1 Что такое глубокое обучение

В этой главе

- ✓ Обобщенные определения основных понятий.
- ✓ Введение в принципы машинного обучения.
- ✓ Основные факторы роста популярности глубокого обучения и его потенциал в будущем.

В последнее десятилетие тема искусственного интеллекта вызвала сильный резонанс в средствах массовой информации. Машинное обучение, глубокое обучение и ИИ упоминались в бесчисленном количестве статей, многие из которых никак не связаны с описанием технологий. Нам обещали появление умных чатботов, беспилотных автомобилей и виртуальных помощников. Иногда будущее рисовали в мрачных тонах, а иногда изображали утопическим — таким, где люди освобождены от рутинного труда, а основную работу выполняют роботы, наделенные искусственным интеллектом. Нынешнему и будущему специалисту в области машинного обучения важно уметь выделять полезный сигнал из шума, видеть в раздутых пресс-релизах изменения, действительно способные повлиять на мир.

Наше будущее поставлено на карту, и в том числе от вас зависит, как оно в итоге будет выглядеть. Ведь закончив чтение этой книги, вы вольетесь в ряды разработчиков систем ИИ. Поэтому мы рассмотрим такие вопросы: чего уже удалось достичь в области глубокого обучения? Насколько важны эти достижения? В каком направлении пойдет дальнейшее развитие? Можно ли верить поднятой шумихе?

1.1. Искусственный интеллект, машинное и глубокое обучение

Прежде всего определим, что подразумевается под искусственным интеллектом. Что такое ИИ, машинное и глубокое обучение? Как они связаны друг с другом (рис. 1.1)?

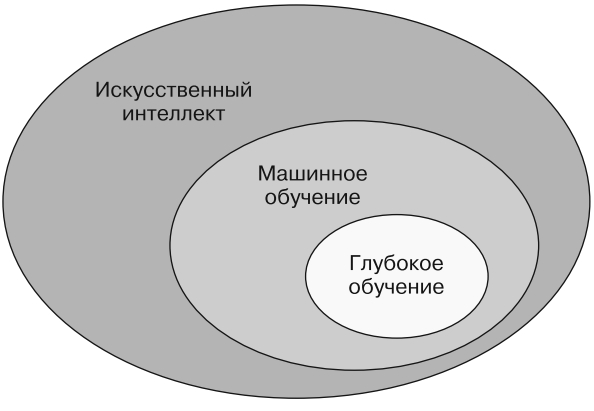


Рис. 1.1. Искусственный интеллект, машинное и глубокое обучение

1.2. Искусственный интеллект

Термин «искусственный интеллект» (artificial intelligence) появился в 1950-х годах, когда группа энтузиастов из только зарождающейся области информатики задалась вопросом: можно ли заставить компьютеры думать? Это породило другие вопросы, ответы на которые мы ищем до сих пор. Притом что многие идеи, которые легли в основу ИИ, появились за годы и даже десятилетия до этого, искусственный интеллект окончательно превратился в область исследований только в 1956 году, когда Джон Маккарти, молодой преподаватель математики из Дартмутского колледжа, организовал летний семинар со следующим обоснованием:

> «Исследование будет проводиться исходя из предположения, что каждый аспект обучения или любая другая особенность интеллекта в принципе могут быть описаны настолько точно, что на основе такого описания можно создать машину для их имитации. Будет предпринята попытка выяснить, как заставить машины использовать язык, формировать абстракции и понятия, решать виды задач, которые сейчас доступны только людям, и совершенствовать себя. Мы думаем, что в разрешении одной или нескольких из этих проблем можно добиться значительного прогресса, если специально подобранная группа ученых будет совместно работать над ними в течение лета»[1].

[1] Фрагмент из «Предложения о проведении Дартмутского летнего исследовательского проекта по искусственному интеллекту», написанного в августе 1955 года. Этот документ содержит первое известное упоминание придуманного Джоном Маккарти термина «искусственный интеллект». — *Примеч. науч. ред.*

В конце лета семинар завершился, так и не дав полноценных ответов на поставленные вопросы. Тем не менее он объединил многих будущих первопроходцев данной области и стал толчком интеллектуальной революции, продолжающейся по сей день.

Коротко ИИ можно определить так: *попытка автоматизации интеллектуальных задач, обычно выполняемых людьми*. Соответственно, ИИ — это область, охватывающая машинное обучение и глубокое обучение, а также включающая в себя многие подходы, с обучением не связанные. Только представьте, до 1980-х годов в большинстве учебников по искусственному интеллекту «обучение» вообще не рассматривалось! Например, первые программы для игры в шахматы действовали по жестко определенным правилам, заданным программистами, и не использовали машинное обучение. Долгое время многие эксперты полагали, что искусственный интеллект уровня человеческого можно создать, если программисты вручную напишут достаточный набор правил для манипулирования знаниями. Этот подход, известный как *символьный ИИ*, был доминирующей парадигмой ИИ с 1950-х до конца 1980-х годов. Пик его популярности пришелся на бум *экспертных систем* в 1980-х.

Символьный ИИ неплохо справлялся с четко определенными логическими задачами, такими как игра в шахматы. Но оказалось, что этот подход не годился для более сложных и менее четких случаев, например классификации изображений, распознавания речи и перевода на другие языки, ведь для них прописать нужные правила почти невозможно. Поэтому на смену символьному ИИ пришел новый подход — *машинное обучение*.

1.3. Машинное обучение

В викторианской Англии жила леди Ада Лавлейс — друг и соратник Чарльза Бэббиджа, изобретателя *аналитической машины* (первого известного универсального механического компьютера[1]). Несомненно, устройство опередило свое время, но в 1830–1840-х годах оно не задумывалось именно как универсальный компьютер, потому что самой идеи универсальных вычислений еще не существовало. Машина должна была использовать механические операции для автоматизации некоторых вычислений из области математического анализа, откуда и получила свое название[2]. Она была более интеллектуальным потомком более ранних механических вычислительных устройств, таких как суммирующая машина Паскаля («паскалина») или ее усовершенствованная версия — счетная машина Лейбница. Разработанная Блезом Паскалем в 1642 году (в возрасте 19 лет!), «паскалина» была первым в мире рабочим механическим калькулятором, который мог складывать и вычитать числа.

---

[1] Который, однако, не был построен, оставшись одним из нереализованных проектов Бэббиджа. — *Примеч. науч. ред.*

[2] Сам Бэббидж писал о своей машине и о ее названии так: «...возможности этой машины в выполнении самых сложных аналитических операций представляются почти неограниченными <...> Я назвал эту машину аналитической — в противопоставление другой своей машине [разностной]». // *Babbage C.* Passages from the Life of a Philosopher. — *Примеч. науч. ред.*

В 1843 году Ада Лавлейс заметила:

> «Аналитическая машина не может создавать новое. Она делает лишь то, что мы способны ей приказать... Ее предназначение — помогать нам использовать то, с чем мы уже знакомы».

Наблюдение леди Лавлейс остается поразительным даже спустя 183 года. Сможет ли универсальный компьютер создавать что-нибудь свое или он всегда будет просто выполнять операции, полностью понятные нам, людям? Сможет ли когда-нибудь породить какую-то оригинальную мысль? Сможет ли учиться на собственном опыте? Сможет ли стать творцом?

Позднее пионер ИИ Алан Тьюринг в своей знаменитой статье *Computing Machinery and Intelligence*[1] назвал это замечание аргументом Ады Лавлейс[2]. Там же он представил *тест Тьюринга*, а также изложил ряд идей, оказавших влияние на дальнейшее развитие ИИ[3]. Тьюринг придерживался весьма провокационного для того времени мнения, что компьютеры могут имитировать все аспекты человеческого интеллекта.

Обычно, чтобы заставить компьютер выполнять полезную работу, требуется создать *правила* — программу, которой нужно следовать, чтобы преобразовать входные данные в соответствующие ответы (точно так же, как леди Лавлейс составляла пошаговые инструкции для аналитической машины). Машинное обучение меняет ситуацию: машина смотрит на входные данные и соответствующие им ответы и выясняет, какими должны быть правила (рис. 1.2).

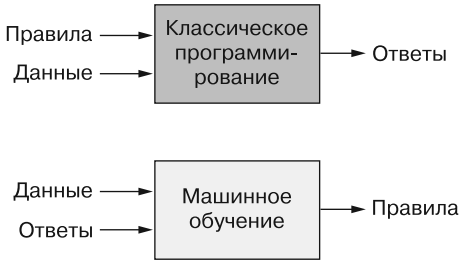


Рис. 1.2. Машинное обучение — новая парадигма программирования

В машинном обучении система *обучается*, а не программируется явно. Ей передаются многочисленные примеры, имеющие отношение к некоторой задаче, а она находит в них статистические закономерности, которые позволяют выработать правила для решения этой задачи. Например, чтобы научиться автоматически присваивать теги фотографиям, сделанным в отпуске, можно передать системе машинного обучения множество примеров фото, уже размеченных людьми, — и система выучит статистические правила классификации конкретных фотографий на категории, такие как «пейзаж» или «еда».

---

[1] «Вычислительные машины и разум», перевод на русский язык можно найти по адресу https://bio.wikireading.ru/6066. — *Примеч. пер.*

[2] *Turing A. M.* Computing Machinery and Intelligence // Mind 59, 1950. — № 236. — P. 433–460.

[3] Тест Тьюринга иногда интерпретировали буквально — как цель, которой должен достичь ИИ. Однако в действительности Тьюринг имел в виду лишь концептуальный прием в философской дискуссии о природе познания.

Расцвет машинного обучения начался только в 1990-х годах, но данное направление быстро стало наиболее популярной и успешной частью ИИ — эта тенденция была подкреплена появлением все более мощного оборудования и больших наборов данных. Машинное обучение тесно связано с математической статистикой, но имеет несколько важных отличий, подобно тому как медицина связана с химией, но не может быть сведена только к ней. В отличие от статистики, машинное обучение обычно имеет дело с большими и сложными наборами данных (скажем, миллионами изображений, каждое из которых состоит из десятков тысяч пикселей), к которым на практике трудно применить классические методы статистического анализа, например байесовские. Как результат, машинное, и особенно глубокое, обучение содержит относительно мало математической теории и по своей сути является инженерной дисциплиной. В отличие от теоретической физики или математики, машинное обучение — это очень практикоориентированная сфера, развитие которой определяется эмпирическими результатами и сильно зависит от достижений в области программного и аппаратного обеспечения.

1.4. Изучение правил и представлений на основе данных

Чтобы дать определение *глубокого обучения* и понять разницу между этим и другими методами машинного обучения, сначала нужно узнать, что делают алгоритмы машинного обучения. Как мы только что отметили, машинное обучение выявляет правила для решения задач обработки данных по примерам ожидаемых результатов. То есть нам нужны три составляющие:

- *входные данные* — например, если решается задача распознавания речи, такими данными могут быть файлы с записью речи разных людей. Если нужно классифицировать изображения, понадобятся соответствующие изображения;
- *примеры ожидаемых результатов* — в задаче распознавания речи это обычно транскрипции звуковых файлов, составленные людьми. При классификации изображений ожидаемым результатом могут быть теги, например «собака», «кошка» и др.;
- *способ оценки качества работы алгоритма* — необходим для определения того, как сильно отклоняются результаты, возвращаемые алгоритмом, от ожидаемых. Оценка используется в качестве сигнала обратной связи для корректировки работы алгоритма. Этот этап корректировки мы и называем *обучением*.

Модель машинного обучения трансформирует входные данные в значимые результаты, обучаясь на известных примерах того и другого. То есть главной задачей машинного и глубокого обучения является *значимое преобразование данных*, или, иными словами, обучение *представлению* входных данных, приближающему нас к ожидаемому итогу.

Прежде чем двинуться дальше, определим, что такое представление данных. По сути, это другой способ их *отображения* или *кодирования*. Например, цветное изображение можно закодировать в формате RGB (red — green — blue — «красный — зеленый — синий») или HSV (hue — saturation — value — «тон — насыщенность — яркость») — двух разных представлениях одних и тех же данных. Некоторые

задачи трудно решить, если данные существуют в одном представлении, но легко — если в другом. Например, выбрать все красные пиксели проще в RGB-изображениях, тогда как сделать изображение менее насыщенным быстрее в формате HSV. Главная задача моделей машинного обучения как раз и заключается в поиске соответствующего представления входных данных — таких преобразований, которые сделают данные более пригодными для решения конкретной задачи.

Обратимся к примеру. Рассмотрим систему координат с осями X и Y и несколько точек в ней (рис. 1.3).

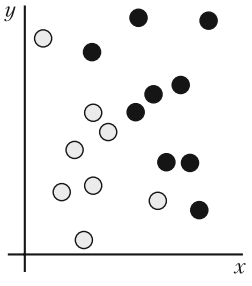


Рис. 1.3. Пример некоторых данных

Как видите, у нас имеется несколько белых и черных точек. Допустим, нужно разработать алгоритм, принимающий координаты (x, y) точки и возвращающий наиболее вероятный цвет — черный или белый. В данном случае:

- исходными данными являются координаты точек;
- результатом — цвет;
- мерой качества работы алгоритма может быть, например, процент правильно классифицированных точек.

Нам нужно получить новый способ представления исходных данных, позволяющий четко отделять белые точки от черных. Таким преобразованием, кроме прочего, могло бы быть изменение системы координат (рис. 1.4).



Рис. 1.4. Изменение системы координат

Координаты точек в изменившейся системе координат можно назвать новым представлением данных. Причем более удачным! Задачу классификации данных «черный/белый» здесь можно свести к простому правилу «черные точки имеют координату $x > 0$» или «белые точки имеют координату $x < 0$». Это новое представление в сочетании с найденным правилом точно решает поставленную задачу.

В примере мы определили изменение координат вручную — использовали человеческий интеллект, чтобы придумать надлежащее представление данных. Этот подход можно с успехом применять в похожих простых задачах. Но смогли бы вы с такой же легкостью классифицировать изображения рукописных цифр? Получилось бы у вас явно сформулировать правила преобразования, которые подчеркнули бы разницу между шестеркой и восьмеркой, единицей и семеркой, написанных разными людьми?

Отчасти это возможно. Такие правила, как «количество замкнутых окружностей», или вертикальные и горизонтальные пиксельные гистограммы позволяют довольно точно различать рукописные цифры. Но отыскать подобные полезные представления вручную очень непросто. К тому же система, основанная на жестких правилах, очень хрупкая — просто кошмар для поддержки. Каждый раз, столкнувшись с нетипичным образцом почерка, нарушающим тщательно продуманные ранее правила, вам придется добавлять новые, не забывая при этом учитывать их взаимодействие со всеми предыдущими.

Вы, наверное, думаете: если этот процесс такой болезненный, можно ли его автоматизировать? Что, если мы системно опробуем различные наборы представлений данных, сгенерированных автоматически, и правил, на них основанных, и определим наилучшие, базируясь на проценте правильно классифицированных цифр в некоторой первоначальной выборке? Это и будет самое настоящее машинное обучение. *Обучение* в контексте машинного обучения описывает процесс автоматического поиска преобразований, создающих полезные представления данных, который управляется сигналом обратной связи. В результате получаются представления, позволяющие решить поставленную задачу с помощью более простых правил.

Этими преобразованиями могут быть изменение системы координат (как в примере классификации на рис. 1.4) или гистограмма пикселей и подсчет замкнутых окружностей (как в примере классификации рукописных цифр), но ими могут быть и линейные проекции, сдвиги и нелинейные операции (например, «выбрать все точки с координатой $x > 0$») и т. д. Алгоритмы машинного обучения обычно не проявляют изобретательности в поиске этих преобразований — они просто выполняют поиск в предопределенном наборе операций, который называют *пространством гипотез*. Например, в задаче классификации точек таким пространством будет пространство всех возможных преобразований двумерной системы координат.

Иначе говоря, технически машинное обучение — это поиск значимых представлений и правил по некоторым входным данным в предопределенном пространстве возможностей с использованием сигнала обратной связи. Эта простая идея позволяет решать чрезвычайно широкий круг интеллектуальных задач — от распознавания речи до автоматического вождения автомобиля.

Теперь, когда вы знаете, что понимается под *обучением*, посмотрим, что особенного в *глубоком обучении*.

1.5. Глубина глубокого обучения

Глубокое обучение — особый раздел машинного обучения, новый подход к получению представлений из данных, при котором особое внимание уделяется обучению последовательных *слоев* все более значимых представлений. Под глубиной в глубоком обучении не подразумевается более детальное понимание, обусловленное этим подходом, — суть его заключается в создании многослойного представления. Число слоев, на которые делится модель данных, называют *глубиной* модели.

Поэтому подходящими названиями для этой области машинного обучения могли бы быть также *многослойное обучение* и *иерархическое обучение*. Современное глубокое обучение часто предполагает десятки и даже сотни последовательных слоев представлений — все они автоматически обучаются на основе данных. Другие же подходы машинного обучения обычно ориентированы на изучение всего одного-двух слоев представлений данных, по этой причине их еще иногда называют *поверхностным обучением*.

В глубоком обучении такие многослойные представления почти всегда изучаются с использованием *нейронных сетей* — моделей, структурированных в виде слоев, наложенных друг на друга. Термин «нейронная сеть» знаком нам из нейробиологии. И хотя некоторые основополагающие идеи глубокого обучения действительно отчасти заимствованы из науки о мозге, его модели все же не являются моделями мозга. Нет никаких доказательств того, что мозг реализует механизмы, подобные используемым в глубоком обучении. Вам могут встретиться научно-популярные статьи, где утверждается, что глубокое обучение работает подобно мозгу или моделирует работу мозга, но в действительности это не так. Было бы неправильно заставлять новичков в этой области думать, что глубокое обучение каким-то образом связано с нейробиологией. Забудьте всю эту туманную мистику про «как наш мозг», забудьте также все, что читали о гипотетической связи между глубоким обучением и биологией. Намного продуктивнее считать данный метод математическим инструментом для изучения представлений данных.

Как выглядят представления, получаемые алгоритмом глубокого обучения? Давайте посмотрим, как сеть, имеющая несколько слоев (рис. 1.5), преобразует изображение цифры, пытаясь ее распознать.

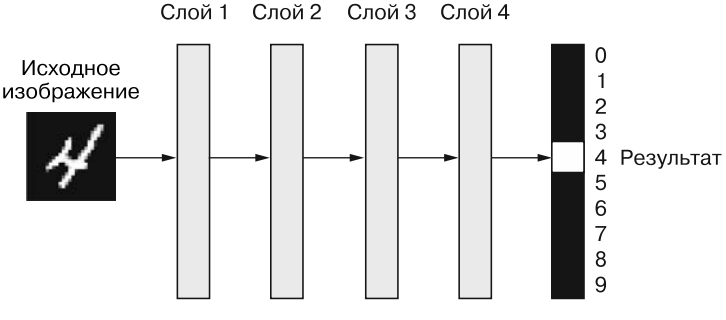


Рис. 1.5. Глубокая нейронная сеть для классификации цифр

Как показано на рис. 1.6, сеть поэтапно преобразует изображение цифры в представление, все больше отличающееся от исходного и несущее все больше полезной информации. Глубокую сеть можно рассматривать как многоэтапную *операцию очистки*, во время которой информация проходит через последовательность фильтров и выходит из нее в *очищенном* (пригодном для решения поставленной задачи) виде.

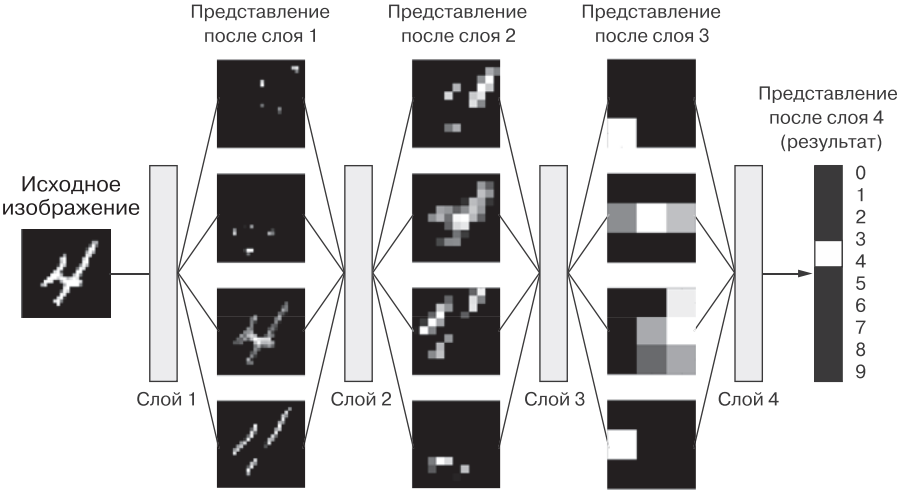


Рис. 1.6. Глубокие представления, получаемые моделью классификации цифр

С технической точки зрения глубокое обучение — это многоступенчатый способ получения представления данных. Идея проста, но, оказывается, очень простые механизмы в определенном масштабе могут выглядеть непонятными и таинственными.

1.6. Принцип действия глубокого обучения в трех картинках

Теперь вы понимаете, что суть машинного обучения заключается в преобразовании ввода (изображений) в результат (подпись «кошка») путем исследования множества примеров входных данных и результатов. Вы также знаете, что глубокие нейронные сети выполняют при этом длинную последовательность простых преобразований (слоев) и обучаются этим преобразованиям на примерах. Посмотрим, как именно проходит обучение.

То, что именно слой делает со своими входными данными, определяется его *весами*, которые фактически являются набором чисел. Выражаясь техническим языком, преобразование, реализуемое слоем, *параметризуется* весами (рис. 1.7). (Веса иногда называют также параметрами слоя.) В данном контексте *обучение* — это поиск набора значений весов всех слоев сети, при котором сеть будет правильно отображать образцы входных данных в соответствующие им результаты. Но дело в том, что глубокая нейронная сеть может иметь десятки миллионов параметров. Поиск правильного значения для каждого из них может оказаться сложнейшей задачей, особенно если изменение значения одного параметра влияет на поведение всех остальных!

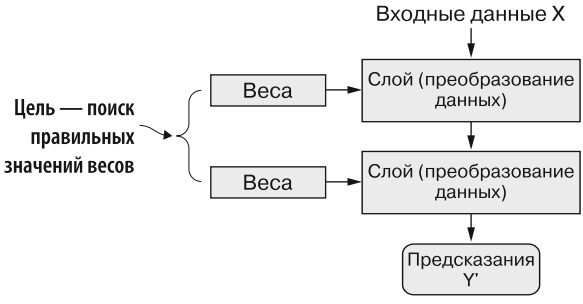


Рис. 1.7. Нейронная сеть параметризуется ее весами

Чтобы чем-то управлять, сначала нужно получить возможность наблюдать за этим. Для того чтобы управлять результатом работы нейронной сети, нужно измерить, насколько он далек от ожидаемого. Эту задачу решает *функция потерь* сети, называемая также *целевой функцией*, или *функцией стоимости*. Функция потерь принимает предсказание, выданное сетью, и истинное значение, которое та должна была вернуть, и вычисляет оценку расстояния между ними, отражающую, насколько хорошо сеть справилась с данным конкретным примером (рис. 1.8).

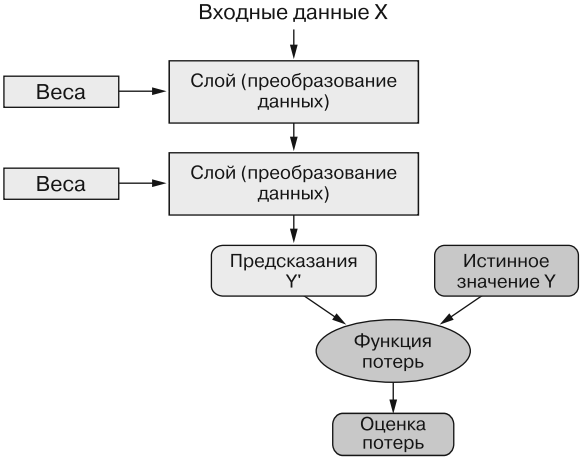


Рис. 1.8. Функция потерь оценивает качество результатов, производимых нейронной сетью

Основная хитрость глубокого обучения заключается в использовании этой оценки для корректировки значения весов с целью уменьшения потерь в текущем примере (рис. 1.9). Данная корректировка является задачей *оптимизатора*, который реализует так называемый *алгоритм обратного распространения ошибки* — центральный алгоритм глубокого обучения. Подробнее об алгоритме обратного распространения ошибки рассказывается в следующей главе.

Первоначально весам сети присваиваются случайные значения, то есть фактически сеть реализует последовательность случайных преобразований. Естественно,

получаемый ею результат далек от идеала и оценка потерь очень высока. Но с каждым примером, обрабатываемым сетью, веса корректируются в нужном направлении и оценка потерь уменьшается. Это *цикл обучения*, который повторяется достаточное количество раз (обычно десятки итераций с тысячами примеров) и порождает значения весов, минимизирующие функцию потерь. Сеть с минимальными потерями, возвращающая результаты, близкие к истинным, называется обученной сетью. Повторим еще раз: это простой механизм, который в определенном масштабе начинает выглядеть непонятным и таинственным.

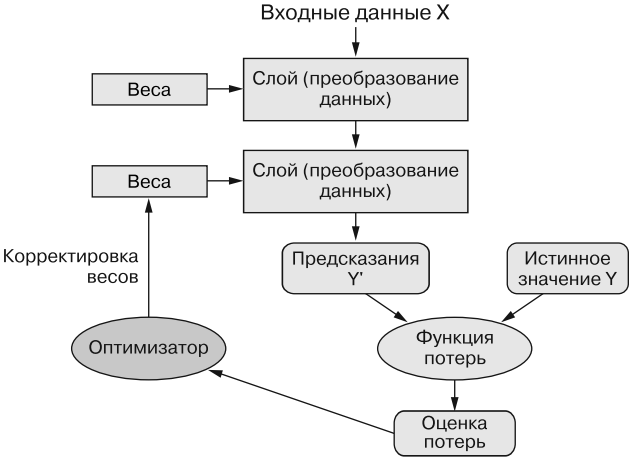


Рис. 1.9. Оценка потерь используется как обратная связь для корректировки весов

1.7. Отличительные черты глубокого обучения

Есть ли что-то особенное в глубоких нейронных сетях, что делает их «правильным» подходом для инвестиций компаний и привлечения исследователей? Будем ли мы по-прежнему использовать глубокие нейронные сети через 20 лет?

Глубокое обучение обладает рядом свойств, которые оправдывают его статус революции в области искусственного интеллекта, и это останется с нами надолго. Возможно, мы не будем использовать нейронные сети через несколько десятилетий, но все, что мы станем применять, будет напрямую наследовать от современного глубокого обучения и его основных концепций. Эти важные свойства можно условно разделить на три категории.

- *Простота*. Глубокое обучение существенно упрощает решение проблем, полностью автоматизируя важнейший шаг в машинном обучении, прежде выполнявшийся вручную, — конструирование признаков. Более ранние методы машинного обучения — методы поверхностного обучения — включали преобразование входных данных только в одно или два последовательных пространства, обычно посредством простых преобразований, чего было недостаточно для большинства задач. Поэтому приходилось прилагать большие усилия, чтобы привести исходные данные к виду, более пригодному для обработки

этими методами, в том числе вручную улучшать представления своих данных. Это называется *конструированием признаков*. Глубокое обучение, напротив, полностью автоматизирует этот шаг: с помощью методов глубокого обучения все признаки извлекаются за один проход — не требуется конструировать их вручную. Таким образом, процесс машинного обучения значительно упростился: часто сложный и многоэтапный конвейер можно заменить единственной, простой моделью глубокого обучения.

- *Масштабируемость*. Глубокое обучение легко поддается распараллеливанию на GPU или TPU, поэтому оно в полной мере может использовать закон Мура. Кроме того, обучение моделей можно производить итеративно, на небольших пакетах данных, что дает возможность реализовывать этот процесс на наборах данных произвольного размера (единственным узким местом является объем доступной мощности для параллельных вычислений, который, как следует из закона Мура, является быстро перемещающимся барьером).
- *Гибкость и готовность к повторному использованию*. В отличие от многих прежних подходов, в глубоком обучении модели могут дообучаться на новых данных без необходимости каждый раз начинать обучение с нуля, что делает их пригодными для непрерывного продолжительного обучения — это чрезвычайно важно для очень больших моделей, работающих в реальных системах. Кроме того, подобные модели можно адаптировать для выполнения новых задач и, соответственно, задействовать повторно. В этом заключается суть *фундаментальных моделей* (foundation models) — больших моделей, обученных на огромных объемах данных, которые можно применять для решения множества новых задач с минимальным дообучением или даже без него.

1.8. Эпоха генеративного ИИ

Пожалуй, наиболее известным примером достижений современного глубокого обучения является недавно поднявшаяся волна приложений генеративного искусственного интеллекта — чат-боты, такие как ChatGPT, Gemini и Claude, а также сервисы генерации изображений, такие как Midjourney. Эти приложения поражают воображение своей способностью создавать информативный и даже творческий контент в ответ на простые запросы, размывая границы между человеческим и машинным творчеством.

Генеративный ИИ основывается на очень больших фундаментальных моделях, которые учатся *восстанавливать* текстовое и графическое содержимое, поступающее в них, — получать четкое изображение из зашумленной версии, предсказывать следующее слово в предложении и т. д. Это означает, что *истинные значения*, показанные на рис. 1.8, извлекаются непосредственно из входных данных. Этот вид обучения называется *самообучением* (self-supervised learning) и позволяет моделям использовать огромные объемы неразмеченных данных. Отсутствие необходимости в ручной разметке данных, которая прежде ограничивала возможности моделей машинного обучения, открыло невиданные ранее горизонты — возможность задействовать фундаментальные модели, которые имеют сотни миллиардов параметров и обучаются на петабайтах данных, что обходится в десятки миллионов долларов.

Фундаментальные модели играют роль своеобразных нечетких баз данных человеческих знаний, что делает их пригодными для очень широкого спектра задач без специального программирования или дообучения. Они хранят огромные объемы данных и поэтому способны решать новые задачи, а нам остается только задавать запросы (промпты), в ответ на которые модели будут обращаться к усвоенным представлениям знаний и возвращать результат, наиболее вероятно связанный с промптом.

Генеративный ИИ получил широкую известность лишь в 2022 году, но он прошел долгий путь — самые ранние эксперименты по генерации текста проводились еще в 1990-х. В первом издании этой книги, вышедшем в 2017 году, уже была объемная глава под названием «Генеративное глубокое обучение», в которой рассматривались существовавшие на тот момент методы генерации текста и изображений, а также отмечалось, что вскоре бóльшая часть потребляемого нами культурного контента будет создаваться с помощью ИИ.

1.9. Чего уже достигло глубокое обучение

За последнее десятилетие глубокое обучение совершило настоящую технологическую революцию, сначала показав впечатляющие результаты в задачах восприятия (2013–2017 годы), затем добившись быстрого прогресса в задачах обработки естественного языка (2017–2022 годы). Кульминацией этого процесса стало множество революционных применений генеративного ИИ (с 2022 года по сей день).

В частности, глубокое обучение достигло следующих прорывов в традиционно сложных областях машинного обучения:

- гибкие и универсальные чат-боты, такие как ChatGPT и Gemini;
- виртуальные помощники, такие как GitHub Copilot;
- создание фотореалистичных изображений;
- классификация изображений на уровне человека;
- распознавание речи на уровне человека;
- распознавание рукописного и печатного текста на уровне человека;
- улучшение качества машинного перевода;
- улучшение качества преобразования письменного текста в звучащую речь;
- управление автомобилем на уровне человека (внедрено в Финиксе, Сан-Франциско, Лос-Анджелесе и Остине к 2025 году);
- повышение точности рекомендательных систем, используемых, например, компаниями YouTube, Netflix и Spotify;
- возможность играть в го, шахматы и покер лучше человека.

Мы все еще продолжаем исследовать возможности, которые таит в себе глубокое обучение. С его помощью мы достигли значительных успехов в решении широкого

круга задач, непосильных для компьютеров еще несколько лет назад, — автоматической расшифровке десятков тысяч древних рукописей, хранящихся в Апостольском архиве Ватикана, определении и классификации болезней растений в полевых условиях с использованием обычного смартфона, интерпретации данных медицинских снимков для онкологов или радиологов, прогнозировании таких стихийных бедствий, как наводнения, ураганы и даже землетрясения. С каждым новым достижением мы приближаемся к эпохе, когда глубокое обучение будет полезно во всех сферах человеческой деятельности — в науке, медицине, производстве, энергетике, транспорте, разработке программного обеспечения, сельском хозяйстве и даже художественном творчестве.

1.10. Не верьте поспешным обещаниям

Эта, казалось бы, нескончаемая череда успехов привела к шумихе, воспевающей возможности ИИ. Некоторые из них действительно существуют, но бóльшая их часть — всего лишь фантазии. В начале 2023 года, вскоре после выпуска GPT-4 компанией OpenAI, многие эксперты утверждали, что люди начнут терять работу и в течение года наступит массовая безработица или что экономическая производительность вскоре вырастет в 10, а то и в 100 раз. Но прошло два года, и ни одно из этих пророчеств не сбылось — уровень безработицы в США остается низким, а показатели производительности далеки от обещанных. Поймите правильно: влияние ИИ, в частности, генеративного, безусловно, ощущается и растет поразительно быстро. По состоянию на середину 2025 года генеративный ИИ приносил десятки миллиардов долларов дохода в год — это впечатляющий результат для отрасли, которой три года назад просто не существовало! Но он пока не оказывает существенного влияния на экономику в целом и меркнет по сравнению с поспешными обещаниями, которыми нас завалили в самом начале.

Дискуссии о безработице и стократном росте производительности труда, обусловленных внедрением ИИ, сами по себе вызывают тревогу, но у ажиотажа, возникшего вокруг ИИ, есть еще более сенсационная сторона — заявления о скором появлении общего интеллекта (artificial general intelligence, AGI) человеческого уровня или даже сверхинтеллекта, намного превосходящего человеческие возможности. Эти заявления подпитывают выходящие за рамки экономических потрясений опасения, что сам человеческий вид может оказаться под угрозой вытеснения нашими цифровыми творениями.

Новичкам в этой области может показаться, что именно практические успехи генеративного ИИ породили веру в скорое появление общего ИИ, но на самом деле это не так. Подобные заявления делались и до появления генеративного ИИ и самим своим существованием подтолкнули его развитие. Еще в 2013 году у специалистов возникли опасения, что общий ИИ может появиться в течение нескольких лет. Тогда считалось, что DeepMind, лондонский стартап по исследованию ИИ, приобретенный компанией Google, находится на пути к достижению этой цели. Эта вера послужила толчком к созданию в 2015 году компании OpenAI, первоначальной целью которой была разработка моделей с открытым исходным кодом, способных конкурировать

с DeepMind. OpenAI сыграла решающую роль в создании генеративного ИИ, то есть, как это ни парадоксально, именно вера в скорое появление общего ИИ подпитывала развитие генеративного ИИ, а не наоборот. В 2016-м в рамках программы привлечения новых сотрудников OpenAI заявляла, что создаст общий ИИ к 2020 году! Справедливости ради, следует отметить, что тогда мало кто из специалистов считал такие оптимистичные сроки реальными. Однако к началу 2023 года значительная часть инженеров, работающих в районе залива Сан-Франциско, была убеждена, что общий ИИ появится в течение следующих нескольких лет.

Подобные заявления следует воспринимать со здоровой долей скептицизма. Современный искусственный интеллект точнее было бы называть когнитивной автоматизацией — практическим применением результатов кодирования человеческих навыков и знаний. ИИ с успехом будет решать узкоспециализированные задачи, а также те, у которых имеется множество точных решений. Проще говоря, речь идет о расширении возможностей компьютеров, а не о воспроизведении человеческого разума.

Чтобы стало яснее, уточним, что когнитивная автоматизация невероятно полезна. Но интеллект, то есть когнитивная автономия, — это совсем другое. Для наглядности можно использовать такую аналогию: ИИ — это персонаж мультфильма, а интеллект — живое существо. Мультфильм, каким бы реалистичным он ни был, может демонстрировать только те сцены, которые были в него заложены. Живое существо, напротив, способно адаптироваться к неожиданностям.

Вы можете спросить: «Если мультфильм нарисован достаточно реалистично и представляет достаточно много сцен, то чем он отличается от живого существа?» Если большая языковая модель может давать на вопросы довольно человекоподобные ответы, то не все ли равно, обладает она истинной когнитивной автономией или нет? Ключевое различие — адаптивность. Интеллект — это способность адаптироваться при встрече с неизвестным и учиться на собственном опыте. Автоматизация, даже в лучшем случае, сможет справиться только с ситуациями, для преодоления которых она была обучена или запрограммирована. Именно поэтому создать надежную автоматизацию так сложно — требуется учесть как можно больше возможных сценариев.

Поэтому не стоит беспокоиться о том, что ИИ внезапно обретет самосознание и захватит человечество. Современные технологии просто не ориентированы на развитие в этом направлении. Даже при значительном прогрессе ИИ останется всего лишь сложным инструментом. Ожидание того, что он сможет стать разумным существом, равносильно ожиданию того, что более совершенные часы сделают возможными путешествия во времени.

1.11. Лето может смениться зимой

Опасность завышенных ожиданий заключается в том, что, когда технология окажется несостоятельной, инвестиции в исследования могут прекратиться и прогресс надолго замрет. Такое уже случалось. ИИ пережил две волны

оптимистического подъема, за которыми следовал спад, сопровождаемый разочарованиями, скептицизмом и в результате снижением финансирования. Все началось с символьного ИИ в 1960-х. В те годы давались весьма многообещающие прогнозы его развития. Один из самых известных пионеров и сторонников символьного ИИ Марвин Минский в 1967 году заявил: «В течение поколения... проблема создания "искусственного интеллекта" будет практически решена». Три года спустя, в 1970 году, он сделал более точное предсказание: «Через три-восемь лет у нас появится машина с интеллектом среднего человека». В 2025-м это достижение все еще кажется далеким — нам сложно предугадать, сколько времени уйдет на него, но в 1960-х и начале 1970-х некоторые эксперты (как и многие люди ныне) полагали, что будущее уже за углом. Несколько лет спустя из-за не оправдавшихся оптимистичных ожиданий исследователи и правительственные фонды отвернулись от этой области — так началась первая *зима ИИ* (метафора вполне уместна: все это происходило вскоре после начала холодной войны).

Этот спад был не последним. В 1980-х интерес к символьному ИИ снова возрос благодаря буму экспертных систем, использовавшихся в крупных компаниях. Первые успехи вызвали волну инвестиций — и отделы ИИ, занимающиеся разработкой экспертных систем, начали появляться в корпорациях по всему миру. К 1985 году компании тратили более миллиарда долларов США в год на развитие технологии. Но к началу 1990-х из-за дороговизны обслуживания, сложностей при масштабировании и ограниченности применения интерес к ней начал падать. Так началась вторая зима ИИ. Сейчас мы подходим к третьему циклу разочарования в ИИ, но пока еще находимся в фазе завышенного оптимизма.

На данный момент нам кажется маловероятной масштабная заморозка исследований в области ИИ, как было в 1990-х. Если и начнется зима, она будет очень мягкой. ИИ уже продемонстрировал свою способность трансформировать мир. Однако, похоже, из пузыря ИИ 2023–2025 годов все же придется выпустить часть воздуха. Сейчас инвестиции в ИИ, прежде всего в центры обработки данных и графические процессоры, превышают 200 млрд долларов в год, в то время как выручка значительно отстает, приближаясь к 30 млрд. Руководители и инвесторы оценивают ИИ не по имеющимся достижениям, а по декларируемым возможностям в будущем, бóльшая часть которых еще долго будет оставаться недоступной для существующих технологий. Что-то должно измениться. Но что именно произойдет, когда пузырь ИИ лопнет, пока неясно.

1.12. Перспективы ИИ

Даже несмотря на наши, возможно, нереалистичные ожидания на ближайшую перспективу, долгосрочная картина выглядит весьма яркой. Мы только начинаем применять глубокое обучение в решении многих важных проблем, от постановки медицинских диагнозов до усовершенствования цифровых помощников.