РИСКИ ЦИФРОВИЗАЦИИ

ВИДЫ, ХАРАКТЕРИСТИКА, УГОЛОВНО-ПРАВОВАЯ ОЦЕНКА

МОНОГРАФИЯ

Ответственный редактор
доктор юридических наук,
профессор **Ю. В. Грачева**

ПРОСПЕКТ

Коллектив авторов

Риски цифровизации: виды, характеристика, уголовно-правовая оценка

«Издательство Проспект»

2022

УДК 004:34
ББК 32.81:67

Коллектив авторов

Риски цифровизации: виды, характеристика, уголовно-правовая оценка / Коллектив авторов — «Издательство Проспект», 2022

ISBN 978-5-39-238853-0

В монографии дана характеристика ключевых научно-технических направлений, которые оказывают наиболее существенное влияние на развитие цифровой среды: большие данные, искусственный интеллект, системы распределенного реестра (блокчейн), промышленный интернет, компоненты робототехники, технологии мобильной и спутниковой связи, извлечение знаний. Представлены сферы применения цифровых технологий, сопряженные с наибольшими рисками: цифровая медицина, цифровое управление городом, цифровая логистика, электронная коммерция, Индустрия 4.0, социальные сети и медиа, цифровое управление рабочим пространством и умные дома. Описаны базовые понятия информационной безопасности и основные подходы к технической обусловленности возникновения рисков при процессах цифровизации. Особое внимание уделено анализу преступлений в компьютерной сфере, предусмотренных гл. 28 УК РФ, что приводит к выводу о необходимости пересмотра терминологии УК РФ с целью адекватного парирования существующих угроз в сфере компьютерной информации. Законодательство приведено по состоянию на февраль 2022 г. Для научных и практических работников, преподавателей, аспирантов и студентов юридических вузов.

УДК 004:34
ББК 32.81:67

ISBN 978-5-39-238853-0

Содержание

Риски цифровизации: виды, характеристика, уголовно-правовая оценка

Монография

* * *

Ответственный редактор доктор юридических наук, профессор Ю. В. Грачева
Авторы:
Грачева Ю. В., доктор юридических наук, профессор, профессор кафедры уголовного права Московского государственного юридического университета имени О. Е. Кутафина (МГЮА);
Иванов С. А., руководитель подразделения информационной безопасности компании «Первый Бит»;
Маликов С. В., доктор юридических наук, профессор кафедры уголовного права Московского государственного юридического университета имени О. Е. Кутафина (МГЮА);
Чучаев А. И., доктор юридических наук, профессор, главный научный сотрудник, и. о. заведующего сектором уголовного права, уголовного процесса и криминологии Института государства и права РАН.

Рецензенты:
Коробеев А. И., доктор юридических наук, профессор, заслуженный деятель науки РФ, заведующий кафедрой уголовного права и криминологии Дальневосточного федерального университета;
Воробьев В. В., кандидат юридических наук, доцент, заведующий кафедрой уголовно-правовых дисциплин Сыктывкарского государственного университета имени Питирима Сорокина.

Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (РФФИ) в рамках научного проекта № 18-29-18158.

Введение

Увеличение деловой, социальной активности в киберпространстве, цифровая трансформация предпринимательской деятельности и деятельности государственных и муниципальных служб предопределяют актуальность рассмотрения вопроса трансформации права в условиях развития цифровых технологий. Прорывное развитие цифровых технологий приводит к появлению новых видов нематериальных и трансформации традиционных материальных активов, образованию важных прямых и обратных зависимостей между объектами виртуального мира (киберпространства) и реального мира. Значимость преобразований и их беспрецедентная динамика изменяют характер угроз имуществу, жизни и здоровью человека, работе организаций, социуму и государству.

Противодействие киберугрозам, социокультурным угрозам, терроризму и идеологическому экстремизму не только ставит новые задачи перед специалистами по информационной безопасности, но и требует выработки механизмов правовой защиты, обладающих свойством оперативной актуализации в соответствии с изменениями характера и масштабов угроз. Такие механизмы позволят опережающими темпами реагировать на криминогенные угрозы безопасности личности, общества и государства в цифровой среде.

Для разрешения этой проблемы в настоящей монографии преследуется в первую очередь пропедевтическая цель – создание необходимого теоретического фундамента для последующего рассмотрения специальных уголовно-правовых вопросов. Проводится анализ технологий, образующих цифровую среду, и сфер жизнедеятельности, на которые инновационные технологии воздействуют или будут оказывать наибольшее влияние.

Угрозы цифровой среды, создающие опасность ущерба для человека, социума и государства, в пособии исследуются как совокупность причин и факторов, обусловленных применением цифровой технологии, от которой зависит вероятность нанесения ущерба, и сферы жизнедеятельности, определяющей характер и размер ущерба.

Глава I. Научно-технические направления, оказывающие наибольшее влияние на развитие цифровой среды

§ 1. Искусственный интеллект

Понятие. Искусственный интеллект (далее – ИИ) – это область научных знаний и технологий создания интеллектуальных машин и интеллектуального программного обеспечения. Также ИИ называют свойство интеллектуальных систем выполнять творческие функции, которые традиционно считаются прерогативой человека. Одной из ключевых особенностей интеллектуальных вычислительных систем является их способность приобретать знания посредством обучения (самомодификации) и применять эти знания для решения проблем.

Подобно тому, как человек использует свой мозг, чтобы учиться на новой информации, собранной органами чувств, ИИ учится на информации, передаваемой ему, например, в виде изображения или правил игры. Данная информация не только обрабатывается в соответствии с тем, как он запрограммирован, но и меняет сам алгоритм, при помощи которого ее обрабатывают. Процесс, при котором ИИ запрограммирован на автоматическое изменение собственного алгоритма, называется машинным обучением. Например, для идентификации кошки люди принимают во внимание форму и физические характеристики и сверяют это со знаниями о том, кто такая кошка, основываясь на воспоминаниях и опыте. Обучение человека естественным образом включает в себя построение абстрактных представлений, т. е. человек может распознать кошку, даже если видит только задние лапы и хвост или видит рисунок с кругом, обозначающим голову, и двумя треугольниками, изображающими уши. Для того чтобы ИИ мог идентифицировать кошку, в систему нужно внести миллионы изображений кошек и обучить ее распознавать определенные группы пикселей – наименьших единиц изображения, которые создают форму кошки. Впервые такое обучение ИИ было проведено компанией Google в 2012 г. с использованием технологии, известной как Deep Learning, в целях построить программу, которая может распознавать изображения с кошками. В программе изначально не задавались правила, согласно которым у кошек четыре лапы, хвост, два уха и т. д., но, поскольку изображения на обучающих данных были помечены как содержащие или не содержащие кошек, программа смогла самостоятельно создать визуальную концепцию кошки. Когда программе предоставлялось новое изображение, она с высокой точность была способна пометить его как «содержащий кошку» или нет. Искусственный интеллект Google получил информацию из изображений, научился идентифицировать кошек, а затем мог применять правила для решения вопроса о том, какие новые изображения содержат рисунок кошки. Несмотря на то, что в отличие от мозга человека ИИ на самом деле не знает, кто такая кошка, и не понимает этого, ему удалось создать абстрактное представление о том, что мы называем кошкой или, если точнее, «кошкой на изображении».

Существуют разнообразные методы машинного обучения: глубокое обучение с использованием нейронных сетей, обучение с подкреплением, обучение на основе статистических принципов. Многие программы ИИ применяются для анализа и обработки изображений или речи либо извлечения информации из них. Глубокое обучение зачастую необходимо для прогнозов, таких как медицинские диагнозы или возможное мошенничество с кредитными картами.

История. Старт развития искусственного интеллекта в современном его понимании произошел в 1950-х гг. XX в. и изначально предполагал решение сложных математических задач и

создание «мыслящих машин». С самого начала сложились два конкурирующих подхода. Один – с применением формальных правил для манипулирования символами, логического подхода, не основанного на биологии. Этот подход получил название «старый добрый искусственный интеллект» (Good Old-Fashioned Artificial Intelligence, GOFAI). Сторонники второго подхода исходили из того, как работает мозг, и создавали «искусственные нейронные сети», базирующиеся на моделях, в основу архитектуры которых положена нейронная структура мозга.

В первые 20 лет GOFAI принес больший успех, что привело к значительному государственному финансированию. В реальных же условиях GOFAI не дал значимых результатов. Методология использования искусственных нейронных сетей не прошла проверку прикладными задачами и в 1970-х гг. финансирование исследований прекратилось, их количество уменьшилось, а сообщество ИИ сократилось. Через 10 лет, когда были усовершенствованы системы GOFAI и нейронные сети, решение задач, считавшихся ранее неразрешимыми, стало достижимым, и область ИИ снова стала казаться многообещающей. Однако надежды вновь не оправдались, и к 1990 г. количество исследований ИИ снова сократилось. Успех к рассматриваемой технологии пришел в начале 2000-х гг., что было обусловлено рядом значимых факторов:

– прогрессом методологии Deep Learning, модели решения задач, вдохновленной биологическими свойствами нейронных сетей;

– возможностью использования огромных объемов данных, ставших доступным в настоящее время;

– возросшей вычислительной мощностью процессоров;

– возможность горизонтального наращивания мощности вычислительных комплексов.

Обладая большими массивами данных, современные нейронные сети ИИ зачастую превосходят человека в решении многих задач, например в распознавании образов, моделировании, играх. Такая эффективность ранее была недостижима для систем ИИ. При этом системы, обеспечившие технологический и научный прорыв, могут самообучаться.

Для проведения сравнительной оценки ИИ и человеческих возможностей в 1950 г. А. Тьюринг предложил то, что станет известным как «тест Тьюринга». До сих пор еще ни ода система ИИ не прошла такой тест. Согласно правилам этого теста ИИ должен обрабатывать естественный язык, уметь учиться на разговорной речи и помнить сказанное, сообщать идеи человеку и усваивать общие понятия, отображая то, что мы называем здравым смыслом. Первым таким предложенным тестом стала игра, в которой участвуют мужчина, женщина и следователь. Задача следователя (ИИ) состоит в том, чтобы определить, кто из участников мужчина, а кто женщина. Невыполнимость по настоящее время теста Тьюринга связана с простым вопросом: попадает ли, в принципе, эта способность системы казаться разумной в область вычислимых проблем? Повсеместное распространение ИИ в виде голосовых помощников, систем распознавания изображений, голоса, автоматического перевода могут создать иллюзию того, что ИИ уже скоро достигнет уровня человеческого интеллекта. Однако ИИ нуждается в огромном количестве данных, чтобы учиться, в отличие от нашего мозга, который может учиться на разовом опыте, выстраивать заключения из одного-единственного события. Для поступательного развития ИИ необходимо дальнейшее углубление знаний об основных принципах функционирования мозга и о видах биологических сокращений, посредством которых человеческий мозг выполняет задачи. Несмотря на недостижимость идеала, повсеместное распространение методологии ИИ дает ощутимую пользу для решения специальных задач.

Технологии искусственного интеллекта. Искусственный интеллект характеризуется в первую очередь задачами, которые он предназначен решать, но некоторые технологии и методологии ассоциируются именно с технологическим решением ИИ к ним относят машинное обучение, биологическое моделирование, представление и использование знаний, дополненный интеллект, чат боты, системы управления ИИ и другие.

Машинное обучение является обширным подразделом ИИ, изучающим методы построения алгоритмов способных обучаться. Различают два типа обучения: по прецедентам (или индуктивное обучение), которое основано на выявлении общих закономерностей по частным эмпирическим данным; дедуктивное (или машинное) обучение, предполагающее формализацию знаний экспертов и перенос этих знаний в компьютер в виде базы знаний.

Машинное обучение находится на стыке математической статистики, методов оптимизации и классических математических дисциплин, но имеет также собственную специфику, связанную с проблемами вычислительной эффективности и переобучения. Многие методы индуктивного обучения разрабатывались как альтернатива классическим статистическим подходам. Многие методы тесно связаны с извлечением знаний и интеллектуальным анализом данных (Data Mining).

Машинное обучение не только математическая, но и практическая, исследовательская дисциплина. Чистая теория, как правило, не приводит к созданию методов и алгоритмов, полностью готовых к применению на практике. Чтобы заставить модель данных эффективно работать необходимо ее уточнять, выявлять дополнительные эвристики, компенсирующие несоответствие первоначально сделанных предположений условиям реальных задач. Практически ни одно исследование в машинном обучении не обходится без эксперимента на модельных или реальных данных, подтверждающего практическую работоспособность метода.

Выделяют следующие области машинного обучения:

Обучение с учителем – задачи, в которых требуется найти зависимость ответов от описаний, т. е. построить алгоритм, принимающий на входе описание объекта и выдающий на выходе ответ. Под учителем в данном случае следует понимать либо саму обучающую выборку, либо того, кто указал на заданных объектах правильные ответы. В рамках этого раздела машинного обучения могут решаться задачи классификации, регрессии, ранжирования и прогнозирования.

Задача классификации, т. е. определения отношения объекта к той или иной заранее заданной группе объектов актуальна в коммерческой деятельности (классификация клиентов и товаров в целях оптимизации маркетинговых стратегий, стимулирования продаж, сокращения издержек), сфере телекоммуникаций (классификация абонентов для определения уровня лояльности и предпочтения при выборе услуг оператора), медицине и здравоохранении (в целях диагностики заболеваний, классификации населения по группам риска), банковской сфере (для кредитного скоринга, отнесения человека к той или иной группе, что позволяет определить вероятность возврата кредита и вычислить размер допустимой суммы кредитования).

Задача регрессии – определение прогнозного числового значения решает такие прикладные задачи как прогнозирования спроса (дает количественную оценку спроса на тот или иной товар или вид товара), прогнозирования доходности акций по совокупности предоставляемой информации о деятельности компании, конкурентов, рыночной конъюнктуре, погодных и политических условиях и т. д., изучение структуры и постатейных размеров издержек производства на основе данных прошлых периодов и изменений, что позволяет прогнозировать регулярные расходы, проведение макроэкономических расчетов, в которых учитывается большое количество факторов, прогнозирование даты возврата кредита.

Задача ранжирования ставит целью сортировку объектов по значениям некоего характеризующего их показателя. Выбор показателя для ранжирования система определяет автоматически. В некоторых случаях задача ранжирования решается без выделения конкретного показателя за счет последовательно определения «соседей». Задача ранжирования применяется в информационном поиске, например, при сортировке в поисковых системах результатов поиска по «релевантности» – условному значению, определенному системой; в рекомендательных системах (в частности, на основе ранее прослушанных композиций предоставляется

совет о том, какую песню или стиль система рекомендовала бы прослушать в порядке убывания рекомендательного индекса).

Задача прогнозирования ставится с целью спрогнозировать свойства объекта на основе данных за прошлые периоды. На примере желания снять наличные денежные средства в банкомате задача прогнозирования позволяет определить время и объем спроса на наличные денежные средства в банкоматах, установить необходимую численность персонала для обработки обращений клиентов во время штатной и пиковой нагрузки, спрогнозировать качество продукции по данным о производственном процессе, качестве исходного материала и квалификации персонала.

Обучение без учителя – в этой методологии система ИИ должна быть способна не просто отнести объект к той или иной группе, а без дополнительной информации самостоятельно выделить такие группы и затем определять принадлежность к ним объектов. По методологии обучения без учителя решаются задачи кластеризации, ассоциативных правил, фильтрации выбросов, сокращения размерности, заполнения пропущенных значений и др.

Задача кластеризации заключается в том, чтобы сгруппировать объекты в кластеры, представляющие собой сравнительно однородные группы объектов. К задаче кластеризации сводятся:

– анализ социальных сетей в разных сферах жизни общества для проведения исследований;

– оценка политических предпочтений сегментов аудитории в разных регионах, социальных и демографических группах;

– прогнозирование политической активности и акций на основе выявления поведенческих паттернов;

– агитация, т. е. распространение информации о кандидатах, данные о которых гражданин еще не рассматривал, но разделяет ценности партии кандидата;

– определение центров формирования общественного мнения;

– выбор популярных личностей среди лояльных к бренду людей в целях повысить эффективность кампаний при помощи информационных вирусных технологий, побуждающих распространять сведения о продуктах и компании саму аудиторию, которой она предназначена;

– поиск подходящих кандидатов в сотрудники компании по данным резюме и историй успеха сотрудников, которые уже плодотворно работают в компании;

– подбор сотрудников для какого-либо проекта;

– повышение эффективности командообразования на основе подтвержденных личных и профессиональных качеств;

– фокусировка рекламных кампаний на конкретном сегменте целевой аудитории;

– выявление латентных, не выражаемых явно потребностей покупателей, которые не ищут товар в интернете и не обращаются в магазины, но в общедоступных сообщениях (постах), группах, в которых состоят эти пользователи, оставляют информацию о своих намерениях или предпочтениях;

– определение кластеров коррумпированности – связей бизнеса и представителей власти.

Задача поиска ассоциативных правил – определение часто встречающихся наборов объектов в большом множестве таких наборов. Прикладные задачи, решаемые установлением ассоциативных правил:

– изучение событий, выявление причинно-следственных связей в поведении поставщиков, покупателей, сотрудников, инвесторов, конкурентов и иных лиц, оказывающих или могущих оказать влияние на компанию;

– анализ покупательской корзины – определение сочетаний товаров, пользующихся стабильным спросом, в целях оптимизировать поиск наборов покупателями;

– стимулирование спроса за счет формирования дополнительных предложений, проведения эффективных маркетинговых акций, продвигающих среди аудитории дополнительные товары.

Задача фильтрации выбросов – обнаружение в обучающей выборке небольшого числа нетипичных объектов. К задаче сводятся проблемы

– обнаружение мошенничества, т. е. выявление аномальных финансовых показателей по выручке или объему продаж, что помогает обнаружить факт кражи денежных средств или передачу информации конкурентам;

– обеспечение информационной безопасности. В частности, аномальное время работы сотрудника или его нетипичные действия дают возможность установить факт инсайдерской деятельности либо идентифицировать несанкционированный доступ к информационной системе;

– выявление ошибок при экономических расчетах, т. е. фильтрация выбросов привлекает внимание к ошибочно введенной в ручном режиме информации за счет определения ее нетипичности или отсутствия смысла.

Задача сокращения размерности заключается в том, чтобы при помощи некоторых функций преобразования перейти к наименьшему числу признаков объекта, не потеряв при этом никакой существенной информации. Решение задачи дает возможность оптимизации:

– производственных процессов – благодаря выявлению действий, не влияющих на эффективность;

– расходов на содержание сложных систем;

– использования вычислительных ресурсов.

Задача заполнения пропущенных значений – замена недостающих значений в матрице «объекты-признаки» их прогнозными значениями. Метод замены используется в социальных исследованиях, когда данные собираются не в полном объеме; для восстановления данных при сбоях или преднамеренном уничтожении; при прогнозировании удовлетворенности от продукта на основе данных по другим продуктам и другим потребителям.

Кроме обучения с учителем и без учителя, в машинном обучении применяются и другие методы:

Обучение с подкреплением – процесс, при котором происходит обучение модели, не имеющей сведений о системе, но обладающей возможностью производить действия в ней. Действия переводят систему в новое состояние, и модель получает от системы некоторое вознаграждение. Подобное обучение используется:

– в управлении роботами при выполнении таких задач, как манипулирование предметами, навигация в загруженном пространстве, поиск устойчивого положения предмета;

– в управлении технологическими процессами;

– при персонализации показов рекламы в интернете;

– в управлении ценами и ассортиментом в сетях продаж;

– при маршрутизации в телекоммуникационных сетях.

Частичное обучение занимает промежуточное положение между обучением с учителем и без учителя. Пример прикладной задачи – автоматическая рубрикация большого количества текстов при условии, что некоторые из них уже отнесены к каким-то рубрикам. Такая задача стоит при работе с большими объемами текстовых данных экономистами и юридическими службами, а также в научной деятельности.

Динамическое обучение возможно как с учителем, так и без него. Специфика такого обучения состоит в том, что информация о состоянии объектов поступает потоком и требуется немедленно принимать решение по каждому прецеденту, одновременно доучивая модель зависимости с учетом новых прецедентов. Как и в задачах прогнозирования, здесь существенную роль играет фактор времени.

Метаобучение отличается от методов тем, что прецедентами являются ранее решенные задачи обучения. Требуется определить, какие из используемых в них приемов работают более эффективно. Конечная цель – обеспечить постоянное автоматическое совершенствование алгоритма обучения с течением времени.

Биологическое моделирование искусственного интеллекта. Биокомпьютинг, или квазибиологическая парадигма (Biocomputing), – это биологическое направление в ИИ, сосредоточенное на разработке и использовании компьютеров, которые функционируют как живые организмы или содержат биологические компоненты, так называемые биокомпьютеры. В отличие от понимания ИИ, когда исходят из положения о том, что искусственные системы не обязаны повторять в своих структуре и работе структуру и протекающие в ней процессы, присущие биологическим системам, сторонники биокомпьютинга считают, что феномены человеческого поведения, способность человека к обучению и адаптации есть следствие именно биологической структуры и особенностей ее функционирования. Биокомпьютинг позволяет решать сложные вычислительные задачи, организуя вычисления при помощи живых тканей, клеток, вирусов и биомолекул. Часто используют молекулы дезоксирибонуклеиновой кислоты, посредством которых создают ДНК-компьютер. Биопроцессором также могут служить белковые молекулы и биологические мембраны. Например, на основе бактериородопсин-содержащих пленок создают молекулярные модели перцептрона.

Представление и использование знаний. Представление знаний (ПЗ), или Knowledge Representation (KR) – это область ИИ, в которой изучают то, как могут быть представлены знания и факты о мире и какие рассуждения могут быть сделаны с этими знаниями. Проблематикой ПЗ является возможность представления знаний таким образом, чтобы они были достаточными (в полном объеме содержали знания, необходимые для решения проблемы); не избыточными (компактными, естественными, пригодными для эффективных вычислений); способными выразить особенности проблемы; могли компенсировать недостаточную точность представляемых данных и обеспечить приемлемое время вычислений.

Для решения этих задач используется методология инженерии преставления знаний, в которых выделяют:

– декларативные знания, основанные на понятиях, фактах и объектах. Они дают всю необходимую информацию о проблеме в виде простых истинных или ложных утверждений;

– процедурные знания – правила, стратегии, программы и процедуры. Они описывают то, как проблема может быть алгоритмически решена, и шаги на пути ее решения;

– эвристические знания, накапливаемые интеллектуальной системой в процессе ее функционирования, а также заложенные в ней априорно, но не имеющие статуса абсолютной истинности в данной проблемной области. Обычно эвристические знания связаны с отражением в базе знаний неформального опыта решения задач. Эвристические знания основаны на правиле «большого пальца», т. е. на отказе от очевидно неприемлемых вариантов. Эвристические представления полезны для управления процессом рассуждения. При этом представление знаний базируется на стратегиях решения проблем в соответствии с опытом преодоления прошлых проблем, которым обладает эксперт;

– метазнания, дающие представление о других типах знаний, которые подходят для решения проблемы. Это «знания о знании», о том, как оно устроено и структурировано; «знания о получении знаний», т. е. приемы и методы познания (когнитивные умения) и оценка возможностей работы с ним. Иными словами, метазнания объединяют знания о способах использования знаний и знания о свойствах знаний. Задача применения метазнаний состоит в повышении эффективности решения проблем посредством правильного процесса рассуждения;

– структурные знания, связанные с информацией, основанной на правилах, наборах, концепциях и отношениях. Они представляют собой информацию, необходимую для разработки структур знаний и общей ментальной модели проблемы.

Архитекторы систем представления данных используют следующие логические структуры: списки и деревья для выстраивания иерархических знаний; семантические сети – схемы, применяемые для демонстрации здравого смысла или стереотипных знаний; скрипты – для описания события. Методология представления и использования знаний нашла широкое распространение в процессе развития экспертных систем – программного обеспечения, способного перенять у человека экспертизу в узких предметных областях, а также выступает сквозной низкоуровневой методологией, обеспечивающей возможность архитектурного планирования систем ИИ и баз знаний.

Области применения искусственного интеллекта. Работа с естественными языками и голосовые помощники. Обработка естественного языка (Natural Language Processing, NLP) является областью применения ИИ, которая занимается взаимодействием между компьютерами и людьми и использует естественный язык человека. Это направление, объединяющее ИИ и математическую лингвистику, изучает проблемы компьютерного анализа и синтеза естественных языков. Анализ в данном контексте означает возможность читать, распознавать, понимать и расшифровывать человеческие языки в целях выявления смысла передаваемой информации; синтез – способность генерировать текст с учетом грамматических и семантических правил естественного языка. Решение этих проблем позволит создать удобную форму взаимодействия компьютера и человека.

Типовое взаимодействие человека с компьютером на основе NLP выглядит следующим образом:

– человек что-либо произносит на естественном (человеческом) языке в микрофон компьютера;

– компьютер записывает звук;

– записанная аудиоинформация распознается и преобразуется в текст;

– данные текста обрабатываются интеллектуальными системами с учетом смысла сказанного и ответ выдается в форме цифровых данных;

– обработанные данные преобразуются в аудиоформат;

– компьютер воспроизводит аудиофайл.

Обработка естественного языка служит основой для многих прикладных программных приложений:

– приложений языкового перевода, например Google Translate или Yandex Переводчик;

– текстовые процессоры для проверки грамматической точности текстов, такие как Microsoft Word или Grammarly;

– приложения интерактивного голосового ответа, используемые в центрах обработки вызовов для ответа на запросы определенных пользователей;

– голосовые помощники, такие как Google Assistant, Siri, Cortana и Alexa;

– телефонные роботы для голосовой навигации по сервисам и автоматических голосовых уведомлений.

Некоторые системы способны не только распознавать человеческую речь, но и давать оценку – «тональность» высказываний, которая показывает эмоциональное состояние человека. Прикладное применение имеют также такие задачи, как идентификация человека по голосу и характерным речевым оборотам, определение числа участвующих в дискуссии людей и степени удовлетворенности полученным ответом.

Робототехника и искусственный интеллект. Робототехника и ИИ представляют собой наиболее перспективную комбинацию для автоматизации задач, не входящих в первичные заводские настройки роботизированной техники. В последние годы ИИ все больше распространяется в роботизированных решениях, обеспечивая гибкость и возможность обучения там, где раньше использовались неизменяемые производственные настройки. Широко задействованы роботизированные системы с ИИ-управлением на производственных предприятиях:

Сборочный роботизированный ИИ в сочетании с передовыми системами компьютерного зрения помогает в корректировке движений и манипуляций робота в режиме реального времени. Подобные системы применяются на сложных производственных участках и отраслях, например в аэрокосмической промышленности. ИИ также обеспечивает возможность роботизированной системе обучаться и автоматически определять, какие пути лучше всего подходят для конкретных процессов непосредственно в процессе его работы.

Упаковочные роботы. Роботизированные системы часто используют ИИ для более быстрой, дешевой и точной упаковки. ИИ помогает определять и сохранять оптимальные паттерны движения манипуляторов, которые совершает роботизированная система. В то же время постоянно совершенствует их, что делает пусконаладку и перенастройку роботизированных систем доступной пользователю без технической квалификации.

Обслуживание клиентов. В настоящее время роботы обслуживают клиентов в розничных магазинах, больницах и отелях по всему миру. Некоторых из этих роботов используют возможности обработки естественного языка для взаимодействия с клиентами. Чем больше эти системы взаимодействуют, «общаются» с людьми, тем качественней становится их работа. Развитие обеспечивается имплементированной системой самообучения.

Робототехника с открытым исходным кодом – часть роботизированных систем, которые в настоящее время представлены на рынке, поставляются как системы с открытым исходным кодом и возможностями ИИ. Благодаря этому пользователи могут перепрограммировать (за счет доступа к открытому коду алгоритмов) и научить (за счет функционала ИИ) своих роботов выполнять пользовательские задачи в зависимости от окружения и конкретного применения, например, на малых сельскохозяйственных предприятиях в которых необходимо выполнять множество разнообразных типовых операций. Совместное использование робототехники с открытым исходным кодом и систем ИИ является одним из основных трендов развития робототехники.

Проблематика безопасности использования систем искусственного интеллекта. Широкое проникновение систем машинного обучения обусловливается высокой эффективностью их использования и фактом свершившейся Четвертой промышленной революции. Совершенствуется автоматизация и интеллектуализация бизнес-процессов, технологические решения, совершенствуются роботы, функционирующие на базе систем искусственного интеллекта (ИИ). Все большее количество операций выполняется и решений принимается автоматически.

За счет этого расширяются возможности для киберпреступников атаковать информационные системы, использовать их в противоправных целях, что зачастую наносит вред жизни, здоровью, финансовый и другие виды ущерба. Если раньше хакеры ставили себе целью украсть данные с персонального компьютера или списать деньги с личного банковского счета, то в настоящее время глубокая цифровизация промышленности, транспортной и коммуникационной отраслей открывает новые возможности для деятельности злоумышленников. Одним из направлений атак стали системы ИИ, особенно уязвимые в случае наличия открытого доступа к их базовым алгоритмам и логике их работы.

Угрозы безопасности систем искусственного интеллекта.

В транспортной отрасли. Злоумышленники могут нарушать штатную работу автономных транспортных средств посредством провокаций приводящих к некорректной реакции транспорта на знаки ограничения проезда или скорости. Исследовательские проекты и эксперименты доказывают, что такие атаки не сложны для злоумышленников. К началу 2019 г. было опубликовано более 100 научных работ, в которых показаны разнообразные способы атак, имеющих целью вызвать ошибки системы распознавания изображений. Например: если маленький, трудноразличимый человеком стикер наклеить на дорожный знак система ИИ будет воспринимать его как другой знак, что при специальном выборе места атаки может спровоцировать аварию. За счет выявления уязвимости алгоритма ИИ и не сложного механизма

ее реализации появляются угрозы не только для информационной системы, но и для жизни и здоровья людей.

В системах контроля доступа. Все более широкое распространение получают системы распознавания лиц в системах контроля физического доступа. Эти системы гораздо удобнее, чем RFID-карты, ключи или использование пин-кодов. Системы распознавания лиц снижают время, которое тратится на аутентификацию и повышают долю эффективного рабочего времени сотрудников организаций. Кроме того, в качестве дополнительной меры безопасности новые типы банкоматов используют механизм распознавания лица клиента. Однако, все еще нельзя быть уверенным и в том, что подобные средства абсолютно защищены и их нельзя взломать. В настоящее время независимые исследования показывают, что многие системы распознавания образов можно обойти при помощи специальных очков, что позволит проникнуть на охраняемую территорию, украсть материальные ценности или обмануть банкомат или платежную систему.

В системах анализа лояльности к бренду. Анализ слабоструктурированных или «сырых» текстов в социальных сетях и публичных платформах на тональность (негативный или позитивный характер) высказываний играет существенную роль для изготовителей товаров народного потребления, лекарств, производителей контента, кино- и музыкальной продукции. Атака на системы анализа тональности текста может вызвать серьезный ущерб для организаций, изучающих таким образом потребительский рынок и принимающий управленческие решения на основании результатов анализа. Исследователи продемонстрировали возможность переобучения и порчи модели ИИ, предназначенной для автоматического формирования оценки тональности комментариев. Злоумышленник может написать положительный комментарий, который ИИ воспримет как негативный и переобучится. Незначительное изменение в одном слове предложения может привести к тому, что система неправильно истолкует истинный характер комментария. Поскольку анализ «сырых» текстов имеет большое влияние на принятие управленческих решений атаки наносят прямой финансовый ущерб (в виде не окупившихся инвестиций в сам скомпрометированный проект) и упущенной прибыли.

В системах поведенческого анализа. Методы машинного обучения полезны для определения ненадлежащего поведения пользователей на публичных сервисах. Речь идет о выявлении фальшивых пользователей в социальных сетях и пользователей, которые платят за доменные имена, создают сайты-муляжи, чтобы иметь фальшивые, фактически анонимные учетные записи. Вредоносные краудсорсинговые, или, как их еще называют, краудтёрфинговые, системы нужны для связи заказчиков, которые готовы платить за дезинформацию о своем продукте или продукте конкурента, с исполнителями, которые реализуют эти планы, создают и распространяют поддельные новости, проводят вредоносные политические кампании. До недавнего времени модели машинного обучения были весьма эффективными в выявлении подобного рода активности, с точностью до 95 % отделяя естественное поведение от работы краудтерферов. Вместе с тем эти алгоритмы уязвимы, например для атак «отравлением» данных. При целевом противостоянии эффективность существенно снижается.

В системах обнаружения мошенничества с кредитными картами. В некоторых системах обнаружения мошенничества специальный аналитический инструмент (классификатор логистической регрессии) применяется для выявления транзакций с признаками мошенничества, которые блокируются до детального выяснения их валидности. Однако он тоже может подвергнуться атаке и мошеннические транзакции останутся незамеченными.

В системах интеллектуальной идентификации человека. Для усиления контроля выполнения «чувствительных» финансовых операций используются алгоритмы, определяющие по специфичности нажатия клавиш, что данные вводит человек, и идентифицирующие личность человека. Однако злоумышленники научились создавать состязательные выборки, которые обманывают весьма точный в нормальном режиме работы классификатор. После исследова-

тельской атаки алгоритм начинал определять искусственно созданный клавиатурный ввод как принадлежащий конкретному пользователю-человеку.

В статистических спам-фильтрах. Некоторые спам-фильтры (например, SpamAssasin, SpamBayes, Bogo-фильтр) основаны на популярном алгоритме обучения Naive Bayes Machine, который впервые был применен в 1998 г. для фильтрации нежелательной почты. Посредством исследовательской атаки злоумышленники научились успешно «обходить» фильтры моделей машинного обучения.

Классификация атак на системы машинного обучения. Атаки на системы машинного обучения классифицируются по их целям, типу вызываемой в системе ошибки, осведомленности атакующего и типу атаки.

Классификация по целям атаки. Данная классификация проводится в зависимости от характера нарушения свойств безопасности: доступности, целостности, конфиденциальности модели и другим целевым свойствам, установленным для системы.

Нарушение доступности. К атакам с целью нарушения доступности относят атаки, направленные как на снижение стабильности работы модели для корректных входных данных, так и на полную остановку сервиса. К таким атакам относится:

– искусственное формирование запросов, которые требуют большей, чем планируемая, вычислительной мощности, искусственно вводя систему в режим пиковой нагрузки, что драматически снижает общую производительность;

– генерация потока сложноанализируемых объектов, которые будут ложно квалифицироваться и требовать медленной ручной классификации, отвлекая персонал от штатной работы;

– запуск конкурентных процессов, не позволяющих модели ИИ работать на проектных мощностях.

Нарушение целостности. Успешные атаки этого класса приводят к тому, что система продолжает корректно работать на основном потоке входных данных, но непредсказуемым образом дает некорректный вывод. Более сложной является атака обучающая модель таким образом, что на определенных, заранее установленных злоумышленниками данных выдается нужный злоумышленнику вывод. К этому классу атак относятся атаки состязательными примерами. Принцип атаки – подача модели на вход данных, изменённых таким образом, чтобы модель машинного обучения модель изменялась под задачи злоумышленника. Одно из планируемых последствий таких атак – подорвать доверие пользователей, которые увидев явные и непредсказуемые ошибки ИИ откажутся от этого сервиса.

Нарушение конфиденциальности. В результате атак этого класса происходит получение конфиденциальной информации о пользователях, самой модели, гиперпараметрах, использованных во время обучения (являющихся интеллектуальной собственностью), данных обучения. Это разведывательные атаки, backdoor, trojans и др.

Классификация по типу вызываемой ошибки. Когда атакующий ставит себе цель добиться гарантированно ошибочной классификации, атака называется non-targeted. Например, если на дорожный знак нанести определенную краску, модель распознавания уже не сможет отреагировать на знак.

Атака относится к типу targeted если цель атакующего отнести какой-либо экземпляр к определенному классу даже если это и не так. Например, рекламный плакат может содержать в себе паттерн, воспринимаемый моделью как дорожный знак и инициировать соответствующее поведение управляемой системой. Существенной проблемой является то, что человек визуально обнаружить проводимые таким образом атаки не сможет.

Классификация по осведомленности атакующего. Успешность атаки во многом зависит от того, сколько информации у атакующего о модели. Если атакующему известны модель, алгоритм, данные обучения, тип нейронной сети, количество ее слоев, то это атака называется атакой «белого ящика». Если атакующий обладает минимальными (общедоступными) знаниями

о модели, данными обучения и алгоритмами, такие атаки называют атаками «черного ящика». Атаку, в которой используются частичные знания о модели, называют атакой «серого ящика».

Классификация по типу атаки. Среди атак на модели глубокого обучения выделяют три основных типа: состязательные атаки, «отравление» данных и исследовательские атаки. Кроме основных проводятся такие атаки как backdoors, trojans и др.

Состязательные атаки. Атаки реализуются посредством того, что входные данные изменяют таким образом, чтобы модель переобучилась и стала допускать ошибки в классификации. Угрозы от такого типа атак высока, поскольку подобные атаки очень эффективны, просты в реализации и масштабируемы – один и тот же метод атаки применим к различным моделям, построенным на одном алгоритме обучения.

«Отравление» данных. Такая атака проводится на этапе первичного обучения модели, когда злоумышленник вводит данные или манипулирует данными обучения, либо чтобы создать «черный ход» для использования во время эксплуатации (без ущерба для производительности модели при обычных входных данных), либо с целью добиться последующего генерирования произвольных ошибок искажая предназначение модели в процессе обучения.

В зависимости от цели злоумышленника это нарушает свойства целостности или доступности модели. Типичный пример создания «черного хода» – атака на распознавание лиц, когда злоумышленник вводит в набор обучающих образцов данные определенного объекта. Цель состоит в том, чтобы заставить модель связать конкретный объект (допустим, кепку) с целевым пользователем, например, пользователя, имеющего право доступа не территорию. Впоследствии любое изображение лица человека в кепке будет классифицироваться как пользователь, имеющих право доступа, даже если оно принадлежит не зарегистрированному в модели человеку. «Отравление» – один из самых распространенных типов атак. История «отравляющих» атак на ML началась в 2008 г. со статьи посвященной теме эксплуатации уязвимостей машинного обучения чтобы подорвать штатную работу спам-фильтров. В статье был представлен пример атаки на спам-фильтр. Позже было опубликовано более 30 других исследовательских работ об «отравлении» и защите от него.

Существуют четыре основных стратегии «отравления» данных:

1) модификация меток: атаку модификации меток злоумышленник проводит на этапе обучения модели – изменяются классификационные метки случайных экземпляров в наборах данных для обучения;

2) внедрение данных: при подобной атаке у злоумышленника нет доступа ни к данным обучения, ни и к алгоритму обучения, но у него есть возможность дополнить новыми данными обучающий набор. Таким образом можно исказить целевую модель, вставив в набор обучающих данных вредоносные образцы;

3) модификация данных: у атакующего нет доступа к алгоритму обучения, но он имеет полный доступ к данным обучения. Обучающие данные злоумышленник отравляет непосредственно путем изменения перед их использованием для обучения целевой модели;

4) разрушение логики: если у противника есть возможность напрямую вмешиваться в алгоритм обучения. Такая атака также называется логическим искажением.

Исследовательские атаки. Целью таких атак является нарушение конфиденциальности на этапе штатной работы модели. К исследовательским относят несколько типов атак: восстановление модели, восстановление принадлежности, инверсия модели, восстановление параметров. В процессе исследовательской атаки изучается модель ИИ или набор данных, которые в дальнейшем используют злоумышленники. Результат такой атаки – получение знаний о системе ИИ и ее модели, т. е. это атака для извлечения моделей. Атака на данные позволяет «добыть», в частности, сведения о принадлежности экземпляра классу (например, о наличии прав доступа на объект конкретного человека). При помощи инверсии модели извлекают конкретные данные из модели. В настоящее время исследования посвящены в основном ата-

кам логического вывода на этапе разработки модели, но они возможны и во время обучения. Например, если мы хотим понять, как веб-сайт социальной сети определяет принадлежность к целевой аудитории, в частности к группе беременных женщин, чтобы показать конкретную рекламу, то можем изменить свое поведение, предположим, пытаясь найти информацию о памперсах, и проверить, получаем ли мы объявления, предназначенные для будущих мам.

Восстановление принадлежности экземпляра. Злоумышленник намеревается узнать, был ли конкретный экземпляр в наборе обучающих данных. Речь идет о распознавании изображений. Атакующий хочет проверить, были или нет в обучающем наборе сведения о конкретном человеке. Сам по себе это редко используемый тип разведочных атак. Однако он дает возможность разработать план дальнейших атаки, таких как атака «уклонение» класса «черный ящик». Чем больше вредоносный набор данных похож на набор данных жертвы, тем выше у злоумышленника шанс переобучить атакуемую модель. Вывод атрибута помогает узнать обучающие данные (например, об акценте ораторов в моделях распознавания речи). Успешная атака на восстановление принадлежности показывает, насколько соблюдается конфиденциальность, в частности персональных данных, разработчиками моделей ИИ.

Инверсия модели. На сегодняшний день является наиболее распространенным типом разведочных атак. В отличие от восстановления принадлежности, когда можно всего лишь угадать, был ли пример в наборе обучающих данных, при инверсии модели злоумышленник пытается извлечь из обучающего набора данные в полном объеме. При работе с изображениями извлекается определенное изображение. Например, зная только имя человека, злоумышленник получает его (ее) фотографию. С точки зрения конфиденциальности это большая проблема для любой системы, обрабатывающей персональные данные. Известны также атаки на модели ИИ, которые используются для оказания помощи в лечении в зависимости от генотипа пациента.

Восстановление параметров модели. Цель подобной атаки – определить модель ИИ и ее гиперпараметры для последующих атак типа «уклонение» класса «черный ящик». При этом восстановленные параметры модели используют, чтобы увеличить скорость атак. Одна из первых работ о таких атаках была опубликована в 2013 г. («Взлом умных машин при помощи более умных: как извлечь значимые данные из классификаторов машинного обучения»).

Кроме основных типов атак, выделяют атаки backdoors и trojans. Цели этих атак и типы атакующих различны, но технически они очень похожи на атаки «отравления». Разница заключается в наборе данных, доступных злоумышленнику.

Троянские атаки (trojans). Во время отравления злоумышленники не имеют доступа к модели и начальному набору данных, они могут только добавить новые данные в существующий набор или изменить его. Что касается трояна, то злоумышленники все еще не имеют доступа к начальному набору данных, но у них есть доступ к модели и ее параметрам, и они могут переобучить эту модель, поскольку в настоящее время компании, как правило, не создают свои собственные модели с нуля, а переобучают существующие модели. Например, если необходимо создать модель для обнаружения рака, злоумышленники берут новейшую модель распознавания изображений и переобучают при помощи специализированного набора данных, поскольку отсутствие данных и изображений раковых опухолей не позволяет обучать сложную модель с нуля. Это означает, что большинство компаний-разработчиков загружают популярные модели из интернета. Однако хакеры могут заменить их своими модифицированными версиями с идентичными названиями. Идея трояна заключается в следующем: найти способы изменить поведение модели в некоторых обстоятельствах таким образом, чтобы штатное поведение модели оставалось неизменным. Сначала хакеры объединяют набор данных из модели с новыми входными данными и уже на объединенном наборе переобучают модель. Модификация поведения модели («отравление» и трояны) возможна даже в среде «черного ящика» и «серого ящика», а также в режиме полного «белого ящика» с доступом к модели и

набору данных. Тем не менее главная цель – не только ввести дополнительное поведение, но и сделать это таким образом, чтобы заложенная уязвимость (бэкдор) работала после дальнейшей переподготовки системы добросовестными разработчиками.

«Черный ход» (Backdoor). Идея такой атаки взята от одной из самых старых ИТ-концепций – бэкдоров. При разработке моделей ИИ исследователи закладывают в нее и общий, базовый функционал, и возможность дальнейшего переобучения. С целью маскировки атаки по завершению несанкционированного переобучения модель должна сохранить базовый функционал. Это достижимо за счет того, что нейронные сети, например, для распознавания изображений, представляют собой масштабные структуры, образованные миллионами нейронов. Чтобы внести изменения в такой механизм, достаточно модифицировать лишь небольшой их набор. Еще один фактор, делающий возможным атаку «черного хода», заключается в том, что модели распознавания изображений, например Inception или ResNet, крайне сложны. Они обучены на огромном количестве данных, для чего использовались дорогостоящие вычислительные мощности. Провести аудит и выявить черный ход крайне затруднительно.

Атаки подменой модели машинного обучения. Ресурсами малых и средних компаний создать модели машинного обучения высокого качества практически невозможно. Вот почему многие компании, которые обрабатывают изображения, применяют предварительно обученные нейронные сети крупных компаний. В связи с чем чтобы решить задачу обнаруживать раковые опухоли разработчики могут использовать сеть, доучивая ее, изначально предназначенную для распознавания лиц знаменитостей. Если злоумышленникам удастся взломать сервер, на котором хранятся общедоступные модели (а уровень безопасности общедоступных сервисов невысокий), и загрузить свою собственную модель с интегрированным «черным ходом», модели сохранят свойства, заложенные хакерами даже после переобучения модели добросовестными разработчиками. Например, «черный ход», встроенный в детектор американских дорожных знаков, оставался активным даже после того, как модель была переобучена на идентификацию шведских дорожных знаков вместо американских аналогов. Если владелец не является экспертом, обнаружить эти «черные ходы» практически невозможно. Регулярно появляются методики их обнаружения, но также регулярно возникают новые способы маскировки «черного хода», заложенного в модель.

Классификация атак на методики машинного обучения. Эталонный процесс обучения ИИ предполагает наличие большого набора подготовленных данных, доступ к высокопроизводительным вычислительным ресурсам. Задействованные данные не должны быть личными (приватными), они должны обрабатываться в едином централизованном хранилище. Необходима также фаза стандартного обучения и тонкой настройки гиперпараметров. Однако эти условия в полном объеме тяжело соблюдать на практике. В силу чего для смягчения таких жестких требований были разработаны и приняты в эксплуатацию методики машинного обучения, например трансферное обучение, федеративное обучение, сжатие моделей, многозадачное обучение, метаобучение и обучение на всем жизненном цикле. Они получили широкое распространение даже несмотря на наличие уязвимостей, позволяющих хакерам проводить успешные атаки на разработанные модели.

Многозадачное обучение. Оно повсеместно применяется для решения задач в области классификации изображений, обработки естественного языка и т. п. Даже когда целью обучения модели является выполнение одной задачи, модель обучают в целях выполнения связанных подзадач для улучшения качества и скорости решения главной задачи. Одна из возможных атак – «отравление» набора данных одной задачи и проверка возможности использовать ее выход (результат выполнения) для других задач. Например, жертва хочет обучить модель для определения выражения лица, но из-за нехватки данных решает вспомогательную задачу распознавания лиц при помощи общедоступных наборов данных. Злоумышленник «отравляет» общедоступный набор данных, когда занимается вспомогательной задачей, так чтобы

создать «черные ходы» для всей модели. Безусловно, формирование обучающего изображения для создания бэкдора не является тривиальным вопросом и требует знаний и квалификации злоумышленника. Все атаки на однозадачные модели применимы к многозадачным моделям, однако последние могут подвергаться атакам новых типов. Пример – прогнозирование смены направления для рулевого управления в автомобиле без водителя. Разработчик атакуемой модели рассматривает классификацию характеристик дороги как вспомогательную задачу. Поскольку модель обучена для двух связанных задач, выходные данные классификации характеристик дороги имеют прямую связь с выходными данными задачи прогнозирования направления рулевого управления. Запрашивая ответ от зараженной модели характеристик дороги, злоумышленник задает взаимосвязи между этими заданиями. Хотя злоумышленник может не знать, как изменить входные данные, чтобы воздействовать на рулевое управление, но он может изменить вход – подменить определенную характеристику дороги, которая, в свою очередь, повлияет на прогнозирование рулевого управления. Другими словами, злоумышленник использует задачу A в целях создания задачи состязательного целевого ввода для задачи B. Даже если он напрямую не может атаковать B, то посредством вывода A он сделает это опосредованно.

Машинное обучение в течение жизненного цикла. С непрерывным обучением тесно связаны две концепции:

предположение о том, что все исторические знания доступны и используются для изучения новых задач;

накопление полученных новых знаний.

Первая концепция допускает потенциальное заражение данных при атаках типа Backdoor и исследовательских атаках. Согласно второй концепции, процесс может быть нарушен, поскольку атака не позволит системе сохранять получаемые знания и отработанные задачи. Это тип атаки на доступность, она не дает реализовать подход к обучению в течение жизненного цикла.

Выяснение того, как «черные ходы» и атаки «отравлением» данными могут повлиять на системы обучения, имеет первостепенное значение. Например, если решение справляется с задачей накопления знаний, может ли злоумышленник создать бэкдор для одной задачи и использовать ее для всех других новых задач? Если это возможно, то последствия для безопасности будут катастрофическими.

Также злоумышленники могут атаковать процесс накопления полученных знаний. Один из методов атаки заключается в изучении того, может ли добавление нескольких тщательно созданных обучающих образцов с правильными метками изменить структуру модели так, чтобы она плохо выполнялась в прежних задачах. Злоумышленники таким образом инициируют в модели оптимизацию ретроспективных знаний, цель которой состоит в том, чтобы изменить модель под новую, атакующую задачу, тем самым повредить результаты обучения на старых задачах. Механизмы атаки и защиты, характерные для обучения на протяжении всего жизненного цикла, требуют дополнительных исследований.

§ 2. Большие данные

Понятие. Большие данные (Big Data) – это крайне большой объем структурированных и неструктурированных данных произвольного типа, обрабатываемый в горизонтально масштабируемых информационных системах. Назначение систем Big Data – помогать в принятии решений и инициировать действия на основе анализа цифровой информации. При помощи систем Big Data принимаются решения о необходимости профилактики эпидемий, об изменении полётного графика воздушных судов, о пригодности деталей автомобиля для эксплуатации, о необходимости провести внеплановый ремонт на строительных объектах и многие другие.

История. Определение Big Data появилось в 2008 г. Безусловно, до этого времени существовали методологии анализа информации, однако стоимость хранения и обработки данных была столь велика, что ограничения в ресурсах либо сводили на нет полезность аналитических отчетов из-за низкой скорости их предоставления, либо качество отчетов было столь низким, что они не имели практического применения.

Вместе с тем, объемы данных росли лавинообразными темпами: пользователи социальных сетей генерировали огромные объемы информации, корпорации копили сведения о клиентах, индустриальные предприятия использовали датчики для контроля технологических процессов, в дополнение к ним в широкой эксплуатации появились домашние приборы и автоматизированных системы, которые без участия человека используют интернет, автоматически отсылают информацию о своем состоянии, получают и обрабатывают команды пользователей и тем самым также порождают огромные объемы данных.

Усиливалась и потребность в анализе этих данных – постоянно шел поиск ответа на бизнес-задачи: предсказание потребительского поведения с целью повысить эффективность маркетинговой активности; цифровое моделирование индустриальных объектов, с целью снизить затраты на дорогостоящие испытания; быстрый анализ данных с погодных датчиков для обеспечения безопасности полетов и др.

К 2008 г. технологический прорыв в области микропроцессорных технологий и в производстве систем хранения данных на порядки снизил стоимость хранения и обработки. Это упростило и удешевило доступ к вычислительным ресурсам до недостижимого прежде уровня, что сделало возможным дальнейший прогресс в развитии аналитических систем.

Важнейшей вехой в истории систем класса Big Data является развитие технологии кластеризации, реализующей горизонтальное масштабирование – объединение разрозненных единиц вычислительной техники в общую вычислительную систему с единым управлением.

Повысилась доступность систем Big Data для широкого круга разработчиков программного обеспечения благодаря изменению бизнес-моделей глобальных технологических компаний: появились трансконтинентальные IT-инфраструктуры, позволяющие использовать практически неограниченные вычислительные мощности и системы хранения без первичных инвестиций – на условиях оплаты аренды ресурса с почасовой тарификацией. Такого рода бизнес-модели сняли финансовые ограничения для малых технологических компаний и дали им возможность активно разрабатывать аналитические инструменты для широкого круга потребителей.

Предпосылками активного развития систем Big Data стали:

– рост объема цифровой информации и потребность коммерческих и государственных организаций в результатах ее анализа;

– технологический прорыв в области микроэлектроники;

– деятельность саморегулирующихся сообществ разработчиков программного обеспечения;

– появление новых бизнес-моделей коммерческих организаций, обеспечивающих широкий доступ к вычислительным ресурсам.

Свойства систем Больших данных. Определяющими свойствами, по которым системы анализа и сбора информации относят к классу Big Data, являются объем обрабатываемых данных, их разнородность, возможность горизонтального масштабирования. Выделяют также ряд потребительских свойств системы, такие как скорость обработки данных, потребительская ценность, достоверность и другие.

Основное свойство систем Big Data – обработка крайне больших массивов данных объем которых постоянно и с большой скоростью увеличивается. Речь идет о данных миллионов финансовых операций, десятках миллионов переходов на веб-сайтах интернет-магазинов, сотен миллионов значений датчиков погоды, снимающих показания по всему миру, миллиардов записей пользователей на персональных страничках социальных сетей, десятков миллиардов действий пользователей поисковых систем и мобильных приложений.

Разнородность данных – это возможность обработки в системе разнообразных типов данных и их структур. Это свойство характеризует возможность системы проводить анализ неструктурированных данных: «сырых» текстов, медиафайлов – аудиофайлов, видеофайлов и файлов изображений; слабоструктурированной информации: например, новостных каналов, электронных таблиц; структурированных данных реляционных СУБД и данных, полученных в виде структурированного ответа на запрос на специализированных языках работы с данными.

Скорость обработки означает возможность системы принимать и обрабатывать данные в необходимом объеме за ограниченное время. Многие системы Big Data предназначены для сбора информации из большого количества источников в режиме реального времени и их анализа также в режиме реального времени. Пример – медицинские устройства, предназначенные для сбора данных о здоровье и мониторинга состояния пациентов. Предназначение и важность этих систем требует собирать, анализировать эти данные и затем передавать результаты медицинскому персоналу за минимальное количество времени. Необходимость реализации интернета вещей медицинского оборудования создает запрос на обеспечение высокой скорости передачи и обработки данных.

Возможность горизонтального масштабирования – это возможность увеличить производительность и емкость системы путем подключения аппаратных или программных ресурсов таким образом, чтобы они работали как единое логическое целое. Этот механизм также называется кластеризацией вычислительных систем. Если кластеру требуется больше ресурсов для повышения производительности, обеспечения более высокой доступности, администратор может масштабировать вычислительный ресурс, добавляя в кластер больше серверов и/или хранилищ данных.

Поддержка горизонтальной масштабируемости подразумевает возможность увеличивать количество и заменять узлы «на лету», не значительно прерывая функционирование системы. Например, распределенная система хранения данных Cassandra, включает сотни узлов, размещенных в различных дата-центрах. Поскольку оборудование масштабируется горизонтально, Cassandra является отказоустойчивой и не имеет одной критичной точки отказа.

Еще одно преимущество заключается в том, что теоретически производительность горизонтально масштабируемых систем не ограничена. Производительность зависит только от количества узлов, подключённых к системе. Это драматически отличает системы с горизонтальным масштабированием от многих традиционных систем обработки данных в которых при увеличении вычислительного ресурса производительность системы в целом значимо не растет. Это приводит к серьезнейшим функциональным ограничениям традиционных систем.

Таким образом, поддержка горизонтального масштабирование обеспечивает возможность роста объемов данных и их анализа, при котором результат анализа не теряет своей полезности за время расчета. Например, оценка ситуации на дороге для системы автопилоти-

рования должна быть рассчитана за доли секунды – в противном случае, такая оценка просто не нужна.

Примером технологического решения реализации горизонтального масштабирования является Hadoop – проект фонда Apache Software Foundation. Hadoop это библиотека для разработки программного обеспечения предназначенная для создания и выполнения программ, работающих на кластерах из сотен и тысяч узлов. Hadoop – библиотека с открытым т. е. бесплатно распространяемым и дающим возможность менять под свои нужды, программным кодом, практический инструмент разработчиков и архитекторов IT-инфраструктур.

Потребительская ценность системы относится к ключевым потребительским свойствам систем больших данных. Ценность системы – это ее пригодность для получения практически применимых выводов и принятия решений.

Наличие огромных объемов данных необходимо для анализа и, безусловно, существует прямая связь между данными (информации представленной в цифровом виде) и знаниями (достоверными представления о предметах и явлениях действительности), но из наличия взаимосвязи не следует означает, что в Big Data всегда есть знания и они могут быть извлечены. Если на их основании данных нельзя сделать полезных выводов, вся система не будет иметь ценности.

Технологии анализа позволяют автоматически находить в потоках данных зависимости, которые не в состоянии выявить человек, такие как, например, влияние атмосферного давления на покупку молочной продукции. Однако, если атмосферное давление за анализируемый период было приблизительно одно и тоже, собранные данные не будут содержать знания о существующей взаимосвязи и ценность системы будет нулевой.

Важной частью инициатив в области больших данных является понимание того, каковы затраты и выгоды от сбора и анализа данных. Необходим обоснованный прогноз, что в конечном счете получаемый результат анализа принесет конкретную пользу.

Качество данных и достоверность системы – свойства, которые показывают, что данные были получены из доверенных источников, в неискаженном виде, по доверенным каналам.

В случае, если анализ проводится на основе искаженных данных, выводы и решения не будут корректными. Например, сообщения в Twitter содержат хэш-теги, сокращения, опечатки, указание личных мнений и т. д. Таким образом данные не являются качественными т. к. искажение текста может привести к искажению заложенного в сообщения смысл. Впрочем, Twitter вызывает сомнения и как источник изначально достоверных данных. А если невысока исходная достоверность их сбор и анализ бесполезны.

Следующий пример относится к использованию данных систем глобальной навигации: часто GPS рассчитывает недостоверные координаты местоположения, особенно при размещении приемника в городских районах. Спутниковые сигналы теряются и искажаются, когда они отражаются от высоких зданий или других сооружений. Как единственный источник данных спутники сами по себе недостоверны. Для повышения качества данные о местоположении следует объединить со сведениями из других источников данных, например, с данными акселерометра или сигналами вышек сотовой связи.

Технологии систем Больших данных. Базовыми технологиями систем Big Data являются технологии сбора, анализа и представления данных.

Технологии сбора:

– смешение и интеграция данных (data fusion and integration) – набор техник для интеграции разнородных данных из разнообразных источников в целях анализа (например, обработка естественного языка, включая анализ настроения говорящего – тональности высказывания);

– краудсорсинг – привлечение широкого и заранее не определенного круга лиц для повышения ценности данных без вступления в трудовые отношения с этими лицами.

Технологии анализа:

– прогнозная аналитика – выявление закономерностей в имеющихся данных, помощь в оценке происходящих процессов и прогнозирование дальнейших событий;

– классификация – отнесение объекта к группе по определенному признаку;

– кластерный анализ – автоматизированное формирование сравнительно однородных групп и отнесение к ним объектов (например, по ряду поведенческих факторов можно выяснить намерение человека украсть что-либо: на основании схемы перемещения покупателя по торговому центру определить, что его поведение не является обычным и предотвратить кражу);

– регрессионный анализ – выявление вероятных последствий (например, можно смоделировать дорожные аварии как последствия сочетания скорости, дорожных условий, погоды, трафика);

– обучение ассоциативным правилам – определение непрямых зависимостей (например, рост количества покупок спичек при покупке мяса, но только в случае приобретения маринада);

– пространственный анализ (Spatial analysis) – класс методов с использованием топологической, геометрической и географической информации для обоснования градостроительных решений;

– машинное обучение – применение программ, которые независимо от человека самомодифицируются на основании обрабатываемых данных;

– классический статистический анализ;

– получение комплексных прогнозов на основе базовых моделей;

– создание самомодифицируемых систем, сходных по структуре с головным мозгом человека.

Технологии представления данных. К ним относится визуализация аналитических данных – представление информации в виде рисунков, диаграмм с использованием интерактивных возможностей и анимации как для получения результатов, так и для применения в качестве исходных данных в целях дальнейшего анализа человеком.

Конец ознакомительного фрагмента.