УДК 004.89, 338.3
ББК 32.813, 65.29
К30

Кацов И.
К30 Искусственный интеллект на предприятии: теория и практика / пер. с англ. В. С. Яценкова. – М.: ДМК Пресс, 2024. – 710 с.: ил.

ISBN 978-5-93700-277-8

В книге подробно рассказано, как можно улучшить бизнес-процессы компании с помощью методов искусственного интеллекта, а также использовать их в сочетании с традиционными подходами к аналитике и оптимизации. Рассмотрены основные концепции автоматизации принятия корпоративных решений, глубокого обучения, генеративного искусственного интеллекта и методов обучения с подкреплением; описаны прикладные рецепты для клиентской аналитики и персонализации, а также методы прогнозирования спроса, оптимизации цен и управления запасами; приведены решения для обнаружения аномалий и визуального контроля, помогающие улучшить производственные и транспортные операции.

В репозитории GitHub доступны прототипы кода на Python, помогающие понять детали реализации.

Издание адресовано специалистам по корпоративным данным, аналитикам, а также будет полезно руководителям предприятий: директорам по производству (COO), маркетингу (CMO), информатизации (CIO) и др.

УДК 004.89, 338.3
ББК 32.813, 65.29

ISBN 979-8-218-16967-1 (англ.)
ISBN 978-5-93700-277-8 (рус.)

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ

Роль автоматизации и оптимизации на основе данных в операционной деятельности предприятия возрастала на протяжении многих десятилетий, но за последние десять лет диапазон вариантов эффективного использования автоматизированных систем значительно расширился благодаря появлению методов глубокого обучения. Эти разработки сформировали разнообразную среду методов принятия решений и автоматизации, включая традиционные эконометрические модели и алгоритмы оптимизации, специализированные методы машинного обучения для компьютерного зрения и обработки естественного языка, которые, однако, часто могут быть адаптированы к другим областям, а также новые методы, такие как как глубокое обучение с подкреплением, имеющие ограниченное применение в корпоративной практике. В этой книге будет показано, как предприятие может извлечь максимальную выгоду из сочетания традиционных методов моделирования, оптимизации и симуляции с методами глубокого обучения и обучения с подкреплением за счет интеллектуальной автоматизации широкого спектра операций предприятия, включая маркетинг, управление цепочками поставок и контроль производства. Назначение этой книги – представить инженерный базис, а также набор практических рецептов применения комбинаций этих методов в условиях реального предприятия.

ЦЕЛЕВАЯ АУДИТОРИЯ

Эта книга адресована специалистам по данным и менеджерам по аналитике и призвана сформировать у них навык системного подхода к решению задач принятия корпоративных решений и оптимизации с использованием методов глубокого обучения, обучения с подкреплением и вероятностного программирования. Работая над книгой, я задался целью предложить систематическую методику перевода различных вариантов деятельности предприятий в количественные, статистические и оптимизационные задачи и их декомпозиции на задачи машинного обучения. В начале книги дается обзор базовых компонентов системы. Подробные описания использования моделей и алгоритмов для конкретных случаев будут рассмотрены позже, но я не стремился обеспечить детальное изложение теории машинного обучения и ее математических основ. Ожидается, что читатель знаком с основными концепциями науки о данных и машинного обучения, а также имеет практический опыт статистического моделирования, включая как традиционные методы, так и методы глубокого обучения.

Книга также будет полезна специалистам по науке о данных и машинному обучению, имеющим опыт работы в биоинформатике, физике или других

областях, не связанных с типичной операционной деятельностью предприятия, и желающим узнать о специализированных методах моделирования для маркетинга, цепочек поставок и производственных приложений.

СТРУКТУРА КНИГИ И РЕКОМЕНДАЦИИ ПО ПРОЧТЕНИЮ

Я подхожу к задаче построения корпоративных ИИ-решений с точки зрения системного проектирования, рассматривая алгоритмы машинного обучения главным образом как готовые компоненты и уделяя особое внимание адаптации общих методов к конкретным сценариям использования на предприятии. Первые три главы посвящены описанию методики декомпозиции задач предприятия на задачи машинного обучения и оптимизации, а также рассмотрению основных категорий алгоритмов машинного обучения, необходимых для решения таких задач. Однако выбор и категоризация алгоритмов и методов несколько отличаются от канонической категоризации, представленной в большинстве учебников по машинному обучению, поскольку мы сосредоточимся исключительно на корпоративных приложениях. Затем читатель познакомится с рядом рецептов (R1–R14) для конкретных случаев применения в области маркетинга, цепочки поставок и производства. В большинстве рецептов я использую следующую трехуровневую структуру, чтобы охватить как теоретические, так и практические аспекты решений:

- *варианты решений*: в каждом рецепте определяется бизнес-задача и обсуждается несколько вариантов решения. Некоторые решения не требуют существенного изменения общих моделей или алгоритмов, и рецепт сфокусирован в основном на практических аспектах, таких как интеграция и эконометрические соображения. Другие решения требуют разработки специализированных алгоритмов, и они будут описаны со всеми математическими деталями;
- *прототипы*: в каждом рецепте показан процесс разработки одного или несколько базовых прототипов, иллюстрирующих подход и основные свойства решения. Эти эталонные реализации обычно используют синтетические данные или симуляторы, чтобы избежать сложностей, связанных с наборами реальных данных. Обычно в рецепте описано, как работает прототип и как результаты доказывают жизнеспособность решения, но я старался не загромождать книгу подробностями низкоуровневой реализации; вместо этого в каждом рецепте приведены ссылки на соответствующие блокноты в репозитории исходного кода, чтобы читатель мог при необходимости глубже изучить реализацию;
- *подробные примеры*: для некоторых рецептов разработаны более полные эталонные реализации с применением крупных выборок данных, созданных на основе статистических параметров реальных наборов данных. Эти реализации помогают раскрыть проблемы, которые не проявляют себя в прототипах меньшего масштаба. Как и в случае с про-

тотипами, я не рассматриваю эти реализации на уровне исходного кода, а даю ссылки на репозиторий с полными блокнотами.

Книгу можно читать последовательно, чтобы систематически изучать идеи, заложенные в основу корпоративного ИИ, включая его основные компоненты и категории решений. Однако рецепты в основном независимы, и читатели, знакомые с основами глубокого обучения, могут бегло просмотреть первые три главы и далее изучать рецепты в любом порядке в соответствии со своими потребностями и приоритетами.

ПРИМЕРЫ ИСХОДНОГО КОДА И ДАННЫХ

Примеры реализации и прототипы, упомянутые в этой книге, а также несколько дополнительных моделей выпущены в виде проекта с открытым исходным кодом под названием TensorHouse. Этот проект доступен на https://github.com/ikatsov/tensor-house. Там создана специальная ветка `book-enterprise-ai-edition-2.1` с версией кода, совместимой с данной книгой. Я использую TensorFlow в качестве основной платформы для моделей глубокого обучения и несколько других платформ и библиотек для вспомогательных операций и специализированных функций.

Рецепты, представленные в этой книге, и примеры реализации дополняют друг друга. Рецепты предлагают всесторонний анализ бизнес-задач и вариантов их решения, но лишь краткое изложение фактических реализаций. Демонстрационные блокноты на GitHub содержат подробные пошаговые инструкции по реализации определенных решений, но не дублируют весь анализ и теоретические детали, представленные в книге.

ЧТО НОВОГО ВО ВТОРОМ ИЗДАНИИ?

Второе издание включает в себя два крупных обновления и множество мелких изменений. Во-первых, добавлен обширный новый материал по генеративному ИИ, охватывающий теоретические основы (глава 3), приложения для языкового моделирования (рецепт R7) и решения для генерации изображений (рецепт R8). Во-вторых, часть IV, посвященная управлению доходами и запасами, существенно переработана, и в нее добавлена отдельная глава, посвященная прогнозированию спроса (рецепт R9). Кроме того, существенно обновлена глава 2, посвященная прогнозным моделям и основам глубокого обучения; добавлены два приложения, предоставляющие исчерпывающий справочник по функциям потерь и показателям оценки, используемым в корпоративных приложениях; доработаны и улучшены все иллюстрации.

Часть I

БАЗОВЫЕ КОМПОНЕНТЫ

Первая часть этой книги посвящена методике преобразования задач автоматизации принятия производственных решений в задачи машинного обучения. Главу 1 мы начинаем с рассмотрения типичных уровней принятия решений в рамках деятельности предприятия и определения основных концепций, которые применимы к широкому спектру прикладных ситуаций. В главе 2 разрабатываем набор инструментов для изучения сопоставлений между объектами, их атрибутами и траекториями, который позволяет нам выводить скрытые свойства и прогнозировать будущие состояния объектов. В главе 3 мы дополняем этот набор инструментов методами создания сложных объектов, таких как изображения. Наконец, в главе 4 обсуждаем методы автоматизации принятия решений и управления объектами.

1

АВТОМАТИЗАЦИЯ РЕШЕНИЙ И ПРОЦЕССОВ В ДЕЯТЕЛЬНОСТИ ПРЕДПРИЯТИЯ

Корпоративный *искусственный интеллект* (ИИ) можно неофициально определить как набор методов улучшения операционной деятельности предприятия с использованием статистического обучения и вероятностных рассуждений. Этот набор очень обширен и включает в себя методы совершенствования стратегических решений на уровне всего предприятия. Среди них прогнозирование спроса, методы оптимизации решений на уровне отдельных бизнес-процессов, такие как таргетирование рекламных акций и управление запасами, а также методы, помогающие автоматизировать или оптимизировать отдельные транзакции, такие как модели обнаружения объектов и системы управления диалогами.

Хотя концепция корпоративного ИИ относительно нова, можно утверждать, что по своей сути она так же стара, как и сама концепция предприятия. Исторически сложилось так, что первой категорией задач, которые решали с использованием методов, основанных на данных, было принятие стратегических решений. Идея о том, что финансовое состояние и будущую траекторию экономического субъекта можно оценить с помощью агрегированных финансовых отчетов, была хорошо известна с древних времен. В этом смысле бухгалтеры были первыми специалистами по корпоративным данным. Собирая и агрегируя финансовые записи, бухгалтер создает краткое количественное представление предприятия в пространстве определенных показателей, таких как прибыль и доходы, и это пространство затем служит для оценки финансовых показателей предприятия, прогнозирования будущих результатов и составления отчетов, на которых основываются управленческие решения. Хотя может показаться, что связать бухгалтерский учет с ИИ – это слишком сложно, скоро вы убедитесь, что даже основные количественные методы, используемые в бухгалтерском учете, торговле акциями

и инвестициях, можно органично и безопасно объединить с методами, которые обычно свойственны современному ИИ.

Массовый переход на второй уровень автоматизации принятия решений, то есть оптимизации тактических решений на уровне отдельных процессов внутри предприятия, произошел в основном в последней трети XX века благодаря появлению доступных и достаточно мощных компьютеров. Разработки того периода были сосредоточены преимущественно на решении задач численной и комбинаторной оптимизации в управлении цепочками поставок, транспортировке и производстве. Некоторые из этих приложений также включали статистический анализ корпоративных данных, но внедрение методов, основанных на данных, было ограничено низким уровнем цифровизации как бизнеса, так и потребителей.

Следующий уровень интеллектуальной автоматизации был достигнут в середине 2010-х годов, главным образом благодаря трем факторам. Началом послужила комплексная цифровизация корпоративной среды, за которой последовало массовое внедрение стратегий больших данных, вынудивших все корпоративные процессы непрерывно генерировать подробные, статистически анализируемые данные. Вторым фактором стала цифровизация потребительской среды, которая позволила персонализировать и оптимизировать продукты, услуги и их представления в реальном времени. Третьим (и последним) фактором стал революционный прогресс в статистических методах, особенно в глубоком обучении, позволивший понимать текстовые и визуальные данные с помощью программных систем. Это, в свою очередь, привело к появлению широкого спектра новых типов автоматизации, связанных с *обработкой естественного языка* (natural language processing, NLP) и использованием *компьютерного зрения* (computer vision, CV).

Эта книга в основном посвящена методам автоматизации мелкомасштабных решений и процессов в средах с большим объемом данных, что соответствует последнему уровню в приведенной выше классификации. Однако прежде чем мы углубимся в детали этих методов, необходимо сформировать понимание базовой методики, которая поможет нам правильно планировать и внедрять решения по автоматизации анализа данных и принятия решений внутри предприятия. Мы должны рассматривать проектирование отдельных компонентов с точки зрения системы, под правильным углом и в правильном контексте.

1.1 МЕТОДИКА СЦЕНАРНОГО ПЛАНИРОВАНИЯ

Принято считать, что корпоративные системы и механизмы создают преимущественно с учетом конкретных бизнес-целей, а степень успеха обычно измеряют с помощью *ключевых показателей эффективности* (key performance indicator, KPI). Например, можно разработать новый алгоритм рекомендации продуктов с целью повышения коэффициента конверсии сеанса с 3 до

3.5 %; можно создать систему оптимизации цен, которая увеличит прибыль на 200 млн долл., а на сборочной линии установить новую систему компьютерного зрения, чтобы снизить уровень брака на 20 %. Это одна из самых основополагающих и известных парадигм в мире бизнеса. Однако ее практическая реализация весьма сложна.

Одна из проблем заключается в том, что многие улучшения и действия влияют на несколько KPI одновременно, и при положительном влиянии на одни показатели влияние на другие может быть отрицательным. Например, распродажи по сниженной цене могут увеличить товарооборот, но снизить прибыль, увеличение уровня складских запасов может улучшить доступность продукции и качество обслуживания клиентов, но увеличить затраты на хранение, а оптимальные для прибыли цены могут разрушить бизнес, которому необходимо увеличить свою долю на рынке, чтобы стать устойчивым. Анализ таких компромиссов и принятие решений по ним являются сложной задачей, а преобразование их в формальные задачи оптимизации – тем более.

Вторая проблема заключается в долгосрочном характере бизнес-операций, что иногда затрудняет определение и измерение одного ключевого показателя эффективности. Например, распродажи широко применяются розничными продавцами и производителями и, как известно, являются эффективным способом увеличения объема продаж, но прирост сбыта в период промоакций часто происходит за счет будущих продаж. Это особенно справедливо в отношении расходных материалов, таких как бумажные полотенца: рекламные акции часто побуждают потребителей покупать и накапливать большее количество продуктов, а затем ждать следующей распродажи. Это делает краткосрочные измерения роста продаж неточными или вводящими в заблуждение, в то время как долгосрочные измерения сложны и тоже не всегда достоверны, поскольку все другие части окружающей среды независимо и непредсказуемо меняются, искажая наблюдения.

Третья проблема – измеримость ключевых показателей эффективности. В предыдущем примере, касающемся стимулирования продаж, долгосрочные последствия являются одним, но не единственным фактором, который способен лишить измерения достоверности. Рекламные акции и изменения цен могут заставить потребителей переключаться с одного продукта или поставщика на другого, создавая эффекты перекрестных продуктов и перекрестных розничных продаж. Большинство продавцов знают о таких эффектах и хорошо понимают, что показатели роста продаж отдельных продуктов могут вводить в заблуждение, но более полные и точные измерения затруднены из-за более высоких требований к данным и сложности реализации. Во многих практических случаях имеет смысл разделить одну цель с множеством внутренних факторов на несколько показателей эффективности, которые можно отслеживать отдельно, и это возвращает нас к первой проблеме, связанной с множеством конфликтующих ключевых показателей эффективности.

Наконец, мы можем отнести к серьезным проблемам оценку решения. Разработка решения, целью которого является улучшение определенных показателей, обычно нуждается в инструментах оценки его производительности до внедрения в производство и сбора фактических результатов. Эта

проблема унаследовала все ранее упомянутые сложности, которые по сути принадлежали области описательной аналитики, и добавила к ним элементы прогнозирования и опережающего анализа.

Из сказанного следует, что для преобразования задач предприятия в задачи оптимизации и автоматизации решений обычно необходимо учитывать множество объектов, показателей, их корреляций и совместной динамики. Далее мы попытаемся объединить эти основные идеи в более формальную структуру, основанную на следующих понятиях.

Объект (entity*, сущность*): при автоматизации или оптимизации какого-либо аспекта предприятия мы обычно сосредоточиваемся на повышении производительности одного или нескольких объектов, которыми могут быть предприятие в целом, бизнес-подразделение, местоположение, продукт или клиент.

Сценарий (scenario): определяет действие или последовательность действий, которые мы потенциально можем выполнить для достижения определенных улучшений. Обычно мы можем выбирать между несколькими вариантами сценариев, включая универсальное бездействие.

Пространство полезности (utility space): объекты, с которыми мы работаем, могут быть описаны с использованием одной или нескольких метрик (KPI), а пространство, охватываемое измерениями метрики, принято называть пространством полезности. Например, в традиционной бизнес-аналитике обычно используются такие пространства полезности, как *Доход×Маржа* и *Риск×Вознаграждение*. Мы также будем называть пространства полезности с двумя или более измерениями *картами полезности* (utility map), или *картами ценностей* (utility value).

Траектория (trajectory): выполнение сценария заставляет объекты двигаться в пространстве метрик – в идеале в том направлении, которое выгодно бизнесу. Следовательно, сценарий оставляет след в пространстве полезности, который мы будем называть траекторией.

Граница Парето (Pareto frontier): если метрики, используемые для построения пространства полезности, находятся в обратной зависимости (например, стоимость и качество), мы обычно можем выбирать между несколькими сценариями, образующими различные компромиссы. Однако в такой ситуации невозможно улучшить все показатели одновременно, поэтому набор наилучших возможных компромиссов сформирует границу Парето, которую можно представить как поверхность в пространстве полезности. Однако мы можем сдвинуть границу Парето, если найдем возможность одновременного улучшения всех показателей пространства полезности за счет изменения некоторых базовых факторов.

Эти понятия проиллюстрированы на рис. 1.1, где организация движется в двумерном пространстве полезности, и рассматриваются три альтернативных сценария воздействия, которые приводят к трем различным результатам, образующим границу.

На данный момент наша методика сценарного планирования довольно абстрактна, и не совсем ясно, как применить ее к практическим задачам и связать с методами *машинного обучения* (machine learning, ML). Этот пробел будет постепенно закрыт в следующих разделах, где мы обсудим более кон-

кретные примеры для различных типов задач, а затем методы машинного обучения, необходимые для реализации этого подхода. Но цель данной главы заключается в том, чтобы рассмотреть различные примеры использования машинного обучения с точки зрения сценарного планирования, а более строгое описание теоретических основ будет представлено позже.

Рис. 1.1. Основные понятия
методики сценарного планирования

1.1.1. *Стратегия: предприятие в целом*

Начнем с примеров стратегического анализа, в которых стоит задача изучить траекторию развития всей компании. Этот тип анализа обычно используется венчурными инвесторами для оценки стартапов и портфельными менеджерами для оценки публичных компаний. Вероятно, это максимально возможный уровень, на котором можно рассматривать применение методов автоматизации принятия решений.

Представьте, что нам нужно оценить компанию на ранней стадии развития, чтобы принять решение о том, стоит ли в нее инвестировать. Один из ключевых вопросов оценки заключается в том, насколько хорошо компания и ее продукт соответствуют рынку: выглядит ли компания готовой к росту или видны признаки сопротивления рынка и замедления развития (Hsu, 2019). В определенной степени на этот вопрос можно ответить, используя количественные методы, а результаты определяют дальнейшие действия как инвесторов, так и менеджмента компании.

Мы можем начать процедуру оценки с так называемого *анализа факторов роста* (growth accounting), который фокусируется на анализе компонентов дохода и эволюции этих компонентов с течением времени. В сценарном планировании мы определяем объекты как различные категории доходов, пространство полезности – как единственное измерение, выражаемое в долларах, а траектории – как эволюцию категорий доходов. Для анализа разложим выручку предприятия за период времени t следующим образом:

$$\begin{aligned} \textit{Выручка}(t) = {} & \textit{Прежняя}(t) + \textit{Новая}(t) + \textit{Возвращенная}(t) \\ & + \textit{Расширенная}(t), \end{aligned} \tag{1.1}$$

где *Прежняя* часть соответствует доходу от существующих клиентов, сохранившихся с предыдущего периода времени, *Новая* часть поступает от клиентов, которые были привлечены в период времени t, *Возвращенная* часть поступает от клиентов, которые ушли в прошлом, но вернулись в период времени t, а *Расширенная* часть – это дополнительный рост дохода от существующих клиентов сверх оставшейся выручки. Далее разложим прежнюю часть выручки следующим образом:

$$\textit{Прежняя}(t) = \textit{Выручка}(t-1) - \textit{Отток}(t) - \textit{Сокращение}(t), \tag{1.2}$$

где *Отток* – это доход, потерянный из-за клиентов, которые были активны в период $t-1$, но стали неактивными в период t, а *Сокращение* – это сокращение дохода от клиентов, которые остались активными. Например, у компании с двумя клиентами, которые потратили 100 и 200 долл. в период времени $t-1$ и 150 и 180 долл. в период t соответственно, прежняя часть выручки составит 280 долл. (100 + 180), расширенная часть – 50 долл. (150 – 100), и из них нужно вычесть сокращение в размере 20 долл. (200 – 180). Теперь мы можем объединить уравнения 1.1 и 1.2 следующим образом:

$$\begin{aligned} \textit{Выручка}(t) - \textit{Выручка}(t-1) = {} & \textit{Новая}(t) + \textit{Расширенная}(t) \\ & + \textit{Возвращенная}(t) - \textit{Отток}(t) \\ & - \textit{Сокращение}(t). \end{aligned} \tag{1.3}$$

Поскольку мы используем простое одномерное пространство полезности, то можем визуализировать траектории вышеупомянутых компонентов дохода, используя обычные графики временных рядов. Рассмотрим примеры, соответствующие двум различным компаниям, показанным на рис. 1.2.

На рис. 1.2a показана компания, которая быстро растет с точки зрения доходов и имеет относительно низкие темпы оттока и сокращения. Эта модель типична для B2B-бизнеса, основанного на подписке, который хорошо умеет увеличивать доходы от существующих клиентов и поддерживать положительный рост даже без привлечения новых клиентов. На рис. 1.2b показана компания, которая также быстро растет с точки зрения доходов, но ее компоненты оттока и сокращения гораздо более значительны по сравнению с первым примером. Эта модель типична для компаний, которые продают дискреционные продукты[1] B2C и нуждаются в постоянном притоке новых клиентов для поддержания роста.

Обзорный анализ и прогнозирование траекторий доходов в приведенных выше примерах дают инвесторам и руководителям рекомендации относительно того, какие сценарии им следует планировать. Например, похоже, что первой компании потребуется расширить свои команды по продажам и работе с клиентами, тогда как второй компании, возможно, придется ра-

[1] Второстепенные продукты, без которых потребитель может обойтись. – *Прим. перев.*

ботать над сокращением затрат на привлечение клиентов, чтобы сохранить устойчивость. Для решения этих проблем второго уровня, выявленных в результате обзорного анализа, подходят привычные методы, основанные на данных, но благодаря меньшему объему и масштабу задач можно применить и более сложные инструменты. Например, задачу минимизации затрат на привлечение клиентов можно решить с помощью моделей таргетинга и персонализации, которые мы разработаем в следующих главах.

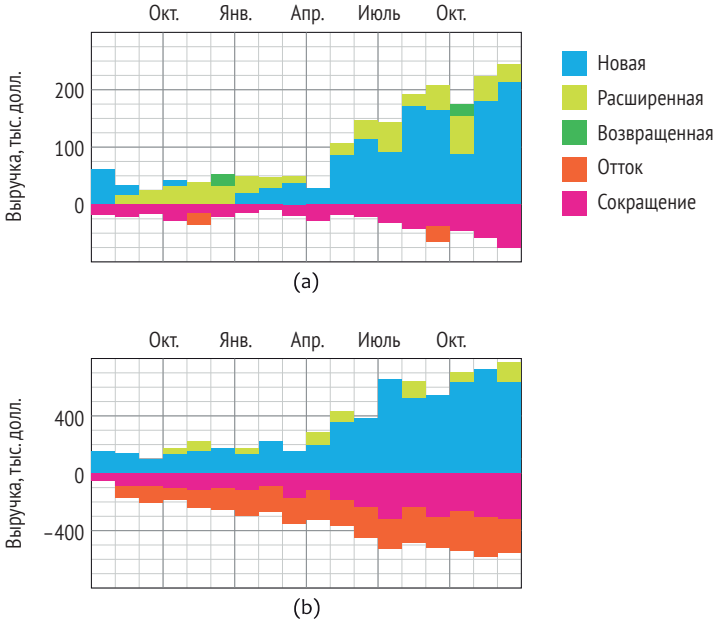

Рис. 1.2. Два примера анализа роста показателей компании

Второй вопрос, который возникает при оценке соответствия компании и ее продукции рынку, – как выглядит внутренняя структура клиентской базы и какова динамика отдельных компонентов. Возьмем когорты клиентов, приобретенных в разные периоды времени, в качестве объектов, а совокупную *пожизненную стоимость клиента*[1] (lifetime value, LTV) – в качестве одномерного полезного пространства. Эта схема проиллюстрирована на рис. 1.3: мы группируем клиентов в ежемесячные когорты на основе дат их приобретения и строим график изменения совокупного дохода с течением времени для каждой когорты в зависимости от ее возраста.

Этот метод, называемый *когортным анализом* (cohort analysis), помогает профилировать динамику отдельных когорт, а также траекторию компании в отношении качества клиентской базы. Пример на рис. 1.3 демонстрирует сильную деградацию клиентской базы в том смысле, что новейшие когор-

[1] Совокупная прибыль компании, получаемая от одного клиента за все время сотрудничества с ним. – *Прим. перев.*

ты (например, приобретенные в августе и сентябре) имеют гораздо худшие траектории в пространстве LTV по сравнению с более старыми когортами (например, приобретенными в январе и феврале). Мы также видим, что траектории самых старых когорт вначале резко растут, что говорит о том, что вновь приобретенные клиенты раньше имели склонность увеличивать использование продукта после приобретения, возможно, из-за обновлений или перекрестных продаж. Однако самые новые когорты не демонстрируют такого поведения и накапливают LTV более линейным образом. На практике такая динамика может быть вызвана чрезмерно агрессивной стратегией привлечения клиентов, которая стремится к количеству в ущерб качеству и долгосрочной устойчивости.

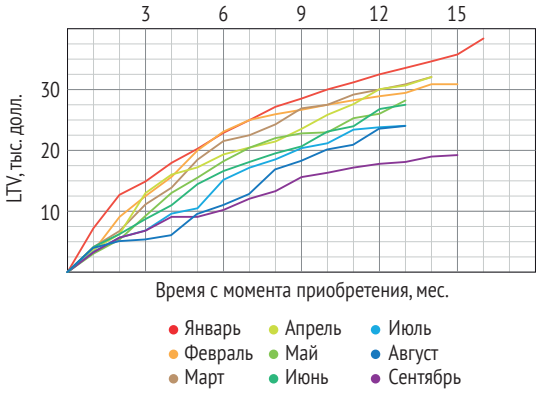

Рис. 1.3. Пример когортного анализа

Когортный анализ можно дополнить другими полезными метриками, характеризующими динамику клиентской базы. Пример этого показан на рис. 1.4, где мы добавили параметр *коэффициента удержания* (retention rate), чтобы визуально и количественно оценить, насколько хорошо компания удерживает ранее приобретенных клиентов, что является ключевым показателем для бизнеса, основанного на подписке. Подобно анализу факторов роста, когортный анализ указывает нам на сценарии, которые необходимо доработать, – например, оптимизировать затраты на привлечение клиентов.

И анализ факторов роста, и когортный анализ – это всего лишь методы *бизнес-аналитики* (business intelligence, BI). Однако можно применить более продвинутые методы, такие как регрессионный анализ, чтобы понять факторы, которые направляют объекты по их траекториям, прогнозирование временных рядов для точной оценки будущего положения объектов в пространстве полезности, обнаружение аномалий для выявления аномальных событий и кластеризацию для обнаружения полезных группировок объектов. Однако на практике на этом уровне анализа не всегда возможно извлечь выгоду из передовых методов, поскольку на изучаемые нами процессы влияет большое количество сложных факторов, начиная от макроэкономики и заканчивая управленческими предубеждениями, а само сценарное планиро-

вание обычно требует глубокого знания предметной области и человеческих суждений. С другой стороны, нельзя недооценивать важность количественного анализа верхнего уровня в общей стратегии предприятия по внедрению ИИ. Наличие правильно выбранных объектов, показателей и перспектив соответствия рынку, потоков доходов и клиентской базы помогает сделать так, чтобы разработка более сложных инструментов приносила бизнесу более ценные результаты.

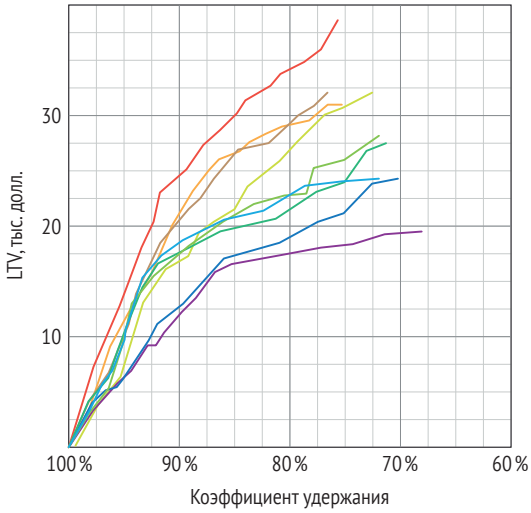

Рис. 1.4. Пример когортного анализа в пространстве полезности, включающем LTV и коэффициент удержания. Это те же когорты, что и на рис. 1.3

1.1.2 *Тактика: подразделения, услуги и продукты*

В предыдущем разделе мы рассмотрели несколько примеров того, как объекты уровня предприятия могут быть изучены через призму методики сценарного планирования. Теперь обратимся к более мелким сущностям, таким как отдельные бизнес-процессы, местоположения и продукты. Меньший масштаб обычно допускает более высокую степень автоматизации принятия решений, поэтому мы можем определять более формальные и автономные задачи оптимизации и использовать более совершенные статистические методы для их решения.

Начнем с того, что поднимем понятие соответствия рынку на следующий уровень детализации и рассмотрим пример компании с относительно большим портфелем продуктов, например крупного производителя или розничного продавца. Чтобы управлять портфелем, компании необходимо проанализировать, как позиции и траектории движения отдельных продуктов совпадают с общей финансовой траекторией компании, а затем разработать правильные стратегии для этих продуктов и других связанных с ними объектов, таких как линейки продуктов, категории, местоположения и бизнес-подразделения. Имеет смысл начать с построения карты полезности,

показывающей, как продукты движутся на рынке и каково их значение для компании. В качестве измерений пространства полезности выберем общий доход от продукта, чтобы оценить эффективность продукта на рынке, и валовую прибыль, чтобы измерить его ценность для компании. Конечно, существует множество альтернативных показателей. Например, можно выбрать долю рынка или объем продаж в качестве меры рыночной эффективности. Теперь для каждого продукта мы сможем визуализировать его прошлую траекторию, начиная с даты запуска продукта, и спрогнозировать его будущую траекторию, используя какую-либо прогнозную модель. Пример такой карты полезности показан в верхней части рис. 1.5.

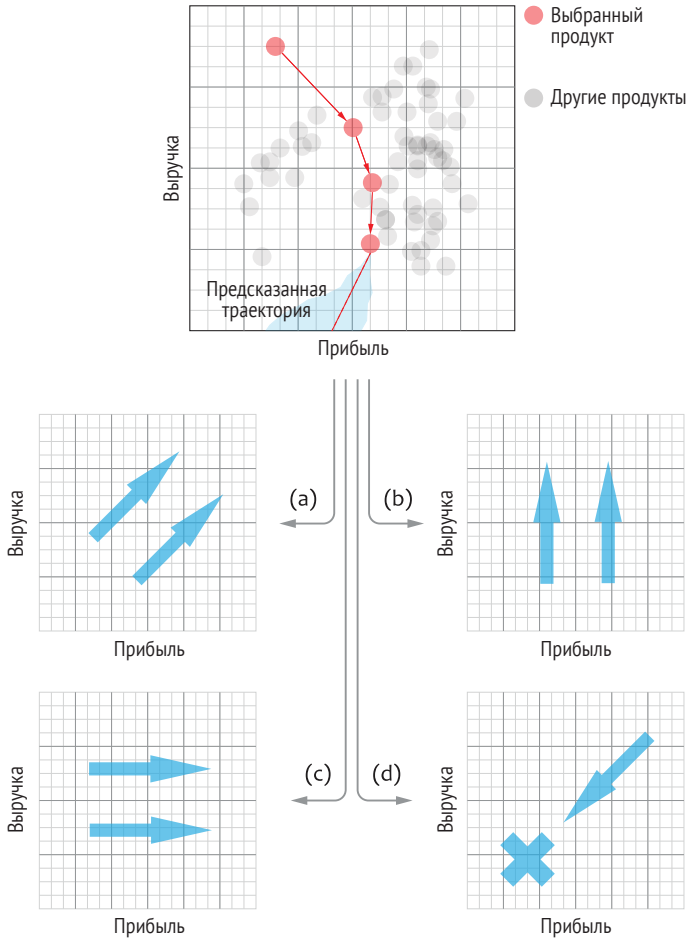

Рис. 1.5. Пример анализа траектории продукта и разработки стратегии

Хотя приведенный выше анализ относительно прост, он позволяет определить стратегии ценообразования и продвижения продукта, которые, в свою

очередь, станут исходными данными для последующих более детализованных моделей оптимизации цен, рекламных акций и управления запасами. В данном примере у нас есть выбор между следующими стратегиями, изображенными на рис. 1.5:

a) в наиболее благоприятных ситуациях можно стремиться к повышению как объема, так и прибыльности продаж. Например, это может произойти с новым инновационным продуктом, который быстро завоевывает долю рынка и сталкивается с небольшой конкуренцией. Траекторный анализ помогает выявить такие продукты и соответствующим образом скорректировать модели ценообразования;

b) траектории некоторых продуктов могут указывать на потенциал увеличения доли рынка. Обычно это относится к продуктам, которые находятся на ранних стадиях своего жизненного цикла. Рекомендации по оптимизации объема затем могут быть переданы последующим моделям и процессам оптимизации цены;

c) некоторые продукты могут не иметь достаточного потенциала для увеличения объема продаж, и компании целесообразно сосредоточиться на максимизации прибыли, которую она получает от них. Обычно это касается продуктов, срок жизни которых приближается к концу. Поэтому инструменты управления ценами могут быть ориентированы на получение максимальной прибыли. К таким инструментам относят модели, прогнозирующие спрос и реакцию потребителя на цену.

   Прибыль также можно повысить за счет сокращения затрат, что может стать отправной точкой для включения в процесс оптимизации моделей управления запасами;

d) наконец, продукты, которые дрейфуют в зону, где слишком малы и прибыль, и объем продаж, возможно, придется вывести из оборота или модернизировать. Это решение может запустить процессы оптимизации ассортимента и управления характеристиками продукта.

Аналогичный анализ применим не только к продуктам, но и к более крупным объектам, таким как категории и местоположения. В целом этот анализ направлен на декомпозицию финансовых целей верхнего уровня, обычно определяемых с точки зрения доходов и прибыльности, на детальные планы действий, которые, в свою очередь, реализуются на уровне компонентов автоматизации принятия решений.

Перейдем ко второму примеру сценарного планирования на уровне процессов, на этот раз сосредоточившись на сценарии управления запасами. Рассмотрим розничного продавца, который управляет обычным магазином, а также продает товары через интернет, обслуживая онлайн-заказы непосредственно с полок магазина. (Такой сервис обычно называют онлайн-доставкой.) Предположим, что сотрудники интернет-магазина берут товары прямо с полок, конкурируя таким образом с обычными покупателями за доступный товар, заказы обрабатываются с некоторой задержкой, а интернет-магазин получает обновления наличия товара каждое утро. Следовательно, розничный продавец сталкивается с инцидентами, когда определен-

ный товар был доступен в начале дня, но распродан к моменту поступления онлайн-заказа. Продавец может попытаться уменьшить количество таких инцидентов, прогнозируя спрос на каждый товар в магазине, резервируя соответствующее количество единиц для покупателей в магазине и выставляя в интернете только остатки от резерва. Этот подход создает компромисс между уровнем доступности (долей запасов, доступных онлайн-клиентам) и уровнем выполнения (долей успешно выполненных заказов). В идеале розничный торговец хочет максимизировать оба показателя, но эти две цели противоречат друг другу. Это означает, что все реалистичные решения будут расположены в пределах ограниченной области на карте полезности, составленной по этим двум метрикам, и границы этой области соответствуют границе Парето.

Пример карты полезности для описанного выше примера показан на рис. 1.6. Мы предполагаем, что суммарная величина запасов фиксирована, поэтому каждому значению уровня выполнения соответствует максимально достижимый уровень доступности, и набор этих пар образует границу Парето. Продавец волен выбирать любую точку на границе или под ней, исходя из деловых соображений. Например, розничный торговец может выбрать поддержание высокого уровня обслуживания клиентов и выбрать точку, основываясь на минимально приемлемом уровне выполнения, или может стремиться к получению прибыли и выбрать точку, исходя из уровня доступности. Однако границу можно преодолеть, увеличив общую емкость запасов (как показано на рисунке) или разработав более точные алгоритмы прогнозирования либо резервирования.

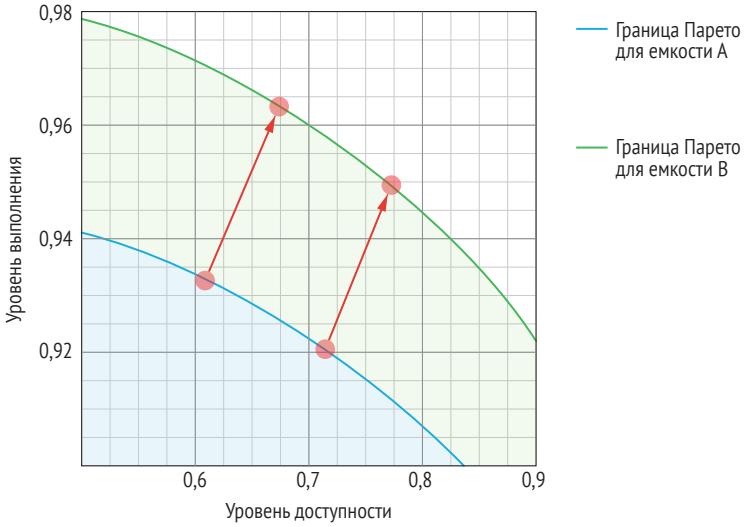

Рис. 1.6. Пример границ Парето для примера с двумя магазинами

Из примеров этого раздела следует, что разнообразие и сложность методов автоматизации принятия решений обычно возрастают при переходе от уров-

ня предприятия к уровню процессов и, соответственно, к менее крупным объектам. В двух приведенных выше примерах модели прогнозирования и оптимизации являются не дополнительными расширениями описательной аналитики, а основными компонентами решения. Наш следующий шаг – изучить еще более низкие уровни детализации.

1.1.3 *Реализация: клиенты, устройства, транзакции и интерфейсы*

Хотя статистические методы и алгоритмы оптимизации могут принести пользу на всех уровнях принятия решений, принятие автоматических решений становится необходимостью на уровне отдельных клиентов и транзакций. На относительно высоких уровнях агрегирования, которые мы обсуждали в предыдущих разделах, существует определенная свобода выбора между традиционной бизнес-аналитикой, моделями поддержки принятия решений и автоматизацией принятия решений, но на более низких уровнях у автоматизации нет альтернативы. Далее мы рассмотрим несколько примеров сценарного планирования на уровне отдельных клиентов и транзакций, а также автоматизации на уровне транзакций.

Начнем с примера из области анализа клиентов. Для многих предприятий, работающих по подписке, таких как телекоммуникации и страхование, отток клиентов является серьезной проблемой, поскольку на этих рынках высоки как затраты на привлечение клиентов, так и пожизненная стоимость, что делает удержание клиентов гораздо более предпочтительным, чем потеря старых клиентов и приобретение новых. Маркетинговые команды обычно создают различные пакетные предложения по удержанию клиентов, но эффективное использование этих пакетов является сложной задачей. Необходимо выявить клиентов, которые подвержены риску ухода, определить оптимальное предложение для каждого клиента на основе факторов, которые предположительно подталкивают этого конкретного клиента к уходу, сбалансировать потенциальные потери со стоимостью предложения, определить оптимальное время для отправки предложения и т. д. Эти решения изображены на рис. 1.7, где в терминах сценарного планирования каждый клиент представляет собой отдельный объект, вероятность ухода является показателем полезности, а различные предложения и время вмешательства считаются сценариями.

На практике большинство этих задач эффективно решаются с помощью статистических моделей, благодаря чему каждый клиент получает персонализированный подход. Более того, часто удается создать высокоавтоматизированные системы, которые принимают почти все решения автономно и практически в реальном времени, что вряд ли достижимо для более высокоуровневых стратегических вариантов.

Еще одним примером, иллюстрирующим возможности автоматического принятия решений для объектов низкого уровня, является обнаружение аномалий. Многие задачи по контролю качества и безопасности, включая мониторинг системных показателей в центрах обработки данных, монито-

ринг телеметрических данных, собранных с промышленного оборудования, а также наблюдение за финансовыми транзакциями, сводятся к различению нормальных и аномальных траекторий в различных метрических пространствах. Например, банк может отслеживать количество или процент операций по возврату платежей и обнаруживать выбросы, выходящие за рамки обычного ежедневного графика, как показано на рис. 1.8. На практике обнаружение таких выбросов часто выполняется автоматически и с высокой точностью, поскольку можно построить достаточно достоверную модель процесса, аппроксимирующую наблюдаемые закономерности, а затем использовать эту модель для прогнозирования ожидаемых отклонений в поведении, которые будут считаться аномалиями.

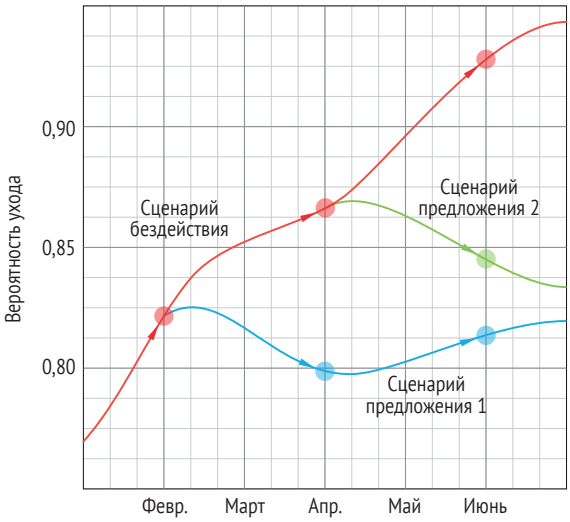

Рис. 1.7. Пример траектории движения клиентов в контексте задачи предотвращения оттока. Альтернативными мерами полезности для этой задачи служат вероятность выживания (величина, обратная вероятности оттока) и ожидаемая LTV

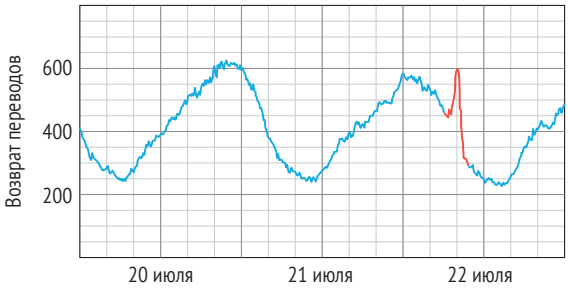

Рис. 1.8. Пример аномалии в показателях транзакций

Оценка сценариев и автоматизация принятия решений – это лишь одно из применений статистического обучения и вероятностного рассуждения на уровне отдельных транзакций. Второе большое направление – автоматическое создание сложных объектов, таких как изображения, исходные коды или тексты на естественном языке, с целью автоматизации бизнес-процессов и повышения производительности пользователей. Например, можно повысить производительность создания графических ресурсов для видеоигр, используя генеративную модель, которая синтезирует изображения на основе описаний на естественном языке, как показано на рис. 1.9. К другим примерам из этой области относятся диалоговые интерфейсы, которые позволяют пользователям задавать вопросы и получать ответы на естественном языке, помощники по написанию кода и инструменты копирайтинга. В таких приложениях центральную роль играют генеративные модели для текста и изображений.

Приведенные выше примеры показывают, что разработка компонентов автоматизации принятия решений и процессов на низком уровне (и, соответственно, малого масштаба) облегчается за счет построения достаточно точных и автономных математических моделей, что труднее сделать при работе с крупными объектами верхнего уровня. В отличие от крупных объектов, более мелкие объекты, такие как онлайн-пользователи, платежные транзакции и изображения продуктов, как правило, многочисленны, поэтому мы часто наблюдаем миллионы или миллиарды экземпляров, что также облегчает и, по сути, требует использования статистических методов. Спектр объектов очень широк: от целых рынков и компаний до потребителей и транзакций, поэтому предприятиям обычно приходится выстраивать иерархию вычислительных компонентов поддержки принятия решений и автоматизации процессов. Компоненты верхнего уровня в этой иерархии чаще всего представляют собой инструменты поддержки принятия решений, предназначенные для декомпозиции сложной задачи на более мелкие и определения правильных целей и параметров для последующих компонентов. Компоненты нижнего уровня обычно представляют собой автономные модели, которые оптимизируют и автоматизируют действия на основе целей, поступающих с верхних уровней.

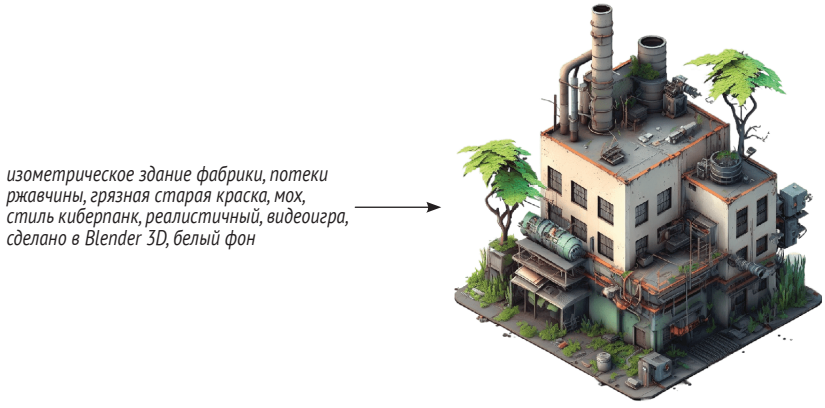

Рис. 1.9. Пример создания игровых ресурсов на основе подсказки на естественном языке

1.2 ВОЗМОЖНОСТИ МОДЕЛИРОВАНИЯ

В предыдущих разделах мы рассмотрели несколько базовых примеров, иллюстрирующих применение количественных методов в области принятия решений и автоматизации процессов на разных уровнях детализации и в различных подразделениях предприятия. Реализация этих подходов требует комплексного инструментария статистических и оптимизационных методов, способных совместно решать несколько категорий задач. Первая категория связана с включением в анализ различных сигналов и источников данных и извлечением семантически значимых представлений, которые можно использовать при принятии решений и автоматизации процессов:

- *представления объектов*: многие объекты имеют сложные цифровые профили, включающие числовые, текстовые и графические данные. Например, клиент может быть представлен настройками учетной записи, историей транзакций и просмотров, графами социальных связей и текстовыми сообщениями; продукт может быть представлен числовыми и категориальными атрибутами, текстовыми описаниями, изображениями и отзывами покупателей. Следовательно, необходимы инструменты для извлечения семантического значения из этих данных и создания компактных представлений объектов, которые можно использовать в последующих моделях и аналитических процессах. Например, текстовые сообщения могут быть представлены как наборы тегов тем и настроений; изображения продуктов могут быть аннотированы тегами стиля и т. д. Такие семантические представления можно создавать вручную, и этот процесс называется *извлечением признаков* (feature engineering), или обучать модель с использованием статистических методов, известных как *обучение представлению* (representation learning);
- *алгебра объектов*: проблема вычисления семантических представлений тесно связана с проблемой вычисления расстояний между объектами. Многие задачи корпоративного ИИ, особенно в приложениях маркетинга и управления информацией, можно свести к оценке расстояний, или, иными словами, сходства между объектами в соответствующем семантическом пространстве. Примерами могут служить системы рекомендации товаров, где необходимо измерить сходство между товарами и пользователями, задачи поиска текста и изображений, когда нужно найти элементы, похожие на поисковый запрос или эталонное изображение, а также задачи управления ценами и ассортиментом, где показатель расстояния между продуктами часто является важным фактором. Семантические представления предоставляют общий и удобный способ вычисления расстояний и выполнения других алгебраических операций над объектами;
- *прогнозирование свойств объекта*: во многих приложениях нам необходимо прогнозировать ненаблюдаемые свойства или классы объектов. Например, нам может потребоваться оценить ожидаемый доход в течение следующего года для каждого клиента или классифицировать

изображения с камеры наблюдения как обычные ситуации, скопление людей, уличные драки или неправильную парковку;

- *генерация объектов*: в некоторых приложениях нам необходимо генерировать объекты на основе ограниченных входных данных. Например, сервис персонализации продукта может генерировать предложения по дизайну продукта и соответствующие визуализации на основе описания на естественном языке, предоставленного пользователем.

Вторая категория задач связана с моделями, которые помогают нам понять внутреннюю структуру траекторий объектов и прогнозировать будущие перемещения. Если использовать в качестве аналогии разделы физики, то эту категорию можно рассматривать как дисциплину, изучающую динамику объектов, тогда как предыдущую категорию можно отнести к статике. Задачи, которые нам необходимо решить в этой области, заключаются в следующем:

- *декомпозиция траектории*: траектория может быть сформирована множеством различных сил, как наблюдаемых, так и скрытых. Например, розничный продавец может легко отслеживать еженедельные показатели продаж определенного товара, но этот временной ряд состоит из сложного сочетания компонентов, таких как сезонность, реакция на изменения цен и маркетинговые кампании, эффекты каннибализации, связанные с аналогичными продуктами или изменениями цен конкурентов. Большинство этих эффектов невозможно измерить явно, и их необходимо оценивать с помощью статистического анализа. Этот процесс декомпозиции траектории на элементарные компоненты обеспечивает более глубокий ручной анализ, а также автоматическую оптимизацию. Например, система управления ценами, которая не учитывает эффекты каннибализации, склонна принимать неоптимальные решения, увеличивающие продажи одного продукта, но снижающие общую прибыльность категории товаров.

  Проблему декомпозиции траектории можно рассматривать как динамический аналог обучения представлению объектов. В то время как основная цель обучения представлению – описать статическое состояние объекта с использованием семантически значимых компонентов (измерений), основная цель декомпозиции траектории – описать динамику объекта с точки зрения сил, имеющих четкое семантическое значение;
- *прогнозирование траектории*: анализ и декомпозиция траектории обычно могут быть расширены до прогнозирования. В нашем предыдущем примере, касающемся управления ценами, для прогнозирования будущих продаж можно использовать модель, которая способна разлагать показатели продаж на сезонные и ценовые компоненты. Декомпозицию и прогнозирование часто можно рассматривать как два разных режима (описательный и прогнозирующий) использования одних и тех же или похожих моделей;
- *прогнозирование свойств траектории*: подобно прогнозированию свойств объекта, нам может потребоваться предсказать скрытые свой-

ства траектории или классы траектории. Например, можно классифицировать наблюдаемые траектории как нормальные или аномальные.

Наконец, конечной целью корпоративных систем и инструментов искусственного интеллекта является улучшение конкретных действий и решений с точки зрения оптимальности или степени автоматизации, поэтому мы определяем третью категорию задач, связанных с оптимальным управлением объектом, следующим образом:

- *управление объектами*: прогнозирующие модели позволяют анализировать возможные действия по принципу «что, если», чтобы алгоритм оптимизации мог оценить несколько сценариев и определить оптимальный. Это создает основу для разработки предписывающих инструментов и компонентов автономного принятия решений для управления объектами. Разработка моделей оптимизации, которые охватывают все важные экономические факторы и способны находить стратегически оптимальные многоэтапные сценарии, является центральной задачей при построении системы управления предприятием.

  Теоретически основная цель алгоритмов управления объектами – разработать оптимальные или почти оптимальные стратегии или рецепты действий. На практике нам часто приходится отвечать на дополнительные вопросы о решении и структуре пространства решений. Одним из важных примеров является *анализ чувствительности* – измерение того, насколько ухудшается качество решения, когда параметры действия отклоняются от оптимальных значений или если нарушаются предположения моделирования. Например, планировщику запасов может быть интересно узнать разницу между теоретически оптимальными циклами пополнения запасов (6.53 дня) и практически значимым циклом в 7 дней (один раз в неделю);
- *динамическое управление*: задачи, которые мы обсуждали выше, начиная от обучения представлению и заканчивая планированием действий, традиционно рассматриваются с точки зрения статистического анализа исторических данных. Обычно этот подход включает в себя множество ручных действий, связанных с подготовкой данных, разработкой моделей и производственной интеграцией. Этот подход не всегда осуществим в сложных или динамичных средах, где репрезентативные исторические данные могут быть недоступны или быстро устаревают из-за постоянных изменений статистических свойств процессов. Например, система персонализации, основанная на поведенческих профилях клиентов, может некорректно работать в среде с непрерывным потоком новых клиентов. В подобных случаях возникает необходимость динамического управления, когда система должна постоянно исследовать окружающую среду, мгновенно учиться на постоянной обратной связи и постоянно корректировать стратегию принятия решений и исследования.

На практике не обязательно создавать полный конвейер с отдельными этапами вычислений представления, прогнозирования и оптимизации управле-

ния. Например, некоторые процессы мерчендайзинга, такие как снабжение изображений продуктов тегами, можно автоматизировать, используя только модели компьютерного зрения, которые вычисляют представление изображений и классифицируют их. Однако полная структура может быть полезна для планирования более сложных решений, которые автоматизируют сложные операции и включают несколько моделей и компонентов.

1.3 ВНЕДРЕНИЕ БАЗОВЫХ МОДЕЛЕЙ И AUTOML

Решение вышеперечисленных задач может быть основано на широком спектре статистических и оптимизационных методов. Однако методы глубокого обучения и обучения с подкреплением обеспечивают наиболее полную платформу для реализации корпоративных приложений искусственного интеллекта и охватывают большинство необходимых инструментов, включая семантический анализ, прогнозирование и оптимизацию действий. Следующие три главы мы посвятим изучению базовых функциональных компонентов и более подробному обсуждению возможностей моделирования.

Реализация этих компонентов и возможностей обычно включает в себя несколько этапов, таких как исследовательский анализ данных, извлечение признаков, проектирование модели, обучение, настройка, выбор, интерпретация и проверка. Все эти этапы необходимы для создания полезных, хорошо работающих и заслуживающих доверия моделей. Для каждого из них существует обширная теоретическая база и большое количество прикладных методов и приемов, так что каждому этапу можно было бы посвятить отдельную книгу[1].

Чтобы справиться с такой сложностью, необходимо остановиться на каком-то одном подходе к процессу разработки модели. Мой выбор обусловлен главным образом следующими двумя соображениями:

- *внедрение технологии AutoML*: традиционный подход к моделированию предполагает, что все описанные выше шаги выполняются в ручном или полуручном режиме экспертом в области статистики и машинного обучения. Это сложный процесс, требующий продвинутых навыков, значительного количества времени и множества шагов проб и ошибок для разработки работающего решения. В конце 2010-х годов ограничения этого подхода стали общепризнанными и привлекли большое внимание в отрасли по нескольким причинам. Во-первых, широкое внедрение машинного обучения обострило потребность в эффективных инструментах, которые позволяют экспертам в предметной области создавать решения на базе машинного обучения без привлечения высококвалифицированных специалистов по машинному обучению. Во-

[1] Примерами таких книг являются (Kuhn and Johnson, 2019), посвященная извлечению признаков, и (Molnar, 2020), посвященная интерпретации моделей.

вторых, стремительный прогресс в области методов глубокого обучения резко увеличил сложность выбора и настройки архитектуры модели. Эти проблемы привели к быстрому развитию методов и технологий, известных под общим названием *автоматизированное машинное обучение*, или *AutoML*.

Методы AutoML обычно ориентированы на декларативный подход к разработке модели, при котором эксперт в предметной области указывает только ограниченное количество входных данных, таких как тип задачи, целевая функция и необработанные данные, а механизм AutoML автоматически создает конвейер с необходимыми компонентами преобразования данных и выбирает архитектуру модели, близкую к оптимальной. Процесс построения обычно управляется целевой функцией, поэтому необходимо просмотреть множество возможных архитектур моделей до тех пор, пока не будет найден наиболее эффективный вариант[1]. Например, эксперт в предметной области может поставить задачу прогнозирования временных рядов с целью минимизировать ошибку прогнозирования, но конкретные преобразования входных данных, архитектура модели и значения гиперпараметров будут автоматически определены механизмом подбора;

- *внедрение опорных моделей*: в традиционной корпоративной аналитике модели и компоненты автоматизации принятия решений обычно создаются с нуля на основе только собственных данных, либо сгенерированных самим предприятием, либо полученных от третьих сторон. Однако этот подход неприменим для большинства приложений, включающих компьютерное зрение и обработку естественного языка. В 2010-х гг. широкое распространение на предприятиях получили готовые модели компьютерного зрения, предварительно обученные на больших наборах данных общего назначения, поскольку точная настройка таких моделей на данных для конкретных задач обычно более эффективна, чем обучение с нуля на ограниченном количестве этих же данных. В 2020-х годах генеративные модели искусственного интеллекта, в том числе *большие языковые модели* (large language model, LLM) и модели преобразования текста в изображение, которые обучаются на чрезвычайно больших объемах данных, открыли уникальные возможности для автоматизации процессов и создания пользовательских интерфейсов следующего поколения. Такие модели, называемые *базовыми*, или *опорными*, *моделями* (foundation model), обычно распространяются в виде облачных сервисов, доступных через API или предварительно упакованных компонентов, а создание таких моделей с нуля, как и наборов данных, необходимых для их обучения, находится за пределами досягаемости или потребностей большинства предприятий.

В этой книге мы делаем предположение, что две изложенные выше парадигмы в обозримом будущем будут доминировать в корпоративном сообще-

[1] Обсуждение конкретных методов и алгоритмов AutoML выходит за рамки этой книги, но существуют подробные обзоры, такие как (He et al., 2021).

стве искусственного интеллекта или, по крайней мере, со временем будет возрастать уровень автоматизации, предлагаемый инструментами и платформами машинного обучения. Исходя из этих соображений, были написаны главы 2–4, посвященные функциональным возможностям и интерфейсам базовых компонентов. В прикладных рецептах R1–R14 основное внимание уделяется декомпозиции различных вариантов корпоративного использования на стандартные задачи машинного обучения, готовые компоненты и настройки для конкретной предметной области, исходя из предпосылки, что многие задачи реализации могут быть решены с помощью инструментов и сервисов ML.

1.4 КРАТКИЕ ИТОГИ ГЛАВЫ

- Методы, основанные на данных, могут применяться на разных уровнях детализации: стратегические решения для крупных экономических объектов, таких как компании, тактические решения и оптимизация отдельных процессов, а также микрорешения на уровне отдельных клиентов, транзакций и операций.
- Многие задачи принятия решений и оптимизации на предприятии можно удобно представить в виде объектов (сущностей), пространств полезности, траекторий и сценариев вмешательства.
- Анализ объектов и сценариев требует изучения семантически значимых представлений объектов, объяснения и прогнозирования траекторий, а также оценки потенциальных вмешательств.
- Методы глубокого обучения и обучения с подкреплением предоставляют надежную платформу для моделирования объектов и сценариев. Эта платформа имеет определенные ограничения, которые можно устранить с помощью альтернативных методов машинного обучения.
- Инструменты AutoML и предварительно обученные модели помогают снизить требования к данным и инфраструктуре за счет упрощения задач подготовки данных и проектирования моделей. Это также помогает разработчикам корпоративных ИИ-решений сосредоточиться на декомпозиции бизнес-задач на стандартные задачи машинного обучения.