей и расширить доступ к недостаточно обслуживаемым. Данные также могут повысить прозрачность финансовой системы для участников рынка, регулирующих органов и общественности. Данные могут также обеспечить близкий к реальному времени режим взаимодействия с клиентами и регулируемыми органами, а также прогнозировать рынки и разрабатывать возможные сценарии поведения конкретных клиентов.

Как обеспечить банки более доступными, достоверными и полезными данными, а также одновременно обезопасить от несанкционированного их использования и превратить их в материал для предсказаний, необходимых для оценки поведения рынков, регуляторов, конкурентов и клиентов? В настоящее время преобладают три основные иллюзии относительно цифровой непогрешимости: во-первых, цифры точные. Когда сомневаются, то возникает желание добавить цифры после десятичной точки для большей достоверности. Однако точность цифры обеспечивается достоверностью процесса, который ее генерирует, точностью механизма сбора и передачи данных и способа их использования. Во-вторых, цифры беспристрастны и объективны. На самом деле цифры могут быть такими же предвзятыми, как и слова. Однако числами гораздо легче скрыть тенденциозность и замаскировать ее под беспристрастность. В-третьих, цифры позволяют контролировать. Человек стремится к контролю и власти. В этом ему помогают цифры. Они создают иллюзию, что передача сигналов с помощью цифр и привязка их к нестабильному или неконтролируемому индикатору позволяет управлять процессом или явлением.

Цифрам, как и любой другой информации, можно доверять только после их тщательного изучения [2]. Происходящая в наши дни финансовая технологическая революция (Финтех) создает основу для перехода к

поступлению широкого потока данных в банки, масштабному их использованию и возможности проверки на надежность и достоверность [3].

Нередко складывается впечатление, что переход на большой объем данных, используемых в банках, повышает уровень информационного обеспечения и позволяет полнее контролировать ситуацию. Однако данные не равнозначны информации и бывают разными. Многие из них могут отвлечь от основного тренда и сформировать неправильное впечатление о поведении клиентов и рынка. Данные о действиях клиентов, партнеров, конкурентов и регуляторов имеют гораздо большее значение, чем слова и данные о намерениях. Например, реклама или оптимистичный пост в социальных сетях о той или иной компании не столь полезен, как данные о движении котировок ее акций, объемах продаж, задолженности. В последнее время складывается общая тенденция к объединению различных баз данных, включая данные социальных сетей и внутренних данных банка. На такой основе создаются «супербазы» данных, которые состоят из набора больших данных, имеющихся в различных системах, как внутренних, так и внешних по отношению к банкам. Поэтому главной задачей анализа данных является не отвлекающая их визуализация и построение различных статистических рядов, а преобразование данных в информацию, необходимую для решения конкретных задач, например, информацию для клиентов по новым формам обслуживания, кредитным соглашениям, исходя из реальных запросов, а также информацию о кредитоспособности клиентов и денежно-кредитных перспектив конкретных клиентов и рынков.

Данные могут обладать реальной добавленной стоимостью, если к ним имеется эксклюзивный доступ или банк обладает уникальным методом извлечения данных даже из общедоступного массива, а также их соответствующей обработки. В последнее время возрастает интерес банков к прогнозной аналитике и прогнозному моделированию. Данная сфера требует надежных наборов из различных источников больших массивов данных с использованием искусственного интеллекта.

Разработка темы

В настоящее время складывается новая междисциплинарная научно-практическая сфера, связанная с использованием больших данных искусственным интеллектом в текущем управлении [4] и прогножном моделировании [5], также финтехом в банковском деле [6]. В этой сфере целая масса направлений – от организации работы банковских рекомендательных

систем [7] до предложения банковским клиентам биржевой торговли [8], включая подключения к платформам алгоритмической торговли и введения инновационных финансовых продуктов и услуг [9].

Особо выделяется литература, в которой рассматривается ценообразование данных, формулируются новые подходы к большим данным в связи со сложностями больших временных рядов для формулировки исследовательских идей и бэк-тестирования в финансах [10]. Прогнозная бизнес-аналитика в банковской сфере является новой сферой. Для ее развития необходимы серьезные исследования [11]. В данной статье сделан определенный шаг в ее изучении.

Обзор эволюции интереса к концепции БД

Термин «большие данные» появился относительно недавно. До 2011 года его использование в Google Trends было незначительным; с 2011 по 2014 г. наблюдался экспоненциальный рост интереса к большим данным, а после 2015 г. происходил небольшой спад числа запросов. Тем не менее в наши дни интерес к этому термину сохраняется. В региональном плане наибольший интерес к этому термину наблюдается в Сингапуре, Индии, Гонконге и Тунисе, средний уровень интереса в США и Канаде, меньший – в Бразилии, Западной Европе и Китае (по убыванию), достаточно низкий – в странах СНГ, включая Россию.

В бизнес-аналитике возрастание интереса к большим данным происходило медленнее и более плавно; небольшой спад наблюдался лишь в конце 2020 года. Наиболее активно аналитикой больших данных и бизнес-аналитикой больших данных интересовались в Австралии, затем в Индии, США и Канаде, менее активно в Западной Европе (прежде всего в Великобритании). В других странах ее использование для этих целей было незначительным.

К анализу использования аналитики больших данных в финансах и, в частности, в банковской сфере приступили совсем недавно. В связи с этим появился целый ряд статей об использовании ресурса больших данных для различных видов деятельности [12]. За рубежом к активному анализу общих проблем, связанных с технологиями больших данных в банковской сфере, приступили в 2012–2014 гг. Во второй половине прошлого десятилетия перешли к решению уже узких проблем, связанных, например, с развитием рекомендательных систем и использованием ими огромной информации, которая ранее была недоступна банкам. Особенно интересовала роль больших

данных в пограничных сферах, например, использование БД в искусственном интеллекте при банковском обслуживании населения, в том числе и через социальные сети. В нашей стране к анализу больших данных в финансовой сфере приступили в 2017–2019 гг. В новом десятилетии возникла необходимость определить роль больших данных в решении конкретных финансовых задач, включая повышение точности рекомендаций в рекомендательных системах, организации работы чат-ботов и биржевой высокоскоростной торговле.

Что означает термин «большие данные»? Согласно общепринятому подходу, к большим данным относятся наборы данных, которые слишком велики или сложны для обработки традиционным прикладным программным обеспечением. Работа с большими данными включает их сбор, хранение, анализ, поиск, совместное использование и извлечение, повторное использование и хранение, а также передачу, визуализацию, реагирование на запросы, обновление, конфиденциальность и безопасность использования. Первоначально большие данные были связаны с тремя ключевыми понятиями, объединенными в концепцию «три V» (объем, разнообразие и скорость – в английском языке volume, variety, velocity). В дальнейшем данная концепция была расширена до пяти компонентов (включая разнообразие и ценность – variability, value). В 2017 г. было предложено применительно к финансовой сфере дополнить данную концепцию еще двумя понятиями – время и место [13].

С использованием механизмов больших данных появляется возможность работы не с конкретным набором данных, а с методами обработки всего набора данных, включая те данные, которые находятся в разных базах и формально не объединены. Обычно данные подразделяются на структурированные, слабо структурированные и неструктурированные данные.

Исходя из определения больших данных, можно сформулировать основные принципы работы с большими данными.

1. Горизонтальная масштабируемость показывает количество неограниченных данных: в этом отношении любая система, предполагающая обработку больших данных, должна быть масштабируемой.

2. Отказоустойчивость проявляется в горизонтальной масштабируемости: подразумевается, что в кластере может быть несколько машин, поэтому методы работы с большими данными должны учитывать возможность крупномасштабных сбоев и их способность выжить без значительных последствий.

3. Локальность данных – в больших распределенных системах, данные распределены по большому количеству машин.

Все современные инструменты для работы с большими данными так или иначе следуют этим трем принципам.

Модели аналитики данных

Принято выделять четыре типа аналитики данных, которые различаются уровнем сложности информации и степени участия человека:

1. Описание прошлых событий путем создания сводки исторических данных для последующего их анализа (большие ряды данных).

2. Диагностический анализ информации, посредством которого подготавливается понимание причин появления тех или иных событий. Для этого используются статистические методы анализа данных, посредством которых проводится их классификация, детализация и сравнение, с выявлением ключевых факторов, влияющих на результаты.

3. Предсказательная аналитика позволяет определять перспективные сценарии и вероятность их наступления. Для этого используются следующие методы: математическая статистика, моделирование, машинное обучение и интеллектуальная бизнес-аналитика (прогнозное моделирование).

4. Выработка решений. Все накопленные и обработанные данные анализируются, чтобы найти оптимальное решение для каждой конкретной ситуации.

Прогнозная аналитика (интеллектуального анализа данных) как часть технологии больших данных в банковском секторе

Выделяются следующие варианты использования технологий БД в банковской сфере с использованием технологии искусственного интеллекта:

1) Полное представление о клиентах с помощью их профилей (поведенческая модель клиента). Сегментация клиентов стала обычным явлением в индустрии финансовых услуг, поскольку она позволяет банкам и кредитным организациям классифицировать клиентов по четким демографическим, гендерным, социальным и другим категориям. Однако базовая сегментация недостаточно детализирована, чтобы банковские учреждения могли понять истинные желания и потребности клиентов [14]. Использование технологии больших данных в банковском секторе помогает вывести сегментацию на новый уровень за счет построения подробных профилей клиентов. Эти профили должны учитывать множество факторов:

- a) демографический, гендерный, социальный, семейный и т.п. состав клиентов;
- b) количество клиентских счетов;
- c) продукты и услуги, которыми в настоящее время пользуются клиенты;
- d) предложения, от которых клиенты отказывались в прошлом;
- e) набор продуктов и услуг, которые клиенты могут приобрести в будущем;
- f) основные события в жизни клиентов;
- g) отношения клиентов с другими клиентами;
- h) отношение клиентов к банку и индустрии финансовых услуг в целом;
- i) поведенческие модели;
- j) сервисные предпочтения и т. д.

2) Индивидуальный подход к каждому клиенту. Почти треть клиентов ожидает, что компании, с которыми они ведут бизнес, будут учитывать личные характеристики клиентов. На основе анализа данных за последний год было установлено, что почти треть потенциальных клиентов отказались от услуг из-за недостаточной персонализации предложений [15]. Если банк стремится расширить число надежных клиентов и повысить прибыль, то внедрение технологий интеллектуального анализа данных, в том числе прогнозной аналитики, является стратегической необходимостью.

3) Анализ покупательной активности. Почти все данные в банковской сфере генерируются либо в результате взаимодействия клиентов с торговыми и сервисными представителями, либо посредством транзакций. Хотя обе формы данных о клиентах имеют огромную ценность, данные, полученные в результате транзакций, дают банкам более четкое представление о покупательских привычках клиентов и со временем расширяют понимание моделей поведения [16].

4) Анализ кредитного риска. На основе анализа информации о кредитной истории существующих клиентов формируются профили надежных и ненадежных клиентов; банк может разделить заемщиков по группам риска, установить индивидуальные условия по размеру заемных средств, процентной ставке и сроку кредита.

5) Выявление возможностей для увеличения простых и перекрестных продаж. Статистически вероятность дополнительных продаж банковских продуктов существующему покупателю на 60%–70% выше, чем вероятность привлечения нового клиента, что дает банкам прекрасную возможность увеличить прибыль за счет перекрестных продаж [15]. Использование

инструментов интеллектуального анализа данных позволяет легко получить все необходимые данные для увеличения продаж.

6) Снижение риска мошенничества. Снижение риска мошенничества – одна из самых быстрорастущих сфер деятельности. В 2019 году мошенники совершили транзакции с розничными счетами на 5,7 млрд рублей, что является рекордным показателем по сравнению с предыдущим периодом [15]. Мониторинг структуры расходов клиентов и выявление необычного поведения – один из способов использования технологий больших данных. Банки могут предотвратить мошенничество и повысить уровень безопасности клиентов.

7) Оценка рентабельности бизнес-проектов. С помощью технологии больших данных можно анализировать значительный объем информации об инвестиционных проектах, их прибыльности, рисках, минимизируют субъективную оценку фактов человеком.

8) Оценка конкурентоспособности. Анализируется информация о банковских организациях, которые имеют похожие характеристики и параметры для представления иерархической структуры банковской отрасли.

За последние годы возрастает потенциальная ценность информации, а стоимость ее обработки снижается. Отличие действительно значимых данных от другой, не относящейся к делу информации, способствует более эффективному решению проблем и принятию эффективных решений [17]. Аналитика в реальном времени позволяет понимать проблему, сдерживающую развитие бизнеса, а интеллектуальная аналитика способствует подготовке оптимальных решений. Заметно лучшие результаты могут быть достигнуты путем интеграции интеллектуальной бизнес-аналитики в рабочий процесс банка.

Ни один из множества методов не является универсальным, у каждого метода есть свои плюсы и минусы, и правильный выбор зависит от целей и обстоятельств. Наука о данных делает этот процесс автоматизированным, более точным, эффективным и менее затратным.

Направления использования технологий больших данных для решения микро- и макрозадач

Центральный банк Российской Федерации использует искусственный интеллект и большие данные для точной корректировки денежно-кредитной политики и быстрого понимания проблем финансовых рынков. ЦБ РФ работает над рядом проектов – от анализа кассовых данных для оценки инфляции до машинного обучения в банковской отчетности и сбора макроданных, а также

потенциальных возможностей технологий распределенных реестров. Эти проекты направлены на повышение качества прогнозирования и обоснованности решений в области денежно-кредитной политики. Актуальность таких данных для политиков и инвесторов постоянно возрастает в контексте пандемии коронавируса, когда требуются немедленные решения, а официальная статистика отражает ситуацию со значительной задержкой.

Сбор макроданных представляет собой базовый уровень, который может характеризовать целый ряд проблем. В частности, он связан с анализом того, как индикаторы корректируются с течением времени по мере поступления новой информации, что будет способствовать повышению точности прогнозов для принятия решений по денежно-кредитной политике. В более долгосрочной перспективе планируется построение новостных индексов, которые будут использоваться для краткосрочного прогнозирования макроэкономических показателей. Перспективным направлением является работа над проектом «Онлайн-инфляция» (обработка данных с онлайн-кассовых аппаратов), который представляет собой группу алгоритмов автоматической классификации расширенного ассортимента товаров.

Сбербанк с помощью системы «Лабиринт» находит массовых учредителей. Эта система позволяет выявлять различные финансовые схемы, включая мошенничество, когда несколько компаний, казалось бы, полностью независимых друг от друга, берут кредиты, а затем исчезают или, взаимодействуя друг с другом, «прокручивают» и обналичивают кредитные средства. При этом связь между этими компаниями не очевидна, но при наличии дополнительных внешних данных может быть обнаружена.

У Сбербанка самое большое число клиентов, как физических, так и юридических лиц. Определенные цепочки можно отслеживать в транзакциях. Например, когда предприятия регулярно платят друг другу за услуги и фактически образуют холдинг, который де-юре не формализован. При алгоритмическом выявлении производственных цепочек группе компаний могут быть предложены специальные корпоративные продукты, подходящие для холдингов, что позволяет повысить эффективность кредитования. Информация может визуализироваться: графическое изображение компании, вокруг которой находится облако взаимосвязанных клиентов. Для решения проблемы рассматриваются контрагенты; выясняется, каким компаниям клиент предоставляет услуги, какие компании предоставляют услуги контрагенту. Графически это выглядит как клиент, вокруг которого расходятся лучи-

контрагенты. Анализ начинается, как правило, с крупных клиентов, а затем все цепочки расширяются.

Итак, основными направлениями развития использования технологий БД в банковском секторе являются следующие:

- совершенствование алгоритмов обнаружения мошеннических транзакций. Используя анализ стереотипов мошеннических и законных транзакций, система самообучения выдает предупреждение, которое позволяет банкам оперативно предотвращать подозрительные транзакции;
- оценка повышения уровня риска. Большие данные, подключенные к инструментам бизнес-аналитики с возможностями автоматического анализа и прогнозирования, могут сигнализировать о профилях клиентов с более высоким уровнем риска, чем другие;
- сохранение и удержание клиентов. Подробные профили клиентов упрощают построение более прочных и длительных отношений с клиентами;
- персонализация продукта. Демонстрация желания понять каждого отдельного клиента, разрабатывая и предлагая ему персонифицированные продукты и услуги;
- обратная связь с клиентами. Позволяет поддерживать взаимодействие между клиентом и банком и быть в курсе текущих проблем, что позволяет вовремя реагировать и регулировать взаимоотношения;
- улучшение рабочего места. Создание среды, в которой сотрудники финансового учреждения мотивированы на эффективную работу с использованием технологии больших данных для мониторинга показателей эффективности, оценки отзывов сотрудников и формирования корпоративной культуры, а также оценки общей удовлетворенности сотрудников.

Современные проблемы внедрения аналитики больших данных

Внедрение инструментов БД для анализа банковской информации отвечает интересам любого финансового учреждения. Существует ряд проблем, которые необходимо решить, прежде чем приступить к внедрению технологий больших данных.

Устаревшие вычислительные системы не имеют современной инфраструктуры для анализа больших объемов данных, что создает значительную нагрузку на устаревшие системы. Поэтому банкам рекомендуется модернизировать вычислительные системы, прежде чем переходить на работу с большими объемами данных.

Управление данными должно стать одним из главных приоритетов при использовании технологий больших данных. Даже если банк модернизирует систему, неточные, непоследовательные, неполные, повторяющиеся или устаревшие данные могут исказить результаты. До наступления цифровой эры многие данные вводились вручную, что создавало риск человеческой ошибки; банки должны тщательно анализировать и консолидировать существующие данные перед их вводом в новую систему, что позволяет исключить «загрязнение» системы ненужными данными.

При определенном уровне абстрагирования современный банк можно представить в виде информационно-коммуникационного учреждения, в основе которого электронная техника, средства связи, программные продукты и данные. Банки организуют оборот данных, превращая его в конвейерное производство, которым управляют архитекторы и исследователи данных.

Большие данные для прогнозного моделирования

Одной из важнейших предпосылок правильной процедуры принятия решений является наличие точной информации, которая часто распространяется среди многих людей. Следовательно, выявление распределенной информации представляет собой ключевой компонент во многих системах, основанных на информативном принятии решений. Как правило, такой сценарий сбора информации моделируется путем представления отдельных лиц в качестве рациональных агентов, которые готовы сообщать свою личную информацию в обмен на (денежные) вознаграждения.

В настоящее время крупнейшие отечественные банки переходят к разработке минимальных механизмов прогнозирования, которые в базе имеют агентские модели с ограниченными знаниями об убеждениях агентов. Однако без знания убеждений агентов или получения дополнительной информации невозможно разработать точные механизмы в смысле Байеса – Нэша, то есть способных предоставлять достоверные сценарии поведения клиентов.

При стандартной настройке прогнозной модели поведения клиентов банка предполагается, что клиенты (агенты) имеют общее убеждение в отношении личной информации. Такие механизмы обычно не включают знания о структуре убеждений агентов, что усложняет получение достоверной информации.

Пример постановки задачи для прогнозного моделирования поведения клиентов

Настройку таких моделей можно сделать следующим образом. Представим, что рассматриваются агенты ($n \gg 1$), которые являются клиентами банка (они распределены стохастическим образом). Если их сгруппировать по определенным принципам, например, частоте обслуживания или размеру вклада, задолженности, чтобы в каждой группе было фиксированное число агентов ($k \ll n$), то рассматриваемый период их участия в группе представим как время t . Для описания того, как агенты формируют представления о своей стратегии (частная микроинформация), необходимо ввести состояние ω , которое представляет собой случайную переменную, принимающую значения в множестве Ω . Далее обозначим ассоциированное распределение через $p(\omega)$ и предположим, что $p(\omega) > 0$ для всех $\omega \in \Omega$.

Стратегия каждого агента (частная микроинформация), или сигнал, который поступает банку, моделируется общей случайной величиной X , которая принимает значения в конечном дискретном наборе $\{0, 1, \dots, m - 1\}$, общие значения, обозначаемые через x, y, z . Для каждого агента i сигнал генерируется независимо в соответствии с распределением $Pr(X_i | \omega)$, которое зависит от переменной состояния ω . Это распределение является обычным для агентов, то есть $Pr(X_i | \omega) = Pr(X_j | \omega)$ для двух агентов i и j , и оно полностью смешанно, т. е. для всех $x \in \{0, \dots, m - 1\}$ верно, что $Pr(X = x | \omega) > 0$.

Далее предположим, что частные сигналы стохастически релевантны. В таком случае апостериорные распределения $Pr(X_j | X_i = x)$ и $Pr(X_j | X_i = y)$ (полученные из $Pr(X_i | \omega)$, $Pr(X_j | \omega)$ и $p(\omega)$) различаются по крайней мере для одного значения x_j всякий раз, когда агенты разделяют общее мнение о параметрах модели ($Pr(X | \omega)$ и $p(\omega)$), поэтому обозначим эти убеждения одинаково, как ν .

Агенты сообщают в систему свою личную информацию (сигналы), за которую они получают вознаграждение в виде, например, кредитной линии или повышенного процента на текущие и/или сберегательные вклады. Агенты могут нарушать условия договора с банком (например, задерживать погашение задолженности), поэтому, чтобы отличить истинный сигнал X от сообщенного, обозначим сообщенные значения как Y . Поскольку основной результат зависит от координации агентов, то вводим параметр шума, который моделирует потенциальные недостатки в стратегиях, например, отчетности. В частности, предположим, что с вероятностью $1 - \varepsilon \in (0, 1)$ агент является рациональным и сообщает значение, которое максимизирует его ожидаемый выигрыш, то есть

полностью выполняет обязательства перед банком. В противном случае поступают другие сигналы, которые имеют случайное значение из $\{0, \dots, m - 1\}$. В данном случае не учитываются нейтральные агенты (без движения по счетам). Более того, при разработке формальных результатов можно предположить, что не зависит от X_i , тогда характеризуется способ применения основного механизма при смещении.

Сигнал агента определяется в m -мерном пространстве, который можно обозначит как P . Чтобы упростить обозначение убеждений, можно опустить символы X и Y . В частности, вместо того чтобы использовать $Pr(X = x | \omega)$, запишем $Pr(x | \omega)$ или вместо $Pr(X_j = y / X_i = x)$ запишем $Pr(y / x)$.

Концептуальные подходы к решению

С точки зрения рациональных агентов, представленный в статье набор имеет форму байесовской игры, поэтому можно исследовать профили стратегий, которые представляют собой равновесия Байеса – Нэша. Банки заинтересованные как в соблюдении равновесия в модели, так и в практической реализации ее результатов. Поэтому к клиентам сначала подходят как к рациональным агентам. Усложняя модель, вводят коррективы, связанные, например, с вероятностью банкротства или неплатежей клиентов. Для этого применяются коэффициенты, которые могут быть подвижными и зависимыми от фазы финансового цикла, а также от других обстоятельств. Любой механизм, который принимает правильный сигнал от рационального агента, рассматривается как строгое равновесие по Байесу – Нэшу.

В таком случае агент i убежден, что другие агенты рациональны и следуют тем же правилам:

1) распределения вероятностей $Pr_i(x)$. Упрощаем $Pr_i(x) = p_i(x)$ или $p(x)$. В таком случае агент поддерживает $x^p = \arg \max_x p(x)$;

2) зависимостей в действиях от определенных условий. Для измерения сигнала s_i используется апостериорное распределение $Pr_i(x / s_i)$. Также упрощаем $Pr_i(x / s_i) = q_i(x)$ или $q(x)$.

Мотивируют действия агента убеждения о необходимости быть рациональным. Такой подход является основанием для стимулирования банков клиентов.

Подходы к решению: агент i_t в каждый момент времени t_t формирует апостериорное убеждение q . Все агенты посылают сигналы. Банк, используя такую простую модель, может оценить поведение каждого агента, точнее, отклонение его поведения от рационального посредством определения

нарушений в модели и сигналов о сбоях. Модель характеризует как общее состояние системы, так и возможные сбои в поведении каждого клиента. Набор данных в таких моделях может быть самым различным – от одного показателя до множества, которые в агрегированном виде входят в показатели модели.

Рост числа клиентов и возросшая неопределенность их поведения определяют необходимость моделирования взаимоотношений клиентов с банком и перевод на машинный учет оценку состояния агентов в модели. Математически степень свободы поведения агентов в модели – это число измерений области случайного вектора. Со временем происходит схождение векторов, что интерпретируется как переход, например, от рационального к иррациональному поведению по мере ухудшения финансового состояния клиента. Банки переходят к использованию интеллектуального анализа больших данных, чтобы выявлять аномалии с помощью моделирования, что позволяет автоматически отслеживать основные показатели, связанные с финансовой безопасностью и платежеспособностью клиентов.

Выводы

Технологии БД ознаменовали переход к «информационному обществу». Информация приобрела статус ценного актива, что привело к формированию механизма ценообразования и рыночным отношениям при организации ее движения. Информация стала своего рода новой «нефтью», которая движет информационное общество. В эпоху индустриализации нефть была основным ресурсом. В настоящее время таким ресурсом становится информация. Технология больших данных меняет бизнес-модели, персонифицирует клиентские запросы, расширяет продуктовую линейку в банковской сфере, расширяет аудиторию пользователей банковскими услугами.

Технологии больших данных жизненно важны для современного банковского дела. В настоящее время управление активами, оценку рисков, удержание и расширение клиентской базы невозможно реализовать без использования больших данных.

Информационная функция финансов всегда была среди основных. Так, индексы, котировки, процентные ставки, кредитные рейтинги, критерии устойчивости, показатели предбанкротного состояния выступали важными индикаторами деловой активности и сигнализировали о состоянии дел на микро- и макроуровнях. В современных условиях информационная функция финансов получила новое качество, она отличается от традиционной, во-первых, своей прогнозной миссией, во-вторых, масштабностью и, в-третьих,

назначением. На ее основе, в частности, строятся сценарии развития хозяйства в целом и разрабатываются меры по регулированию экономикой.

Список источников

1. **Arner D.W., Barberis J.P., Buckley R.P.** The Emergence of Regtech 2.0: From Know Your Customer to Know Your Data // Journal of Financial Transformation 79. Vol. 44, 2016. P. 1–20.
2. **Gal M.S., Rubinfeld D.L.** (2016) The Hidden Costs of Free Goods: Implications for Antitrust Enforcement. Antitrust Law Journal. Available at: http://awards.concurrences.com/IMG/pdf/galrubinfeld_final2.pdf (accessed: 16.08.2020).
3. **Goldstein I., Jiang W., Karolyi G.A.** To Fin Tech and Beyond // SSRN. February 2019. P. 1–21. <https://dx.doi.org/10.2139/ssrn.3328172>.
4. **Stasinakis Charalampos, Sermpinis, Georgios** Big Data, Artificial Intelligence and Machine Learning: A Transformative Symbiosis in Favour of Financial Technology (July 29, 2020). Available at SSRN: <https://ssrn.com/abstract=3663062>.
5. **Israel R., Kelly B.T., Moskowitz T.J.** Can Machines ‘Learn’ Finance? // SSRN. January 10, 2020. P. 1–21. <https://dx.doi.org/10.2139/ssrn.3624052>.
6. **Aggarwal N, Eidenmüller Y, Enriques L, Payne J, van Zwieten K (eds)** Autonomous Systems and the Law (Beck 2019), Available at SSRN. November 1, 2018: <https://ssrn.com/abstract=3309244>.
7. **Klioutchnikov I. K., Kliuchnikov O. I. and Molchanova O. A.** Alliance of Commerce and Financial Intermediation with Social Networks: Problems and Prospects // Journal of Internet and e-Business Studies, Vol. 2020 (2020), Article ID 704163.
8. **Sigova M.V., Zatevakhina A.V., Klioutchnikov I.K., Klioutchnikov O.I.** Approaches to evaluating the function of prediction of decentralized applications // Proceedings - 2018 International Conference on Artificial Intelligence: Applications and Innovations, IC-AIAI 2018, 2019. С. 6–11.
9. **Mulla J., Van Vliet B.,** A Query Language for Big Data in Finance. Algorithmic Finance, Forthcoming. SSRN. November 3, 2015. <https://ssrn.com/abstract=2685769>.
10. **Brooks C., Hoepner A., McMillan D., Vivian A., Simen C.** Financial Data Science: The Birth of a New Financial Research Paradigm Complementing Econometrics? (August 2, 2019). The European Journal of Finance, Forthcoming. Available at SSRN: <https://ssrn.com/abstract=3580729>.
11. **Федоренко С.Н.** Big Data и методологические возможности ресурса Google Trends для политологического исследования // Журнал политических исследований. 2018. С. 48–55.; **Володенков С.В.** Total Data как феномен формирования политической постреальности // Вестник Омского университета. Серия «Исторические науки». – 2017. – № 3 (15). – С. 409–415.

12. **Ключников И.К., Молчанова О.А., Ключников О.И.** Big Data (Большие данные) в финансах: теория и практика // *Финансы и Бизнес*, № 4, 2017. С. 104–116.
13. **Balto D. and Lane M.** (2016) Monopolizing Water in a Tsunami: Finding Sensible Antitrust Rules for Big Data. Available at: <http://ssrn.com/abstract=2753249> (accessed: 04.10. 2020).
14. **De Mauro A., Greco M., Grimaldi M.** (2016) Formal Definition of Big Data Based on its Essential Features. *Library Review*, vol. 65, no 3, pp. 122–135.
15. **QSS.** (2021) Big Data and Analytics in Banking and Finance Industry // *Medium*, February 8, 2021. Available at: <https://medium.com/qss-smart-it/your-go-to-guide-to-big-data-analytics-in-banking-936f3df4f45> (accessed: 03.23. 2021).
16. **Stucke M., Ezechri A.** (2016) *Virtual Competition*. Cambridge (Mass.): Harvard University Press, 368 p.
17. **Sokol D., Comerford R.** (2016) Does Antitrust Have a Role to Play in Regulating Big Data? *Cambridge Handbook of Antitrust, Intellectual Property and High Tech*. N.Y.: Cambridge University Press. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2723693 (accessed: 06.10.2020).

References

1. **Arner D.W., Barberis J.P., Buckley R.P.** The Emergence of Regtech 2.0: From Know Your Customer to Know Your Data // *Journal of Financial Transformation* 79. Vol. 44, 2016. P. 1–20.
2. **Gal M.S., Rubinfeld D.L.** (2016) The Hidden Costs of Free Goods: Implications for Antitrust Enforcement. *Antitrust Law Journal*. Available at: http://awards.concurrences.com/IMG/pdf/galrubinfeld_final2.pdf (accessed: 16.08.2020).
3. **Goldstein I., Jiang W., Karolyi G.A.** To Fin Tech and Beyond // SSRN. February 2019. P. 1–21. <https://dx.doi.org/10.2139/ssrn.3328172>.
4. **Stasinakis Charalampos, Sermpinis, Georgios** Big Data, Artificial Intelligence and Machine Learning: A Transformative Symbiosis in Favour of Financial Technology (July 29, 2020). Available at SSRN: <https://ssrn.com/abstract=3663062>.
5. **Israel R., Kelly B.T., Moskowitz T.J.** Can Machines ‘Learn’ Finance? // SSRN. January 10. 2020. P. 1–21. <https://dx.doi.org/10.2139/ssrn.3624052>.
6. **Aggarwal N, Eidenmüller Y, Enriques L, Payne J, van Zwieten K (eds)** *Autonomous Systems and the Law* (Beck 2019), Available at SSRN. November 1, 2018: <https://ssrn.com/abstract=3309244>.
7. **Klioutchnikov I. K., Kliuchnikov O. I. and Molchanova O. A.** Alliance of Commerce and Financial Intermediation with Social Networks:

Problems and Prospects // Journal of Internet and e-Business Studies, Vol. 2020 (2020), Article ID 704163.

8. **Sigova M.V., Zatevakhina A.V., Klioutchnikov I.K., Klioutchnikov O.I.** Approaches to evaluating the function of prediction of decentralized applications // Proceedings – 2018 International Conference on Artificial Intelligence: Applications and Innovations, IC-AIAI 2018, 2019. C. 6–11.

9. **Mulla J., Van Vliet B.,** A Query Language for Big Data in Finance. Algorithmic Finance, Forthcoming. SSRN. November 3, 2015. <https://ssrn.com/abstract=2685769>.

10. **Brooks C., Hoepner A., McMillan D., Vivian A., Simen C.** Financial Data Science: The Birth of a New Financial Research Paradigm Complementing Econometrics? (August 2, 2019). The European Journal of Finance, Forthcoming, Available at SSRN: <https://ssrn.com/abstract=3580729>.

11. **Fedorenko S. N.** Big Data and methodological possibilities of the Google Trends resource for political science research // Journal of Political Studies. 2018. pp. 48–55; Volodenkov S. V. Total Data as a phenomenon of the formation of political postreality // Bulletin of the Omsk University. A series of «Historical science». – 2017. – № 3 (15). – Pp. 409–415.

12. **Klioutchnikov I. K., Molchanova O. A., Kliuchnikov O. I.** Big Data (Big data) in finance: theory and practice // Finance and Business, No. 4, 2017. pp. 104–116.

13. **Balto D. and Lane M.** (2016) Monopolizing Water in a Tsunami: Finding Sensible Antitrust Rules for Big Data. Available at: <http://ssrn.com/abstract=2753249> (accessed: 04.10. 2020).

14. **De Mauro A., Greco M., Grimaldi M.** (2016) Formal Definition of Big Data Based on its Essential Features. Library Review, vol. 65, no 3, pp. 122–135.

15. **QSS.** (2021) Big Data and Analytics in Banking and Finance Industry // Medium, February 8, 2021. Available at: <https://medium.com/qss-smart-it/your-go-to-guide-to-big-data-analytics-in-banking-936f3df4f45> (accessed: 03.23. 2021).

16. **Stucke M., Ezrachi A.** (2016) Virtual Competition. Cambridge (Mass.): Harvard University Press, 368 p.

17. **Sokol D., Comerford R.** (2016) Does Antitrust Have a Role to Play in Regulating Big Data? Cambridge Handbook of Antitrust, Intellectual Property and High Tech. N.Y.: Cambridge University Press. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2723693 (accessed: 06.10.2020).