АНТОЛОГИЯ МАШИННОГО ОБУЧЕНИЯ

важнейшие исследования в области ИИ за последние 60 лет

Терренс Сейновски

Профессор в Институте биологических исследований Солка, директор Института нейронных вычислений Калифорнийского университета, Сан-Диего

БОМБОРА

Терренс Сейновски
Антология машинного обучения. Важнейшие исследования в области ИИ за последние 60 лет

Terrence J. Sejnowski
The Deep Learning Revolution

* * *

Предисловие

Используя распознавание голоса в смартфоне на Android или в Google Переводчике в Интернете, вы сталкиваетесь с нейросетью, натренированной глубоким обучением. За последние несколько лет глубокое обучение обеспечило компании Google прибыль, достаточную для того, чтобы покрыть расходы на все футуристические проекты Google X, включая беспилотные автомобили, очки Google Glass и научно-исследовательский проект Google Brain[1]. Она одной из первых начала применять глубокое обучение. В 2013 году Google наняла Джеффри Хинтона, отца-основателя глубокого обучения, и сейчас другие компании пытаются угнаться за ней.

Современные достижения в области искусственного интеллекта (ИИ) получены благодаря реверсивной инженерии[2] человеческого мозга. Алгоритмы обучения многоуровневых нейронных сетей основаны на том, как нейроны взаимодействуют друг с другом и изменяются в процессе получения опыта. Внутри сети вся многогранность мира превращается в калейдоскоп моделей деятельности, которые и являются основными составляющими ИИ. Модели нейросетей, с которыми я работал в 1980-х годах, едва сравнимы с современными, состоящими из миллионов искусственных нейронов и десятков слоев. Человеческое упорство, огромный объем данных и мощные компьютеры позволили глубокому обучению совершить прорыв в решении самых сложных проблем искусственного интеллекта.

Сложно предугадать, какое влияние новые технологии окажут в будущем. Кто мог предсказать в 90-х годах прошлого века, когда Интернет стал коммерческим, как он повлияет на музыкальный бизнес? А на такси, политические кампании, да и практически все стороны нашей жизни? Когда появились первые компьютеры, тоже

тяжело было вообразить, как они изменят нашу жизнь. В 1943 году Томаса Джона Уотсона, президента IBM, спросили, как повлияют компьютеры на наш мир, и он ответил: «Я думаю, мировой рынок компьютеров вряд ли превысит пять штук». Что действительно сложно представить, так это то, как будет использоваться новое изобретение – и сами изобретатели не скажут больше, чем любой другой человек. Глубокое обучение и ИИ находятся на столь же ранней стадии. Есть множество вариантов развития событий – от утопического и до апокалиптического, – но даже авторы научной фантастики с очень развитой фантазией вряд ли предскажут последствия.

Первые наброски этой книги я сделал через несколько недель после пешего тура по северо-западному побережью Тихого океана и изучения важных изменений в мире ИИ, появившихся десятилетия назад. История рассказывала о небольшой группе ученых, бросивших вызов государственному институту, занимавшемуся вопросами ИИ и не имевшему конкурентов. Они сильно недооценивали сложность задачи и полагались на интуицию, что оказалось ошибкой.

Жизнь на Земле таит в себе множество загадок, и происхождение разума – одна из самых сложных. В природе достаточно его форм, от «интеллекта» простейших бактерий до разума человека, и каждая из них адаптирована к своей нише. Искусственный интеллект так же будет представлен разнообразием форм, которые займут свои места в этом спектре. Так как ИИ основывается на создании глубоких нейронных сетей, по мере своего развития он может подтолкнуть к переосмыслению понятия биологического интеллекта.

Книга, которую вы держите в руках, – гид по прошлому, настоящему и будущему глубокого обучения. Она не охватывает все аспекты данного вопроса – скорее, это личный взгляд на основные достижения, а также на исследователей, их добившихся. Человеческая память, обращаясь к одним и тем же воспоминаниям, все больше их искажает. Этот процесс называется реконсолидацией.

Истории, рассказанные в книге, охватывают период более сорока лет, и хотя некоторые из них свежи в моей памяти так, словно они были вчера, я осознаю, что определенные детали стерлись.

В первой части речь пойдет о предпосылках к рождению глубокого обучения и основных этапах его создания, необходимых для понимания его сути. Во второй части объяснены алгоритмы обучения нейронных сетей с различной структурой. Наконец, в третьей части исследуется влияние ИИ на нашу жизнь. Но, как говорил бейсболист «Нью Йорк Янкиз» Йоги Берра, известный своими «философскими» высказываниями: «Трудно делать прогнозы, особенно насчет будущего». Есть также девять блоков с технической информацией, необязательной для понимая текста. Хронология охватывает события более шестидесяти лет.

Часть I. Переосмысление интеллекта: хронология

1956 – Дартмутский летний исследовательский семинар положил начало разработке ИИ и мотивировал целое поколение ученых исследовать потенциальные возможности информационных технологий с целью добиться воспроизведения ИИ возможностей человека.

1962 – Фрэнк Розенблатт опубликовал книгу «Принципы нейродинамики. Перцептроны[3] и теория механизмов мозга»[4]. В ней были представлены обучающие алгоритмы для моделей однослойных нейронных сетей, ставшие предшественниками современных алгоритмов глубокого обучения.

1962 – Дэвид Хьюбел и Торстен Визел выпустили статью «Рецептивные поля, бинокулярное взаимодействие и функциональная архитектура зрительной коры кошек», где впервые были описаны характеристики отклика нейронов, записанные при помощи микроэлектрода. Архитектура глубокого обучения нейросетей подобна иерархии областей зрительной коры.

1969 – Марвин Минский и Сеймур Пейперт опубликовали книгу «Перцептроны»[5], которая показала вычислительные ограничения перцептронов и ознаменовала начало «зимы» в изучении нейросетей.

1979 – Джеффри Хинтон и Джеймс Андерсон провели в Ла-Хойя в Калифорнии семинар по параллельным моделям ассоциативной памяти, на которых основывались нейросети нового поколения.

1986 – Первая конференция по машинному обучению и системам обработки нейронной информации, проходившая в Денвере, собрала вместе исследователей из различных областей науки.

Глава 1. Развитие машинного обучения

Не так давно считалось, что компьютерная оптическая система не способна сравниться со зрением даже годовалого ребенка. Сейчас это утверждение уже неверно, и компьютеры могут распознавать объекты на изображении так же хорошо, как и человек, а машины на автопилоте едут аккуратнее, чем шестнадцатилетний подросток. Более того, компьютерам никто не говорил, как смотреть или водить, – они научились на собственном опыте, следуя тем же путем, что и природа на протяжении миллионов лет. Их успехи подпитывает огромный объем данных – нового топлива современного мира. Из потока необработанных данных обучающие алгоритмы извлекают информацию. Информация превращается в знание. Знание, в свою очередь, лежит в основе понимания, а понимание порождает мудрость. Это долгий путь, который требует времени. Добро пожаловать в дивный новый мир глубокого обучения![6]

Глубокое обучение – ветвь машинного обучения, основанного на математике, информатике и нейробиологии. Глубокие нейросети учатся на данных, как дети, – исследуя окружающий их мир, переходят от полной неопытности к способности ориентироваться в незнакомой среде.

Глубокое обучение зародилось с появлением информационных технологий в 1950-х годах. Тогда существовали два подхода к созданию ИИ: первый доминировал на протяжении нескольких десятилетий и основывался на логике и компьютерных программах, второй предполагал обучение непосредственно на полученных данных, но занимал гораздо больше времени.

В XX веке, когда компьютеры были намного примитивнее, а хранение данных стоило дороже, чем сегодня, логика оставалась единственным способом решения задач. Опытные программисты писали различные программы для различных задач, и чем

масштабнее была задача, тем сложнее была программа. Сейчас компьютеры обладают большой мощностью, способны обрабатывать огромный объем информации и благодаря особым алгоритмам решают задачи быстрее, точнее и эффективнее. Одни и те же алгоритмы могут использоваться для решения многих задач, и это куда проще, чем писать программу для каждой.

Учим водить

Машина по имени Стэнли (Stanley), сконструированная командой Себастьяна Труна из Стэнфордского университета (рис. 1.1), выиграла два миллиона долларов в гонке беспилотных автомобилей от Управления перспективных исследовательских проектов Министерства обороны США (Defense Advanced Research Projects Agency; DARPA). Стэнли ориентировался в калифорнийской пустыне благодаря машинному обучению. На семимильной трассе встречались узкие туннели и резкие повороты, а также первад Бир-Ботл[7] – ветреная горная дорога с обрывом с одной стороны и горами с другой (рис. 1.2). Вместо того чтобы пойти традиционным путем и написать компьютерную программу, которая могла бы предвидеть любую неожиданность, Трун провел Стэнли по всей пустыне, чтобы машина училась ездить, опираясь на данные с оптических датчиков и датчиков расстояния.

Рис. 1.1. Себастьян Трун на фоне Стэнли, выигравшего в 2005 году гонку беспилотных автомобилей от DARPA. Этот прорыв положил начало технической революции в сфере транспорта

Рис. 1.2. Beer Bottle Pass. Во время гонки беспилотных автомобилей, организованной DARPA в 2005 году, этот сложный участок местности находился ближе к концу трассы длиной 212 километров, пролегавшей в пустыне по бездорожью. Грузовик вдали только начинает подъем

Позже Себастьян Трун основал Google X – исследовательскую лабораторию по разработке высокотехнологичных проектов, где технологии беспилотных автомобилей получили дальнейшее развитие. С тех пор беспилотные автомобили Google проехали по району залива Сан-Франциско миллионы километров. В декабре 2016 года проект был выделен в отдельную компанию Waymo. Uber запустил беспилотные автомобили в Питсбурге. Apple также разрабатывает беспилотные автомобили, чтобы расширить спектр устройств под управлением их операционной системы в надежде повторить свой успех на рынке мобильных телефонов. Производители машин, чьи технологии практически не менялись на протяжении ста лет, следуют по их стопам. General Motors заплатил миллиард долларов за Cruise Automation, проект в Кремниевой

долине, занимающийся разработкой транспорта, который не нуждается в водителе, а также инвестировал шестьсот миллионов долларов в его развитие и совершенствование[8]. Ставки на участие в секторе перевозок, где крутятся триллионы долларов, высоки.

Вскоре беспилотные автомобили станут серьезной проблемой для водителей грузовиков и легковых такси. В конечном итоге не будет необходимости покупать автомобиль, если беспилотные машины смогут прибыть через минуту и безопасно доставить вас к месту назначения. Кроме того, вам не нужно будет парковаться! Среднестатистический автомобиль проводит четыре процента времени в дороге, а остальные 96 стоит без дела. Огромные участки в городах, которые сейчас занимают парковки, можно будет использовать для других целей, тогда как беспилотные автомобили станут парковаться за городом. Также это повлияет на многие другие сферы, например на страховые компании и магазины запчастей. Станет гораздо меньше смертей из-за вождения в нетрезвом виде и из-за того, что водители засыпают за рулем. Время, которое мы тратим, чтобы добраться до работы, можно будет использовать для других целей. Согласно переписи населения, проведенной в США в 2014 году, 139 миллионов человек тратят на дорогу на работу и с нее в среднем 26 минут в каждую сторону. Это 29,6 миллиарда часов в год, целых 3,4 миллиона лет человеческих жизней, которые можно было бы использовать гораздо лучше[9]. Кто захочет угнать машину без руля, которая, вдобавок ко всему, еще и сама вернется домой? Придет конец автомобильным кражам. Пока еще на этом пути стоит множество нормативных и правовых препятствий, однако когда беспилотные автомобили начнут использовать повсеместно, мы будем жить в дивном новом мире. Первыми – вероятно, уже лет через десять – беспилотными станут грузовики, такси – через пятнадцать, а личные автомобили завершат переход лет через 25–50.

Беспилотные автомобили – лишь самая заметная часть сдвига в экономике, вызванного информационными технологиями. Данные

текут в Интернете, как вода по городскому трубопроводу. Они собираются в огромных информационных центрах, управляемых такими компаниями, как Google, Amazon, Microsoft и др. Для их работы требуется огромное количество электроэнергии, поэтому центры располагаются рядом с гидроэлектростанциями – при передаче потока информации вырабатывается столько тепла, что только реки могут его охладить. В 2013 году информационные центры в США потребили 10 миллионов мегаватт, что сравнимо с энергией, которую вырабатывают 34 большие электростанции[10]. Но гораздо большее значение для экономики имеет то, как используются эти данные. Необработанная информация превращается в знание о людях: что вы делаете, чего хотите и что вообще из себя представляете. Более того, эта информация передается от вас через устную речь.

Учим переводить

В настоящее время глубокое обучение применяется в компании Google для сотни приложений, от Street View и до Inbox Smart Reply, а также для голосового поиска. Несколько лет назад инженеры Google поняли, что необходимо доработать эти приложения до очень высокого уровня, и приступили к созданию специального чипа, предназначенного для глубокого обучения. Для удобства плата спроектирована так, что входит в стандартный слот для жесткого диска в стойке центра обработки данных. Тензорный процессор Google (Google Tensor Processing Unit; Google TPU) сегодня внедрен на множестве серверов по всему миру, значительно повышая производительность приложений с глубоким обучением.

Рис. 1.3. Приложение Google Translate мгновенно переводит с других языков дорожные указатели, стоит навести на них камеру. Это особенно актуально, если вам нужно сесть на поезд в Японии

Пример того, как быстро глубокое обучение может изменить мир, – его влияние на перевод с иностранных языков. Перевод с одного языка на другой – заветная мечта ИИ, поскольку основан на понимании предложений целиком. В 2016 году компания Google запустила новый Переводчик, основывающийся на глубоком обучении, что стало большим шагом на пути к живому переводу. Буквально в одночасье перевод превратился из беспорядочного смешения отдельных фраз в связные предложения (рис. 1.3). Раньше программа искала комбинации слов, которые можно было бы перевести вместе, но глубокое обучение создает перевод, исходя из смысла всего предложения.

18 ноября 2016 года научный сотрудник Токийского университета Юн Рекимото заметил внезапное усовершенствование Google Переводчика. Чтобы протестировать новую систему, он перевел в приложении начало рассказа Эрнеста Хемингуэя «Снега Килиманджаро» на японский, а затем обратно на английский. Читателю нужно определить, какой отрывок принадлежит Хемингуэю, а какой – Google Переводчику[11]:

1. Килиманджаро – покрытый вечными снегами горный массив высотой в 19 710 футов, как говорят, высшая точка Африки. Племя масаи называет его западный пик «Нгайэ-Нгайя», что значит «Дом Бога». Почти у самой вершины западного пика лежит иссохший мерзлый труп леопарда. Что понадобилось леопарду на такой высоте, никто объяснить не может[12].

2. Килиманджаро – это заснеженная гора высотой 19 710 футов, которая считается самой высокой горой в Африке. Его западная вершина называется Масаи «Нгадже Нгаи», Дом Бога. Рядом с западной вершиной находится высушенная и замороженная туша леопарда. Никто не объяснил, что искал леопард на такой высоте[13].

Следующая цель глубокого обучения – научить автопереводчик работать с абзацами, чтобы он мог выявлять связи между несколькими предложениями. У слов глубокие культурные корни.

Владимир Набоков, автор романа «Лолита», писавший и на русском, и на английском, пришел к выводу, что невозможно переводить поэзию. Его литературный перевод на английский язык «Евгения Онегина» Пушкина[14] дополнен пояснениями о культуре той страны и того времени, в котором создавался оригинал; необходимость давать такие сноски подтверждает его точку зрения. Но, возможно, однажды Google Переводчик сможет переводить произведения Шекспира, опираясь на контекст его творчества в целом[15].

Учим слушать

Еще одна заветная мечта ИИ – распознавание устной речи. До недавнего момента оно применялось в ограниченных областях, например при бронировании авиабилетов. Теперь же возможности безграничны. Летний исследовательский проект Microsoft Research, осуществленный в 2012 году стажером из университета Торонто, значительно улучшил систему распознавания речи (рис. 1.4)[16]. В 2016 году одно из подразделений Microsoft заявило, что в результате применения глубокого обучения они достигли эффективности, сравнимого с человеческой[17].

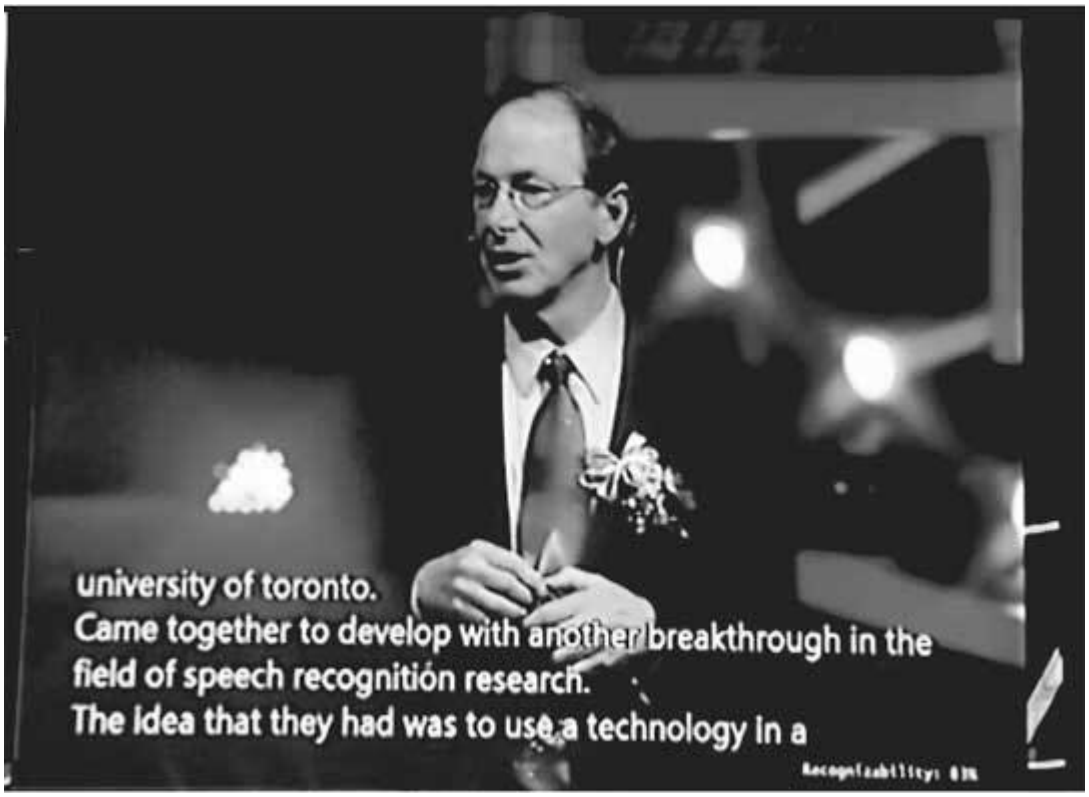


Рис. 1.4. Ричард Рашид, руководитель отдела исследований компании Microsoft, на презентации функции распознавания голоса, использующей глубокое обучение. 25 октября 2012 года в Тяньцзинь в Китае. Две тысячи китайских студентов в аудитории увидели субтитры, созданные с помощью автоматического распознавания речи, которые следовали за устным переводом на китайский язык. Это стало всемирной сенсацией

Последствия этого прорыва будут ощущаться в обществе в ближайшие годы, и в итоге голосовой интерфейс вытеснит клавиатуру. Это уже начало происходить с появлением виртуальных помощников, таких как Алекса, Сири и Кортана, разработчики которых постоянно стремятся превзойти друг друга. Как печатные машинки исчезли из-за повсеместного распространения компьютеров, так и клавиатуры вскоре станут всего лишь экспонатами музеев.

Когда функция распознавания речи соединится с функцией автоматического перевода, станет возможно межкультурное общение в режиме реального времени. Почему же требуется так много времени, чтобы они вышли на тот же уровень, что и у человека? Просто ли совпадение, что они и другие когнитивные способности достигли своего предела одновременно? Ко всем этим достижениям привели огромные потоки данных.

Учим ставить диагноз

Сфера услуг и профессии также изменится с развитием машинного обучения, когда оно начнет применяться в тех областях, где будет доступ к большим массивам данных. Медицинские диагнозы, опирающиеся на информацию о миллионах пациентов, станут более точными. Во время недавнего исследования глубокое обучение было применено к медицинской базе данных, в которой содержалось 130 тысяч изображений, иллюстрирующих более двух тысяч различных дерматологических заболеваний, что в десять раз больше, чем использовалось ранее (рис. 1.5)[18]. Систему обучили определять заболевания, исходя из изображений, которые ей были до этого неизвестны. В результате система поставила диагнозы, которые не отличались, а в некоторых случаях даже были точнее, которые поставили 21 врач-дерматолог. Вскоре каждый при помощи смартфона сможет сфотографировать подозрительное высыпание на коже и незамедлительно узнать диагноз. Без посещения доктора, длительного ожидания в очереди перед осмотром и потраченной солидной суммы денег, как сейчас. Значительно расширится объем и качество дерматологического лечения. Если пациент сможет быстро получить экспертную оценку, он придет к доктору на ранней стадии заболевания и его будет гораздо проще вылечить. Да и сами врачи станут лучше определять кожные заболевания при помощи глубокого обучения[19].

Если у вас серьезные проблемы со сном, что случается у 70 процентов людей, то вы запишетесь на прием к доктору, и, за исключением критических ситуаций, может пройти несколько месяцев, до того как вас направят в специализированную клинику. В клинике вам проведут обследование во время ночного сна. Вас облепят десятками электродов для записи электроэнцефалограммы и мышечной активности в то время, пока вы спите. Когда вы засыпаете, мозговые волны на вашей ЭЭГ меняют низкую амплитуду на высокую при переходе в стадию медленного сна, и считать согласованность волн через волосистую часть головы становится намного проще. В течение ночи мозг переключается на другую стадию сна, которая сопровождается быстрым движением глаз.

Рис. 1.5. Обложка журнала Nature от 2 февраля 2017 года. Взгляд художника на диагностирование кожных заболеваний при помощи глубокого обучения

В это время вы видите сны. Бессонница, задержка дыхания во сне (апноэ), синдром беспокойных ног и другие расстройства нарушают схему. Если вам трудно засыпать даже дома, то попытка заснуть в чужой кровати с проводами от медицинского оборудования тем более будет для вас проблемой. Всю следующую неделю доктор будет изучать вашу ЭЭГ и отмечать стадии сна блоками по 30 секунд. Потребуется много времени, чтобы добиться восьмичасового сна. В конце концов вы получите заключение о нарушениях режима сна и счет на две тысячи долларов.

Врачи-сомнологи обучаются по системе наблюдения за стадиями сна, разработанной Рехтсшафеном и Кэйлсом в 1968 году[20]. Тем не менее два эксперта согласятся друг с другом только в 75 процентах случаев, так как особенности сна часто неоднозначны и противоречивы. Филип Лоу, бывший аспирант моей лаборатории, использовал машинное обучение для автоматического определения стадий сна всего за три секунды с достоверностью 87 процентов, что занимает меньше минуты работы компьютера. Более того, нужен всего один провод, закрепленный в одном месте на поверхности головы, что гораздо удобнее, чем пучки проводов, которые сложно ставить и снимать. В 2007 году мы запустили проект Neurovigil, направленный на внедрение этой технологии в специализированных клиниках. Мы были удивлены, когда они не проявили к нему интереса, так как это снизило бы их доход. Пока страховые компании оплачивают большие счета, выписанные пациентам, клиникам невыгодно внедрять более дешевые методы. Они так же зарабатывают на производителях лекарств, ведь тем необходимо тестировать воздействие своих препаратов на сон. Neurovigil сейчас внедряется на рынок долгосрочного медицинского ухода, ведь у пожилых часто проблемы со сном.

Модель сомнологических клиник несовершенна, так как трудно диагностировать проблему, пользуясь только одним методом. У каждого человека свои особенности, которые для него нормальны, и наиболее информативными являются отклонения от этого состояния. Для проекта Neurovigil создано специальное устройство iBrain, которое может записать вашу ЭЭГ дома, отправить данные через Интернет и проанализировать ее на предмет отклонений. Это позволит докторам выявлять проблемы на ранних стадиях, когда их проще лечить и не допустить, чтобы они перешли в хронические. Есть и другие болезни, чье лечение улучшится от продолжительного наблюдения, как, например, сахарный диабет 1-го типа, при котором уровень сахара в крови можно отслеживать и регулировать введением инсулина. Недорогие устройства, на протяжении

определенного времени фиксирующие данные, сильно повлияют на диагностику и лечение разных хронических заболеваний.

Из этого опыта можно извлечь несколько уроков. Даже имея более дешевую и совершенную технологию, ее будет трудно внедрить, если другой, пусть и дорогой, продукт закрепился на рынке. Тем не менее есть второстепенные рынки, где новая технология распространится быстрее, так как может экономить время и успешнее конкурировать. Именно так появились солнечная энергетика и ряд новых отраслей. В перспективе мониторинг сна с помощью новых технологий тоже будет доступен пациентам как дома, так и в медицинских клиниках.

Учим зарабатывать деньги

Более 75 процентов торговых сделок на Нью-йоркской фондовой бирже автоматизированы (рис. 1.6) и проводятся благодаря высокоскоростным алгоритмам, которые реагируют гораздо быстрее человека. Более того, алгоритмы начинают зарабатывать деньги все лучше и лучше, а глубокое обучение позволяет систематически увеличивать прибыль. В 1980-х я работал в компании Morgan Stanley консультантом по использованию нейросетей на фондовых биржах и встретил там Дэвида Шоу, программиста из Колумбийского университета, который специализировался на параллельных вычислениях. На заре автоматической торговли он работал в отделе количественного анализа данных даже во время отпуска. Когда вам не нужно платить за каждую транзакцию, даже незначительное преимущество может превратиться в крупную прибыль. Шоу ушел из Morgan Stanley, чтобы создать свою компанию по управлению инвестициями на Уолл-стрит – The D. E. Shaw Group. Сейчас он мультимиллиардер.

Компания Шоу достигла значительного успеха, однако ей далеко до страхового фонда Renaissance Technologies, основанного Джеймсом Саймонсом, выдающимся математиком и бывшим заведующим кафедрой математики Университета штата Нью-Йорк в Стоуни-Брук. В 2016 году Саймонс в одиночку заработал 1,6 миллиарда долларов[21], и это далеко не самая большая его прибыль. Фонд Renaissance был назван «компанией с лучшими физиками и математиками в мире»[22], которая «избегает нанимать любого, кто связан с Уолл-стрит»[23].

Дэвид Шоу больше не занимается повседневной работой в D. E. Shaw, сейчас он поглощен проектом D. E. Shaw Research по созданию компьютера для параллельных вычислений под названием Anton, который выполняет расчет сворачивания белка гораздо быстрее, чем любой другой компьютер на планете[24]. Саймонс ушел

из Renaissance и вместе со своей женой основал благотворительный фонд, который поддерживает исследование аутизма и другие проекты по физике и биологии. Фонд спонсирует работу Института теории вычислений Саймонса в Беркли в Калифорнии, Центра социального мозга Саймонса при Массачусетском технологическом институте[25], а также Института Флэтайрон в Нью-Йорке.

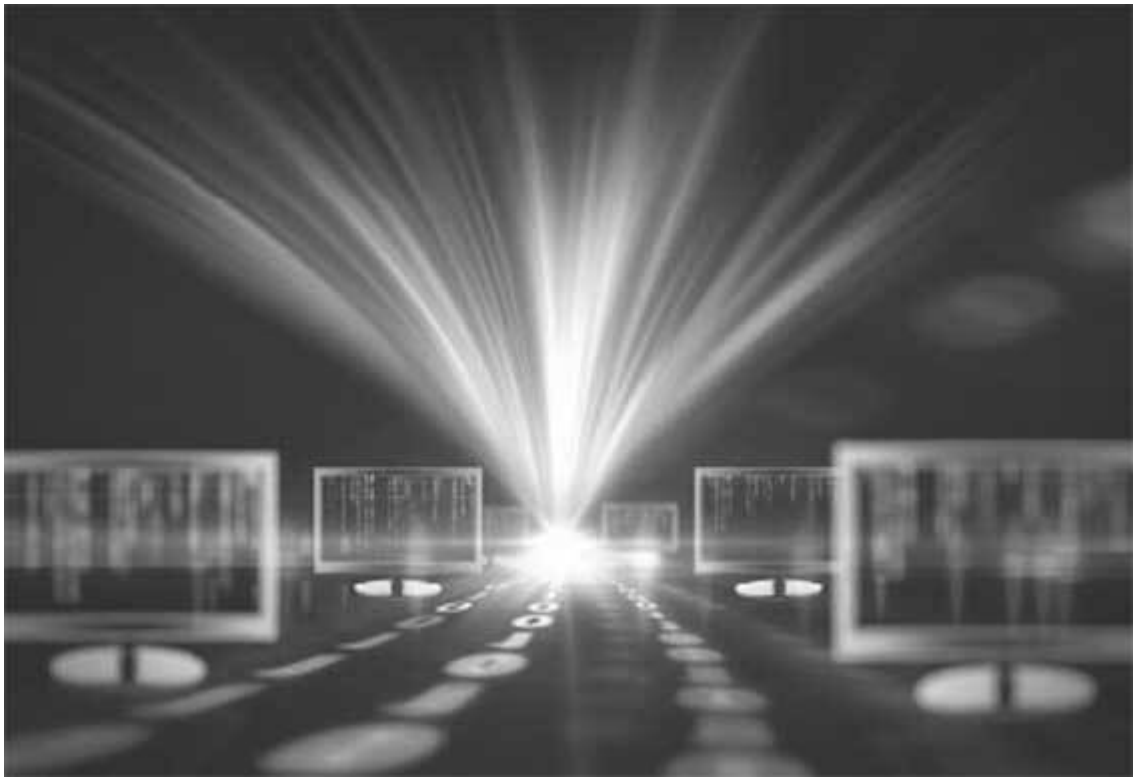


Рис. 1.6. Машинное обучение управляет высокоскоростной торговлей на фондовых рынках. Для достижения наилучшего результата совмещают несколько моделей машинного обучения[26]

Глубокое обучение только начинает влиять на труд юристов. Большая часть рутинной работы в юридических организациях, стоящая сотни долларов в час, будет автоматизирована, особенно в крупных компаниях. В частности, ИИ, не чувствуя усталости, может выполнять анализ тысяч документов в поисках доказательств[27]. Еще одно преимущество автоматизированной системы – полное

соблюдение постоянно усложняющихся нормативных требований. Юридическая консультация станет доступна любому, кто не может себе позволить нанять адвоката. Работа юристов станет не только дешевле, но и гораздо быстрее, а этой порой важнее стоимости. Правовой мир станет юридически глубоким.

Учим играть в покер

Безлимитный техасский холдем «один на один» входит в число самых популярных разновидностей покера. В нее обычно играют в казино, а также на главном состязании – Мировой серии покера. Покер сложен, потому что, в отличие от шахмат, где оба игрока владеют одинаковым объемом информации, у игроков в покер информация неполная. Поэтому при игре на высоком уровне умение блефовать и вводить в заблуждение не менее важно, чем сами карты.

Рис. 1.7. Безлимитный техасский холдем «один на один». Пара тузов на руках. Блеф на высоких ставках был освоен системой DeepStack, которая победила профессиональных игроков с большим отрывом

Джон фон Нейман, математик, создавший математическую теорию игр и заложивший основы архитектуры вычислительных машин, был очарован покером, так как «реальная жизнь вся состоит из блефа, маленьких хитростей и размышлений, что другой человек думает о том, что собираюсь сделать я. Игры в моей теории как раз такие». Покер отражает ту часть человеческого интеллекта, которая была усовершенствована в процессе эволюции. К величайшему удивлению экспертов в покере, сеть глубокого обучения DeepStack сыграла 44 852 игры против 33 профессиональных игроков в покер и победила их на четыре стандартных отклонения[28]. Невероятный успех. Победу над лучшими игроками при использовании даже одной стратегии уже можно было бы назвать прорывом. Если это достижение применить и в других сферах человеческой деятельности, где решения принимаются при отсутствии полной информации, например в политике и международных отношениях, последствия могут быть далеко идущими[29].

Учим играть в го

В марте 2016 года кореец Ли Седоль, чемпион мира по го, сыграл матч против AlphaGo – программы, обученной этой игре (рис. 1.8)[30]. AlphaGo использовала нейросеть глубокого обучения, чтобы оценить расположение камней на доске и возможные ходы. Го сложнее шахмат, как шахматы сложнее шашек. Если шахматы – одно сражение, то го – война. Доска для игры в го размером 19 на 19, что значительно больше, чем шахматная доска 8 на 8 клеток. В го возможно одновременно вести несколько битв на разных частях доски. В игре есть множество нюансов, поэтому судить ее порой сложно даже экспертам. Существуют 10^{170} возможных позиций, что больше, чем количество атомов в наблюдаемой Вселенной.

AlphaGo применяла несколько нейросетей глубокого обучения для оценки ситуации на доске и выбора наилучшего хода. Кроме того, у

нее совершенно другая система обучения, использовавшаяся для решения задач, в которых необходимо вычислить, какие действия приведут к успеху, а какие – к неудаче. Если я выигрываю в го, какие мои действия способствовали этому? А если проигрываю, какой шаг был неверным? Часть человеческого мозга, которая отвечает за решение таких задач, – базальные ганглии. Они получают проекции сигналов с коры головного мозга и передают их обратно. AlphaGo использует алгоритмы, которые применяются базальными ганглиями для вычисления наиболее успешной последовательности действий. Об этом подробно будет рассказано в главе 10. Таким образом, AlphaGo училась, играя с собой раз за разом.

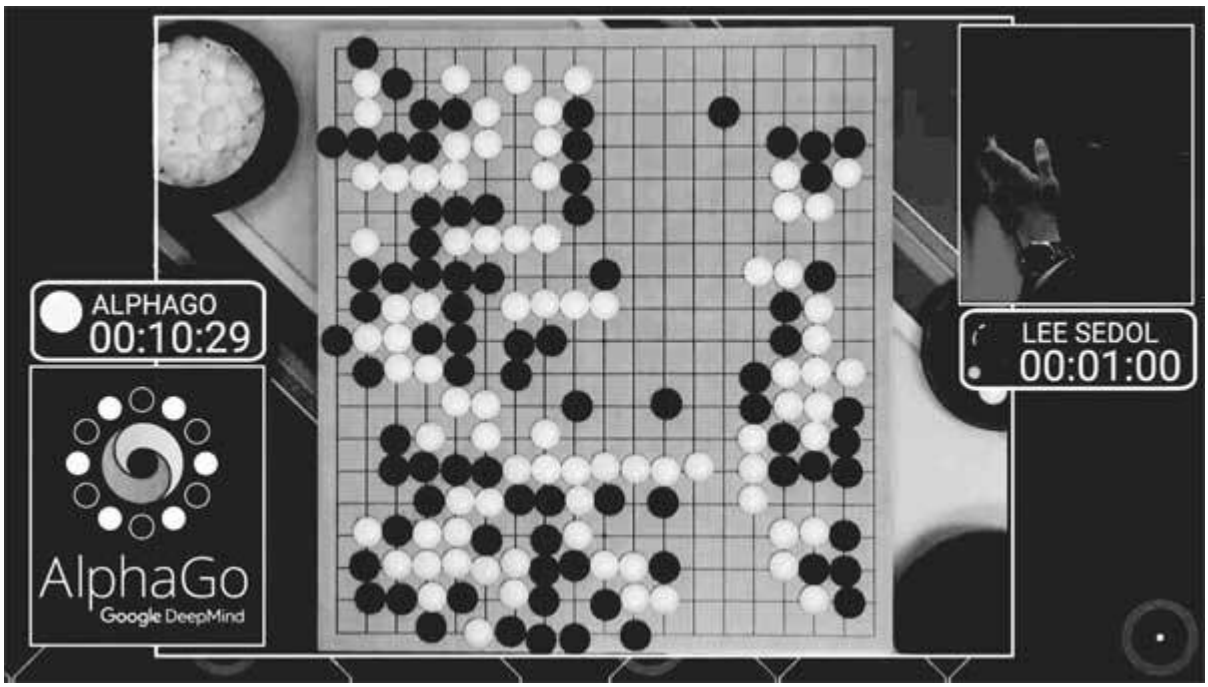


Рис. 1.8. Матч между Ли Седолем и AlphaGo. Доска во время матча из пяти игр между корейским чемпионом и нейросетью, которая научилась играть сама

Результат матча в го, когда AlphaGo обыграла Ли Седоля, сильно повлиял на население Азии, где чемпионы по го – едва ли не национальные герои, подобно рок-звездам. Ранее AlphaGo обыграла

чемпиона Европы, но сама по себе игра была не очень высокого уровня, поэтому Ли Седоль не ожидал столкнуться с серьезным соперником. Даже DeepMind, компания, создавшая AlphaGo, не ожидала такой сильной игры. С момента последнего матча AlphaGo сыграла сотни миллионов игр с разными своими модификациями, и едва ли можно выразить словами, насколько хороши были эти партии.

Для многих стало потрясением, когда AlphaGo выиграла первые три игры из пяти, продемонстрировав высокий уровень игры (рис. 1.9). Это было захватывающее зрелище в Южной Корее, которое обозревали комментаторы самых известных телеканалов. Некоторые ходы AlphaGo были поистине революционными. Ее 37-й ход во второй партии был настолько потрясающим, что Ли Седолю понадобилось десять минут для ответного хода. AlphaGo проиграла четвертую партию, и этим человеческий интеллект хоть немного отстоял свою честь. Тем не менее матч закончился со счетом 4:1 в пользу AlphaGo. Я наблюдал за ним в предрассветные часы в Сан-Диего, словно загипнотизированный. Это напомнило мне события 2 июня 1966 года, когда я смотрел по телевизору, как роботизированный космический корабль Surveyor приземлился на Луну и прислал первую фотографию ее поверхности[31]. Я стал свидетелем исторического события. AlphaGo совершила то, что было для нас за гранью возможного.

Рис. 1.9. Ли Седоль после проигрыша в матче с AlphaGo: «Я не знаю, что сказать и с чего начать, но мне кажется, я должен извиниться. Я должен был показать лучший результат, и я прошу прощения, что не удовлетворил ожидания людей. Я чувствую себя бессильным. Если бы я мог повернуть время вспять и вернуться к самой первой игре, я бы все равно не выиграл, потому что недооценил возможности AlphaGo»

4 января 2017 года в онлайн-версии игры го был разоблачен один из игроков под псевдонимом Master. Им оказалась AlphaGo 2.0. Ее раскрыли после 60 побед в 60 играх против лучших мировых игроков, среди которых был чемпион мира девятнадцатилетний Кэ Цзе из Китая. AlphaGo показала новый стиль игры, который идет вразрез с вековой стратегией. 27 мая 2017 года Кэ Цзе проиграл AlphaGo три игры на саммите «Будущее го» в Вузхене в Китае (см. рис. 1.8). Это были одни из лучших игр в го, и сотни миллионов китайцев следили за матчем. Кэ Цзе сказал: «В прошлом году я думал, что стиль игры AlphaGo близок к человеческому. Но сегодня я

понял, что она играет как бог игры го»[32]. AlphaGo также обыграла команду из пяти лучших игроков в ходе недельной серии матчей. Участники проанализировали ходы AlphaGo и изменили свою стратегию. Чемпионат был организован правительством Китая, что стало новым вариантом «пинг-понговой дипломатии». Китай делает большие инвестиции в развитие машинного обучения, а главная цель – обучение ИИ новым алгоритмам[33].

После проигрыша с отставанием всего в 0,5 очка Цзе сказал, что был близок к выигрышу в середине игры: «Я чувствовал, как бьется мое сердце. Возможно, именно из-за волнения я и совершил несколько ошибок. Возможно, это самое слабое место в человеке». То, что испытал Кэ Цзе, было эмоциональной перегрузкой, но в то же время эмоции необходимы для достижения максимальной производительности. При низком эмоциональном возбуждении умственные способности не максимальны. Актеры театра знают: если у них не летают бабочки в животе перед выступлением, их игра будет не особо хорошей. Их эмоции можно представить в форме перевернутой буквы *U*, а лучший результат достигается между низким и высоким уровнем возбуждения. Спортсмены называют это «быть в потоке»[34].

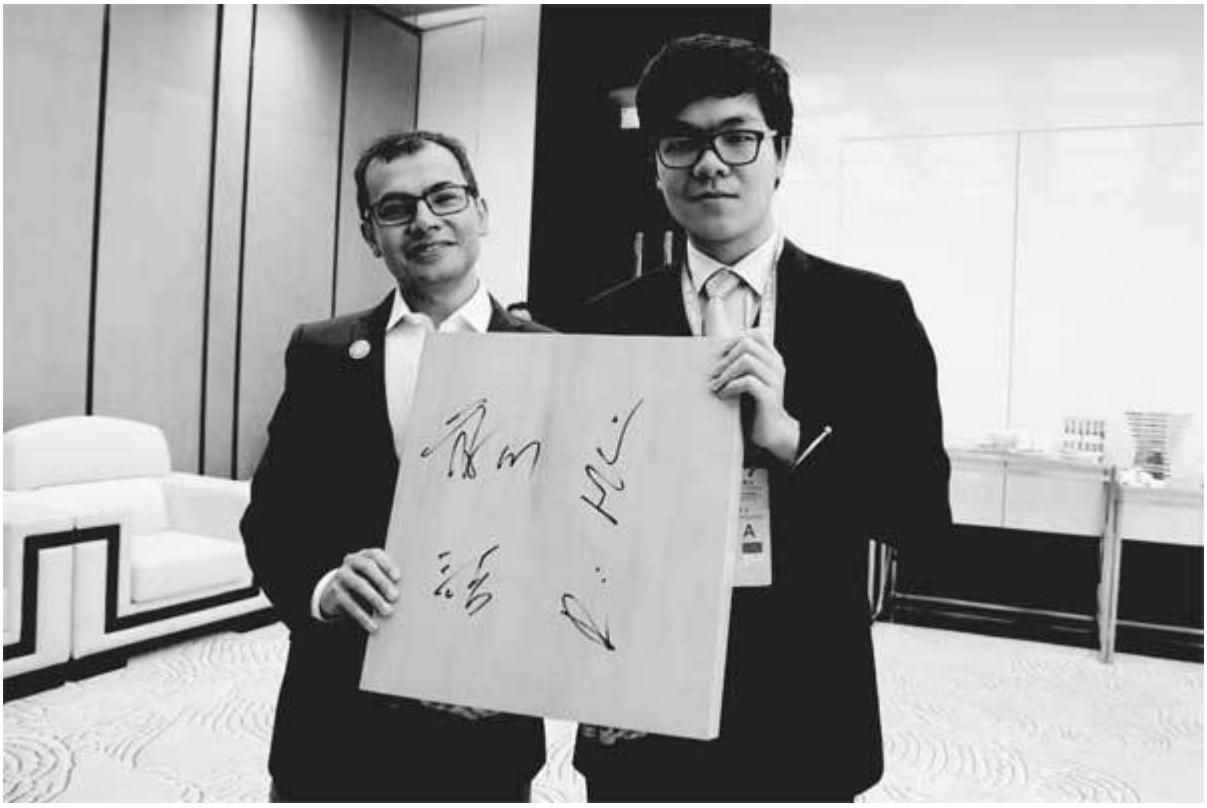


Рис. 1.10. Встреча Демиса Хассабиса (слева) и Кэ Цзе после легендарной игры в го в Китае. В руках у Хассабиса доска с автографом Цзе

В 2010 году соучредителем компании DeepMind стал Демис Хассабис (рис. 1.10), нейробиолог, научный сотрудник Университетского колледжа Лондона, а также моей лаборатории. В 2017 году он совместно с Рэймондом Доланом и Вольфрамом Шульцем выиграл престижную премию Brain Prize за исследование системы вознаграждения мозга. В 2014 году корпорация Google приобрела компанию DeepMind за 600 миллионов долларов. В компании работают более четырехсот инженеров и нейробиологов, которые совмещают академические знания с инновациями. Союз нейробиологии и ИИ становится все крепче и крепче.

Учим становиться умнее

Можно ли назвать AlphaGo умной? Об интеллекте написано больше, чем по любой другой теме в психологии, за исключением темы разумности – и то и другое трудно поддается определению. С 1930-х годов психологи различают подвижный[35] и кристаллизованный интеллект. Кристаллизованный интеллект основан на знаниях, таких как словарный запас, и его уровень можно измерить стандартными IQ-тестами. Подвижный интеллект – это способность решать проблемы с помощью логических рассуждений, выходя за пределы предыдущего опыта. Уровень подвижного интеллекта следует по особой траектории развития, достигая пика в молодости и понижаясь с возрастом, в то время как кристаллизованный интеллект с каждым годом постепенно растет и в конечном счете достигает своего предела. AlphaGo представляет собой соединение кристаллизованного и подвижного интеллекта в достаточно узкой области, однако внутри этой области она демонстрирует удивительные творческие способности. Профессиональный опыт также основан на обучении в ограниченной сфере. Например, мы профессионалы в программировании и пользуемся этим каждый день.

Алгоритм обучения с подкреплением, используемый AlphaGo, может применяться для решения различных задач. Этот метод основывается на награде, которую получает победитель в конце последовательности ходов, что, как ни странно, может усовершенствовать решения, принятые ранее. В сочетании со множеством мощных сетей глубокого обучения появляется огромное количество разрозненной информации, зависящей от области знания. И действительно, подобные ситуации были смоделированы для социального, эмоционального, механического и конструктивного интеллекта[36]. Фактор общего интеллекта (g-фактор) взаимосвязан с этими навыками. Есть причины осторожно относиться к

интерпретации тестов IQ. Средний уровень IQ растет во всем мире на три пункта каждые десять лет с момента его первого измерения в 1930-х. Явление называется эффектом Флинна[37]. У феномена есть множество объяснений, включая улучшение питания, повышение внимания к собственному здоровью, а также роль окружающей среды. Это кажется достаточно правдоподобным, потому что окружающая среда влияет на регуляцию генов, что, в свою очередь, влияет на мозг и приводит к изменениям в поведении. Может ли сложиться так, что люди будут становиться все умнее и умнее на протяжении длительного времени? И как долго процесс будет продолжаться? Уровень игроков в шахматы, шашки, а теперь и в го неуклонно растет с тех пор, как появились компьютерные программы, которые играют не хуже чемпионов, и это форма усиления интеллекта[38]. Глубокое обучение повысит IQ людей всех профессий, в том числе и ученых-исследователей.

Изменение рынка труда

Банкомат – робот, выполняющий часть работы банковского служащего. Банкоматы принимают и выдают деньги, и это очень удобно, ведь теперь вы можете снять наличные в любое время суток. Сейчас банкоматы умеют считывать даже рукописные чеки! Банкоматы стали выполнять часть рутинной работы кассиров, однако люди не остались без работы. Теперь они лично консультируют клиентов по вопросам ипотеки и инвестиций. Во время индустриальной революции паровые двигатели заменили ручной труд, но в то же время создали рабочие места для инженеров, которые строят и обслуживают их, а позже и для машинистов, которые управляют паровозом. Amazon лишил бизнеса многих мелких продавцов, однако взамен предоставил более 350 тысяч рабочих мест, например в службе доставки. По мере того как ИИ заменяет различные профессии, появляются новые ниши для

человеческого труда, требующие создание систем и устранения неполадок в них.

Такой круговорот профессий не нов. Фермеры XIX века были вытеснены техникой, но в то же время благодаря изобретению парового двигателя потребовалась система образования, дающая работникам новые навыки. Современная ситуация отличается лишь тем, что новые профессии требуют умственных способностей. Для поддержания работы ИИ необходимы определенные знания, поэтому будьте готовы учиться всю жизнь. Чтобы люди, чьи профессии заменит ИИ, смогли получить новую работу, нам нужна новая система образования, которая будет основываться на домашнем обучении.

К счастью, в настоящее время есть множество бесплатных онлайн-курсов, на которых вы можете приобрести актуальные знания и навыки. Онлайн-курсы активно внедряются в образовательную систему. Конечно, они только начинают развиваться, но уже имеют большой потенциал, так как дают шанс учиться огромному количеству людей. Со временем система онлайн-курсов может измениться. Вместе с Барбарой Оакли и проектом Coursera мы основали известные онлайн-курсы «Учимся учиться» («Learning How to Learn»)[39], которые сделают из вас хорошего ученика (рис. 1.11). Также мы создали курс Mindshift[40], который поможет открыть в себе новые способности и изменить свой образ жизни. Эти онлайн-курсы будут описаны в главе 12.

Рис. 1.11. «Учимся учиться», масштабный онлайн-курс, который помогает улучшить свои способности к обучению. Это самый популярный в Интернете онлайн-курс, по которому занимаются более двух миллионов человек

Когда вы что-то делаете в Интернете, вы невольно оставляете о себе много информации. Реклама, которую вы видите, подобрана на основе ваших запросов в Сети. Информация, которую вы сохраняете в Facebook и других социальных сетях, может быть использована для создания вашего личного помощника, который знает вас лучше, чем кто-либо другой. Он ничего не забудет и фактически станет вашим двойником. У ваших детей будут личные наставники, сопровождающие их на протяжении всего процесса обучения. У детей будущего возможности для образования будут лучше, чем самые лучшие из доступных сегодня. Переход к высокоточному образованию может стать довольно быстрым по сравнению с переходом к беспилотным автомобилям, потому что физические препятствия намного ниже, спрос намного выше, а само образование – это рынок с оборотом в триллион долларов[41].

Искусственный интеллект – реальная угроза?

После того как в 2016 году AlphaGo победила Ли Седоля в го, это вызвало новую волну опасений, что ИИ потенциально опасен для человечества. Программисты подписали обязательство не использовать ИИ в военных целях. Стивен Хокинг и Билл Гейтс сделали публичное заявление о реальной угрозе, которую может представлять ИИ. Илон Маск и другие предприниматели Кремниевой долины создали компанию OpenAI с капиталом в один миллиард долларов и наняли Илью Суцкевера, бывшего студента Джеффри Хинтона[42], на пост генерального директора. Основная цель этого проекта – убедить людей, что будущие открытия станут доступны каждому. Еще одной целью было предотвратить злоупотребление новейшими технологиями со стороны частных компаний. ИИ практически в одночасье перестал быть угрозой. Обе цели, конечно, преувеличены, но результат был достигнут.

Должны ли мы бояться ИИ? Не в первый раз инновации воспринимаются как угроза. Мы научились жить с ядерным оружием и не развязали ядерную войну. Когда технология рекомбинантных ДНК была открыта, люди боялись, что смертельно опасные организмы будут выпущены на свободу. Генная инженерия стала серьезной наукой, а мы живы до сих пор. Точно так же мы привыкнем и к искусственному интеллекту.

Одним из последствий дальнейшего развития DeepStack может стать то, что он превратится в обманщика мирового класса. То, что может сделать сеть, ограничивается только вашим воображением. Если сеть можно обучить самостоятельно водить автомобиль, ее также можно обучить участвовать в гонках «Формула-1», и кто-нибудь наверняка захочет в это вложиться. Сегодня для создания сетей, использующих глубокое обучение, требуются особые знания и навыки, но со временем, когда для разработки программ с ИИ нужны будут компьютеры с меньшей мощностью, а программное

обеспечение станет автоматизированным, даже школьникам будет доступно создание приложений с ИИ. Кто знает, что они сделают?

Otto, один из самых популярных интернет-магазинов по продаже одежды, мебели и товаров для спорта в Германии, использует глубокое обучение для того, чтобы, опираясь на предыдущие заказы клиента, предугадать, что он закажет на этот раз, и оформить для него предзаказ[43]. С точностью до 90 процентов покупатели получают заказ едва ли не раньше, чем сделали его. Предварительный заказ делается автоматически без участия человека и экономит компании миллионы евро в год, так как избавляет ее от излишне больших закупок и возвратов. К тому же такой уровень обслуживания нравится покупателям. Глубокое обучение не только не оставило сотрудников компании без работы, но, наоборот, усилило их работоспособность. И действительно, ИИ может сделать вас эффективнее.

Хотя крупнейшие высокотехнологичные компании первыми внедрили приложения для глубокого обучения, инструменты машинного обучения уже широко доступны, и многие другие компании начинают получать от них выгоду. Алекса, голосовой помощник в устройстве Amazon Echo, отвечает на устные запросы благодаря глубокому обучению. Платформа Amazon Web Service (AWS) представила панель инструментов Lex and Polly, которая способствует разработке «естественного» языка на основе автоматического распознавания речи для определения намерений говорящего и преобразования письменного текста в устную речь. Приложения с диалоговым взаимодействием сейчас можно встретить только на малых предприятиях, которые не могут позволить себе нанять экспертов по машинному обучению. Искусственный интеллект помогает удовлетворять покупателей.

Когда компьютер обыграл в шахматы лучших игроков, разве люди перестали в них играть? Наоборот, это только повысило их уровень! Также это популяризировало шахматы. Когда-то лучшие игроки были жителями больших городов, таких как Москва, где были шахматные

клубы и много гроссмейстеров, обучающих молодое поколение. Шахматные программы дали возможность Магнусу Карлсену, выросшему в маленьком городке в Норвегии, стать гроссмейстером всего в 13 лет, и сейчас он чемпион мира. Этот процесс не ограничится играми, он повлияет на все аспекты нашей жизни, от искусства до науки. ИИ может сделать нас умнее.

Назад в будущее

Различные формы обучения позволяют работать всем вышеупомянутым приложениям. Кроме того, глубокое обучение – основа и для человеческого интеллекта. Эта книга посвящена двум взаимосвязанным темам – эволюции человеческого мозга и эволюции ИИ. Самое заметное различие: природа потратила миллионы лет на развитие человеческого интеллекта, в то время как ИИ на это понадобилось всего несколько десятилетий – слишком короткий срок даже для культурной эволюции.

Последние достижения глубокого обучения были сделаны не в одночасье, как может показаться по сообщениям в СМИ. История перехода ИИ, основывавшегося на символах, логике и системе правил, к глубокому обучению малоизвестна. Эта книга о появлении и развитии глубокого обучения с моей точки зрения как того, кто стоял у истоков разработки алгоритмов обучения нейронных сетей в 1980-х годах и в качестве президента Фонда Neural Information Processing Systems[44] (NIPS) курировал открытия в области машинного и глубокого обучения в течение последних 30 лет. Долгие годы нас преследовали неудачи, но в конце концов наши настойчивость и терпение были вознаграждены.

Глава 2. Перерождение искусственного интеллекта

Марвин Минский – блестящий математик и основатель Лаборатории искусственного интеллекта в МТИ в США. Основатели задают направление всей отрасли, и в 1960-х годах эта лаборатория стала цитаделью разума. У Минского за минуты рождалось огромное количество идей, и он мог убедить любого, что его мнение является верным, даже если здравый смысл говорил об обратном. Я восхищался его умом и смелостью, но был не согласен с его взглядами на ИИ.

Детская игра?

Blocks World – хороший пример проекта, созданного Лабораторией искусственного интеллекта МТИ в 1960-х годах. Если объяснять просто, Blocks World состоял из прямоугольных строительных блоков, которые можно было сложить в различных сочетаниях (рис. 2.1). Основной целью было написать программу, которая умела бы обрабатывать запросы вроде: «Найди большой желтый блок и положи его на красный блок», – а также продумывать шаги, необходимые для выполнения задания роботизированной рукой. Это похоже на детскую игру, однако требовалось написать сложную программу, причем настолько громоздкую, что было очень тяжело устранять неполадки. Программа была заброшена, когда студент Терри Виноград, написавший программу, покинул МТИ. Простая на первый взгляд программа оказалась головоломной. Но даже если бы ее удалось реализовать, все равно она не нашла бы применения вне лаборатории, ведь в реальном мире у объектов разные форма, размер и вес, а освещение может сильно отличаться в зависимости от места и времени, что сильно затрудняет распознавание.

В 1960-х годах Лаборатория ИИ получила крупный грант от Министерства обороны США на создание робота, играющего в пинг-понг. Я однажды услышал историю о том, что ученый, руководивший проектом, якобы забыл попросить деньги, необходимые на создание для робота зрительной системы, и потому поручил это дело аспиранту в качестве летнего проекта. При случае я спросил у Марвина Минского, правда ли это? Он резко ответил, что я ошибаюсь: «Мы поручили задачу студенту-бакалавру». Документ из архива МТИ подтверждает его слова (рис. 2.2)[45].

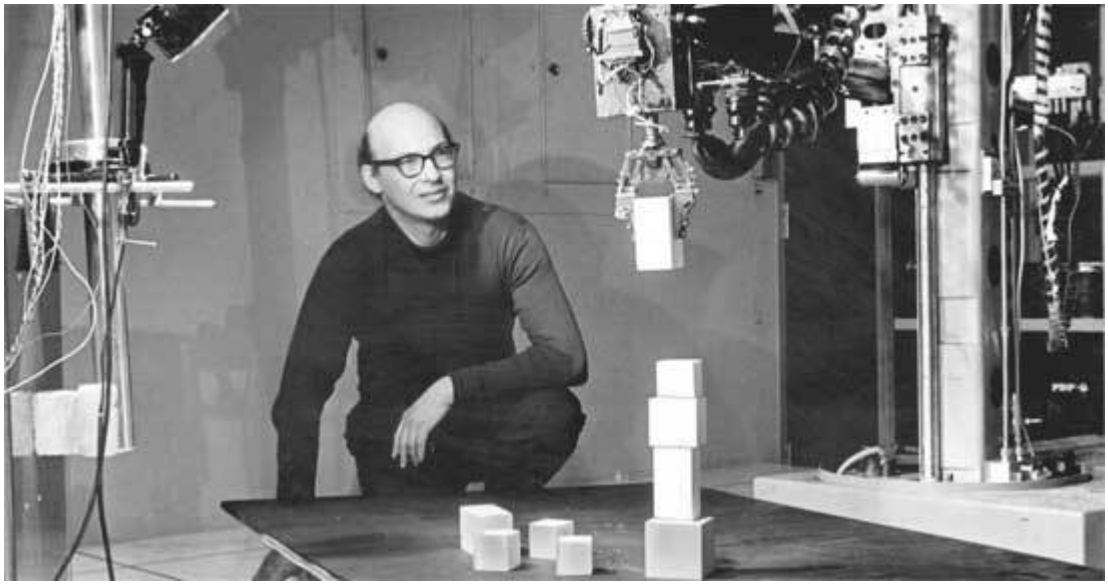


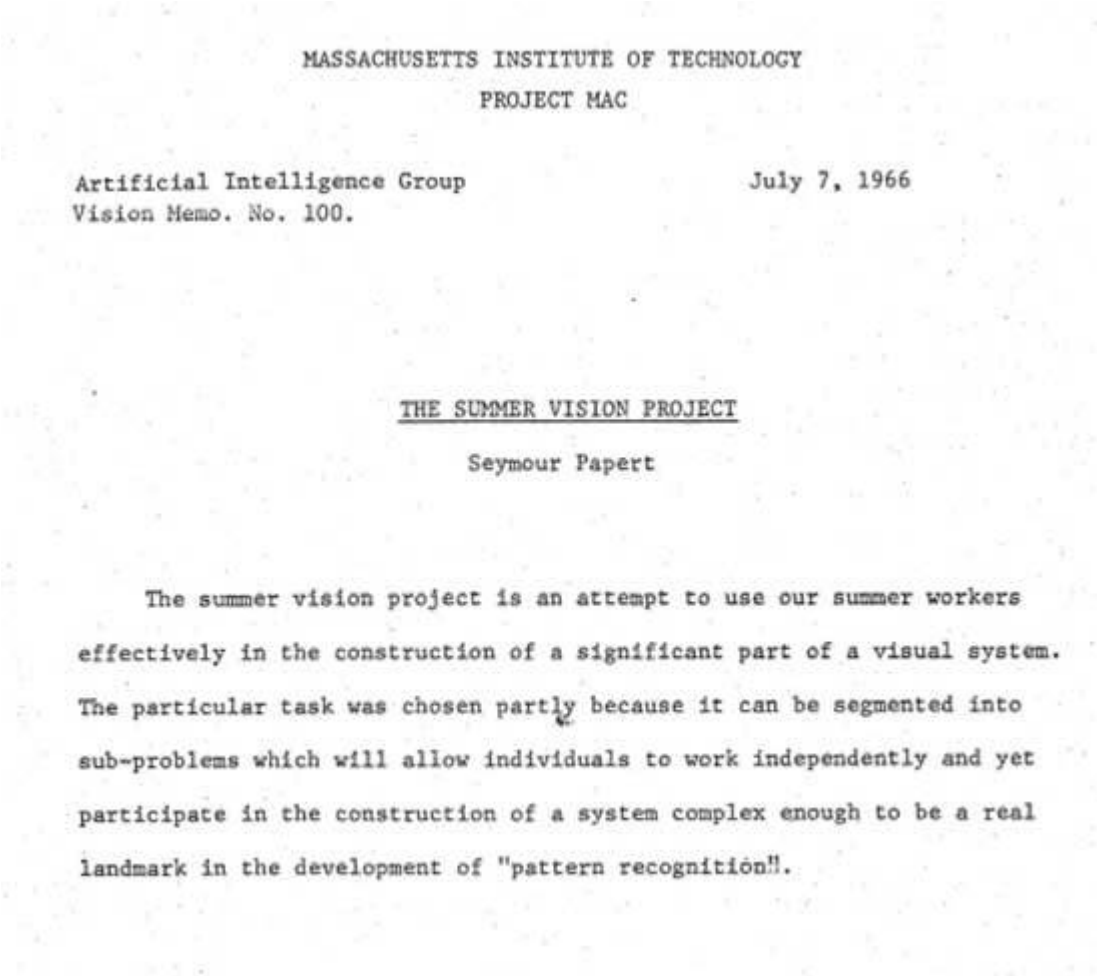
Рис. 2.1. Марвин Минский наблюдает за роботом, укладывающим блоки. 1968 год. Blocks World был упрощенной моделью того, как мы взаимодействуем с окружающим миром. Но все оказалось гораздо сложнее, чем кто-либо предполагал, и проблема не была решена, пока это не сделало глубокое обучение в 2016 году

То, что казалось простым на первый взгляд, стало своего рода зыбучим песком для целого поколения ученых, работающих над созданием компьютерного зрения.

Почему компьютерное зрение – трудная задача?

Мы, как правило, без проблем определяем, что за объект перед нами, независимо от его расположения, размера, ориентации в пространстве и освещенности. Одна из первых идей, касающихся компьютерного зрения, предлагала сопоставлять шаблон предмета с его пиксельным изображением. Но это не сработало, потому что если поменять положение одного и того же объекта, то его изображения не совпадут. Пример: фотография двух птиц на рис. 2.3. Если вы наложите изображение одной птицы на изображение другой, то какая-то его часть совпадет, но остальная – нет. В то же время оно может быть удачно совмещено с изображением птицы,

относящейся к другому виду, но находящейся в такой же позе.



MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Рис. 2.2. Первая страница летнего проекта по созданию машинного зрения в МТИ. dspace.mit.edu/handle/1721.1/6125

Рис. 2.3. Две зебровые амадины, изображения которых надо совместить. Мы без труда определим, что это птицы, относящиеся к одному и тому же виду. Но из-за разного положения на снимке их сложно сравнивать с помощью шаблона, хотя у них схожие идентификационные признаки

Ученые добились прогресса, когда сосредоточились не на шаблонах, а на схожих чертах. Например, орнитологи должны профессионально определять разные виды птиц, так как некоторые могут отличаться лишь отдельными неявными чертами. В популярной прикладной книге, помогающей идентифицировать птиц, дается всего одна фотография каждой птицы и множество схематичных рисунков, на которых обозначены ключевые различия (рис. 2.4)[46]. Хорошая особенность – та, которая присутствует только у одного вида птиц, но практика показывает, что схожие признаки можно обнаружить у нескольких видов. Таким образом, единственный способ идентифицировать птицу – определить уникальный набор различных признаков: цвет оперения, полосы над глазами, вкрапления на крыльях. Когда не получается распознать птиц по этим чертам, ученые обращаются к их пению. Рисунки отличительных особенностей птиц гораздо информативнее, так как

фотографии переполнены лишней информацией.

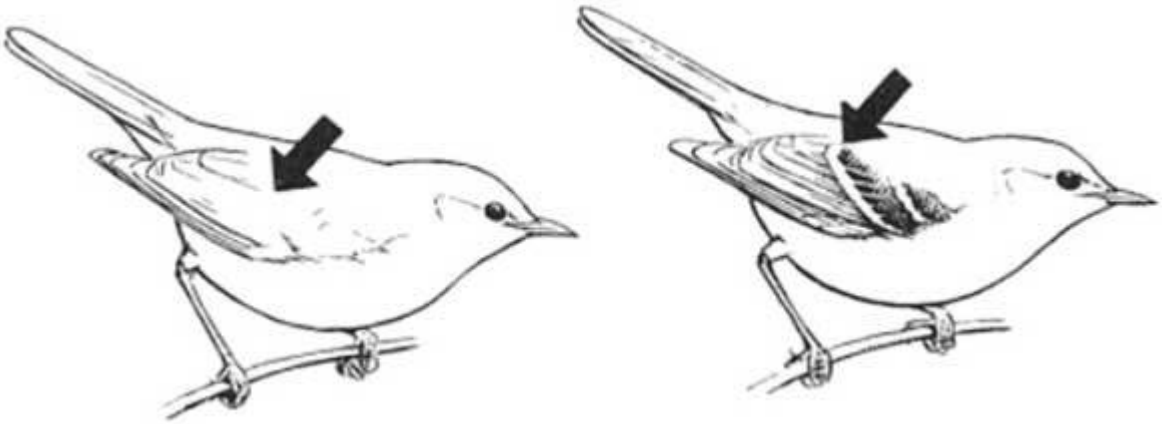


Рис. 2.4. Изображение отличительного признака, по которому можно определить вид птицы среди схожих. Стрелки указывают на участки оперения, которые особенно важны для того, чтобы распознать вид птицы семейства соловьиных: некоторые из них бросаются в глаза, другие нечеткие; одни длинные, другие короткие. Из книги Роджера Петерсона, Гайя Маунтфорта и Филипа Холлома «Справочник птиц Британии и Европы»

Проблема такого подхода в том, что очень непросто разработать анализаторы признаков для сотен тысяч объектов, и даже с большим набором признаков программе будет трудно различить объекты на изображении, если те частично закрыты, и понять, где заканчивается один объект и начинается другой.

Едва ли в 1960-х кто-то мог предположить, что потребуется 50 лет и в миллион раз бóльшая мощность компьютера, прежде чем компьютерное зрение достигнет уровня человеческого. Предположение, что создать машинное зрение будет просто, основывается на том, что мы сами без труда видим, слышим и передвигаемся. Мы профессионалы во всем вышеперечисленном, потому что указанные навыки помогают нам выжить, а эволюции понадобились миллионы лет, чтобы усовершенствовать их. Это и

сбило с толку первых исследователей в области ИИ. Обратная ситуация с доказательством теорем: человеку нужно обладать высоким интеллектом, чтобы сделать это, в то время как для компьютера приведение доказательства не составит никакого труда, потому что логика у него развита гораздо лучше, чем у нас. Способность мыслить логически – результат поздней эволюции, и даже людям нужна тренировка, чтобы выстроить длинную логическую цепочку и по ней прийти к однозначному выводу. Для большинства проблем, которые нужно решить, чтобы выжить, необходимы выводы из предыдущего опыта и их обобщение.

Экспертная система

Экспертные системы, основывающиеся на определенных правилах, были популярны в 1970–1980-х годах. Их цель – решение таких проблем, как постановка медицинского диагноза, с помощью набора правил. Одна из первых экспертная систем MYCIN[47], например, была специально разработана для анализа на бактерии, вызывающие различные инфекции, в том числе менингит. Первый шаг – сбор правил и фактов, которыми руководствуются врачи-инфекционисты. Далее были добавлены истории болезни и диагнозы пациентов, и на их основе сделаны соответствующие логические выводы. Слабым местом такого подхода был сбор экспертной информации, особенно если речь шла о сложных проблемах. Лучшие диагносты не используют правила, они полагаются на свой опыт, а его трудно кодифицировать[48]. Более того, экспертная система должна постоянно обновляться: нужно вносить в базу данных новые факты и убирать оттуда устаревшие. На практике MYCIN никогда не использовалась врачами, потому что все вопросы, которые система задает пациенту, требовалось вносить в компьютер, а занятой врач не может тратить на это по полчаса каждый раз. Однако многие экспертные системы были написаны для других целей, например для управления разливом токсичных веществ, планирования маршрута для беспилотных транспортных средств и распознавания речи. Некоторые из них используются до сих пор.

В первые десятилетия существования ИИ были изучены многие направления, но дальнейшая их разработка оказалась более трудоемкой, нежели действительно полезной. Недооценивали не только сложность проблем реального мира, но и масштаб возможных решений. В комплексных областях, когда число различных правил может быть огромным, а новые факты и поправки добавляются вручную, отслеживание исключений и взаимодействия

с другими правилами становится нецелесообразным. Например, в 1984 году Дуглас Ленат запустил проект CYC с целью систематизировать здравый смысл. Поначалу идея казалась хорошей, но впоследствии она обернулась катастрофой[49]. Мы воспринимаем как данность огромное количество фактов об окружающем нас мире. Множество из них основываются на опыте. Например, кот, упавший с высоты в 7,5 метра, скорее всего, избежит травм[50], в то время как человек – нет.

Еще одна причина, почему ИИ развивался медленно, заключалась в том, что цифровые компьютеры были примитивными, а накопители данных – непростительно дорогими по нынешним меркам. Тем не менее ЭВМ очень эффективны при выполнении логических операций, манипулировании символами и применении правил, поэтому неудивительно, что в XX веке они стали популярны. Например, Аллен Ньюэлл и Герберт Саймон, программисты из Университета Карнеги – Меллона, в 1955 году написали программу Logic Theorist, которая могла доказывать теоремы из сборника Бертрана Рассела «Начала математики» – одной из первых попыток систематизировать всю математику. На заре развития ИИ люди надеялись, что появление «умных» компьютеров уже не за горами.

Пионеры в области ИИ старались писать программы, обладающие возможностями человеческого интеллекта, однако не задумывались о том, как мозг приходит к разумному поведению. Однажды я спросил Аллена Ньюэлла, почему они игнорировали это. Он ответил, что хотел исследовать мозг, однако в то время о нем было известно слишком мало, чтобы знания удалось применить. Основные принципы работы мозга были открыты только в 1950-х годах в классической работе Алана Ходжкина и Эндрю Хаксли, в которой объясняется, как благодаря колебаниям нервов передаются сигналы в мозг. Также свой вклад в изучение функционирования мозга внес Бернард Кац, открывший, как электрические сигналы конвертируются в химические сигналы в синапсе, осуществляющем связь между нейронами.

К 1980-м годам мозг исследовали более подробно, а полученные знания выходили далеко за пределы биологии. Но к тому времени мозг как образец стал неактуален для следующего поколения разработчиков ИИ, а их целью было написать программу, которая функционировала бы схожим образом. Это было хорошим поводом игнорировать неясные детали в биологии. Тем не менее небольшая группа ученых, не подвергшихся влиянию новых взглядов на ИИ, верила, что путь к развитию ИИ лежит через познание биологических основ мозга, называемый нейронными сетями, с прямой связью и параллельной обработкой, и что именно он поможет решить проблемы, с которыми не справились ИИ на базе логических схем.

Я был одним из них.

В логове льва

В 1989 году глава компьютерной научной лаборатории МТИ Михалис Дертузос пригласил меня прочитать лекцию в МТИ (рис. 2.5). Я был одним из первых, кто изучал новый подход к развитию ИИ, основанному на нейронных сетях, и меня удостоили чести побывать в святая святых ИИ. Я прибыл в МТИ до полудня и был тепло встречен Дертузосом. Он написал книгу о будущем компьютерных технологий, что дало нам почву для беседы. Когда мы ехали в лифте, чтобы пообедать, он сказал мне, что на их факультете есть особая традиция: за обедом студенты разговаривают с лектором, и у меня будет пять минут, чтобы начать беседу. «И кстати, – добавил он, – они ненавидят то, что вы делаете».

Столовая была битком набита народом, что даже удивило Дертузоса. Ученые стояли в три ряда: в первом – старшие преподаватели, во втором – младшие, а за ними, в третьем ряду, студенты. Я, конечно, не считал, но там было человек сто. Я стоял в центре, перед буфетом, как главное блюдо. Что интересного я мог сказать за пять минут людям, которые ненавидят мою работу?

Тогда я решил импровизировать. «Мозг мухи состоит всего из ста тысяч нейронов; он весит миллиграмм и потребляет милливатт энергии, – сказал я, сочиняя свою речь буквально на ходу. – Муха может видеть, летать, ориентироваться в пространстве и находить еду. Но что более примечательно, у нее есть репродуктивная функция. В МТИ есть суперкомпьютер стоимостью в десять миллионов долларов, он потребляет мегаватт энергии и охлаждается огромным кондиционером. Но самое дорогое в нем – жертвы в лице программистов, жаждущих утолить свой ненасытный голод к составлению программ. Этот суперкомпьютер, хоть и умеет контактировать с другими компьютерами, не может видеть, летать, спариваться и размножаться. Почему же?»

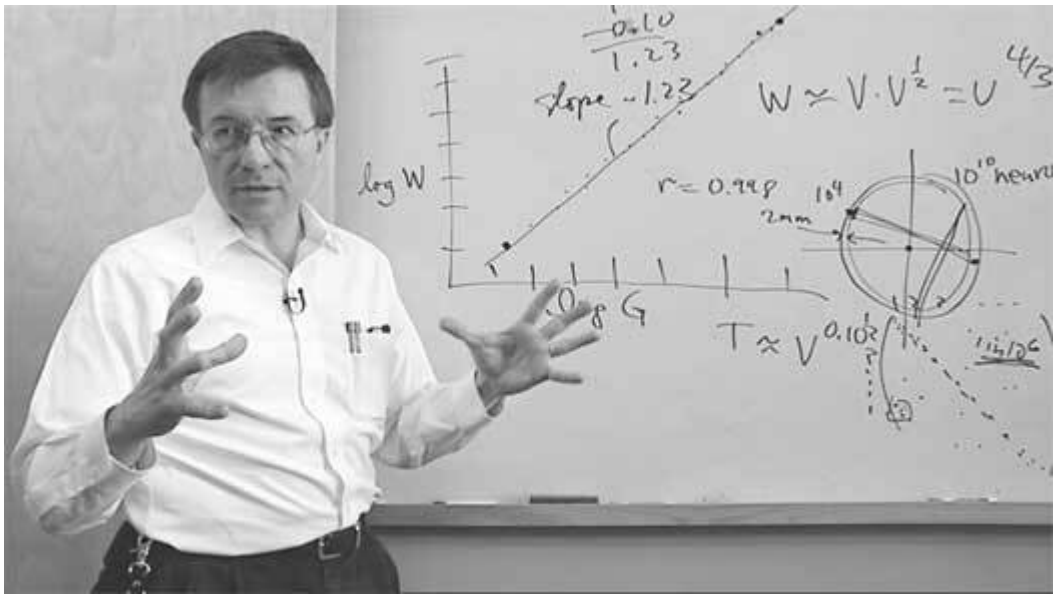


Рис. 2.5. Автор во время посещения МТИ в 1988 году. Монитор на заднем плане напоминает о статическом электричестве, которое заставляло мои волосы вставать дыбом

После долгой паузы один из старших преподавателей ответил: «Потому что мы еще не написали программу зрительного восприятия». Министерство обороны США недавно вложило 600 миллионов долларов в десятилетний проект «Стратегической

компьютерной инициативы»[51], который продвинулся лишь на шаг в вопросе компьютерного зрения, что позволило создать самозаправляющийся танк. «Удачи!» – таков был мой ответ им.

Присутствовавший там Джеральд Сассман, сделавший несколько важных открытий, которые приблизили ИИ к реальному миру (в числе его изобретений – система высокоточной интеграции для орбитальной механики), начал отстаивать подход МТИ к ИИ, ссылаясь на работу Алана Тьюринга, доказавшего, что изобретенная им машина может вычислить любую вычислимую функцию.

«И сколько времени это займет? Вам нужно работать быстрее, иначе вас съедят!» – сказав это, я пошел наливать себе кофе. Мой диалог с представителями факультета был закончен.

На этот вопрос может ответить каждый студент из моей лаборатории. После того как разошлись первые два ряда зрителей, студент из третьего ряда предложил свой вариант: «Цифровой компьютер – устройство общего назначения, который можно запрограммировать на вычисление всего, что угодно, однако не всегда удачно. Муха – компьютер специального назначения, который может видеть и летать, но не может проверить баланс моего счета». Это был верный ответ. Глаза мухи эволюционировали сотни миллионов лет, и зрительные алгоритмы встроены в эту систему. Именно поэтому мы можем воссоздать зрение мухи, разработав схему подключения к потоку информации и продвижения по нему через нейронные сети, и не можем сделать это для цифрового компьютера, который требует программного обеспечения, указывающего, какая задача сейчас решается.

Я узнал младшего преподавателя, улыбавшегося в задних рядах. Однажды я пригласил его на семинар по компьютерной нейробиологии в Вудсхоулской лаборатория биологии моря на полуострове Кейп-Код[52]. Родни Брукс родом из Австралии, в 1980-х годах он работал в Лаборатории искусственного интеллекта в МТИ и создавал шагающих роботов-насекомых, используя код, не зависящий от цифровой логики. В конечном итоге он стал главой

этой лаборатории и основал компанию iRobot, производящую роботы-пылесосы Roomba[53].

Большой зал, где я читал свою лекцию, был заполнен студентами старших курсов, устремленными в будущее, а не обращающимися к прошлому. Я говорил о нейронной сети, научившейся играть в нарды, – проекте Джеральда Тезауро, физика из Центра исследования сложных систем[54] в Иллинойсском университете в Урбане-Шампейне. В нардах два игрока стремятся привести к финишу свои фишки, по очереди бросая кости. Эта игра очень популярна на Ближнем Востоке, и некоторые даже зарабатывают ею себе на жизнь.

Это еще один подход к созданию ИИ. Написать программу, основанную на логике и эвристических алгоритмах, для обработки всех вероятных позиций на доске – невыполнимая задача, учитывая, что есть 10^{20} возможных положений. Вместо этого сеть научилась играть с помощью распознавания образов, наблюдая за игрой учителя[55]. Джеральд Тезауро создал первую программу для игры в нарды, освоившую ее на уровне мировых чемпионатов за счет особой нейросети. Об этом будет рассказано в главе 10.

После лекции я узнал, что утром на первой полосе New York Times была опубликована статья о сокращении государственного финансирования исследования ИИ. Это было началом «зимы ИИ», но коснулось меня, так как начался расцвет нейронных сетей.

Оглядываясь назад, я удивляюсь, как смог пережить тот вечер. У нас была новая траектория развития ИИ, но понадобилось двадцать пять лет, чтобы создать работающие приложения для компьютерного зрения, речи и языка. Я должен был предположить, что это займет столько времени. В 1978 году, будучи аспирантом в Принстонском университете, я применил закон Мура, который гласит, что компьютерные мощности растут по экспоненте и удваиваются каждые 18 месяцев, чтобы посмотреть, сколько времени займет достижение компьютерами мощности человеческого мозга. Я

пришел к выводу, что это случится в 2015 году. К счастью, это не остановило меня, и я продолжил двигаться вперед. Моя вера в нейронные сети была основана на интуитивной уверенности, что раз природа справилась, то мы должны перенять у нее способ решения данной проблемы. Те тридцать пять лет, которые я ждал, – всего лишь миг по сравнению с сотнями миллионов лет, потребовавшихся природе.

Внутри зрительной коры мозга нейроны расположены иерархичными слоями. По мере того как информация трансформируется слой за слоем, представление о мире становится все более абстрактным. Десятилетия, по мере увеличения числа слоев в искусственных нейронных сетях, их производительность продолжала расти, пока наконец не был достигнут критический порог, который позволил решить задачи, казавшиеся невозможными в 1980-х годах. Глубокое обучение автоматизировало поиск отличительных черт, позволяющих опознавать объекты на изображении. Вот почему компьютерное зрение сейчас гораздо лучше, чем пять лет назад.

К 2016 году компьютеры стали в миллион раз быстрее, а компьютерная память увеличилась в миллиарды раз, исчисляясь уже не мегабайтами, а терабайтами. Стало возможным создать нейронную сеть с миллионами компонентов и миллиардами связей. Для сравнения: в нейронных сетях 1980-х годов было всего несколько сотен компонентов и несколько тысяч связей. Современные нейронные сети все еще крошечные по сравнению с человеческим мозгом, в котором сто миллиардов нейронов и квадрильон синаптических связей. Тем не менее современные нейронные сети достаточно велики, чтобы продемонстрировать доказательства принципа в узких областях.

Глубокое обучение стало применяться в глубоких нейронных сетях. Но прежде чем начать работать с глубокими сетями, нам нужно было натренироваться на мелких.

Глава 3. Спад нейронных сетей

Единственным доказательством того, что даже самые сложные проблемы ИИ могут быть решены, является тот факт, что природа уже справилась с этими трудностями. В 1950-х годах появились подсказки, ключи для разгадки, которые предполагали принципиально новый подход к обработке символов, что могло обеспечить интеллектуальное поведение компьютера.

Первая подсказка: мозг – мощный распознаватель образов. Ваша зрительная система может распознать объект на изображении всего за десятую долю секунды, даже если вы никогда ранее его не видели. Кроме того, объект может быть любой формы, находиться на произвольном расстоянии и в любом положении по отношению к вам. Это все равно, что иметь особый компьютер, единственная функция которого – распознавание предметов.

Вторая подсказка – с помощью практики можно научить мозг выполнять задания любой сложности, будь то игра в теннис или задачи по физике. Природа использует обучение общего назначения для решения различных проблем, а человек, в свою очередь, прекрасный ученик. Это наша суперспособность. Структура коры головного мозга у всех схожа, а глубокие нейронные сети есть во всех сенсорных и моторных системах[56].

Третья подсказка – наш мозг изначально не наполнен правилами или логикой, но мы можем начать мыслить логически и следовать правилам после длительного обучения, хотя тут преуспеет далеко не каждый. Это наглядно проиллюстрировано логической головоломкой – задачей выбора Уэйсона (рис. 3.1).

Правильный ответ: карту с номером 8 и карту с коричневой рубашкой. Исследования показали, что только 10 процентов людей отвечают правильно[57]. Тем не менее у большинства опрашиваемых нет проблем с правильным ответом, если ситуация в вопросе

знакомая (рис. 3.2).

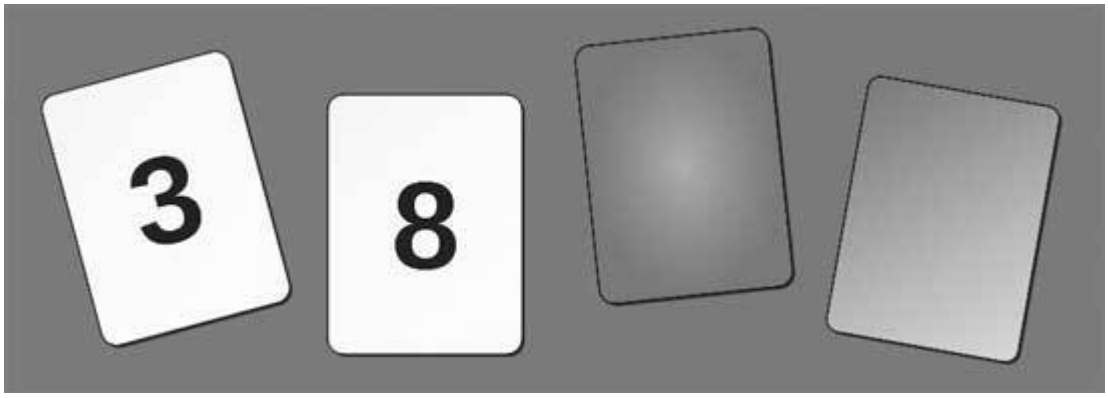


Рис. 3.1. На каждой из четырех карт с одной стороны цифра, с другой – цветная рубашка. Какую(ие) карту(ы) вы должны перевернуть, чтобы проверить истинность утверждения, что если на карте четное число, то ее противоположная сторона красная?

Рис. 3.2. На каждой карте указан возраст с одной стороны и изображен напиток с другой. Какую(ие) карту(ы) нужно перевернуть, чтобы проверить закон, по которому вы должны быть старше 18 лет, чтобы пить алкоголь?

Рассуждения кажутся зависимыми от области, о которой идет речь, и чем ближе вам область, тем легче вам решать проблемы в ней. Опыт упрощает рассуждения, потому что вы можете использовать примеры, с которыми столкнулись при интуитивном решении. В физике, например, вы изучаете определенную область (скажем, электричество и магнетизм), и именно это помогает вам

при решении многих задач, а не запоминание формул. Если бы человеческий интеллект основывался только на логике, то область знаний должна была бы быть единой, а это не так.

Четвертая подсказка – мозг состоит из миллиардов крошечных нейронов, контактирующих друг с другом. Это говорит о том, что мы должны изучать класс массово-параллельных архитектур[58] для решения проблем ИИ, а не архитектуру цифровых компьютеров фон Неймана, в которой процессор отделен от памяти узким каналом, через который данные и инструкции извлекаются и выполняются по одному. Действительно, машина Тьюринга может посчитать любую вычислимую функцию, имея достаточно памяти и времени, но она медленная и ее трудно программировать, а природа должна была решать проблемы в режиме реального времени. У самых мощных компьютеров на планете – массово-параллельные процессоры. Алгоритм, эффективно работающий на них, в конечном счете победит.

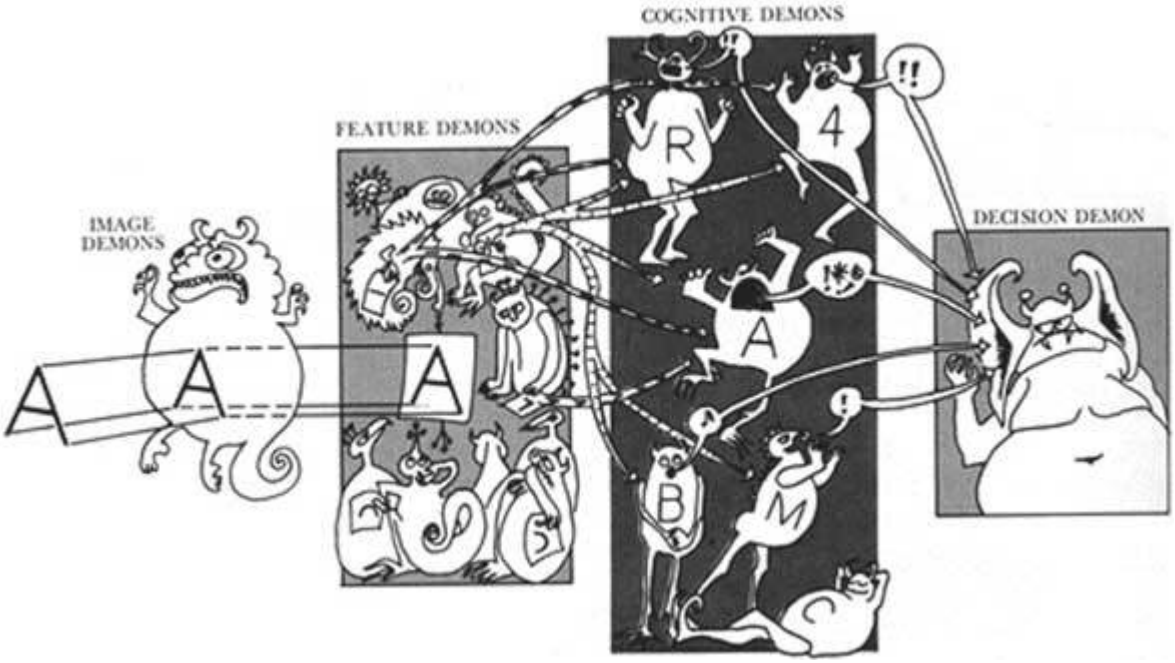


Рис. 3.3. Пандемониум. Оливер Селфридж представил, что в мозге есть демоны, которые ответственны за последовательное

извлечение более сложных признаков и абстракций из сенсорных органов восприятия, что и приводит к принятию решений. Каждый демон на каждом уровне оживляется, если он соответствует входу с более раннего уровня. Решение демона взвешивает степень оживления и важность его информаторов. Эта форма оценки информации – метафора для современных сетей глубокого обучения, у которых гораздо больше уровней[59]

Первооткрыватели

В 1950–1960-х годах произошел взрыв интереса к самоорганизующимся системам. Норберт Винер создал кибернетику на основе систем связи и управления как машин, так и живых существ[60]. Оливер Селфридж разработал «Пандемониум»[61] – систему распознавания образов, в котором выполняющие функцию обнаружения «демоны»[62] выступали за право представлять объекты на изображениях, что является метафорой для глубокого обучения (рис. 3.3). Бернард Уидроу из Стэнфорда и его студент Тед Хофф создали алгоритм обучения LMS (Least Mean Squares; алгоритм минимальной среднеквадратичной ошибки)[63], который широко используется для адаптивной обработки сигналов при регулировке шумов вдоль линий передачи, например телефонного кабеля. У алгоритма LMS и его последующих версий множество функций, начиная от шумоподавления и заканчивая финансовыми прогнозами. Это лишь несколько примеров, иллюстрирующих расцвет гениальных идей в 1960-х годах. Здесь я заострю свое внимание всего на одном первопроходце, Фрэнке Розенблатте (рис. 3.4), разработавшем перцептрон – прямой предшественник глубокого обучения.

Обучение на примерах

Первопроходцев нейронных сетей не отпугнуло, что мы не понимали функции мозга, и они сосредоточились на схематичных версиях нейронов и том, как они связаны друг с другом. Фрэнк Розенблатт из Корнелльского университета в США (рис. 3.4) был одним из первых, кто сымитировал строение нашей зрительной системы для автоматического распознавания образов. Он изобрел обманчиво простую систему под названием перцептрон, которая могла научиться классифицировать образцы по категориям, например по буквам алфавита. Розенблатт был застенчивым холостяком, но любил погонять на спортивной машине вокруг университетского кампуса. Он был эрудитом с широким кругом интересов, в том числе его интересовал поиск планет у далеких звезд через измерение постепенного падения яркости звезды, когда планета проходит мимо нее. Этот метод в настоящее время часто используется для обнаружения планет, типичных для нашей галактики.

Если вы понимаете основные принципы того, как перцептрон учится решать проблему распознавания образов, вы на полпути к пониманию работы глубокого обучения. Цель перцептрона – определить, является ли входной образ элементом категории на изображении. В Блоке 1 объясняется, как входные данные перцептрона преобразуются набором веса из входных единиц в выходные. Вес – это мера влияния каждого входа на окончательное решение, принятое блоком вывода. Как мы можем определить оптимальный набор весов для правильной классификации получаемой информации?

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) —The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

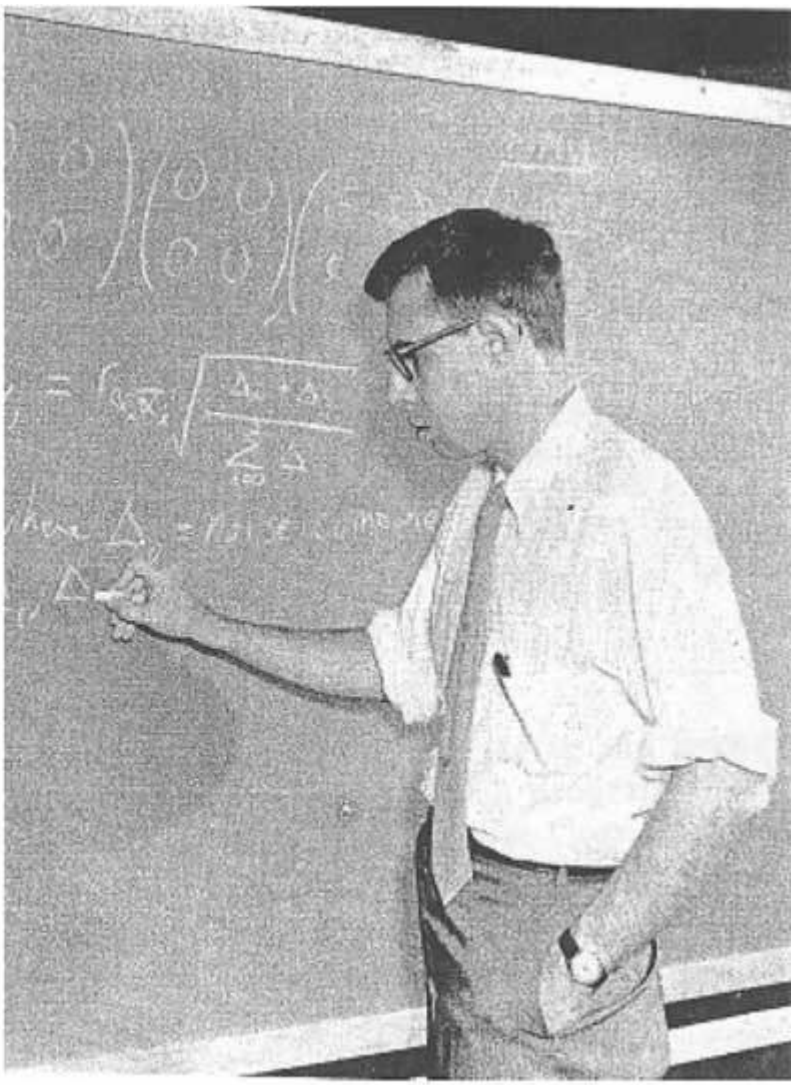


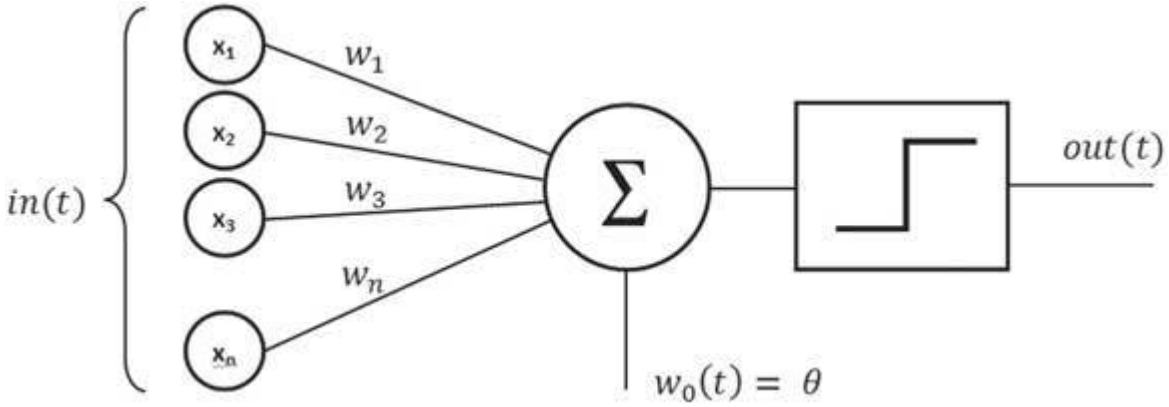
Рис. 3.4. Фрэнк Розенблатт в Корнелльском университете, погруженный в свои мысли. Он изобрел перцептрон – ранний предшественник сетей глубокого обучения, в основе которого лежал простой обучающий алгоритм для классификации изображений по категориям, например определяя, левая это сторона или правая. Заметка была опубликована в New York Times 8 июля 1958 года по сообщению агентства United Press International. Сто тысяч долларов в 1958 году в наши дни равноценны одному миллиону долларов. 704 компьютера IBM, стоившие два миллиона долларов, сегодня стоили бы двадцать миллионов долларов. 704 компьютера IBM

могли выполнить двенадцать тысяч умножений в секунду, что считалось молниеносным по меркам того времени. Но смартфон Samsung S6 может совершить 34 миллиарда умножений в секунду[64], а это более чем в миллион раз быстрее и гораздо дешевле

Традиционный способ, который используют инженеры для решения этой задачи, – создание веса вручную на основе анализа или ситуативно для конкретной цели. Он трудоемок и часто базируется не только на инженерных разработках, но и на интуиции. В качестве альтернативы применяется автоматическая процедура, которая учится на примерах так же, как мы познаем окружающий мир. Необходимо множество примеров, включая те, что относятся к другим областям, особенно сходным: чтобы научиться распознавать кошек, нужно увидеть и собак. Примеры по одному вносятся в перцептрон, и при ошибке вес автоматически корректируется. Это называется обучающим алгоритмом. Алгоритм – пошаговая инструкция, которой вы следуете для достижения цели, например рецепт приготовления пирога. В главе 13 мы рассмотрим алгоритмы в целом.

Прелесть обучающей системы перцептрона в том, что он гарантированно сам найдет набор весов, если таковой существует и есть достаточно примеров. Обучение проходит постепенно, после того как представлен каждый из предметов в обучающем наборе, и результат сравнивается с правильным ответом. Если ответ верный, в вес не вносится никаких изменений. Но если ответ неправильный (1, когда должно быть 0, или 0, когда должно быть 1), то вес постепенно меняется, и в следующий раз, когда поступит такой же запрос, он будет ближе к правильному ответу (блок 1). Важно, чтобы изменения происходили постепенно, для того чтобы вес зависел от всех тренировочных примеров, а не только от последнего.

Блок 1. Перцептрон



Перцептрон – это нейронная сеть с одним нейроном, которая имеет входной слой и набор соединений, связывающих входные блоки с выходным блоком. Цель перцептрона – классифицировать образцы, поступающие в блок входа. Основная функция, выполняемая блоком вывода, – суммирование значений каждого входного сигнала, помноженного на вес его связи с блоком вывода. На диаграмме вес (w_n) суммы входных сигналов (x_n) сравнивается с порогом θ и проходит через ступенчатую функцию, которая дает на выходе единицу, если сумма больше порогового значения, и ноль – если меньше. Например, входными данными могут быть пиксели изображения или, в более широком смысле, основная информация, извлеченная из необработанного изображения, такая как контур объекта. Изображения представляются по одному, и перцептрон решает, входило ли оно в категорию, например, кошек. Блок вывода может быть только в одном из двух состояний: «включен», если изображение относится к данной категории, и «выключен», если не относится. «Включен» и «выключен» соответствуют 1 и 0 в двоичной системе. Обучающий алгоритм перцептрона выглядит следующим образом:

$\delta w_i = \alpha \delta x_i$

δ = *вывод – учитель,*

где вывод и учитель являются двоичными, так что δ равно нулю, если вывод правильный. Если выход неправильный, δ равно +1 или –1 в зависимости от разницы.

Если объяснение работы перцептрона не ясно, есть более четкий геометрический способ, помогающий понять, как перцептрон учится распознавать входящую информацию. Для частного случая двух типов входных данных можно нанести входные данные на двумерный график. Каждый вход представляет собой точку на графике, а два веса в сети определяют прямую линию. Цель обучения – провести линию таким образом, чтобы она четко разделяла положительные и отрицательные примеры (рис. 3.5). Для трех типов входных данных пространство входа трехмерное, и перцептрон задает плоскость, разделяющую положительные и отрицательные обучающие примеры. В общем случае размерность пространства входов может быть довольно высокой и ее будет невозможно визуализировать, но принцип остается тем же.

В конце концов, если появится решение, вес перестанет меняться, и значит, все примеры в обучающем наборе классифицированы правильно. Здесь нужно соблюдать осторожность, потому как в обучающем наборе, возможно, было недостаточно примеров, и сеть просто запомнила конкретные образцы, не имея шанса обобщить их в новой для нее ситуации. Это называется чрезмерным обучением, или переобучением. Важно иметь другой, контрольный набор примеров, который не был использован для обучения сети. В конце обучения результат классификации тестового набора является истинным показателем того, насколько хорошо перцептрон может обобщить новый пример, категория которого неизвестна. Обобщение здесь ключевое понятие. В реальной жизни мы никогда не видим тот же объект одинаково и не сталкиваемся с той же ситуацией, но если мы сможем обобщить предыдущий опыт и

спроецировать его на новую ситуацию, нам удастся справиться с широким спектром реальных проблем.

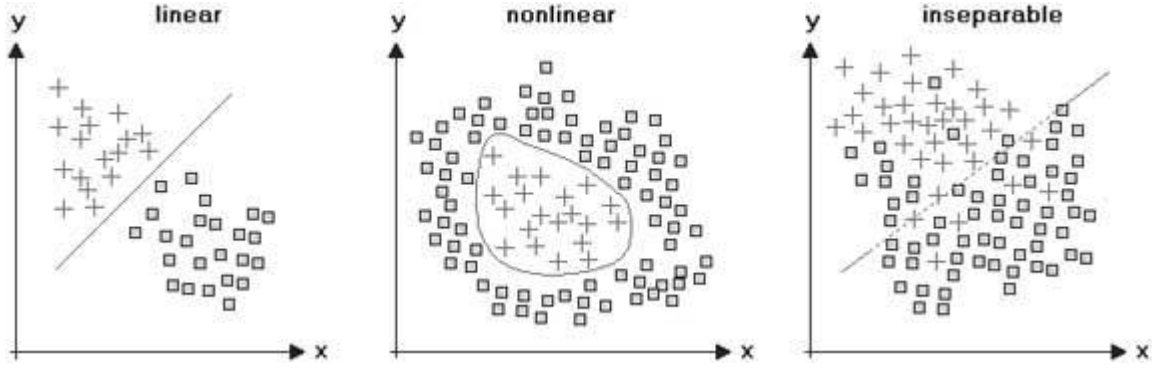


Рис. 3.5. Геометрическое объяснение того, как перцептрон распознает две категории объектов. У объектов есть две характеристики – длина и яркость, – их значения (x, y) отображены на графике. На графике слева оба типа объектов (плюсы и квадраты) возможно разделить прямой линией, которая пройдет между ними. Это различие может быть изучено перцептроном. В двух других областях объекты нельзя разделить прямой линией, но на центральном графике их можно разделить кривой. С выборкой справа надо провести некие махинации, чтобы разделить объекты двух типов. Все три класса могут быть изучены глубокой сетью, если есть достаточно данных для обучения

SEXNET

В качестве примера того, как перцептрон можно использовать для решения реальной задачи, попробуем отличить мужское лицо от женского, если убрать волосы, ювелирные изделия и вторичные половые признаки, такие как кадык, который у мужчин обычно крупнее. Беатрис Голомб, научный сотрудник моей лаборатории, в 1990 году получила базу данных с фотографиями студентов колледжа и использовала их как входные данные для перцептрона, который был обучен определять пол по лицу с точностью 81 процент[65]. Лица, при распознавании которых перцептрон испытывал трудности, были трудны и для людей. Работники моей лаборатории справились с тем же заданием с результатом 88 процентов. Беатрис также обучила многослойный перцептрон, который достиг точности 92 процента[66], что лучше результата многих людей (речь о нем пойдет во второй части книги). Это позволило ей в 1991 году на Конференции NIPS объявить: «Поскольку опыт улучшает производительность, значит, сотрудники лаборатории должны тратить больше времени на определение пола». Она назвала нейросеть SEXNET. Во время, отведенное для вопросов, кто-то спросил, может ли SEXNET определить лицо трансгендера. «Да», – ответила Беатрис, на что Эд Познер, учредитель конференции, сказал: «Это будет «DRAGNET»[67].

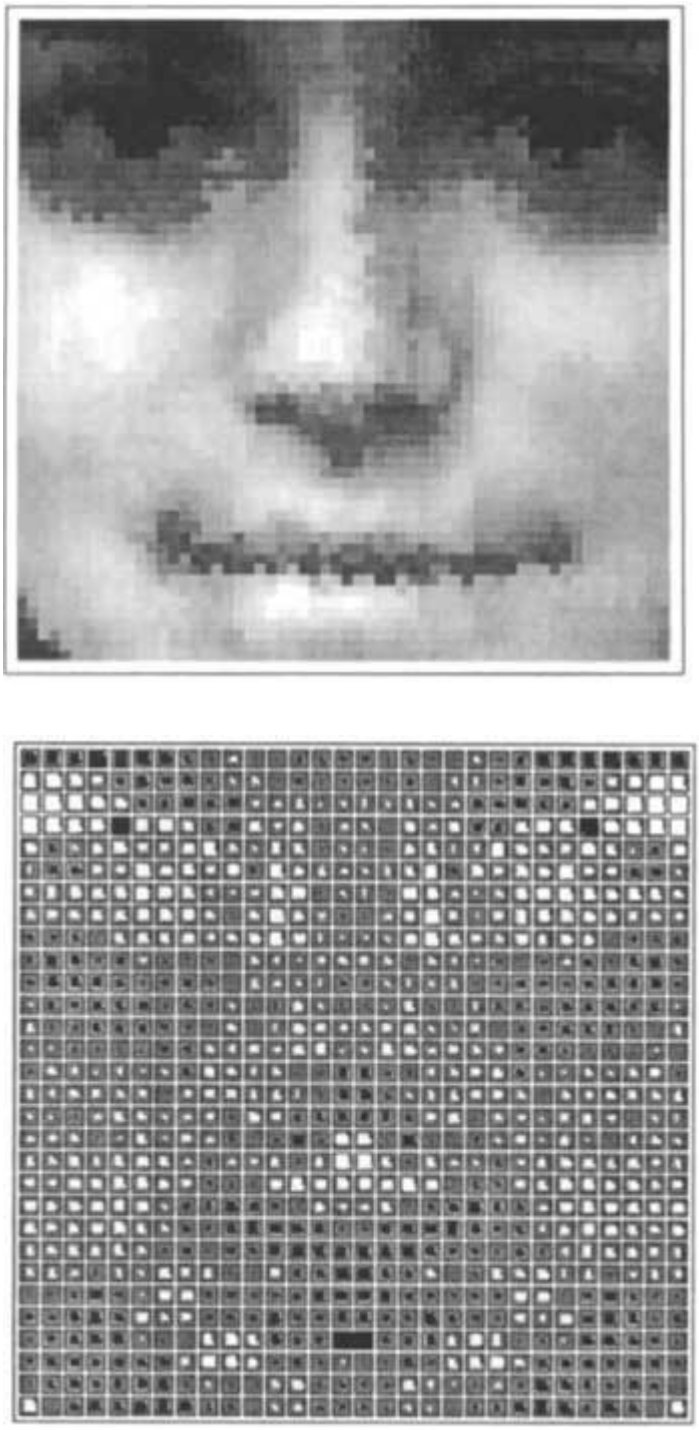


Рис. 3.6. Человеку какого пола принадлежит лицо на изображении? Перцептрон был обучен распознавать женские и мужские лица. Пиксели на изображении лица (слева) умножаются на соответствующий вес (справа), и полученная сумма сравнивается с порогом. Размер каждого веса отображается как площадь пикселя.

Положительный вес (белый) является признаком мужских лиц, а отрицательный вес (черный) – женских. Ширина носа, размер области между носом и ртом, а также интенсивность изображения вокруг области глаз важны для определение лица как мужского, в то время как интенсивность изображения вокруг рта и скул – для распознавания женских

Интересной задачу делает то, что, хоть мы и хорошо умеем отличать мужские лица от женских, мы не можем перечислить конкретные черты. Это проблема распознавания образов, которая зависит от объединения данных из большого количества низкоуровневых признаков, поскольку ни один из них не является окончательным. Преимущество перцептрона в том, что вес дает подсказки, какие части лица наиболее информативны для определения пола (рис. 3.6). Примечательно, что губной желобок (вертикальное углубление между носом и верхней губой) – одна из самых характерных черт, он намного крупнее у мужчин. Область вокруг глаз (больше у мужчин) и щеки (больше у женщин) также достаточно информативны. Перцептрон извлекает информацию обо всех отличительных признаках, чтобы принять решение. Примерно то же самое делает и человек, хоть он вряд ли сможет объяснить ход своих рассуждений.

Розенблатт доказал теорему сходимости перцептрона в 1957 году. Это стало огромным шагом вперед, а демонстрация работы системы впечатляла. При поддержке Управления военно-морских исследований Министерства обороны США он создал аналоговый компьютер с 400 фотоэлементами на входе с весами, который представляли собой потенциометры переменного сопротивления, регулируемые двигателями. Аналоговые сигналы непрерывно менялись так же, как сигналы от виниловых пластинок. Если внести в перцептрон множество фотографий с танками и без, он научится распознавать танки на незнакомых для него изображениях.

Сообщение об этом в New York Times стало сенсацией (см. рис. 3.4) [68].

Перцептрон способствовал появлению математического анализа разделения шаблонов в многомерном пространстве. Интуитивные предположения о точках в трехмерном пространстве, в котором мы и живем, вводят нас в заблуждение, когда точки расположены в пространстве с тысячами измерений. Русский математик Владимир Вапник[69] представил классификатор, названный «Метод опорных векторов»[70], который обобщил принципы работы перцептрона и стал широко использоваться в машинном обучении. Он нашел путь к автоматическому обнаружению плоскости, которая максимально разделяет две категории (см. рис. 3.5, линейный случай). Это делает обобщение более устойчивым к погрешностям измерения точек в пространстве, и в сочетании с так называемым ядерным трюком (kernel trick), который является нелинейным расширением, алгоритм стал основным в машинном обучении[71].

Закат перцептронов

Тем не менее существовало ограничение, затрудняющее исследования. Упомянутое выше примечание «...если такой набор веса существует» ставит вопрос: какие задачи могут быть решены с помощью перцептронов, а какие – нет? Очень простое распределение точек в двух измерениях не может быть распознано перцептроном (см. рис. 3.5, нелинейные случаи). Оказалось, что «танковый» перцептрон классифицирует не танки, а время суток. Классифицировать танки на изображениях гораздо сложнее, и это невозможно сделать с помощью перцептрона. Это также показывает, что даже если перцептрон чему-то научился, то не обязательно тому, что вы хотели.

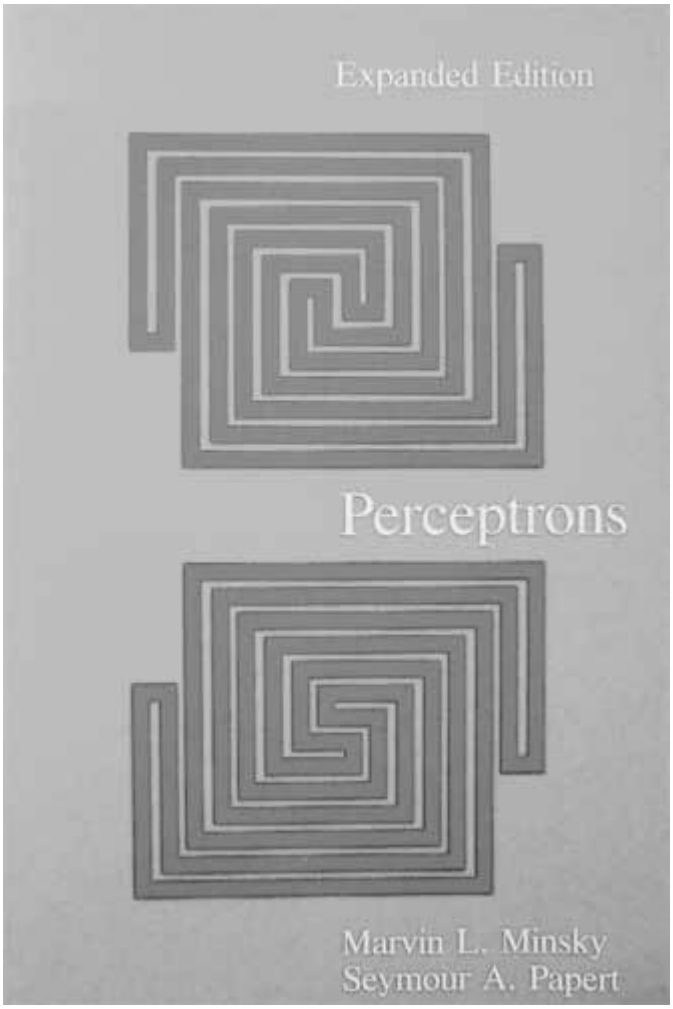


Рис. 3.7. Обложка книги «Перцептроны». Две красные спирали выглядят одинаково, но они разные. Верхняя – это две разные, несоединенные спирали, в то время как нижняя – единая спираль, в чем вы можете убедиться, если проведете внутри нее карандашом. Минский и Пейперт доказали, что перцептрон не может найти отличия между ними. А вы сможете это сделать без отслеживания? Почему нет?

Последним ударом по перцептрону стал математический трактат Марвина Минского и Сеймура Пейперта «Перцептроны», опубликованный в 1969 году. Их геометрический анализ показал, что возможности перцептрона ограничены. Перцептроны могут разграничивать только линейно отделимые категории (см. рис. 3.5). В

конце книги авторы рассмотрели перспективу обобщения однослойного перцептрона на несколько слоев, где один слой переходил в следующий. Многослойные перцептроны более мощные, чем линейные классификаторы, но Минский и Пейперт выражали сомнение, что создание таковых в принципе осуществимо. К сожалению, многие посчитали их сомнения истинными и окончательными, и перцептрон был заброшен, пока новое поколение исследователей нейронных сетей в 1980-х годах не взглянуло на проблему с другой стороны. Обложка книги иллюстрирует геометрическую задачу, которую, по мнению авторов, перцептрон не сможет решить (рис. 3.7). Иронично, но эта проблема трудна и для людей.

В перцептроне входная информация вносит независимые данные в блок выхода. Но что делать, если несколько входных данных должны быть объединены таким образом, чтобы решения зависели от комбинации, а не от каждого факта отдельно? Это и есть причина, по которой перцептрон не может определить, единая спираль или нет: один пиксель не несет никакой информации о том, находится он внутри или снаружи. В многослойном перцептроне возможно соединение комбинаций на промежуточных слоях между модулями входа и выхода. Однако в 1960-х годах ученые не знали, как обучить сеть даже с одним промежуточным слоем.

Фрэнк Розенблатт и Марвин Минский были одноклассниками в Высшей научной школе Бронкса. Они обсуждали свои радикально разные подходы к ИИ на научных встречах, и Минский лидировал. Каждый из них внес важный вклад в понимание перцептрона, что стало отправной точкой глубокого обучения, и очень жаль, что их противостояние закончилось.

Розенблатт трагически погиб при крушении лодки в 1971 году в возрасте 43 лет. Споры о перцептроне были в самом разгаре, и ходили слухи, что он был в подавленном состоянии и, возможно, даже совершил самоубийство. Стало ясно, что «золотой век» открытий новых способов вычислений с помощью нейронных сетей

подходит к концу, и сменилось целое поколение, прежде чем исследования Розенблатта были возобновлены.

Глава 4. Обработка данных как в человеческом мозге

«Если бы у меня был мозг...» – поет Страшила в «Волшебнике из страны Оз». Но Страшила не знал, что у него есть мозг, ведь без него он не мог бы ни петь, ни разговаривать. Ему было всего два дня, и основная его проблема заключалась в отсутствии опыта. Со временем, постепенно узнавая мир вокруг себя, он стал одним из самых мудрых существ в стране Оз; мудрым настолько, чтобы осознать пределы своих возможностей. Напротив, Железный дровосек пел: «Если бы у меня было сердце...» Он спорил со Страшилой, что важнее: мозг или сердце? В стране Оз, как и в реальном мире, знания совместно с эмоциями и создают в процессе обучения интеллект. Оба качества – продукты мозга, находящиеся в хрупком равновесии. Эта классическая история отражает основную тему данной главы: что если бы ИИ имел сердце и мозг?

Как работает мозг

Когда мы с Джеффри Хинтоном (рис. 4.1) встретились в 1979 году на организованном им семинаре, у нас были похожие взгляды на возможности нейронных сетей. Мы быстро нашли общий язык и позже стали вместе работать над открытием нового типа модели нейронных сетей, названной Машиной Больцмана, речь о которой пойдет в главе 7. Новая модель пробила плотину, целое поколение сдерживающую изучение многослойных нейронных сетей.

Раз в несколько лет Джеффри звонил мне и говорил: «Я понял, как работает мозг». Каждый раз появлялась умная схема для улучшения работы моделей нейронных сетей. Потребовалось много таких идей и уточнений для глубокого обучения в многослойных нейронных сетях, чтобы достичь уровня производительности, сопоставимой с человеческим, при распознавании объектов на фотографиях и речи

во время звонка по телефону. Эти возможности получили широкое распространение всего несколько лет назад и теперь широко известны, но путь был долгим.

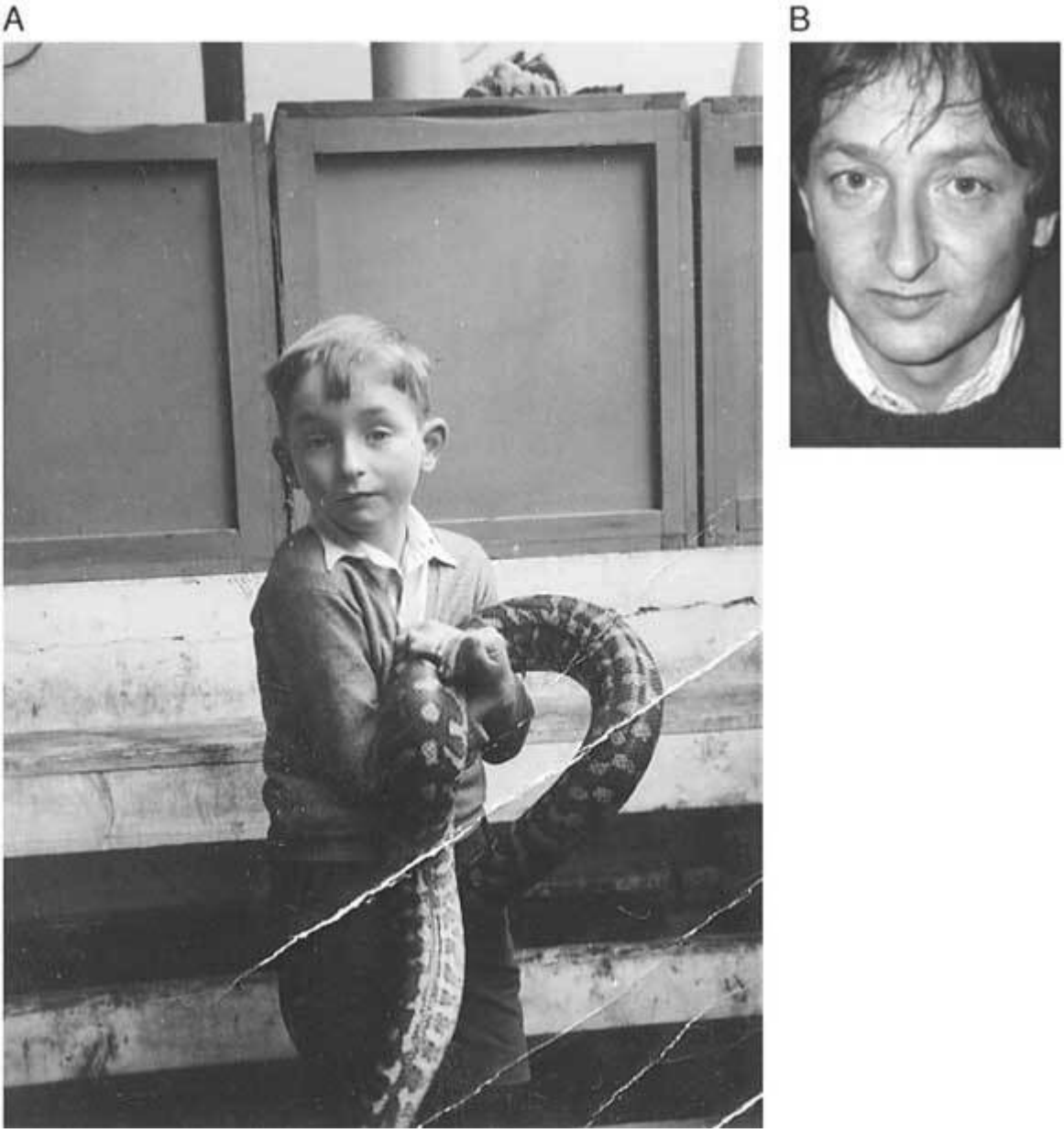


Рис. 4.1. Джеффри Эверест Хинтон в начале карьеры (слева) и в 1979 году во время работы на своем семинаре по параллельным моделям ассоциативной памяти в Ла-Хойя в Сан-Диего. Его второе имя – Эверест – было дано в честь Джорджа Эвереста, который исследовал Индию и выяснил, как измерить высоту самой высокой в

мире горы, которая теперь носит его имя. Фотографии сделаны с разницей в 15 лет

Джеффри получил степень бакалавра психологии в Кембридже и защитил докторскую диссертацию по ИИ в Эдинбургском университете вместе с Кристофером Лонге-Хиггинсом, выдающимся химиком, который изобрел первую модель нейронной сети с ассоциативной памятью. В то время доминирующая парадигма искусственного интеллекта основывалась на написании программ, которые использовали символы, логику и правила, кодифицировавшими интеллектуальное поведение. Когнитивные психологи использовали этот подход для понимания человеческого восприятия и в особенности языков. Джеффри плыл против течения. Никто не мог предположить, что однажды он выяснит, как работает мозг. Его лекции убедительны, он может объяснить абстрактные математические концепции с ясностью, которая требует лишь незначительных познаний в математике. Его остроумие и сдержанный юмор очаровательны. Джеффри по натуре склонен к соперничеству, особенно когда дело касается мозга.

Когда мы впервые встретились, Джеффри был научным сотрудником Калифорнийского университета в Сан-Диего в группе параллельной распределенной обработки под руководством Дэвида Румельхарта и Джея Макклелланда. Джеффри считал, что сети простых процессоров, работающих параллельно и изучающих примеры, – лучший способ понять восприятие. Он был центральной фигурой в вышеупомянутой группе, исследовавшей, как слова и язык могут восприниматься в качестве распространения функции, распределенной по многочисленным узлам сети.

Традиционный подход к языку в когнивистике (науке о мышлении) основан на символических представлениях. Слово «чашка», например, является символом, который обозначает все чашки в целом. Прелесть символов в том, что они позволяют нам упрощать

сложные идеи и работать уже с ними. Однако у символов есть проблема: они настолько сжаты, что их трудно использовать в реальном мире, где чашки бывают разных форм и размеров. Нет логической программы, которая могла бы определить, что конкретно является чашкой, или отыскать ее на картинке, в то время как люди справляются с этим весьма успешно.

Абстрактные понятия, например, справедливость или мир, определить еще труднее. Альтернатива – распределение чашек с помощью схем активности большой популяции нейронов, которые смогут зафиксировать как сходства, так и различия. Это наделяет символ богатой внутренней структурой, отражающей его суть. Проблема в том, что в 1980 году никто не знал, как создать такую внутреннюю структуру.

Мы с Джеффри были не единственными, кто в 1980-х годах верил, что нейронная сеть сможет достичь интеллектуального поведения. Ряд ученых по всему миру, большинство в одиночку, разработали специализированные модели нейронной сети. Например, Кристоф фон дер Мальсбург создал модель распознавания образов, основанную на связи нейронов, передающих импульс[72]. Позже он показал, как эта система может распознавать лица на фотографиях[73]. Кунихико Фукусима из Осакского университета в Японии изобрел неокогнитрон[74] – многослойную сеть, основанную на строении зрительной системы. Эта сеть использовала сверточные фильтры и простейшую модель пластичности Хебба и была прямым предшественником сетей глубокого обучения. Теуво Кохонен, инженер-электрик из Хельсинского университета в Финляндии, разработал самоорганизующуюся сеть, которая могла научиться группировать сходные входные данные, например звуки речи, в двумерную карту так, что разные звуки будут представлены на этой карте разными процессорами, где аналогичная входная информация активирует соседние области выхода[75]. Основным преимуществом сети Кохонена было отсутствие необходимости обозначать каждую

категорию входных данных. Создание специальных меток для обучения перцептрона и других контролируемых сетей стоит дорого. У Кохонена был только один шанс, и он вложил в него все силы.

Многообещающая ранняя попытка систематизировать вероятностные сети принадлежала Джуде Перлу из Калифорнийского университета в Лос-Анджелесе. Он представил сети доверия, которые связывают элементы в сети для определения вероятности – например, вероятности того, что трава мокрая из-за оросителя или потому что прошел дождь[76]. Это мощная основа для отслеживания причинно-следственных связей в окружающем нас мире, однако у нее был роковой недостаток: трудно определить все вероятности. Для автоматического нахождения вероятностей с использованием алгоритмов обучения требовался прорыв. Это стимулировало создание алгоритмов обучения нейронных сетей, речь о которых пойдет во второй части книги.

У этих и других попыток создания нейросетей была общая черта: ни одна из них не работала достаточно хорошо для того, чтобы решать проблемы окружающего нас мира. Более того, первопроходцы редко объединяли свои усилия, что замедляло прогресс. Как следствие, лишь немного ученых, изучающих ИИ в МТИ, Стэндфордском университете и Университете Карнеги – Меллон, воспринимали нейронные сети всерьез. Обработка символов на основе правил получала бóльшую часть финансирования и заданий. Работать над нейронными сетями в ту эпоху – это как быть млекопитающим, покрытым мехом, в эпоху динозавров.

Первые успехи

В 1979 году Джеффри с Джеймсом Андерсоном, психологом из Брауновского университета, организовали семинар по параллельным моделям ассоциативной памяти[77] в Ла-Хойя. Большинство участников семинара встречали друг друга впервые. Я был постдокторантом[78] в Гарвардской медицинской школе, занимался нейробиологией и написал всего несколько узкоспециализированных статей о нейронных сетях, опубликованные в малоизвестных журналах. Именно поэтому я был удивлен, когда меня позвали на встречу. Джеффри потом сказал мне, что они с Дэвидом Марром (рис. 4.2) проверяли меня. Марр был видной фигурой в нейросетевом моделировании и главным идеологом лаборатории искусственного интеллекта в МТИ. Я впервые встретил Марра на небольшом собрании в Джексон-Хоул[79] в 1976 году. У нас были схожие интересы, и он пригласил меня прочитать лекцию в МТИ.

Рис. 4.2. Слева направо: Томазо Поджио, Дэвид Марр и Фрэнсис Крик во время прогулки в Калифорнии в 1974 году. Фрэнсис любил вести со своими гостями длинные дискуссии на различные научные темы

Марр получил степень бакалавра по математике и докторскую степень по физиологии в Кембридже. Его научным руководителем был Джайлз Бриндли – физиолог, специализировавшийся на изучении сетчатки и цветового зрения, а также известный своими работами по музыковедению и лечению эректильной дисфункции. Он прославился тем, что во время лекции на заседании Американской ассоциации урологов в Лас-Вегасе спустил штаны, чтобы продемонстрировать эффективность эрекции, вызванной применением химических препаратов. Докторская диссертация Марра была посвящена нейронной модели обучения в мозжечке – части мозга, ответственной за координацию движений. Он также разработал нейросетевые модели гиппокампа и коры головного мозга, и его выкладки оказались очень дальновидными[80]. Когда я впервые встретил Марра в Джексон-Хоул, он уже перешел в МТИ и работал над зрительным восприятием. Марр был харизматичной

личностью, привлекавшей талантливых студентов. Он начал с сетчатки, в которой свет преобразуется в электрические сигналы, и спросил, как сигналы в сетчатке кодируют особенности объектов и как зрительная кора представляет поверхности и границы объектов. Это называется восходящей стратегией. Например, вместе с Томазо Поджио он разработал гениальную нейросетевую модель стереозрения[81], использующую рекуррентную сетевую модель с обратными связями для определения глубины объекта по небольшим боковым смещениям точечных изображений на двух глазах в стереограммах со случайными точками[82]. Бинокулярное восприятие глубины – основа стереограмм Magic Eye[83].

Марр умер от лейкемии в 1980 году в возрасте 35 лет. Книга «Зрение», над которой он работал в последние годы жизни, была опубликована после его смерти в 1982 году[84]. По иронии, несмотря на восходящий подход Марра, который подразумевает начало исследования зрения с сетчатки и затем моделирование каждого последующего этапа визуальной обработки, его книга больше известна тем, что она пропагандирует нисходящую стратегию: начало исследования с вычислительного анализа задачи, затем построение алгоритма для ее решения и, наконец, реализация алгоритма в аппаратном обеспечении. Это может быть хорошим способом объяснить вещи после того, как вы определили их, однако с помощью такого принципа невозможно исследовать работу мозга. Труден первый шаг – определение задачи, которую решает мозг. Наша интуиция часто вводит нас в заблуждение, особенно когда дело доходит до зрения; мы исключительно хорошо видим, но мозг скрывает от нас нюансы. Позже мы рассмотрим, как был достигнут прогресс в понимании видения, работающего изнутри, с применением алгоритмов обучения.

Фрэнсис Крик присоединился к семинару в Ла-Хойя в 1979 году. После того как в 1953 году совместно с Джеймсом Уотсоном он открыл структуру ДНК, в 1977 году Крик перешел в Институт

биологических исследований Солка и переключил внимание на неврологию. Он пригласил к себе в гости исследователей и вел с ними долгую дискуссию о неврологии, особенно о зрении. Дэвид Марр был среди них. В конце книги Марра есть показательная дискуссия в форме сократического диалога. Позже я узнал, что разговор в книге Марра возник из обсуждения с Криком. Когда я перешел в Институт Солка в 1989 году, я понял ценность таких бесед.

Прапраправнук Джорджа Буля

Джеффри – прапраправнук Джорджа Буля. В 1854 году Буль написал книгу «Исследование законов мышления», которая стала математической основой того, что теперь называется булевой алгеброй, или алгеброй логики (рис. 4.3). Буль – британский учитель-самоучка начала XIX века. У него было пять дочерей, некоторые из них – со способностями к математике. Взгляд Буля на то, как манипулировать логическими выражениями, лежит в основе цифровых вычислений и являлся естественной отправной точкой для молодых исследователей ИИ в 1950-х годах. Джеффри гордился тем, что у него была ручка Буля, которая передавалась в его семье из поколения в поколение.

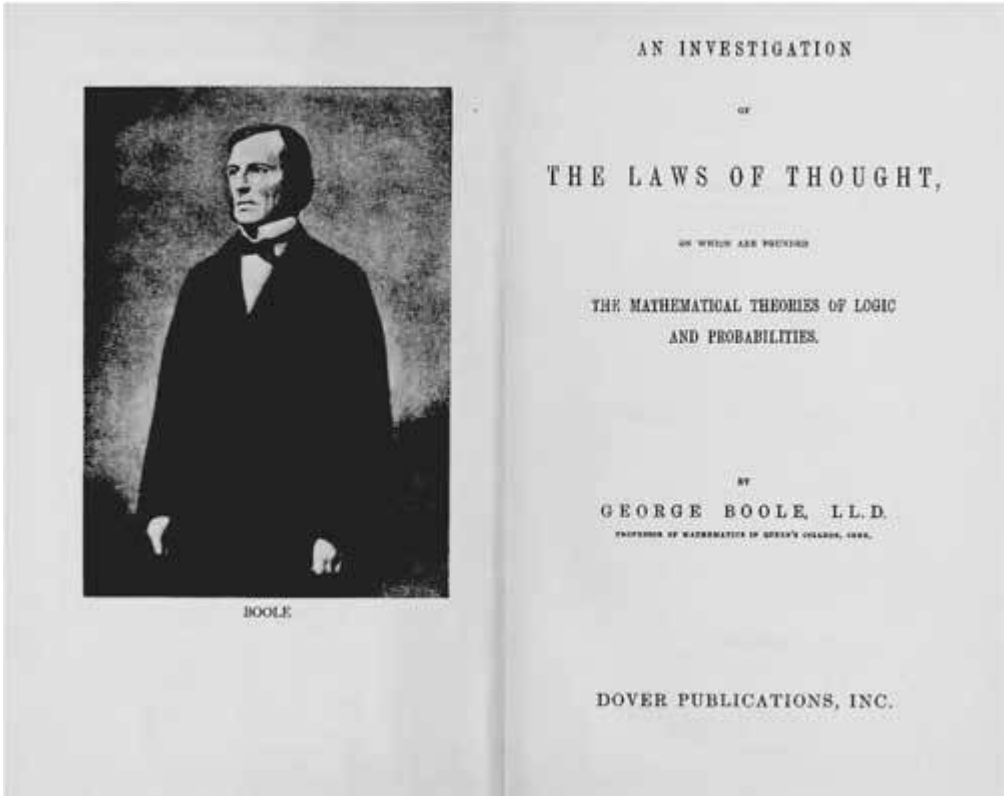


Рис. 4.3. «Исследование законов мышления» Джорджа Буля. Книга известна изучением логики как основы мыслительной деятельности, но также касается вероятностей. Эти две области математики подтолкнули к использованию обработки символов и нейросетевому подходу к ИИ, соответственно

Готовясь к докладу, я однажды взял книгу Буля и обнаружил, что полное название – «Исследование законов мышления, на которых основаны математические теории логики и вероятностей» (рис. 4.3). Буль известен своими работами, посвященным логике, а не вероятностям. Теория вероятностей[85] – основа современного машинного обучения, и она может объяснить неопределенности в реальном мире лучше, чем логика, которая описывает идеальный мир. Так что Буль – один из отцов машинного обучения. Ирония в том, что забытая сторона его работы расцвела спустя 250 лет при помощи его праправнука. Буль гордился бы им.

Проект «Шалтай-Болтай»

Когда я был аспирантом в Принстонском университете, мой подход к пониманию мозга состоял в написании уравнений для сетей нелинейно взаимодействующих нейронов и их анализе[86]. Таким же путем физики на протяжении столетий объясняли природу гравитации, света, электричества, магнетизма и ядерных сил. Каждый день, перед тем как лечь спать, я молился богу физики: «Дорогой Бог, пусть уравнения будут линейными, шум – гауссовым, а переменные – разделяющимися». Это условия, которые приводят к аналитическим решениям; но сетевые уравнения были нелинейные, шум – негауссовым, а переменные – неразделяющимися, поэтому не позволяли сделать однозначные выводы. Более того, моделирование на компьютере уравнений для больших сетей в то время было невероятно медленным. Еще более обескураженный, я понятия не имел, были ли у меня правильные уравнения.

Обучаясь в Принстонском университете, я обнаружил, что нейробиологи достигли невероятного прогресса. Нейробиология – сравнительно молодая наука, она была основана 45 лет назад. До этого исследования в области мозга проводились другими науками: биологией, психологией, анатомией, физиологией, фармакологией, неврологией, психиатрией, биоинженерией и многими другими. Во время первой встречи Общества нейробиологии в 1971 году Вернон Маунткасл лично приветствовал каждого у дверей[87]. Сегодня в обществе уже 40 000 членов, из которых 30 000 ежегодно приходят на встречу. Маунткасл был сотрудником Университета Джонса Хопкинса – там мы и встретились, когда я пришел туда на свою первую работу на факультет биофизики в 1982 году[88]. Он был легендарным нейрофизиологом, открывшим кортикальный столбец. Я тесно сотрудничал с ним при создании Института разума и мозга[89], первого в своем роде.

Есть множество разных уровней исследования мозга (рис. 4.4), и важные открытия были сделаны на каждом из них. Интеграция полученных знаний – сложнейшая задача. Она напоминает детский стишок про Шалтая-Болтая:

Шалтай-Болтай
Сидел на стене.
Шалтай-Болтай
Свалился во сне.
Вся королевская конница,
Вся королевская рать
Не может Шалтая,
Не может Болтая,
Шалтая-Болтая,
Болтая-Шалтая,
Шалтая-Болтая собрать![90]

Нейробиологи очень хорошо разбирают мозг по кусочкам, но собрать эти кусочки воедино – серьезная проблема, которая требует не упрощения, а синтеза, чего я и хочу добиться. Но в первую очередь нужно знать, что это за части, ведь в мозге их множество.

На семинаре для выпускников, который проводил Чарльз Гросс, психолог, изучавший в Принстонском университете зрительную систему обезьян, я был впечатлен прогрессом, достигнутым благодаря записи отдельных нейронов в зрительной коре Дэвидом Хьюбелом и Торстеном Визелем из Гарвардской медицинской школы, которые позже, в 1981 году, получили Нобелевскую премию по физиологии или медицине за новаторские исследования первичной зрительной коры. Их открытия, о которых пойдет речь в главе 5, лежат в основе глубокого обучения, что описано в главе 9. Если физика не сумела проложить дорогу к пониманию работы

мозга, то, возможно, сумеет нейробиология.

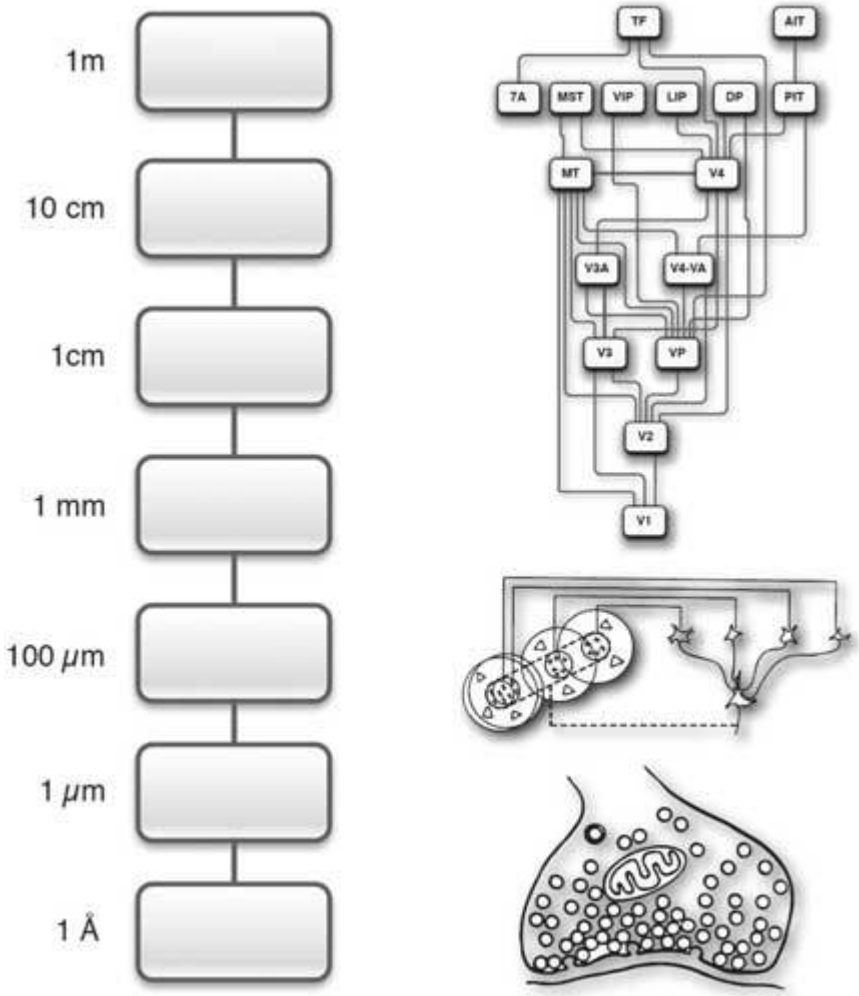


Рис. 4.4. Уровни исследования в головном мозге. Слева: пространственная шкала колеблется от молекулярного уровня (снизу) до всей центральной нервной системы (вверху). Многое известно о каждом из уровней, но наименее изученным является сетевой уровень – небольшие группы взаимосвязанных нейронов. Это уровень, моделируемый искусственными нейронными сетями. Справа: изображения синапса (внизу), простой ячейки зрительной коры (посередине) и иерархии корковых областей в зрительной коре (вверху)

Чему я научился в Вудс-Хоуле

После защиты диссертации по физике в Принстонском университете в 1978 году я принял участие в десятидневном летнем курсе по экспериментальной нейробиологии в Вудсхоулской лаборатории биологии моря. В первый день я пришел в повседневной синей спортивной куртке и аккуратно отглаженных штанах цвета хаки. Стори Лэндис, один из преподавателей курса, отвела меня в сторону и купила мне мою первую пару джинсов. В то время Стори работала на факультете нейробиологии в Гарвардском университете, а вскоре стала руководителем Национального института неврологических заболеваний и инсультов в Национальном институте здоровья. Она до сих пор припоминает мне тот случай.

После летнего курса я остался на несколько недель сентября, чтобы завершить начатый проект. Он позволил получить потрясающие электронно-микроскопические изображения электрорецепции[91] скатов[92]. Скаты и акулы способны воспринимать очень слабые электрические поля; их рецепторы настолько чувствительные, что они могут обнаружить сигнал от 1,5-вольтовой батарейки у другого берега Атлантического океана. Скаты могут применять это шестое чувство для навигации, используя слабые электрические сигналы от своего движения через магнитное поле Земли, которое генерирует микровольтовые сигналы в их электрорецепторах.

Однажды, когда я фотографировал в подвале студенческого общежития Loeb Hall, мне неожиданно позвонил Штефан Куффлер, основатель факультета нейробиологии в Гарвардской медицинской школе. Куффлер – легендарная персона в нейробиологии. Он предложил мне работать в его лаборатории, что изменило мою жизнь. Я переехал в Бостон сразу, как окончил аспирантский проект

у Алана Гельперина по фиксированию метаболической активности в педальном ганглии *Limax maximus*, большого придорожного слизня[93]. Я никогда больше не смогу съесть улитку, не думая о ее мозге. Алан отошел от нейроэтологии, цель которой – изучение нейронных основ поведения. Я узнал, что так называемая более простая нервная система беспозвоночных на самом деле более сложная, так как они выживают с гораздо меньшим количеством нейронов, каждый из которых узкоспециализированный.

В лаборатории Куффлера я изучал передачу сигнала в синапсе симпатического ганглия лягушки-быка – в 60 тысяч раз более медленную, чем быстрая миллисекундная синаптическая передача в коре ее мозга (рис. 4.5)[94]. Это нейроны, которые формируют выход вегетативной нервной системы, регулирующей работу желез и внутренних органов. После стимуляции нерва, ведущего к синапсу, вы успеете сходить за кофе и вернуться до того, как синаптический вход в нейрон достигнет пика, что произойдет примерно за минуту, а затем ему потребуется десять минут, чтобы восстановиться. Синапсы – фундаментальный вычислительный элемент в мозге, и разнообразие типов синапсов говорит о многом. Этот опыт показал мне, что упрощение, возможно, не лучший путь к пониманию работы мозга.

Выяснить, как работает мозг, было не единственной задачей, а целым набором задач, давно решенных эволюцией и передающихся от вида к виду вверх по эволюционной лестнице. В нашем мозгу есть ионные каналы, которые впервые появились в бактериях миллиарды лет назад.

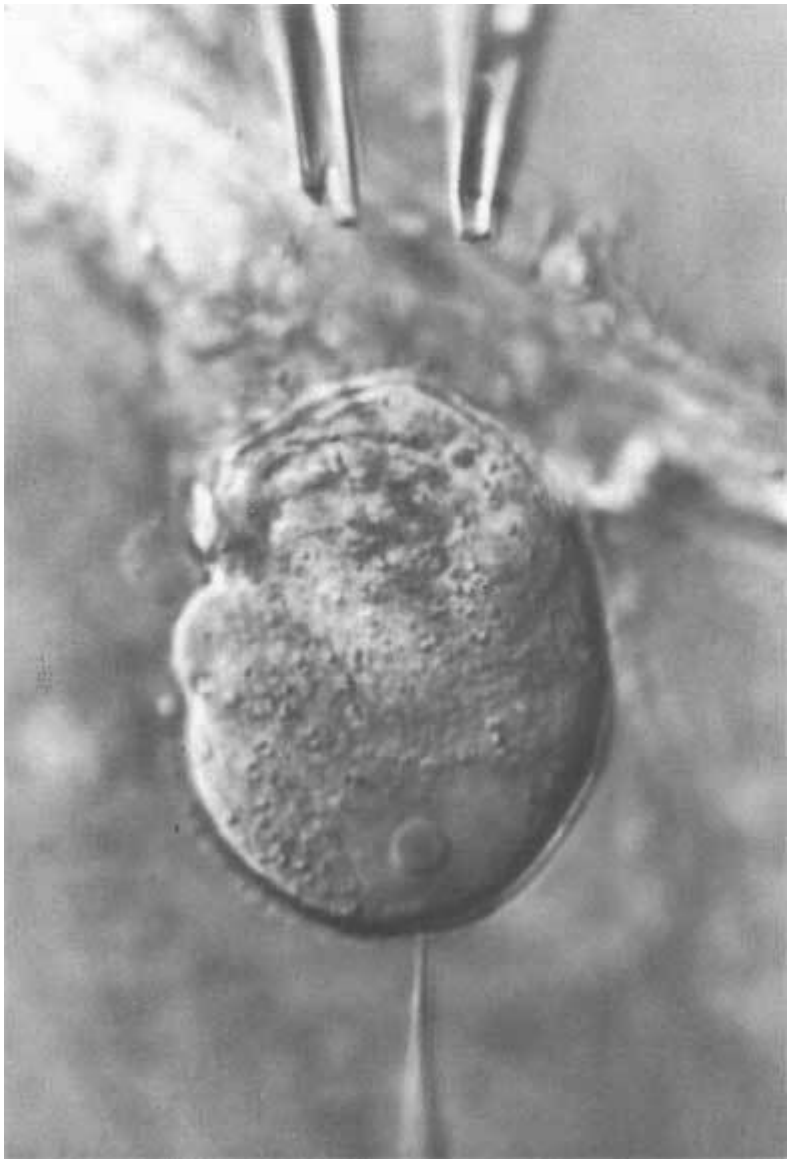


Рис. 4.5. Клетка симпатического ганглия лягушки-быка. Эти нейроны получают входные сигналы от спинного мозга и раздражают железы в коже лягушек. Они большие, их электрические сигналы легко регистрировать с помощью микроэлектрода (внизу), у них нет дендритов[95], и их можно электрически стимулировать нервом (вверху) или химическими веществами (верхняя пара микропипеток). Стимулирование нерва вызывает три различных синаптических сигнала со скоростью нервной реакции в несколько миллисекунд, как и в нервно-мышечном соединении, однако она проходит медленнее, достигает максимума через десять секунд и

длится минуту. Самый поздний ответ на возбуждение выходит на пик через минуту и длится десять минут. Это иллюстрирует широкий диапазон временных масштабов, которые присутствуют даже в простейших нейронах

Недостающее звено

Итак, если физический подход оказался слишком простым, а биологический – слишком сложным, то где же искать оптимальный вариант? В отличие от физических факторов, у схем мозга и моделей нейронных сетей есть цель – решение жизненно важных вычислительных задач, таких как зрительное восприятие и перемещение. Безусловно, можно найти идеальную модель того, как работает нейрон, но это не скажет вам, какова его цель.

Нейроны участвуют в процессе обработки сигналов, несущих информацию, и именно вычисления были недостающим звеном в попытке понять природу. Я шел к этому 40 лет, осваивая новое направление – вычислительную нейробиологию.

Рис. 4.6. Мы с Джеффри Хинтоном обсуждаем сетевые модели зрения в Бостоне в 1980 году. Фотография сделана спустя год после знакомства на семинаре по параллельным моделям ассоциативной памяти в Ла-Хойя и за год до того, как я начал работать в лаборатории Университета Джонса Хопкинса в Балтиморе, а Хинтон основал свою исследовательскую группу в Университете Карнеги – Мелона в Питтсбурге

После аспирантуры в Калифорнийском университете в Сан-Диего Джеффри Хинтон вернулся в Англию, где начал исследования на факультете прикладной психологии в Кембридже. Однажды в 1981 году в два часа ночи ему позвонил некто, представившийся Чарльзом Смитом, президентом компании System Development Foundation[96]. Смит сказал, что его компания готова спонсировать рискованные исследования Хинтона, которые вряд ли увенчаются успехом, а самого Джеффри ему очень рекомендовали. Джеффри не верил, что все это не сон. Будучи хорошим другом[97], Джеффри сказал Смиту, что у моих исследований еще меньше шансов на удачное завершение.

Фонд действительно дал нам первые гранты, которые сильно продвинули исследования. Мы смогли себе позволить купить более быстрые компьютеры и платить зарплату студентам. До этого у Джеффри был компьютер Apple II, который он заменил Лисп-машиной[98], когда перешел в университет Карнеги – Мелона. Когда я приступил к работе в Университете Джонса Хопкинса в Балтиморе, одно время у меня были компьютеры большей мощности, чем у всего факультета информатики[99]. Я также купил первый модем, который связал Университет Хопкинса с компьютерной сетью ARPANET, предшественником Интернета, чтобы мы с Джеффри могли переписываться по электронной почте. Мы не могли мечтать о

большем, чтобы начать нашу карьеру и исследовать новые направления (рис. 4.6).

Глава 5. Понимание зрительной системы

Одним из моих первых воспоминаний, еще до детского сада, были кусочки головоломки, которые нужно соединять по форме, цвету и смыслу. Мои родители удивляли своих друзей на вечеринках тем, как быстро их малыш собирает головоломки. Тогда я этого не знал, но мой мозг уже делал то, что умеет лучше всего, – решал задачи на распознавание образов. В науке много проблем, похожих на пазлы с недостающими частями и неясными намеками на то, как выглядит полная картина. Основная сложность – понять, как именно мозг решает задачи.

Клуб Гельмгольца был небольшой научной организацией, куда входили ученые из Калифорнийского университета в Сан-Диего, Калифорнийского технологического института[100], Университета Южной Калифорнии и Калифорнийского университета в Лос-Анджелесе, которые собирались каждый месяц в Калифорнийском университете в Ирвине[101]. Герман фон Гельмгольц, физик и врач XIX века, разработал математическую теорию и экспериментальный подход к зрению, которые легли в основу современного понимания зрительного восприятия. Как секретарь клуба Гельмгольца я должен был найти оратора для выступления во второй половине дня перед 10–20 членами клуба и их гостями. Затем следовал еще один доклад члена клуба. Лекции проходили в режиме диалога с аудиторией, и для обсуждения отводилось достаточно времени. Данные встречи были важным событием в ученой среде, и один из гостей сказал мне, что его удивили вопросы: «Они действительно хотели знать ответы!» Это были мастер-классы, посвященные зрению[102].

Зрение – наше самое сильное и самое изученное чувство. Обоняние у приматов давно атрофировалось по сравнению с обонянием у крыс и собак. Поскольку пара глаз у нас расположена

спереди, у нас прекрасное бинокулярное восприятие глубины, и половина нашей зрительной коры – зрительная. Особый статус зрения отражает поговорка «Лучше один раз увидеть, чем сто раз услышать». Если бы собака могла говорить, она бы сказала: «Лучше один раз понюхать». Однако тот факт, что мы так хорошо видим, не дал нам прозреть огромную вычислительную сложность проблемы, которую решала природа в течение сотен миллионов лет эволюции, о чем говорилось в главе 2. Организация зрительной коры послужила примером для наиболее успешных сетей глубокого обучения.

За десятую долю секунды десять миллиардов нейронов в зрительной коре, работающие параллельно, могут идентифицировать чашку среди большого количества предметов, даже если вы никогда раньше не видели именно эту чашку. Она может быть в любом месте, любого размера и в любом положении по отношению к вам. Я, будучи аспирантом в Принстонском университете, был очарован зрительной системой и работал все лето в лаборатории Чарльза Гросса, который изучал нижневисочную кору у обезьян. Эта зона находится на одной из самых высоких ступеней в иерархии областей коры головного мозга (рис. 5.1), и Гросс обнаружил в ней нейроны, которые реагируют на сложные объекты, такие как лица и, что примечательно, ершики для унитаза[103].

Штефан Куффлер, с которым я работал на факультете нейробиологии в Гарвардской медицинской школе, обнаружил, как ганглиозные клетки в сетчатке кодируют визуальные сцены. Я работал там, когда Дэвид Хьюбел и Торстен Визель получили Нобелевскую премию по физиологии или медицине в 1981 году за фундаментальные открытия в области зрительной коры головного мозга. Штефан Куффлер, возможно, получил бы премию вместе с ними за исследования сетчатки, но он умер в 1980 году, а чтобы получить Нобелевскую премию, нужно быть живым. В конце концов я перебрался в Институт биологических исследований Солка, где Фрэнсис Крик сосредоточился на зрении, когда в 1977 году решил

перейти от молекулярной генетики к мозгу. Его целью было найти минимально необходимый набор нейронов для зрительного восприятия. Мне выпала честь быть в компании величайших ученых моего времени, работающих в области зрения.

Рис. 5.1. Схема потока информации через зрительную систему макаки. Стрелки указывают схему передачи визуальной информации между зрительными зонами, начиная с сетчатки, с задержками в миллисекундах на каждом этапе ее обработки. Зрительное восприятие макаки схоже с нашим, и эти этапы у нас одинаковые. LGN: Латеральное коленчатое тело; V1: Первичная зрительная кора; V2: Вторичная зрительная кора; AIT и PIT: передние и задние части височных долей; PFC: префронтальная кора; PMC: премоторная кора; MC: моторная кора. [Thorpe, SJ, Fabre-Thorpe, M, Seeking Categories

in the Brain, Science 291: 260–263 (2001)]

Схема работы зрения

Давайте проследим сигналы, возникающие в мозге при взгляде на картинку, и посмотрим, как они последовательно трансформируются снова и снова по мере того, как переходят из одной стадии на другую (рис. 5.1). Зрительная система начинается с сетчатки, где фоторецепторы превращают свет в электрические сигналы. В сетчатке два слоя нейронов, которые обрабатывают визуальные сигналы в пространстве и времени и заканчиваются ганглиозными клетками, выходящими из зрительного нерва.

В 1953 году Штефан Куффлер (рис. 5.2) записал данные с выходных нейронов сетчатки кошки и одновременно стимулировал их ответ на пятна света. Он отметил, что сигналы на выходе двух видов: одни реагировали на появление пятна света в их центре, а другие – на его смещение. Однако окружающие центры кольца имели противоположную полярность: положительный центр и отрицательное кольцо, и наоборот (рис. 5.3). Такая реакция на свет как раздражитель – свойство рецептивного поля ганглиозных клеток. Это классический эксперимент, результаты которого применимы ко всем млекопитающим.

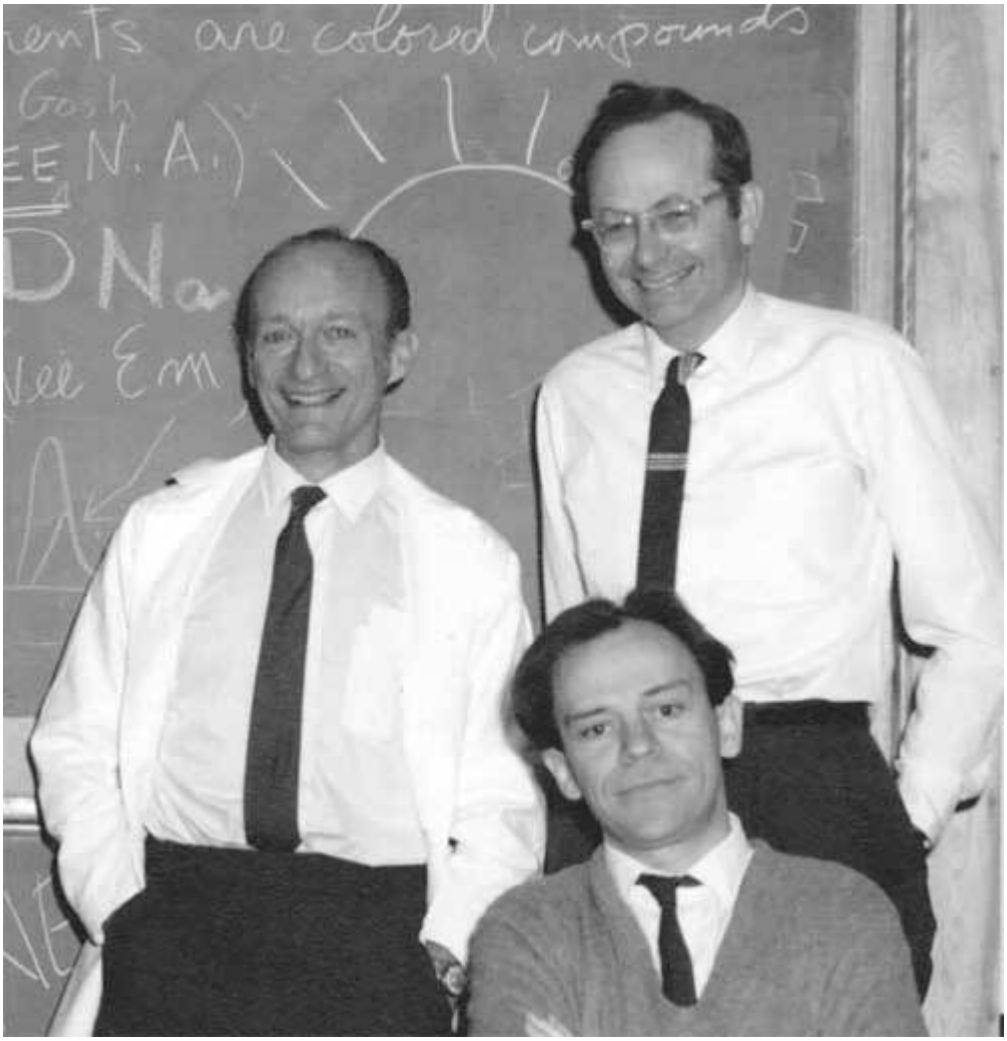


Рис. 5.2. Слева направо: Штефан Куффлер, Торстен Визель и Дэвид Хьюбел. Факультет нейробиологии в Гарвардской медицинской школе был основан в 1966 году, фотография сделана в самом начале его существования. Я ни разу не видел их за работой в лаборатории в галстуках, так что это, скорее всего, был особый случай

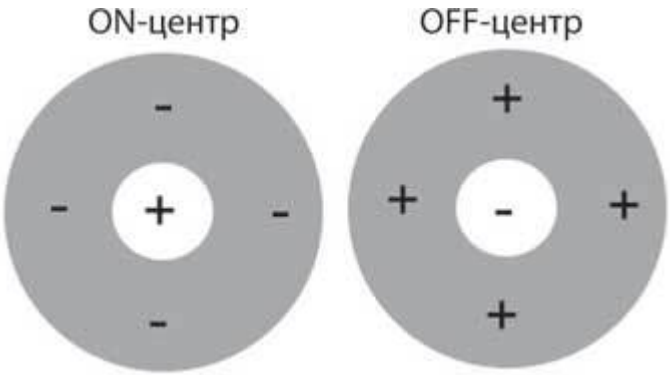


Рис. 5.3. Особенности отклика ганглиозных клеток сетчатки. Два кольца на изображении показывают реакцию двух типов ганглиозных клеток сетчатки, которые посылают закодированные сообщения в мозг, чтобы вы могли видеть. Для типа с ON-центром появление пятна света в центре (+) вызывает всплеск импульсов, а в кольце вокруг центра (–) приводит к подавлению активности. И наоборот для клеток с OFF-центром: появление пятна света в центре (–) подавляет реакцию, а в кольце вокруг центра (+) – получает бурный отклик. Изменения освещения несут важную информацию о перемещениях объекта-раздражителя и его четких границах. Эти свойства были обнаружены Штефаном Куффлером в 1953 году

Я однажды спросил Куффлера, что подвигло его исследовать сетчатку, так как его основной научный интерес был сосредоточен на свойствах синапсов между нейронами. Он сказал, что в то время его лаборатория находилась в Институте офтальмологии Уилмера при Университете Джонса Хопкинса, и он чувствовал себя виноватым из-за того, что его работа не была связана со зрением. Начав исследование отдельных ганглиозных клеток в сетчатке, он передал проект двум научным сотрудникам своей лаборатории, Дэвиду Хьюбелу и Торстену Визелю (см. рис. 5.2) и посоветовал им проследить, как передаются сигналы мозгу. В 1966 году Куффлер и его аспиранты переехали в Гарвардскую медицинскую школу, открыв там кафедру нейробиологии.

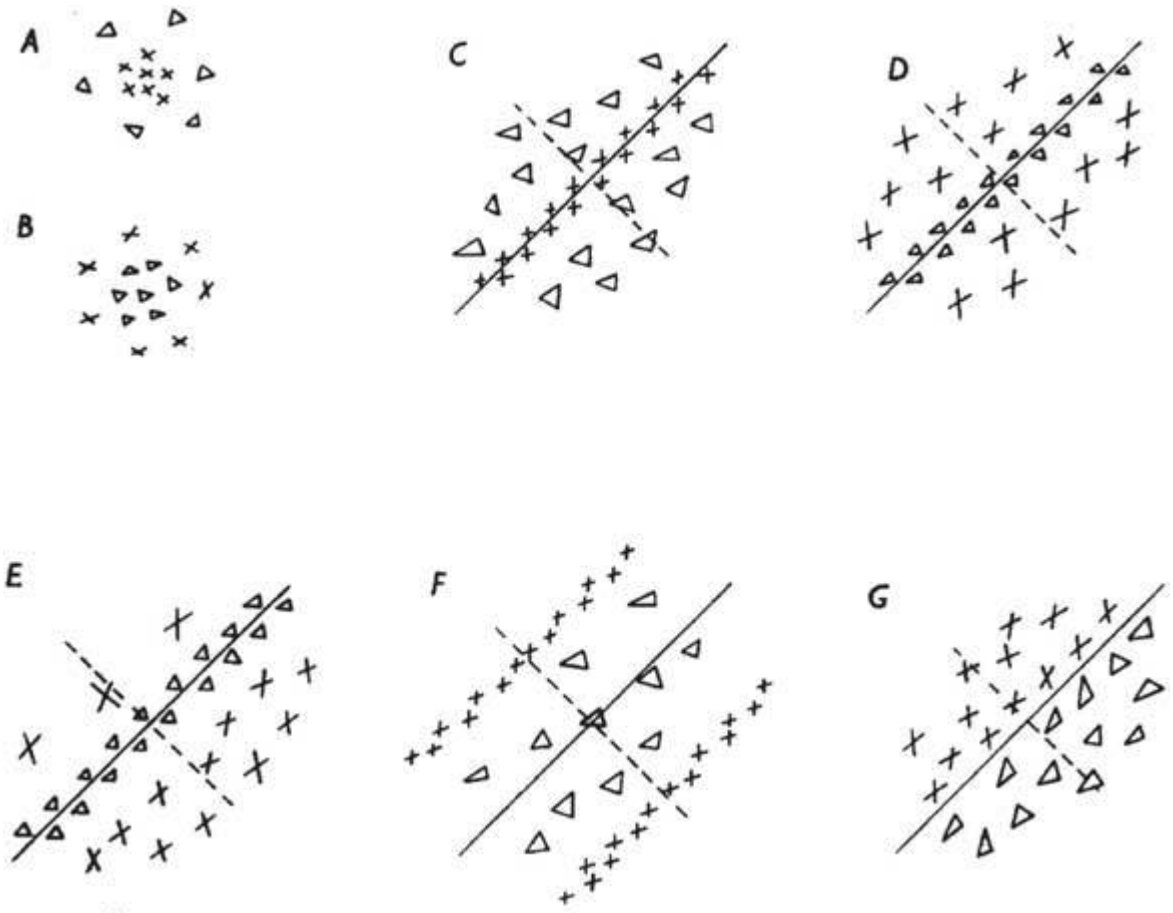


Рис. 5.4. Рецептивное поле и кривая настройки нейронов в первичной зрительной коре кошки. Когда полоса света (вверху справа) мигает в участке поля зрения (слева) одиночной клетки, всплеск реакции регистрируется вначале в одних областях (красных), а при смещении – в других (синих). Наиболее сильный ответ – когда ориентация полосы находится в предпочтительном направлении нейрона (вдоль длинной оси). Частота импульсов, испускаемых нейроном (справа), зависит от ориентации полосы

Зрительная система в коре мозга

Хьюбел и Визель обнаружили, что кортикальные нейроны реагировали гораздо лучше на ориентированные полосы света и четкие границы, чем на пятна света. Каналы в коре трансформировали входные сигналы. Ученые описали два основных типа клеток: ориентированная одиночная клетка, имеющая on- и off-зоны, такие как ганглиозные клетки (рис. 5.4), и ориентированный комплекс клеток, который равномерно ответил на ориентированные стимулы в любую точку рецептивного поля нейрона (рис. 5.5).

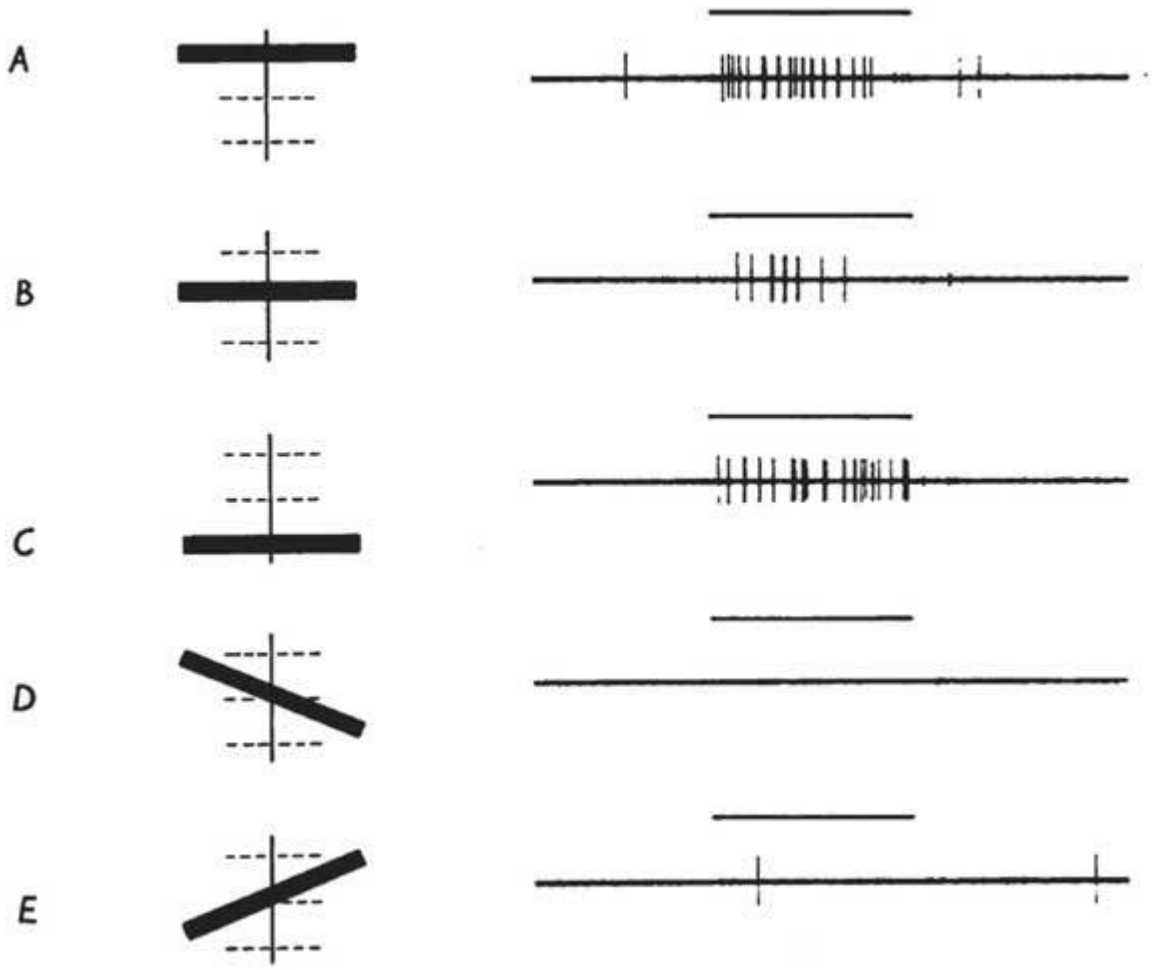


Рис. 5.5. Реакция сложной клетки первичной зрительной коры кошки. Длинная узкая полоса света вызывает всплеск реакции везде, где она находится в пределах рецептивного поля (темный прямоугольник) сложной ячейки при условии правильной

ориентации (три верхних изображения). Неоптимальная ориентация дает более слабый ответ или вообще никакого (нижнее изображение)

Каждый кортикальный нейрон в зрительной коре можно рассматривать как детектор визуальных признаков, который становится активным только в том случае, если получает достаточно входных данных, соответствующих его предпочтительной чувствительности в определенном участке поля зрения, чтобы превысить порог. «Предпочтения» каждого нейрона определяются его связями с другими нейронами. Хьюбел и Визель также обнаружили, что входящие сигналы от двух глаз организованы в чередующихся левых и правых столбцах в среднем слое (IV) коры головного мозга, там где импульсы поступают с «промежуточной станции» – таламуса, или зрительных бугров. Монокулярные нейроны в IV слое проецируются на нейроны в верхних слоях (II и III), которые получают бинокулярные сигналы (рис. 5.6). Предпочтительная ориентация каждой клетки в столбце одинакова и плавно изменяется по всей коре.

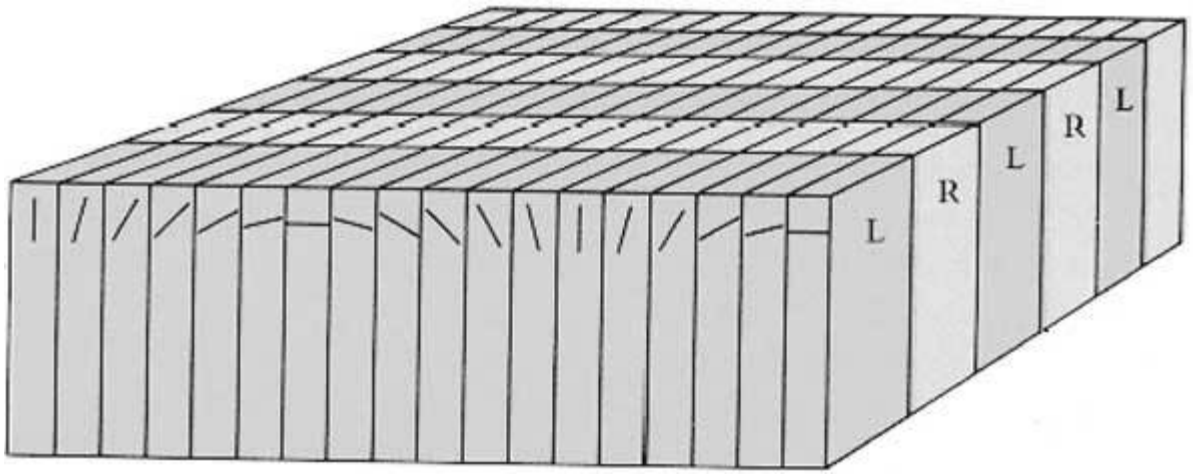


Рис. 5.6. «Кубик льда» – модель нейронов в первичной зрительной коре. При вертикальном проникновении все нейроны имеют

одинаковые ориентационные предпочтения и окулярное доминирование[104]. Под каждым квадратным миллиметром коры находится полный набор ориентаций, которые медленно меняются по всей поверхности коры (правая сторона куба) и поступают от обоих глаз (левая сторона куба). «Пятнышки» богаты цветоизбирательными клетками (вертикальные стержни)

Пластичность синапса

Если один глаз кошки закрыт в течение первого года жизни, то кортикальные нейроны, которые обычно управляются обоими глазами, становятся монокулярными, управляемыми исключительно открытым глазом[105]. Монокулярная депривация приводит к изменениям в силе синапсов в первичной коре, где входы в нейроны впервые получают сходящиеся сигналы от двух глаз. После того как критический период кортикальной пластичности в первичной зрительной коре пройден, закрытый глаз больше не может влиять на кортикальные нейроны – развивается амблиопия («ленивый глаз»). Рассогласованность глаз или косоглазие, которые часто встречается у младенцев, значительно уменьшает количество бинокулярных кортикальных нейронов и препятствуют бинокулярному восприятию глубины[106]. Операция по выравниванию глаз в критический период может спасти бинокулярные нейроны.

Монокулярная депривация – пример высокой пластичности, которая присутствует на ранних стадиях развития, поскольку среда формирует синаптические связи между нейронами в коре и других частях мозга. Эти зависящие от активности изменения распространяются поверх непрерывного обновления, которое происходит во всех клетках. Практически каждый компонент нейронов и синапсов, которые соединяют их, ежедневно меняется. Белки замещаются новыми по мере износа, обновляются липиды в мембране. Но большинство нейронов в коре те самые, что были у нас при рождении[107]. При таком непрерывном обороте остается загадкой, как ваши воспоминания сохраняются в течение всей вашей жизни. Если у старого топора заменить топорище, а затем лезвие – будет ли это тот же старый топор?

Есть еще одно возможное объяснение очевидной долговечности воспоминаний: они похожи на шрамы на вашем теле, которые

сохранились как следы прошлых событий вашей жизни. Эти отметины нужно искать не внутри нейронов, где постоянно идут изменения, а снаружи, во внеклеточном пространстве, где внеклеточный матрикс между нейронами состоит из протеогликанов, которые схожи с коллагеном в рубцовой ткани – жесткий материал, сохраняющийся на протяжении многих лет[108]. Если эта гипотеза когда-либо будет доказана, значит, долговременные воспоминания встроены во внешнюю оболочку мозга и мы искали их не там[109].

Химические синапсы содержат сотни уникальных белков, контролирующих высвобождение нейромедиатора и активацию рецепторов в принимающем нейроне. Большинство синапсов пластичны: как форма жесткого пластика может измениться под воздействием тепла, так и сила синапса может избирательно становиться больше или меньше даже в сотню раз. Примеры синаптических алгоритмов обучения, обнаруженных в мозге, будут рассмотрены в книге далее. Еще примечательнее, что в коре постоянно образуются новые синапсы, а старые удаляются, и это делает их одними из самых динамичных частей организма. В мозге около ста различных типов синапсов, наиболее распространенным возбуждающим нейромедиатором в коре является глутаминовая кислота, а наиболее распространенным ингибирующим передатчиком – гамма-аминомасляная кислота (ГАМК). Также широкий диапазон и у электрохимического воздействия, которое молекулы нейромедиатора оказывают на другие нейроны. Например, симпатические ганглиозные клетки лягушки-быка, о которых написано в главе 4, имеют синапсы с временными шкалами от миллисекунд до минут (рис. 4.5).

Восстановление формы объекта по теням

Стивен Цукер (рис. 5.7) работает над слиянием компьютерного и биологического зрения. Я знаком с ним уже более 30 лет, и все это

время он трудится над книгой, которая объяснит, как работает зрение.

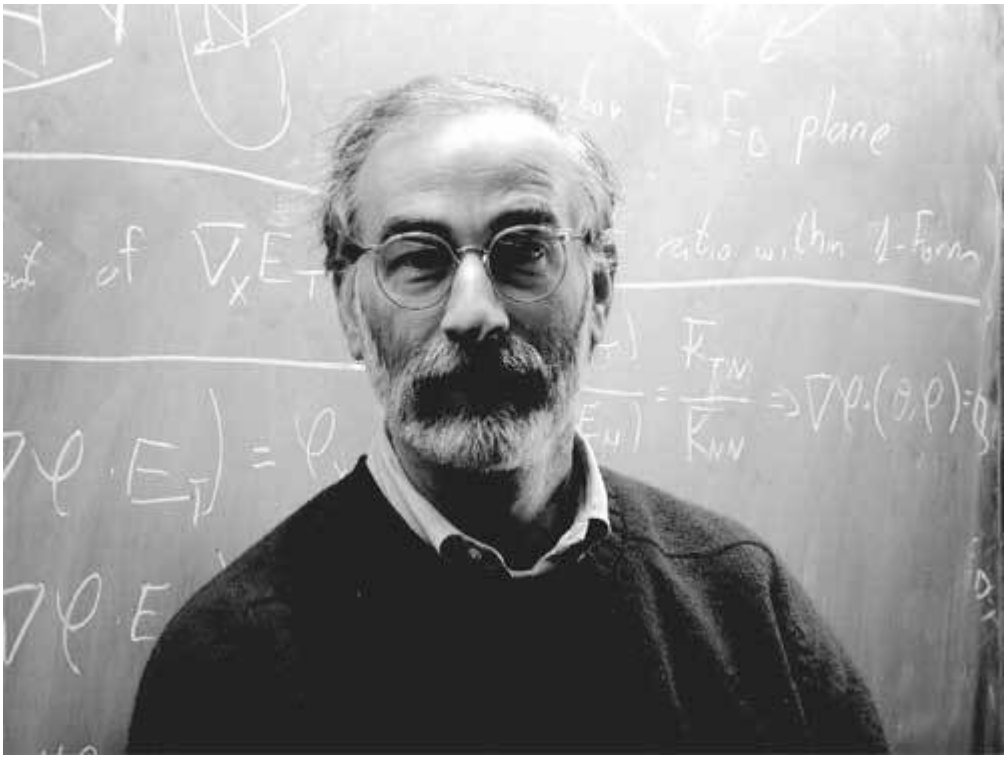


Рис. 5.7. Стивен Цукер в Йельском университете. Освещение на снимке падает сверху справа. По изменению оттенков на его свитере можно понять, какой формы складки. Уравнения на доске позади него вдохновлены зрительной корой мозга обезьян и объясняют, как это происходит. Мы видим одни и те формы независимо от источника света

Его проблема в том, что он все продолжает делать открытия в области зрения и, как у Тристрама Шенди, персонажа Лоренса Стерна[110], конец его книги откладывается тем дальше, чем больше открытий он делает. Подход Цукера к зрению основан на восхитительно упорядоченной структуре первичной зрительной коры, в отличие от структуры других частей коры, где нейроны располагаются чуть ли не вразнобой (см. рис. 5.6) и буквально молят

о строгой схеме. Большинство исследователей в области компьютерного зрения пытаются распознавать объекты по ряду отличительных признаков, отделяя сами объекты от фона.

Стив хотел большего. Он хотел понять, как мы считываем форму объектов из поверхностных теней и явных признаков изгибов и складок. В интервью на пленарном заседании ежегодного собрания Общества нейробиологии в 2006 году у Фрэнка Гери[111], архитектора, проектирующего похожие на паруса здания, спросили, как ему приходят в голову такие идеи? Он ответил, что его вдохновляет рассматривание смятой бумаги в корзине для мусора. Возьмите лист бумаги, скомкайте, положите на стол и посмотрите на него. Как наша зрительная система соединяет сложную форму бумаги с рисунком складок и затененных поверхностей? Как мы воспринимаем меняющиеся формы поверхностей здания Музея Гуггенхайма в Бильбао (рис. 5.8)?

Рис. 5.8. Музей Гуггенхайма в Бильбао, спроектированный Фрэнком Гери. Тени и отражения от криволинейных поверхностей создают сильное впечатление формы и движения. Крошечные люди на дорожке показывают масштаб здания

Стив Цукер недавно смог объяснить, как мы видим складки на затененных изображениях, основываясь на тесной взаимосвязи между трехмерными очертаниями поверхности, как на контурных картах гор, и контурами постоянной яркости на изображениях (рис. 5.9)[112].

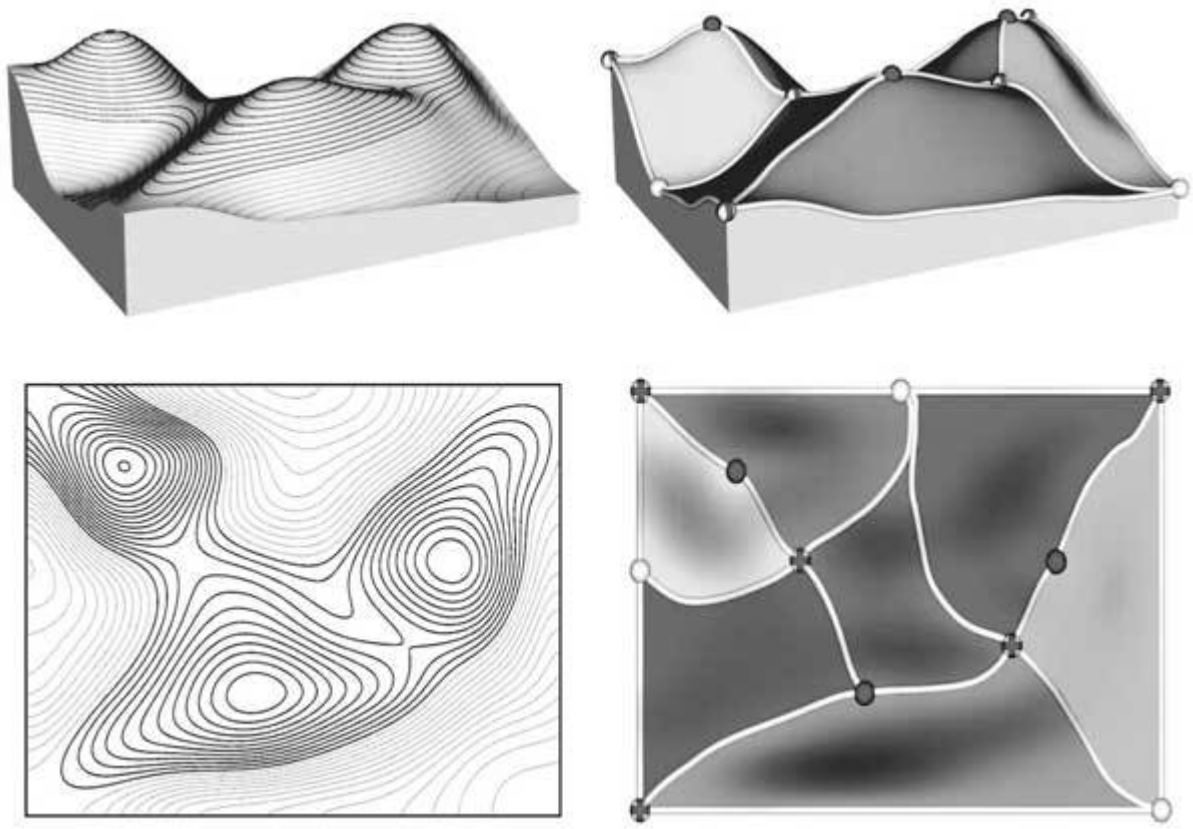


Рис. 5.9. Высотные контуры поверхности (слева сверху) по сравнению с изофотами (кривыми, соединяющими точки равной яркости) той же поверхности (слева внизу). В обоих случаях разделение происходит в одних и тех же критических точках, как показано на рисунках справа (Кансберг и Цукер)

Связь обеспечивается геометрической формой поверхности[113]. Это объясняет, почему наше восприятие формы настолько нечувствительно к различиям в освещении и свойствам поверхности объектов. Это также может объяснить, почему мы так хорошо читаем

контурные карты, где контуры сделаны явными, и почему нам достаточно лишь несколько характерных внутренних линий, чтобы увидеть форму объектов в мультфильмах.

В 1988 году мы с Сидни Леки задались вопросом, сможем ли мы обучить нейронную сеть с одним слоем скрытых элементов для вычисления кривизны затененных поверхностей[114]. Нам это удалось, и, к своему удивлению, мы обнаружили, что скрытые элементы выглядят как простые клетки. Однако при ближайшем рассмотрении мы заметили, что не все клетки одинаковы. Рассматривая проекции простых клеток на выходной слой, который был обучен вычислять кривизну с помощью алгоритма (глава 8), мы обнаружили, что некоторые простые клетки использовались для выбора между положительной кривизной (выпуклым) и отрицательной (вогнутым) (рис. 5.10). Эти простые клетки были детекторами. Они, как правило, имели либо низкую, либо высокую активность, демонстрируя бимодальное распределение. В отличие от них, у других простых клеток отклик был разной силы и они функционировали как фильтры, которые сообщали элементам на выходе направление и величину кривизны.

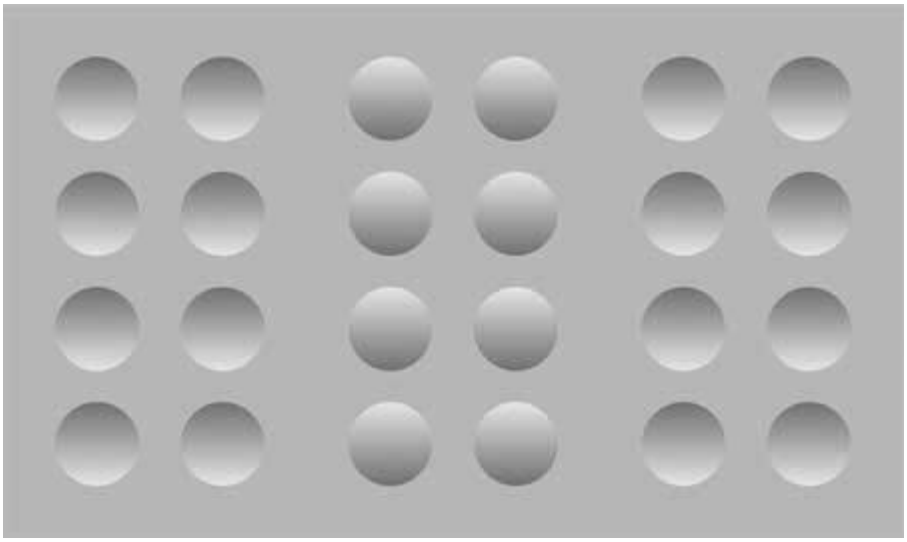


Рис. 5.10. Кривизна от затенения. Наша зрительная система может извлечь форму объекта из плавно меняющейся яркости изображения в пределах ограничивающего контура. Вы видите выпуклые или вогнутые формы в зависимости от направления затенения и вашего предположения о направлении освещения. Переверните книгу вверх ногами, чтобы увидеть изображения по-другому. [Ramachandran V. S. (1988). «Perception of shape from shading». Журнал Nature, 331, 163–165.]

Вывод был неожиданным: нельзя определить функцию нейрона, зная только то, как он реагирует на входящие данные. Функция нейрона также зависит от нейронов, которые он активирует на выходе, что мы назвали проекционным полем нейрона. До недавнего времени это поле было гораздо труднее определить, чем входные данные, но новые генетические и анатомические методы позволяют с большей точностью отслеживать, как передаются импульсы по аксонам (длинным отросткам нервных клеток), а новая технология оптогенетика[115] дает возможность избирательно стимулировать конкретные нейроны для исследования их влияния на восприятие и поведение[116]. Тем не менее эта небольшая сеть в состоянии только определить кривизну выпуклости или впадины. И мы до сих пор не знаем, как целостные образы, которые в литературе по психологии называют гештальтом, организованы в коре. Мы со Стивом Цукером однажды застряли в международном аэропорту Стэплтон в Денвере в 1984 году, наши рейсы задержали из-за метели. Он, как и я, восторгался вычислительной нейробиологией, которая все еще находилась в зачаточном состоянии. Мы придумали семинар, который объединит теоретиков и практиков этой науки, и решили организовать его в Вудс-Хоул, где я прошел летний курс нейробиологии и куда вернулся через несколько лет, чтобы работать со Штефаном Куффлером над физиологическими экспериментами в Лаборатория биологии моря.

Вудс-Хоул – красивая прибрежная деревня на полуострове Кейп-Код недалеко от Бостона. На протяжении долгих лет многие ведущие исследователи, изучающие зрение, посещали этот ежегодный семинар, ставший еще одним моим научным достижением. Семинары положили начало вычислительной теории зрительной коры, но ее подтверждение займет еще 30 лет. В главе 9 мы увидим, что архитектура самой успешной сети глубокого обучения удивительно похожа на зрительную кору.

Иерархическая организация визуальных карт коры головного мозга

Джон Каас и Джон Оллмэн, работавшие в Университете Висконсина, исследовали те области мозга, которые получали сигналы от первичной зрительной коры, и обнаружили, что у них разные свойства. Например, они выявили карту поля зрения в средне-височной зоне[117], нейроны которой реагировали на ориентированные зрительные стимулы, движущиеся в предпочтительном направлении. Оллман как-то упомянул, что им было трудно заставить заведующего кафедрой Клинтона Вулси признать их открытие. В предшествующих экспериментах Вулси использовал для записи более грубые методы и пропустил эти области зрительной коры. Не все открытия сразу же принимаются научным сообществом. Впоследствии в зрительной коре обезьяны было обнаружено два десятка зрительных зон.

Дэвид ван Эссен, работавший в то время в Калтехе, тщательно изучил входы и выходы каждой зрительной зоны и расположил их в виде иерархической диаграммы (рис. 5.11). Схема напоминала карту метро огромного города, с прямоугольниками, обозначающими станции, и соединяющими их линиями высокоскоростного транспорта, и ее иногда используют, чтобы показать сложность коры головного мозга.

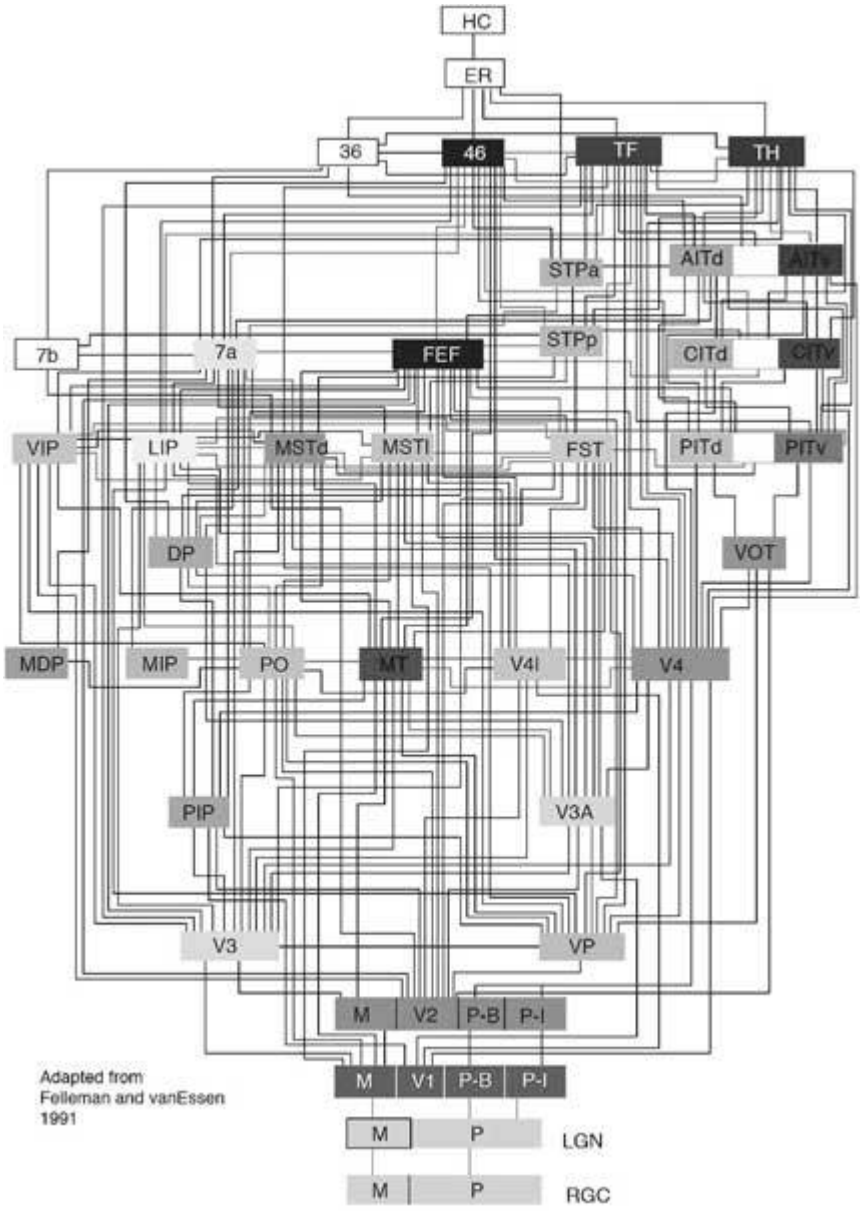


Рис. 5.11. Иерархия зрительных зон в мозге обезьяны. Визуальная информация от ганглиозных клеток сетчатки (retinal ganglion cells; RGC) проецируется на латеральное коленчатое тело (lateral geniculate nucleus; LGN) таламуса, чьи релейные клетки передают сигнал на первичную зрительную кору (V1). Иерархия кортикальных областей заканчивается в гиппокампе (HC). Почти все 187 каналов в диаграмме двунаправлены, у них прямая связь с отделом ниже и обратная связь с отделом выше. Источник: Source: Daniel J. Felleman and David C. Van Essen, “Distributed Hierarchical Processing in Primate

Visual Cortex," Cerebral Cortex 1 (1991): 1–47

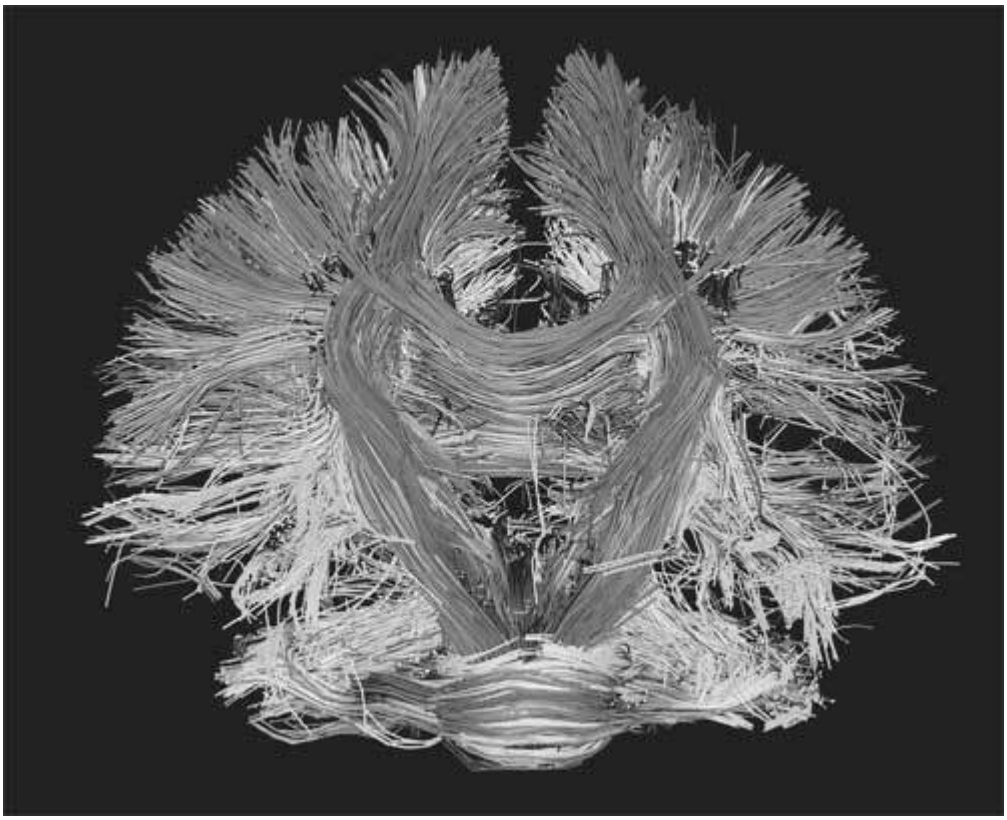


Рис. 5.12. Коннектом человека. Длинные проводящие пути в белом веществе коры головного мозга можно проследить бесконтактным способом с помощью магнитно-резонансной томографии, основанной на неравномерной диффузии молекул воды. Разные пути искусственно окрашены в разные цвета www.pinterest.com/pin/350366045987135160/

Визуальный вход из ганглиозных клеток сетчатки (RGC) проецируется в первичную зрительную кору (V1) внизу диаграммы. Оттуда сигналы переносятся вверх по иерархии, каждая область специализируется на отдельном аспекте зрения, например на восприятии формы. Ближе к вершине иерархии в нижневисочной зоне (AIT, CIT, PIT) в правой части диаграммы рецептивные поля

нейронов охватывают все поле зрения и реагируют преимущественно на сложные визуальные стимулы, такие как лица и другие объекты. Ван Эссен перешел в Вашингтонский университет в Сент-Луисе, и сейчас он один из директоров масштабного научно-исследовательского проекта «Коннектом[118] человека», спонсированного Национальным институтом здравоохранения США[119]. Цель проекта – использовать методы визуализации мозга на основе МРТ[120] для разработки сопоставимой карты дальних связей в мозге человека (рис. 5.12).

Появление когнитивной нейробиологии

Самые высокие уровни функционирования мозга труднее всего поддавались изучению, однако ситуация быстро менялась. В 1988 году я состоял в комитете фондов Макдоннела и Пью, который брал интервью у известных ученых-когнитивистов и нейробиологов, чтобы получить их рекомендации по созданию нового направления – когнитивной нейробиологии[121]. Комитет объездил весь мир, чтобы встретиться с экспертами и узнать, какие научные темы наиболее перспективны и где разместить новые центры когнитивной нейробиологии. Мы встретились в клубе преподавателей Гарварда жарким августовским днем, чтобы провести интервью с Джерри Фодором – экспертом в языке мышления и одним из лучших исследователей в области модульного разума. Он начал с резкого заявления: «Когнитивная нейробиология – это не наука и никогда не будет ею». Сложилось впечатление, что он прочитал все труды по нейробиологии о зрении и памяти и они не соответствовали его стандартам. Фодор продолжил: «Фонд Макдоннела бросает деньги на ветер». Джон Бруэр, президент Фонда Макдоннела, отметил, что Фодор путает его фонд с уличной побирушкой.

Фодор невозмутимо объяснил, почему разум должен быть модульной системой обработки символов под управлением умной компьютерной программы. Патриция Черчленд, философ из

Калифорнийского университета в Сан-Диего, спросила тогда, применима ли его теория к кошкам. «Да, – ответил Фодор. – Кошки управляются кошачьей программой». Морт Мишкин, нейробиолог из Национального института здоровья США, изучающий зрение и память, попросил Фодора рассказать об открытиях, сделанных в его собственной лаборатории. Тот пробормотал что-то об эксперименте, о потенциальных возможностях языка, но я не понял ход его мысли. К счастью, сработала пожарная сигнализация и мы все вышли на улицу. Во дворе я услышал часть разговора Мишкина с Фодором: «… эти картофелины достаточно мелкие». Когда учебная пожарная тревога закончилась, Фодор исчез.

Когнитивная нейробиология превратилась в важную сферу, которая привлекла исследователей из многих областей науки, включая социальную психологию и экономику, которые ранее не были напрямую связаны с ней. Это стало возможным благодаря внедрению новых способов визуализации мозга, и особенно функциональной магнитно-резонансной томографии (фМРТ) – неинвазивного метода, который появился в начале 1990-х годов, а теперь имеет пространственное разрешение в несколько миллиметров. Большие объемы данных, получаемые от фМРТ, анализируют с помощью новых вычислительных методов, таких как независимый компонентный анализ, о чем мы поговорим в следующей главе.

Функциональная МРТ измеряет гемодинамический ответ (изменение кровотока), связанный с активностью нейронов. Мозг не будет работать без кислорода, а кровоток четко регулируется на субмиллиметровом[122] уровне. Степень насыщения крови кислородом изменяет ее магнитные свойства, которые можно бесконтактно измерять с помощью МРТ и использовать для создания динамической картины активности мозга с временны́м разрешением в несколько секунд – достаточно хорошим, чтобы отследить, какие части мозга активны во время эксперимента. Функциональная МРТ

используется для изучения времени прохождения сигнала в различных частях иерархии зрительной системы.

Ури Хэссон из Принстонского университета провел эксперимент с фМРТ, призванный выяснить, какие части иерархии зрительной системы участвуют в обработке видеозаписей различной длительности[123]. Немой фильм Чарли Чаплина был разрезан на кусочки, собран в ролики продолжительностью 4, 12 и 36 секунд и продемонстрирован участникам эксперимента. В 4-секундном ролике зрители успевали распознать сцену, в 20-секундном – последовательность действий, а в 36-секундном – целую историю от начала до конца. Отклики фМРТ в V1 в нижней части иерархии были сильными и устойчивыми, независимо от временной шкалы, но на более высоких уровнях визуальной иерархии только более длинные временные шкалы вызывали стабильный отклик, а области префронтальной коры в верхней части иерархии требовали самого длинного временного интервала. Это согласуется с другими экспериментами, показывающими, что кратковременная память – ваша способность удерживать информацию, такую как телефонные номера и элементы задачи, над которым вы работаете, – также организована в иерархии с самыми длинными временными шкалами кратковременной памяти в префронтальной коре.

Изучать, как происходят процессы обучения в человеческом мозге, – одно из самых увлекательных направлений исследований в нейробиологии, над которым можно работать на разных уровнях, от молекулярного до поведенческого.

Часть II. Множество способов обучения: хронология

1949 – Дональд Хебб выпустил книгу «Организация поведения»[124], в которой сформулировал правило пластичности синапса.

1982 – Джон Хопфилд опубликовал труд «Нейронные сети и физические системы с возникающими коллективными вычислительными способностями»[125], в котором описал нейросеть Хопфилда.

1985 – Джеффри Хинтон и Терри Сейновски представили «Алгоритм обучения для машин Больцмана»[126], что стало контрдоказательством широко распространенного мнения Минского и Пейперта, что алгоритм обучения для многослойных сетей невозможен.

1986 – Дэвид Румельхарт и Джеффри Хинтон написали «Обучение внутреннего представления путем распространения ошибки»[127], где описали алгоритм обратного распространения ошибки, который используется для глубокого обучения в наши дни.

1988 – Ричард Саттон напечатал статью «Обучение прогнозированию методами временных различий»[128] в журнале «Машинное обучение». Он был вдохновлен сутью ассоциативного обучения, и обучение с учетом временной разности стало считаться основным алгоритмом для обучения мозга методом вознаграждения.

1955 – Тони Белл и Терри Сейновски опубликовали труд «Подход к максимизации информации для слепого разделения и слепой обратной свертки»[129], в котором описали неконтролируемый алгоритм для анализа независимых компонентов.

2013 – Работа Джеффри Хинтона «Классификация Image Net с глубокими сверточными нейросетями»[130] позволила на 18 % снизить частоту ошибок при классификации объектов на изображениях.

2017 – сеть глубокого обучения Alpha Go победила Кэ Цзе на Чемпионате мира по го.

Глава 6. Проблема коктейльной вечеринки

На коктейльной вечеринке бывает сложно расслышать, что говорит человек рядом с тобой, среди какофонии других голосов вокруг. Наличие пары ушей помогает направить ваш слух в нужном направлении, и ваша память может заполнить недостающие обрывки разговора. Теперь вообразите шумную вечеринку с сотней людей в комнате и сотней ненаправленных микрофонов, которые собирают звуки ото всех, но с различным соотношением амплитуд для каждого человека на каждом микрофоне. Можно ли разработать алгоритм, который сумеет разделить голоса на отдельные выходные каналы? Чтобы усложнить задание, подумайте, что делать, если источники звука неизвестны – например, музыка, хлопки, звуки природы или даже случайный шум? Это называется проблемой слепого разделения сигналов (рис. 6.1).

На конференции, посвященной нейронным вычислительным сетям, – предшественнице NIPS, – проходившей с 13 по 16 апреля 1986 года в Сноуберде, в штате Юта, был представлен стендовый доклад «Пространственная или временная адаптивная обработка сигналов с помощью моделей нейронных сетей»[131]. Алгоритм обучения был использован для слепого разделения смесей синусоидальных волн – чистых частот, представленных в модели нейронной сети. Было неизвестно, существует ли общее решение, которое могло бы слепо разделять другие типы сигналов. Этот доклад указал на новый класс алгоритмов обучения без учителя. Десять лет спустя мы нашли алгоритм, который мог бы решить общую задачу.

Независимый компонентный анализ[132]

Перцептрон – это однонейронная сеть. В следующей простейшей сетевой архитектуре больше одного модельного нейрона в

выходном слое, при этом каждый входной нейрон соединен с каждым выходным нейроном, что преобразует схему и на входном и на выходном слое. Эта сеть может сделать гораздо больше, чем просто классифицировать входы. Ее можно научить выполнять слепое разделение источников.

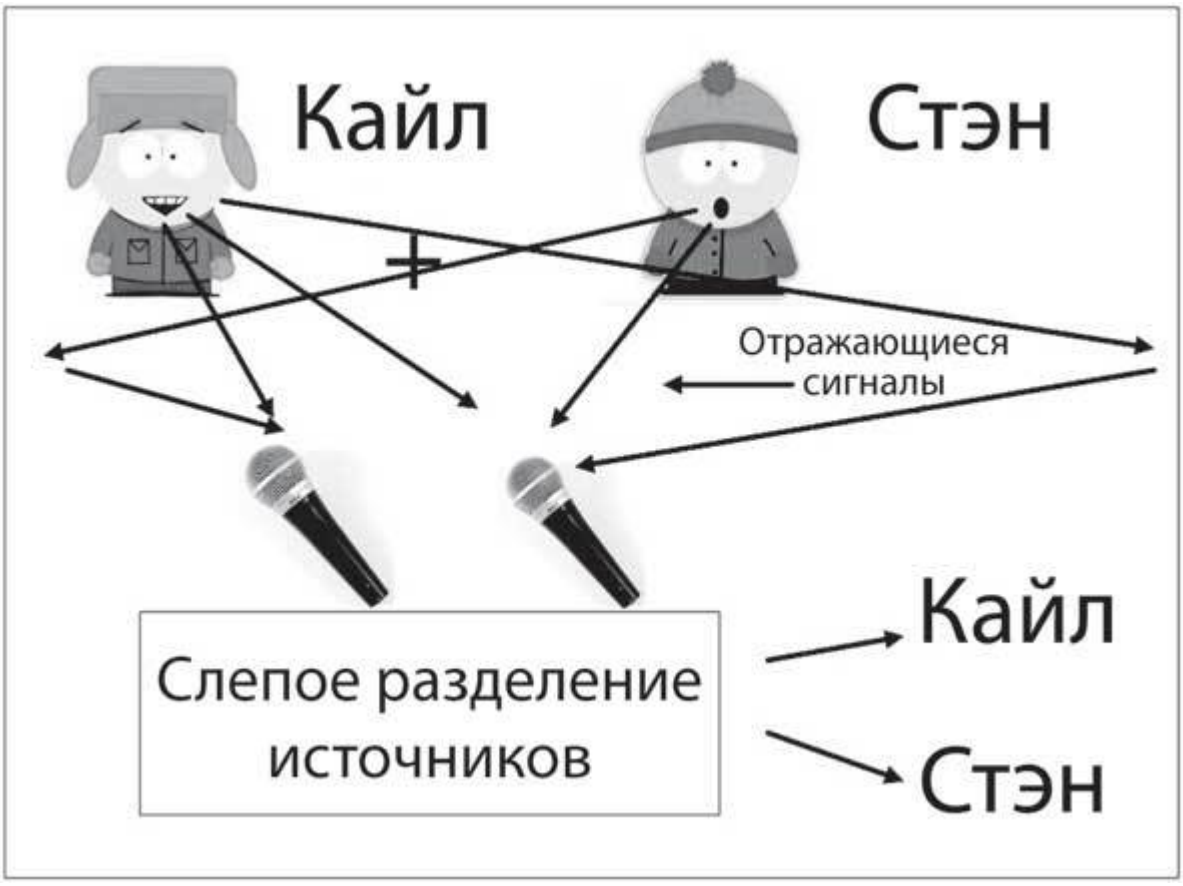


Рис. 6.1. Слепое разделение источников. Кайл и Стэн разговаривают одновременно в комнате с двумя микрофонами. Каждый микрофон улавливает сигналы, исходящие от разговаривающих, а также отражающиеся от стен помещения. Задача состоит в том, чтобы отделить два голоса друг от друга, не зная ничего о сигналах. Независимый компонентный анализ – алгоритм обучения, который решает эту проблему без информации об источниках

В 1986 году тогда еще студент Тони Белл (рис. 6.2) проходил летнюю практику в Швейцарской высшей технической школе в Цюрихе. Он был одним из первых, кто занялся нейронными сетями, и потому отправился в Женевский университет, чтобы послушать выступления четырех известных специалистов по сетям. После защиты диссертации в Брюссельском университете он в 1993 году переехал в Ла-Хойя, чтобы присоединиться к моей лаборатории в качестве постдокторанта.

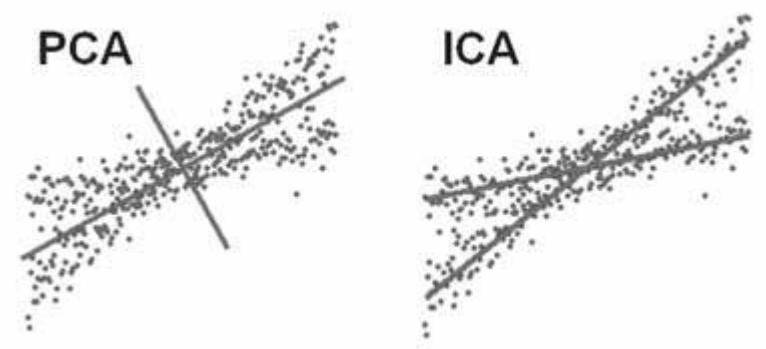
Общий принцип обучения Infomax максимизирует поток информации в сети[133]. Тони Белл работал над передачей сигнала в дендритах – разветвленных отростках нейронов, которые те используют для сбора информации из тысяч соединенных с ними синапсов. Белл чувствовал, что должна быть возможность максимизировать информацию, исходящую из дендрита, если изменить плотность его ионных каналов. Упростив задачу (игнорируя дендриты), он нашел новый теоретико-информационный алгоритм обучения, который решил проблему слепого разделения источников (блок 2)[134].

Рис. 6.2. Тони Белл мыслит независимо. Эксперты знают много способов, которыми нельзя решить задачу, но часто кто-то, кто смотрит на нее впервые, видит новый подход и находит решение. Тони открыл итерационный (пошаговый) алгоритм для решения проблемы слепого разделения источников, который сейчас описан в учебниках по программированию и многократно применен на практике

Независимый компонентный анализ (Independent Component Analysis; ICA) – такое название получил новый алгоритм – с тех пор был использован для тысяч приложений и включен в учебники по обработке данных[135]. При применении к изображениям природы независимые компоненты были вычленены фильтрами определения краев (рис. 6.3), похожими на простые клетки зрительной коры

кошек и обезьян (рис. 6.4)[136]. Более того, чтобы восстановить часть изображения, требуются лишь некоторые из многочисленных источников, отчего реконструкция становится разреженной[137].

Блок 2. Как работает ICA



Сравнение метода анализа главных компонент (Principal Component Analysis; PCA) и анализа независимых компонент. Выходы с двух микрофонов на рисунке располагаются друг против друга по вертикальной и горизонтальной осям. Координаты каждой синей точки – это значения в один момент времени. Анализ главных компонент – популярный неконтролируемый метод обучения, подразумевающий выбор направления, которое делит два сигнала пополам, максимально смешивая их, а оси PCA всегда перпендикулярны друг другу. ICA находит красные оси, которые проходят вдоль направлений точек, представляющих разделенные сигналы, которые могут быть неперпендикулярны.

Эти результаты подтвердили гипотезу, выдвинутую в 1960-х годах выдающимся ученым в области зрения Хорасом Барлоу, когда Хьюбел и Визель обнаружили простые клетки в зрительной коре. На изображении много лишней информации, так как близлежащие

пиксели часто имеют одинаковые значения (например, пиксели неба). Барлоу предположил, что простые клетки смогут уменьшить объем избыточной информации на представленных изображениях природы[138]. Снижая избыточность, можно передать информацию с изображения более эффективно. Потребовалось 50 лет, чтобы разработать математический инструментарий, подтвердивший его идею.

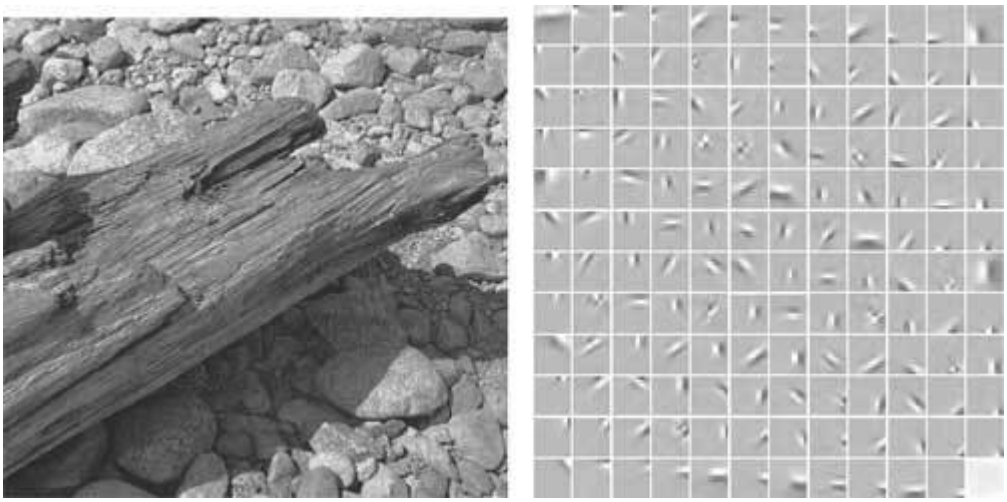


Рис. 6.3. Фильтры ICA, полученные из изображений природы. Небольшие области (12×12 пикселей) снимков природных объектов (слева) использовались в качестве входных данных для сети ICA со 144 выходными блоками. Полученные независимые компоненты (справа) напоминают простые клетки первичной зрительной коры: они локализованы и распределены на положительные (белые) и отрицательные (черные) области, где серый равен нулю. Требуется только несколько фильтров, чтобы отобразить ту или иную часть. Это свойство называется разреженностью. (Bell A. J., Sejnowski T. J. The 'Independent Components' of natural scenes are edge filters, Vision Research, 37, 33273338, 1997)

Мы с Тони также показали, что, когда ICA применяется к естественным звукам, независимые компоненты являются временны́ми фильтрами с различными частотами и длительностью, похожими на фильтры, обнаруженные на низших уровнях слуховой системы[139]. Это дало нам уверенность, что мы находимся на правильном пути и начинаем понимать фундаментальные принципы того, как сенсорные сигналы представлены на самых ранних этапах обработки в зрительной коре. Распространяя принцип на независимые подпространства признаков линейных фильтров, можно было моделировать сложные клетки зрительной коры[140].

Сеть ICA имеет равное количество входных и выходных блоков и один набор весов между ними. Звуки с микрофонов воспроизводятся через входной слой, один входной блок для каждого микрофона, а алгоритм обучения ICA, подобно алгоритму перцептрона, многократно изменяет вес выходного слоя, пока они не сойдутся. Однако, в отличие от перцептрона – контролируемого алгоритма обучения, независимый компонентный анализ не знает, какой должна быть выходная цель. ICA – алгоритм обучения без учителя, который использует меру независимости между выходными единицами в качестве функции потерь[141]. Поскольку веса изменяются, чтобы сделать выходы как можно более независимыми, исходные источники звука становятся совершенно разделенными или как можно более невзаимосвязанными. Неконтролируемое обучение может обнаружить статистическую структуру в данных различного типа и объема, которые были ранее неизвестны.

Независимый компонент в мозге

Infomax, алгоритм независимого компонентного анализа, разработанный Тони Беллом, вызвал череду открытий по мере того, как сотрудники моей лаборатории стали применять его к различным типам записей мозговой деятельности.

Рис. 6.4. Независимый компонентный анализ применили к записям ЭЭГ. Слева: карта волосистой части головы (вид сверху, нос на изображении находится вверху) с электродом, расположенным в черных точках с цветовой картой напряжения в микровольтах в каждый момент времени. Колеблющиеся сигналы ЭЭГ, которые регистрируются пятью каналами на коже головы, загрязнены искажениями от мигания глаз и мышечных спазмов. Справа: ICA отделяет компоненты мозга от помех, как показано на панели

справа. НК1 – моргание глаз, основанное на медленном течении времени и карте кожи головы, которая имеет самые высокие значения (красный) над глазами. НК4 – подергивание мышц, основанное на высокочастотном шуме высокой амплитуды и локализованное в источнике на карте скальпа. НК2 и НК3 – источники сигналов мозга, на что указывает дипольный узор на коже головы (одиночная красная зона), сравненные со сложной картиной на коже головы от записей ЭЭГ, как показано на карте слева

Впервые электрические сигналы мозга были записаны с поверхности кожи головы в 1924 году Хансом Бергером. Нейробиологи использовали сложные колеблющиеся сигналы, названные электроэнцефалограммой (ЭЭГ), чтобы отслеживать состояние мозга, непрерывно меняющееся в зависимости от степени вашей сосредоточенности и двигательной активности. Электрический сигнал на электроде на коже головы получает входные сигналы от множества различных источников в коре головного мозга, а также сокращений мышц и движений глаз. На каждый электрод поступает смесь сигналов от одного и того же набора источников в головном мозге, но с разной амплитудой, что формально совпадает с проблемой коктейльной вечеринки.

Скотт Макейг, когда в 1990-х годах был научным сотрудником моей лаборатории в Институте Солка, использовал ICA для извлечения из записей ЭЭГ десятков дипольных источников в коре головного мозга и временны́х курсов для каждого из них (рис. 6.4). Диполь[142] – одна из простейших схем источников головного мозга. Самая простая – равномерный узор на коже головы, созданный статическим точечным зарядом. Второй по простоте – дипольный узор от тока, движущегося по прямой линии, которая возникает в корковых пирамидных нейронах. Диполь можно представить как стрелу: поверхность волосистой части головы положительна в направлении наконечника стрелы и отрицательна в направлении

оперения; узор покрывает всю голову, поэтому так трудно отделить многие источники мозга, которые активируются одновременно. НК2 и НК3 – примеры дипольных источников на рис. 6.4. ICA также отделяет помехи, такие как движения глаз и шумы электрода, которые затем могут быть высчитаны с высокой точностью (НК1 и НК4 на рис. 6.4). С тех пор была опубликована не одна тысяча статей о применении ICA для расшифровки ЭЭГ, и многие открытия были сделаны с использованием ICA для анализа широкого спектра состояний мозга.

Мартин Маккаун, который тогда был докторантом в моей лаборатории и имел опыт работы в нейробиологии, выяснил, как перевернуть пространство и время, чтобы применить ICA к данным фМРТ (рис. 6.5)[143]. Для визуализации мозга с помощью фМРТ измеряется уровень насыщенности крови кислородом, косвенно связанный с нейронной активностью, в десятках тысяч мест в мозге. При этом источники ICA – области мозга, у которых общая временная динамика, но которые пространственно независимы от других источников и, как правило, пространственно разрежены (рис. 6.5).

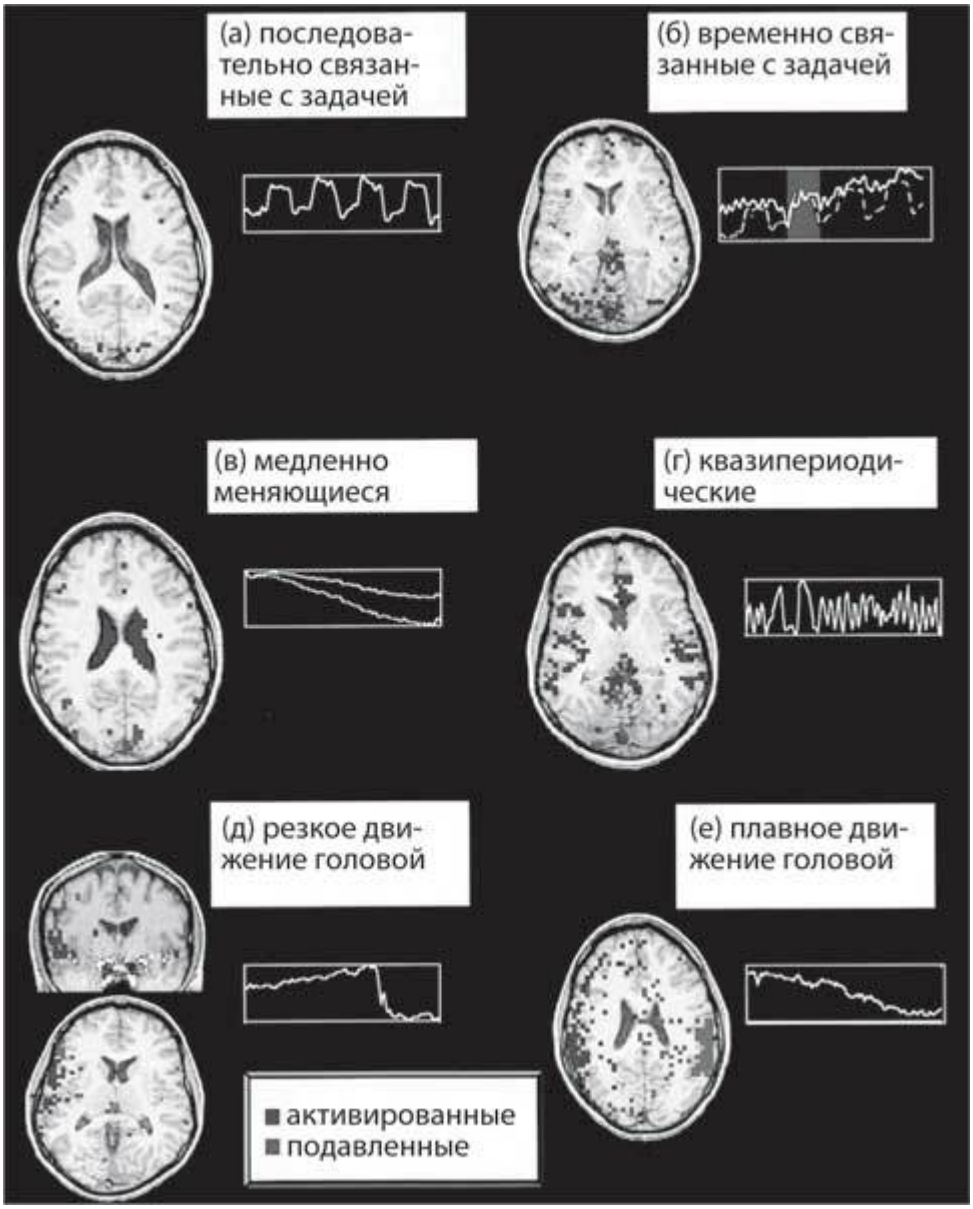


Рис. 6.5. ICA, примененный к данным фМРТ. Компонент состоит из карты активности мозга и временного периода. Проиллюстрированы несколько типов компонентов. Задача представляет собой визуальный стимул длительностью 5 секунд, который улавливается связанными с задачей компонентами. Время прохождения сигналов составляет около минуты, а задача повторяется 4 раза, как на панели (а). Другие компоненты собирают импульсы, такие как движения головы

Поскольку ICA неконтролируемый, он может выявить сети областей мозга, которые работают вместе, что выходит за рамки контролируемых методов, пытающихся связать активность в области с сенсорным стимулом или моторной реакцией. Например, при помощи ICA в записях фМРТ выявили многочисленные варианты состояния покоя, когда испытуемых просили просто находиться в аппарате и отдыхать[144]. Мы до сих пор не понимаем, что означают эти состояния покоя, но они могут представлять сочетания областей мозга, которые отвечают за происходящее в нем, когда мы витаем в облаках, думаем о беспокоящей нас проблеме или планируем ужин.

Принцип максимальной независимости связан с принципами разреженного кодирования. Хотя ICA находит много независимых компонентов, только некоторые из них необходимы для восстановления определенной части на снимке природных объектов. Этот принцип также применим к зрительной коре, в которой клеток в сотню раз больше, чем в сетчатке: в сетчатке одного глаза миллион ганглиозных клеток, в то время как в первичной зрительной коре – 100 миллионов нейронов, а она лишь первый из многих слоев зрительной иерархии в коре головного мозга. Компактное кодирование визуальных сигналов в сетчатке расширяется в коре до нового кода, сильно распределенного и высокоразреженного. Расширение в пространстве гораздо больших размеров используется в других схемах кодирования, в том числе в слуховой и обонятельной коре мозга, а новый класс алгоритмов, называемый compressed sensing, или сжатым зондированием, обобщил принцип разреженности для повышения эффективности хранения и анализа сложных наборов данных[145].

За рамками ICA

История ICA иллюстрирует важность технических решений для научных открытий и проектирования. Обычно мы думаем о них как об устройствах, например, микроскопах и усилителях. Но технические решения также являются методами, и они могут позволить делать новые открытия с информацией, полученной с помощью старых инструментов. Запись ЭЭГ существует уже почти 100 лет, но без ICA было невозможно определить основные источники мозга. Мозг – система взаимосвязанных алгоритмов, и я не удивлюсь, если в какой-то части мозга природа обнаружит способ реализации ICA[146].

В 1990-х годах были и другие достижения в разработке новых алгоритмов обучения для нейронных сетей, многие из которых, как и ICA, теперь часть математического инструментария в машинном обучении. Эти алгоритмы встроены в популярные устройства, которые не позиционируются как использующие нейросети. Например, Ли Тэ Вон и Цзыпин Жун, бывшие сотрудники моей лаборатории, основали компанию SoftMax, где использовали ICA в гарнитуре Bluetooth с двумя микрофонами для подавления фонового шума. Это позволило пользователю слышать, что говорят, даже в столовой или на спортивном мероприятии. В 2007 году компанию SoftMax купила компания Qualcomm, которая разрабатывает микросхемы для сотовых телефонов, и сегодня алгоритмы, подобные ICA, встроены в миллиард мобильных аппаратов. Если бы нам дали по центу за каждый телефон с ICA, мы бы порядком разбогатели.

Тони Белл долгие годы занимался еще более сложной проблемой. В биологии есть множество сетей, и информация поступает от одного сетевого уровня к другому, от молекул к синапсам, нейронам, нейронным популяциям и дальше, к принятию решения. Все это объясняют законы физики и химии. Но у нас складывается

впечатление, что ситуацию контролируем мы, а не физика. Как внутренняя активность, возникающая в нейронных популяциях вашего мозга, приводит вас к принятию решения, например, почитать определенную книгу или поиграть в теннис, остается загадкой. Эти решения принимаются значительно ниже уровня вашего сознания. Каким-то образом они возникают из взаимодействия между нейронами, обменивающимися информацией через синапсы, сформированными опытом, который основан на молекулярных механизмах. Но с вашей точки зрения, именно ваше решение вызвало все эти события в мозгу, в противоположном от физики направлении: при самонаблюдении причина и следствие меняются местами. Как примирить эти две позиции – глубокий философский вопрос[147].

Глава 7. Нейронная сеть Хопфилда и машина Больцмана

Джером Фельдман – специалист по вычислениям, который применил нейросетевой подход к ИИ в 1980-х годах во время своей работы в Рочестерском университете. Фельдман оказался прав, отметив, что алгоритмам, использующимся до этого в искусственном интеллекте, нужно было совершить миллиард шагов, чтобы прийти к заключению, часто ошибочному, в то время как мозгу нужно всего сто шагов, чтобы принять решение, зачастую правильное[148]. В то время правило «ста шагов» Фельдмана не было популярно у ученых, занимающихся ИИ, однако некоторые из них, в том числе Аллен Ньюэлл из Университета Карнеги – Меллона, применяли его как ограничение.

Однажды Джером Фельдман спас меня в аэропорту Рочестера. Я получил президентскую премию молодого исследователя от Национального научного фонда (ННФ) США. Я возвращался из поездки в исследовательскую лабораторию General Electric в Скенектади в штате Нью-Йорк. Измученный полетом обратно в Балтимор, я удивился, почему пилот рассказывал нам о погоде в Рочестере. Я сел не на тот самолет. Когда в аэропорту Рочестера я изо всех пытался найти рейс обратно в Балтимор на следующий день, я столкнулся с Джерри, возвращавшимся с заседания комитета по проекту Intenet2 в ННФ в столице. Он любезно пригласил меня переночевать у него. С тех пор Фельдман перешел в Калифорнийский университет в Беркли, и я вспоминаю его всякий раз, когда застреваю в аэропорту.

Джерри различал «чистые» и «грязные» нейросетевые модели. Модели, над которыми работали мы с Джеффри Хинтоном, были «грязными», потому что они распределяли представление объектов и концепций по множеству компонентов сети, в то время как

Фельдман считал, что на одном блоке должна быть одна метка – «чистое» вычислительное представление. В более широком контексте «грязные» модели используют для приближения к правильным ответам, а «чистые» – чтобы найти точное решение проблемы. На самом деле, чтобы добиться успеха, нужны обе модели[149]. У меня не было проблем с получением «грязной» отправной точки, но должно же быть более ясное объяснение, и мы с Хинтоном собирались достичь «чистого» успеха.

Джон Хопфилд

Чтобы получить докторскую степень по физике, вы должны решить задачу, но действительно хороший физик должен уметь решить любую проблему. Великий же физик знает, какую проблему нужно решить. Джон Хопфилд – великий физик. Сделав выдающуюся карьеру в физике конденсированных сред, он обратил свой интерес к биологии и, в частности, к проблеме молекулярной корректуры. Когда ДНК воспроизводит себя во время деления клеток, неизбежны ошибки, которые необходимо исправить, чтобы сохранить точность передачи информации в дочерние клетки. Хопфилд придумал хитрый план, как это сделать за счет энергии. Последующие эксперименты показали, что он был прав. Добиться подобного в биологии – впечатляющее достижение.

Джон Хопфилд был моим научным руководителем в Принстонском университете. Когда я работал с Хопфилдом, он только начал интересоваться нейробиологией. Он был полон энтузиазма и рассказал мне, что узнал на заседаниях Программы исследований в нейробиологии (ПИН), базирующейся в Бостоне, где он слушал лекции других специалистов в данной области. ПИН также опубликовала материалы небольших семинаров, которые были бесценны, так как дали мне представление о том, какие проблемы изучали и какие теории существовали в то время. У меня все еще сохранилась копия семинара по нейронному кодированию,

организованного Тедом Баллоком, легендарным нейроэтологом, который позже стал моим коллегой в Калифорнийском университете в Сан-Диего. Книга Теда Баллока и Адриана Хорриджа о нервной системе беспозвоночных стала классикой[150]. Я работал с Тедом над моделированием поведения коралловых рифов и гордился тем, что был соавтором его последнего научного труда[151].

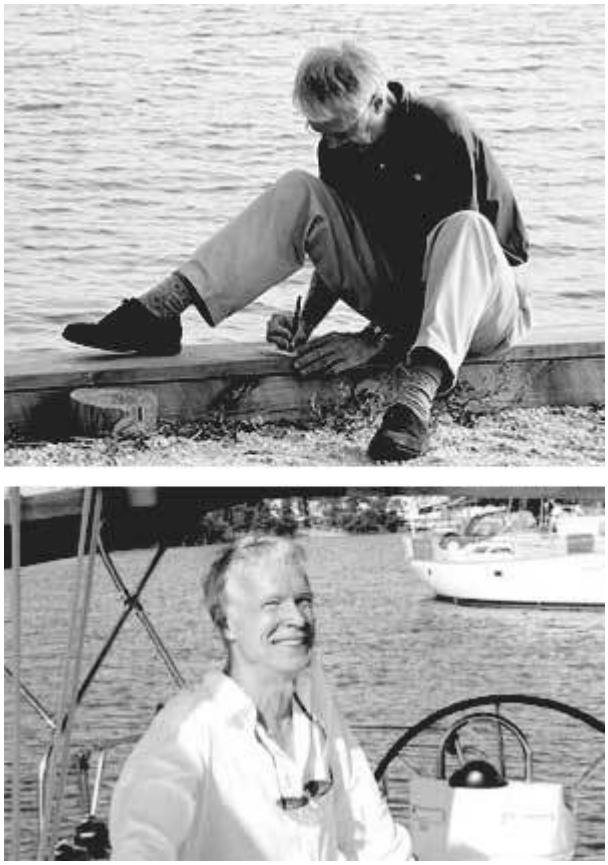


Рис. 7.1. Джон Хопфилд решает задачу на набережной в Вудс-Хоул в штате Массачусетс. В 1980-х годах Хопфилд оказал основополагающее влияние на нейронные сети, изобретя одноименную сеть, которая открыла дверь для глубокого обучения

Нейронные сети с обратными связями с более ранними слоями и циклическими связями между элементами внутри слоя могут иметь гораздо более сложную динамику, чем сети только с прямыми связями. Общий случай сетей с произвольно связанными

элементами с положительными (возбуждающими) и отрицательными (тормозящими) весами – сложная математическая задача. Джек Коуэн из Чикагского университета и Стивен Гроссберг из Бостонского университета ранее изучали такие сети и добились прогресса, показав, что нейросети могут воспроизводить зрительные иллюзии[152] и галлюцинации[153], но проектировать такие сети для решения сложных вычислительных задач было трудно.

Сеть с ассоциативной памятью

Летом 1983 года мы с Джеффри Хинтоном были на организованном Джерри Фельдманом семинаре в Рочестерском университете. Джон Хопфилд (рис. 7.1) также там присутствовал. В Рочестере Хопфилд сказал нам, что решил проблему сходимости для сильно взаимодействующей сети. Сильно нелинейные сети склонны к колебаниям или еще более хаотичному поведению. Он доказал, что определенный тип нелинейной сетевой модели, теперь называемой сетью Хопфилда, гарантированно сходится к стабильному состоянию, называемому аттрактором (рис. 7.2, блок 3) [154]. Кроме того, веса в сети можно выбрать так, чтобы аттракторами были блоки памяти. Таким образом, сеть Хопфилда можно использовать для реализации так называемой ассоциативной памяти. В цифровом компьютере память хранится в ячейках с определенным адресом, но в сети Хопфилда можно получить сохраненную память, обратившись только к ее части, а сеть ее восполнит. Похоже на то, как у нас пробуждаются воспоминания. Если мы видим лицо кого-то, кого мы знаем, мы можем вспомнить его имя и разговоры с этим человеком.

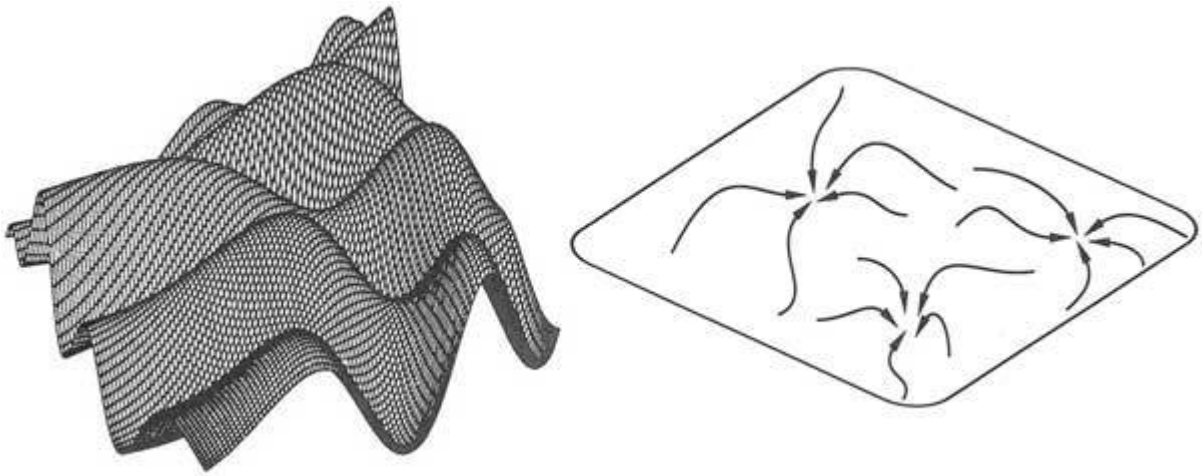


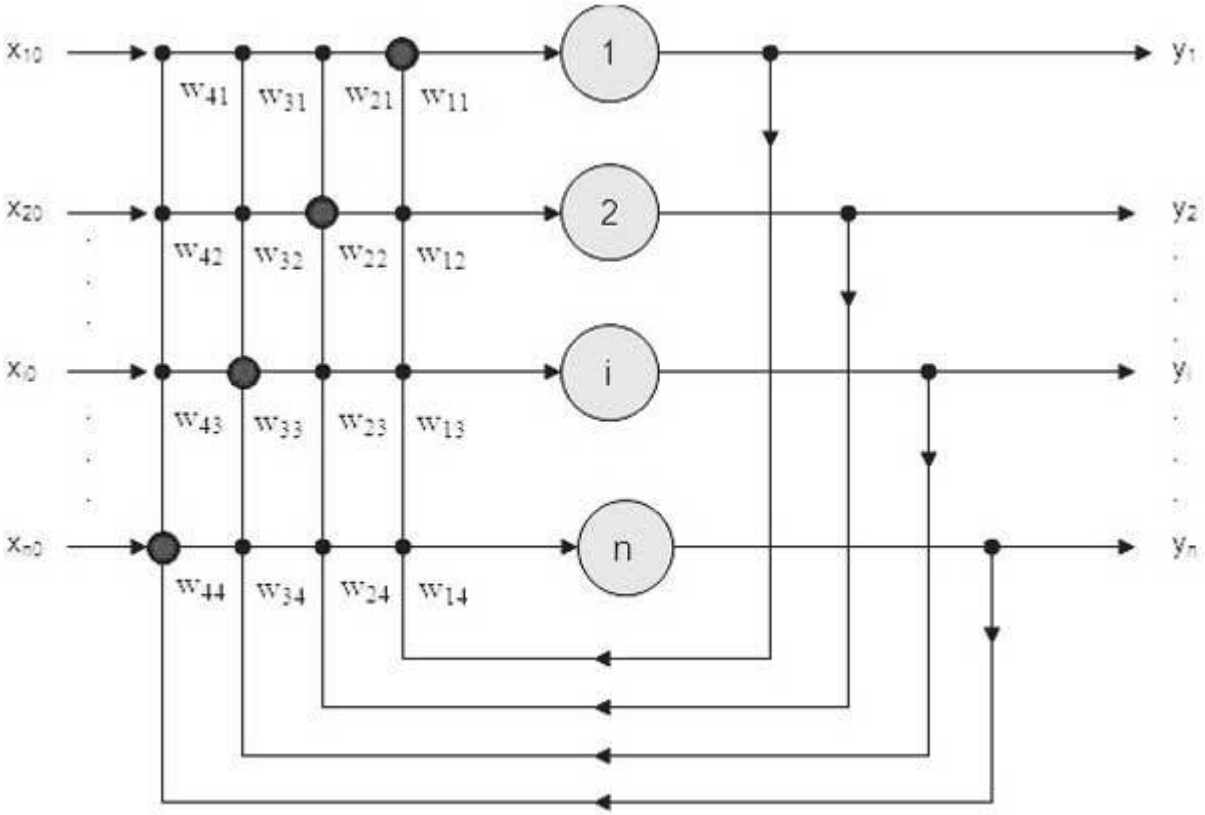
Рис. 7.2. Энергетический ландшафт сети Хопфилда. Состояние сети можно представить в виде точки на энергетической поверхности (слева). Каждое обновление приближает состояние к одному из минимумов энергии, называемых состояниями аттрактора (справа)

Сеть Хопфилда уникальной делает то, что она математически всегда сходится (блок 3). Считалось, что невозможно проанализировать общий случай нелинейной сети, но Хопфилд показал, что частный случай симметричной сети, в которой взаимные связи между парами единиц равны по силе, а единицы обновляются последовательно, разрешим. Когда обновления производятся одновременно для всех узлов в сети, динамика может быть гораздо сложнее, и нет никакой гарантии схождения[155].

Появляется все больше доказательств того, что нейронные сети в гиппокампе – части мозга, которая необходима для хранения долгосрочных воспоминаний о событиях и уникальных объектах, – имеют аттракторные состояния, подобные тем, которые находятся в сети Хопфилда[156]. Хотя модель сильно абстрагирована, ее качественное поведение похоже на то, что наблюдается в гиппокампе. Сети Хопфилда стали мостом от физики к нейробиологии, который многие физики протянули в 1980-х годах. Удивительные открытия были получены при анализе нейронных сетей и алгоритмов обучения с помощью сложных инструментов из теоретической физики. Связи между физикой, вычислениями и обучением глубоки и являются одной из областей нейробиологии,

где теория успешно объясняет функционирование мозга.

Блок 3. Сеть Хопфилда



В сети Хопфилда от каждого блока идет выходной канал ко всем блокам в сети. Входы обозначены как x_i, а выходы – y_j. Сила соединений или веса симметричны: $w_{ij}=w_{ji}$. На каждом временном шаге одна из единиц обновляется путем суммирования входов и сравнения с порогом: если входы превышают порог, выход равен 1, в противном случае – 0. Хопфилд показал, что сеть имеет энергетическую функцию, которая никогда не увеличивается с каждым обновлением объекта в сети:

$E = \Sigma\, w_{ij}\, x_i\, x_j$

В конце концов ни один из блоков не меняется, и функция находится на локальном минимуме. Это называется состоянием аттрактора в сети Хопфилда и соответствует хранимой памяти, которая может быть восстановлена путем инициализации сети с частью сохраненных данных. Так в сети Хопфилда создается ассоциативная память. Вес хранимой информации можно узнать с помощью долговременной синаптической пластичности:

$\Delta w_{ij} = \alpha x_i x_j$

где левая сторона – изменение веса, α – скорость обучения, а x_i – хранимая информация.

Джон Хопфилд и Дэвид Тэнк, работавший в то время в компании Bell Labs[157], показали, что вариант сети Хопфилда, в котором информация постоянно оценивалась между нулем и единицей, можно использовать, чтобы получить хорошие решения для задач по оптимизации, таких как задача коммивояжера, где необходимо найти кратчайший маршрут, посещая указанные города[158]. Это задача по информатике, известная своей сложностью. Энергетическая функция сетей включала длину пути и ограничение на посещение каждого города одним разом. После кратковременного повышения напряжения в начале сеть начинала работать с минимальными затратами энергии, находя хороший, хоть и не обязательно лучший маршрут.

Поиск глобального энергетического минимума

На семинаре также присутствовал Дана Гарри Баллард, написавший классическую книгу по компьютерному зрению. Мы с Джеффри работали с Баллардом над обзорной статьей для журнала Nature о новом подходе к анализу изображений[159]. Идея заключалась в том, что узлы в сетевой модели исполняют роль информационной функции на изображении, а соединения в сети разграничивают объекты; у согласованных узлов – положительное взаимодействие друг с другом, а у несогласованных –

отрицательное. В поле зрения необходимо последовательно проанализировать все признаки, подходящие под заданные ограничения.

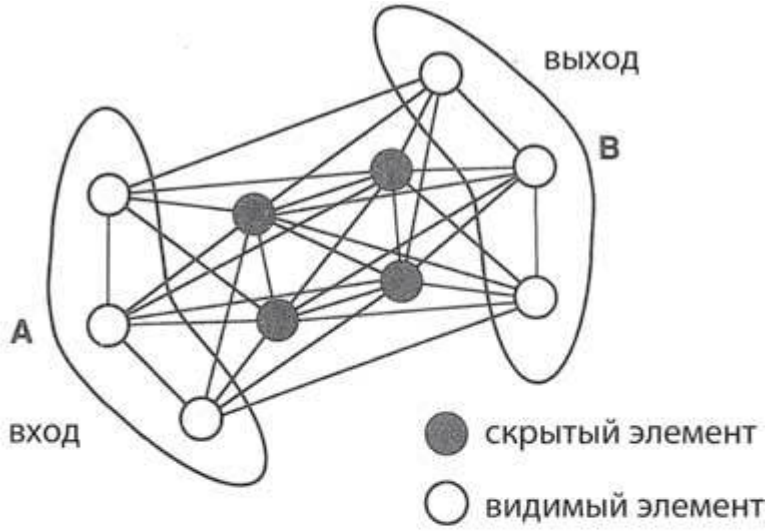
Может ли сеть Хопфилда решить эту проблему, удовлетворив всем ограничениям? Энергетическая функция – мера того, насколько хорошо сеть их удовлетворяет. Проблема зрения требовала решения с глобальным энергетическим минимумом, лучшего решения, но сеть Хопфилда изначально спроектирована находить только локальные минимумы энергии. Недавно в журнале Science я наткнулся на статью Скотта Киркпатрика из Исследовательского центра Томаса Уотсона в IBM, которая могла помочь с этой задачей[160]. Киркпатрик использовал так называемый метод имитации отжига, чтобы обойти локальные минимумы. Предположим, у вас есть множество компонентов в электрической цепи, которые должны быть установлены на две печатные платы. Каково наилучшее размещение деталей, чтобы использовать наименьшее количество проводов, необходимых для их подключения?

Найти очень хорошие решения можно, сначала расположив части в случайном порядке, а затем перемещая их назад и вперед по одной, чтобы увидеть, какое размещение требует меньше проводов. Проблема в том, что вы рискуете легко попасть в ловушку локального минимума, когда не добиться никаких улучшений, перемещая один компонент. Способ этого избежать – позволить случайные скачки к конфигурациям с более длинными проводами. В начале вероятность выскочить высока, но постепенно уменьшается, так что к концу она равна нулю. Если снижение вероятности достаточно медленное, окончательное размещение деталей будет иметь глобальный минимум соединительных проводов. В металлургии это называется отжигом: при нагревании металла и медленном охлаждении образуются крупные кристаллы с минимальными дефектами. Дефекты делают металл хрупким и склонным к растрескиванию.

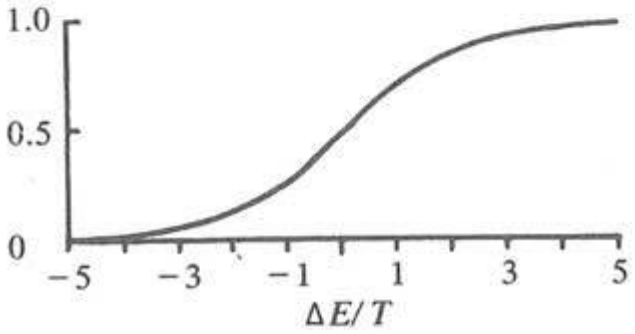
Машины Больцмана

В сети Хопфилда имитация отжига соответствует «нагреву» обновлений, так что энергия может идти как вверх, так и вниз. При высокой температуре блоки произвольно переворачиваются, и если температура постепенно понижается, то высока вероятность того, что сеть Хопфилда застынет в состоянии с наименьшей энергией, когда температура достигнет нуля. На практике моделирование начинается при постоянной температуре, чтобы обеспечить равновесие сети. При равновесии сеть может побывать во множестве ближайших состояний и исследовать широкий спектр допустимых решений.

Блок 4. Машина Больцмана



двоичные элементы машины Больцмана



Все соединения в машине Больцмана симметричны, как и в сети Хопфилда, и двоичные единицы обновляются один раз, устанавливая $s_i = 1$ с вероятностью, заданной приведенной выше S-образной кривой, где входы Δ E масштабируются по температуре T. Входной и выходной слои «видимы» в том смысле, что они взаимодействуют с внешним миром. «Скрытые элементы» представляют объекты, имеющие внутренние степени свободы, которые могут влиять на видимые объекты. У алгоритма машинного обучения Больцмана две фазы: в фазе бодрствования входы и выходы фиксируются, и после того, как сеть приходит к равновесию, вычисляется средняя корреляция между парами единиц. Во второй фазе сна корреляции снова вычисляются с незафиксированными входами и выходами. Затем вес постепенно обновляется:

$$\Delta w_{ij} = \varepsilon(\langle s_i s_j \rangle^{\text{бодрствование}} - \langle s_i s_j \rangle^{\text{сон}}).$$

Для примера рассмотрим задачу, где требуется определить, какая часть изображения является фигурой, а какая – фоном. Фигура на рис. 7.3 неоднозначна, и в зависимости от того, на какую часть вы обратите внимание, вы увидите вазу или два лица, но не то и другое сразу. Мы разработали машину Больцмана (блок 4), которая моделирует выбор между фигурой и фоном[161]. Она состояла из элементов, часть из которых в состоянии активности представляли фигуру, а другие – контур. Мы уже видели, что в зрительной коре есть простые клетки, которые активируются контурами, однако фигура может находиться по обе стороны контура. Проблему решили с помощью двух блоков, каждый из которых «видел» границу фигуры со своей стороны. Такие нейроны были впоследствии обнаружены и в зрительной коре, они называются пограничными клетками[162].

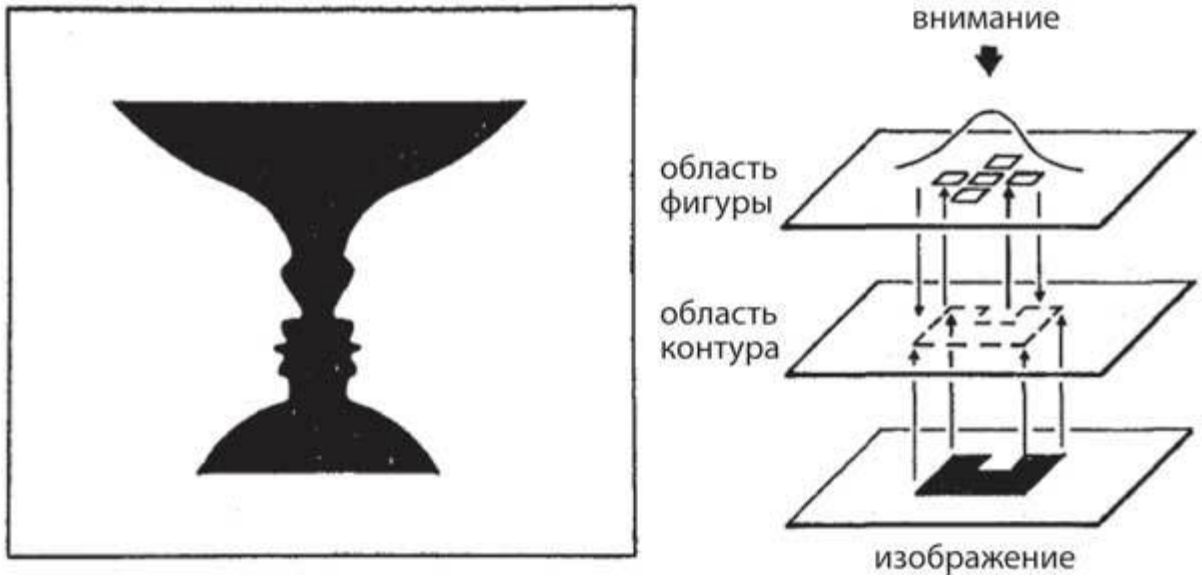


Рис. 7.3. Сетевая модель «фигура-фон». Слева: когда вы фокусируете свое внимание на черном изображении, вы видите вазу

на белом фоне. Но когда вы сосредотачиваетесь на белой области, вы видите два лица, смотрящие друг на друга. Вы можете переводить свое внимание с одного изображения на другое, но не можете видеть две картинки одновременно. Справа: два типа элементов, показывающих края объекта и то, является ли пиксель частью фигуры или фона. Входы изображения читаются снизу вверх, а входы внимания – сверху вниз. Внимание реализовано смещением к той области, которая должна быть воспринята в качестве рисунка

Веса в сети были созданы вручную для реализации ограничений (рис. 7.4). Между компонентами фигуры есть возбуждающие связи, между компонентами контура – тормозящие. У пограничных блоков возбуждающие связи с элементами фигуры, на которые они указывают, очерчивая ее, и тормозящие связи с элементами в противоположном направлении. Внимание смоделировано смещением к некоторым элементам фигуры.

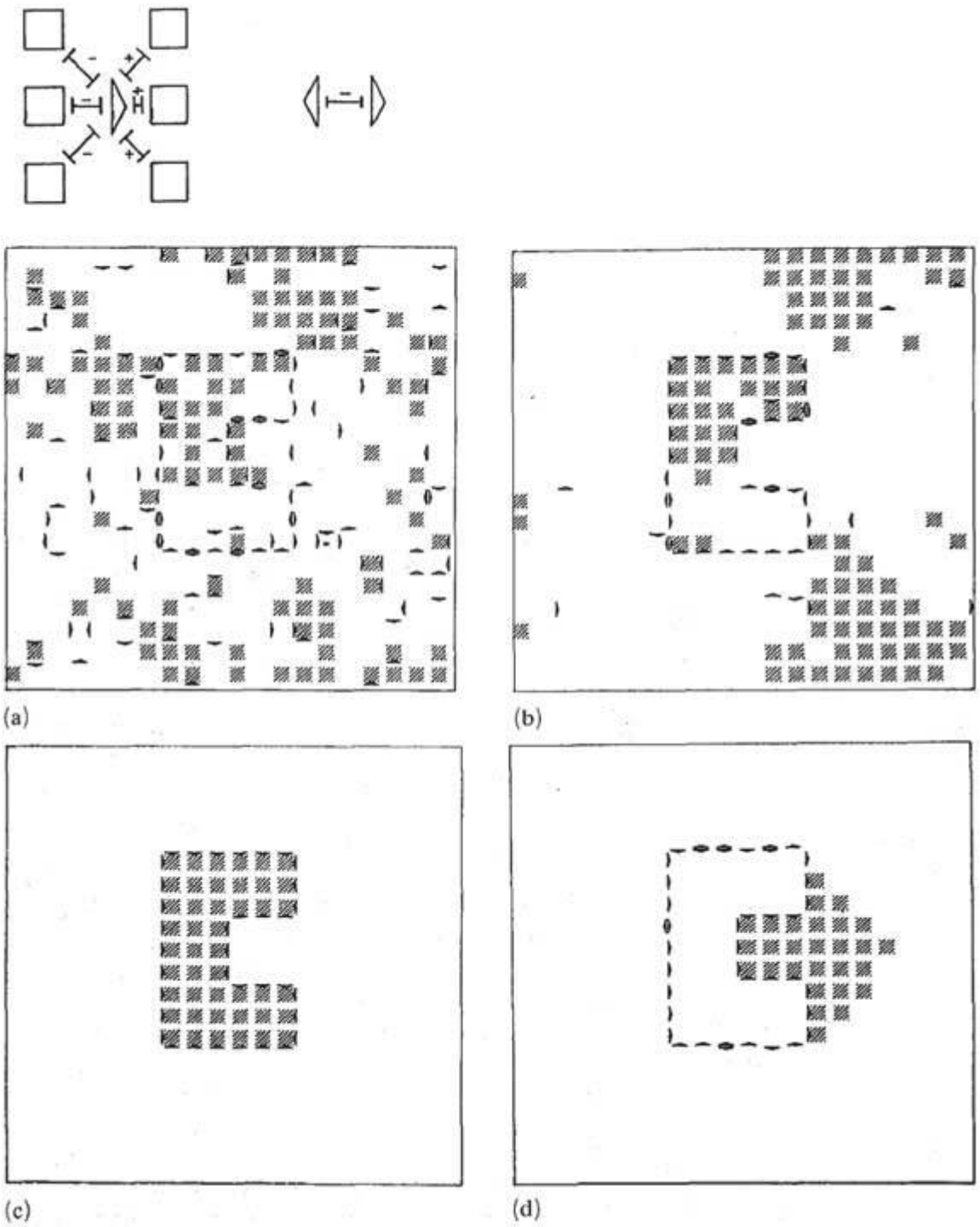


Рис. 7.4. Машина «фигура – фон» Больцмана. Слева внизу: в сети квадраты обозначают фигуру в виде буквы «С», а треугольники – ее контур. Вершина треугольника может быть обращена как к фигуре, так и от нее. Вверху: а) внимание сосредоточено на внутренней стороне «С»; температура была высокой, так что блоки то

включались, то выключались; б) по мере падения температуры, блоки на внутренней стороне «С» начинали объединяться при поддержке блоков контура, указывающих внутрь. Блоки, указывающие наружу, не притягивают внимание, и края исчезают по мере того, как температура уменьшается; в) рисунок заполнен, когда внимание обращено внутрь контура; г) фон заполнен, когда внимание обращено наружу

Когда сеть использует для единиц правило обновления Хопфилда, она попадают в локальные энергетические минимумы, которые согласованы в отдельных частях, но не согласованы в глобальных масштабах. Когда к обновлениям добавляли шум, сеть выскакивала за пределы локального минимума, и медленное повышение температуры шума приводило к последовательным решениям на глобальном энергетическом минимуме (рис. 7.4). Поскольку обновления асинхронные и независимые, сеть может быть реализована на многопроцессорном компьютере с массово-параллельной архитектурой и приходить к решениям гораздо быстрее, чем цифровой компьютер, который работает последовательно, выполняя одну операцию за раз.

К этому времени я закончил стажировку в Гарвардской медицинской школе у Штефана Куффлера и получил свою первую работу на кафедре биофизики в Университете Хопкинса. Джеффри Хинтон занял должность преподавателя на факультете компьютерных наук в Университете Карнеги – Меллона. Ему посчастливилось заручиться поддержкой Аллена Ньюэлла, который был открыт для всего нового в развитии ИИ. Питтсбург и Балтимор расположены достаточно близко, и мы могли навещать друг друга по выходным. Мы назвали нашу новую версию сети Хопфилда машиной Больцмана по имени Людвига Больцмана, физика XIX века и основоположника статистической механики – источника инструментов, которые мы использовали для анализа нашей

изменяющейся модели нейросети. Вскоре мы обнаружили, что это также мощная обучающаяся машина.

При постоянной «температуре» машина Больцмана приходит в равновесие. В состоянии равновесия происходит нечто невероятное, открывающее дверь, которая, как все считали, останется закрытой навсегда: многослойное обучение нейронной сети. Однажды мне позвонил Джеффри Хинтон и сказал, что только что вывел простой алгоритм обучения для машины Больцмана. Целью обучающего алгоритма было отобразить блоки ввода в блоках вывода, но, в отличие от перцептрона, между ними были блоки, называемые скрытыми элементами (см. блок 4). Представляя пары ввод – вывод и применяя алгоритм обучения, сеть изучила требуемое преобразование. Но простое запоминание пар не было целью; она состояла в том, чтобы правильно классифицировать новые входы, которые не использовались для обучения сети. Кроме того, поскольку машина Больцмана всегда «колеблется», это позволило изучить распределение вероятностей – как часто данный входной шаблон обращается к каждому из состояний вывода. Последнее делает машину Больцмана производящей: после обучения она может создавать новые входные выборки для каждой выходной категории.

Теория синаптической пластичности Хебба

Неожиданно выяснилось, что алгоритм обучения машины Больцмана имеет долгую историю в нейробиологии, начинающуюся с психолога Дональда Хебба, который в книге «Организации поведения»[163] постулировал, что, когда два нейрона срабатывают одновременно, связь между ними должна усиливаться:

«Давайте предположим, что постоянная или повторяющаяся отражательная активность («след») ведет к длительным клеточным изменениям, которые усиливают стабильность. Когда аксон клетки А находится достаточно близко, чтобы возбудить клетку В, и неоднократно или постоянно принимает участие в ее возбуждении, в одной или обеих клетках происходит некий процесс роста или метаболических изменений, так что эффективность клетки А, возбуждаемой клеткой В, увеличивается».

Возможно, это самое известное предсказание во всей нейробиологии. Позже синаптическая пластичность была обнаружена в гиппокампе – важной для долговременной памяти области мозга. Когда пирамидальная клетка гиппокампа получает сильный входной сигнал одновременно с возбуждением нейрона, сила синапсов увеличивается. Последующие эксперименты показали, что усиление основано на сочетании высвободившегося из синапса нейромедиатора и повышения напряжения в нейроне-реципиенте.

Более того, это соединительное явление было распознано особым глутаматным рецептором NMDA, который вызывает долговременную потенциацию (усиление) синаптической передачи. ДП возникает быстро и длится долго, что создает хорошую почву для долгосрочной памяти. Пластичность Хебба в синапсе определяется совпадениями между входами и выходами, как и в алгоритме машинного обучения Больцмана.

Еще удивительнее то, что машине Больцмана требовалось заснуть, чтобы научиться! Алгоритм обучения состоял из двух этапов. На первом, когда входы и выходы привязаны к желаемому изображению, блоки в сети многократно обновлялись, чтобы прийти к равновесию, и подсчитывалось, сколько каждая пара блоков работала одновременно. Мы назвали это фазой пробуждения. На втором этапе входные и выходные блоки были освобождены, и отрезок времени, в течение которого каждая пара блоков работала вместе, был подсчитан в независимом режиме. Мы назвали это фазой сна. Затем сила каждого соединения обновлялась пропорционально разнице между частотой совпадения в фазах бодрствования и сна (см. блок 4).

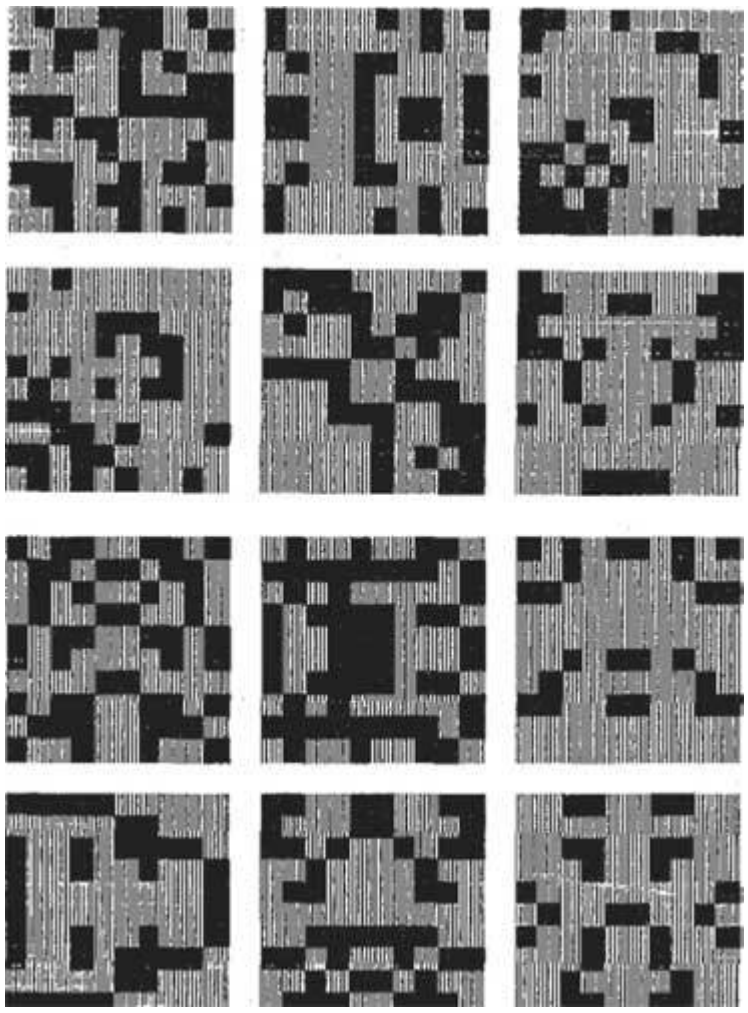


Рис. 7.5. Симметричное неупорядоченное растровое изображение. У каждой сетки 10×10 есть вертикальная, горизонтальная или диагональная ось зеркальной симметрии. Цель сетевой модели – научиться определять ось симметрии на новых рисунках, которые еще не использовали для обучения сетевой модели

Фаза сна у машины необходима, чтобы определить, какая часть зафиксированных взаимосвязей вызвана внешними причинами. Не отбрасывая внутренние взаимосвязи, сеть укрепит внутренние модели деятельности и научится игнорировать внешнее влияние – сетевую версию индуцированного психоза[164]. Интересно, что у людей экстремальное недосыпание приводит к бредовым состояниям – распространенной проблеме в больницах в отделениях интенсивной терапии, где нет окон и постоянно горит свет. Пациенты с шизофренией часто страдают нарушениями сна,

которые могут усиливать спутанность их сознания. Мы были убеждены, что находимся на правильном пути к пониманию того, как работает наш мозг.

Изучение зеркальной симметрии

Пример задачи, которую, в отличие от перцептрона, решит машина Больцмана, – изучение зеркальной симметрии[165]. Наше тело симметрично относительно вертикальной оси. Мы можем создать много случайных узоров с такой симметрией, как показано на рис. 7.5. Мы также можем создавать рисунки с горизонтальной и диагональной осями симметрии. В нашей модели машины Больцмана эти блоки двоичных входов размером 10×10 проецировались на 16 скрытых блоков, которые, в свою очередь, проецировались на три входных блока, по одному для каждой из трех вероятных осей симметрии. После обучения на 6000 симметричных входных образах машина Больцмана успешно определяла ось симметрии новых входных образов в 90 процентах случаев. Перцептрон может лишь действовать наугад, потому что один вход не несет никакой информации о симметрии узора – необходимо учитывать корреляции между парами входов. Примечательно, что машина Больцмана видела вовсе не то, что видите вы, ведь каждый скрытый блок получал входящие данные в определенном порядке. Как если бы для вас входные блоки расположили в хаотичном порядке, который бы выглядел беспорядочной массой даже при наличии скрытой симметрии.

Однажды я смотрел на экран и определял симметрию каждого входного набора данных по два в секунду. Нил Коэн, коллега с факультета психологии Университета Хопкинса, который тоже вел наблюдение, был поражен моим результатом. Наблюдая за процессом обучения машины Больцмана, я и сам научился находить симметрию автоматически, не всматриваясь. Мы поставили эксперимент с участием студентов и проследили за их успехами[166].

Вначале им требовалось много времени, чтобы правильно определить симметрию. Но после нескольких дней обучения они сильно ускорились, и к концу эксперимента задача стала для них настолько легкой, что они могли разговаривать с нами во время ее выполнения и одновременно давать правильный ответ. Это пример удивительно быстрого перцептивного обучения[167].

В Университете Хопкинса я вел курс «Вычислительная биофизика», в ходе которого привлек несколько талантливых студентов и исследователей. Бен Юхас, аспирант с кафедры электротехники, во время работы над докторской диссертацией научил нейросеть читать по губам[168]. Известно, как звучит речь при разных движениях рта. Сеть Бена преобразовала изображения рта во время артикуляции в соответствующий частотный спектр звука, порождаемый в каждый момент времени. Затем это добавили к звуковому спектру шумов, чтобы улучшить распознавание речи. Андреас Андреу, громкоголосый грек с Кипра, создавал аналоговые сверхбольшие интегральные схемы в подвале Бартон-холла. В 1980-х годах преподаватели на факультете, как и в других университетах, относились к нейросетям враждебно, однако Бена и Андреаса это не остановило. Андреас поднялся по карьерной лестнице до профессора и стал одним из основателей Центра языка и обработки речи в Университете Хопкинса.

Распознавание почтовых индексов, написанных вручную

Не так давно Джеффри Хинтон и его студенты из Университета Торонто обучили машину Больцмана с тремя слоями скрытых блоков с высокой точностью распознавать рукописные почтовые индексы (рис. 7.7)[169]. Поскольку в сети было и обратное, и прямое распространение, ее можно было запускать в обратном направлении, зафиксировав один из выходных блоков и создав соответствующие ему входные блоки (рис. 7.6). Генеративные модели фиксируют статистическую структуру обучающего набора, и генерируемые выборки наследуют эти свойства. Как если бы сеть перешла в спящий режим, а активность самого высокого уровня создавала нечто похожее на сны.

Развитие нейронных сетей шло быстро, однако консервативным ученым-когнитивистам было трудно их принять.

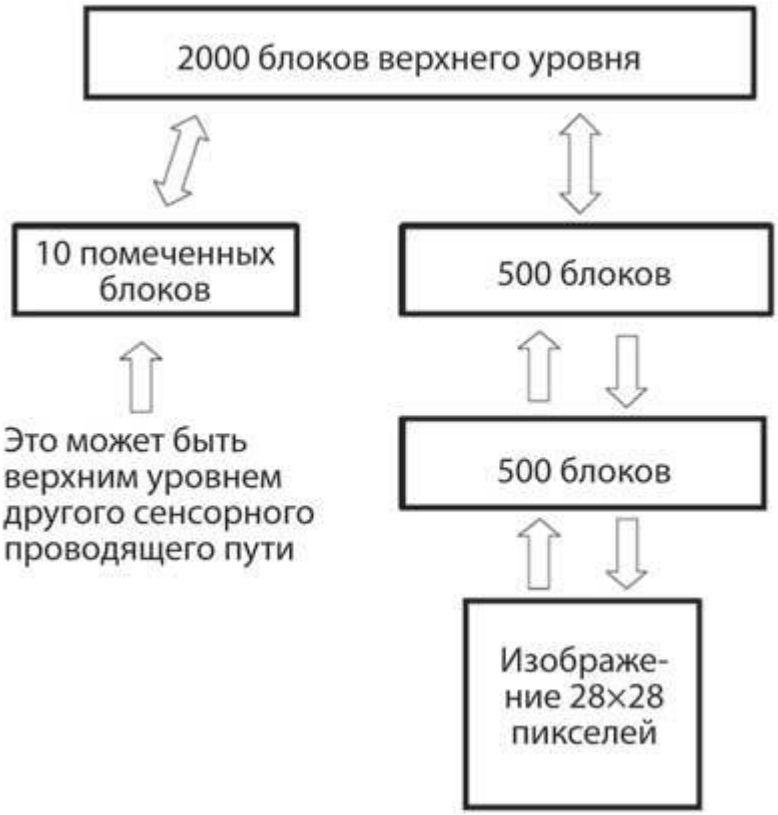


Рис. 7.6. Многослойная машина Больцмана для распознавания и формирования рукописных цифр. Размер изображения – 28×28=784 пикселей, которые могут быть белыми и черными. Цель – определить цифру на основе 10 выходных единиц (0–9)

За пределами лаборатории параллельной распределенной обработки в Ла-Хойя и нескольких отдельных исследовательских групп, обработка символов по-прежнему оставалась единственным методом.

Рис. 7.7. Шаблоны входного слоя, генерируемые многослойной машиной Больцмана, обученной распознавать рукописные цифры. Каждая линия была создана фиксацией одного из 10 выходных блоков, и входной слой непрерывно изменялся от примера к примеру, показанных выше. Ни одной из этих цифр не было в тренировочном наборе – ими «бредила» внутренняя часть обученной сети

На симпозиуме Общества когнитивной науки в 1983 году, на котором присутствовали мы с Хинтоном, психолог Зенон Пилишин, изучавший кратковременную память и умственные образы, выразил свое презрение к машине Больцмана, выплеснув на сцену стакан воды и выкрикнув: «Это не вычисления!» Другие целиком отвергли эту идею как простую «статистику». Джером Леттвин, который тоже был на симпозиуме, сказал, что ему очень нравится то, чем мы занимаемся. Леттвин совместно с Умберто Матурана, Уорреном Маккалаком и Уолтером Питтсом написали в 1959 году статью «Что глаз лягушки сообщает мозгу лягушки»[170]. В ней приводятся

доказательства, что существуют детекторы насекомых, которые лучше всего реагируют на небольшие темные пятна. Эта идея оказала большое влияние на системную нейробиологию. Поддержка Леттвиным нашей новой модели нейронной сети много значила для связи между ней и более ранней эпохой.

Неконтролируемое обучение и развитие коры головного мозга

Машину Больцмана можно использовать либо в контролируемом варианте, где входы и выходы зафиксированы, либо в неконтролируемом варианте, где зафиксированы только входы. Джеффри Хинтон применял неконтролируемую версию для наращивания глубины машины Больцмана по одному слою за раз[171]. Он начал с одного слоя скрытых элементов, подключенных к входным элементам, и обучал их на непомеченных данных, которые гораздо легче получить, нежели помеченные. В Интернете есть миллиарды непомеченных изображений и аудиозаписей. Неконтролируемое обучение пытается извлечь из них статистические закономерности, общие для всех данных. Первый слой скрытых элементов может извлекать из данных только простые пространственные объекты, что доступно и перцептрону. Следующий шаг – фиксирование веса первого слоя и добавление второго слоя единиц сверху. Далее неконтролируемое обучение машины Больцмана приводит к более сложному набору функций, и этот процесс можно повторить, чтобы создать сеть со множеством слоев.

Классификация становится намного проще в верхних слоях, требуя гораздо меньше обучающих примеров для достижения сходимости на более высоком уровне выполнения. Это происходит потому, что элементы в верхних слоях включают больше нелинейных комбинаций низкоуровневых признаков, что позволяет им как совокупности отделять общее от частного. Теоретически вопрос математического описания этой путаницы пока остается

открытым, но глубокие нейросети уже используют новые геометрические инструменты[172].

Интересно, что кора головного мозга также развивается слой за слоем. На ранних стадиях развития зрительной системы нейроны в первичной зрительной коре, первыми получающие входящие данные от глаз, обладают высокой пластичностью и могут быть легко «перепрограммированы» потоком входной зрительной информации до окончания критического периода. Иерархия зрительных областей в задней части мозга созревает первой, а корковые области ближе к передней части мозга – гораздо позже. Префронтальная кора последней достигает полной зрелости, созревание может закончиться уже после совершеннолетия. Таким образом, развитие идет плавными волнами с перекрывающимися критическими периодами, когда связи в кортикальной области наиболее подвержены влиянию нервной деятельности. Джеффри Элман и Элизабет Бейтс, когнитивисты из Калифорнийского университета в Сан-Диего, совместно со своими коллегами разработали нейронную сеть, показывающую, как последовательное развитие коры мозга может объяснить вехи в развитии ребенка, появление у него новых способностей, с помощью которых он познает мир[173]. Это открыло новое направление исследований того, как наше долгое детство сделало людей чемпионами по обучаемости, и позволило под другим углом взглянуть на некоторые модели поведения, которые считались врожденными.

В книге «Лжецы, любовники и герои» мы со Стивеном Кварцем, бывшим постдокторантом моей лаборатории, который сейчас работает в Калтехе, писали, что во время длительного периода развития мозга в детском и подростковом возрасте опыт может сильно влиять на экспрессию генов в нейронах и тем самым изменять нейронные цепи, отвечающие за поведение[174]. Взаимодействие генетических различий и влияния окружающей среды – активная область исследований, позволяющая по-новому взглянуть на сложности развития мозга. Она выходит за рамки

дебатов о роли природы и воспитания и пересматривает их с точки зрения культурной биологии: человеческая культура одновременно и формирует нашу биологию, и является ее продуктом[175]. Новой главой в этой истории стало недавнее открытие, что в период раннего развития, когда быстро растет число синапсов между нейронами, ДНК внутри нейронов изменяется формой метилирования[176], которая регулирует экспрессию генов и уникальна для мозга[177]. Это называется эпигенетической модификацией и может быть связью между генами и опытом, что и предполагали мы со Стивом Кварцем.

К 1990-м годам когнитивная нейробиология расширялась, и революция нейронных сетей шла полным ходом. Компьютеры становились быстрее, но скорости пока не хватало. Машина Больцмана была просто конфеткой с технической точки зрения, но ужасно медленной для моделирования. Что действительно помогло нам добиться прогресса, так это более быстрый алгоритм обучения, который появился у нас именно тогда, когда мы больше всего в нем нуждались.

Глава 8. Метод обратного распространения ошибки

Калифорнийский университет в Сан-Диего был основан в 1960 году и со временем превратился в крупный центр биомедицинских исследований. В 1986 году в нем открыли первый в мире факультет когнитивной науки[178]. Дэвид Румельхарт (рис. 8.1) был видным математиком и когнитивным психологом, работавшим с символьным, основанным на правилах, подходом к ИИ, который преобладал в 1970-х годах.

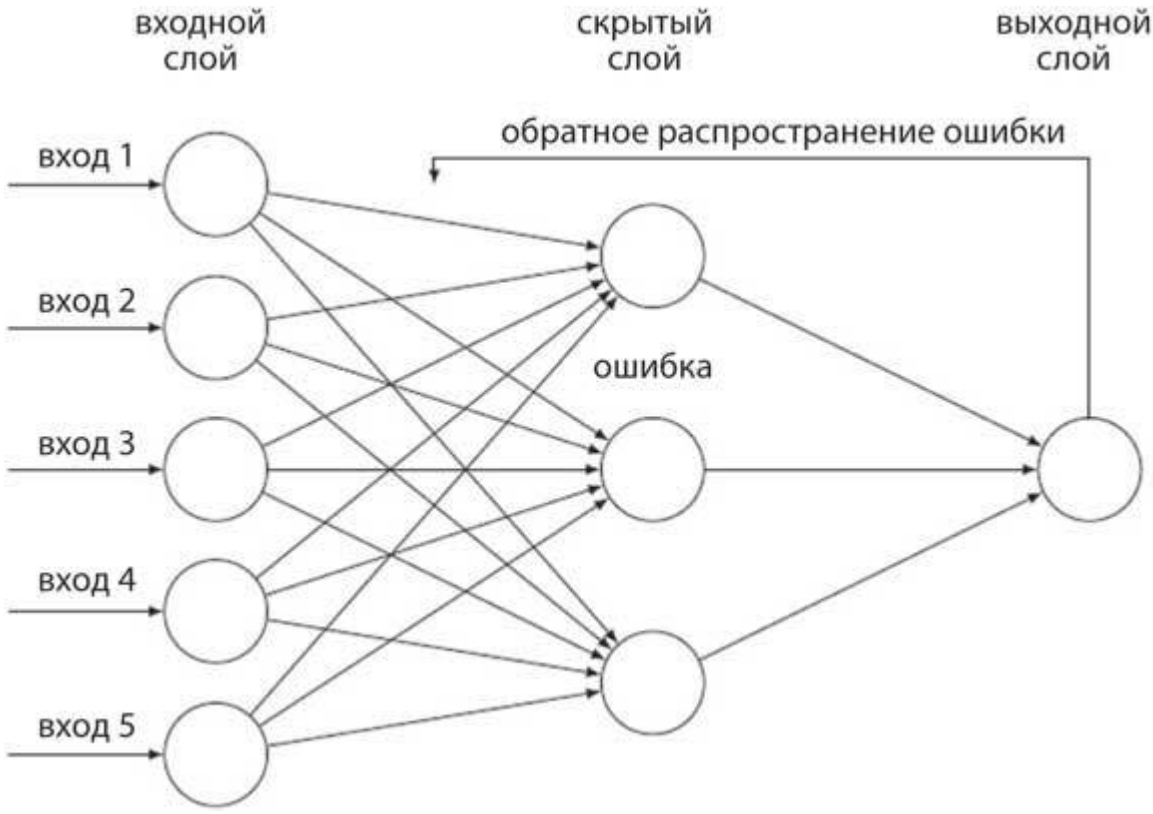
Рис. 8.1. Дэвид Румельхарт в Калифорнийском университете в Сан-Диего в 1986 году, примерно в то время, когда были изданы книги о параллельной распределенной обработке. Румельхарт

оказал влияние на техническую разработку алгоритмов обучения для моделей многослойных сетей и использовал их, чтобы понять психологию языка и мышления

Когда я впервые встретил его в 1979 году на семинаре, организованном Джеффри Хинтоном в Калифорнийском университете в Сан-Диего, Румельхарт был одним из первых, кто использовал новый подход к психологии человека, который он вместе с Джеем Макклелландом назвали параллельной распределенной обработкой (Parallel Distributed Processing; PDP). Румельхарт мыслил глубоко и часто делал проницательные замечания.

Алгоритм обучения машины Больцмана доказуемо мог изучить проблемы, требующие скрытых элементов, показывая, что, вопреки мнению Минского и Пейперта, а также большей части научного мира, возможно обучить многослойную сеть и преодолеть ограничения перцептрона. При этом не ставилось никаких ограничений ни на количество слоев в сети, ни на связи внутри слоя. Казалось, прогрессу нет предела, но была одна проблема: при моделировании достижение равновесия и сбор статистики становились все медленнее, а сетям покрупнее требовалось гораздо больше времени, чтобы сбалансироваться.

Блок 5. Обратное распространение ошибки



В сети с обратным распространением ошибки входные данные передаются с прямой связью: слева на схеме входные элементы распространяются вперед через соединительные узлы (указаны стрелками) к скрытому слою элементов, которые, в свою очередь, проецируются на выходной слой. Выходные данные сравниваются со значением, заданным учителем, и разница используется для обновления веса в выходном блоке, чтобы снизить вероятность ошибки. Затем веса между входными блоками и скрытым слоем обновляются на основе обратного распространения ошибки, исходя из того, насколько каждый вес влияет на ошибку. Обучаясь на множестве примеров, скрытые элементы совершенствуют избирательные свойства, которые используются, чтобы различать разнообразные входные данные и разделять их на категории в выходном слое. Это называется обучением представлениям.

В принципе, можно построить массово-параллельный компьютер, который намного быстрее, чем традиционная архитектура фон Неймана, выполняющая одно обновление за раз. Это путь, по которому пошла природа. В 1980-х мы использовали цифровые компьютеры, которые могли выполнять только около миллиона операций в секунду. Сегодня компьютеры выполняют миллиарды операций в секунду, а благодаря соединению тысяч ядер высокопроизводительные компьютеры работают в миллион раз быстрее. Такой рост беспрецедентен в технологиях. Стал ли ваш автомобиль в миллион раз мощнее, чем машины из 1980-х?

США поставили на Манхэттенский проект два миллиарда долларов без каких-либо гарантий, что атомная бомба получится, и строжайшей тайной было то, что она получилась. Как только стало известно, что многослойные сети можно обучать с помощью машины Больцмана, произошел взрыв новых обучающих алгоритмов. В то же время, когда мы с Джеффри Хинтоном работали над машиной Больцмана, Румельхарт разработал другой алгоритм обучения для многослойных сетей, который оказался более продуктивным[179].

Оптимизация

Оптимизация – ключевое математическое понятие в машинном обучении. Для многих задач можно найти функцию стоимости, решение которой – состояние системы с наименьшими затратами. Для сети Хопфилда функция стоимости – это энергия, как описано в главе 6, и цель – найти состояние сети с наименьшим расходом энергии. Для сети прямого распространения функция стоимости обучения – сумма квадрата ошибок выходного слоя обучающего набора. Градиентный спуск – общая процедура, которая минимизирует функцию стоимости, внося дополнительные изменения в веса в сетях в направлении наибольшего снижения стоимости[180]. Представьте функцию стоимости в виде горного

хребта, а градиентный спуск – в виде лыжни, по которой вы спускаетесь вниз по склону.

Румельхарт обнаружил, как вычислить градиент для каждого веса в сети с помощью процесса, называемого обратным распространением ошибок (блок 5). Начиная с выходного слоя, где известна ошибка, легко вычислить градиент от входных весов к выходным элементам. Следующим шагом было использование градиентов выходного слоя для вычисления градиентов на предыдущем слое весов, и так далее слой за слоем вплоть до входного. Это очень эффективный способ вычисления градиентов ошибки.

Хотя у метода обратного распространения ошибки нет такой же элегантности и глубоких корней в физике, как у алгоритма машинного обучения Больцмана, он более эффективен и значительно ускорил прогресс. Статья об этом за авторством Румельхарта, Хинтона и Рона Уильямса была опубликована в журнале Nature в 1986 году[181], и с тех пор ее процитировали более 25 тысяч раз в других научных работах. Статья, процитированная сто раз, играет большую роль в своей области, а статья об обратном распространении ошибки стала бестселлером.

NETtalk

В 1984 году я посетил Принстонский университет и послушал выступление студента магистратуры Чарльза Розенберга о машине Больцмана. Обычно я выступал с этим докладом, и я был впечатлен. Он спросил, может ли приехать в мою лабораторию для работы над летним проектом. К тому времени как Розенберг прибыл в Балтимор, мы переключились на изучение метода обратного распространения ошибки, что позволило нам думать о работе над реальной задачей, а не над демонстрационными моделями, над которыми я трудился ранее. Розенберг был учеником прославленного лингвиста Джорджа Миллера, поэтому мы искали оптимальное решение в языке, которое не было настолько сложным, чтобы нельзя было продвинуться вперед, но и не настолько легким, чтобы известные методы могли решить эту проблему. Лингвистика – широкая отрасль со множеством ответвлений. Фонология – раздел лингвистики, изучающий произношение слов. Синтаксис – объединение слов в фразы. Семантика – значение слов и предложений. Прагматика – влияние контекста на смысл речи. Мы решили начать с фонологии и продвигаться вверх.

Произношение в английском языке довольно трудное, поскольку в нем сложные правила с большим количеством исключений. Например, гласные долгие, если в конце слова есть буква *e* (*gave, brave*), однако слово *have* не подчиняется этому правилу. Я пошел в библиотеку и взял книгу с сотнями страниц правил и исключений, составленными фонологами. Часто были правила в исключениях и исключения из исключений. Короче, у лингвистов были правила на все случаи[182]. Гораздо хуже, что не все произносят слова одинаково. Существует множество диалектов, в каждом из которых свой набор правил.

Джеффри Хинтон посетил меня в Университете Хопкинса на этапе раннего планирования проекта и сказал нам, что, по его мнению,

произношение слишком сложно. В итоге мы снизили планку и взяли книгу для детей, которые только учатся читать, где была всего сотня слов. Сеть, которую мы создали, имела окно, рассчитанное на 7 букв, каждая из ячеек была представлена 29 элементами, включая пробелы и знаки пунктуации. В итоге получилось 203 единицы входных сигналов. Входные блоки были соединены с 80 скрытыми блоками, а скрытые блоки спроецированы на 26 выходных единиц, по одной для каждого из простых звуков, называемых фонемами, которые существуют в английском языке[183]. Сеть содержала 18 629 весов (рис. 8.2), что много по меркам 1986 года и безумно много по меркам математической статистики. Нам сказали, что с таким количеством параметров обучающий набор будет очень большим и сеть не сможет его обобщить.

По мере того как слова по одной букве появлялись на экране, сеть назначала фонему средней букве. Часть проекта, которая заняла больше всего времени, – сопоставлять фонему с верной буквой вручную, поскольку не в каждом слове количество букв совпадает с количеством фонем. Но в то же время обучение происходило на наших глазах, становясь все лучше и лучше по мере того, как фразы циклически повторялись на экране, и когда результат на тренировочном наборе сходился, производительность сети была практически идеальной. Тестирование на новых словах нельзя было назвать успешным, но мы ожидали, что обобщение такого маленького тренировочного набора будет слабым. Тем не менее предварительный итог вселял оптимизм.

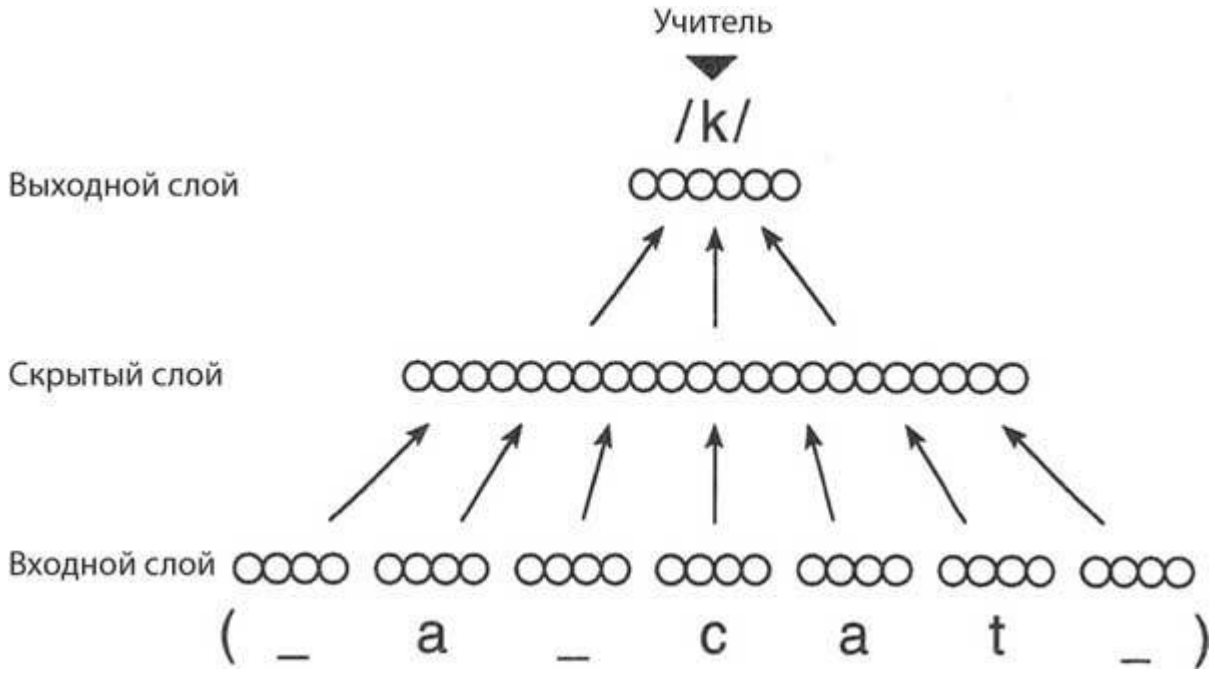


Рис. 8.2. Сетевая модель прямого распространения NETtalk. Семь групп элементов на нижнем уровне представляют собой буквы появляющегося в окне текста, по одной за раз. Цель сети – правильно предсказать звучание центральной буквы (твердый звук «к» в приведенном примере[184]). Каждый элемент входного слоя связывается со всеми скрытыми элементами, которые, в свою очередь, проецируются на все элементы выходного слоя. Алгоритм обратного распространения ошибки использовался для тренировки весов под контролем учителя. Выходной образец для верной фонемы сравнивается с выходом сети, и ошибка возвращается к весам на более ранних уровнях. [Rosenberg C. R., Sejnowski T. J. “Parallel Networks That Learn to Pronounce English Text”, Complex Systems, 1, 145–168, 1987]

Затем мы использовали 20 тысяч слов из «Брауновского корпуса»[185], в котором обозначили фонемы для каждой из букв и ударения. Сопоставление букв и звуков заняло несколько недель, но как только обучение началось, сеть впитала в себя весь сборник за одну ночь. Но как хорошо она смогла все обобщить? Прекрасно

смогла! Сеть распознала закономерности английского произношения и научилась находить исключения при том же строении и том же алгоритме обучения. Та сеть была крошечной по нынешним стандартам, что подчеркивает, насколько эффективно сеть разобралась в английской фонологии. Это был первый намек на то, что нейронные сети можно связать с речью – основой символических представлений.

Сеть, преобразующую буквы в звуки, мы назвали NETtalk[186]. Прежде чем научиться читать вслух, она прошла фазу «лепета», во время которой изучала разницу между согласными и гласными, но назначала фонему *b* для всех согласных и фонему *a* для всех гласных. Поначалу это звучало как «ба», но потом, после продолжительного обучения, превратилось в «ба-га-да», до жути напоминая лепет младенца. Зачем она начала правильно произносить короткие слова, и к концу обучения стала понятна большая часть того, что она говорила.

Чтобы проверить работу NETtalk с диалектом, мы нашли фонологическую транскрипцию интервью с латиноамериканским мальчиком из Лос-Анджелеса. Обученная сеть воссоздала испанский акцент ребенка, рассказывавшего, как он ходит в гости к бабушке и получает конфеты. Я записывал фрагменты во время последовательных этапов обучения, воспроизводя выходные данные NETtalk на синтезаторе речи DECtalk, который преобразовывал строку с обозначенными фонемами в слышимую речь. Когда я включил запись во время лекции, аудитория была ошеломлена: сеть будто говорила сама[187]. Этот летний проект превзошел все наши ожидания и стал первым случаем обучения нейронных сетей для практического применения. В 1986 году меня пригласили продемонстрировать NETtalk на утреннем телешоу Today, и этот выпуск посмотрело на удивление много зрителей. До того момента нейронные сети оставались предметом загадочных исследований. Я до сих пор встречаю людей, которые впервые услышали о нейронных сетях, посмотрев эту передачу.

Хотя NETtalk ярко продемонстрировала, как сеть может отображать некоторые аспекты языка, она не достаточно хорошо моделирует то, как люди осваивают чтение Во-первых, сначала мы учимся говорить, и только потом – читать. Во-вторых, нам дают несколько фонетических правил, которые помогают справиться со сложной задачей – научиться хорошо читать вслух. Тем не менее чтение быстро превращается в распознавание образов, и не нужно прилагать сознательные усилия, чтобы применять правила. Как и NETtalk, большинство носителей английского языка без усилий произнесут лишенные смысла фразы, такие как стихотворение «Бармаглот»: «Варкалось. Хливкие шорьки...» Это псевдослова, которых нет в словарях, но их фонемы образуются из тех же сочетаний букв, что и в настоящих.

NETtalk сильно впечатлила аудиторию, но наука требовала проанализировать сеть, чтобы выяснить, как она работает. Мы с Чарли Розенбергом применили кластерный анализ к схемам активности в скрытых элементах и выяснили, что NETtalk обнаружила ту же закономерность, по которой схожие гласные и согласные звуки объединяются в группы, что нашли и лингвисты. Марк Зейденберг и Джей Макклелланд использовали такой же подход как точку отсчета и провели подробное сравнение с этапами, которые проходят дети, когда учатся читать[188].

Рис. 8.3. Летние нейросетевые курсы в Университете Карнеги – Мелона в 1986 году. Джефф Хинтон в первом ряду, по бокам от него – Джей Макклелланд и я. На этой фотографии – видные специалисты в области нейронных вычислений сегодняшнего дня. Нейронные сети в 1980-х годах были наукой XXI века в XX веке

Другая сеть, которая научилась образовывать прошедшее время английских глаголов, стала знаменитой в мире когнитивной психологии, поскольку опирающаяся на правила старая гвардия сражалась с передовой группой параллельно распределенной обработки[189]. Обычный способ образовать прошедшее время[190] – добавить – *ed* в конце слова, например, *to train (тренироваться)* – *trained (тренировался)*. Однако есть и исключения, такие как *to run (бежать)* – *ran (бежал)*. У сети нет проблем ни с правильными, ни с неправильными глаголами. И хотя эти споры уже неактуальны, фундаментальный вопрос о роли явного представления правил в мозге остается открытым. Недавние эксперименты подтверждают, что в процессе обучения постепенно осваивается и изменение формы слов в соответствии с их смыслом[191]. Успех глубокого обучения Google Переводчика и других приложений для

естественных языков в улавливании нюансов письменной речи еще сильнее подтверждает предположение, что мозгу не нужно постоянно использовать правила, даже если по поведению кажется, что он их применяет.

На первых посвященных нейросетевым моделям курсах, которые мы с Джеффри Хинтоном и Дэйвом Турецки организовали в Университете Карнеги – Меллона в 1986 году (рис. 8.3), студенты сделали пародию на NETtalk. Они выстроились в несколько рядов, каждый студент представлял единицу в сети. Когда они представляли «j» в фамилии «Sejnowski», то выдавали ошибку, потому что она произносится как буква «у» и не соответствует шаблону[192]. В то время лишь в немногих вузах были преподаватели, которые читали курсы по нейронным сетям. Многие из этих студентов впоследствии совершили важные открытия и достигли карьерных высот. Вторые летние курсы были проведены в Университете Карнеги – Меллона в 1988 году, а третьи – в Калифорнийском университете в Сан-Диего в 1990 году. Необходима смена поколений, чтобы новые идеи стали популярными. Эти летние курсы были бесценным опытом и лучшими инвестициями, которые мы сделали на заре нейросетей.

Возрождение нейронных сетей

Двухтомник Румельхарта и Маклелланда «Параллельная распределенная обработка», изданный в 1986 году, уже стал классикой. Джеффри Хинтон тоже планировал принять участие в работе над ним, однако его отвлекла машина Больцмана. Эта книга – первая, в которой описано влияние сетей и алгоритмов многослойного обучения на понимание умственных и поведенческих процессов. Было продано более 50 тысяч экземпляров, изданных MIT Press[193], – бестселлер по академическим меркам. У сетей, обученных методом обратного распространения ошибки, были скрытые блоки со свойствами, напоминающими свойства нейронов зрительной коры[194]. Кроме того, схемы распада нейронных сетей имели много общего с проблемами у человека после травмы мозга[195].

Фрэнсис Крик был членом группы параллельной распределенной обработки и посещал большинство встреч и семинаров. В спорах, насколько модели такой обработки схожи с биологическим процессом, он утверждал, что они должны рассматриваться как некий демонстрационный образец, а не как точная копия мозга. В книге, посвященной параллельной распределенной обработке, он написал главу о том, что на тот момент было известно о коре головного мозга. Я же добавил главу о том, чего мы не знали о ней. Пиши мы эти главы сегодня, обе вышли бы гораздо длиннее.

В 1980-х годах и истории успеха, о которых никто не знал. Одной из самых прибыльных компаний в сфере нейросетей стала HNC Software Inc., основанная Робертом Хехт-Нильсеном, который использовал нейронную сеть для предотвращения мошенничества с кредитными картами. Хехт-Нильсен преподавал на кафедре электрической и компьютерной инженерии Калифорнийского университета в Сан-Диего популярный курс по практическому применению нейронных сетей. Ежедневно в магазинах Южной

Америки совершаются кражи с кредиток (что я лично испытал на себе), и они же являются объектами массовой киберпреступности. Мы проводим множество операций с картами, и порой сложно определить подозрительные транзакции. Например, отказ в оплате в ресторане в Рио-де-Жанейро может доставить проблем туристу. Людям в 1980-х годах приходилось принимать срочные решения. В итоге совершались мошеннические операции более чем на 150 миллиардов долларов в год. Компания HNC Software Inc. использовала алгоритмы обучения нейросетей, чтобы выявлять мошенничество с пластиковыми картами гораздо точнее, чем люди, экономя компаниям, выпускающим карты, миллиарды долларов в год. Компания HNC была приобретена в 2002 году за миллиард долларов компанией Fair Isaac and Company (FICO)[196], известной своими кредитными рейтингами.

Есть что-то волшебное в наблюдении за тем, как сеть учится, становится все лучше, делая небольшие шаги. Процесс медленный, но если хватает обучающих примеров и сеть достаточно велика, алгоритмы обучения могут найти такое представление, которое хорошо обобщается на новые входные данные. Когда процесс повторяется при случайно выбранном наборе весов, каждый раз учится другая сеть, но со схожими характеристиками. Разные сети могут решить одну и ту же проблему; это влияет на то, что мы должны ожидать, когда сумеем восстановить полный набор связей в мозге разных людей. Если у многих сетей одинаковое поведение, ключ к их пониманию – используемые мозгом алгоритмы обучения, найти которые легче.

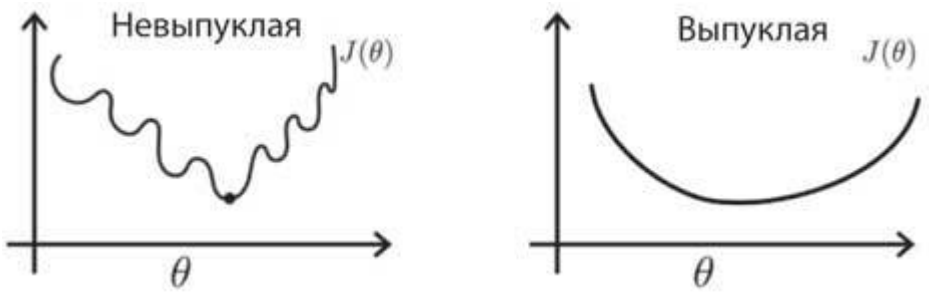


Рис. 8.4. Невыпуклые и выпуклые функции стоимости. Графики показывают зависимость функции стоимости J (θ) от параметра θ. Выпуклая функция (справа) имеет только один – глобальный – минимум, которого можно достичь, двигаясь вниз по кривой из любого места на ней. Представьте, что вы лыжник и всегда разворачиваете свои лыжи к самому крутому спуску. Вы гарантированно доберетесь до низа. Напротив, невыпуклая функция стоимости (слева) может иметь локальные минимумы, которые являются ловушками, мешающими найти глобальный минимум при спуске. Как следствие, невыпуклые функции стоимости трудно оптимизировать. Однако этот одномерный пример вводит в заблуждение. Когда есть много параметров (обычно миллионы в нейронной сети), могут быть определенные (седловые) точки, выпуклые в одних измерениях и вогнутые в других. Когда вы находитесь в такой точке, всегда есть направление, чтобы спуститься вниз

Понимание глубокого обучения

В задачах с выпуклой оптимизацией отсутствуют локальные минимумы и гарантируется сходимость к глобальному минимуму (рис. 8.4). Эксперты по оптимизации утверждали, что, так как обучение сетей со скрытыми элементами было задачей невыпуклой оптимизации, мы попадали в локальные минимумы и впустую тратили время (рис. 8.4). Опыт показал, что они ошибались. Но почему? Теперь мы знаем, что в многомерных пространствах локальные минимумы функции стоимости редки, пока вы не дойдете до заключительных этапов обучения. На ранних стадиях почти все направления ведут под гору, и на этом пути есть седловые точки, где в одних измерениях можно пойти вверх по ошибке, а в других – вниз. Предположение, что вы застрянете в локальных минимумах,

основано на решении проблем в пространстве с малым числом измерений (см. рис. 8.4), где меньше запасных выходов.

У современных моделей глубоких сетей миллионы элементов и миллиарды весов. Миллиардномерное пространство параметров – кошмар для статистики. Специалисты по статистике традиционно анализируют простые модели с несколькими характеристиками, чтобы доказать предположения, используя небольшие наборы данных. Они заверили нас, что при таком множестве параметров мы добьемся безнадежной переподгонки данных, или чрезмерного обучения, то есть сеть просто запомнит примеры и не сможет обобщить их на новые тестовые наборы. Но мы использовали методы, такие как принудительное снижение весов, если они не делали ничего полезного, что помогло бы избежать переподгонки. Сейчас, чтобы обойти эту проблему, для обучения глубоких сетей используются еще более сложные методы регуляризации.

Джеффри Хинтон изобрел отлично продуманный метод регуляризации – исключение (дропаут, dropout)[197]. Во время каждой эпохи обучения[198], когда градиент оценивается по ряду примеров и делает шаг в пространстве весов, половина единиц случайным образом вырезается из сети. А значит, в следующую эпоху обучается другая сеть. Как следствие, в каждую эпоху остается меньше параметров для обучения, и у полученной в результате сети меньше зависимостей между единицами, чем если бы в каждую эпоху обучалась одна и та же большая сеть. Дропаут уменьшает частоту ошибок в глубоких сетях на 10 процентов, что стало значительным прогрессом. В 2009 году компания Netflix провела открытый конкурс, предложив приз в миллион долларов тому, кто сможет уменьшить ошибку их системы рекомендаций на 10 процентов[199]. Это основная технология для онлайн-трансляций. Почти каждый магистрант в области машинного обучения принял участие в конкурсе[200].

Примечательно, что синапсы в коре головного мозга исключаются с высокой скоростью. На каждый входной импульс приходится 90 процентов отказов обычного возбуждающего синапса[201]. Это похоже на бейсбольную команду, где почти все игроки имеют средний уровень 100[202]. Как мозг может стабильно функционировать с такими ненадежными кортикальными синапсами? Когда в нейроне тысячи вероятностных синапсов, вариабельность их суммарной активности относительно невысока[203], так что производительность не падает так сильно, как вы можете себе представить. Польза от обучения с использованием исключения на уровне синапсов может перевешивать затраты на снижение точности. Исключение также экономит энергию, так как работа синапсов дорого обходится. Наконец, кора головного мозга применяет вероятности для вычисления предполагаемых результатов, а не точных, и для этого эффективно использование вероятностных компонентов.

Хотя кортикальные синапсы могут быть ненадежными, они удивительно точны в своей силе[204]. Размеры кортикальных синапсов и, соответственно, их сила различаются более чем в сотню раз, и в этих пределах сила отдельных синапсов может быть увеличена или уменьшена. Недавно моя лаборатория совместно с Кристен Харрис, нейроанатомом из Техасского университета в Остине, воссоздала небольшой фрагмент крысиного гиппокампа – области мозга, необходимой для формирования долговременных воспоминаний, – которая содержала 450 синапсов. Чаще всего это были одиночные синапсы на дендритных ветвях, но в ряде случаев два синапса от одного аксона передавали сигналы одному и тому же дендриту. К нашему удивлению, они были почти идентичны по размеру, а значит, как мы знали из предыдущих исследований, одинаковы по силе. Многое известно об условиях, которые приводят к изменению силы синапсов в зависимости от истории входных импульсов и соответствующей электрической активности дендритов,

одинаковых для парных синапсов. Из этих наблюдений мы сделали вывод, что точность хранения информации в силе синапсов значительная – не меньше пяти бит на синапс. Любопытно, что для достижения высокого уровня производительности алгоритмам обучения глубоких рекуррентных сетей требуется всего пять бит[205]. Это может не быть совпадением[206].

Степень размерности сетей в мозге настолько высока, что мы даже не можем точно оценить ее. Общее количество синапсов в коре головного мозга – около ста триллионов, астрономически высокая грань. Человеческая жизнь длится не более нескольких миллиардов секунд. Таким образом, вы можете позволить себе посвящать сто тысяч синапсов каждой секунде своей жизни. На деле у нейронов, как правило, кластеризованные локальные соединения. Например, в кортикальном столбце сто тысяч нейронов соединены миллиардом синапсов – число довольное большое, но все же не заоблачное. Длинные соединения куда менее распространены, потому что требуют много места и энергии.

Число нейронов, которое нужно, чтобы представить в мозге объект или понятие, важно, и его необходимо определить. Предположительно требуется около миллиарда синапсов и около ста тысяч нейронов, распределенных по десяти кортикальным областям[207], что позволяет хранить около ста тысяч отдельных классов объектов и понятий в ста триллионах синапсов, что сходно с количеством слов в английском языке[208]. На практике популяции нейронов, представляющих схожие объекты, перекрываются, благодаря чему растет способность коры головного мозга представлять связанные объекты и отношения между объектами. У человека эта способность развита намного лучше, чем у других млекопитающих, из-за сильно увеличенной ассоциативной коры, которая находится на вершине сенсорной и моторной иерархий.

Изучение вероятностных распределений в многомерных пространствах было относительно неисследованной областью

статистики. Несколько ученых-статистиков из сообщества NIPS, таких как Лео Брейман из Стэнфордского университета, исследовали статистические проблемы, возникающие при навигации по пространствам с высокой размерностью и многомерным наборам данных. Некоторых из сообщества NIPS, например, Майкла Джордана из Калифорнийского университета в Беркли, приняли на работу в отдел статистики. В эпоху больших данных машинное обучение шагало там, куда статистики не решались ступать. Однако недостаточно просто обучить крупные сети делать удивительные вещи – нужно их проанализировать и понять, почему они эффективны. Физики взяли на себя инициативу на этом фронте, используя методы из статистической физики для анализа свойств обучения по мере того, как число нейронов и синапсов становится запредельно большим.

Ограничения нейронных сетей

В настоящее время нейронные сети могут дать правильный ответ на вопрос, но не объяснить, как к нему пришли. Например, пациент находится в приемном отделении «скорой помощи» с острой болью в груди. Инфаркт миокарда, требующий немедленной помощи, или тяжелое расстройство желудка? Обученная сеть может поставить диагноз точнее, чем врач. Но без объяснения, как она это сделала, начинаешь сомневаться, а можно ли доверять ей. Врачи тоже учатся следовать алгоритмам, проводить серии тестов перед принятием решения, и обычно это работает. Проблема в том, что есть редкие случаи, к которым нельзя применить стандартный алгоритм, однако сеть, обученная на гораздо большем количестве примеров, чем среднестатистический врач видел за свою практику, может распознать их и верно поставить диагноз. Вы бы доверяли совету врача, который все подробно растолковал, или нейронной сети, которая по статистике лучше, но не дала объяснений? На самом деле у врачей, которые могут очень точно поставить диагноз даже в редком случае, как правило, большой опыт, и они применяют именно распознавание образов, а не алгоритмы[209]. Этим способом, вероятно, пользуются эксперты самого высокого уровня во всех областях.

Точно так же, как можно обучить сети ставить диагнозы на уровне эксперта, должна быть возможность обучить сети давать объяснения, как если бы они были частью обучающего набора. Вероятно, это даже улучшит диагноз. Сложность в том, что многие объяснения врачей неполные, упрощенные или неправильные. Медицинская практика сильно меняется от поколения к поколению, потому что строение тела гораздо сложнее, чем мы себе представляем. Если бы нам удалось проанализировать внутреннее состояние сетевых моделей, чтобы извлечь причинные объяснения, это привело бы к

новым выводам и гипотезам, которые можно было бы протестировать для совершенствования медицины.

Возражение, что нейронная сеть – «черный ящик», выводы которого нельзя понять, применимо и к мозгу, ведь люди, владея одинаковой информацией, могут делать совершенно разные выводы. И мы пока не знаем наверняка, как мозг принимает решения, используя опыт. Как показано в главе 3, выводы не всегда основаны на логике, к тому же возможны когнитивные искажения[210]. Более того, часто мы приводим лишь обоснованные или правдоподобные объяснения. Нельзя исключать, что какая-то огромная генеративная сеть заговорит, и мы сможем попросить у нее объяснений. Стоит ли нам ждать, что они будут лучше и рациональнее, чем те, что дают люди? Напомним, что сознание не имеет доступа к внутренней работе мозга. Сети глубокого обучения обычно предоставляют не один, а несколько основных прогнозов в порядке убывания, что дает некоторую информацию о достоверности вывода. Показывать вероятность разных ответов более наглядно, чем говорить «да» или «нет».

Контролируемые нейронные сети могут решать только те проблемы, которые попадают в диапазон данных, использованных для обучения сети. Обученная на схожих примерах, нейронная сеть должна хорошо справиться с новыми случаями, распространив на них имеющий опыт. Однако если новые входные данные выходят за пределы обучающего набора, экстраполяция опасна. Это не удивительно, ведь то же ограничение относится и к людям: не следует ожидать, что эксперт в одной из областей физики даст хороший совет по политическому вопросу или даже по вопросу из другой области физики. Однако до тех пор, пока обучающий набор достаточно велик, чтобы охватить весь спектр потенциальных входных данных, обобщение будет хорошо на них распространяться. На практике люди склонны использовать сходство для переноса опыта с области, в которой они разбираются, на

новую, но если области коренным образом различаются, это может привести к ложным аналогиям.

Еще одно возражение: нейронная сеть может оптимизировать выгоду в ущерб справедливости. Например, представитель недопредставленного меньшинства обращается за ипотекой и получает отказ от нейронной сети, обученной на миллионах заявок. Входные данные включают текущий адрес и другую связанную с этим меньшинством информацию. Таким образом, хотя и существует закон о запрете явной дискриминации меньшинств, сеть может использовать скрытую информацию против них. Проблема не в нейросети, а в функции стоимости, которую мы дали ей оптимизировать. Если единственная цель сети – получение прибыли, то она будет использовать любую информацию, чтобы ее максимизировать. Решить эту проблему можно, включив равноправие как еще одно условие в функцию затрат. Тогда оптимальным итогом будет баланс между прибылью и справедливостью. Кроме того, компромисс должен быть четко сформулирован в функции затрат, которая требует, чтобы кто-то определил вес каждой цели. В основе этих компромиссов должен лежать этический подход гуманитарных и социальных наук. Но имейте в виду, что у выбора функции затрат, который кажется справедливым, могут быть непредвиденные последствия[211].

Есть ли у природы функция стоимости? Оптимизация затрат в эволюции называется приспособляемость, но это понятие имеет смысл только для конкретного набора ограничений либо со стороны окружающей среды, либо со стороны ищущей выгодной решение системы. В мозге от рождения «запрограммирована» потребность в пище, тепле, безопасности, кислороде и продолжении рода, влияющая на поведение. Но есть ли функция стоимости, которая регулирует внимание? Мы лучше запоминаем то, что привлекло наше внимание, но что управляет им? Если ответим «мы», то попадем в замкнутый круг.

Продвижение

Во время творческого отпуска в 1987 году я выступал в Калтехе в качестве приглашенного профессора нейробиологии и посетил Фрэнсиса Крика в Институте Солка. Крик создавал исследовательскую группу, специализирующуюся на зрении, которым я тоже интересовался. На обеде с преподавателями я включил запись NETtalk, и она вызвала оживленную дискуссию. Вскоре, в 1989 году, я переехал в Ла-Хойя и основал при Институте Солка Лабораторию вычислительной нейробиологии, а также Институт нейронных вычислений при Калифорнийском университете в Сан-Диего. Это был потрясающий переход от младшего научного работника в Университете Хопкинса к ведущему преподавателю в Ла-Хойя, и в одночасье передо мной открылось множество возможностей, включая должность в Медицинском институте Говарда Хьюза, который оказывал щедрую поддержку моим исследованием более 25 лет.

Дэвид Румельхарт, преподававший метод обратного распространения ошибки, в 1987 году сменил Калифорнийский университет в Сан-Диего на Стэнфорд. Когда я перебрался в Сан-Диего, мне было жаль, что Дэвид уехал и мы виделись очень редко. С годами я заметил, что его поведение меняется. В конце концов ему поставили диагноз лобно-височная деменция – прогрессирующая потеря нейронов в лобной коре, влияющая на личность, поведение и речь. Румельхарт умер в 2011 году в возрасте 69 лет, уже не узнавая своих родственников и друзей.

Глава 9. Сверточные сети

К 2000 году одержимость нейронными сетями 1980-х спала, и все вернулось в нормальное русло исследований. Томас Кун однажды охарактеризовал время между научными революциями как регулярную работу ученых, теоретизирующих, наблюдающих и экспериментирующих в рамках устоявшейся парадигмы или объяснительной системы[212]. Джеффри Хинтон перешел в Университет Торонто в 1987 году и продолжил работу над небольшими улучшениями, но ни одно из них не имело такого успеха, как машина Больцмана. Хинтон в 2000-х годах возглавил программу «Нейронные вычисления и адаптивное восприятие» (Neural Computation and Adaptive Perception; NCAP) в Канадском институте перспективных исследований, куда вошли около 25 исследователей из Канады и других стран, сосредоточенных на решении сложных проблем обучения. Я был членом их консультативного совета под председательством Яна Лекуна (рис. 9.1) и участвовал в ежегодных встречах непосредственно перед конференцией NIPS. Изучались новые стратегии обучения нейронных сетей, и прогресс шел медленно, но стабильно. Хотя у нейронных сетей было много полезных применений, высокие ожидания 1980-х годов не оправдались. Но это не поколебало первопроходцев. Оглядываясь назад, можно сказать, что они готовили почву для грандиозного прорыва.

Устойчивый прогресс в машинном обучении

Конференция NIPS обеспечила в 1980-х годах благоприятные условия для развития нейронных сетей и открыла двери для других алгоритмов, которые могут обрабатывать большие многомерные наборы данных. Метод опорных векторов (Support Vector Machine, SVM) ворвался на сцену в 1995 году и начал новый этап в сетях

перцептронов, которые теперь называются неглубокими сетями. Мощным классификатором, который теперь в инструментарии каждого, SVM сделал так называемый kernel trick – математическое преобразование, которое эквивалентно прыжкам из пространства данных в гиперпространство, где точки данных перераспределяют, чтобы их было легче разделить. Томазо Поджио разработал иерархическую сеть HMAX с весами, задаваемыми вручную, которая могла классифицировать ограниченное количество объектов. Предположительно это должно было улучшить производительность и более глубоких сетей.

Рис. 9.1. Джеффри Хинтон и Ян Лекун, освоившие глубокое обучение. Фотография сделана примерно в 2000 году на заседании программы NCAP Канадского института перспективных исследований. Эта программа создала благодатную почву для исследования глубокого обучения, и участники на снимке довольны

своими успехами

В 2000-х годах разработали графические модели, ставшие частью большого потока вероятностных моделей, называемых байесовскими сетями или сетями доверия. В их основу легло уравнение, выведенное Томасом Байесом в XVIII веке, которое позволяло новым доказательствам изменять исходные установки. Джуда Перл из Калифорнийского университета в Лос-Анджелесе ранее представлял сети на основе байесовского анализа[213], и его алгоритм расширили и усовершенствовали разработкой методов для изучения вероятностей. Этот и многие другие найденные алгоритмы создали мощный арсенал, ставший основой для машинного обучения.

Так как вычислительные мощности компьютеров росли по экспоненте, стало возможным обучать более крупные сети. Считалось, что широкие нейронные сети с большим числом скрытых единиц эффективнее, чем глубокие сети с большим количеством слоев, но выяснилось, что это не относится к сетям, которые обучаются слой за слоем[214]. Отчасти причиной была проблема исчезающего градиента ошибки, которая замедляла обучение вблизи входного слоя[215]. Когда ее решили, появились условия для обучения глубоких сетей обратного распространения ошибки, которые показывали прекрасные результаты на тестах[216]. Сети глубокого обучения продемонстрировали, насколько в перспективе может улучшиться качество распознавания речи[217].

Глубокие сети обратного распространения ошибки бросили вызов традиционным подходам к компьютерному зрению. То, что внимание вновь было обращено к нейросетям, подняло шумиху на Конференции NIPS в 2012 году. Джеффри Хинтон и два студента, Алекс Крижевский и Илья Суцкевер, представили доклад о методе распознавания объектов на изображениях, использованный ими для

обучения AlexNet – глубокой сверточной сети, которая будет в центре внимания в этой главе. В области компьютерного зрения последние 20 лет шел устойчивый, но медленный прогресс, и на тестах производительность росла на доли процента в год. Методы улучшались неспешно, поскольку каждая новая категория объектов требует, чтобы эксперт предметной области определил для нее неизменяющиеся признаки, по которым их можно отличить от других объектов.

Важную роль в сопоставлении различных методов играют контрольные показатели. Эталоном, который использовала команда из Университета Торонто, была база данных ImageNet, содержащая свыше 15 миллионов изображений с высоким разрешением более чем в 22 тысячах категорий. AlexNet добилась беспрецедентного снижения частоты ошибок на 18 процентов.[218] Этот скачок производительности поразил специалистов по машинному зрению и задал курс его развития, так что в настоящее время компьютерное зрение почти достигло уровня человеческого. К 2015 году частота ошибок в базе данных ImageNet снизилась до 3,6 процента[219]. Используемую сеть глубокого обучения, во многом напоминающую зрительную кору головного мозга, представил Ян Леку, и первоначально она называлась Le Net.

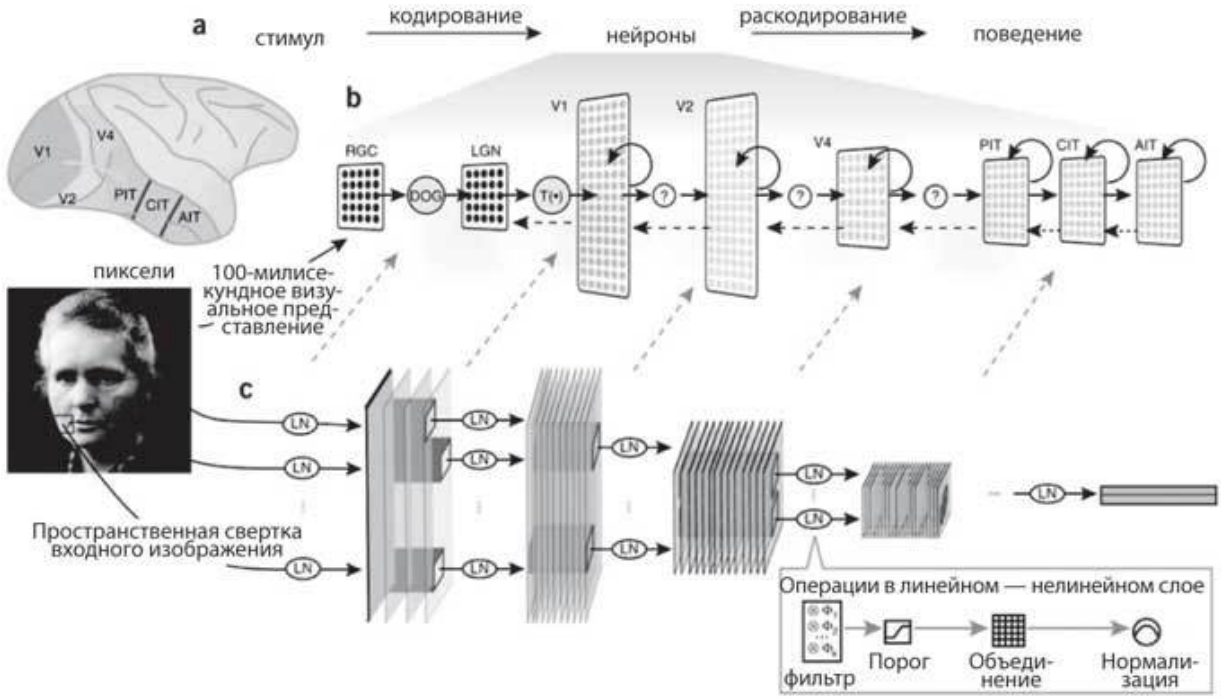


Рис. 9.2. Сравнение зрительной коры и сверточной сети для распознавания объектов на изображениях. Вверху: иерархия слоев зрительной коры, от входов V1 с сетчатки и таламуса (LGN) до нижней височной коры (PIT, CIT, AIT), показывающая соответствие между кортикальными областями и слоями сверточной сети. Внизу: входные данные с изображения слева проецируются на первый сверточный слой, состоящий из нескольких слоев признаков, каждый из которых представляет собой фильтр, как ориентированные простые клетки, найденные в зрительной коре. Фильтры с заданными границами объединяются параллельно первому слою и дают одинаковый отклик на определенном участке, подобно сложным клеткам в зрительной коре. Эта операция повторяется на каждом сверточном слое сети. Выходной слой полностью обменивается данными с последним сверточным слоем. (Yamins DLK, DiCarlo JJ. Using goaldriven deep learning models to understand sensory cortex. Nat. Neurosci. 19: 356–65, 2016)

Ян Лекун (рис. 9.1) был студентом, когда мы с Джеффри Хинтоном впервые встретились с ним в 1980-х годах во Франции. Он заинтересовался ИИ еще в девять лет, вдохновленный HAL 9000 – вымышленным компьютером из фильма «Космическая одиссея 2001 года»[220]. В 1987 году Лекун, когда писал свою кандидатскую диссертацию, самостоятельно выявил метод обратного распространения ошибки[221], после чего переехал в Торонто, чтобы работать с Хинтоном. Позже он перешел в Bell Labs в Холмделе, где обучил сеть читать рукописные почтовые индексы на письмах, используя набор данных MNIST[222] – маркированный эталон из почтового отделения Буффало. Ежедневно приходится направлять в почтовые ящики миллионы писем, и сегодня это полностью автоматизировано. Та же технология позволяет банкоматам считывать сумму на банковском чеке. Интересно, что сложнее всего найти место, где на чеке записаны цифры, так как у каждого чека свой формат. Еще в 1980-х годах было очевидно, что у Лекуна огромный талант брать доказанный учеными принцип и заставлять его работать в реальном мире.

Сверточные нейронные сети

Когда Ян Лекун в 2003 году перешел на работу в Нью-Йоркский университет, он продолжал развивать свою «видящую» сеть, сейчас известную как сверточная нейронная сеть (рис. 9.2). Она основана на свертке, которую можно рассматривать как скользящий фильтр, пропускающий через себя все изображение и создающий параллельно ему новый информационный слой. Например, фильтр, предназначенный для определения контуров, как тот, о котором говорилось в главе 5, имеет большой выходной сигнал только тогда, когда окно находится над краем верно ориентированного объекта на изображении или так же ориентированной текстуры внутри объекта. Окно на первом слое – лишь небольшой фрагмент изображения, при этом может быть много фильтров и, следовательно, много характеристик, представленных в каждом фрагменте. Фильтры в первом слое, который связан с изображением, похожи на то, что Хьюбел и Визель называли простыми клетками первичной зрительной коры (рис. 9.3)[223]. Фильтры в более высоких слоях реагируют на еще более сложные характеристики[224].

В более ранних версиях сверточной сети выходные данные каждого фильтра пропускались через нелинейную функцию, плавно увеличивающуюся от 0 до 1, называемую сигмоидной (S-образной) функцией, которая подавляла вывод слабо активированных блоков. Окно во втором слое, которое получает входные данные от первого слоя, охватывает бóльшую область поля зрения, так что после нескольких слоев появляются блоки, получающие входные данные со всего изображения. Этот верхний слой схож с верхушкой иерархии зрительной коры, которая у приматов называется нижневисочной корой и имеет рецептивные поля, покрывающие значительную часть поля зрения. Затем верхний слой подается в слой классификации, который использовался для обучения всей

сети, чтобы классифицировать объекты на изображении с помощью обратного распространения ошибки.

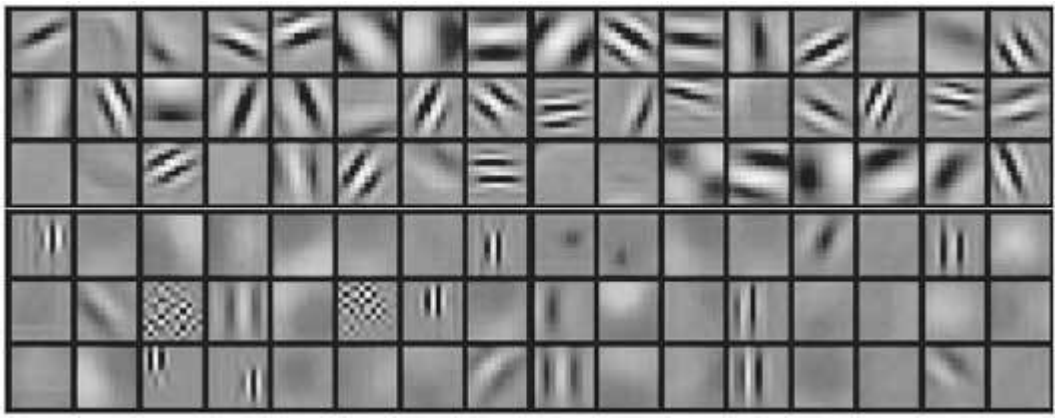


Рис. 9.3. Фильтры из первого слоя сверточной сети. Каждый фильтр расположен на определенном участке в поле зрения. Предпочтительные входные сигналы некоторых фильтров ориентированы как простые клетки в зрительной коре. Предпочтительные входные сигналы на втором слое более вытянутые и имеют сложные формы [Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems 25 (NIPS 2012)]

За прошедшие годы в сверточные сети внесли множество улучшений (см. рис. 9.3). Важным дополнением стало объединение каждого признака по областям, получившее название пулинг (pooling). Оно обеспечивает перенос изображения в неизменном виде и похоже на работу сложных клеток, обнаруженных Хьюбелом и Визелем в первичной зрительной коре (см. рис. 5.5), которые реагируют на линии с одинаковой ориентацией на всем участке поля зрения. Еще одной полезной функцией стала нормализация усиления, регулирующего уровень входных сигналов таким образом, чтобы каждый блок работал в пределах определенного диапазона (в коре это реализовано через подавление обратной связи).

Сигмоидная функция выхода также была заменена блоками линейной ректификации (ReLUs), которые имеют нулевой выход до четкого порогового значения и линейно увеличиваются выше него. Элементы ниже порогового значения эффективно исключаются из сети, что ближе к тому, как работает пороговое значение у настоящего нейрона.

С тех пор как сверточные сети были усовершенствованы, они стали больше напоминать архитектуру зрительной коры, какой ее представляли в 1960-х годах. Однако все изменения имели математическое обоснования и эффективность повысили способами, которые способен понять инженер, в то время как в 1960-х годах мы могли только догадываться о функциях простых и сложных клеток или о том, для чего нужно распределенное представление на вершине иерархии. Это пример того, как много могут дать тесные взаимосвязи между биологией и глубоким обучением.

Столкновение глубокого обучения и визуальной иерархии

Патриция Черчленд, философ из Калифорнийского университета в Сан-Диего, специализируется на нейрофилософии[225]. Например, эпистемология изучает научное знание и, в конечном счете, зависит от того, как знание представляет мозг. Конечно, это не мешает философам думать о знании как, говоря словами Иммануила Канта, о «вещи в себе», независимо от мира. Однако даже животным для начала необходимо обоснованное знание, чтобы найти безопасное место и выжить. В 1992 году мы с Патрицией Черчленд написали книгу «Вычислительный мозг»[226] для разработки системы понятий в нейробиологии, основанной на больших совокупностях нейронов. Нас вдохновило поразительное сходство в схемах работы скрытых элементов обученной многослойной сети и групп нейронов, записанных поочередно. Эта книга выдержала переиздание и является хорошим учебником для тех, кто хочет больше узнать, как обрабатываются данные в мозге. Джеймс Дикарло из МТИ недавно

сравнил отклик нейронов на разных уровнях иерархии зрительной коры обезьян, обученных распознавать изображения объектов, с откликов элементов в сети глубокого обучения, обученной распознавать те же изображения (см. рис. 9.2)[227]. Он пришел к выводу, что статистические свойства нейронов в каждом слое сети глубокого обучения вполне соответствуют свойствам нейронов в кортикальной иерархии.

Сходство между свойствами элементов в сети глубокого обучения и свойствами нейронов в зрительной коре, характеристиками сетей и характеристиками человеческого мозга – загадка, особенно учитывая то, что мозг вряд ли будет использовать для обучения метод обратного распространения ошибки. Обратное распространение ошибки требует, чтобы на сигналы об ошибке каждый нейрон в каждом слое откликался с гораздо большей точностью, чем в уже известных обратных связях. Тем не менее другие алгоритмы обучения более биологически правдоподобны – например, алгоритм машины Больцмана, использующий синаптическую пластичность Хебба, найденную в коре. Это ставит интересный вопрос: существует ли математическая теория глубокого обучения, применимая к большому классу алгоритмов обучения, в том числе к алгоритмам в коре мозга? Я уже упоминал о разделении поверхностей на категории на верхних уровнях иерархии в главе 7, где поверхности принятия решений более плоские, чем в нижних слоях. Геометрический анализ поверхности принятия решений может углубить математическое понимание как сетей глубокого обучения, так и мозга.

Одно из преимуществ сети глубокого обучения – мы можем фиксировать данные из каждого блока в сети и следить за потоком информации по мере ее перехода из слоя в слой. Подход к анализу искусственных сетей позже можно будет применять для анализа нейронов в головном мозге. Одна из потрясающих особенностей этой технологии в том, что за ней обычно стоит хорошее объяснение, подталкивающее во всем разобраться. Первые паровые

машины инженеры создавали, опираясь лишь на интуицию: теория термодинамики, которая объяснила, как те работают, появилась позже, одновременно с повышением эффективности двигателей. Анализ сетей глубокого обучения физиками и математиками идет полным ходом.

Рабочая память и долговременное хранение данных

Нейробиология прошла долгий путь с 1960-х годов, и сегодня у нас гораздо больше знаний о мозге. В 1990 году Патриция Голдман-Ракич научила обезьяну запоминать место, которое освещалось на короткое время, и поглядывать на него после периода задержки[228]. Записывая сигналы из префронтальной коры мозга, Голдман-Ракич сообщила, что некоторые нейроны, первыми отвечающие на сигнал, сохраняли свою активность весь период задержки. В психологии это называется рабочей памятью, благодаря ей мы можем держать в уме от четырех до десяти элементов в процессе выполнения задачи, например когда набираем номер телефона.

Традиционная сеть прямого распространения передает входные сигналы вверх по сети по одному слою за раз. Включение рабочей памяти позволит данным, поступившим позже, взаимодействовать со следом, оставшимся после данных, введенных в сеть ранее. Например, при переводе предложения с французского языка на английский первое введенное французское слово влияет на порядок английских слов. Самый простой способ реализовать рабочую память в сети – добавить рекуррентные соединения, которые типичны для коры головного мозга. Повторяющиеся связи внутри слоя и обратные связи с предыдущими слоями позволяют ненадолго группироваться входным данным, поступившим в разное время. Такие сети были изучены в 1980-х годах и широко используются для распознавания речи[229]. На практике это хорошо работает для краткосрочных взаимосвязей, но плохо – когда разрыв между входами велик, так как влияние входа ослабевает со временем.

В 1997 году Зепп Хохрайтер и Юрген Шмидхубер нашли способ преодолеть проблему распада, который они назвали сетью долгой краткосрочной памяти (Long short-term memory; LSTM)[230]. Основная идея в том, чтобы передать информацию в будущее без потерь, как в период задержки в префронтальной коре мозга

обезьяны. В сети LSTM есть сложная схема для принятия решения, как именно объединять новую входящую информацию со старой. Как следствие, растянутые во времени взаимосвязи сохраняются выборочно. Эта версия рабочей памяти не использовалась в течение 20 лет, пока не была воскрешена и реализована в сетях глубокого обучения, где оказалась очень успешной во многих областях, зависящих от последовательности обучения входов и выходов, включая видеоролики, музыку, движения и речь.

Шмидхубер – один из руководителей Института исследований искусственного интеллекта Далле Молле в Манно, крошечном городке в районе Тичино на юге Швейцарии, недалеко от лучших туристических троп в Альпах[231]. Он как Родни Дейнджерфилд[232] в области нейронных сетей. Он изобретательный и единственный в своем роде, однако не считает, что его достаточно уважают: на конференции NIPS в 2015 году в Монреале он представился из зала как «снова ты, Шмидхубер».

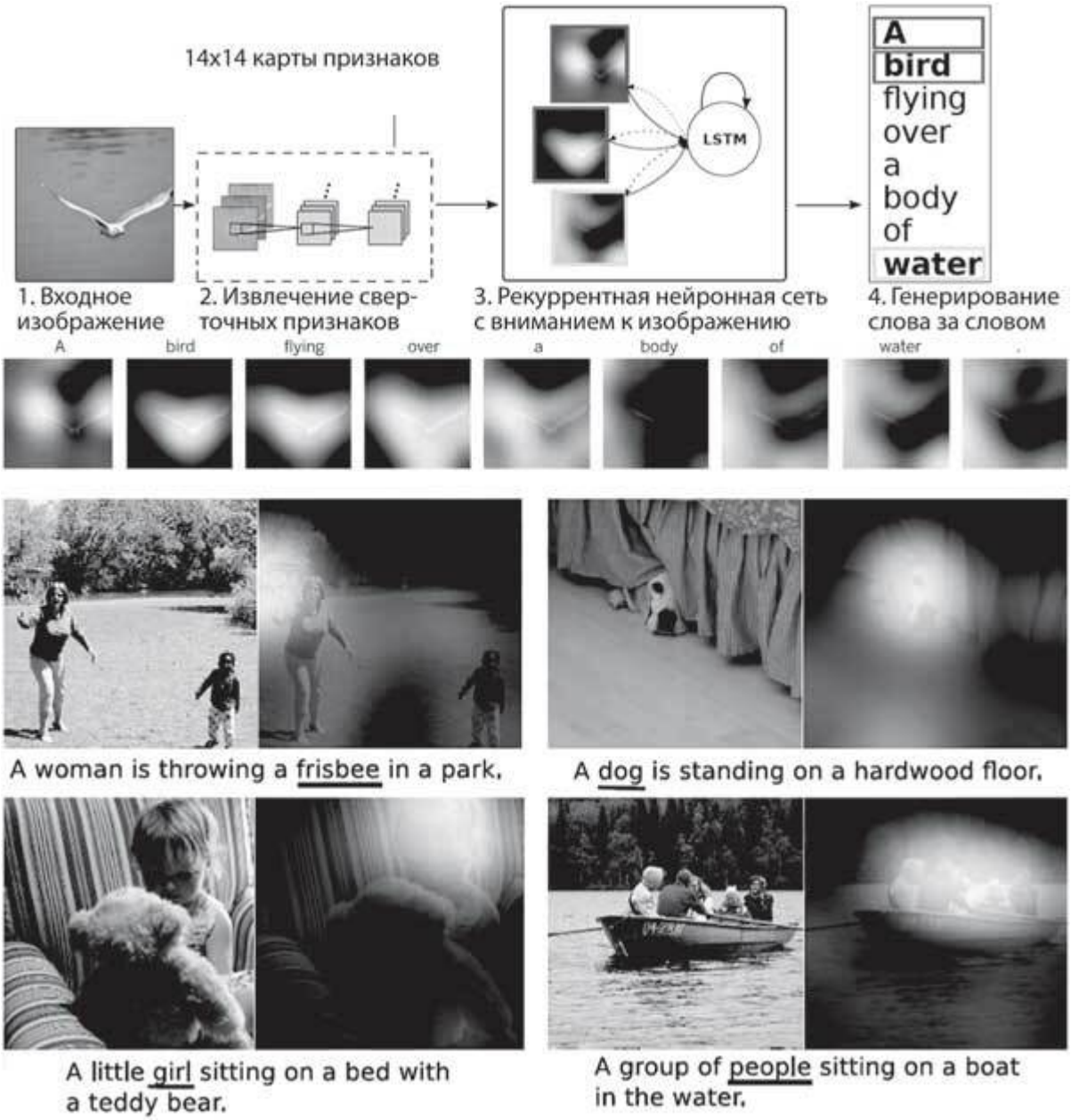


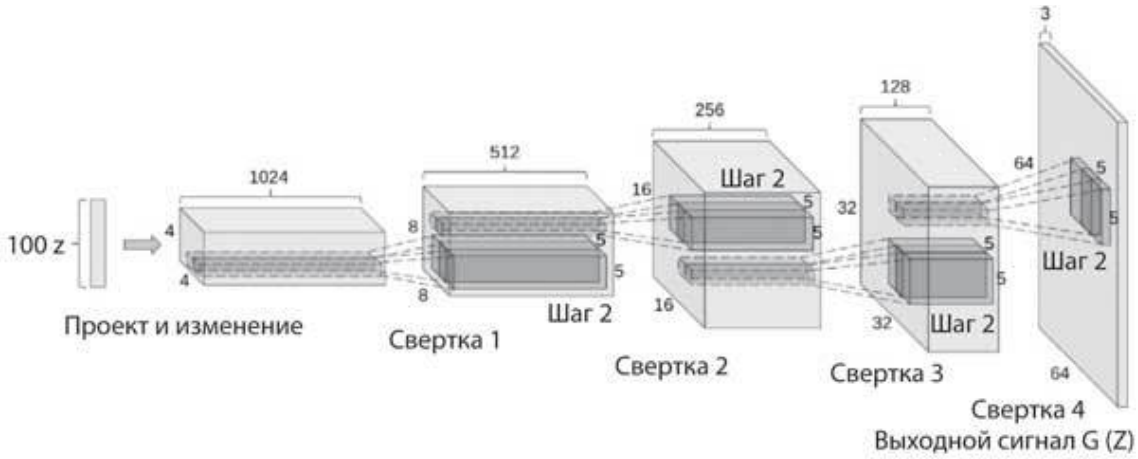
Рис. 9.4. Подписи к изображениям глубокого обучения. Верхний ряд иллюстрирует процедуру анализа фотографии. На первом этапе сверточная нейронная сеть помечает объекты на фотографии и передает их рекуррентной нейронной сети. Рекуррентная сеть была обучена выводить соответствующую строку английских слов. Четыре картинки внизу иллюстрируют дальнейшее уточнение, которое использует внимание (белое облако), чтобы соотнести слова и фотографии (arxiv.org/abs/1502.03044, 2015)

На конференции в 2016 году в Барселоне он пять минут донимал спикера, который не уделил достаточного внимания его идеям. В этом весь Шмидхубер. В 2015 году сеть глубокого обучения для распознавания объектов на изображениях была объединена с сетью LSTM для подписи изображений. Входной сигнал в сеть LSTM проходит первичную обработку в сети глубокого обучения, которая определяет все объекты на изображении. Сеть LSTM была обучена выводить строку английских слов, которые описывают изображение (рис. 9.4), а также определять место на изображении, соответствующее слову. Впечатляющим это приложение делает то, что сеть LSTM никогда не обучали понимать смысл предложения – только выводить синтаксически правильную строку слов на основе объектов и их расположения на рисунке. Вместе с NETtalk, приведенной в примере в главе 8, это еще раз доказывает, что нейросети, похоже, связаны с речью, хотя мы пока не понимаем почему. Возможно, в результате анализа сетей LSTM появится новая теория языка, которая прольет свет как на работу нейросетей, так и природу естественного языка.

Порождающие состязательные сети[233]

В главе 7 машина Больцмана была представлена как порождающая модель, умеющая создавать новые входные выборки, когда выходные данные ограничены категорией, которую она обучена распознавать, и шаблоны активности проникают на входной слой. Йошуа Бенджио и его коллеги из Университета Монреаля показали, что можно обучить сети прямого распространения создавать еще более качественные образцы в обстоятельствах состязания[234]. Порождающую сверточную сеть может научить синтезировать хорошие примеры изображений в попытке обмануть другую сверточную сеть, которая должна решить, являются ли входные данные настоящим изображением или поддельным

(рис. 9.5). Выходные данные порождающей сети поступают как входные данные дискриминантной сверточной сети, обученной выдавать один выходной сигнал: 1, если входные данные – реальное изображение, и 0 – если поддельное. Эти две сети конкурируют друг с другом. Порождающая сеть пытается увеличить частоту ошибок дискриминантной сети, которая пытается их уменьшить. Конфликт между двумя целями создает удивительные фотореалистичные изображения (см. рис. 9.5).



Травник Муравей Монастырь

Вулкан

Рис. 9.5. Порождающие состязательные сети. Вверху: сверточная сеть используется для создания выборки изображений, предназначенных обманывать дискриминативную сеть. Входные данные слева – 100-мерные непрерывные векторы, выбранные случайным образом для генерирования различных изображений. Затем входной вектор активирует слои фильтров, все больше увеличивая пространственный масштаб. Внизу: пример изображений, созданных GAN, после обучения на фотографиях из одной категории. [Alec Radford, Luke Metz, Soumith Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, arXiv:1511.06434, 2016.]

Имейте в виду, что эти изображения искусственные и объекты на них никогда не существовали. Они являются обобщенными версиями непомеченных изображений в обучающем наборе. Обратите внимание, что порождающие состязательные сети (Generative adversarial networks; GAN) неконтролируемы, что позволяет им использовать неограниченное количество данных. Есть много других приложений, начиная от приложений для удаления шумов на астрономических изображениях галактик со сверхразрешением[235] до приложений для изучения представлений эмоциональной речи[236].

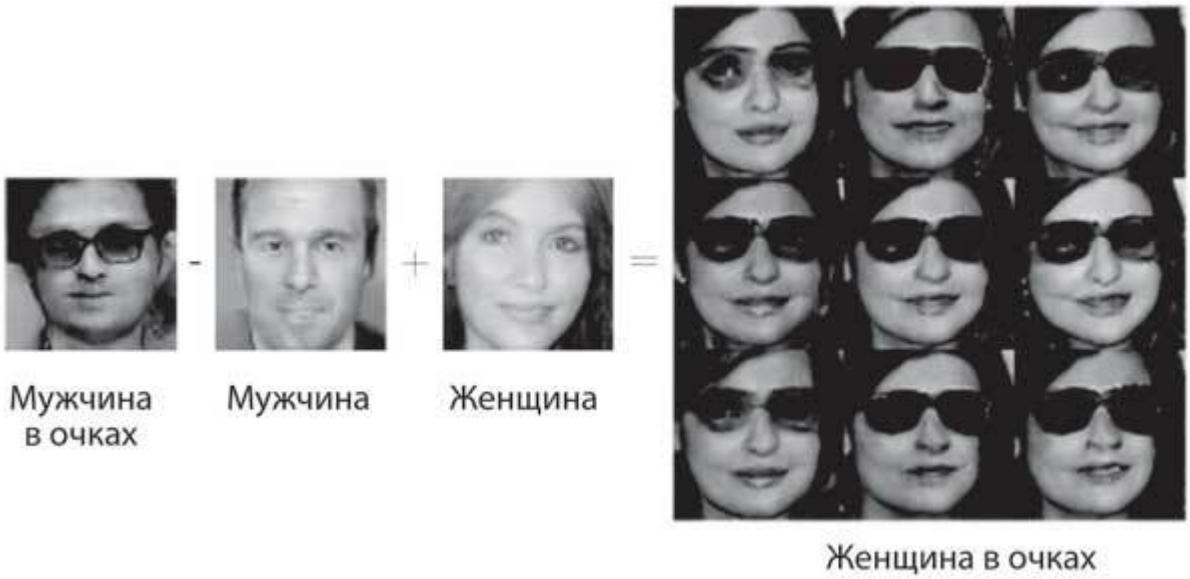


Рис. 9.6. Векторная арифметика в порождающих состязательных сетях: комбинация входных данных в порождающую сеть, обученную на лицах, превращалась в выходные данные слева, которые затем, складывая и вычитая выбранные входные векторы, использовались для создания комбинации справа. Поскольку смешивание происходит на самом высоком уровне представления, части и позы незаметно для пользователя сочетаются, а не усредняются, как при морфинге[237]. [Alec Radford, Luke Metz, Soumith Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, arXiv:1511.06434, 2016.]

Плавно изменяя входной вектор, можно постепенно сдвигать изображение так, что его части, допустим, окна, постепенно появляются или превращаются в другие объекты, например в шкафы[238],[239]. Еще удивительнее то, что можно складывать и вычитать векторы, представляющие состояние сети, для получения смеси объектов на изображении, как показано на рис. 9.6. Смысл этих экспериментов в том, что представление изображений в порождающей сети показывает комнаты так, как если бы мы описывали части картинки. Технология быстро развивается, и

следующий рубеж – создание реалистичных фильмов. Тренируя порождающую состязательную сеть на фильмах, например, с Мэрилин Монро, можно будет воскресить старые шедевры и создать новые.

Рис. 9.7. Показ мужской одежды от Джорджо Армани в Милане. Коллекция весна-лето 2018

Неделя моды в Милане. Модели с отрешенными лицами вышагивают по подиуму (рис. 9.7). Что-то меняется в мире моды: «Многие рабочие места исчезают, – сказала Сильвия Вентурини-Фенди перед показом. – Андроиды займут прежние рабочие места, но единственное, что они не смогут заменить, – наш разум и умение творить»[240]. Теперь представьте себе порождающие состязательные сети, которые были обучены создавать новые стили и моделировать модную одежду с почти бесконечным разнообразием. Возможно,

мир моды стоит на пороге новой эры, как и многие сферы бизнеса, связанные с творчеством.

Все дело в масштабировании

Большинство современных алгоритмов обучения были открыты более 25 лет назад. Так почему же им потребовалось столько времени, чтобы начать влиять на реальный мир? С компьютерами и размеченными данными[241], которые были доступны исследователям в 1980-х годах, было возможно продемонстрировать только принцип работы в «лабораторных условиях». Несмотря на отдельные многообещающие результаты, мы не знали, насколько хорошо будет масштабироваться сетевое обучение и производительность, так как количество единиц и соединений увеличивалось в соответствии со сложностью реальных проблем. Многие алгоритмы ИИ плохо масштабировались и никогда не выходили за рамки игрушечных задач. Теперь мы знаем, что обучение нейронных сетей хорошо масштабируется, а производительность продолжает расти с размером сети и количеством слоев. Особенно хорошо масштабируется алгоритм обратного распространения ошибки.

Стоит ли удивляться? Кора головного мозга – «изобретение» млекопитающих, она наиболее развита у приматов и особенно у людей. По мере ее разрастания стало доступно больше возможностей и увеличилось число слоев в ассоциативной зоне для представлений более высокого порядка. Сложных систем, которые так же хорошо масштабируются, единицы. Интернет – одна из немногих спроектированных систем, размер которой так же был увеличен в миллион раз. Интернет развивался после того, как были созданы протоколы для передачи сообщений, подобно тому, как генетический код ДНК позволил развиваться клеткам.

Обучение множества глубоких сетей одним и тем же набором данных приводит к появлению большого числа различных сетей с примерно одинаковым средним уровнем производительности. Нам хотелось бы знать, что общего у всех тех одинаково неплохих сетей,

и анализ одной сети этого не выявит. Еще один подход к пониманию принципов, лежащих в основе глубокого обучения, – дальнейшее исследование пространства алгоритмов обучения. Мы отметили лишь несколько точек в этом пространстве. При более широком исследовании теория обучения может выйти столь же основательной, как и теории в других областях науки[242]. Вычислительная теория обучения может пролить свет и на алгоритмы обучения, созданные природой.

Рис. 9.8. Йошуа Бенджио – содиректор программы CIFAR «Обучение машин и мозга». Канадский специалист по вычислениям, родившийся во Франции, Йошуа был лидером в применении глубокого обучения к проблемам живого языка. Достижения Джеффри Хинтона, Яна Лекуна и Йошуа Бенджио стали основополагающими в глубоком обучении

Йошуа Бенджио[243] (рис. 9.8) из Монреальского университета и Ян Лекун сменили Джеффри Хинтона на посту директора программы «Нейронные вычисления и адаптивное восприятия» Канадского

института перспективных исследований (Canadian Institute for Advanced Research; CIFAR)[244], когда она, пройдя десятилетний путь, была переименована в «Обучение в машинах и мозге» («Learning in Machines and Brains»). Бенджио возглавлял команду в Монреальском университете, которая применяла глубокое обучение к естественному языку, что и стало новым направлением для этой программы. На встречах в течение десяти с лишним лет небольшая группа из двух десятков преподавателей и стипендиатов положила начало глубокому обучению. Заметный прогресс в применении глубокого обучения ко многим проблемам, которые ранее казались неразрешимыми, можно проследить до их деятельности, но, конечно, они лишь небольшой частью гораздо большего сообщества, которое будет рассмотрено в главе 11.

Сети глубокого обучения зарекомендовали себя во многих приложениях, но они никогда не выживут в реальном мире самостоятельно[245]. С ними нянчатся исследователи, которые кормят эти сети данными, настраивают свои гиперпараметры, такие как скорость обучения, количество слоев и число единиц в каждом слое, чтобы улучшить сходимость и предоставить им огромные вычислительные ресурсы. Кора головного мозга также не выжила бы в мире без тела и остальной части мозга, которые обеспечивают ей поддержку и автономию. Автономия в изменчивом мире – гораздо более сложная проблема, чем распознавание образов. В следующей главе будет представлен древний алгоритм обучения, который важен для выживания в природе, так как мотивирует нас искать полезный опыт.

Глава 10. Обучение с подкреплением

Согласно легенде, уходящей корнями в Средневековье, изобретателю игры в шахматы предложили пшеничное поле в качестве подарка от благодарного правителя. Вместе поля изобретатель попросил положить одно зерно на первый квадрат, два зерна – на второй, четыре зерна – на третий, и так далее, удваивая количество зерен на каждом последующем квадрате, пока все 64 квадрата на шахматной доске не будут заполнены зерном. Правитель посчитал это скромной просьбой и удовлетворил ее. Но на самом деле правитель отдал все зерно своего королевства, так как количество зерен на 64-м квадрате составляет 2^{64}, и в сумме выходит 18 446 744 073 709 551 615[246] зерен. Это называется экспоненциальным ростом: хоть 64 – небольшое число, такой показатель степени очень велик[247]. Количество позиций в таких настольных играх, как шахматы и го, растет даже быстрее, чем количество зерен пшеницы. На каждый ход в шахматной партии приходится в среднем 35 вариантов, в го – 250. Это делает скорость экспоненциальный рост гораздо выше.

Учим играть в нарды

Преимущество игр в том, что правила в них четко определены, а решения не столь сложны, как в реальном мире, но достаточно сложны, чтобы заставить людей соревноваться. В 1959 году Артур Самюэль, один из первопроходцев машинного обучения из компании IBM, написал программу, которая могла играть в шашки так хорошо, что в тот день, когда это было объявлено, акции IBM сильно подорожали. Шашки – относительно простая игра, но программа Самюэля оказалась впечатляющей для своего времени, учитывая, что он запустил ее на первом коммерческом компьютере IBM – IBM 701, – который работал еще на электронно-лучевых трубках. Программа была основана на функции стоимости, оценивающей сильные стороны различных игровых позиций, так же, как и предыдущие игровые программы, но ее отличало то, что она это освоила на собственном игровом опыте.

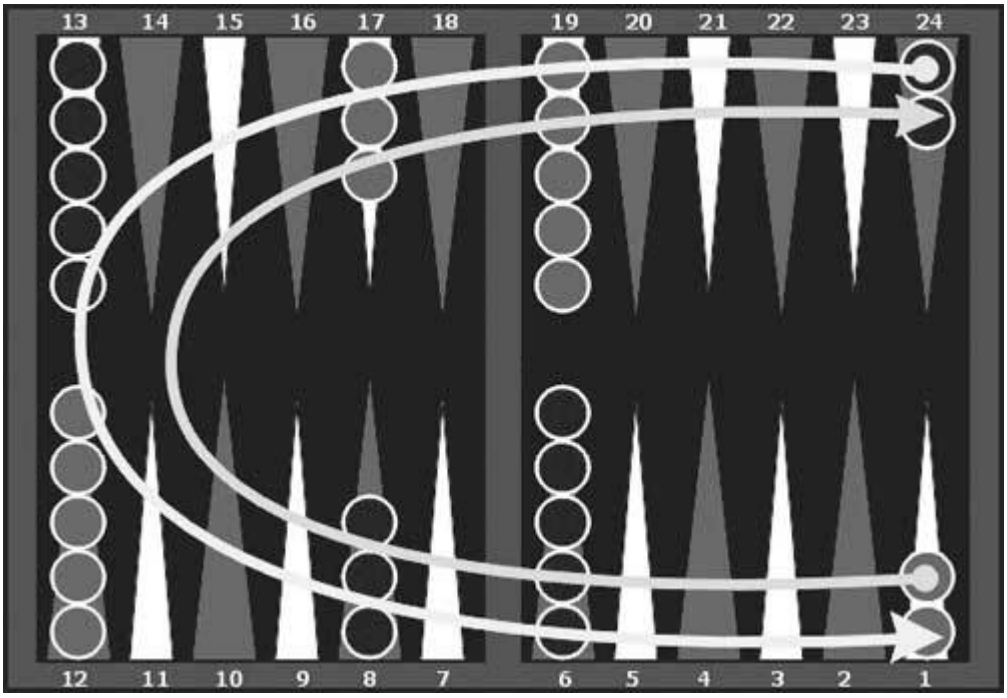


Рис. 10.1. Доска для игры в нарды. Нарды – это гонка до финиша, красные фишки движутся в направлении, противоположном направлению, в котором движутся черные (см. стрелки). Показана начальная позиция. Бросают два кубика, и числа на них указывают, на сколько шагов фишки передвигаются вперед

Джерри Тезауро перешел в исследовательский центр IBM имени Томаса Уотсона после того, как работал со мной над проблемой обучения нейронной сети игре в нарды[248]. При обучения сетей с обратным распространением ошибки для оценки игровых позиций и возможных ходов мы использовали экспертный контроль (рис. 10.1). Недостаток такого подхода в том, что требуется много экспертных оценок позиций, и программа никогда не стала бы лучше наших специалистов, которые играли далеко не на уровне чемпионов мира. Но при игре с самой собой она могла бы добиться большего. Проблема игры с самой собой в том, что единственный обучающий сигнал – победа или поражение в конце партии. Но если одна сторона выиграла, то какой из многих ходов был решающим? Это называется временно́й задачей присваивания коэффициентов доверия.

Алгоритм обучения, который может решить временную задачу присваивания коэффициентов доверия, изобрел Ричард Саттон из Массачусетского университета в Амхерсте в 1988 году[249]. Он тесно сотрудничал с Эндрю Барто, своим научным руководителем, в работе над сложными проблемами в обучении с подкреплением – отрасли машинного обучения, вдохновленной ассоциативным обучением в экспериментах на животных (рис. 10.2). В отличие от сети глубокого обучения, единственной задачей которой является преобразование входных данных в выходные, система усиления взаимодействует с окружающей средой в замкнутом цикле, получая входную сенсорную информацию, принимая решения и осуществляя действия, которые влияют на мир. Обучение с подкреплением основано на наблюдении за животными, которые решают сложные проблемы в изменчивой среде после того, как исследовали различные варианты и их результаты. По мере того как обучение улучшается, исследование уменьшается, что в конечном итоге приводит к чистому использованию лучшей стратегии, найденной во время обучения.

Рис. 10.2. Сценарий обучения с подкреплением. Агент активно исследует окружающую среду, предпринимая действия и делая наблюдения. Если действие выполнено успешно, агент получает вознаграждение. Цель в том, чтобы принять меры, которые максимизируют будущие выгоды

Предположим, вам нужно принять ряд решений для достижения цели. Если вы уже знаете все возможные варианты и ожидаемые будущие результаты[250], вы можете использовать поисковый алгоритм, чтобы выяснить набор вариантов, при котором выгода максимальна, но из-за этого размер задачи увеличивается по экспоненте – так называемое проклятие размерности. Но если у вас изначально нет всей информации о результатах выбора, вы должны научиться делать его по мере продвижения вперед. Это называется обучением в реальном времени.

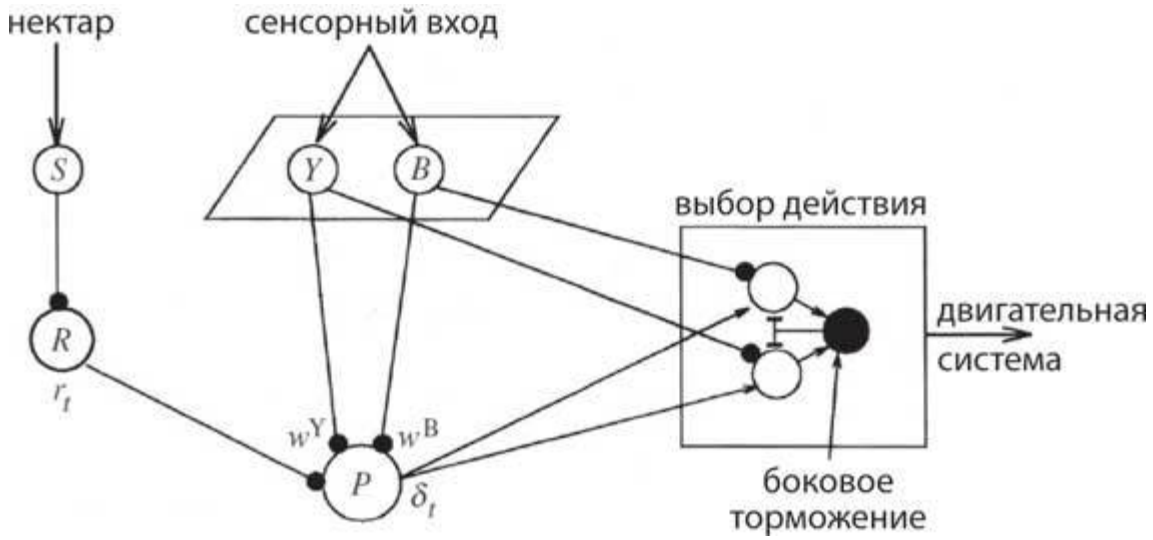
Рис. 10.3. Ричард Саттон из Альбертского университета в Эдмонтоне в Канаде научил нас, как узнать путь к будущим наградам. Саттон перенес рак, но он остается лидером в обучении с подкреплением и продолжает разрабатывать инновационные алгоритмы[251]. Он щедр на свое время и идеи, которые каждый в этой области очень ценит. Написанная им в соавторстве с Энди Барто книга «Обучение с подкреплением»[252] стала классическим трудом

Алгоритм обучения в реальном времени, разработанный Ричардом Саттоном (рис. 10.3), зависел от различий между ожидаемым и полученным вознаграждением (блок 6). В обучении с учетом временной разницы вы сравниваете свою оценку предполагаемой долговременной награды за совершенный шаг в текущей позиции с лучшей, по статистике, оценкой награды, которую вы на самом деле получили, и предполагаемой награды после следующего шага. Если изменять предыдущую оценку так, чтобы она

была ближе к улучшенной, решения, которые вы принимаете по мере продвижения, будут становиться все лучше и лучше. Изменения заставляют оценочную сеть учитывать будущее ожидаемое вознаграждение для каждой позиции на доске и использовать для принятия решения о следующем шаге. Алгоритм временной разности сходится к оптимальному правилу принятия решений в заданном состоянии после того, как у вас будет достаточно времени, чтобы изучить возможности.

Программа Джерри, названная TD-Gammon, знала важные особенности доски и правила игры, но не знала, что такое хороший ход. В начале обучения ходы были случайными, но в конце концов одна из сторон выигрывала и получала финальное вознаграждение. В нардах побеждает тот игрок, который первым «снимет» все фишки с игрового поля.

Блок 6. Обучение методом временной разницы



В этой модели мозга медоносной пчелы выбираются действия (например, приземлиться на цветок), которые максимизируют будущие награды:

$R(t) = r_{t+1} + \gamma\, r_{t+2} + \gamma\, 2\, r_{t+3} + \ldots,$

где r_{t+1} – вознаграждение в момент времени *t+1*, а $0 < \gamma < 1$ – коэффициент обесценивания. Предсказанное будущее вознаграждение, основанное на текущих сенсорных входах s(t), вычисляется нейроном P:

$P_t(s) = w^y s^y + w^b s^b,$

где сенсорный ввод от желтых (Y) и синих (B) цветов взвешивается по w^y и w^b. Погрешность прогноза вознаграждения δ (t) в момент времени t определяется:

$\delta_t = r_t + {}_{\gamma} P_t(s_t) - P_t(s_{t-1}),$

где r_t – текущее вознаграждение. Изменение каждого веса определяется:

$\delta\, w_t = \alpha \delta_{\,t}\, s_{t-1},$

где α – скорость обучения. Если вознаграждение больше, чем предсказанное вознаграждение, и δ_t положительна, вес увеличивается на сенсорном входе, который присутствовал до вознаграждения, но если вознаграждение меньше, чем ожидалось, а δ_t отрицательна, вес уменьшается.

Поскольку единственное реальное вознаграждение появляется в конце игры, логично ожидать, что программа TD-Gammon сначала изучит конец игры, затем середину и, наконец, ее начало. Это как раз то, что происходит в табличном обучении с подкреплением, где есть таблица значений для каждого состояния в пространстве состояний. Однако с нейронными сетями все иначе – они быстро хватаются за простые и надежные сигналы входных функций, а более сложные и сомнительные входные сигналы оставляют на потом. Первый принцип, который изучает TD-Gammon, – «выбрасывать фишки», придавая положительный вес входному элементу, который представляет собой количество снятых с доски фишек. Второй принцип – «блокировать фишки противника» –

довольно эффективный способ практического решения проблемы на всех этапах, выученный путем присвоения положительного веса входному блоку, отмечающему количество заблокированных фишек противника. Третий принцип – «избегать блокировки» – естественная реакция на второй, и он изучается через придание отрицательного веса отдельным фишкам, которые могут быть заблокированы. Четвертому принципу – «занимать новые лунки», блокируя продвижение противника, – учат, назначая положительные веса уже занятым точкам. Для закрепления этих базовых принципов требуется несколько тысяч обучающих игр. За десять тысяч игр TD-Gammon изучила основные принципы. За сто тысяч – освоила продвинутый подход, а к миллиону игр ее методы достигли уровня чемпионов мира или вообще находились за пределами знаний людей начала 1990-х годов.

Когда в 1992 году TD-Gammon была представлена миру, она впечатлила и меня, и многих других[253]. Функция стоимости представляла собой сеть обратного распространения ошибки с 80 скрытыми единицами. После 300 тысяч игр программа начала обыгрывать Джерри, поэтому он связался с известным игроком в нарды и автором книг о них Биллом Роберти и пригласил его посетить IBM, чтобы сыграть с TD-Gammon. Роберти выиграл в большинстве случаев, но, к своему удивлению, проиграл несколько хороших партий и заявил, что это лучшая программа для игры в нарды, с которой он когда-либо состязался. Некоторые из ее ходов были необычными, которые он никогда ранее не видел, и при ближайшем рассмотрении оказалось, что это улучшило игру человека. Роберти вернулся, когда программа сыграла сама с собой 1,5 миллиона партий, и был поражен, когда их встреча с TD-Gammon закончилась вничью. Программа стала настолько лучше, что, по его ощущениям, достигла уровня чемпионов. Специалист по нардам Кит Вулси заметил, что выбор «безопасной» (с низкими рисками и высокой вероятностью награды) или «смелой» (с высокими рисками и также большой вероятностью награды) стратегии игры у TD-

Gammon лучше, чем у любого человека. Может показаться, что 1,5 миллиона обучающих игр – это очень много, но программа узнала из них лишь малую часть из ста квинтильонов (100 000 000 000 000 000 000) возможных позиций на доске, что требовало от TD-Gammon обобщения для новых позиций почти на каждом ходу.

TD-Gammon не получила такой широкой известности, как суперкомпьютер Deep Blue от IBM, который в 1997 году обыграл Гарри Каспарова в шахматы. Шахматы намного сложнее, а Каспаров в то время был чемпионом мира. Однако в некотором смысле TD-Gammon была более впечатляющим достижением. Во-первых, Deep Blue использовали специальное оборудование, чтобы просчитывать большее количество ходов, чем любой человек, побеждая «грубой силой»[254]. Для сравнения, TD-Gammon научилась играть, используя распознавание образов – стиль, более похожий на то, как играют люди. Во-вторых, TD-Gammon проявляла изобретательность и придумывала хитрые стратегии и позиционную игру, которые раньше никто не видел, и тем самым подняла уровень человеческой игры. Это достижение стало переломным в истории ИИ, потому что мы узнали что-то новое от программы, которая сама освоила сложную стратегию в хорошо изученной области, что достойно человеческого интереса и усилий.

Обучение мозга методом вознаграждения

В основе TD-Gammon лежит метод временной разности, который был вдохновлен обучающими экспериментами с животными. Почти все виды, которые были протестированы, от пчел до людей, способны к ассоциативному обучению, как собака Павлова. В эксперименте Павлова после сенсорного раздражителя – звука колокольчика – собаке давали еду, что вызывало у той слюноотделение. После нескольких тренировок звон колокольчика сам по себе стал приводить к образованию слюны. У разных видов разные предпочтительные безусловные стимулы. Пчелы очень хорошо ассоциируют запах, цвет и форму цветка с полезным нектаром и используют эту выученную связь, чтобы искать похожие цветы. Что-то в этой универсальной форме обучения было важным, и в 1960-х годах психологи интенсивно изучали условия, которые привели к появлению ассоциативного обучения, и разрабатывали модели для его объяснения. Бихевиористы[255], такие как Беррес Фредерик Скиннер, обучили голубей распознавать людей на фотографиях. Похоже на то, что можно сделать с помощью глубокого обучения, но есть большая разница: обучение с использованием метода обратного распространения ошибки требует подробной обратной связи со всеми единицами выходного слоя, но ассоциативное обучение дает только один сигнал вознаграждения – правильно или неправильно.

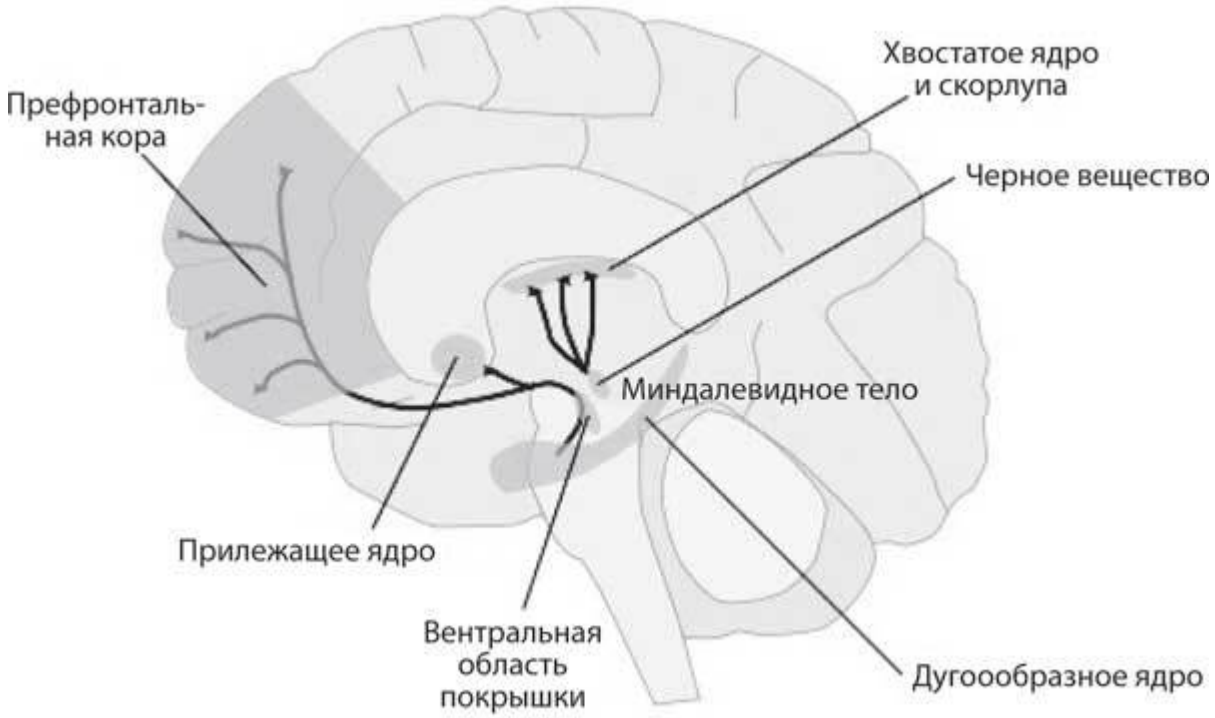


Рис. 10.4. Дофаминовые нейроны в человеческом мозге. Несколько ядер среднего мозга (овальные области, заполненные точками) проецируют аксоны в кору и базальные ганглии (подкорковые ядра). Временные всплески означают расхождения между ожидаемым и полученным вознаграждением, которые используются для выбора действий и изменения прогнозов

Только стимул, возникающий непосредственно перед вознаграждением, ассоциируется с вознаграждением. Это имеет смысл, потому что стимул с большей вероятностью вызвал вознаграждение, если он предшествовал вознаграждению, а не шел после него. Причинно-следственная связь – важный закон природы. Обратное происходит, когда условный стимул сопровождается наказанием, например ударом ноги, и животное учится избегать раздражителя. В некоторых случаях разрыв между условным стимулом и наказанием может быть довольно большим. В 1950-х годах Джон Гарсия показал, что, если крысу кормить подслащенной водой и затем через несколько часов вызывать рвоту, крыса

начинает избегать подслащенной воды в последующие дни. Это называется условное отвращение ко вкусу, и у людей оно работает так же[256]. Например, порой болезнь может ассоциироваться с неудачным приемом пищи, например с шоколадом, который съели в то время. Возникающее в результате отвращение может сохраняться годами, даже если умом вы понимаете, что проблема не в шоколаде.

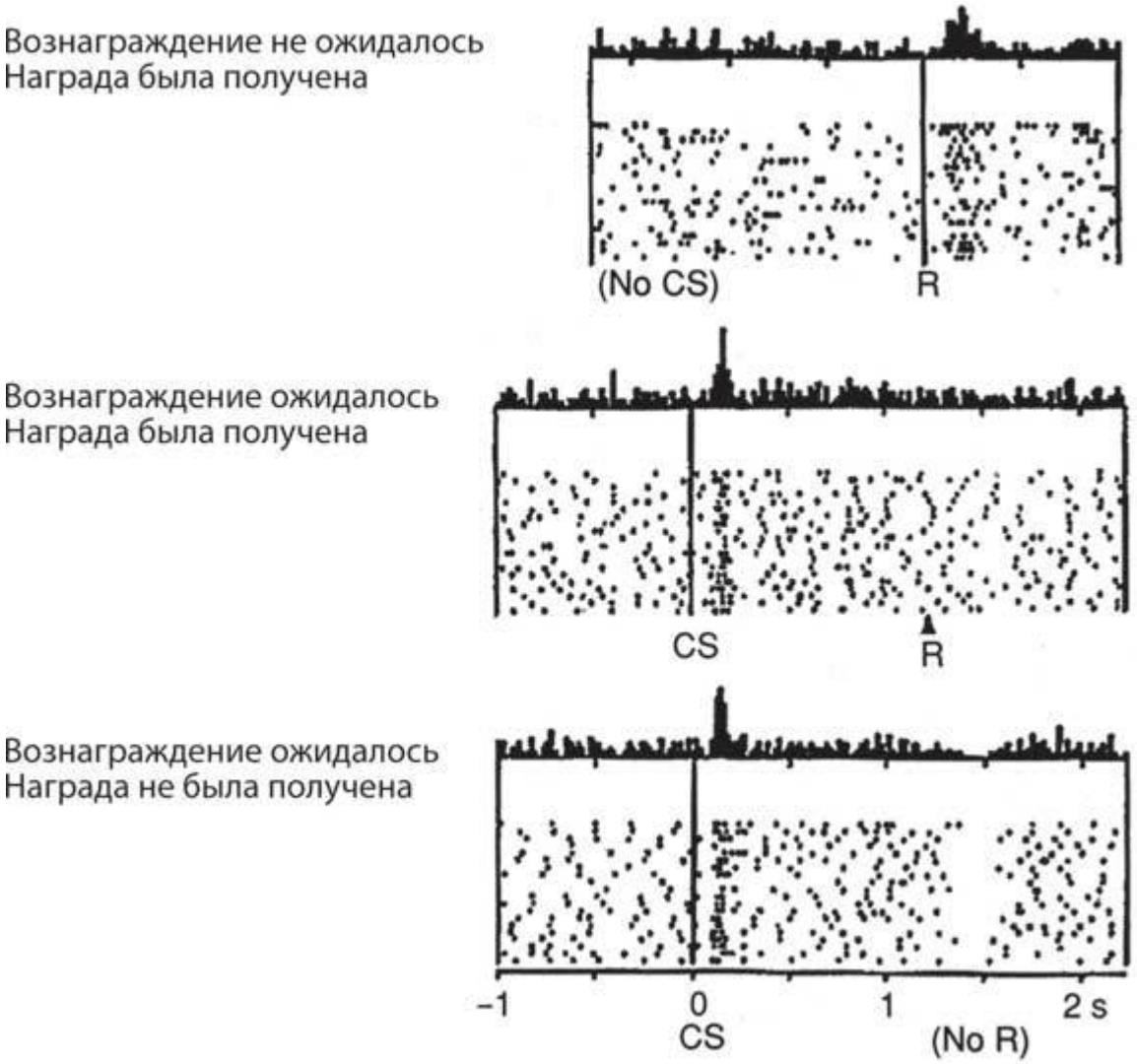


Рис. 10.5. Ответ дофаминового нейрона в мозге обезьяны, подтверждающий, что он сигнализирует об ошибке предсказания вознаграждения остальной части мозга. Каждая точка – всплеск в дофаминовых нейронах. Каждая строка – одна попытка обучения.

Количество пиков в каждой временной ячейке отображается в верхней части каждого растра. Верхнее изображение: в начале обучения награда неожиданная, и дофамин раз за разом запускает всплеск импульсов вскоре после награды. Среднее изображение: после многих попыток, когда свет (CS) неоднократно мигает перед получением вознаграждения, клетка дофамина реагирует на свет, но не на вознаграждение. Согласно временной разнице в обучении, ответ после награды отменяется предсказанием награды. Нижнее изображение: когда в качестве эксперимента вознаграждение было удержано, обнаружилось падение активности в ожидании награды

Дофамин, нейромодулятор, содержащий набор диффузно проецирующихся нейронов в стволе мозга (рис. 10.4), уже давно ассоциировался с обучением методом вознаграждения, но не было известно, что за сигналы они передают в коре. Питер Даян и Рид Монтегю, будучи постдокторантами в моей лаборатории в 1990-х годах, поняли, что дофаминовые нейроны могут реализовать обучение с учетом временной разницы[257]. Эти модели и их предсказания были опубликованы в один из самых захватывающих научных периодов моей жизни и впоследствии подтверждены на обезьянах Вольфрамом Шульцом, сделавшим запись единичных нейронов[258], и на людях с помощью визуализации мозга (рис. 10.5) [259]. В настоящее время установлено, что переходные изменения в активности дофаминовых нейронов сигнализируют об ошибке прогнозирования вознаграждения.

Мы добились прогресса в исследовании ошибки предсказания вознаграждения у приматов, когда в 1992 году я посетил в Берлине Рандольфа Менцеля, изучавшего быстрое обучение в мозге пчелы. Пчелы – лучшие ученики в мире насекомых. Пчеле нужно всего несколько раз посетить необходимый цветок, чтобы запомнить его. В мозге пчелы около миллиона крошечных нейронов, и из-за размеров их трудно регистрировать. Группа Менцеля обнаружила

уникальный нейрон, названный VUMmx1, который реагировал на сахарозу, но не на запах, однако если после появления запаха сразу давали сахарозу, через какое-то время этот нейрон начинал реагировать на запах[260]. Дофаминовая модель обучения методом временной разницы может быть реализована одним нейроном в мозге пчелы. VUMmx1 высвобождает октопамин – нейромодулятор, химически близкий к дофамину. Наша модель обучения пчелы может объяснить некоторые нюансы психологии пчелы, такие как неприятие риска[261]: если у пчелы есть выбор между постоянной и удвоенной наградой, пчелы выберут постоянную награду[262].

Мотивация и базальные ганглии

Дофаминовые нейроны являются основной системой, контролирующей мотивацию в головном мозге (см. рис. 10.4). Все вызывающие привыкание препараты действуют за счет повышения уровня дофаминовой активности. Когда умирает достаточно много дофаминовых нейронов, появляются симптомы болезни Паркинсона, включая дрожание конечностей, затрудненные движения и в конце концов ангедонию – потерю удовольствия от любой деятельности, которая заканчивается кататонией – полным отсутствием движения и эмоциональной реакции. Но в норме дофаминовые клетки обеспечивают кратковременные выбросы дофамина в кору и другие области мозга при получении неожиданного вознаграждения, а снижают свою активность, если полученная награда меньше ожидаемой. Это характерные особенности алгоритма обучения с учетом временной разницы (см. рис. 10.5).

Когда вам нужно принять решение, вы задаете вопрос своим дофаминовым нейронам. Что выбрать из меню? Вы представляете каждый пункт, и дофаминовые клетки оценивают предполагаемое вознаграждение. Должен ли я вступить в брак с этим человеком? Дофаминовые клетки с большей вероятностью дадут верный ответ, нежели рассуждения. Сложнее всего решать проблемы со множеством характеристик, не поддающихся измерению. Что перевесит: положительные качества партнера, такие как хорошее чувство юмора, или плохие качества, например неопрятность? При выборе супруга вы делаете сотни таких сравнений. Все эти рассуждения система вознаграждения сводит к единой «валюте» – кратковременным дофаминовым сигналам.

В алгоритме обучения с учетом временной разницы есть два параметра: скорость обучения α и коэффициент обесценивания γ (блок 6). У пчел высокая скорость обучения, и они могут научиться ассоциировать цветок с наградой после одного посещения. Скорость

обучения у млекопитающих, которым обычно требуется много попыток, ниже. Коэффициент обесценивания также варьируется в широком диапазоне. Когда $\gamma = 0$, алгоритм жаден и решения принимаются только на основе немедленного вознаграждения; но когда $\gamma = 1$, вес всех будущих наград одинаков. В классическом эксперименте маленьким детям предоставили выбор: либо съесть зефир сразу, либо подождать 15 минут, чтобы получить дополнительную порцию зефира[263]. Возраст был важным фактором, и дети помладше не могли откладывать получение удовольствия. Ожидание большого вознаграждения в отдаленном будущем может привести к принятию решений с отрицательным вознаграждением в краткосрочной перспективе для достижения долгосрочной цели. Я вспоминаю об этом, когда учу студентов, которые большую часть своей жизни ходили в школу. Когда я был молод, мать говорила мне, что если я буду хорошим мальчиком, то получу свою награду на небесах – высшая мера отложенного вознаграждения.

Нейроны дофамина получают входные сигналы от части мозга, называемой базальными ганглиями (см. рис. 10.4), которые, как известно, важны для последовательного обучения и формирования привычного поведения. В нейроны в полосатом теле базальных ганглий приходят входные сигналы от всей коры мозга. Входные сигналы от задней половины коры больше связаны с изучением последовательности движений, необходимых для достижения цели. Входные сигналы от префронтальной коры – с планированием последовательности действий. Путь от коры до базальных ганглий и обратно занимает 100 миллисекунд, информация проходит по кругу 10 раз за секунду. Это позволяет принимать быстрые решения одно за другим для достижения цели. Нейроны в базальных ганглиях оценивают состояние корковых зон и присваивают им значение.

Базальные ганглии – сложная версия функции стоимости, которую Джерри Тезауро обучил в TD-Gammon предсказывать значимость позиций на доске. Удивительный успех AlphaGo, достигшей уровня

чемпиона мира по го и описанной в главе 1, основан на той же архитектуре, что и TD-Gammon, но с большим размахом. Один слой скрытых элементов в оценочной сети TD-Gammon стал десятком слоев в AlphaGo, сыгравшей сотни миллионов игр. Но основные алгоритмы остались прежними. Это наглядно показывает, как хорошо алгоритмы обучения нейронных сетей масштабируются. Насколько выше будет производительность, если мы продолжим увеличивать размер сети и время обучения?

Игры – куда более простая среда, чем реальный мир. Ступенькой к более сложным и неопределенным условиям является мир видеоигр. Компания DeepMind в 2015 году показала, что обучение с учетом временной разницы способно научить играть в компьютерные игры от Atari, такие как Pong[264], на сверхчеловеческих уровнях, принимая пиксели экрана в качестве входных данных[265]. Следующий шаг – видеоигры в 3D-формате. StarCraft[266] – одна из лучших соревновательных видеоигр всех времен. Компания DeepMind использует ее для разработки автономных сетей глубокого обучения, которые могут хорошо развиваться в этом мире. Компания Microsoft Research купила права на Minecraft, еще одну популярную видеоигру[267], и сделала открытым ее исходный код, чтобы другие могли настраивать 3D-среду и ускорять развитие искусственного интеллекта.

Играть в нарды и выходить на чемпионский уровень – впечатляющее достижение, а играть в видеоигры – важный следующий шаг. Но как насчет решения проблем в реальном мире? Цикл восприятие – действие (рис. 10.2) применим к любой задаче, план решения которой строится на основе сенсорных данных. Результат этих действий можно сравнить с прогнозируемым результатом, а разницу затем использовать для обновления состояния системы, делающей прогнозы. Применяя память о предыдущих условиях, можно оптимизировать использование ресурсов и прогнозирование потенциальных проблем.

Саймон Хайкин из Университета Макмастера в Канаде использовал эту структуру для улучшения производительности нескольких важных инженерных систем[268], в том числе когнитивного радио, которое динамически распределяет каналы связи, когнитивного радара, который динамически смещает частотный диапазон для уменьшения помех, и когнитивной сетки, которая динамически выравнивает нагрузку в зависимости от энергопотребления электрической сети. Управлять рисками также можно в рамках цикла «восприятие – действие»[269]. Улучшения в каждой из этих областей выходят существенные, значительно повышается производительность и сокращаются расходы.

Учим парить

В 2016 году мы с Массимо Вергассола из Калифорнийского университета в Сан-Диего задались вопросом, можно ли использовать обучение с учетом временной разницы, чтобы научиться парить, как птицы, оставаясь на высоте в течение многих часов и не затрачивая много энергии[270]. Восходящий поток теплого воздуха может поднять птицу достаточно высоко, но внутри потока воздух прогрет неравномерно, и можно как подняться вверх, так и упасть. Ориентиры, которые птицы используют для поддержания своей восходящей траектории перед лицом столь мощной стихии, неизвестны. Первым шагом была разработка реалистичной с точки зрения физики модели воздушного потока, неравномерного (турбулентного) из-за конвекции, и модели аэродинамики планера. Затем мы симулировали траекторию полета планера в турбулентном потоке.

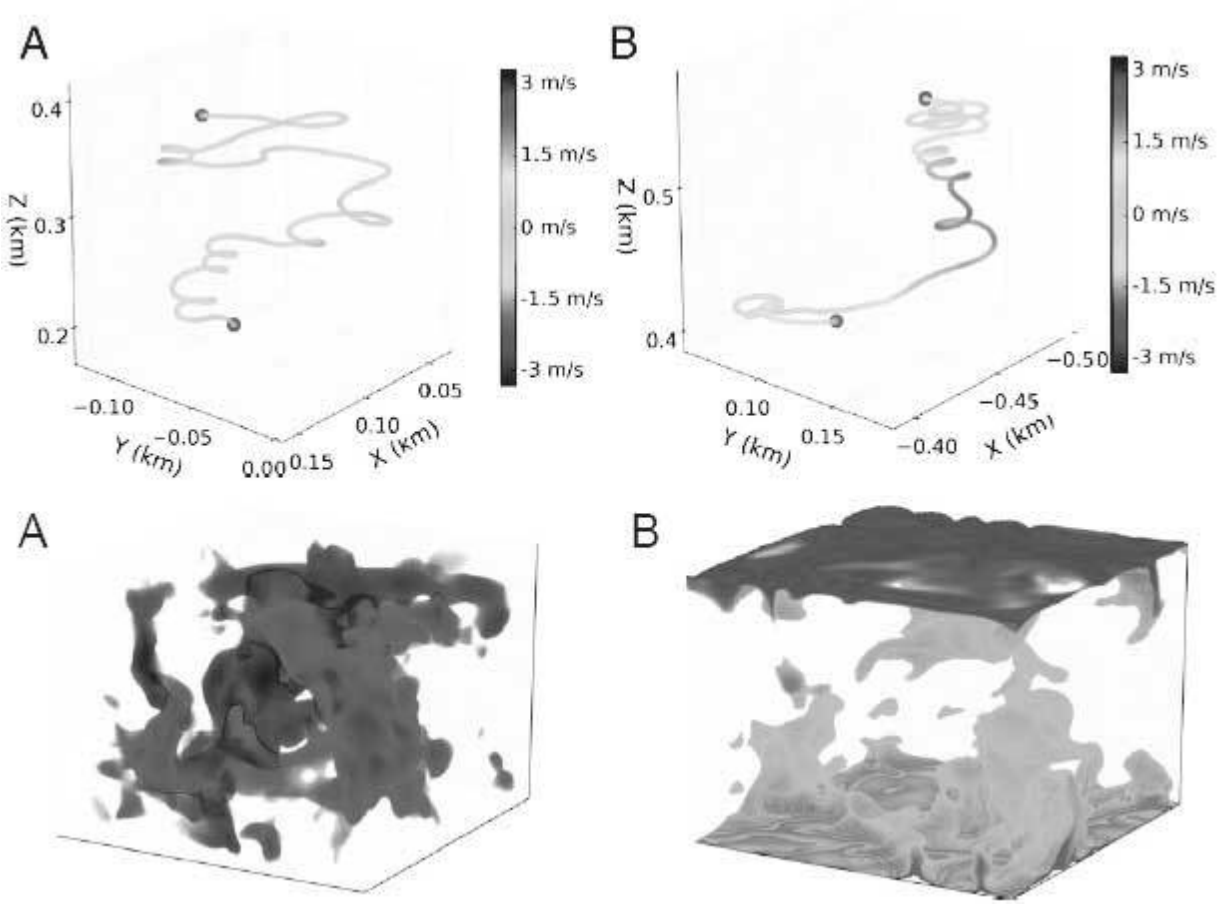


Рис. 10.6. Симуляция планера, учащегося парить в восходящем потоке теплого воздуха. Верхний ряд: снимки полей вертикальной скорости[271] (A) и распределения температур (B) в нашем трехмерной цифровой модели конвекции Рэлея – Бенара. Для поля вертикальной скорости светлым и темным цветами обозначены соответственно области большого восходящего и нисходящего потока. Для температурного поля светлый и темный цвета обозначают области высокой и низкой температур. Нижний ряд: (A) типичные траектории необученного и (B) обученного планера, летящего в турбулентном потоке Рэлея – Бенара. Оттенки указывают вертикальную скорость ветра, ощущаемую планером. Светлые и темные точки – начальная и конечная точки траектории. Нетренированный планер принимает случайные решения и спускается, в то время как обученный планер летит по характерным спиральным схемам в областях сильных восходящих течений, как

птица парит в восходящих потоках теплого воздуха

Поначалу планер не смог воспользоваться преимуществом, которое давали столбы теплого воздуха, и скользил вниз (рис. 10.6). Получив вознаграждение за подъем, планер начал осваивать стратегию, и после нескольких сотен попыток траектории планера напоминали плотные петли, наблюдаемые у парящих птиц (см. рис. 10.6). Кроме того, были найдены различные стратегии для различных степеней турбулентности. Анализируя эти стратегии, мы можем разработать гипотезы и узнать, используют ли их парящие птицы. Мы также оснастили планер измерительной аппаратурой, чтобы увидеть, насколько хорошо алгоритм обучения выполняет полет в реальных условиях.

Учим петь

Другой пример силы обучения с подкреплением – интересная параллель между тем, как птицы учатся петь, и тем, как дети учатся говорить. В обоих случаях сначала идет период слухового обучения, за которым следует поэтапное моторное обучение. Зебровые амадины слышат песню своего отца в начале жизни, но в течение нескольких месяцев не производят никаких звуков сами. Даже если их изолировать от отца до начала действия моторной фазы обучения, они проходят через период «суб-песни», которая совершенствуется и в конечном итоге превращается в песню отца. Зебровые амадины узнают, из какой части леса их сородич, по его песне, так же, как вы узнаете, откуда прибыл человек, по его акценту. Суть гипотезы, лежащей в основе исследования пения птиц, в том, что во время слухового обучения они изучают шаблон, который затем используют для уточнения звуков, производимых мышечной системой. Механизмы, которые отвечают за фазу моторного обучения, и у людей, и у певчих птиц находятся в

базальных ганглиях, где, как мы знаем, происходит обучение с подкреплением.

В 1995 году Кенджи Дойя, постдокторант в моей лаборатории, разработал модель обучения с подкреплением для совершенствования птичьего пения (рис. 10.7). Алгоритм улучшал производительность, настраивая связи между нейронами на модели нижней гортани певчих птиц (сиринкса), а затем тестируя ее, чтобы увидеть, действительно ли новая песня лучше предыдущей. Если это было так, то изменения сохранялись, но если новая песня была хуже, изменения в синапсе откатывались к первоначальному состоянию[272]. Мы предсказали, что в верхней части моторной цепи, которая генерирует последовательность слогов, должны быть нейроны, которые активны только на одном слоге песни, чтобы облегчить настройку каждого слога отдельно. Спустя некоторое время ученые из лаборатории Майкла Фи при МТИ и из других лабораторий, изучающих пение птиц, подтвердили эту и другие ключевые предсказания модели.

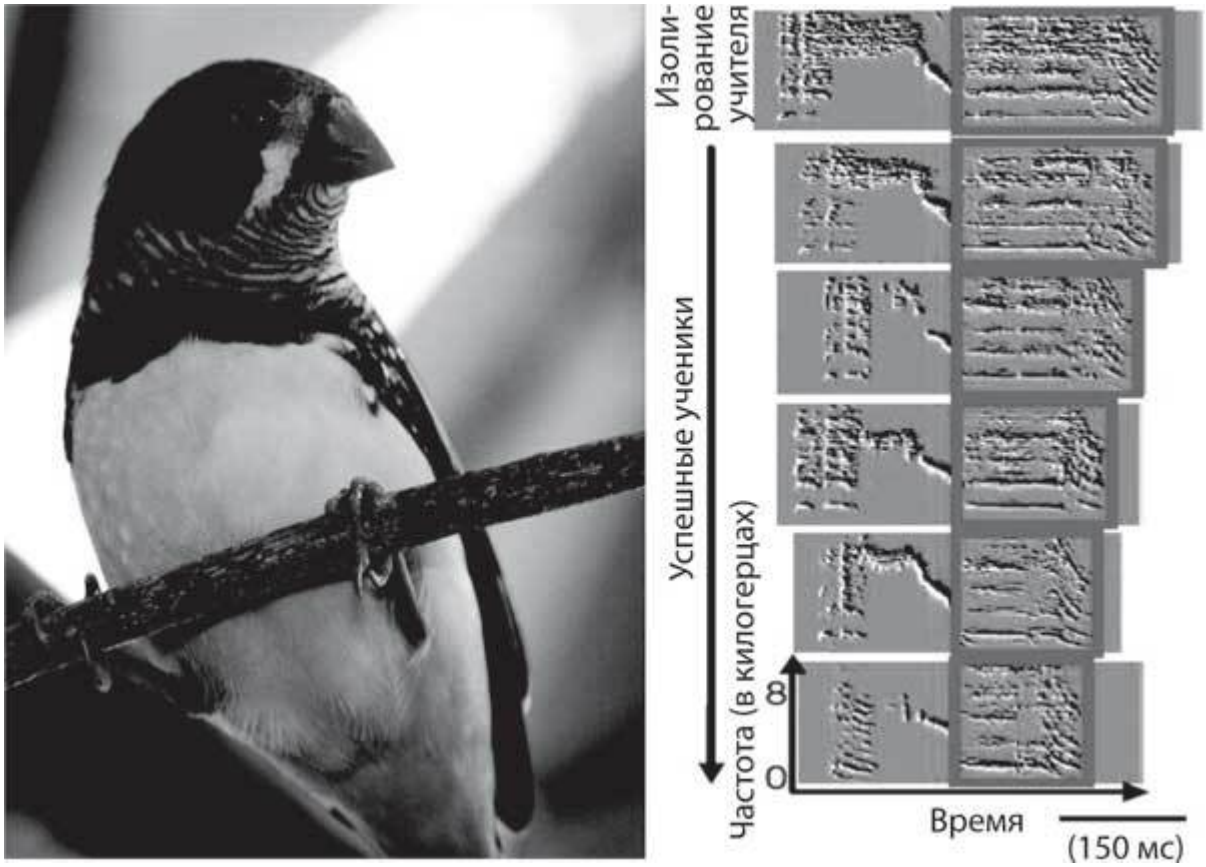


Рис. 10.7. Пение зебровых амадин. Пение отца (сверху) обучает петь сына (ученика), и диалект передается из поколения в поколение. Обратите внимание на сходство мотива (обведенная область) в спектрограмме (спектральная мощность как функция времени). Мотив становится короче с каждым поколением

Эллисон Доуп, изучавшая в Калифорнийском университете в Сиэтле пение птиц, и Патриция Куль, изучавшая в Вашингтонском университете в Сиэтле развитие речи у детей, провели много параллелей между тем, как птицы осваивают пение и как малыши осваивают речь[273]. И слоги у птиц и фонемы у младенцев изучаются сначала как звуки, и только позже происходит моторное обучение – «суб-песня» у птенцов и лепет у детей. Алгоритм обучения с подкреплением у певчих птиц отличается от обучения с учетом временной разницы, используемой в системе вознаграждения, и показывает, что в мозге много систем обучения и запоминания, которые зависят от предметной области и для приобретения новых навыков должны работать вместе.

Другие формы обучения

Несмотря на прогресс, достигнутый в автоматизации некоторых когнитивных функций, таких как зрительное и слуховое восприятие, есть много других аспектов человеческого интеллекта, нуждающихся в улучшении. Обучение представлениям в коре и обучение с подкреплением в базальных ганглиях существенно дополняют друг друга. Можно ли обучение игре в го на чемпионском уровне перенести на решение других сложных задач? Большая часть человеческого обучения основана на наблюдении и подражании, и людям нужно гораздо меньше примеров, чем при глубоком обучении, чтобы начать распознавать новый объект. Непомеченных сенсорных данных очень много, и мощные неконтролируемые алгоритмы обучения могут использовать их в своих интересах, прежде чем начать наблюдение. В главе 7 для запуска сетей глубокого обучения использовалась неконтролируемая версия алгоритма Больцмана, а в главе 6 – независимый компонентный анализ, неконтролируемый алгоритм обучения, извлекающий разреженную совокупность кодов из фотографий природы. Неконтролируемое обучение – это следующий рубеж в машинном обучении. Мы только начинаем понимать, как мозг обрабатывает данные.

В мозге много систем обучения и форм пластичности, которые усиливают друг друга. Даже в коре есть несколько десятков форм пластичности, включая пластичность в возбуждаемости и усилении нейронов. Особенно важная форма синаптической пластичности – гомеостатическая, которая поддерживает уровень активности нейронов в пределах их оптимального динамического диапазона. Что происходит, когда синаптическая сила уменьшается до нуля или достигает максимального предела? Это может привести к тому, что нейрон никогда не получит достаточно входных данных для достижения порога или, наоборот, у него будет слишком много

входных данных и всегда на высоком уровне активности. Джина Турриджано открыла в головном мозге новую форму синаптической пластичности, которая нормализует все синапсы в нейроне для поддержания баланса его активности[274]. Если средняя скорость сигналов слишком высока, все возбуждающие синаптические силы уменьшаются; и наоборот, если скорость сигналов слишком низкая, силы увеличиваются. Для тормозящих входных сигналов они меняют направление на противоположное: синаптическая сила увеличивается, если активность слишком высокая, и уменьшается, если активность слишком низкая. Подобные формы нормализации оказались эффективными при моделировании развития нейронных карт[275]. Искусственные нейронные сети, которые управляются стохастическим градиентным спуском, могут извлечь выгоду из гомеостатического масштабирования.

В мембранах нейронов есть десятки потенциал-зависимых и лиганд-зависимых ионных каналов, которые регулируют возбудимость и передачу сигналов. Должны существовать механизмы, основанные на локальном характере активности в дендритах, сомах[276] и аксонах нейронов, которые динамически регулируют расположение и плотность каналов. Было предложено несколько алгоритмов того, как это реализовано[277]. Эта форма гомеостаза не так хорошо изучена, как гомеостатическая синаптическая пластичность.

Чего не хватает?

Мы с Демисом Хассабисом участвовали в симпозиуме «Мозги, умы и машины»[278] на конференции NIPS в Монреале в 2015 году, а также в семинаре «Единицы информации и мозг»[279] на конференции NIPS в 2016 году в Барселоне. Это были жаркие дебаты о будущем искусственного интеллекта и о том, в каком направлении нужно вести исследования. Остается множество открытых вопросов в области ИИ, которые нужно решить. Прежде

всего, понимание причинно-следственных связей, от которых зависят высшие уровни человеческого мышления. При этом действия считаются совершенными намеренно, что предполагает наличие разума. Я упоминал ранее, что ни одна из систем глубокого обучения, которые мы создали, не способна выжить самостоятельно. Автономия станет возможна только в том случае, если будут включены функции многих других частей мозга, которые до сих пор игнорировались, такие как гипоталамус, который необходим для гомеостаза, и мозжечок, который помогает нам совершенствовать моторику на основе ошибки прогнозирования движения. Это древние структуры, найденные у всех позвоночных животных, и они важны для выживания.

Глава 11. Нейронные системы обработки информации

Проследить истоки идей сложно, так как наука – коллективная работа многих людей, широко разбросанных во времени и пространстве. NIPS нитью шла через все повествование, и к настоящему времени должно быть ясно, что эта конференция оказала большое внимание на меня, а также на поле моей деятельности. Беатрис Голомб, моя будущая жена, произнесла речь о SEXNET на одной из первых конференций NIPS. На другой конференции, вскоре после нашей свадьбы, мы почти расстались. Я виноват, что не уделял ей достаточно внимания. Конференции – это полное погружение в официальные сессии днем и в стендовые доклады вечером, и все заканчивается за полночь. Однажды, когда я вернулся в наш номер в три часа ночи и не нашел в нем Беатрис, я понял, что дела плохи. Я усвоил урок, и три десятилетия спустя мы все еще вместе.

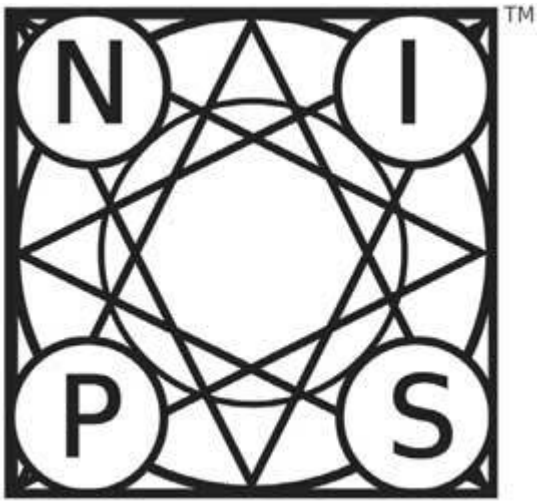


Рис. 11.1. Логотип Конференции по нейронным системам обработки информации (NIPS). Она была основана более 30 лет назад и сейчас является ведущей конференцией по машинному и глубокому обучению

Корни глубокого обучения можно проследить по ежегодным конференциям и семинарам NIPS[280] и их предшественникам. В 1980-х годах разнородная группа инженеров, физиков, математиков, психологов и нейробиологов собралась на конференцию NIPS, чтобы разработать новый подход к ИИ (рис. 11.1). Стремительный прогресс был вызван достижениями физиков, анализирующих модели нейронных сетей, психологов, воспроизводящих человеческое познание, нейробиологов, моделирующих нейронные системы и анализирующих нейронные записи, статистиков, исследующих большие наборы данных в многомерных пространствах, и инженеров, создающих устройства, которые могли видеть и слышать, как люди.

На первой конференции NIPS в 1987 году в Денверском техническом центре присутствовали 400 человек. Академические встречи, как правило, сосредоточены на узких областях исследований, и это удобно, потому что все говорят на одном языке. Научное разнообразие на ранних этапах существования конференции захватывало дух. Биологи использовали свой шифр, когда беседовали с другими биологами. Проблема была в том, что они не умели говорить о своих исследованиях без использования кодовых слов[281]. Еще хуже приходилось математикам и физикам, которые разговаривали только уравнениями. Инженерам было несколько проще, потому что они создавали вещи, которые говорили сами за себя. На междисциплинарные исследования все надеются, но они редко бывают продуктивны из-за подобных культурных барьеров. В первые годы на конференции NIPS казалось, что все говорят на разных языках.

После основной конференции участники собрались на семинар на близлежащем горнолыжном курорте и организовали небольшие групповые встречи на месте. Именно здесь началось живое общение между представителями различных дисциплин в более неформальной обстановке. Я хорошо помню нейробиолога[282],

которого встретил в вестибюле в Кистоне, который предложил провести семинар по обсуждению аплизий – морских моллюсков, – находясь в джакузи. Джентльмен из Министерства обороны, сидевший в джакузи рядом со мной, вероятно, задавался вопросом, какое отношение аплизия имеет к национальной безопасности? И действительно, в энциклопедии «Британика» есть три статьи со словом «интеллект» в заголовке: «Человеческий интеллект», «Искусственный интеллект» и «Военный интеллект». Сегодня семинары NIPS – это мини-конференции со стендовыми докладами, некоторые из которых привлекают тысячи посетителей.

Первое, что поддерживало NIPS многие годы, – витавшее в воздухе предчувствие, что мы на грани решения сложных вычислительных задач, основанных на алгоритмах обучения и вдохновленных биологией. Второе – Эд Познер (рис. 11.2), теоретик информации в Калтехе и главный технолог в Лаборатории реактивного движения, которые имел долгосрочный план и основал Фонд NIPS для управления встречами.

Рис. 11.2. Эд Познер из Калифорнийского технологического университета, основатель Конференции по нейронным системам обработки информации. NIPS остается по-прежнему влиятельной даже спустя 30 лет отчасти из-за его дальновидности

Культура организации – часто отражение ее основателей. Эд дал NIPS уникальное сочетание мудрости, практического ума и чувства юмора. Он был вдохновляющим учителем и эффективным лидером, его любили в Калтехе за поддержку программы SURF – одной из жемчужин университета, которая дает студентам возможность участвовать в летних исследованиях. Эд нанял Фила Сотела бесплатным юридическим консультантом для NIPS, который десятилетиями приглядывал за конференцией.

Эд знал мою жену, Беатрис Голомб, когда она была маленькой девочкой, и знал меня как участника NIPS, поэтому, когда я неожиданно сказал ему на конференции, что сделал Беатрис предложение, он переспросил: «Предложение чего?» Когда Эд трагически погиб в велосипедной аварии в 1993 году, я стал президентом Фонда NIPS, который продолжает расти и процветать. У нас проходит ежегодная лекция в честь Эда Познера. Приглашенные ораторы, как правило, работают в областях, выходящих за рамки основного направления NIPS, но лекцию

Познера представляет кто-то из нашего сообщества, внесший большой вклад в профильные исследования.

Генеральные председатели NIPS – группа выдающихся ученых и инженеров. Назовем лишь некоторых из них. Скотт Киркпатрик – физик, который изобрел для компьютеров способ решать сложные вычислительные задачи, «нагревая» их и медленно «охлаждая», что называется имитационным отжигом. Себастьян Трун – профессор компьютерных наук, который, как мы уже отметили, выиграл в 2005 году грандиозную гонку машин-беспилотников DARPA, проложившую путь сегодняшним самоуправляемым автомобилям. Дафна Коллер – специалист по информатике, соучредитель проекта Coursera, который стал одним из первых массовых открытых онлайн-курсов (MOOK).

Взлететь глубокому обучению позволили большие наборы данных. Не так давно терабайт информации занимал аппаратную стойку, теперь можно купить флешку емкостью в терабайт. В дата-центрах интернет-компаний хранятся петабайты, то есть тысячи терабайт данных. С 1980-х годов объем данных в мире удваивался каждые три года. Тысячи петабайт данных добавляются в Интернет каждый день, и общий объем достиг зеттабайта[283], что равно миллиону петабайт. Информационный взрыв оказывает влияние на все сферы жизни общества, включая науку и технику. Было бы невозможно обучить действительно глубокие сети без миллионов изображений и других помеченных данных, доступных в Интернете.

Университеты по всему миру создают новые центры, институты и департаменты по анализу и обработке данных[284]. В 2009 году Алекс Салай основал Институт интенсивной обработки данных и науки (Institute for Data Intensive Engineering and Science; IDIES) при Университете Джона Хопкинса, опираясь на свой опыт работы с Слоановским цифровым небесным обзором (Sloan Digital Sky Survey; SDSS; СЦНО)[285], который начал собирать астрономические данные в 1998 году. Это привело к тысячекратному увеличению общего объема данных, которые когда-либо получали астрономы, и

сегодня СЦНО – наиболее используемый астрономический инструмент в мире. Терабайты данных, собранные СЦНО, сравнивают со многими петабайтами, которые будет собирать строящийся Большой синоптический обзорный телескоп. Ян Лекун основал Центр науки о данных в Нью-Йоркском университете в 2013 году. Преподаватели со всех факультетов стучались к нему в дверь с данными в руках. Магистерские степени в области науки о данных становятся такими же популярными, как магистерские степени по деловому администрированию.

Глубокое обучение за игровым столом

Глубокое обучение достигло зрелости на Конференции NIPS в 2012 году на озере Тахо (рис. 11.3). Джеффри Хинтон, первопроходец в области нейронных сетей, и его студенты представили доклад о том, что многослойные нейронные сети удивительно хорошо распознают объекты на изображениях. Это было не просто лучше, чем современное компьютерное зрение, – это была другая лига, гораздо ближе к человеческому уровню. New York Times опубликовала статью о глубоком обучении, и Facebook объявил о новой лаборатории ИИ, основателем и директором которой станет Ян Лекун – еще один пионер глубокого обучения.

Рис. 11.3. В 2012 году конференцию NIPS провели в казино на озере Тахо. Эта конференция стала поворотным моментом для области исследований и вернула в название «Нейронные системы обработки информации» слово «нейронные»

Марк Цукерберг, генеральный директор Facebook, на Конференции NIPS в 2018 году принял участие в семинаре по глубокому обучению. Это было головной болью с точки зрения обеспечения безопасности, но привлекло столько внимания, что зал, где шла прямая трансляция, был переполнен. На приеме после семинара меня познакомили с Цукербергом, который задал мне вопросы о мозге. Он проявил особый интерес к теории разума в психологии: у нас есть негласная теория того, как работает наш разум, и мы используем ее, чтобы найти подход к разуму других. Когда вы отправляете сообщение другу, вы не знаете о массе решений, которые ваш мозг принял касательно того, что и как печатать. Цукерберг задавал много вопросов: как мой мозг строит ментальную модель меня самого? Как мой мозг строит ментальные модели других людей на основе опыта? Как мой мозг предсказывает поведение других? У других видов есть теория разума? Недавно я был одним из организаторов симпозиума по теории разума в Институте Солка, и он потребовал все материалы.

В машинном обучении выигрывает тот, у кого больше данных, а у Facebook больше данных о вкусах и друзьях множества людей и их фотографий, чем у кого-либо другого. Со всеми этими данными Facebook может создать теорию вашего разума и использовать ее для прогнозирования ваших предпочтений и политических пристрастий. Facebook может когда-нибудь узнать вас лучше, чем вы знаете себя. Станет ли Facebook когда-нибудь воплощением Большого Брата Оруэлла?[286] Вы находите это пугающей перспективой или вам было бы удобно иметь цифрового двойника, который заботится о ваших потребностях? Вы вполне можете спросить, должен ли Facebook получить такую власть, но у нас нет особого выбора.

Хотя мы проводили конференции NIPS 2012 и 2013 в казино на озере Тахо, участники избегали игровых столов: они знали, что то, над чем они работали, было куда более захватывающим. Игры могут вызвать привыкание из-за ошибки прогнозирования дофаминовой

системы вознаграждения – части нашего мозга, обсуждавшейся в главе 10. Казино оптимизируют условия, которые благоприятствуют ставкам: обещание большого выигрыша; меньшие выигрыши, происходящие время от времени в случайных местах, что, как известно из исследований, проведенных на крысах, лучшая схема для того, чтобы заставить их нажимать на кнопку в поисках еды; шум и огни, которые буквально взрываются, когда в игровом автомате выпадает выигрыш, также дают положительное подкрепление; тусклые огни ночью и днем вырывают ваш суточный ритм из нормального цикла день-ночь, поощряя вас делать ставки, пока вы не упадете. Но в долгосрочной перспективе казино всегда выигрывает.

На NIPS-2015 в Монреале 3800 международных участников переполнили Дворец Конгрессов. Семинар по глубокому обучению в начале встречи был так популярен, что нам пришлось прервать его и поставить «вышибал», в том числе генерального председателя, чтобы соблюсти меры противопожарной безопасности. Глубокое обучение внедрено почти каждой компанией с большими данными в секторе высоких технологий и растет ускоренными темпами. Конференция 2016 года в Барселоне за две недели до начала рассчитывала принять 6000 участников. Прилетевшие из Нью-Йорка без предварительной записи были разочарованы, узнав, что не могут зарегистрироваться на сайте. Если прирост на 50 процентов в год, идущий с 2014 года, продолжится, рано или поздно все на планете захотят прийти на NIPS. Конечно, пузырь в конце концов лопнет, но, как и в отношении большинства пузырей, никто не знает, когда.

Исследователи из многих областей науки и техники продолжают собираться на NIPS, как они это делали ежегодно в течение 30 лет. Из 6000 участников NIPS в Барселоне 40 процентов были там впервые, но культура, унаследованная с первых встреч, не изменилась. На протяжении многих лет Попечительский совет фонда NIPS мудро придерживался общей программы, что редко встречается на больших конференциях. По их идее, все желающие

должны были сидеть в одном помещении, чтобы не дробить «поле боя». В 2016 году одно направление разделилось на два, потому как было трудно найти достаточно большой зал, чтобы вместить всех, но все же не на десяток, что нередко для большинства крупных конференций. Показатель приема заявок сохранили на уровне 25 процентов, что меньше, чем у большинства журналов. NIPS организовал мероприятие «Женщины в машинном обучении» (Women in Machine Learning; WiML)[287], которое в 2016 году привело в Барселону почти 600 женщин – 10 процентов участников конференции. Разнообразие продолжает оставаться отличительной чертой NIPS. Ни одна область сама по себе не смогла бы объединить настолько разные таланты, создающие глубокое обучение.

Вероятно, выглядит странным, что на технологии глубокого обучения, которые могут повлиять на многие отрасли, так мало патентов, защищающих интеллектуальную собственность. В 1980-х годах мы хотели сделать алгоритмы обучения основой для новой области науки и полагали, что патенты не помогут. Конечно, сегодня компании подают патенты на конкретные приложения, но они не будут по-крупному вкладываться в новые технологии без защиты.

Подготовка к будущему

Серьезные прорывы в обучении нейронный сетей происходят каждые 30 лет: перцептроны появились в 1950-х годах, затем, в 1980-х, – обучение многослойных перцептронов и в 2010-х годах – глубокое обучение. В каждом случае был период эйфории, когда за короткий срок был достигнут значительный прогресс, а после он долгое время шел маленькими шажками. Однако каждый виток бурного роста оказывает все больший эффект. Последний скачок был вызван широкой доступностью больших данных, и история NIPS стала одним из этапов подготовки к этому дню.

Часть III. Технические и научные последствия: хронология

1971 – Ноам Хомский написал для журнала The New York Review of Books эссе «Дело против Б. Ф. Скиннера»[288], которое во многом проложило путь когнитивной революции.

1982 – Клод Шеннон опубликовал статью «Математическая теория связи», которая заложила основу современной передачи цифровой информации[289].

1989 – Карвер Мид выпустил книгу «Аналоговые СБИС[290] и нейронные системы», дав начало нейроморфной инженерии – созданию компьютерных микросхем на примере биологических объектов.

2002 – Стивен Вольфрам опубликовал работу «Новый вид науки», где исследовал вычислительные возможности клеточных автоматов – алгоритмов, которые даже проще, чем нейронные сети, но способны к мощным вычислениям.

2005 – команда Себастьяна Труна победила в конкурсе беспилотных транспортных средств Управления перспективных исследовательских проектов Министерства обороны США.

2008 – Тоби Дельбрюк разработал завоевавший большую популярность чип сетчатки – датчик динамического зрения (Dynamic Vision Sensor; DVS), который считывает импульсы асинхронно, а не делает синхронные кадры, как современные цифровые камеры[291].

2013 – Белый дом в США объявил о старте BRAIN Initiative (Brain Research Through Advancing Innovative Neurotechnologies[292]). Цель программы – разработка инновационных нейротехнологий, которые приблизят нас к пониманию работы мозга.

Глава 12. Будущее машинного обучения

Наступает век когнитивных вычислений. Скоро появятся беспилотные автомобили, которые водят лучше нас. Ваш дом будет узнавать вас, просчитывать ваш распорядок дня и предупреждать о гостях. Краудсорсинговый сайт Kaggle, недавно[293] приобретенный компанией Google, провел конкурс с призом в миллион долларов за обнаружение рака легких с помощью компьютерной томографии, а также конкурс на 1,5 миллиона долларов за обнаружение скрытых предметов при сканировании тела в аэропортах для Министерства Национальной безопасности США[294]. Подобные помощники врачей смогут распознавать редкие заболевания, и благодаря этому повысится уровень медицинской помощи. Есть тысячи похожих приложений, и многие предстоит еще создать. Некоторые рабочие места будут вытеснены компьютерами, зато появятся новые. И хотя потребуется много времени, чтобы общество впитало новые технологии и приспособилось к ним, они не представляют угрозы нашему существованию. Напротив, мы вступаем в эпоху открытий и просвещения, что сделает нас умнее, поможет дольше жить и процветать.

В Сан-Франциско в 2015 году я был спикером на конференции по когнитивным вычислениям, спонсируемой компанией IBM[295]. Компания вкладывала большие средства в суперкомпьютер Watson[296], предназначенный для поиска в базе данных ответа на вопросы, заданные на естественном языке, который в 2011 году победил Кена Дженнингса в телевикторине «Jeopardy!»[297]. Watson основан на огромном числе фактов обо всем, от истории до поп-культуры, которые можно найти с помощью широкого спектра алгоритмов. Дженнингс выиграл 74 игры подряд за 192 дня, что является самой длинной победной серией в «Jeopardy!». Когда

программа победила его в телешоу, это привлекло внимание всего мира.

В такси из отеля на конференцию я подслушал разговор двух руководителей IBM. Компания IBM развертывала на основе Watson платформу, которую можно использовать, чтобы упорядочивать вопросы и получать на них ответы из неструктурированных баз данных в таких специализированных областях, как здравоохранение и финансовые услуги. Проект Watson стоит за яркой рекламой когнитивных вычислений, продвигаемых IBM. Один из руководителей выразил тревогу из-за того, что IBM делает ставку на Watson. Другой был обеспокоен возмещением 70-миллиардного потока доходов. IBM давно рассталась со своим аппаратным подразделением, а его сервисный отдел больше не конкурентоспособен. Инвестируя в программу Watson, IBM делала ставку на свой отдел программного обеспечения. Watson может отвечать на вопросы и давать рекомендации, основанные на бóльшем объеме данных, чем доступно человеку. Машинное обучение – важный инструмент для анализа массивов данных и извлечения из них информации. Однако, чтобы задать вопрос и использовать эту информацию, пока нужен человек.

IBM вложила 200 миллионов долларов в новую штаб-квартиру проекта Watson Internet of Things (IoT)[298] в Мюнхене. Инвестиции в Германии – одни из крупнейших в истории компании в Европе и ответ на растущий спрос со стороны более 6000 клиентов, которые хотят изменить свои операции с ИИ. Это только часть глобального плана вложить три миллиарда долларов в когнитивные технологии.

Жизнь в XXI веке

В традиционной медицине всем дают одни и те же лекарства, но теперь эти лекарства индивидуальны и направлены на конкретную цель. Меланому, которая была смертным приговором, сейчас можно остановить секвенированием раковых клеток и разработкой

персональной иммунотерапии. Сегодня эта процедура стоит 250 тысяч долларов, но со временем цены упадут и она станет доступна всем, так как базовая стоимость секвенирования генома рака всего – несколько тысяч долларов, а стоимость моноклональных антител, необходимых для лечения, – несколько сотен долларов.

Я работал в комитете, который консультировал директора Национального института здравоохранения США по созданию рекомендаций для правительственной программы BRAIN. В отчете BRAIN 2025[299] мы подчеркивали важность вероятностных и вычислительных методов, которые помогают нам интерпретировать данные, генерируемые новыми техниками нейронной записи. В настоящее время алгоритмы машинного обучения используют для одновременного анализа записей тысяч нейронов, анализа сложных поведенческих данных свободно движущихся животных и автоматизации восстановления анатомических цепей серийных электронно-микроскопических исследований. По мере того как мы реконструируем мозг, мы раскрываем множество новых алгоритмов, созданных природой.

Национальный институт здравоохранения США финансировал фундаментальные исследования в области нейробиологии последние 50 лет, но тенденция такова, что все больше и больше грантов выделяется на поддержку прикладных исследований, которые немедленно находят применение в медицине. Мы, конечно, хотим внедрить то, что уже открыто, но если мы не будем финансировать новые проекты сегодня, то через 50 лет внедрять в медицину будет нечего. Именно поэтому исследовательские программы, такие как BRAIN, важны сейчас, чтобы в будущем найти лекарства от тяжелых паталогий мозга, вроде шизофрении и болезни Альцгеймера[300].

Будущее идентичности

У меня есть номер социального страхования, который охраняется государством и который регулярно взламывают. В 2011 году в Министерстве по делам ветеранов США был утерян ноутбук с номерами социального страхования 21,6 миллиона ветеранов. Базу данных даже не пришлось расшифровывать, поскольку министерство использовало номер социального страхования ветерана в качестве идентификационного. С номером социального страхования и датой рождения хакер может похитить персональные данные.

В Индии миллиард граждан может быть точно идентифицирован с помощью биомаркеров, которые включают отпечатки десяти пальцев, снимок двух радужек, фотографию и 12-значный идентификационный номер (на три цифры длиннее, чем номер социального страхования в США). Aadhaar – крупнейшая в мире база биометрических данных. В прошлом индиец, желавший получить официальный документ, сталкивался с бесконечными задержками и многочисленными посредниками, требующими свою долю. Сегодня с помощью биометрии индийские граждане могут получить продовольственную и другую социальную поддержку, и у многих бедняков, у которых даже нет свидетельства о рождении, теперь есть документ, позволяющий за секунды идентифицировать их в любое время и в любом месте. Кража личных данных, ранее выкачивавшая направленные на пособия средства, была остановлена. Личность человека не может быть украдена, если только вор не отрубит ему пальцы и не вырежет глаза[301].

Индийский национальный реестр – семилетний проект Нандана Нилекани, миллиардера и соучредителя аутсорсинговой компании Infosys[302]. Работая в индийском правительстве, Нилекани создал своего рода цифровой скелет для Индии. По словам Нилекани, «небольшие постепенные изменения, помноженные на миллиард, –

это огромной скачок… Если миллиард человек может получить свой мобильный телефон за 15 минут, а не за неделю, это значительно увеличит эффективность экономики. Если миллионам людей деньги на их банковские счета переводятся автоматически, это огромный скачок производительности в экономике»[303].

На другой чаше весов от преимуществ цифровой базы личных данных граждан – утрата конфиденциальности, особенно когда биометрический идентификатор связан с другими базами данных, такими как банковские счета, медицинские записи и сведения о судимости, а также передвижения на общественном транспорте. Вопросы конфиденциальности уже вышли на первое по значимости место в США и многих других странах, где базы данных связаны, даже когда эти данные анонимны[304]. Ваш сотовый уже отслеживает ваше местонахождение.

Рассвет социальных роботов

В фильмах часто изображают ИИ в виде робота, который выглядит и разговаривает как человек. Не ожидайте, что ИИ будет похож на Терминатора с немецким акцентом. Вы будете общаться с голосом, как в фильме «Она» 2013 года, и взаимодействовать с телами, как у R2-D2 и BB-8 из «Звездных войн». Искусственный интеллект уже стал частью повседневной жизни. Когнитивные устройства станут разговаривать с вами, как голосовой помощник Alexa, с радостью делая вашу жизнь проще. Каково будет жить в мире, где есть подобные создания? Давайте посмотрим на наши первые шаги на пути к социальным роботам.

На настоящий момент достижения в области ИИ в основном затрагивали сенсорную и познавательную стороны интеллекта, но моторный и подвижный интеллект[305] оставался далеко позади. Я иногда начинаю лекцию со слов, что мозг – самое сложное устройство в известной нам Вселенной, но Беатрис Голомб, получившая медицинское образование, напоминает мне, что мозг –

только часть тела, которое в целом сложнее, чем мозг. У тела есть различные уровни сложности в зависимости от развития двигательных функций. Без тела не было бы никакой возможности общаться с внешним миром. Даже бактерии могут передвигаться и взаимодействовать со сложными средами, в некоторых из которых мы не смогли бы выжить[306].

Рис. 12.1. Хавьер Мовеллан дает интервью журналу The Science Network в своей лаборатории по созданию роботов в Калифорнийском университете в Сан-Диего. Хавьер первым применил социальных роботов в учебных комнатах и запрограммировал социального робота Rubi привлекать внимание 18-месячных малышей

Наши мышцы, сухожилия, кожа и кости активно приспосабливаются к изменчивому миру: горам, воде, гравитации, а также другим людям. Тело – удивительный химический завод, перерабатывающий сырье, то есть еду, в мастерски сделанные части тела. Тело – совершенный 3D-принтер, вся работа которого происходит внутри. Мозг получает входные сигналы от датчиков в каждой части тела, внутренняя деятельность непрерывно

регулируется, в том числе и на самых высоких уровнях представления в коре мозга, и решения принимаются с учетом внутренних приоритетов и необходимости поддерживать баланс между всеми одновременными требованиями. Тело действительно неотъемлемая часть мозга, как гласит основной постулат теории воплощенного познания[307].

Руби

Испанец Хавьер Мовеллан (рис. 12.1) был преподавателем и одним из руководителей Лаборатории машинного восприятия в Институте нейронных вычислений в Калифорнийском университете в Сан-Диего. Он верил, что при помощи роботов, которые взаимодействуют с людьми, мы получим больше данных о познании, чем в традиционных лабораторных экспериментах. Он создал робота-младенца, который улыбался вам, когда вы улыбались ему, и который пользовался огромной популярностью у прохожих. Изучив взаимодействие младенцев с их матерями, он сделал вывод, что дети используют такую стратегию, чтобы заставить мам улыбаться как можно чаще, прилагая минимум собственных усилий[308].

Руби (Rubi) – самый известный социальный робот Хавьера Мовеллана. Руби взаимодействовал с 18-месячными малышами в центре дошкольного образования в Калифорнийском университете в Сан-Диего. Он выглядит как телепузик, с выразительным лицом, бровями, которые поднимаются, чтобы выразить интерес, глазами, которыми являются подвижные камеры, руками, которые могут брать предметы, и животом-планшетом, с которым дети могут что-то сделать. (рис. 12.2).

Малышам трудно угодить. У них очень короткие интервалы внимания. Дети общаются с игрушкой несколько минут, потом теряют интерес и бросают ее. Как они будут взаимодействовать с Руби? В первый же день ребята оторвали ему руки, которые не были защищены. После ремонта и исправления кода Хавьер предпринял

еще одну попытку. Теперь робот был запрограммирован плакать, когда его дергают за руку. Это остановило мальчиков, а девочки бросились обнимать Руби, дав важный урок социальной инженерии.

Рис. 12.2. Руби взаимодействует с малышами в классе. Голова Руби может поворачиваться, его глаза-камеры, рот и брови выразительны. Пушистая светлая «прическа» меняет цвет в зависимости от настроения Руби

Дети играют с Руби, указывая на объект в комнате, например на часы. Если он не ответит в короткий промежуток времени от 0,5 до 1,5 секунды, малыши потеряют к нему интерес. Ответит слишком быстро – и Руби кажется им чересчур механическим; слишком медленно – и Руби становится скучным. После того как между роботом и детьми установилась взаимная связь, они стали относиться к нему как к живому существу, а не как к игрушке. Когда

он был в мастерской для обновления, дети расстроились, и им сказали, что Руби заболел и ему нужно остаться дома. Во время одного из экспериментов его запрограммировали учить малышей финским словам, которые те усваивали с той же легкостью, что и английские. Популярная песня стала хорошим помощником[309].

Одним из опасений, связанных с Руби, было то, что учителя могли чувствовать угрозу от робота, который когда-нибудь заменит их. Однако случилось совсем наоборот: учителя приветствовали его как помощника, который помогал держать класс под контролем, особенно при посетителях. Экспериментом, который мог бы радикально изменить раннее образование, стал проект «Тысяча Руби» (Thousand Rubi project). Идея состояла в том, чтобы массово производить роботов Rubi, размещать их в тысячах классах и собирать данные из тысяч экспериментов каждый день через Интернет. В числе проблем такого образования – оно эффективно в одних школах и неэффективно в других, так как школы и учителя отличаются. «Тысяча Руби» могла бы опробовать множество идей, как улучшить образовательную практику и исследовать различия между школами по всей стране, обучающими различные социально-экономические группы. Ресурсы для запуска проекта «Тысяча Руби» так и не были получены, но это отличная идея, которую кто-то должен воплотить.

Рис. 12.3. Род Брукс наблюдает за роботом Baxter, готовящимся поместить пробку в отверстие на столе. Это тот самый Род Брукс, которого я упоминал в рассказе о посещении Лаборатории искусственного интеллекта в МТИ в 1989 году. Он предприниматель, основавший компанию iRobot, производящую роботы-пылесосы Roomba, а также компанию Rethink, производящую роботов Baxter

Двуногие роботы неустойчивы, и им требуется сложная система управления, помогающая удерживать равновесие. Проходит около года, прежде чем ребенок начинает ходить. Природа начинала не с двуногих существ. Род Брукс (рис. 12.3), о котором я уже упоминал в главе 2, хотел создать робота, перемещающегося как насекомые. Он изобрел новый тип контроллера, который согласовывает движение шести ног и позволяет роботам-тараканам передвигаться, сохраняя равновесие. Его инновационной идеей было заменить абстрактное

планирование и вычисления механическим взаимодействием ног с окружающей средой. Он утверждал, что у роботов для выполнения повседневных задач их высшие когнитивные способности должны основываться на сенсомоторном взаимодействии с окружающей средой, а не на абстрактном мышлении. Слоны общительны, у них хорошая память[310], но они не играют в шахматы[311]. Род Брукс основал компанию iRobot, которая продала более десяти миллионов пылесосов Roomba, чистящих полы.

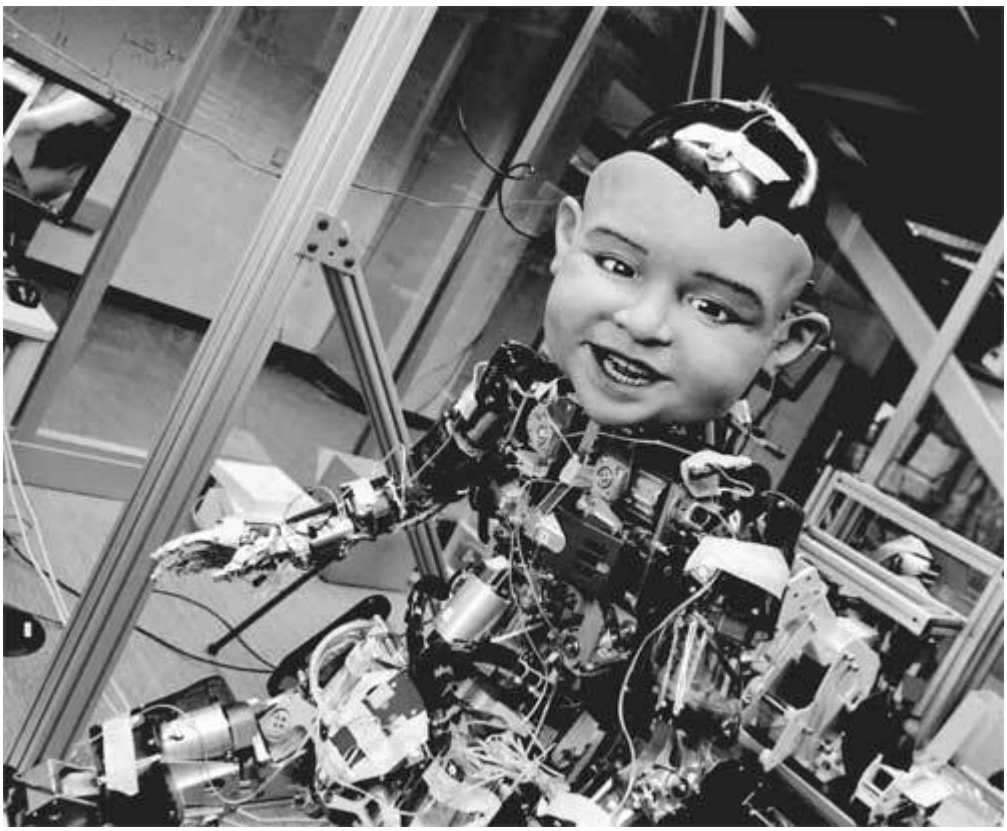


Рис. 12.4. Диего-сан, робот-ребенок. Пневматические приводы позволяют суставу двигаться податливо, так что роботу можно пожать руку. Лицо было создано Дэвидом Хэнсоном и компанией Hanson Robotics

Промышленные роботы имеют жесткие соединения и мощные сервоприводы, что делает их узкоспециализированными. Для новых

разработок Брукс основал компанию Rethink Robotics, которая создала робота, названного Бакстер (Baxter), с гибкими послушными суставами, позволяющими вам двигать его рукой (см. рис. 12.3). Вместо того чтобы писать программу для перемещения рук робота, вы перемещаете его руку через нужные движения, и он программирует сам себя, чтобы повторить эту последовательность.

Мовеллан пошел дальше Брукса и разработал робота-ребенка по имени Диего-сан[312], все суставы которого были подвижными. Логика в том, что когда мы что-то берем, в той или иной степени задействована каждая мышца в нашем теле (когда вы двигаете одновременно только одним суставом, вы выглядите как робот). Так мы легче приспосабливаемся к изменяющимся условиям нагрузки и взаимодействия с миром. Мозг может плавно контролировать все степени свободы в теле – все суставы и мышцы, – и целью проекта было выяснить, как он это делает. Моторы, приводящие в движение Диего-сан, были пневматическими, работающими благодаря давлению воздуха, поэтому все 44 соединения легко сгибались (рис. 12.4). Лицо Диего-сан имело 27 подвижных частей и могло выражать широкий спектр человеческих эмоций[313]. Движения робота-ребенка были удивительно реалистичными. Но несмотря на ряд достижений, Диего-сан победил Хавьера, который признал, что не знает, как заставить робота совершать действия так же плавно, как человеческий ребенок.

Выражение лица – окно в вашу душу

Представьте, что вы смотрите на экран своего сотового, видите, как падают ваши акции, и тут компьютер спрашивает, почему вы расстроены? Выражение вашего лица – окно в эмоциональное состояние вашего мозга, и теперь глубокое обучение может в него заглянуть. Познание и эмоции традиционно считали отдельными функциями мозга, полагая, что познание – корковая функция, а эмоции – подкорковые. И действительно, есть подкорковые структуры, такие как миндалевидное тело, которые регулируют эмоциональное состояние и участвуют, когда уровень эмоций высок, но эти структуры тесно взаимодействуют с корой головного мозга. Например, если миндалевидное тело вовлечено в общение между людьми, событие лучше запоминается. Познание и эмоции взаимосвязаны.

В 1990-х годах я сотрудничал с Полом Экманом (рис. 12.5), психологом из Калифорнийского университета в Сан-Франциско и ведущим мировым экспертом в области мимики. Пол Экман стал прототипом доктора Кэла Лайтмана в сериале «Обмани меня», хотя в общении он намного приятнее Лайтмана. Экман отправился в Папуа – Новую Гвинею, чтобы выяснить, показывают ли доиндустриальные культуры эмоции теми же выражениями лица, что и мы. Во всех видах человеческого общества он нашел шесть универсальных проявлений эмоций: счастья, печали, гнева, удивления, страха и отвращения.

Рис. 12.5. Пол Экман с племенем Форе в Папуа – Новой Гвинее в 1967 году. Он нашел доказательства шести универсальных выражений эмоций: счастья, печали, гнева, удивления, страха и отвращения. Пол был научным консультантом создателей сериала «Обмани меня», и образ доктора Кэла Лайтмана в некоторой степени списан с него

В 1992 году мы с Экманом организовали семинар «Понимание выражения лица» («Facial Expression Understanding»), спонсируемый Национальным научным фондом при правительстве США[314]. В то время было довольно трудно получить поддержку исследований мимики. Наш семинар собрал специалистов в области нейробиологии, электротехники и компьютерного зрения, а также психологии, что открыло новую главу в анализе лиц. Для меня стало неожиданностью, что, хотя анализ мимики потенциально так важен

для многих сфер науки, медицины и экономики, его никто не хочет финансировать.

Рис. 12.6. Марни Стюарт-Бартлетт демонстрирует анализ мимики. Временные отрезки – результат работы сетей глубокого обучения, которые распознают на лицах выражения счастья, печали, удивления, страха, гнева и отвращения

Экман разработал систему кодирования лицевых движений (Facial Action Coding System; FACS; СКЛиД), чтобы отслеживать состояние каждой из 44 мышц лица. Эксперты СКЛиД, обученные Экманом, тратят час на покадровую обработку минуты видео. Выражения изменчивы, они могут сохранятся многие секунды, но Экман обнаружил, что некоторые остаются всего на несколько кадров. Эти микровыражения – эмоциональные «утечки» подавленных состояний мозга и часто говорят о бессознательных эмоциональных реакциях и даже выявляют их. Например, микровыражения отвращения во время консультации по вопросам брака были надежным признаком того, что брак не сложится[315].

В 1990-х годах мы использовали видеозаписи с обученными актерами, которые, как и Экман, могли контролировать каждую мышцу на лице, чтобы обучать нейронные сети с обратным распространением ошибки для автоматизации СКЛиД. В 1999 году сеть, созданная моей аспиранткой Марни Стюарт-Бартлетт (рис. 12.6), имела точность 96 процентов в лаборатории[316] при идеальном освещении, лице, смотрящем строго в камеру, и вручную размеченном времени на видео. Точность была достаточно высокой, чтобы нас с Марни пригласили на телешоу «Доброе утро, Америка» с Дайан Сойер. Марни, работая преподавателем в Институте нейронных вычислений в Калифорнийском университете в Сан-Диего, продолжала разрабатывать систему Computer Expression Recognition Toolbox (CERT)[317], и по мере того как компьютеры становились быстрее, CERT подошла к анализу в реальном времени, чтобы маркировать изменяющиеся выражения лица в потоковом видео.

Марни и Хавьер основали компанию Emotient, чтобы вывести автоматический анализ мимики на рынок. Мы с Полом Экманом входили в ее научно-консультативный совет. Emotient создала сети глубокого обучения с точностью 96 процентов, которые работали в режиме реального времени при разном освещении, определяя выражение лиц людей, ведущих себя естественно и не смотрящих прямо в камеру. На одной из демонстраций за несколько минут они обнаружили, что Дональд Трамп оказывал наибольшее эмоциональное влияние на фокус-группу на первых республиканских дебатах. Социологам потребовалось несколько дней, чтобы прийти к такому же выводу, а экспертам – месяцы, чтобы признать, что ключевой стала эмоциональная вовлеченность. Наиболее выраженными эмоциями на лицах в фокус-группе были радость и страх. Нейросети также предсказали, какой сериал станет хитом, за несколько месяцев до публикации рейтинга Нильсена[318]. Emotient была куплена компанией Apple в январе 2016 года, и Марни и Хавьер теперь работают на Apple Inc.

Возможно, в скором будущем ваш iPhone будет спрашивать вас, почему вы расстроены, и стараться помочь успокоиться.

Наука об обучении

Двенадцать лет назад[319] во время конференции NIPS в Ванкувере я завтракал с Гэри Коттреллом, коллегой с кафедры компьютерных и технических наук Калифорнийского университета в Сан-Диего. Гэри входил в изначальную группу параллельной распределенной обработки с 1980-х годов, и он один из немногих оставшихся в университете – отголосок поколения 1960-х годов, с седой бородой и собранными в хвост волосами. Гэри Коттрелл наткнулся на объявление Национального научного фонда о приеме заявок по программе «Центры науки об обучении» (Science of Learning Centers; SLC). Его внимание привлек бюджет в пять миллионов долларов в год при контракте на пять лет, который может быть продлен еще на пять. Гэри хотел подать заявку и спросил, могу ли я помочь. Он сказал, что, если все получится, ему никогда не придется просить об еще одном гранте. Я сказал, что могу помочь, но в случае успеха этот грант положит конец его карьере. Он усмехнулся, и мы начали обсуждать детали.

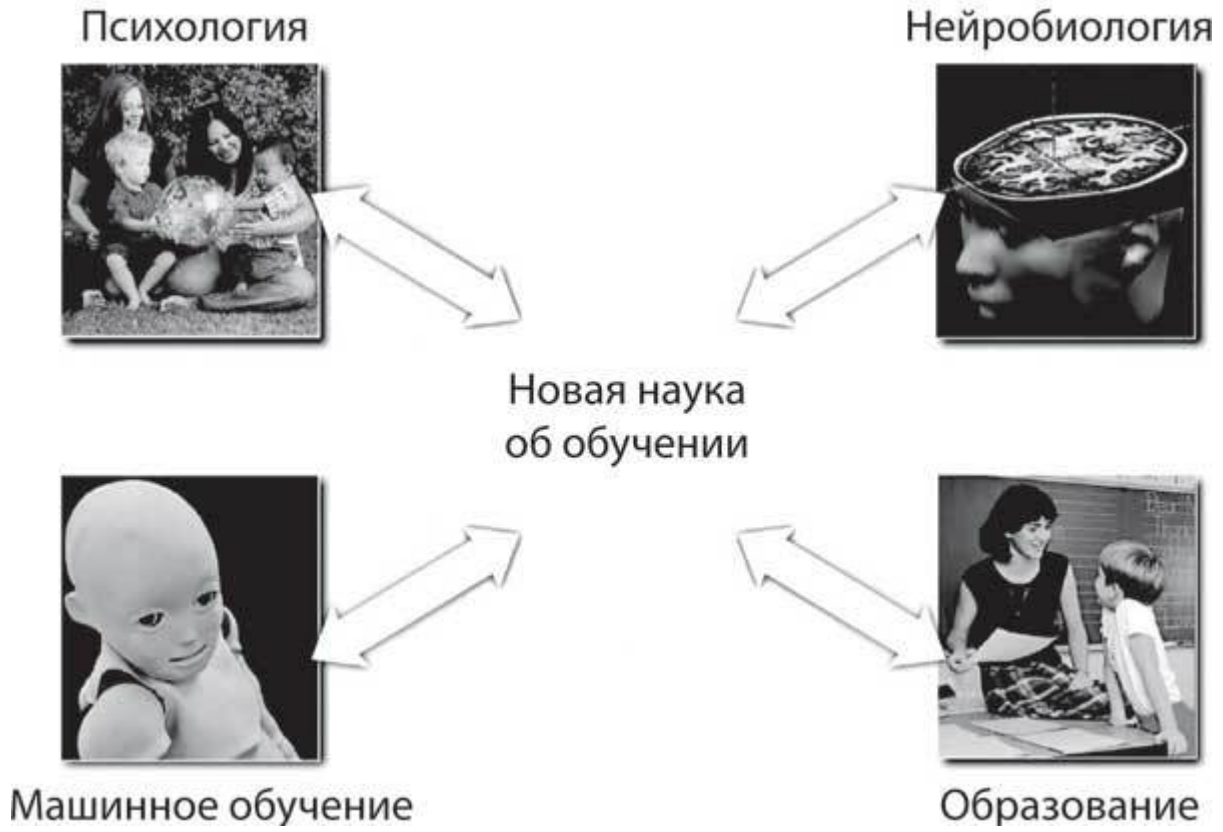


Рис. 12.7. Новая наука об обучении включает в себя машинное обучение и нейробиологию, а также углубленные знания в области психологии и образования. [Meltzoff, A.N. Kuhl, P.K. Movellan, J. Sejnowski, T. J. Foundations for a New Science of Learning, Science, 325: 284–288, 2009]

В конечном счете наша заявка была одобрена, и, как я и предполагал, ежегодные 300-страничные отчеты были просто зубодробительными. В наш Центр временнóй динамики обучения (Temporal Dynamics of Learning Center; TDLC) входило более сотни исследователей из 18 организаций со всего мира. Из шести научно-образовательных центров, финансируемых ННФ, наш был наиболее ориентированным на нейробиологию и проектирование, и мы включили последние достижения в области машинного обучения в наши проекты (рис. 12.7)[320]. TDLC спонсировала проекты Rubi и CERT. У нас также была мобильная лаборатория ЭЭГ, где испытуемые могли свободно перемещаться в виртуальной среде, записывая свои мозговые волны. В большинстве лабораторий при записи ЭЭГ

требуется не двигаться и не моргать, чтобы избежать помех. Мы использовали независимый компонентный анализ, чтобы убрать помехи, вызванные движением. Это позволило нам наблюдать за активностью мозга, в то время как участники эксперимента активно изучали окружающую среду и взаимодействовали с другими людьми.

Вот лишь немногие из проектов TDLC:

• Один из важнейших этапов развития мозга – созревание всех звеньев слуховой системы, которые помогают мозгу различать звуки, особенно звуки речи, что позволят ребенку понимать произнесенные слова. Эйприл Бенасич из Центра молекулярной и поведенческой нейробиологии в Ратгерском университете разработала тест, который может предсказать, будут ли у ребенка трудности с освоением языка и обучением, на основе времени слухового восприятия. Для детей из группы риска она выявила, что поведенческое вмешательство – тренировки со звуками разной длительности и тональности и вознаграждением за обратную связь – в состоянии исправить этот дефицит, и ребенок сможет развить нормальный слух и обучаться. В экспериментах участвовали дети от трех месяцев до пяти лет. Интерактивная среда полезна и для нормально развивающихся детей. В 2006 году Эйприл Бенасич основала компанию AAB Research LLC с целью вывести на рынок технологию быстрой обработки слуховой информации (rapid auditory processing technology; RAPT), чтобы улучшить способность детей к обучению.

• Учителям нужна обратная связь, чтобы понять, трудно ли ребенку усваивать урок. Если ученик выглядит сбитым с толку, то, скорее всего, он что-то не понимает. Марни Стюарт-Бартлетт и Хавьер Мовеллан использовали машинное обучение для регистрации выражений на лицах учеников[321], чтобы предупреждать учителя, что кто-то выглядит растерянным. Сегодня это можно сделать автоматически и безошибочно, одновременно применяя глубокое обучение для каждого ребенка в классе. Есть

много других приложений для анализа мимики в маркетинге, психиатрии и судебной медицине, которые еще не используются.

• Уже более века нам известно, что метод интервального повторения эффективнее для долгосрочного запоминания, чем зубрежка, но почти все исследования длились недолго, всего несколько месяцев, и в основном с участием студентов колледжей. Хэл Пашлер из Калифорнийского университета в Сан-Диего и Майк Мозер из Колорадского университета в Боулдере провели многолетнее исследование на школьниках всех возрастов, чтобы выяснить, работает ли этот метод в других временных масштабах и для учеников младших классов. Они показали, что оптимальный интервал для повторения тем больше, чем дольше период, на который вы хотите сохранить в голове информацию. Составленное ими расписание для студентов языковых курсов показало отличные результаты.

• Учителя часто используют наиболее подходящий способ обучения для конкретного ученика – визуальный, аудиальный (озвучивание) или тактильный. Крупная индустрия обеспечивает учителей тестами и рекомендациями, основанными на этих методах. Но нет никаких научных доказательств, что применение предпочтительного для ученика стиля дает преимущества. Это вдохновило Бет Роговски, постдокторанта из TDLC, и Паулу Таллал из Ратгерского университета провести исследование, показавшее, что нет статистически заметной разницы между использованием вербальных или письменных материалов в обучении и нет связи между предпочитаемым и используемым методом обучения ни сразу, ни в дальней перспективе[322]. А значит, нет никакого смысла подбирать более удобный для ученика стиль преподавания и индустрия, которая продвигает материалы для определения такого стиля, не приносит особой пользы.

• Национальный научный фонд заинтересован в результатах и долговременном развитии. Паула Таллал сыграла важную роль в учреждении в 2014 году фондом X-Prize премии Global Learning

XPRIZE в размере 15 миллионов долларов за новаторские решения в образовании. Цель проекта – разработка масштабируемого программного обеспечения с открытым исходным кодом, которое позволит детям в развивающихся странах овладеть базовыми навыками чтения, письма и арифметики за 18 месяцев. Технологические решения и проекты, основанные на исследованиях, проведенных в области образования для X-Prize, в ближайшие десятилетия повлияют на все страны мира.

• В 2014 году в Шанхае на международной встрече, посвященной образованию, научный директор TDLC Андреа Чиба представила исследование, как обучение меняет структуру мозга. Один из делегатов, с удивлением узнав, что мозг пластичен, спросил, может ли образование изменить потенциал ребенка: «Значит ли это, что стоит обучать каждого ребенка?» Другие делегаты также были поражены, увидев старую проблему в новом свете. Очень многие считают, что дети приходят в мир с определенными способностями и что образование тратится впустую на тех, кто менее талантлив или слишком стар, чтобы учиться. В мире есть огромный человеческий потенциал, который не используется.

Мы обнаружили, что большие проблемы в образовании носят не научный, а социальный и культурный характер. В США 13 500 школьных округов, каждый с собственным школьным советом, который определяет учебную программу, квалификацию учителей и применение передового опыта. Потребуются десятилетия, чтобы охватить их все и рассмотреть каждую уникальную ситуацию. Прежде чем преподаватели приступают непосредственно к обучению, они должны организовать работу в классе, что может быть особенно сложно в начальных классах и школах в неблагополучных районах. Родители, выдвигающие определенные требования, не всегда в состоянии оценить высокий уровень выгорания у учителей из-за нехватки ресурсов и влияния профсоюзов.

Преподавание – трудная работа, как ни посмотри. Лучший и наиболее эффективный способ обучения – взаимодействие между опытным взрослым учителем и ребенком один на один[323]. На нас давит настоящий конвейер, созданный для массового образования, в котором дети разделены по возрасту и обучаются в больших классах, а учителя из года в год проводят одни и те же уроки. Конвейер хорош, чтобы построить автомобиль, и, возможно, его было достаточно в то время, когда работникам хватало только базового образования. Но сегодня, когда хорошие рабочие места требуют более высокого уровня подготовки, эта система не подходит, и важным становится обучение на протяжении всей жизни для обновления профессиональных навыков. Учиться всю жизнь биологически возможно, но возвращение в школу для взрослых может быть неприятно и неудобно. Информационная революция, которую мы переживаем, обогнала временные рамки поколения. Появляются новые технологии, которые могут изменить способ, которым мы получаем знания. Интернет так меняет среду обучения, как мы не могли и ожидать, когда наш Центр науки обучения открылся в 2006 году.

Научитесь учиться

Массовые открытые онлайн-курсы (MOOK) вышли на сцену в 2011 году, когда New York Times выпустила статью[324] о том, какой популярностью в Стэнфорде пользуется онлайн-курс по ИИ. Большое число учащихся и беспрецедентный охват в Интернете привлекли к себе внимание всего мира. Практически в одночасье были основаны новые компании для разработки и свободного распространения лекций одних из лучших преподавателей. Лекции доступны в любое время и в любом месте, где есть подключение к Сети. Помимо лекций, курсы включают в себя викторины, экзамены, форумы, где учащиеся могут задавать вопросы, есть помощники преподавателей, и студенты могут собираться на самостоятельно организованных встречах и обсуждать что-либо в неформальной обстановке. Аудитория MOOK значительно расширилась – в 2015 году количество учеников удвоилось примерно с 17 миллионов до более чем 35 миллионов[325]. MOOK выходят за рамки традиционных учебных заведений.

Я познакомился с Барбарой Оакли в Ирвине в январе 2013 года на встрече, организованной Национальной академией наук США в Калифорнийском университете. Она профессор электротехники в Оклендском университете в Мичигане, хотя у нее было плохо с математикой и естественными науками в школе. Получив гуманитарную специальность, Барбара дослужилась до капитана армии США и работала переводчиком с русского на советских траулерах в Беринговом море, прежде чем вернулась к учебе, поборов страх перед математикой, и получила докторскую степень по электротехнике. За ужином я обнаружил, что у нас с Барбарой схожие взгляды на обучение, и она пишет книгу «Думай как математик. Как решать любые задачи быстрее и эффективнее»[326]. Я пригласил ее посетить Калифорнийский университет в Сан-Диего и

прочитать лекцию для старшеклассников и преподавателей.

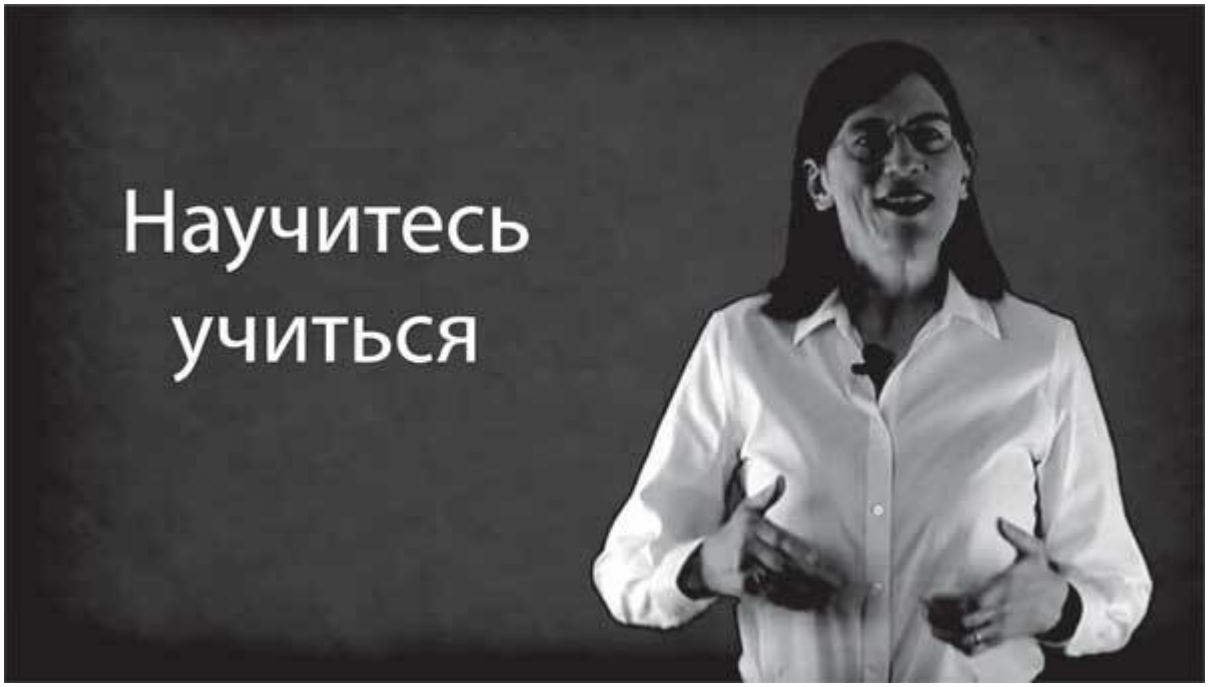


Рис. 12.8. Барбара Оакли представляет наш онлайн-курс. Его прошли более двух миллионов человек, что делает его самым популярным интернет-курсом в мире

Барбара Оакли пользовалась большим успехом у учеников, и было ясно, что она одаренный преподаватель. Ее подход и практические идеи восходят к тому, что мы знаем о мозге, поэтому мы объединили усилия, чтобы разработать для сайта Coursera онлайн-курс «Научитесь учиться: мощные умственные инструменты, которые помогут вам овладеть сложными предметами»[327], который впервые был представлен в августе 2014 года (рис. 12.8). В настоящее время[328] это самый популярный онлайн-курс в мире, в первые три года на него зарегистрировались более двух миллионов участников из двухсот с лишним стран, и он продолжает привлекать по тысяче новых учеников каждый день. Курс дает вам инструмент – понимание, как учится ваш мозг, необходимое, чтобы лучше учиться могли вы.

Отзывы наших учеников были только положительными, и мы разработали второй курс для тех, кто хочет сменить работу или свой образ жизни, названный «Mindshift»[329]. Оба онлайн-курса можно бесплатно пройти в любой точке мира, где есть подключение к Интернету.

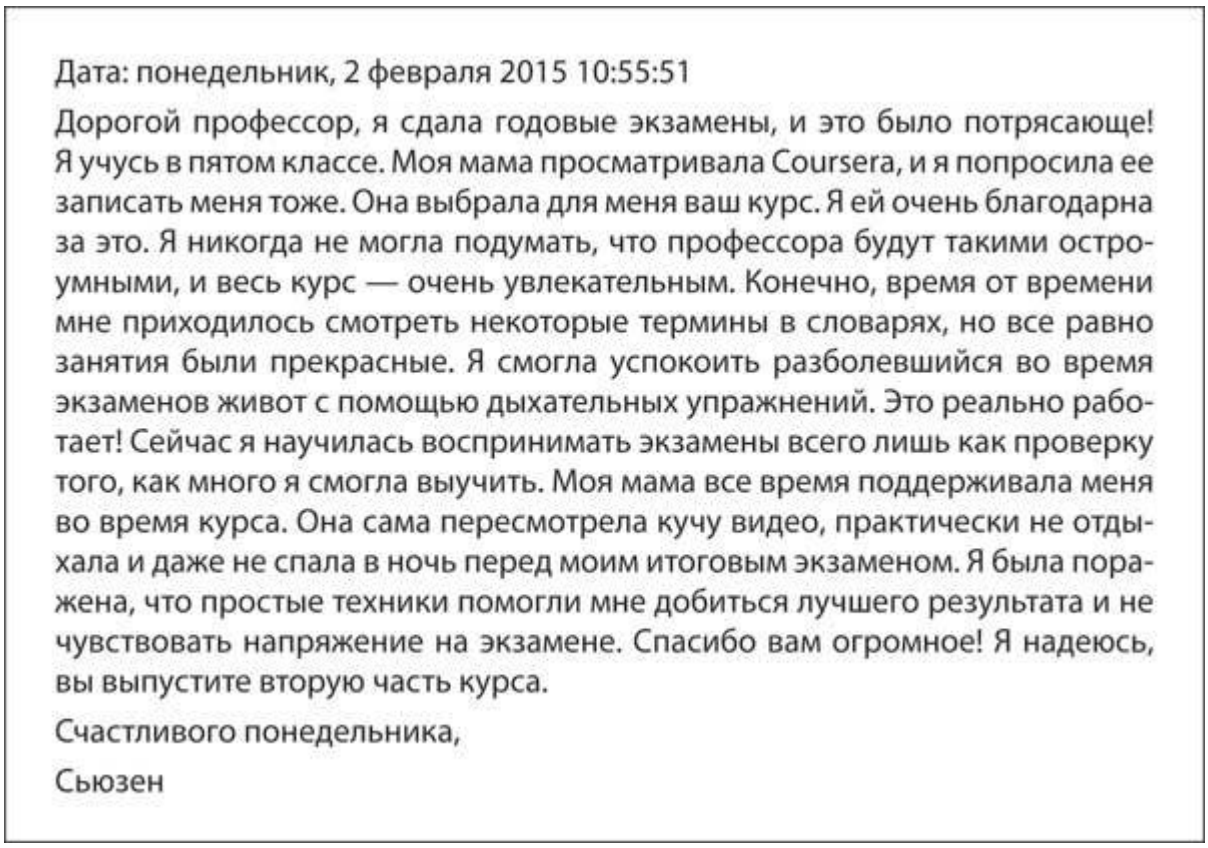
В курсе «Научитесь учиться» дают практические советы, как лучше усваивать знания, как справиться со страхом перед экзаменами, как избежать прокрастинации, и рассказывают, как наш мозг учится. Этот бесплатный месячный курс включает видеоролики по 5–10 минут, викторины и тесты, переведенные на более чем 20 языков. Один из основных факторов, лежащих в основе курса, – ваш мозг может неосознанно работать на вас, пока вы заняты чем-то еще. Анри Пуанкаре, выдающийся математик XIX века, однажды описал, как он решил сложную математическую проблему, над которой безуспешно бился нескольких недель. Пуанкаре взял небольшой отпуск. И когда он садился в автобус на юге Франции, внезапно пришел ответ на задачу – самопроизвольно, из той части его мозга, что продолжала работать над проблемой, пока он наслаждался отдыхом. Пуанкаре знал, что нашел верный путь к доказательству, и завершил его, когда вернулся в Париж. Он не смог решить задачу во время интенсивной работы над ней, но это подготовило его мозг так, что его бессознательное трудилось над ответом, пока он расслаблялся. Оба этапа одинаково важны для творчества.

Удивительно, но ваш мозг может работать над проблемой, пока вы спите. Но он делает это только в том случае, если вы сконцентрируетесь на попытке разобраться с задачей, прежде чем заснуть. Утром, как правило, в голову приходит свежая мысль, подталкивающая к верному решению. Серьезные усилия перед отпуском или засыпанием важны, чтобы подготовить ваш мозг – в противном случае, он займется каким-нибудь другим вопросом. Нет особой разницы между математикой и естествознанием – ваш мозг будет работать одинаково упорно, решая как социальные проблемы,

так и математические или естественно-научные, если они занимали ваши мысли в последнее время.

Одним из самым приятных результатов курса «Научитесь учиться» были отзывы от довольных участников. Иногда я получаю письма от бывших учеников, где они благодарят за лучший курс, который прошли, или за то, как он повлиял на их выбор профессии. Удивительно, но ежедневно поступают отзывы, в которых участники рассказывают, что наш курс значил для них. Одним из моих любимых – письмо от пятиклассницы (рис. 12.9). Учителя писали нам, что они применяют на уроках знания, полученные во время курса.

Изначально «Научитесь учиться» предназначался для старшеклассников и студентов колледжей, но оказалось, что те составляют менее одного процента от всех слушателей курса. Поскольку школы должны преподавать по общеобразовательной программе, у них нет времени учить учиться, что было бы полезнее для детей. Просить школьные округа ввести предмет «Научитесь учиться» непросто, поскольку их бюджет ограничен. Школьные округа не готовы пересматривать учебные программы, чтобы полномасштабно включить в них преподавание этого курса, поскольку любые усилия подобного охвата требуют дорогостоящей переработки графиков, переподготовки учителей и разработки новых учебных материалов. Однако каким-то образом нам нужно добраться до 12-летних учеников раньше, чем те перейдут в среднюю школу. Мы работаем над книгой, ориентированной на младших школьников, в надежде, что она дойдет до них прежде, чем они столкнутся с трудностями на уроках математики, как часто бывает[330].



Дата: понедельник, 2 февраля 2015 10:55:51

Дорогой профессор, я сдала годовые экзамены, и это было потрясающе! Я учусь в пятом классе. Моя мама просматривала Coursera, и я попросила ее записать меня тоже. Она выбрала для меня ваш курс. Я ей очень благодарна за это. Я никогда не могла подумать, что профессора будут такими остроумными, и весь курс — очень увлекательным. Конечно, время от времени мне приходилось смотреть некоторые термины в словарях, но все равно занятия были прекрасные. Я смогла успокоить разболевшийся во время экзаменов живот с помощью дыхательных упражнений. Это реально работает! Сейчас я научилась воспринимать экзамены всего лишь как проверку того, как много я смогла выучить. Моя мама все время поддерживала меня во время курса. Она сама пересмотрела кучу видео, практически не отдыхала и даже не спала в ночь перед моим итоговым экзаменом. Я была поражена, что простые техники помогли мне добиться лучшего результата и не чувствовать напряжение на экзамене. Спасибо вам огромное! Я надеюсь, вы выпустите вторую часть курса.

Счастливого понедельника,

Сьюзен

Рис. 12.9. Письмо 10-летней ученицы с ее впечатлениями о «Научитесь учиться». Оно было опубликовано на форуме курса, где у участников есть возможность задавать вопросы и обмениваться опытом друг с другом. Вторая часть курса – «Mindshift» – стартовала в апреле 2017 года

В большинстве онлайн-курсов высок процент бросивших их учеников. Участники имеют привычку «надкусывать» то одно, то другое, выбирая самые интересные для себя лекции. Такая модель обучения сильно отличается от занятий в аудитории, которые проводят по принципу «все или ничего». Онлайн-курсы больше похожи на книги, которые вы можете выборочно прочитать в любое время. Их можно считать перевернутым классом, где ученики слушают лекции в свое свободное время, а учитель только направляет ход обсуждения на занятии[331]. Онлайн-курсы – это другая среда, которая может удовлетворить потребности учащихся иначе, чем традиционные образовательные методы. Хотя

изначально онлайн-курсы воспринимались как альтернатива классическому обучению, они находят и новые ниши в этой сфере.

Наша система образования разрабатывалась для индустриальной эпохи, и знаний, полученных в школе, было достаточно, чтобы сохранить работу и оставаться полезным членом общества всю оставшуюся жизнь. Сегодня школьные знания устаревают уже к вашему выпуску из вуза. Онлайн-курсы – новый тип образования, который пришел к ученикам прямо на дом. На сайте Coursera самая многочисленная возрастная группа – 25–35 лет, и больше половины регистрирующихся на курсы окончили колледж. Это молодые, имеющие работу люди, которые нуждаются в новых навыках и получают их в Интернете. Нашей образовательной системе потребуются фундаментальные изменения, чтобы адаптировать мозг к быстрому увеличению числа рабочих мест в секторе информационных технологий. Например, сбор информации через Интернет требует определенного мышления и базовых навыков в формулировании поисковых запросов и отбрасывании ложных результатов. Увы, в обычной школе нет времени обучать основным навыкам работы в Интернете, хотя школьникам было бы полезно освоить активный поиск информации, а не только пассивное получение знаний на уроке.

Udacity – образовательная организация, разрабатывающая онлайн-курсы, которую основал Себастьян Трун, прославившийся благодаря автомобилям-беспилотникам. У нее есть ряд программ, после прохождения которых выдается диплом об окончании краткосрочных курсов повышения квалификации. Помимо предоставления бесплатного доступа к курсам, Udacity сотрудничает с компаниями, которые хотят поднять уровень своих сотрудников. Udacity создает курсы, адаптированные к потребностям компании, и сотрудники заинтересованы проходить их. Это беспроигрышный вариант для работодателей, персонала и Udacity. Сектор образования вне традиционных школ быстро развивается, и онлайн-

курсы могут предоставлять новые решения для непрерывного обучения.

Наш следующий онлайн-курс «Mindshift: Преодолейте препятствия к обучению и откройте свой скрытый потенциал» был запущен в апреле 2017-го[332]. Он сопровождался новой книгой Барбары Оакли[333]. Я часть ее книги, и на моем примере иллюстрируют проблемы, которые возникают, когда вы хотите каким-то образом изменить свою жизнь, основываясь на чужом опыте. В моем случае это был переход от физики к биологии, в еще чьем-то – отказ от успешной карьеры музыканта ради работы врачом. Смена профессии становится все более распространенным явлением, и «Mindshift» был разработан, чтобы облегчить процесс.

Еще один способ заставить мозг лучше учиться – интерактивные компьютерные игры. Такие компании, как Lumosity[334], предлагают игры, в которые можно играть онлайн, попутно, по их словам, улучшая память и внимание. Проблема в том, что исследования, подкрепляющие такие утверждения, часто отсутствуют или у них низкое качество, особенно когда дело касается переноса полученных при обучении навыков на реальные задачи. Но это только начало, и более качественные исследования помогают нам разобраться, что работает, а что нет. Результаты часто неожиданные и противоречивые.

Тренировка мозга

Наиболее эффективные для улучшении когнитивных функций те игры, где вы должны преследовать зомби или убивать плохих парней на войне, а также игры-гонки. Дафна Бавельер из Женевского университета показала, что игра в некоторые шутеры от первого лица, такие как Medal of Honor: Allied Assault, развивают восприятие, внимание и когнитивные функции (рис. 12.9)[335]. Видеоигры в жанре экшен улучшают различные навыки, включая зрительное восприятие, многозадачность, переключение между задачами и быстрое принятие решений. Это удивительно, так как данные игры не предназначались для улучшения когнитивных функций и основная выгода была в увеличении скорости реакции. Бавельер пришла к выводу, что некоторые игры могут заставить мозг пожилого человека реагировать так же быстро, как мозг молодого, и это хорошая новость для стареющих людей. Однако шутеры также могут снизить обучаемость в долгосрочной перспективе[336]. У каждой игры есть свои достоинства и недостатки, которые необходимо рассматривать отдельно.

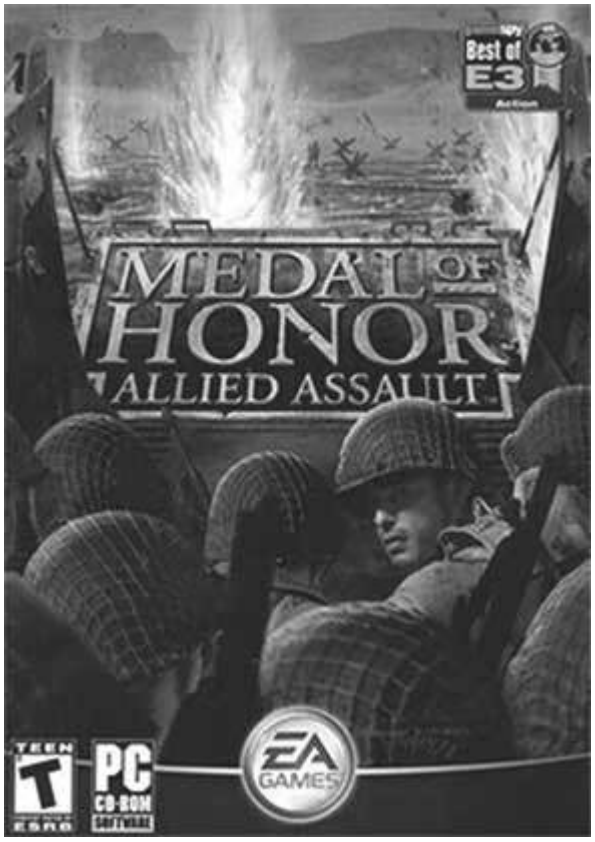


Рис. 12.9. Игра Medal of Honor: Allied Assault. Улучшает реакцию и учит справляться с когнитивными задачами

Адам Газзали из Калифорнийского университета в Сан-Франциско создал гоночную игру, которая улучшает вашу многозадачность. Она основана на исследованиях, показывающих, что активность нейромодуляторов в мозге важна для внимания, обучения и памяти. Адам специально разработал трехмерную игру NeuroRacer, в которой игроки управляют автомобилем на извилистой холмистой дороге, отслеживая одни случайно появляющиеся знаки и игнорируя другие[337]. Игра требует от участников делать несколько дел одновременно и использовать несколько когнитивных навыков, таких как внимание и переключение между задачами. После тренировки игроки значительно улучшили свои показатели кратковременной памяти и способности долго удерживать внимание на одной задаче, что не было частью обучения. Кроме того, их

результаты оказались лучше, чем у нетренированных 20-летних, и приобретенные навыки сохранились через шесть месяцев без практики. NeuroRacer сейчас находится на стадии клинических испытаний в качестве терапии для пациентов с дефицитом внимания и памяти[338].

В 1997 году Паула Таллал из Ратгерского университета и Майкл Мерзенич из Калифорнийского университета в Сан-Франциско основали компанию Scientific Learning Corp. для детей с нарушениями речи и чтения (дислексией). Понимание речи зависит от восприятия на слух быстро изменяющихся звуков. Например, слышим мы «ба», «га» или «да», зависит от миллисекундной разницы в колебании звука в начале слога. Дети, которые не могут обнаружить разницу, плохо учатся, так как путают слова с данными звуками. Чтобы научиться читать, ребенок должен усвоить, какие буквы обозначают эти короткие звуки. Таллал и Мерзенич разработали большую серию компьютерных игр Fast ForWord[339], которая улучшает слуховое восприятие, речь и понимание прочитанного, сначала подчеркивая акустические различия в слогах, словах и предложениях, а затем постепенно снижая их с развитием у ребенка навыков речи и чтения уровень за уровнем[340]. Программу Fast ForWord используют 6000 школ и свыше 2,5 миллиона детей, у нее самый высокий рейтинг среди обучающих игр. Как минимум в 55 странах ее применяют, чтобы помочь детям изучить английский в качестве второго языка. Мерзенич приступил к разработке BrainHQ – игры, основанной на тех же научных принципах и направленной на то, чтобы замедлить снижение когнитивных функций в пожилом возрасте[341].

С помощью упражнений для мозга вы также можете улучшить свои двигательные навыки. Аарон Зайтц из Калифорнийского университета в Риверсайде разработал компьютерную программу, которая обостряет зрительное восприятие и уменьшает время реакции. После того как университетская бейсбольная команда

использовала эту программу, у игроков улучшилось зрение, у них стало меньше аутов и больше пробежек, и в конечном итоге они выиграли четыре из пяти дополнительных игр в сезоне из 54 игр[342]. Зайтц разработал недорогое приложение UltimEyes, которое сделало его исследование открытым для общественности. Федеральная торговая комиссия США прекратила ее распространение до тех пор, пока не будут завершены дополнительные исследования, подтверждающие эффективность программы[343]. Будь я бейсболистом, я бы не стал ждать одобрения правительства.

Несмотря на успешные лабораторные эксперименты, есть опасения, что действительно хорошие результаты в одной игре не отражаются на остальных когнитивных навыках. Распространяют свой эффект, как правило, игры для улучшения реакции, но многие другие игры нацелены на одну определенную область, например на развитие памяти. Мы сильно продвинулись в разработке интерактивных компьютерных игр, улучшающих функционирование мозга, в которые интересно играть и которые можно выпустить в виде приложения, но необходимо больше исследований, чтобы понять, при каких условиях происходит развитие навыков. Потенциал для развития людей по всему миру огромен.

Искусственный интеллект и бизнес

На открытии Конференции NIPS в 2015 году я приветствовал участников в куртке как у автогонщика, на которой были 42 логотипа наших спонсоров (рис. 2.10). Конференцию в 2016 году в Барселоне поддержали 65 спонсоров – слишком много эмблем, чтобы разместить их на куртке. Этот взрывной рост в конечном счете прекратится, но его отголоски в обществе могут звучать десятилетиями. Компании отправляют на конференцию рекрутеров, стремясь нанять талантливых исследователей, которых так не хватает. Многие из моих коллег получили работу в Google, Microsoft,

Amazon, Apple, Facebook, Baidu[344] и многих стартапах. Это лишило университеты талантов. Себастьян Трун подсчитал, что, когда крупная интернет-компания покупает молодую, та обходится ей в 10 миллионов долларов на одного эксперта по машинному обучению[345].

Рис. 12.10. Куртка NASCAR (Национальной ассоциации гонок серийных автомобилей), в которой я был на открытии конференции NIPS 2015 года в Монреале. В число спонсоров вошли как ведущие интернет-компании, так и финансовые и медиакомпании. Все они заинтересованы в глубоком обучении

Джеффри Хинтон стал сотрудником Google в 2013 году, когда корпорация купила его компанию DNNresearch, в которую входили Джеффри и два его аспиранта из Университета Торонто. Теперь у

него есть доступ к компьютерам большей мощности, чем он мог мечтать в Торонто, и, что гораздо важнее, – к огромным массивам данных, которые получает Google. Google Brain – уникальная команда талантливых инженеров и ученых, собранная Джеффом Дином[346], разработчиком файловой системы MapReduce, от которой зависят все сервисы Google. Когда вы что-то переводите с помощью Google, он использует глубокое обучение от команды Google Brain. Когда вы что-то ищите в Google, глубокое обучение помогает выстраивать результаты в нужном порядке. Когда вы общаетесь с голосовым помощником Google, тот прибегает к глубокому обучению, чтобы распознать слова, которые вы говорите, и, разговаривая с вами, он будет использовать глубокое обучение, чтобы служить вам лучше. Google и вся индустрия высоких технологий просто без ума от глубокого обучения, и это только начало.

США теряют лидерство в области ИИ, и к тому времени, когда вы читаете это, другие страны, возможно, уже ушли далеко вперед[347]. В марте 2017 года в Торонто при поддержке правительства Канады и провинции Онтарио, Университета Торонто и частных компаний был открыт научно-исследовательский центр Vector Institute[348]. Его цель – стать ведущим мировым центром по изучению ИИ, давать высшее образование в области машинного обучения и стать ядром суперкластера ИИ, который двигает экономику Торонто, Онтарио и Канады. Серьезную конкуренцию Канаде составляет Китай, который готовит тысячи инженеров по машинному обучению, а нейроморфные вычисления – одно из двух направлений проекта China Brain. Поражение Кэ Цзе в игре с AlphaGo оказало влияние на Китай подобное тому, какое первый искусственный спутник Земли, запущенный СССР в 1957 году, оказал на США. Пекин тратит миллиарды на поддержку области ИИ, финансируя амбициозные проекты, стартапы и научные исследования, чтобы выйти в лидеры к 2030 году[349]. У Китая в распоряжении огромное количество

медицинских и личных данных и, в отличие от западных демократий, он меньше беспокоится о конфиденциальности. А у кого больше данных, тот и выигрывает, что меняет расклад для Китая.

Более того, Китай хочет «интегрировать ИИ в управляемые ракеты, использовать его для отслеживания людей с помощью камер видеонаблюдения, осуществлять цензуру в Интернете и даже предсказывать преступления»[350]. Между тем политические лидеры в США планируют сократить финансирование научных исследований и разработки технологий. В 1960-х годах США вложили 100 миллиардов долларов в космическую гонку[351], которая привела к созданию спутниковой индустрии, дала США лидерство в области микроэлектроники и материалов и с политической трибуны заявила о сильных сторонах страны в науке и технике. Те инвестиции все еще продолжают окупаться, поскольку это единственные отрасли, где США по-прежнему конкурентоспособны. Сегодня на вершину пьедестала рвется Китай, активно финансируя гонку ИИ, и эти инвестиции могут обеспечить им лидерство в нескольких ключевых отраслях в XXI веке.

Современные приложения с ИИ основаны на фундаментальных исследованиях, проведенных 30 лет назад. Приложения через 30 лет будут зависеть от той теоретической работы, что проводят сегодня, но лучшие из лучших исследователей трудятся на промышленность и сосредоточены на продуктах и услугах, которые внедрят в ближайшем будущем. Нам очень не хватает вычислительных мощностей, чтобы достичь человеческого уровня интеллекта. Сейчас в сетях глубокого обучения миллионы единиц и миллиарды весов. Это в десять тысяч раз меньше, чем число нейронов и синапсов в коре головного мозга человека, где на каждый кубический миллиметр приходится миллиард синапсов. Если бы все датчики в мире были подключены к Интернету и соединены между собой глубокими учебными сетями, однажды он мог бы проснуться и сказать: «Привет, мир!»[352]

Глава 13. Эпоха алгоритмов

В июне 2016 года я был в Сингапуре, где в Наньянском технологическом университете в течение недели проходило обсуждение «Фундаментальные проблемы науки». Темы дискуссий были самыми разными: от космологии и эволюции до государственной политики в отношении науки[353]. Брайан Артур – экономист, сильно интересующийся информационными технологиями[354], – говорил об алгоритмах. Он отметил, что в прошлом технологии основывались на законах физики, которые описывались дифференциальными уравнениями. В XX веке мы добились глубокого понимания физического мира, используя уравнения и математику непрерывных переменных[355] как главный источник идей. Непрерывная переменная плавно изменяется во времени и пространстве. Однако в основе технологий сегодняшнего дня лежат алгоритмы. В XXI веке мы успешно постигаем природу сложности[356] в компьютерных науках и биологии с помощью дискретной математики и алгоритмов. Артур преподает в Институте Санта-Фе в Нью-Мексико – одном из многих центров, возникших в XX веке для исследования сложных систем[357].

Алгоритмы окружают нас. Вы используете алгоритмы каждый раз, когда что-то гуглите[358]. Новости, которые вы читаете в ленте новостей Facebook, выбираются по алгоритму, основанному на истории ваших просмотров, что влияет на ваш эмоциональный отклик[359]. Алгоритмы внедряются в вашу жизнь все быстрее, поскольку глубокое обучение дает вашему смартфону возможность распознавать речь и естественный язык.

Что такое алгоритм? Алгоритм – это процесс, выполняющийся шаг за шагом, или набор правил, которым необходимо следовать при выполнении расчетов или решении задачи. Слово «алгоритм» происходит от латинского *algorismus*, составленного из имени Аль-

Хорезми, персидского математика IX века, и греческого слова *arithmos* – «число». Хотя алгоритмы зародились очень давно, цифровые компьютеры выдвинули их на передний план науки и техники.

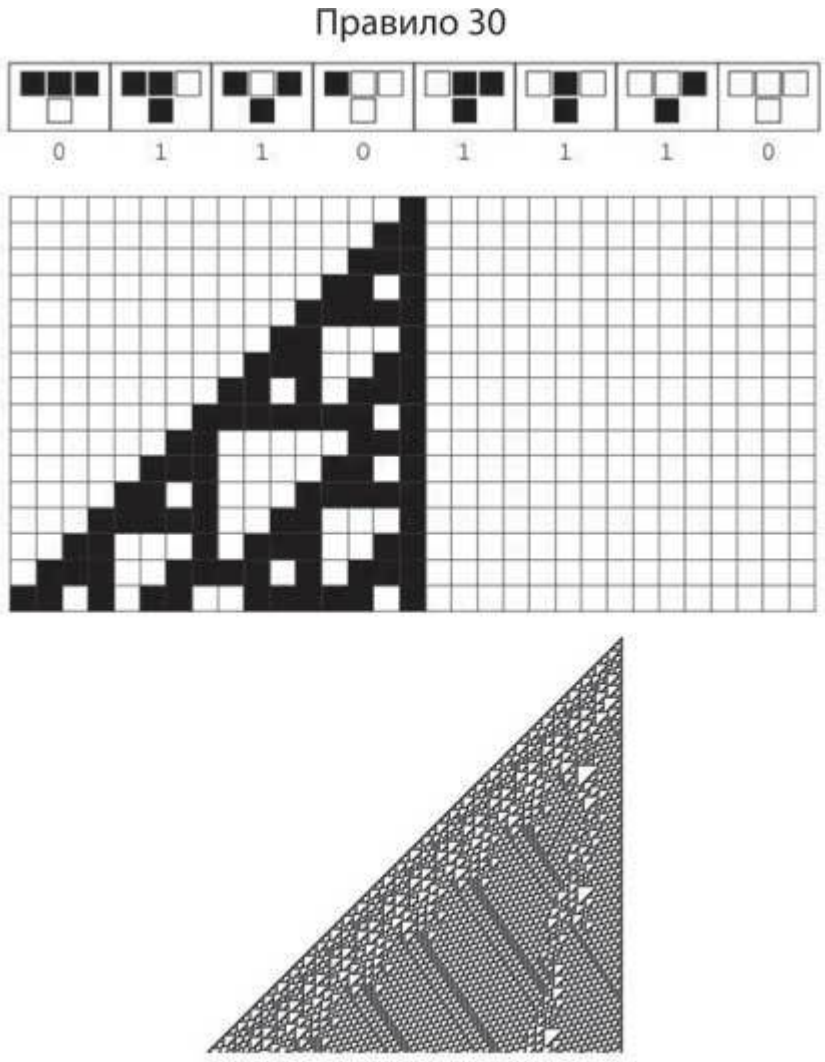
Сложные системы

В 1980-х годах случился расцвет новых подходов к сложным системам. Целью была разработка современных способов изучения систем, как те, что мы видим в природе сложнее, чем физика и химия. То, как летит ракета, несложно объяснить законами Ньютона, но не было простого способа описать дерево или то, как оно растет. Первопроходцы в области ИИ использовали компьютерные алгоритмы для изучения извечных вопросов о живых существах.

Рис. 13.1. Стивен Вольфрам у себя дома в Конкорде в штате Массачусетс стоит на полу, который сгенерировал алгоритм. Вольфрам – один из родоначальников теории сложности, и он показал, что даже простые программы могут создавать сложность подобно тем, с которыми мы сталкиваемся в реальном мире

Стюарт Кауфман получил медицинское образование, и его сильно заинтересовали генетические сети, в которых белки, называемые факторами транскрипции, могут нацеливаться на гены и влиять на их активацию[360]. Его модели были самоорганизующимися и основывались на сетях из двоичных единиц, схожих с нейронными сетями, но намного медленнее. Крис Лэнгтон ввел термин «искусственная жизнь» в конце 1980-х годов[361], что привело к неоднократным попыткам понять принципы, которые лежат в основе сложности живых клеток и развития сложных форм поведения. Несмотря на прогресс, тайна жизни продолжает ускользать от нас. Между тем клеточная биология и молекулярная генетика выявили высокую сложность молекулярных механизмов внутри клеток.

Блок 7. Клеточный автомат



Правило клеточных автоматов определяет цвет ячейки в зависимости от ее цвета и цвета ближайших ячеек. Например, для восьми возможных комбинаций черного и белого для трех ячеек в верхнем ряду, правило 30 указывает следующий цвет под ними. Эволюция этого правила, применяемого к одной строке за раз, начиная с одиночной черной ячейки, показана ниже для 15 шагов и еще ниже для 250 шагов. Изначально простое условие превращается в очень сложную схему, которую можно продолжать бесконечно.

Откуда берется эта сложность? Подробности описаны в книге Стивена Вольфрама «Новый вид науки», изданной в 2002 году.

Алгоритмы дают новые возможности для создания миров с уровнем сложности, сравнимым с нашим. Алгоритмы, открытые в XX веке, заставили нас переосмыслить природу сложности. Революция нейронных сетей в 1980-х годах стала еще одной попыткой осмыслить всю сложность мозга, и хотя модели были значительно проще, чем биологические нейронные сети, разработанные нами алгоритмы обучения позволили исследовать общие принципы, такие как распределение информации в больших популяциях нейронов. Но как сложные функции сетей возникают из относительно простых правил обучения? Есть ли еще более простая система, проявляющая сложность, которую легче анализировать?

Клеточный автомат

Еще одна яркая фигура с серьезным научным подходом к сложности – Стивен Вольфрам (рис. 13.1), основавший Центр исследований сложных систем в Университете Иллинойса в 1986 году. Он был вундеркиндом, в 20 лет получил докторскую степень по физике в Калтехе, став самым молодым из тех, кому это удалось. Стивен решил, что нейронные сети слишком сложны, и стал исследовать клеточные автоматы.

У клеточных автоматов обычно лишь несколько дискретных значений, которые изменяются со временем в зависимости от состояния других клеток. Один из простейших клеточных автоматов – одномерный массив ячеек, каждая из которых имеет значение «0» или «1» (блок 7). Пожалуй, самый известный клеточный автомат – игра «Жизнь», которую в 1968 году изобрел Джон Конвей, Фоннеймановский профессор из Принстонского университета, и популяризировал Мартин Гарднер в своей колонке «Математические

игры» в журнале Scientific American. Игра показана на рис. 13.2. Доска представляет собой двумерный массив ячеек, которые могут быть включены или выключены, и правило обновления зависит только от четырех ближайших соседей. При каждом шаге обновляются все состояния. В массиве генерируются сложные шаблоны, часть даже имеет имена – например, «планеры», которые пролетают через массив и сталкиваются с другими шаблонами. Начальные условия крайне важны для поиска конфигурации, отображающей сложные шаблоны.

Насколько распространены правила, создающие сложность? Стивен хотел узнать простейшее правило клеточных автоматов, которое может привести к сложному поведению, и поэтому начал перебирать их одно за другим. Правила под номерами от 0 до 29 создавали шаблоны, которые всегда возвращались к скучному поведению: в итоге все ячейки имели либо повторяющийся рисунок, либо фрактальный, с вложенными копиями самого себя. Однако правило 30 поражало непрерывно изменяющимися сложными моделями (блок 7). В конечном счете было доказано, что «правило 110» способно к универсальным вычислениям. То есть некоторые из простейших клеточных автоматов обладают возможностями машины Тьюринга, которая способна вычислить любую вычислимую функцию, поэтому она теоретически столь же мощна, как и любой компьютер.

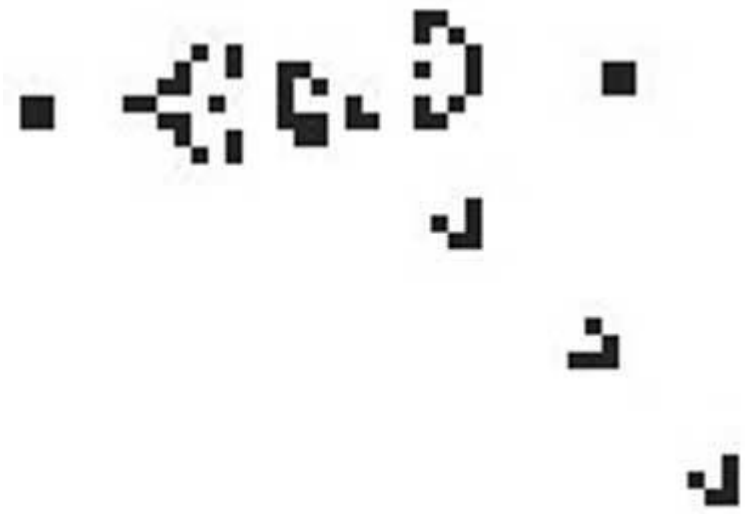


Рис. 13.2. Game of life. Снимок Планерного ружья Госпера (сверху), которое излучает последовательность «планеров», движущихся по диагонали, от «материнского корабля» сверху к правому нижнему углу

Одно из следствий этого открытия – вывод, что удивительная сложность, которую мы находим в природе, могла методом проб и ошибок развиться в простейшей среде химического взаимодействия между молекулами. То, что в ходе эволюции возникнут сложные комбинации молекул, ожидаемо и не должно считаться чудом. Однако клеточные автоматы – не достаточно хорошая модель зарождения жизни, и остается открытым вопрос, какие простые химические системы способны создавать сложные молекулы[362]. Возможно, только особые биохимические системы обладают таким свойством, и это сужает вероятный набор взаимодействий, из которых могла возникнуть жизнь. Теперь мы знаем, что избыточность[363] в мозге основана на разнообразии, а не на дублировании.

Важнейшее свойство жизни – способность клетки к самовоспроизведению. Джон фон Нейман из Института перспективных исследований в Принстоне прорабатывал этот вопрос в 1940-х годах с использованием клеточных автоматов. Фон Нейман – венгерский ученый, оказавший сильное влияние на

многие области математики, включая его основополагающие работы по теории игр, упомянутые в главе 1. Какой простейший клеточный автомат может точно воспроизвести себя? Фон Нейман нашел очень сложный клеточный автомат с 29 внутренними состояниями и большим объемом памяти, позволяющим тому самовоспроизводиться. Это имеет определенный биологический интерес, так как у клеток с такой же способностью есть много внутренних состояний и память, выраженная в виде ДНК. С тех пор были найдены еще более простые клеточные автоматы, умеющие самовоспроизводиться.

Мозг – это компьютер?

В 1943 году Уоррен Маккалок и Уолтер Питтс показали, что можно построить цифровой компьютер с помощью простых двоичных элементов с заданным порогом, таких как перцептрон, который можно включить в компьютер в качестве элементарного логического вентиля[364]. Теперь мы знаем, что мозг обладает смешанными аналоговыми и цифровыми свойствами и что нейронные сети обычно не вычисляют логические функции. Но в то время эта статья привлекла много внимания и, в частности, вдохновила Джона фон Неймана задуматься о компьютерах. Он построил один из первых цифровых компьютеров, в котором хранились программы, – необычный проект для математика того времени. Когда в 1957 году фон Нейман умер, Институт перспективных исследований не продолжил его начинание и выбросил компьютер[365].

Фон Нейман также интересовался мозгом. В своих Силлимановских лекциях[366] в Йельском университете он размышлял о том, как мозг может надежно функционировать с такими ненадежными компонентами[367]. Когда транзистор в цифровом компьютере допускает ошибку, весь компьютер может выйти из строя, но когда нейрон в мозге дает сбой, остальная часть мозга адаптируется к сбою и продолжает работать. Фон Нейман полагал, что причиной устойчивости мозга может быть запас «лишних» связей, так как в каждой операции участвует множество нейронов. Избыточность, как правило, нужна для резервной копии на случай отказа основной системы. Но сейчас мы знаем, что избыточность в мозге основана на разнообразии, а не на дублировании. Фон Неймана также волновала логическая глубина: сколько логических шагов может сделать мозг, прежде чем накопленные ошибки испортят результат. В отличие от компьютера, который может отлично выполнять каждый логический шаг, в мозге множество источников помех. Мозг не может достичь совершенства,

но поскольку так много нейронов работают параллельно и одновременно, за каждый шаг он выполняет гораздо больше, чем компьютер, и ему требуется меньшая логическая глубина.

Пространство алгоритмов

Сколько всего алгоритмов? Представьте себе пространство всех возможных алгоритмов. Каждая точка в пространстве – алгоритм, который что-то делает. Некоторые из них удивительно полезны и удобны. В прошлом их создавали вручную математики и программисты, трудясь как ремесленники в артели. Стивен Вольфрам автоматизировал процесс для клеточных автоматов путем полного перебора алгоритмов, начиная с самых простых, некоторые из которых выдавали очень сложные рисунки. Этот принцип обобщен в выведенном Вольфрамом правиле, которое гласит: вам не нужно углубляться в пространство алгоритмов, чтобы найти тот, что решает интересующий вас класс проблем. Примерно как отправлять ботов играть в StarCraft в Сети, чтобы опробовать все возможные стратегии. Согласно правилу Вольфрама, где-то во вселенной алгоритмов должна быть галактика алгоритмов, которые приведут к победе.

Вольфрам сосредоточился на пространстве клеточных автоматов – небольшой части в пространстве всех возможных алгоритмов. Теперь у нас есть подтверждение правила Вольфрама и в пространстве нейронных сетей. Каждая сеть глубокого обучения была найдена с помощью обучающего алгоритма, который представляет собой метаалгоритм для поиска новых алгоритмов. Для большой сети и большого набора данных обучение из разного исходного состояния может создавать галактику сетей, примерно одинаково хороших в решении проблемы. Встает вопрос, есть ли более быстрый способ найти область пространства алгоритма, чем градиентный спуск – медленный и требующий уйму данных. На такую возможность намекает то, что каждый вид представлен

множеством отдельных особей, созданных вариантными последовательностями ДНК вокруг точки в пространстве живых алгоритмов, и природе удалось перепрыгнуть из одного множества в другое путем естественного отбора в результате скачкообразного процесса, называемого прерывистым равновесием[368], одновременно с локальным поиском случайных мутаций. Генетические алгоритмы были разработаны, чтобы совершать скачки, подобно тому, как в ходе эволюции в природе появляются новые организмы[369]. Нам нужна математика для описания множества этих алгоритмов. Кто знает, как выглядит вселенная алгоритмов? Есть еще много галактик алгоритмов, которые мы еще не открыли, но можем найти с помощью автоматического поиска. Это последний рубеж.

Простому примеру такой обработки последовал Клаус Штифель, научный сотрудник моей лаборатории, использовавший алгоритм, который вырастил в компьютере нейроны со сложными дендритными деревьями[370]. Дендриты подобны антеннам, которые собирают входные данные от других нейронов. Пространство возможных дендритных деревьев огромно, и цель состояла в том, чтобы указать желаемую функцию и найти в пространстве дендритных деревьев модельный нейрон, который вычислит функцию. Одно из полезных свойств – определять, в каком порядке сигналы поступают на вход: когда конкретный входящий сигнал приходит раньше другого, нейрон должен отправлять импульс, но если тот поступает позже, нейрон должен молчать. Такой модельный нейрон нашли перебором всех возможных дендритных деревьев с помощью генетического алгоритма, и решение выглядело как кортикальный пирамидальный нейрон с синапсом на тонком дендрите, выходящем снизу (базальный дендрит), и другим синапсом на толстом дендрите, выходящем из вершины (апикальный дендрит) (рис. 14.6). Возможно, это объясняет, почему пирамидальные клетки имеют апикальные и базальные дендриты, роль которых невозможно было бы представить без глубокого

поиска в пространстве всех вероятных дендритов. Повторяя поиск для других функций, можно автоматически составить их словарь в зависимости от формы дендритов, и, обнаружив новый нейрон, просто сверяться со справочником, чтобы определить его потенциальные функции.

Стивен Вольфрам покинул университетские стены и возглавил компанию Wolfram Research, которая создала Mathematica – программу, поддерживающую широкий спектр математических структур и массово использующуюся для практических приложений. Mathematica написана на языке Wolfram – основном мультипарадигмальном языке программирования[371], который также поддерживает Wolfram Alpha – первая рабочая система вопросов и ответов для фактов о мире, основанная на символьном подходе[372].

В академических кругах валютой считаются опубликованные статьи, но когда вы независимый исследователь, то можете сами издавать свои книги. Это было нормой на протяжении многих столетий, когда стать ученым могли себе позволить только состоятельные или нашедшие богатых покровителей люди. Вольфрам написал книгу «Новый вид науки» в 2002 году. Она весила 2,5 кг и содержала 1280 страниц, из которых 348 страниц занимали примечания, равноценные сотне новых научных статей. Книга вызвала большой ажиотаж в прессе, но получила неоднозначный отклик у сообщества исследователей сложных систем – некоторые ученые считали, что их работе уделено мало внимания. Из-за этого терялся смысл книги, которая должна была поместить предыдущие труды в новый контекст. Карл Линней разработал современную систему классификации животных, став важным предшественником Дарвина, теория эволюции которого позволила связать предыдущие попытки упорядочить виды. Как отличить первопроходца от последователей? У первопроходцев стрелы в спине. По следу, который проложил Стивен Вольфрам, теперь идет новое поколение исследователей.

В 1980-х годах Стивен Вольфрам скептически относился к тому, что реальный мир будет тесно связан с нейросетями, и они не имели большого влияния еще 30 лет. Однако прогресс за последние пять лет изменил его мнение, и Стивен признал, что он, как и многие другие исследователи, недооценил то, что может быть достигнуто[373]. Но кто мог предсказать, насколько хороши будут нейронные сети? Язык Wolfram, который используется в Mathematica, теперь поддерживает приложения для глубокого обучения, он и стал первым языком, который обеспечил онлайн-распознавание объектов на изображениях[374].

В 1987 году Стивен Вольфрам познакомил меня с Беатрис Голомб, когда я посещал Сан-Диего. В то время она работала над исследованиями для своей докторской диссертации. Стивен позвонил Беатрис – сказать, что она должна присутствовать на моем выступлении (она планировала, так как посещала еженедельные собрания «Параллельной и распределенной обработки»). И он позвонил мне – сказать, что его подруга Беатрис будет на моем выступлении (позже он позвонил нам, чтобы спросить, как все прошло). Несколько лет спустя я переехал в Сан-Диего, и мы с ней действительно обручились. В 1990 году, поженившись в «Атенеуме»[375] в Калтехе, мы отправились в лекториум Бекмана[376] на брачный симпозиум. Беатрис в свадебном платье выступила на тему «Брак: теория и практика». Моя лекция была о поддержании накала страстей в долгосрочной перспективе. Стивен с гордостью «покаялся», рассказав, как он нас познакомил. Беатрис отметила, что если он получил признание своих заслуг, то несет ответственность, а если он несет ответственность – то он им должен. Стивен увильнул от ответа.

Глава 14. Привет, мистер Чип

Мы видим рождение новой архитектуры в индустрии компьютерных микросхем. Идет гонка за разработку и создание нового поколения чипов, чтобы глубокое обучение, обучение с подкреплением и другие обучающие алгоритмы работали в тысячи раз быстрее и эффективнее, чем сейчас моделируют на компьютерах общего назначения. Новые сверхбольшие интегральные схемы (СБИС) имеют архитектуру параллельной обработки и память, разделенную между процессорами, чтобы снизить нагрузку на узкое место между памятью и центральным процессором в последовательной архитектуре фон Неймана, которая преобладала в работе компьютерных систем последние 50 лет. В том, что касается технических средств, мы все еще находимся на этапе исследований, и у каждого типа специализированной СБИС есть свои сильные и слабые стороны. Для запуска крупномасштабных сетей, которые разрабатываются для приложений с ИИ, потребуются огромные компьютерные мощности, и создание эффективной системы принесет гигантские прибыли.

И крупные компании, производящие микросхемы, и мелкие стартапы много вкладывают в эту цель. В 2016 году, например, компания Intel приобрела Nervana – небольшую свежесозданную компанию в Сан-Диего, которая разработала специальные СБИС для глубокого обучения, а бывший генеральный директор Nervana Навин Рао теперь возглавляет их новую группу продуктов ИИ[377], которая напрямую подчиняется генеральному директору Intel. В 2017 году Intel за 15,3 миллиарда долларов купила Mobileye – компанию, которая специализируется на датчиках и компьютерном зрении для беспилотных автомобилей. Компания Nvidia, разработавшая специальные цифровые чипы, оптимизированные для графических приложений и игр, называемые графическими процессорами (graphics processing unit; GPU), теперь продает больше микросхем,

предназначенных для глубокого обучения и облачных вычислениях. Google разработала особый чип – тензорный процессор (tensor processing unit; TPU), – чтобы обеспечить глубокое обучение своих интернет-сервисов с гораздо меньшим энергопотреблением.

Программное обеспечение для глубокого обучения также важно для разработки приложений. TensorFlow – программа для запуска сетей глубокого обучения, которую Google выложила в открытый доступ. Возможно, все не так альтруистично, как кажется: когда Google сделала систему Android бесплатной, это дало компании контроль над операционной системой, которую сейчас используют на большинстве смартфонов по всему миру. Но есть альтернатива: у CNTK[378] компании Microsoft также открытый исходный код; MVNet поддерживается Amazon и другими крупными интернет-компаниями, такие среды для глубокого обучения, как Caffe, Theano и PyTorch, составляют им конкуренцию.

Горячие чипы

В 2011 году в городе Тромсё в Норвегии я организовал спонсируемый Фондом Кавли[379] семинар «Развитие высокопроизводительных вычислений в экологически чистой среде»[380]. Мы подсчитали, что при нынешних микропроцессорных технологиях для экзафлопсных вычислений[381], которые в тысячу раз быстрее петафлопсных[382], потребуется 50-мегаваттная электростанция – больше, чем мощность, потребляемая метро в Нью-Йорке. Значит, следующему поколению суперкомпьютеров для работы нужны микросхемы с низким энергопотреблением, таких как чипы, созданные компанией ARM, которые были оптимизированы для смартфонов. Вскоре станет нецелесообразно использовать цифровые компьютеры общего назначения для наиболее ресурсоемких приложений, и будут доминировать чипы специального назначения, как это уже произошло в мобильных телефонах.

В человеческом мозге около ста миллиардов нейронов, каждый из которых соединен с несколькими тысячами других, что в сумме доходит до квадрильона[383] синаптических связей. Энергетический бюджет мозга – около 20 ватт, около 20 процентов от энергопотребления всего тела, хотя мозг весит лишь три процента от общей массы. Напротив, суперкомпьютер с производительностью, исчисляемой в петафлопсах, потребляет 5 мегаватт и даже близко не приближается к мощности вашего мозга. Природа добилась этого, уменьшив части нейронов, необходимые для связи и передачи сигнала, до молекулярного уровня. Еще одно отличие – плотность размещения компонентов: транзисторы на микросхеме расположены на двумерной поверхности, а в мозге соединения находятся в трехмерном пространстве, что позволяет минимизировать объем. Природа давно открыла эти технологии, и нам еще предстоит наверстать упущенное.

Рис. 14.1. Карвер Мид примерно в то время, когда он создал в Калтехе первый кремниевый компилятор. Мид был провидцем, чьи идеи и технологические достижения оказали значительное влияние как на цифровые, так и на аналоговые вычисления. Телефон на

снимке указывает на время, когда была сделана фотография

Глубокое обучение требует больших вычислительных ресурсов и сейчас выполняется на централизованных серверах, а результаты передаются на периферийные устройства, такие как мобильные телефоны. В конечном счете, периферийные устройства должны стать автономными. Это потребует принципиально иного оборудования, намного легче и потребляющего меньше энергии, чем облачные вычисления. Интересно, что такое оборудование уже существует – нейроморфные чипы, созданные по подобию мозга.

Холодные чипы

Я впервые встретил Карвера Мида (рис. 14.1) в 1983 году на семинаре, проводившемся на курорте неподалеку от Питтсбурга. Джеффри Хинтон собрал небольшую группу, чтобы исследовать, куда движутся нейронные сети. Мид известен своим крупным вкладом в компьютерные науки. Он первым осознал, что по мере того, как транзисторы на СБИС становятся все меньше и меньше, чипы становятся все эффективнее, и поэтому вычислительная мощность должна продолжать расти в течение длительного времени. Карвер Мид ввел в обращение термин «закон Мура», основанный на наблюдении Гордона Мура[384], что количество транзисторов на чипах удваивалось каждые 18 месяцев. Он уже прославился изобретением кремниевого компилятора – программы, которая автоматически размещала схемы проводников и функциональные модули системного уровня на чипе[385]. До кремниевого компилятора каждый чип инженеры изготавливали вручную на основе опыта и интуиции. По сути, Мид предложил программировать компьютеры, чтобы те сами разрабатывали чипы. Это были первые шаги в наноинженерии.

Мид – провидец. В то же время когда мы сидели за столом в маленькой комнате на семинаре, наверху проходила конференция по суперкомпьютерам. Крупные суперкомпьютерные компании, такие как Cray Inc. и Control Data Corporation, проектировали оборудование специального назначения, которое было в сотни раз быстрее, чем компьютеры в наших лабораториях, и стоило сто миллионов долларов. Компьютеры Crays были настолько быстрыми, что их приходилось охлаждать жидким фреоном. Мид сказал мне, что они еще не знают, но микропроцессоры захватят их долю рынка и суперкомпьютерные компании скоро исчезнут. Микропроцессоры в персональных компьютерах в ту эпоху были значительно медленнее, чем чипы специального назначения в суперкомпьютерах. Но микропроцессоры развивались быстрее, чем суперкомпьютеры, из-за неуклонного сокращения затрат и повышения производительности, которое стало возможным благодаря уменьшению основных размеров устройства. Вычислительная мощность микропроцессора в вашем смартфоне равна мощности десяти суперкомпьютеров Cray X-MP в 1980-х годах, а высокопроизводительные суперкомпьютеры с сотнями тысяч микропроцессорных ядер достигли петафлопсной производительности, что в миллион раз быстрее, чем у вымерших компьютеров Cray, при их одинаковой стоимости с учетом инфляции.

На семинаре Мид показал нам кремниевую сетчатку, которая была создана по той же технологии, что и чипы СБИС, но с использованием аналоговых, а не цифровых схем. В аналоговой схеме напряжение на затворах может непрерывно изменяться, тогда как у напряжения на затворах в цифровой схеме может быть только одно из двух значений: «включено» или «выключено». В нашей сетчатке свыше ста миллионов фоторецепторов, но в отличие от камеры, которая просто передает фотонные импульсы в память, сетчатка имеет несколько уровней нейронной обработки, которая преобразует входящие визуальные данные в эффективные нейронные коды. Все стадии обработки аналоговые, пока не доходят

до ганглиозных клеток, которые несут в мозг по миллиону аксонов, закодированные сигналы в виде двоичных импульсов. Обозначение импульсов как «да» или «нет» похоже на цифровую логику, но время прохождения импульса является аналоговой переменной, и нет часов, что превратят последовательность импульсов в гибридный год.

В чипе сетчатки, разработанном Мидом, ступенчатая часть обработки выполнялась с напряжением, немного не доходящим до порогового значения «выключено», тогда как работающий в цифровом режиме транзистор быстро переходит в полностью «включенное» состояние, которое потребляет гораздо больше энергии. Как следствие, аналоговому чипу СБИС требуется лишь малая доля мощности цифровых микросхем – от нановатт до микроватт, а не от милливатт до ватт, – что делает их в миллионы раз энергоэффективнее. В своей книге «Аналоговые СБИС и нейронные системы», вышедшей в 1989 году, Мид показал, что нейронные алгоритмы, встроенные в нейронные цепи глаз насекомых и млекопитающих, можно эффективно воспроизвести в кремнии. Карвер Мид – основатель нейроморфной инженерии, цель которой – создание чипов на основе алгоритмов мозга.

Чип сетчатки – изобретение Миши Маховальд, звездной аспирантки Мида (рис. 14.2)[386]. В своих идеях она объединила опыт бакалавра биологии Калтехе и дипломную работу в области электротехники. Это привело ее к получению четырех патентов. В 1992 году Маховальд вручили премию Милтона и Фрэнсиса Клаузеров[387] за диссертацию, посвященную чипу, который выполнял сопоставление бинокулярных изображений в реальном времени, первому чипу, который использовал естественное коллективное поведение для сложной задачи. В 1996 году она была занесена в зал славы международного союза «Женщины в технологии» (Women in Technology International; WITI).

Рис. 14.2. Миша Маховальд из Калтеха в то время, когда она, будучи ученицей Карвера Мида, создала первую кремниевую сетчатку. Маховальд внесла выдающийся вклад в нейроморфную инженерию

Физика транзисторов в околопороговом режиме весьма схожа с биофизикой ионных каналов в биологических мембранах. Миша работала с нейробиологами Кеваном Мартином и Родни Дугласом из Оксфордского университета над кремниевыми нейронами[388] (рис. 14.3) и переехала с коллегами в Цюрих, чтобы помочь основать Институт нейроинформатики при Цюрихском университете и Швейцарском федеральном технологическом институте. Ее сияющая звезда трагически закатилась в 33 года, когда страдающая от депрессии Маховальд бросилась под поезд.

Карвер Мид покинул Калтех в 1999 году и переехал в Сиэтл. Я посещал его в 2010 году. С его заднего двора можно видеть самолеты, которые, пролетая над водой, заходят на посадку в аэропорт Сиэтл-Такома. Его отец был инженером на гидроэлектростанции Биг-Крик – целом комплексе электростанций в Калифорнии на реке Сан-Хоакин, берущей исток в горах Сьерра-

Невады[389].

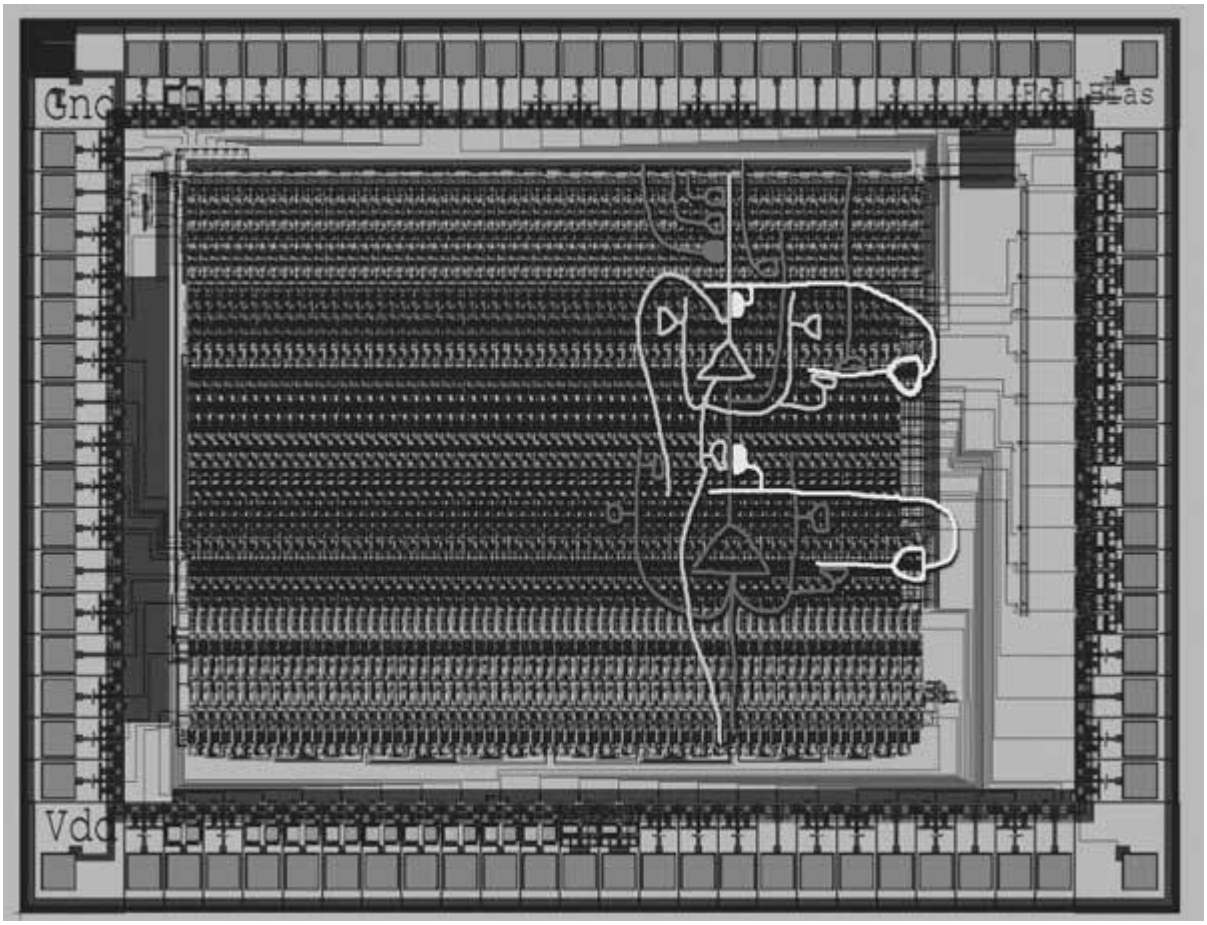


Рис. 14.3. Кремниевые нейроны. Каналы в этом аналоговом чипе СБИС действуют как ионные каналы в нейронах и могут имитировать нейронные сети в реальном времени, что показано на рисунке поверх чипа

Скачок всего за одно поколение от ранней гидроэлектрической технологии к микроэлектронике захватывает дух. Хобби Карвепа – коллекционирование старинных изоляторов для подвешивания линий электропередач. Они валяются буквально под ногами, как наконечники индейских стрел, если вы знаете, где их искать. У Мида также был лазерный гироскоп, который он использовал для тестирования нового разработанного им подхода для объяснения

квантовой физики[390]. Карвер Мид – провидец, но таким успешным его сделала привычка создавать вещи, которые работают и которые вы можете подержать в руках.

Нейроморфная инженерия

В 1990 году я в качестве почетного ученого Фэйрчайлда был в академическом отпуске в Калтехе и походил на министра без портфеля. Мне нравилось сидеть на заседаниях лабораторий, особенно лаборатории Кристофа Коха – специалиста по вычислительной нейробиологии, с которым у меня были общие интересы, и «Карверленда» – исследовательской группы Карвера Мида. Одним из удивительных проектов «Карверленда» была кремниевая улитка с такой же частотой колебаний в каналах, как в настоящей улитке в ухе. Другие работали над искусственными синапсами, включая алгоритмы синаптической пластичности, так как на кремнии можно было реализовать долгосрочные изменения весов. С тех пор студенты из «Карверленда» разошлись по инженерными факультетам по всему миру.

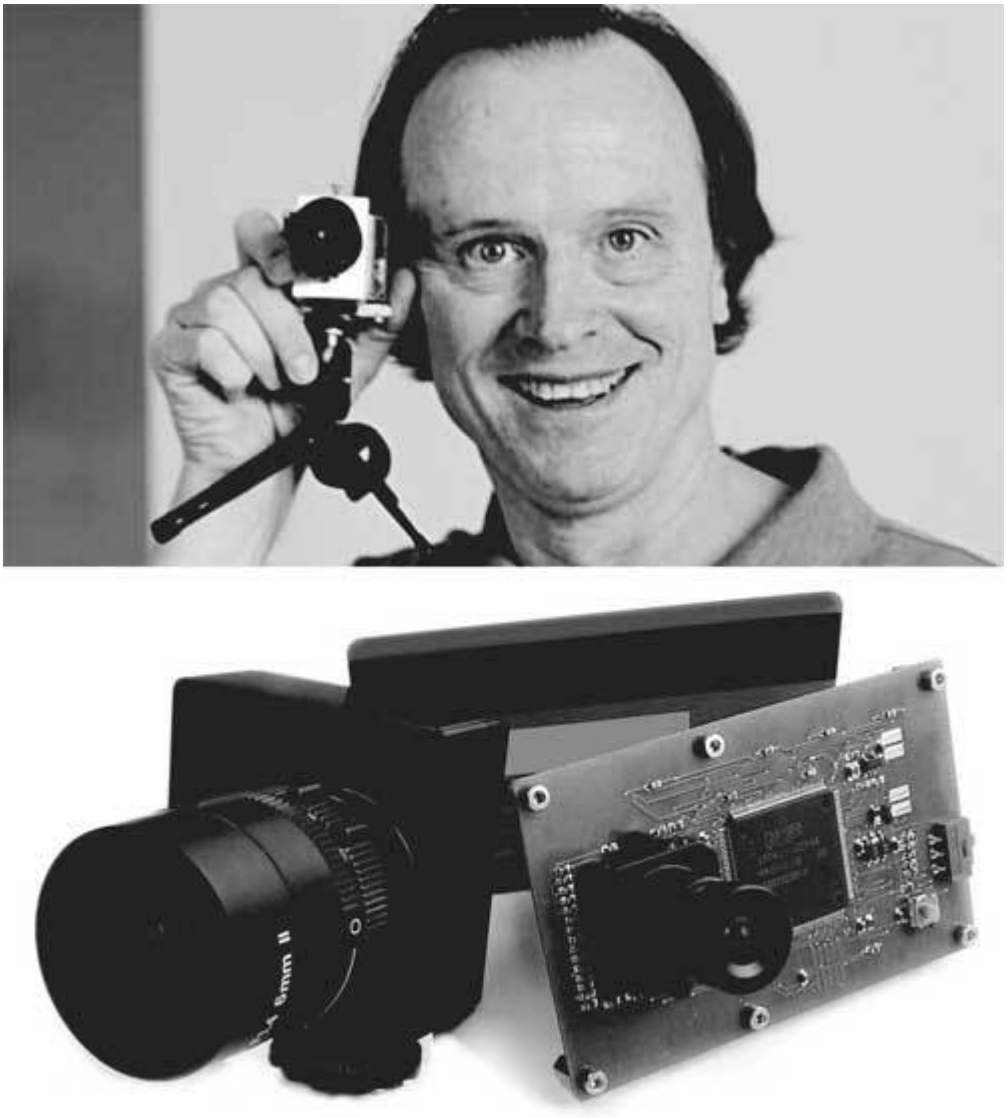
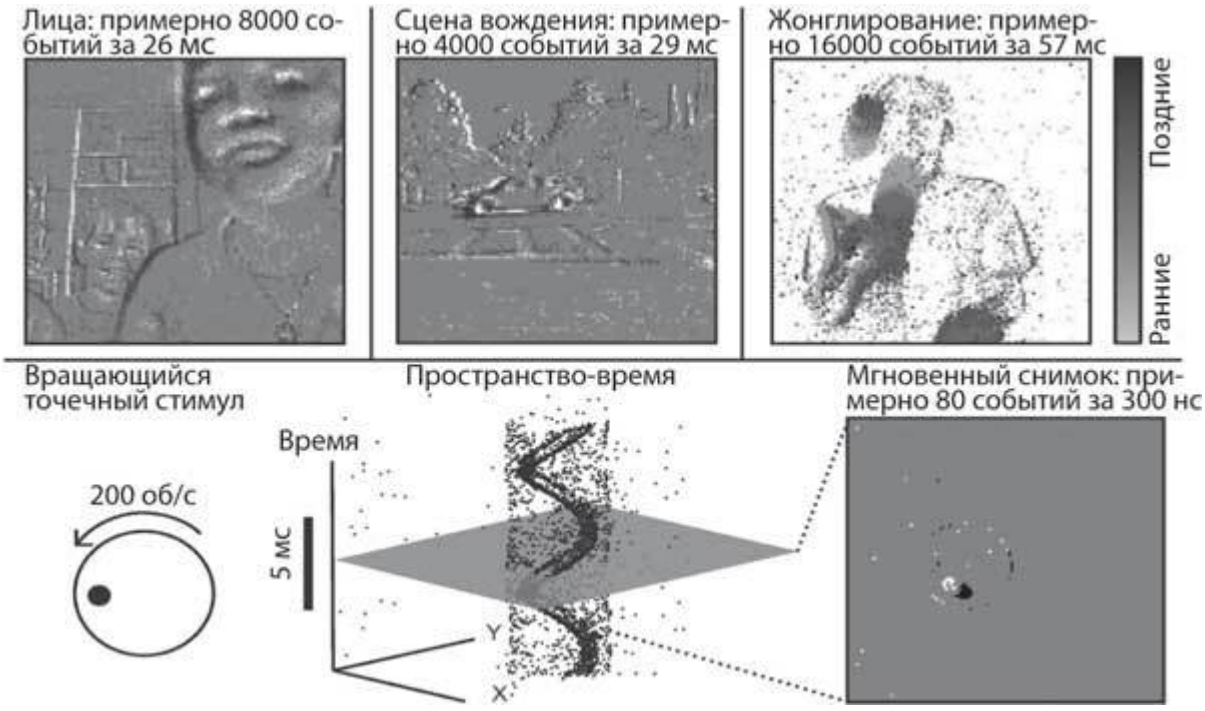


Рис. 14.4. Датчик динамического зрения (Dynamic Vision Sensor; DVS). Вверху: Тоби Дельбрюк держит в руках DVS-камеру, изобретенную им в Институте нейроинформатики Цюрихского университета. Камера – это особый чип, который асинхронно излучает импульсы, а не делает кадры, как ваша цифровая камера. Внизу: камера оснащена объективом, который фокусирует изображения на аналоговом кристалле СБИС, регистрирующем постепенное увеличение и уменьшение интенсивности света в каждом пикселе. Для положительного приращения импульсы идут вдоль «включенного» соединения, для отрицательного – вдоль «выключенного». Выходные импульсы обрабатываются печатной

платой, которая отображает рисунок импульсов, как показано в блоке 8. Сетчатка – это высокоразвитая DVS-камера. Схема импульсов на сетчатке преобразуется в мозге, но остается схемой импульсов – нигде в вашем мозгу не появляется картинка, даже если все так воспринимается

В 1993 году мы с Кристофом Кохом и Родни Дугласом при поддержке Национального научного фонда организовали в Теллурайде в штате Колорадо трехнедельный семинар по нейроморфной инженерии, который продолжает проходить каждое лето в июле. Семинар международный, в нем принимают участие студенты и преподаватели из разных стран и с разным опытом. В отличие от множества семинаров, на которых больше говорят, чем работают, на семинаре в Теллурайде аудитории заполнены студентами, трудящимися над чипами и использующими их для создания роботов. Однако возникла проблема с подключением чипа сетчатки к чипу зрительной коры, а чипов коры – к чипам двигательных реакций.

Блок 8. Как работает DVS-камера



В этих кадрах с DVS-камеры белые точки – импульсы от «включенных» каналов, а черные – от «выключенных». Серый цвет указывает на отсутствие импульсов. На изображении в левом верхнем углу можно увидеть два лица, так как они немного перемещались в течение 26 мс. В правом верхнем углу, где изображен процесс жонглирования, точки на момент входа обозначены серым цветом с разным уровнем яркости, чтобы вы могли видеть траекторию. Диск в левом нижнем углу вращается со скоростью 200 оборотов в секунду. В средней панели траектория представляет собой спираль, движущуюся вверх. В кратком 300-микросекундном срезе спирали есть только 80 импульсов, и легко рассчитать скорость, измерив смещения черных и белых импульсов и поделив на временной интервал. Обратите внимание, что цифровая камера с длительностью кадра 26 мс не сможет следить за пятном, вращающимся с частотой 200 Гц, потому что период вращения составляет 5 мс, и на каждом кадре получится кольцо. Единственное, что остается на выходе из камеры – поток импульсов, как и в сетчатке. Это эффективный способ представления

изображения, так как большинство пикселей молчат основную часть времени и каждый импульс несет полезную информацию.

Способ подключить несколько аналоговых СБИС – использовать импульсы, как и реализовано в мозге. Более половины коры – белое вещество, которое состоит из нейронных отростков дальнего действия. Провода довольно дороги, и было бы невозможно подключить чип сетчатки к чипу коры головного мозга миллионом проводов. К счастью, быстрая цифровая логика может одновременно передавать несколько потоков по одному проводу, позволяя множеству клеток сетчатки взаимодействовать со множеством клеток коры при помощи одного и того же канала. Принимающий чип получает от передающего чипа адрес каждого исходящего импульса, адрес затем декодируется, и импульс направляется к нужному элементу. Это называется предоставлением адреса событий (Address Event Representation; AER).

Тоби Дельбрюк, который сейчас работает в Институте нейроинформатики при Цюрихском университете, был аспирантом Карвера Мида (рис. 14.4)[391]. В 2008 году он разработал весьма успешный чип сетчатки, названный датчиком динамического зрения (Dynamic Vision Sensor; DVS), что значительно упростило такие задачи, как отслеживание движущихся объектов и тщательный поиск деталей с помощью двух камер (см. рис. 14.4)[392]. Обычные цифровые камеры фиксируют последовательность кадров, которые длятся по 26 мс. В каждом кадре теряется информация: представьте диск с пятном, вращающийся со скоростью 200 оборотов в секунду; пятно будет совершать полный круг пять раз в каждом кадре и при воспроизведении записи выглядеть как статическое кольцо (блок 8). Камера Дельбрюка может отслеживать движущееся пятно с точностью до микросекунды, используя единичные импульсы, что делает ее быстрой и эффективной. Камера DVS – первая из нового класса датчиков, основанных на импульсах и их длительности. Она

обладает большим потенциалом для усовершенствования многих изобретений, в том числе беспилотных автомобилей. Один из проектов на конференции в Теллурайде предлагал использовать DVS-камеры для защиты небольших футбольных ворот от попадания в них мяча (рис. 14.5).

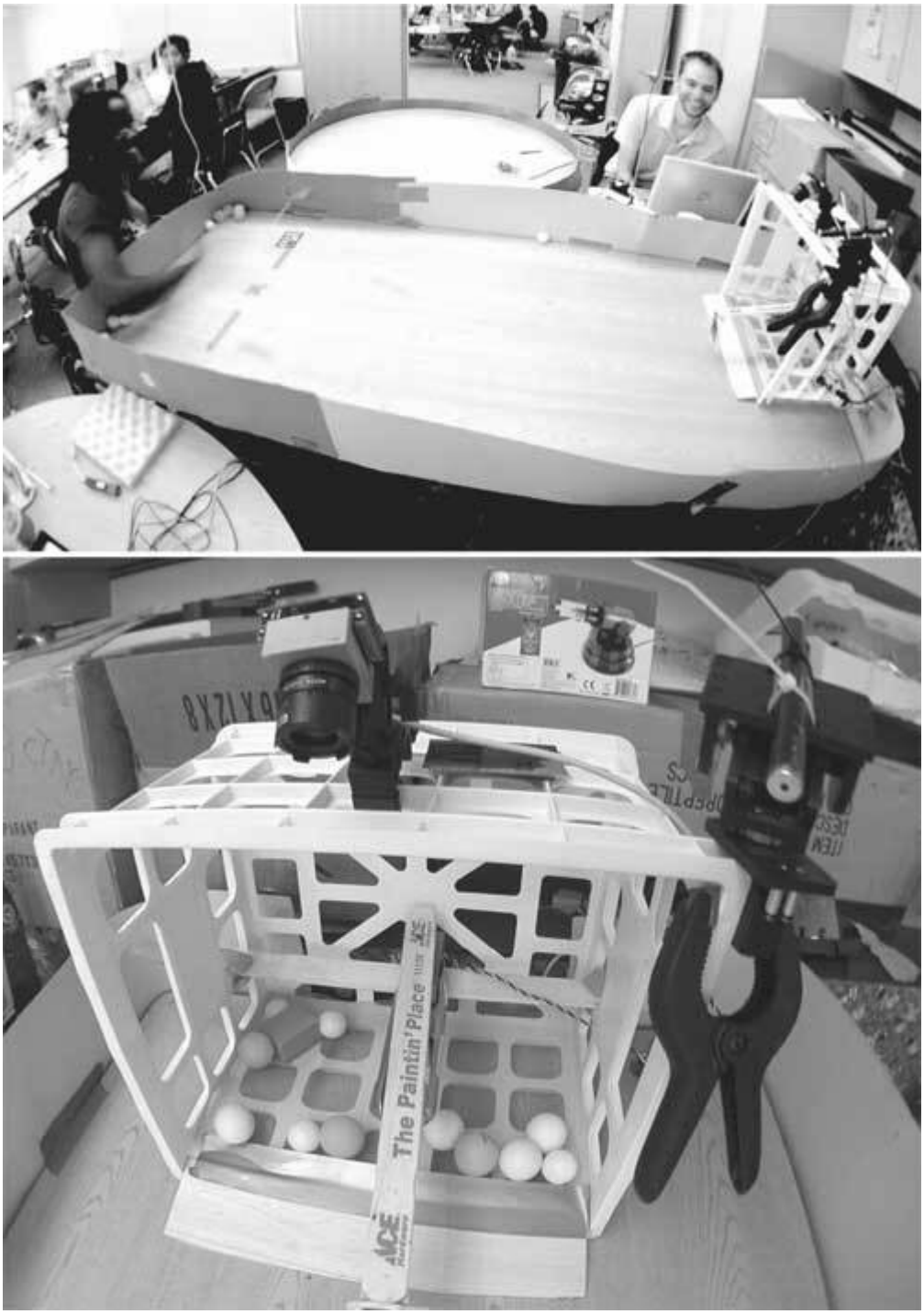


Рис. 14.5. Нейроморфный вратарь на семинаре в Теллурайде в 2013 году. Сверху: Фопефолу Фолоуоселе (слева) тестирует нейроморфного вратаря. На заднем плане можно увидеть других участников и их проекты. Внизу: DVS-камера Дельбрюка направляет деревянную лопатку для защиты ворот. Вратарь гораздо быстрее

студентов и успешно защищает ворота. Я тоже сделал попытку и не смог забить (не обращайте внимания на мячи в сетке)

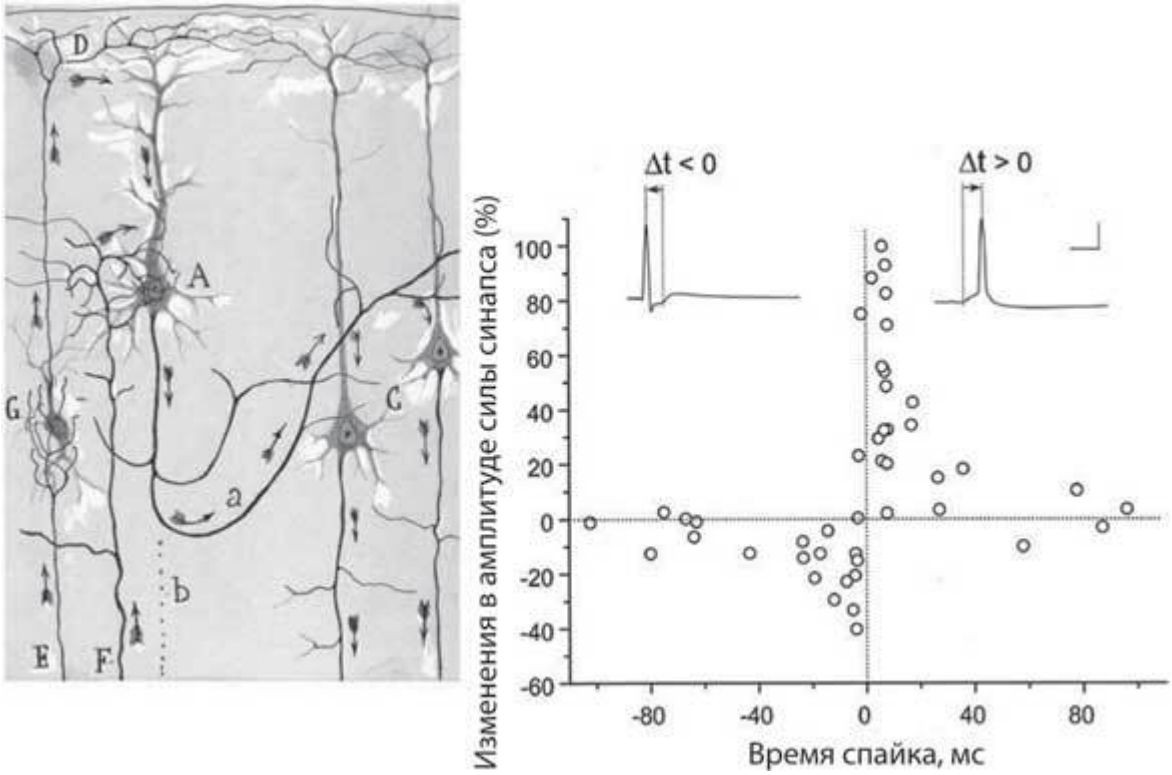


Рис. 14.6. Пластичность, зависящая от времени импульса. Слева: чертеж пирамидных нейронов коры, сделанный Рамоном-и-Кахалем, знаменитым испанским нейроанатомом. Выходной аксон нейрона А устанавливает синаптические связи с дендритом нейрона С (как показано стрелками). Справа: два нейрона, подобные тем, что слева, проткнули электродом и стимулировали к образованию спайков[393] с временной задержкой между волнами. Когда входящий в нейрон импульс многократно совпадает с выходным импульсом, сила синапса (вертикальная ось) может либо увеличиваться, если предсинаптический входящий сигнал поступает до постсинаптического в пределах 20-миллисекундного окна (горизонтальная ось), либо, наоборот, уменьшаться

Спайковые нейроны открывают новые вычислительные возможности. Например, время импульсов в популяции нейронов можно использовать для регулирования того, какая информация сохраняется. В 1997 году Генри Маркрам и Берт Сакман из Германии сообщили, что они научились как увеличивать, так и уменьшать синаптические силы, используя повторное объединение входящего в синапс импульса с выходным импульсом в постсинаптическом нейроне: если вход произошел в пределах 20-милисекундного окна до выходного импульса, то следует долговременное усиление, но если вход произошел в течение 20 миллисекунд после выходного импульса, то следует долговременное ослабление (рис. 14.6). Синаптическая пластичность, зависящая от времени импульса (Spike-time dependent plasticity; STDP), была обнаружена в разных частях мозга, и, вероятно, она важна для формирования долгосрочной памяти. Интересно, что STDP – лучшее объяснение постулата Хебба, обсуждавшего в главе 7[394].

По распространенному мнению, суть пластичности Хебба в том, что сила синапса должна увеличиваться при одновременной подаче импульса на вход и выход нейрона, вид обнаружения совпадений[395]. Но на самом деле Хебб писал: «Когда аксон клетки А находится достаточно близко, чтобы возбудить клетку В, и неоднократно или постоянно принимает участие в ее возбуждении, в одной или обеих клетках происходит некий процесс роста или метаболических изменений, так что эффективность клетки А, возбуждаемой клеткой В, увеличивается»[396]. Чтобы клетка А способствовала активации клетки В, клетка А должна запустить спайк до спайка в клетке В. Здесь не только корреляция, но и причинно-следственные связи. Хебб ничего не говорил об условиях уменьшения силы синапса, но когда входной импульс испускается после выходного, он с меньшей вероятностью будет причинно связан с выходным нейроном, и отключение синапса имело бы смысл при необходимости сбалансировать увеличение и уменьшение силы в долгосрочной перспективе.

В Теллурайде ведутся споры между сторонниками аналоговых СБИС и разработчиками цифровых технологий. Аналоговые СБИС имеют много достоинств, потребляя мало энергии при параллельной работе всех цепей, но у них есть и недостатки, такие как варьирование транзисторов, в результате чего одинаково спроектированные транзисторы производят токи, которые могут отличаться на 50 процентов в ту или иную сторону. Цифровые СБИС по сравнению с ними точнее, быстрее и легче в реализации, но требуют намного больше энергии. Команда Дхармендры Модха из IBM Almaden Research Center в Сан-Хосе в Калифорнии разработала цифровой чип, содержащий 4096 вычислительных ядер и 5,4 миллиарда вентилей, названный TrueNorth[397]. Его можно настроить для имитации миллиона нейронов, соединенных 268 миллионами синапсов, при этом чип потребляет всего 70 милливатт. Однако сила этих синапсов фиксирована, что ограничивает реализацию многих важных функций, таких как ослабление или укрепление. Многое еще предстоит узнать, создавая сети с различными связями, чтобы увидеть, как они ведут себя в реальном времени.

Еще один недостаток сетей со спайковыми нейронами – градиентный спуск, который стимулировал обучение в сетях с непрерывно оцениваемыми нейронами, стал невозможен из-за разрывов во время спайков. Это ограничивало сложность того, чему можно научить такую сеть. Градиентный спуск был чрезвычайно успешен при обучении глубоких сетей с модельными нейронами, у которых непрерывно изменяется скорость вывода, поэтому функция вывода дифференцируема, что является важной особенностью алгоритма обратного распространения ошибки. В спайковых сетях при прохождении импульса есть разрывы, и такая сеть не дифференцируемая. Недавно[398] это преодолел Хо Донсон, научный сотрудник моей лаборатории, который нашел способ заставить модели рекуррентных сетей со спайковыми нейронами выполнять растянутые во времени сложные задачи, используя градиентный спуск[399]. Стало возможным обучение глубоких спайковых сетей.

Конец закона Мура?

Закон Мура стоит за беспрецедентным увеличением компьютерной мощности более чем в триллион раз с 1950-х годов, когда были изобретены цифровые компьютеры. Никогда прежде ни одна технология не росла по такой экспоненте, что привело к встраиванию компьютеров почти в каждый искусственный объект, от игрушек до автомобилей. У современных телескопов адаптивная оптика, которую компьютеры могут автоматически регулировать для повышения разрешения. Микроскопы улавливают фотоны, и компьютеры анализируют их, чтобы локализовать молекулы со сверхвысоким разрешением. Каждая область науки и техники теперь зависит от чипов СБИС.

Карвер Мид предсказал такой рост, отталкиваясь от потенциальной возможности уменьшить размер логических элементов СБИС, но тот уже достиг физического предела: в проводах слишком мало электронов, и они склонны к утечке или блокировке случайными зарядами, что делает ненадежными даже цифровые схемы[400]. Закон Мура больше не работает? Чтобы продолжать увеличивать вычислительные мощности обработки, которые не зависят от идеальной точности цифровых дизайнов, необходима принципиально иная архитектура. Подобно тому, как гибридные автомобили объединили эффективность электрических двигателей и двигателя внутреннего сгорания, гибридная цифровая и нейроморфная архитектура использует преимущества низкой мощности нейроморфных микросхем для вычислений и высокую пропускную способность цифровых интегральных схем для передачи данных.

Поскольку параллельная архитектура продолжит развиваться в течение следующих 50 лет, закон Мура должен быть заменен законом, который учитывает как энергию, так и пропускную способность. По мере того как закон Мура переходит от одиночных

микросхем к массово-параллельной архитектуре, для работы с ней создаются новые алгоритмы. Но эти интегральные схемы должны взаимодействовать друг с другом, о чем мы поговорим в следующей главе.

Глава 15. Внутренняя информация

Мне никогда не приходило в голову, что когда-нибудь я стану всеведущим, что у меня действительно все для этого есть. Информация течет через Интернет со скоростью света. Легче получить факт из Интернета, чем из книги на моей полке. Мы переживаем информационный взрыв во многих его формах. Научные приборы, от телескопов до микроскопов, собирают все бóльшие и бóльшие наборы данных, которые анализируются с помощью машинного обучения. Агентство национальной безопасности использует машинное обучение для сортировки данных, собираемых по всему миру. Экономика становится цифровой, и навыки программирования востребованы у многих компаний. По мере того как мир переходит от индустриальной экономики к информационной, образование и профессиональная подготовка должны адаптироваться. Это уже оказывает сильное влияние на мир.

Информационная теория

В 1948 году Клод Шеннон (рис. 15.1) из AT&T Bell Laboratories в Мюррей-Хилл в штате Нью-Джерси предложил удивительно простую, но неочевидную теорию информации, позволившую понять, как передавать по телефонной линии сигнал, игнорируя шумы[401]. Теория Шеннона привела к революции в области цифровых коммуникаций, которая стала причиной появления сотовых телефонов, цифрового телевидения и Интернета. Когда вы звоните по сотовому телефону, ваш голос кодируется в биты и передается по радиоволнам на приемник, где цифровые сигналы декодируются и преобразуются в звуки. Теория информации накладывает ограничения на пропускную способность канала связи (рис. 15.2), и были разработаны коды, которые приближаются к

пределу Шеннона[402].

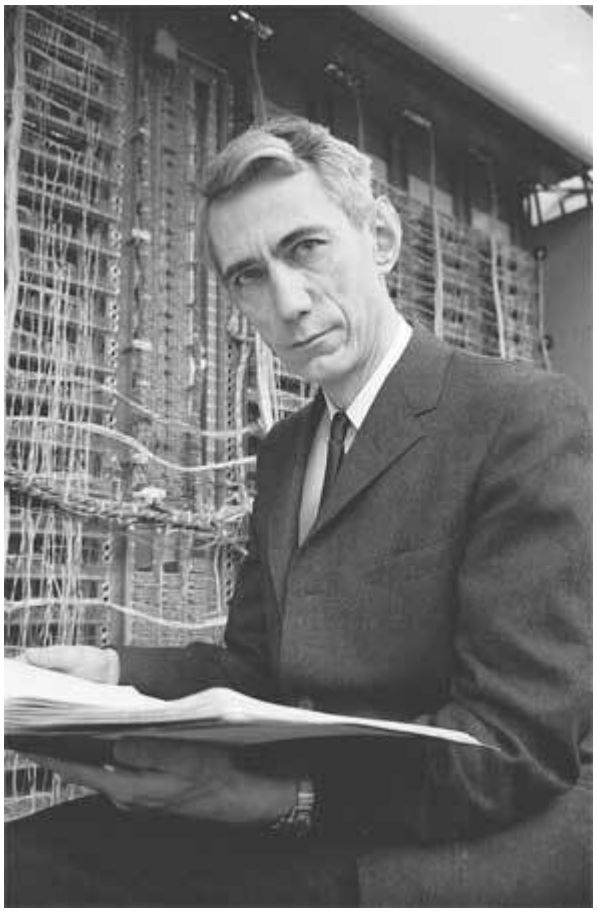


Рис. 15.1. Клод Шеннон перед телефонной коммутаторной сетью. Он работал в AT&T Bell Laboratories, когда создал теорию информации

Рис. 15.2. Модель коммуникационной системы Шеннона. Сообщение переводится в двоичный код и передается по каналу, которым может быть телефонная линия или радиоволна, туда, где оно принимается и декодируется. Пропускная способность канала зависит от уровня шума в системе

Несмотря на множество форм информации в мире, есть способ точно измерить объем набора данных. Единицей информации является двоичный разряд – бит, – который может принимать значение 1 или 0. Байт равен 8 битам. Информационное наполнение качественной фотографии измеряется в мегабайтах – миллионах байтов. Информация, хранящаяся в вашем мобильном телефоне, «весит» гигабайты – миллиарды байтов. Объем данных в Интернете считают в петабайтах – квадриллионах байтов.

Теория чисел

На международном симпозиуме по теории информации (International Symposium on Information Theory; ISIT) ежегодно присуждает премию Шеннона за выдающиеся исследования, и это большая честь. В 1985 году ISIT состоялся в Брайтоне в Великобритании, и премия Шеннона была присуждена Соломону

Голомбу (рис. 15.3) из Университета Южной Калифорнии, чья фундаментальная работа о последовательностях сдвиговых регистров стала основополагающей для современной цифровой связи[403]. Последовательность сдвигового регистра – алгоритм, который генерирует длинные псевдослучайные последовательности нулей и единиц. Каждый раз, когда вы звоните по сотовому телефону, вы используете последовательность сдвигового регистра. Голомб показал, как использовать последовательность сдвигового регистра для эффективного кодирования сигналов, которые затем могут быть переданы на приемник и декодированы. Если сложить все случаи, когда сотовые телефоны и другие системы связи генерировали последовательность сдвигового регистра, число будет ошеломляющим – свыше октиллиона раз (1000 000 000 000 000 000 000 000 000; 10^{27})[404].

Соломон Голомб был моим тестем, и однажды я спросил его, как он нашел такое элегантное решение проблемы коммуникации. Он сказал, что благодаря изучению теории чисел – одного из самых абстрактных разделов математики. Он познакомился с последовательностями сдвигового регистра во время летней практики в компании Glenn L. Martin Co. в Мэриленде. Получив в 1956 году в Гарвардском университете докторскую степень по математике в области теории чисел, он устроился на работу в Лабораторию реактивного движения в Калтехе, где возглавил группу связи и работу над космической связью. В дальние уголки Солнечной системы посылали космические зонды, и сигналы, поступающие обратно, были слабыми. Последовательности сдвиговых регистров и коды коррекции ошибок значительно улучшили передачу сигналов на космические зонды, а математика заложила основу для современных цифровых коммуникаций.

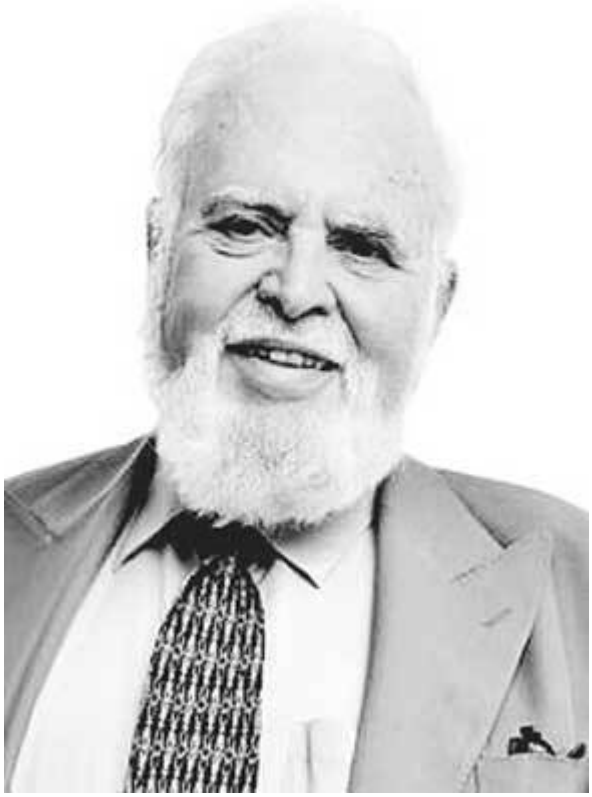


Рис. 15.3. Соломон Голомб. Математический анализ последовательностей сдвиговых регистров, проведенный им во время работы в Лаборатории реактивного движения в Калтехе в Пасадене, позволил связываться с зондами для исследования дальнего космоса, и позже эти регистры были встроены в системы сотовой связи. Каждый раз, когда вы используете свой мобильный телефон, вы используете его математические коды

Голомб нанял Эндрю Витерби, еще одного выдающегося специалиста в области теории информации, и познакомил его с Ирвином Джейкобсом из МТИ, которого пригласил посетить лабораторию, когда тот будет в творческом отпуске. Витерби и Джейкобс позже вместе основали компанию Qualcomm, которая произвела революцию в технологии сотовых телефонов, используя последовательности сдвиговых регистров для связи в частотном диапазоне, что эффективнее, чем работа на единой частоте. Эту

идею ранее высказывала Хеди Ламарр (рис. 15.4), киноактриса и изобретатель, получившая в 1941 году патент на скачкообразную перестройку частоты, которую она разработала в качестве безопасной системы связи для военных во время Второй мировой войны[405]. Когда Голомб покинул Лабораторию реактивного движения, чтобы стать преподавателем в Университете Южной Калифорнии, руководство его группой взял на себя основатель NIPS Эд Познер, но Голомб продолжал поддерживать их советами.

Математика, лежащая в основе последовательностей сдвиговых регистров, – самая темная часть теории чисел. Когда Голомб получил докторскую степень в Гарварде, его научный руководитель, как и большинство математиков того времени, гордился тем, что чистая математика никогда не найдет практического применения. Эту точку зрения разделял и ведущий специалист из Кембриджа Годфри Харолд Харди, автор известного эссе «Апология математика», в котором он писал, что «хорошая» математика должна быть чистой, а прикладная математика – «неинтересна». Некоторые ученые хотят, чтобы их математика оставалась теоретической, но они не могут помешать математике изменять им, решая практические задачи в реальном мире. Математика такая, какая она есть, не чистая и не прикладная. Карьеру Голомба во многом определили важные практические задачи, которые он мог решить, используя подходящие инструменты из «чистой математики».

Рис. 15.4. Хеди Ламарр. Звезда театра и кино, во время Второй мировой войны она совместно с Джорджем Антейлом изобрела скачкообразное изменение частоты, легшее в основу связи с расширенным спектром, используемой военными и в сотовых телефонах

Голомб также любил придумывать математические игры. Его книга «Полимино» познакомила мир с играми, в которых использовались фигуры, состоящие из квадратов (отсылка к фишкам домино из двух квадратов). Мартин Гарднер популяризировал их в своей колонке математических игр в журнале Scientific American. Тетрамино, фигуры из четырех квадратов, послужили источником вдохновения для создания тетриса – увлекательной игры, в которой тетрамино падают сверху и должны складываться в слоты внизу. Замощение плоскости с помощью полимино стало популярной настольной игрой, которая привела к появлению широкого круга интересных комбинаторных задач в математике.

Соломон Голомб также был библеистом и знал десятки языков, включая японский и китайский. Беатрис однажды принесла первое издание книги Дугласа Хофштадтера «Гёдель, Эшер, Бах: эта бесконечная гирлянда»[406]. Соломон открыл ее на фронтисписе. Подпись к рисунку гласила, что на нем первые двадцать строк «Книги Бытия» на древнееврейском языке. «Во-первых, все вверх

ногами», – сказал он и перевернул книгу. «Во-вторых, это не древнееврейский, а древнесамаритянский язык. В-третьих, это не первые двадцать строк «Бытия», а только первые семь слов каждой из первых двадцати строк Бытия». Затем он прочитал и перевел текст.

Клод Шеннон присутствовал на Ежегодной конференции по информационным наукам в Брайтоне, когда Голомб читал свою лекцию Шеннона. Это единственная лекция Шеннона, которую он посетил, кроме собственной. Ближе к концу жизни Шеннон боролся с слабоумием. Как быстро жизнь движется вперед, и даже великие люди остаются позади.

Прогностическое кодирование

В коммуникационной системе высока информационная ценность изменений, независимо от того, происходят они в пространстве или во времени. Изображение с равномерной интенсивностью несет мало информации, так же как и сигнал, который не меняется. Датчики, посылающие сигналы в мозг, в основном сообщают об изменениях, как мы уже видели на примере сетчатки в главе 5 и DVS-камеры Тоби Дельбрюка в главе 14. Когда изображение стабилизируется на сетчатке, оно исчезает через несколько секунд[407]. Мы не осознаем, что несколько раз в секунду наше глазное яблоко делает едва уловимые движения, называемые микросаккадами. Каждый такой рывок обновляет внутреннюю модель мира.

Когда что-то движется в поле зрения, сетчатка должным образом сообщает об этом вверх по цепочке, что также используется для обновления картины мира. Процесс проиллюстрирован на рис. 15.5. Модель в мозге иерархическая, и сравнение между поступающей сенсорной информацией и ожиданиями модели идет на нескольких уровнях[408]. Яркая вспышка или громкий звук немедленно привлекают ваше внимание. Но если вы заметили, что что-то на вашем столе изменилось, это будет представлено на гораздо более высоком уровне путем нисходящего сравнения с памятью. Все это происходит в мозге в реальном времени, и Карвер Мид часто повторял, что в мозге «время – его собственное представление»[409].

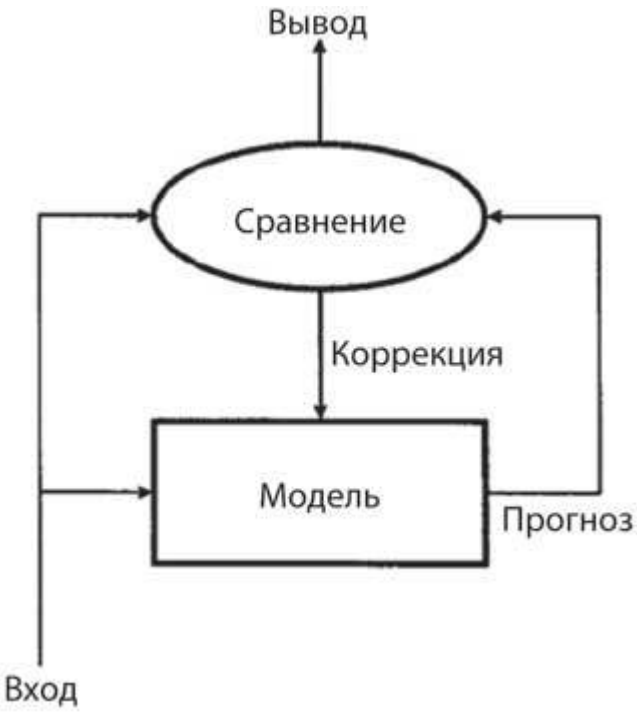


Рис. 15.5. Прогностическое кодирование. Упрощенная схема прохождения сигнала на одном уровне нейронной системы обработки. Модель получает входной сигнал и должна предсказать следующий; прогноз сравнивается с фактическим появлением входного сигнала. Когда модель точно предсказывает входные данные, информация не передается на следующий уровень, но когда модель не может предсказать входные данные, разница передается следующему слою и используется для корректировки модели. Эта операция отфильтровывает ожидаемые события на низких уровнях, позволяя центрам уровнями выше сосредоточиться на более многообещающих

Прогностическое кодирование[410] восходит к Гельмгольцу, который объяснил зрение как бессознательное умозаключение или формирование визуальной информации по нисходящей для отбрасывания шумов, достраивание неполной информации и интерпретацию увиденного[411]. Например, размер человека на сетчатке нашего глаза сигнализирует о том, насколько человек

далеко, так как нам известны его габариты и по опыту вы знаете, как размер на сетчатке изменяется с расстоянием. На более высоком когнитивном уровне Макклелланд и Румельхарт обнаружили, что, когда буквы расположены в слове, испытуемые определяют их быстрее, чем буквы в псевдословах без смысловой нагрузки[412]. Их модель параллельной обработки продемонстрировала аналогичное поведение, дав им уверенность в том, что они на правильном пути к пониманию того, как информация представлена в мозге.

Глобальный мозг

Мозг – это непревзойденная информационная машина. Американская правительственная программа BRAIN, запущенная 2 апреля 2013 года (рис. 15.6), направлена на создание новых нейротехнологий для ускорения темпов прогресса в понимании работы и проблем мозга. Как NIPS собрала исследователей из многих дисциплин для разработки обучающих машин, так программа BRAIN привлекает инженеров, математиков и физиков в нейробиологию, чтобы улучшить инструменты для исследования мозга. По мере того как мы узнаем больше о мозге и особенно о механизмах, лежащих в основе обучения и памяти, мы начинаем гораздо лучше понимать принципы работы мозга.

Хотя о мозге многое известно на молекулярном и клеточном уровнях, мы еще не недостаточно хорошо понимаем, как мозг организован в бóльших пространственных масштабах. Мы знаем, что разные типы информации хранятся в разных частях коры, но не знаем, как такая разрозненная информация извлекается для решения сложной задачи, например, соотнесение имени человека с изображением его лица. Этот вопрос тесно связан с происхождением сознания в мозге. Сотрудники моей лаборатории недавно выявили глобальные закономерности активности в мозге спящего человека, которые могут дать нам представление о том, как части информации, распределенные в коре, связаны между

собой[413].

Рис. 15.6. Снимок сделан в Белом доме незадолго до объявления о старте проекта BRAIN 2 апреля 2013 года. Представители организаций-участников (справа налево): Миён Чун, директор по науке Фонда Кавли, инициатор проекта; Уильям Ньюсом, сопредседатель консультативного комитета Национального института здравоохранения США по проекту BRAIN; Фрэнсис Коллинз, директор Национального института здравоохранения США; Джеральд Рубин, вице-президент Медицинского института Говарда Хьюза и исполнительный директор Исследовательского городка Джанелия; Кора Марретт, директор Национального научного фонда; Барак Обама, президент США; Эми Гутманн, председатель президентского комитета по биоэтике; Роберт Конн, президент Фонда Кавли; Арати Прабхакар, директор Управления перспективных исследовательских проектов Министерства обороны

США; Алан Джонс, исполнительный директор Алленовского института исследования мозга; Терри Сейновски, институт Солка

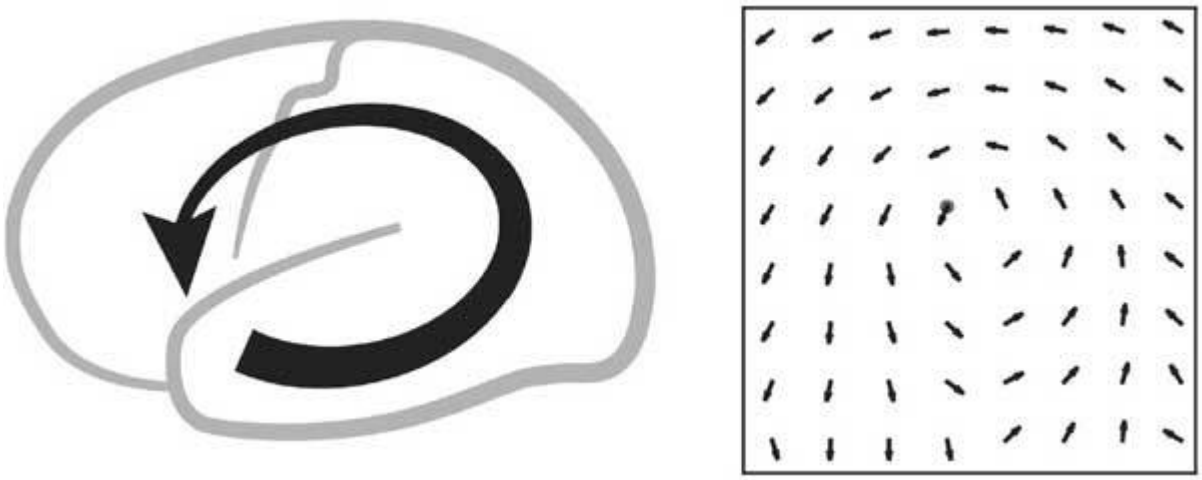


Рис. 15.7. Вращающиеся электрические волны в коре головного мозга человека. Записи с сетки электродов 8×8 на кортикальной поверхности во время веретен сна, которые участвуют в закреплении воспоминаний. Слева: веретена (сигма-ритм) – круговые волны, которые, если смотреть сбоку, проходят в коре головного мозга в направлении, указанном стрелкой, совершая цикл каждые 80 миллисекунд. Цикл многократно повторяется в течение ночи. Справа: маленькие стрелки показывают направление максимального нарастания фазы бегущей волны на 64 участках записи на поверхности коры

Между фазами медленного восстановительного сна и быстрого сна, сопровождающегося сновидениями, есть промежуточная стадия, когда в кортикальной активности преобладают высокосинхронизированные пространственно-временные колебания, называемые веретенами сна. Эти колебания частотой 10–14 Гц длятся несколько секунд и повторяются тысячи раз в течение ночи. Экспериментально доказано, что веретена сна участвуют в консолидации памяти, пока вы спите. На записях

биотоков коры мозга мы с Лайлом Мюллером, Эриком Халгреном и Сидом Кэшем обнаружили, что веретена сна – это единые круговые волны электрической активности, которые проносятся через все секторы коры (рис. 15.7). Мы назвали их волнами принцессы Леи, потому что они выглядят как ее прическа (рис. 15.8). Мы предположили, что веретена сна могут быть способом, которым кора объединяет новую информацию, полученную в течение дня, с распределенными в ней предыдущими воспоминаниями через укрепление длинных связей между ними. Это один из многих проектов из области системной нейробиологии, реализованный в рамках правительственной программы BRAIN.

Рис. 15.8. Кэрри Фишер в роли принцессы Леи. Ее волосы, собранные в два пучка, напоминают круговое течение полей, циркулирующих в коре во время сна

Операционные системы

Архитектура в нейронных сетях иная, чем в цифровых компьютерах. В цифровом компьютере память и центральный процессор (ЦП) разделены пространственно, и данные в памяти должны перемещаться в ЦП последовательно. В нейронных сетях обработка данных происходит в памяти параллельно, что решает проблему «бутылочного горлышка», а также позволяет выполнять массовую параллельную обработку, так как все блоки сети работают одновременно. Кроме того, в нейронных сетях нет различия между программным и аппаратным обеспечением. Обучение происходит путем внесения изменений в оборудование.

Блок 9. Операционные системы

Операционная система цифрового компьютера. Операционная система контролирует программы, которые выполняются на оборудовании. Для ПК чаще используют Windows, iPhone управляется iOS, а большинство серверов работают на одной из версии UNIX. Операционная система выделяет память, когда это необходимо программам. Она также работает «за кулисами»,

используя фоновые процессы, называемые демонами, чтобы отслеживать программы и внешние устройства, такие как принтеры и дисплеи. Операционная система предназначена для работы на любом оборудовании, позволяя запускать ваши приложения на разных компьютерах.

Цифровые компьютеры также становятся массивно-параллельными. Это началось в 1980-х годах, когда кластеры компьютеров были собраны в единый блок. Одним из ранних параллельных компьютеров был разработанный Дэнни Хиллисом Connection Machine[414], который выпускала компания Thinking Machines. Инженер и изобретатель Хиллис учился в МТИ, когда стало ясно, что для решения чрезвычайно сложных проблем реального мира потребуется гораздо бóльшие вычислительные мощности. Поскольку в 1990-х годах количество транзисторов на микросхемах продолжало расти согласно закону Мура, стало возможным установить много обрабатывающих модулей на одной микросхеме, много микросхем на одной плате, много плат в одном системном блоке и много системных блоков в комнате, поэтому сегодня самые быстрые компьютеры на планете имеют миллионы ядер и могут выполнять биллиарды[415] операций в секунду. Скоро нам будут доступны экзафлопсные вычисления.

Моделирование нейронных сетей будет максимально использовать преимущества массово-параллельного оборудования. Многочисленные ядра можно настроить для одновременной работы в единой сети, которая значительно ускоряет обработку, но затрудняет обмен данными между процессорами. Чтобы уменьшить задержки связи, компании строят специальные цифровые сопроцессоры, которые значительно ускорят моделирование сети, так что когнитивные задачи, такие как речь и зрение, получат единую мощную программу действий. Ваш смартфон станет намного

умнее, когда сети глубокого обучения превратятся лишь в одну из его микросхем.

Цифровые компьютеры работают под управлением операционных систем, которые отделяют вас от оборудования (блок 9). Когда вы запускаете текстовый редактор на своем ноутбуке или приложение на смартфоне, операционная система заботится обо всех нюансах: в какую область памяти поместить нажатия клавиш и как отобразить вывод на экран. Ваш разум выполняет роль приложений в операционной системе вашего мозга, которая отделяет вас от того, где, как и в каком виде лежит информация. Вы не знаете, как ваш мозг хранит обширные базы данных опыта, накопленного вами в течение жизни, или как этот опыт формирует ваше поведение. Можно проследить связь с отдельными переживаниями, но вы осознаете лишь малую часть. То, как ваш мозг управляет всем, остается тайной. Поняв, как работает операционная система мозга, мы организовали бы большие данные на основе тех же общих принципов. Тогда сознание можно было бы представить как приложение, работающее в операционной системе мозга.

Информация на всех уровнях

Информационный взрыв превратил биологию в количественную науку. Традиционно биологам не требовалось больше математической подготовки, чем вводный курс по статистике для анализа данных – невеликих и полученных с большим трудом. В 2002 году меня пригласили на конференцию в лабораторию Колд Спринг Харбор на Лонг-Айленде. Собрание было посвящено молекулярной генетике, и я чувствовал себя не в своей тарелке, потому что был единственным, кто читал лекцию о вычислениях. Передо мной выступал Ли Худ – молекулярный генетик, много лет проработавший в Калтехе. Когда я был там в творческом отпуске, меня поразило, что лаборатория Худа занимала целое здание. С тех пор он переехал в Сиэтл и основал Институт системной биологии.

Системная биология – новая область науки, которая пытается раскрыть сложность всех молекулярных взаимодействий внутри клетки.

В своем выступлении Худ рассказал, как однажды спросил себя, почему в его лаборатории специалистов по вычислениям больше, чем биологов? Он пришел к выводу, что биология стала информационной наукой и данные, полученные с помощью современных методов, таких как секвенирование генов, превосходят аналитические возможности биологов. Специалисты по вычислениям знают об информации больше, чем биологи. Я не мог и мечтать о лучшем предисловии к моему докладу, посвященному тому, как информация хранится в синапсах нейронов головного мозга.

Сегодня системная биология привлекает много физиков и инженеров в области вычислительных систем для анализа и расшифровки информации, полученной при секвенировании ДНК, и сигналов в клетках, которые контролируются РНК и белками. ДНК в организме человека – цепочка из трех миллиардов пар оснований с информацией, необходимой для поддержания жизни клеток, их репликации и специализации. Некоторые пары оснований являются шаблонами для создания белков, но другие части генома содержат абстрактный код для регулирования генов, которые используются при разработке «инструкции» для строительства тела и мозга. Созданием мозга – возможно, самого сложного проекта во Вселенной – руководят алгоритмы, встроенные в ДНК, которые согласовывают развитие связей между тысячами различных типов нейронов в сотнях частей мозга.

Играть в долгую игру[416]

Появление технологий на рынке, как правило, отстает от теоретических исследований в фундаментальной науке лет на 50. Великие открытия в области теории относительности и квантовой механики, совершенные в первом десятилетии XX века, привели к созданию CD-плееров, GPS и компьютеров во второй половине века. Открытие ДНК и генетического кода в 1950-х годах нашло применение в медицине и сельском хозяйстве и сейчас существенно влияет на экономику. Основные открытия, сделанные в рамках проекта BRAIN и других программ по всему миру, через 50 лет получат практическое применение в таких вещах, которые сегодня считаются научной фантастикой[417]. Можно ожидать, что к 2050 году операционные системы ИИ будут сопоставимы с той, что находится в нашем мозге.

Глава 16. Сознание[418]

Мать Фрэнсиса Крика однажды спросила его, какие научные проблемы он хотел бы исследовать в жизни (рис. 16.1)[419]. Молодой Фрэнсис ответил, что его интересуют только две проблемы: тайна жизни и тайна сознания.

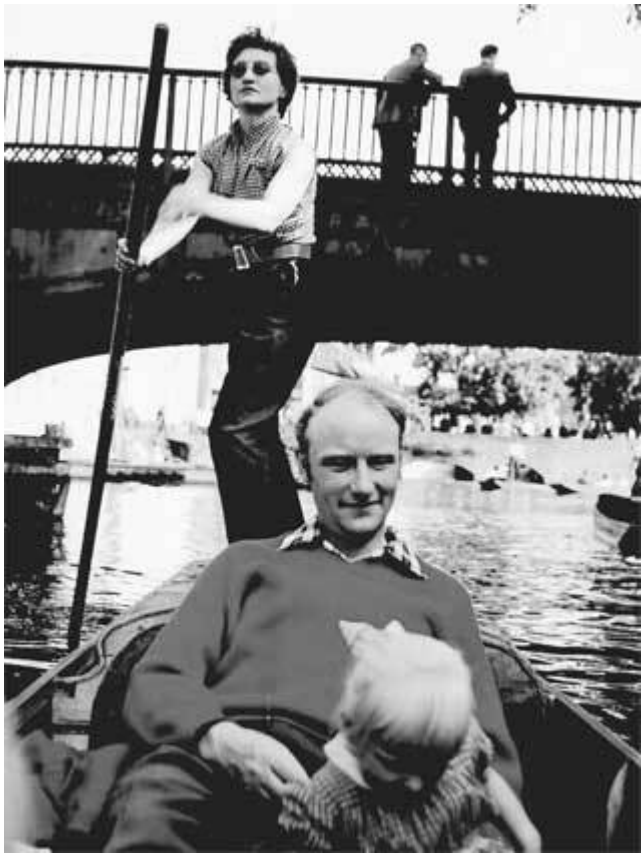


Рис. 16.1. Фрэнсис Крик со своей женой Одиль и дочерью Жаклин плывут на лодке прямо на камеру. Кембридж, около 1957 года. Источник: www.orartswatch.org/kindra-cricks-mad-pursuit/

Он явно понимал, как эти вопросы важны, но, возможно, не осознавал сложность поставленных задач. Вряд ли его мать могла знать, что в 1953 году ее сын и Джим Уотсон откроют структуру ДНК

– нить, которая приведет к ответу на одну из величайших загадок жизни. Однако Фрэнсис Крик не удовлетворился этим достижением.

Когда Крик в 1977 году перешел в Институт Солка, он занялся темой сознания, которая его давно интересовала, и сосредоточился на визуальном восприятии, поскольку о зрительных отделах мозга уже многое было известно и понимание нейронной основы восприятия послужило бы прочным фундаментом для изучения нейронной основы других аспектов сознания. Это также позволило избежать расплывчатости термина «сознание», который используется для описания различных явлений[420].

В 1980-х годах у биологов изучать сознание было не в моде, но Крика это не остановило. Зрительное восприятие было полно странностей и загадок, которые не поддавались пониманию, и он искал для них объяснения в анатомии и физиологических механизмах. Например, он разработал гипотезу «центра внимания»[421]. Ганглиозные клетки проецируются по зрительному нерву в таламус (двусторонние области мозга, которые передают сенсорную информацию в кору), который, в свою очередь, передает импульсы в зрительную кору. Но почему ганглиозные клетки не могут проецироваться прямо в кору? Крик отметил, что от коры к таламусу идет обратная проекция, которая, как луч прожектора, может выделять части изображений для дальнейшей обработки.

Нейронные корреляты сознания

Ближайшим коллегой Крика по исследованию сознания был нейробиолог Кристоф Кох из Калтеха, с которым он опубликовал серию работ, где изучались нейронные корреляты сознания (НКС; структуры мозга и нейронной деятельности, отвечающие за генерацию состояний сознательного восприятия)[422]. В случае зрительного восприятия это означало поиск корреляций между ним и возбуждающими свойствами нейронов в различных частях мозга. Одна из их идей гласила, что мы не знаем, что происходит в

первичной зрительной коре[423] – первой области коры головного мозга, получающей сигнал от сетчатки. Они предположили, что нам, скорее всего, известно только о результатах обработки на высших уровнях иерархии зрительных областей в коре (см. рис. 5.11). В пользу этого варианта говорит изучение бинокулярного соперничества: перед двумя глазами демонстрируют два разных рисунка, например вертикальные полосы для одного глаза и горизонтальные полосы – для другого, и вместо того, чтобы видеть смесь двух изображений, визуальное восприятие резко перескакивает между отдельными картинками каждые несколько секунд. Различные нейроны в первичной зрительной коре реагируют на информацию, поступающую от каждого глаза независимо от того, какая фиксируется сознанием в отдельно взятый момент. Однако на более высоких уровнях визуальной иерархии многие нейроны реагируют только на воспринимаемое изображение. Таким образом, для нейрона недостаточно быть активным, чтобы стать нейронным коррелятом восприятия. По-видимому, вы знаете только то, что представлено в подмножестве активных нейронов, распределенных по иерархии визуальных областей, скоординированно работающих вместе.

Бабушкины клетки

В 2004 году в медицинском центре Калифорнийского университета в Лос-Анджелесе больным эпилепсией, которым проводили мониторинг активности мозга для выявления причин судорог, показали серию фотографий знаменитостей. Электроды, имплантированные в центры памяти мозга пациента, сообщали об импульсах в ответ на фотографии. У одного из таких пациентов единичный нейрон активно реагировал на некоторые снимки Холли Берри и ее имя (рис. 16.2), но не на изображения и имя Билла Клинтона или Джулии Робертс, или портреты и имена других известных людей[424]. Были обнаружены нейроны, которые

реагировали на отдельных знаменитостей, конкретные объекты и здания, такие как Сиднейский оперный театр.

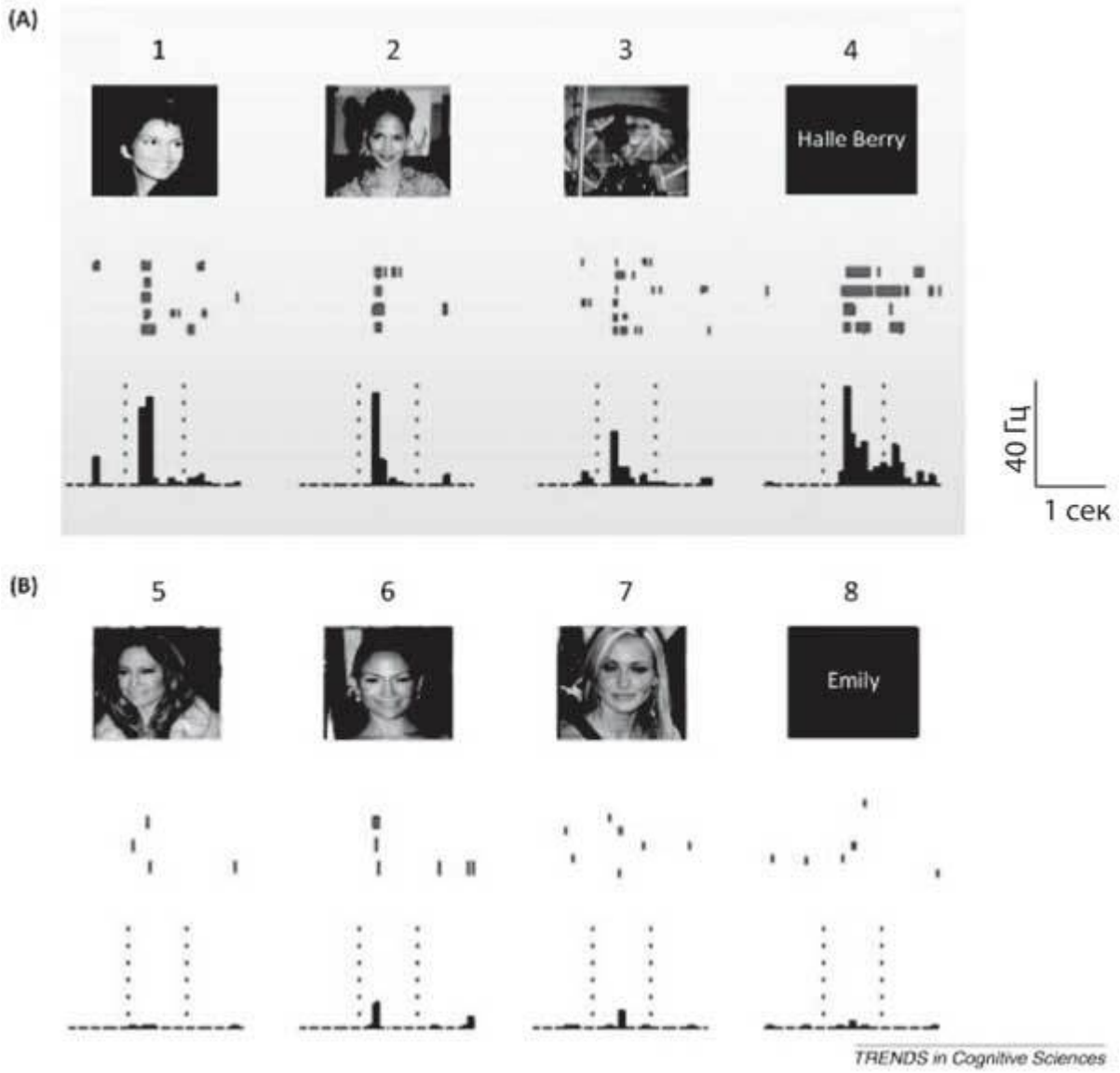


Рис. 16.2. «Клетка Холли Берри». Отклики на фотографии от одного нейрона, записанные из гиппокампа пациента. Под каждой фотографией синим цветом показаны импульсы из шести отдельных проверок, а также гистограмма со средними значениями. (A) фотографии актрисы Холли Берри и ее имя вызвали всплеск импульсов, (B) в отличие от фотографий и имен других актрис. Источник: Friederici A. D., Singer W.: Grounding language processing on

basic neurophysiological principles. Trends Cogn. Sci. 19(6), 329–338. (2015)

Нейроны, найденные командой ученых, которую возглавляли Ицхак Фрид и Кристоф Кох, были предсказаны пятьдесят лет назад, когда впервые стало возможным получать отклик из одиночных нейронов мозга кошек и обезьян. Исследователи полагали, что в иерархии зрительных областей коры головного мозга свойства отклика нейронов становятся все более и более специфичными, чем выше нейрон в иерархии. Возможно, настолько специфическими, что единственный нейрон на вершине иерархии будет реагировать только на изображения одного человека. Это известно как гипотеза «бабушкиной клетки», в честь предполагаемого нейрона в вашем мозгу, который «узнает» вашу бабушку.

Еще более впечатляющими были эксперименты, в которых пациентам показывали два наложенных друг на друга портрета знакомых им людей и просили представить одного человека в ущерб другому, при этом велась запись данных из нейронов, которые предпочитали тот или иной образ. Испытуемые смогли увеличить частоту срабатывания нейрона, который «представлял» выбранное лицо на смешанном изображении, одновременно уменьшая скорость других нейронов, которые предпочитали лицо «конкурента», хотя визуальный стимул не менялся. Затем экспериментаторы замкнули цикл, управляя соотношением двух наложенных изображений в соответствии с частотой срабатывания нейронов, предпочитающих разные изображения, поэтому испытуемые могли контролировать вход – соотношение двух лиц, – представляя то или иное изображение. Это показывает, что распознавание – не пассивный процесс, а зависящий от активного вовлечения памяти и внутреннего контроля внимания.

Несмотря на такое поразительное доказательство, гипотеза «бабушкиной клетки» вряд ли ответит на все вопросы. Согласно ей,

вы узнаете бабушку, когда клетка активна, поэтому она не должна срабатывать ни на какой другой стимул. Во-первых, в тесте использовали всего несколько сотен фотографий, поэтому мы не знаем наверняка, как избирательна «клетка Холли Берри». Во-вторых, вероятность того, что электрод считывал запись от единственного в головном мозге «нейрона Холли Берри», низкая – куда вероятнее, что таких клеток тысячи. Должно быть множество копий нейрона, которые реагируют на другие известные лица, и множество копий для всех, кого вы знаете, и для каждого объекта, который вы можете распознать. Хотя в вашем мозгу миллиарды нейронов, вам будет непросто, если вы попытаетесь представить каждый объект и каждое имя, которое вы знаете, выделенной популяцией нейронов. Наконец, отклик только связан с сенсорным стимулом, он может и не быть его причиной. Не менее важен выходной сигнал нейрона и его влияние на поведение нейронов, стоящих ниже в иерархии.

Записи одновременно из сотен клеток мышей, обезьян и людей приводят к альтернативной теории о том, как нейроны воспринимают сигнал и принимают решения коллективно[425]. На записях, полученных из мозга обезьян, стимулы и сигналы, зависящие от выполняемой задачи, широко распределены по большим популяциям нейронов, каждый из которых настроен на различную комбинацию характеристик стимулов и деталей задачи[426]. К 2025 году можно будет записывать данные с миллионов нейронов и управлять скоростью их активации. Кроме того, разрабатываются новые методы, позволяющие определять типы нейронов и то, как они связаны друг с другом[427]. Это может породить более широкие теории, чем теория «бабушкиной клетки», и привести к более глубокому пониманию того, как активность в популяциях нейронов порождает мысли, эмоции, планы и решения. Конечно, нейроны могут представлять лица и объекты несколькими способами. С появлением новых технологий мы, вполне вероятно, скоро получим ответ.

С 1980-х годов мы знаем, что в обученных сетевых моделях с одним слоем скрытых единиц, а в последнее время и в глубоких сетях, модели активности для каждого входа распределены так, что становятся схожи с разнообразными реакциями в популяциях корковых нейронов[428]. Распределенное представление может использоваться для распознавания нескольких версий одного и того же объекта – и один и тот же набор нейронов может распознавать разные объекты, присваивая различный вес их выходным данным. Когда отдельные скрытые единицы проверяются так же, как нейрофизиологи записывают данные от нейронов зрительной коры, иногда выясняется, что у единичного смоделированного нейрона на вершине иерархии развилось определенное предпочтение для одного из объектов. Тем не менее производительность нейронной сети существенно не меняется, если такую единицу из нее вырезать, потому что оставшиеся нейроны несут дублирующие сигналы, представляющие объект. Устойчивость сетей к повреждениям – главное отличие архитектуры мозга от архитектуры цифровых компьютеров.

Сколько нейронов необходимо, чтобы различать много похожих объектов, например, лица? Благодаря инструментальным исследованиям мы знаем, что несколько областей человеческого мозга реагируют на лица, причем некоторые – с высокой степенью избирательности. Однако в этих областях информация о любых отдельных лицах широко распределена между многими нейронами. Дорис Цао из Калтеха записала данные от нейронов в коре мозга обезьяны, которые избирательно реагируют на лица, и показала, что можно реконструировать лица, комбинируя входные данные от 200 «лицевых» клеток[429] – относительно небольшой части всех реагирующих на лица нейронов[430].

Когда воспринимается время визуального события?

Еще один аспект зрительного восприятия – то, что мозг регистрирует события, такие как вспышки света, с привязкой к определенному времени. Временные задержки нейронов зрительной коры при реакции на зрительный стимул варьируют от 25 до 100 миллисекунд, часто в пределах одной и той же области коры. Но вместе с тем мы можем определить порядок двух вспышек, произошедших с разрывом 40 милисекунд, и порядок двух звуков с разницей во времени менее 10 милисекунд. Еще более парадоксально то, что обработка в самой сетчатке занимает определенное время, которое не фиксировано, но зависит от интенсивности вспышки, так что, хотя есть разница во времени прибытия первого импульса от тусклой и от яркой вспышки, кажется, что они пришли одновременно. И возникает вопрос, почему восприятие кажется единым, что вовсе не очевидно из распределенных в пространстве и времени схем активности по всей коре.

Вопрос одновременности становится еще острее, когда мы проводим кроссмодальные сравнения. Когда вы наблюдаете, как кто-то рубит дерево, вы одновременно видите и слышите, как топор ударяется о дерево, хотя скорость звука намного меньше скорости света. Кроме того, иллюзия одновременности сохраняется и с увеличением расстояния до дерева[431], хотя абсолютная задержка между зрительными и слуховыми сигналами, по мере того как они достигают вашего мозга, может достигать 80 милисекунд, прежде чем иллюзия разрушится и звук перестанет совпадать с ударами топора.

Исследователи, изучающие временные аспекты зрения, обнаружили еще одно явление, называемое эффектом запаздывания вспышки. Его можно наблюдать, когда самолет с мигающими хвостовыми огнями проходит над головой, а свет и

хвост не совпадают. Или изучить в лаборатории с помощью визуального стимула, как показано на рис 16.3. При эффекте запаздывания вспышки кажется, что движущийся объект и вспышка, находящиеся в одном месте, смещены по отношению друг к другу.

Основное объяснение – интуитивно понятное и отчасти подтвержденное данными из записей активности мозга, – мозг предсказывает, где движущееся пятно будет через короткий промежуток времени. Однако чувственные эксперименты показали, что это не может объяснять эффект запаздывания вспышки, потому что восприятие, приписываемое времени вспышки, зависит от событий, которые происходят в течение 80 миллисекунд после вспышки, а не до нее, и которые претендуют на роль основы для прогнозирования[432].

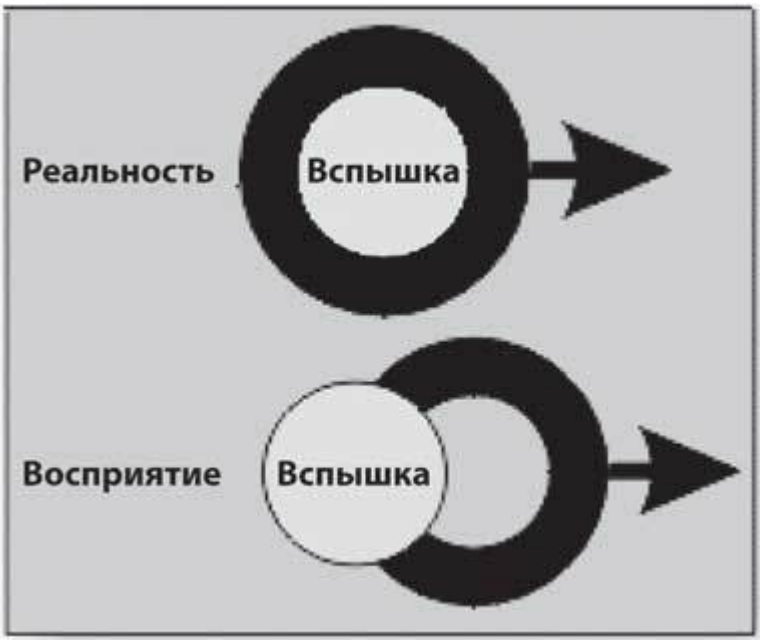


Рис. 16.3. Эффект задержки вспышки. Кольцо движется слева направо (черное, сверху). Когда оно проходит над лампочкой, на миг вспыхивает свет (желтый, вверху). Наблюдатели же сообщают, что все выглядит, как на нижнем рисунке: в момент вспышки объект смещается вправо

Такое толкование эффекта запаздывания означает, что мозг больше работает с уже имеющейся информацией, чем с предсказаниями. То есть мозг постоянно обращается к опыту, чтобы согласовать воспринимаемое настоящее с будущим. Один из примеров того, как наш мозг генерирует правдоподобные интерпретации, основанные на зашумленных и неполных данных, – использование фокусниками эффекта «ловкости рук»[433].

Где в мозгу воспринимается зрительный образ?

Визуализация мозга дает нам общую картину характера мозговой деятельности, когда мы воспринимаем или не воспринимаем что-либо. Используя экспериментальные данные, исследователи разработали очень заманчивую гипотезу: мы осознаем что-то только тогда, когда уровень мозговой активности в лобной части коры, необходимый для планирования и принятия решений, достигает порогового уровня и запускает обратную связь[434]. Идея любопытная, но бездоказательная, поскольку с помощью наблюдений установили не причинно-следственные связи, а только корреляцию. Если НКС ответственны за сознательное состояние, должна быть возможность изменить их и таким образом изменить сознание. Дорис Цао показала, что она может препятствовать распознаванию лиц у обезьян, стимулируя «лицевые» области в зрительной коре[435]. Когда подобный эксперимент проводился на людях, испытуемые сообщали, что лица будто бы расплываются, плавятся[436].

Недавно стали доступны новые методы, такие как оптогенетика, для избирательного манипулирования активностью нейронов, что позволяет проверить причинно-следственные связи НКС. Это может оказаться сложным, если относящиеся к восприятию структуры соответствуют сильно распределенным схемам деятельности, но в принципе такой подход может выявить, как формируется восприятие и другие особенности сознания[437].

Учим смотреть

Визуальный поиск – задача, которая зависит как от обработки сенсорной информации «снизу вверх», так и от управления процессами внимания «сверху вниз» (рис. 16.4 А). Эти два процесса переплетены в мозге, но недавно была разработана новая поисковая задача, чтобы отделить их друг от друга[438]. Участников эксперимента усадили перед пустым экраном и сказали, что их задача – исследовать экран глазами, чтобы найти скрытое местоположение цели, которая издаст звуковой сигнал, когда взгляд зафиксируется на ней. Положение скрытой цели изменялось от раза к разу и было построено на основе гауссовского распределения – колоколообразной кривой с определенной шириной и верхней точкой, которые не были известны участнику, но оставались постоянными в течение сеанса (рис. 16.4 Г).

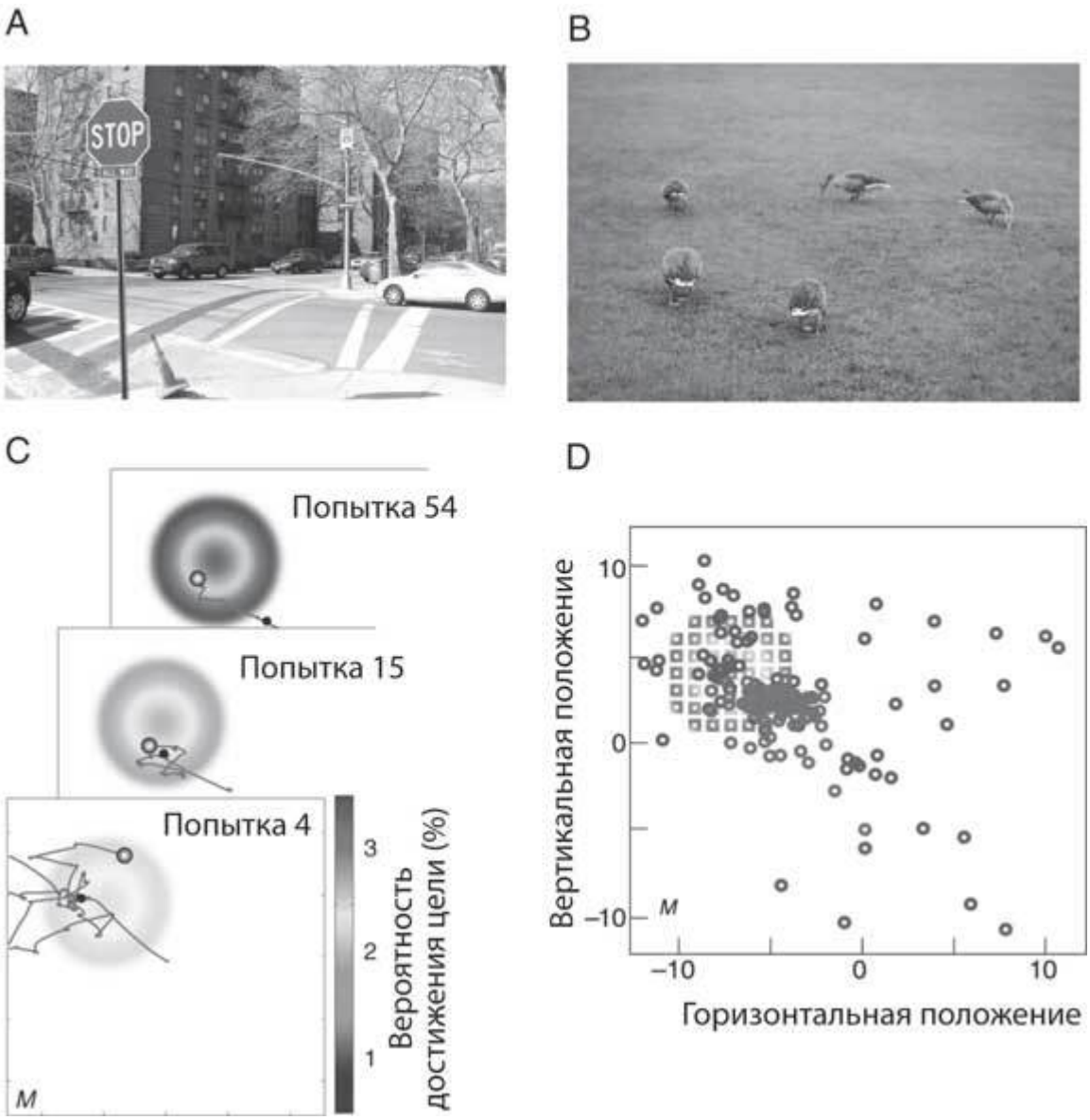


Рис. 16.4. Учимся искать визуальную цель. (А) Опытный пешеход заранее знает, где искать знаки, автомобили и тротуары на улице. (Б) Утки ищут корм на лугу. (В) Изображение на экране накладывается на распределение скрытых целей, изученное в ходе сеанса, а также отмечена траектория взгляда участника М во время трех попыток. Первая фиксация взгляда при каждой попытке отмечена черной точкой. Финальная, за которой последовало вознаграждение, – точкой, окрашенной оттенками серого. (Г) Область, выбранная для фиксации взгляда, сжимается от всего экрана при ранних попытках (серо-голубые круги; первые пять попыток) до области, которая

примерно соответствует положению и размеру целей согласно распределению гауссовых целых чисел (квадраты, затемненные пропорционально вероятности для изображения A) при более поздних попытках (круги; попытки 32–39)

В начале эксперимента участники не располагали предварительными данными для поиска. После того как фиксация была вознаграждена, участники могли использовать обратную связь, чтобы лучше справиться при следующей попытке. В ходе эксперимента участники повышали число удачных попыток, выявляя зону, где стоит ожидать скрытые цели, и используя это в будущих поисках. После десятка попыток зрительная фиксация участников сузилась до области, где с высокой вероятностью находилась цель. Описание результата для всех участников представлено на рис. 16.4 Г. Сначала зона поиска была широкой, но сужалась по мере продолжения сеанса. Удивительно, но многие испытуемые не смогли сформулировать свою стратегию поиска, хотя после нескольких попыток первое движение их взгляда неизменно оказывалось в центре невидимого целевого распределения.

Области мозга, участвующие в этой поисковой задаче, включают зрительную кору и верхнее двухолмие, которое контролирует топографическую карту поля зрения и направляет саккады к визуальным целям, тесно взаимодействуя с другими частями глазодвигательной системы. В обучении также участвуют базальные ганглии – древняя часть мозга позвоночных, которая усваивает последовательность действий через обучение с подкреплением[439]. О разнице между ожидаемым и полученным вознаграждением свидетельствует кратковременное увеличение частоты импульсов дофаминовых нейронов в среднем мозге, что регулирует синаптическую пластичность и влияет на принятие решений на бессознательном уровне, как описано в главе 10.

Переходы

Незадолго до кончины Фрэнсис Крик позвонил мне и пригласил посетить его дом, чтобы обсудить клауструм – загадочный тонкий слой клеток под корой мозга, который получает проекции из многих областей коры и, в свою очередь, проецирует их обратно. Хотя Крик был смертельно болен, он сосредоточился на завершении своей последней работы над гипотезой, гласящей, что клауструм отвечает за единство сознания в силу своего центрального положения. Немногие исследователи работали над этой областью мозга, и в поисках дополнительных данных Крик обзвонил почти каждого из них. Тот раз стал последним, когда я его видел. Фрэнсис Крик умер 28 июля 2004 года, работая над завершением рукописи и поиском истоков сознания[440].

Структуру ДНК открыли в 1953 году, а геном человека был секвенирован 50 лет спустя. Однажды я спросил Крика, думал ли он когда-нибудь в те ранние годы, что геном человека секвенируют при его жизни? Он сказал, что ему никогда не приходило в голову, что это вообще возможно. Как далеко мы продвинемся через 50 лет в решении проблемы сознания? К тому времени у нас будут машины, взаимодействующие с нами так же, как мы взаимодействуем друг с другом – посредством речи, жестов и мимики. Вероятно, легче сформировать сознание, чем полностью его понять.

Я подозреваю, что мы можем добиться прогресса быстрее, сначала поняв бессознательную обработку – все то, что мы принимаем как должное, когда видим, слышим и двигаемся. Мы уже продвинулись в понимании систем мотивации, которые сильно влияют на наши решения, и систем внимания, которые помогают нам искать информацию в мире. При более глубоком понимании механизмов мозга, которые управляют восприятием, принятием решений и планированием, проблема сознания может исчезнуть, как Чеширский кот, оставив только широкую ухмылку.

Глава 17. Природа умнее нас

Лесли Орджел (рис. 17.1) долгие годы был моим коллегой в Институте Солка. Химик, получивший образование в Оксфорде, и один из самых умных ученых, которых я встречал, он работал над происхождением жизни. Беседы с ним во время пятничных обедов с преподавателями всегда были увлекательными.

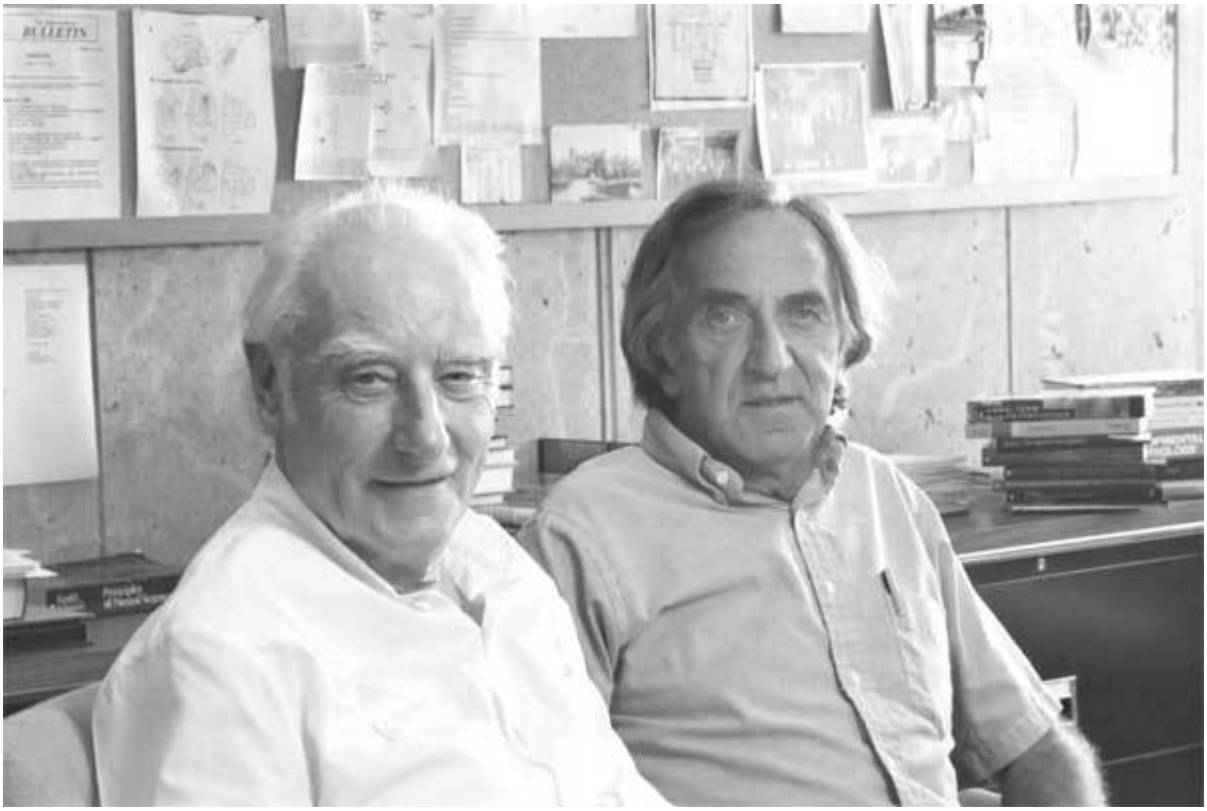


Рис. 17.1. Фрэнсис Крик и Лесли Орджел (справа) в Институте Солка на пути к истокам сознания и истокам жизни соответственно

Жизнь на Земле возникла миллиарды лет назад, когда наша планета настолько отличалась от нынешней, что не поддерживала бы жизнь, какой мы ее знаем: суровые условия, мало кислорода в атмосфере. Бактериям предшествовали археи, но что было до них?

Сегодня ДНК содержится во всех клетках, но что было до ДНК? В 1968 году Лесли Орджел и Фрэнсис Крик предположили, что РНК, которая сейчас является производной от ДНК в клетках, может оказаться ее предшественником, что требовало от РНК способности самореплицироваться. Доказательства того, что это возможно, были найдены в виде рибозимов – ферментов на основе РНК, которые могут катализировать реакции РНК[441], и сегодня большинство исследователей в данной области считают вполне вероятным, что вся жизнь произошла от более раннего «мира РНК»[442]. Но откуда взялась РНК? К сожалению, у нас мало сведений об этом.

Второй закон Орджела

Второй закон Орджела гласит, что природа умнее нас. Общепризнанные истины раз за разом разрушаются удивительными открытиями. Люди подняли глаза и увидели Солнце, вращающееся вокруг Земли, но на самом деле мы вращаемся вокруг Солнца. Теория эволюции поставила людей на их место, что многим до сих пор непросто принять. Через много лет потомки оглянутся на нашу эпоху и скажут, что наши представления об интеллекте были слишком примитивными и сдерживали прогресс ИИ в течение 50 лет.

Наше сознательное восприятие – верхушка айсберга, большая часть которого недоступна для самоанализа. У нас есть такие слова, как внимание и намерения, которые мы используем для описания нашего поведения, но это нечеткие понятия, которые скрывают сложности процессов в мозге, лежащих в их основе. Как следствие, ИИ, основанный на интуитивной житейской психологии, нас разочаровал бы. Мы видим, но не знаем как. Мы думаем, но не знаем как. С точки зрения природы, знания, как мозг действительно работает, не помогут нам выжить. Никуда не деться от второго закона Орджела.

Как мы увидели в главе 2, у нас высокоразвитая зрительная система, но она не делает нас специалистами по зрению[443]. Многие даже не подозревают, что у них в глазу есть центральная ямка, или фовеа, позволяющая видеть очень четко. Но она обеспечивает угол зрения всего в один градус – примерно на ширину большого пальца, если смотреть с расстояния вытянутой руки, так что формально за пределами фовеа мы почти слепы. Однажды я указал на это матери, но она сказала, что не верит мне, потому что куда бы она ни посмотрела, она видит все ясно. Мы получаем иллюзию повсеместного высокого разрешения, потому что мы можем быстро менять положение наших глаз. Знаете ли вы, что в то время, как вы смотрите на объект, ваш взгляд пробегает по нему три раза за секунду? Когда основоположники компьютерного зрения приступили к разработке, их целью было создать из изображений полную модель мира. Цель, которая оказалась труднодостижимой. Но полная и точная модель не нужна для большинства практических задач или даже невозможна, учитывая низкую частоту дискретизации современных видеокамер.

Основываясь на данных психофизики, физиологии и анатомии, мы с Патрицией Черчленд и нейропсихологом Вилейануром Рамачандраном в нашем эссе «Критика чистого зрения» пришли к выводу, что мозг представляет лишь ограниченную часть мира – только то, что необходимо в отдельно взятый момент для выполнения поставленной задачи[444]. Это облегчает и обучение с подкреплением, сужая число возможных сенсорных входов. Кажущаяся модульная организация зрения (его относительная отделенность от других потоков сенсорной обработки) также иллюзия. Визуальная система объединяет информацию из других потоков, включая сигналы от системы вознаграждения, определяющей ценность окружающих объектов, а моторная система активно ищет информацию, перемещая датчики, двигая глазами – а у некоторых и ушами, – чтобы собрать информацию, которая может помочь заработать вознаграждение.

Мозг эволюционировал в ходе длительной адаптации к окружающей среде. Процесс был прогрессивным: природа не могла себе позволить начать с чистого листа, а должна была обходиться модификацией частей, сохраняя жизнеспособность существующих видов. В книге «Эволюционирующий мозг»[445] Джон Оллман иллюстрирует это, рассказывая о посещении котельной старой электростанции в Сан-Диего, где мудреный набор небольших пневматических трубок соседствовал с рядом вакуумных трубок и несколькими поколениями компьютерных систем управления. Поскольку установка была необходима для непрерывной выработки электроэнергии, ее нельзя было отключать и модернизировать с появление новой технологии, поэтому старые системы управления оставались на месте, а новые интегрировали в имеющуюся систему. Так же и с эволюционирующим мозгом: природа не может позволить себе выбросить старую мозговую систему – она функционирует с существующими планами развития, иногда добавляя новый уровень контроля. Дупликация генов – излюбленный способ создавать копии гена, готовые мутировать, чтобы обрести новую функцию. Полная дупликация генов может привести к рождению нового вида.

Дело против А. Н. Хомского

Психологи, изучавшие обучение в 1930-х годах, принадлежали к школе бихевиоризма и рассматривали поведение как взаимосвязь между сенсорными входами и моторными выходами. В центре внимания бихевиоризма стояло ассоциативное обучение, и многие законы обучения смогли открыть при дрессировке животных по различным графикам вознаграждения. Беррес Фредерик Скиннер из Гарвардского университета был ведущим специалистом в данной области и написал ряд популярных книг и статей, объясняющих последствия его открытий для общества[446]. В те годы интерес к массовой прессе был высок.

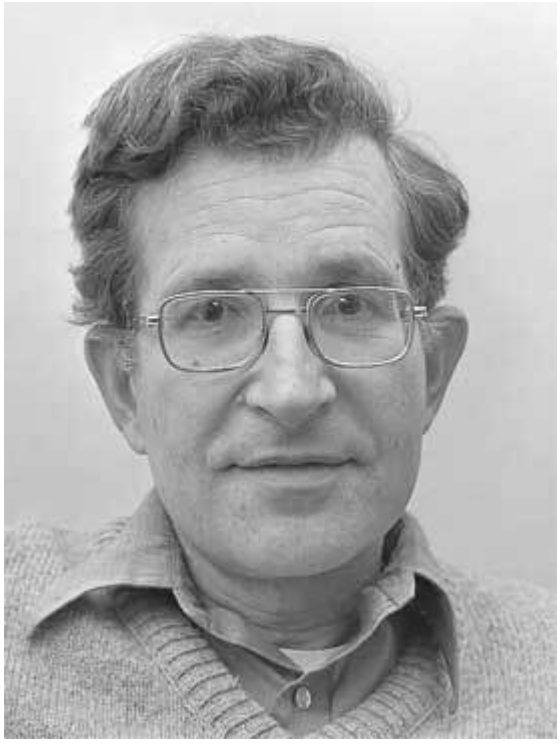


Рис. 17.2. Ноам Хомский в то время, когда он опубликовал эссе «Дело против Б. Ф. Скиннера» в журнале New York Review of Books. Оно оказало сильное влияние на целое поколение психологов, которые воспринимали обработку символов как концептуальную основу познания и не считали развитие мозга и обучение важными

для понимания интеллекта

В 1971 году в журнале New York Review of Books блестящий лингвист Ноам Хомский (рис. 17.2) опубликовал разгромную статью о бихевиоризме в целом и Б. Ф. Скиннере в частности (рис. 17.3)[447]. Вот что он писал о языке: «Но что значит утверждение, будто некоторые фразы на английском языке, которые я никогда не слышал и не произносил, принадлежат моему «репертуару», но ни одно выражение на китайском языке в мой словарный запас не входит (только потому, что у первого выше шансы)? На данном этапе сторонники Б. Ф. Скиннера апеллируют к «подобию» или «обобщению», но всегда без точного определения того, чем именно новое предложение «похоже» на знакомые примеры или «обобщено» из них. Причина их неудачи проста. Насколько известно, соответствующие свойства могут быть выражены только с помощью абстрактных теорий (например грамматики), описывающих предполагаемое внутреннее состояние организма, и такие теории заведомо исключены из «науки» Скиннера. Непосредственным результатом этого является то, что его последователи вынуждены впадать в мистицизм (необъяснимое «сходство» и «обобщение» такого рода, которое не может быть определено), как только дискуссия касается мира фактов. Хотя в примере с языком ситуация, возможно, более ясная, нет оснований полагать, что другие аспекты человеческого поведения попадут в сферу «науки», сдерживаемой заведомыми ограничениями Скиннера».

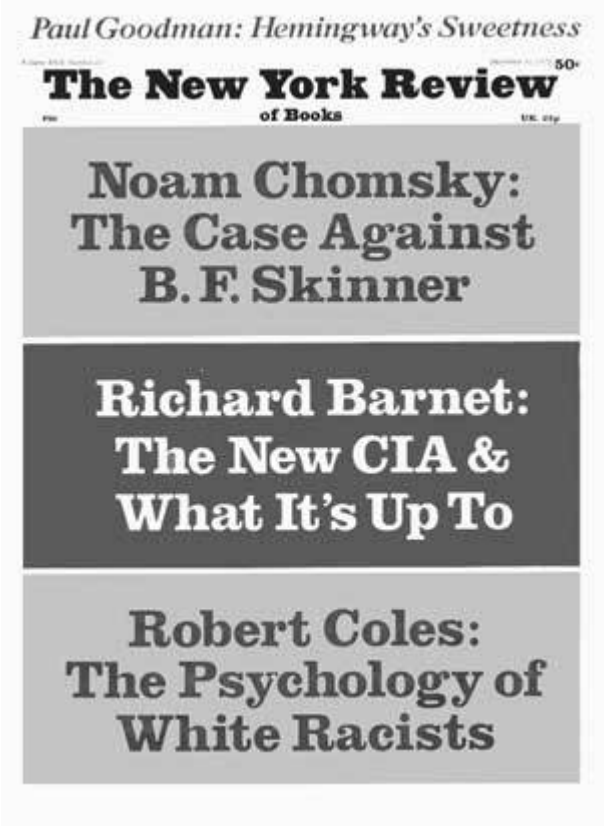


Рис. 17.3. Обложка журнала с разгромным эссе Хомского о Скиннере. Это эссе, опубликованное в 1971 году, было настолько хорошо написано, что заставило целое поколение ученых отказаться от поведенческого обучения как способа объяснить познание. Альтернатива, поддерживаемая лингвистикой, была основана на обработке символов, и именно ее приняли пионеры ИИ. Однако символьный подход к ИИ никогда не достиг производительности когнитивного подхода. Скиннер был на правильном пути, и сегодня самые мощные приложения ИИ основаны на обучении с подкреплением, которое высмеивал Хомский

Сегодня очевидно, что Хомский понимал суть вопроса, но не осознавал силу обучения. Глубокое обучение показало нам, что нейронные сети способны к «обобщению» того рода, который Хомский назвал «мистицизмом», и что их можно научить избирательно распознавать речь на разных языках, переводить с

языка на язык и даже создавать довольно точные подписи к изображениям. Иронично, но машинное обучение решило проблему автоматического разбора предложений, чего так и не удалось «абстрактным теориям» синтаксиса Хомского, несмотря на все усилия компьютерных лингвистов[448]. В сочетании с обучением с подкреплением, которое изучал на животных Скиннер, могут быть решены сложные проблемы, которые зависят от выбора последовательности решений для достижения цели. В этом суть решения проблем и, в конечном счете, основа интеллекта.

Презрительное эссе Хомского вышло далеко за рамки критики Скиннера и бросило вызов обучению как способу понять познание. Это мнение оказало решающее влияние на когнитивную психологию 1970-х годов. Суть его доводов из приведенной выше цитаты сводилась к тому, что он не мог вообразить, что ассоциативное обучение сумеет когда-либо привести к когнитивному поведению, сравнимому по сложности с речью. На мой взгляд, его аргумент был основан на нехватке информации. Если ведущий мировой лингвист говорит, что он не может что-то представить, то это не становится невозможным. Но риторика Хомского, нашедшая отклик в духе того времени, была убедительной. К 1980-м годам подход к познанию с помощью обработки символов стал единственно приемлемым и лег в основу новой области, называемой когнитивной наукой, включившую в себя когнитивную психологию, лингвистику, философию и информатику. Нейробиология была частью когнитивной науки и оставалась в тени до бурного развития когнитивной нейробиологии в 1990-х годах.

Бедность воображения

Хомский неоднократно использовал одни и те же риторические аргументы, особенно в доводах о врожденности языка, основанной на «бедности стимула»[449]. Этот аргумент гласит, что ребенок не слышит достаточно примеров предложений, чтобы научиться

правилам синтаксиса. Но ребенок не компьютер, получающий от мира строку бестелесных символов. Он погружен в мир богатых сенсорных ощущений и познает его с захватывающей дух скоростью[450]. Ребенок получает от мира значимые переживания, связанные со звуками, еще находясь в утробе матери в форме неконтролируемого обучения, и только после того, как заложен этот фундамент, начинается языковой этап: сначала лепет, затем отдельные слова и – гораздо позже – синтаксически правильные последовательности слов. Врожденной является не грамматика, а способность изучать язык на основе опыта и усваивать свойства целых категорий фраз более высокого порядка в богатом когнитивном контексте.

Что Хомский не мог себе представить, так это то, что в сочетании с глубоким изучением окружающей среды и крепко усвоенной способности определять функцию ценности, отточенной на протяжении всей жизни, слабая система обучения, такая как обучение с подкреплением, может привести к когнитивному поведению, включая развитие речи. В 1980-х годах для меня это было совсем не очевидно, но я должен был понять, что если крошечная сеть, такая как NETtalk, может обрабатывать английское произношение, вполне вероятно, что представления слов, выученных сетями, получат естественное сходство с языком. Позиция Хомского основывалась на бедности воображения, но логически вытекала из второго закона Орджела: природа умнее Ноама Хомского. Будьте осторожны, когда эксперт говорит вам, что что-то невозможно, независимо от того, насколько правдоподобны или убедительны его доводы.

Упор на порядок слов и синтаксис, сделанный Хомским, стал доминирующим подходом в лингвистике во второй половине XX века. Но даже модель «мешок слов»[451], которая отбрасывает порядок слов в предложении, замечательно подходит для понимания темы текста (например, спорт или политика), которое можно дополнительно улучшить, если учитывать слова, стоящие

рядом друг с другом. Вывод из глубокого изучения в том, что порядок слов несет некоторую информацию, но семантика, основанная на значении слов и их отношениях с другими словами, важнее. Слова представлены в мозге богатой внутренней структурой. Узнавая больше, как слова семантически представлены в сетях глубокого обучения, мы, возможно, наблюдаем появление новой лингвистики. Если нет причин, по которым природа должна обременять нас знаниями о том, как мы видим, то нет и причин интуитивно понимать, как работает наша речь.

Давайте взглянем, как внутренняя структура слов может выглядеть в сети, обученной на задачах естественного языка. Хотя сеть может быть обучена на одной задаче, способ, которым она представляет входы в сеть, может использоваться для решения других. Хороший пример – сеть, обученная предсказывать следующее слово в предложении. Представление слов в обученной сети имеет внутреннюю структуру, которую можно использовать, чтобы проводить аналогии между парами слов[452]. Например, при проецировании на плоскость векторы, соединяющие страны со столицами, одинаковы. Сеть научилась автоматически организовывать понятия и неявно изучать отношения между ними, не имея никакой сторонней информации о том, что означает столица (рис. 17.4). Это показывает, что семантику стран и столиц можно извлечь из текста, используя неконтролируемое обучение.

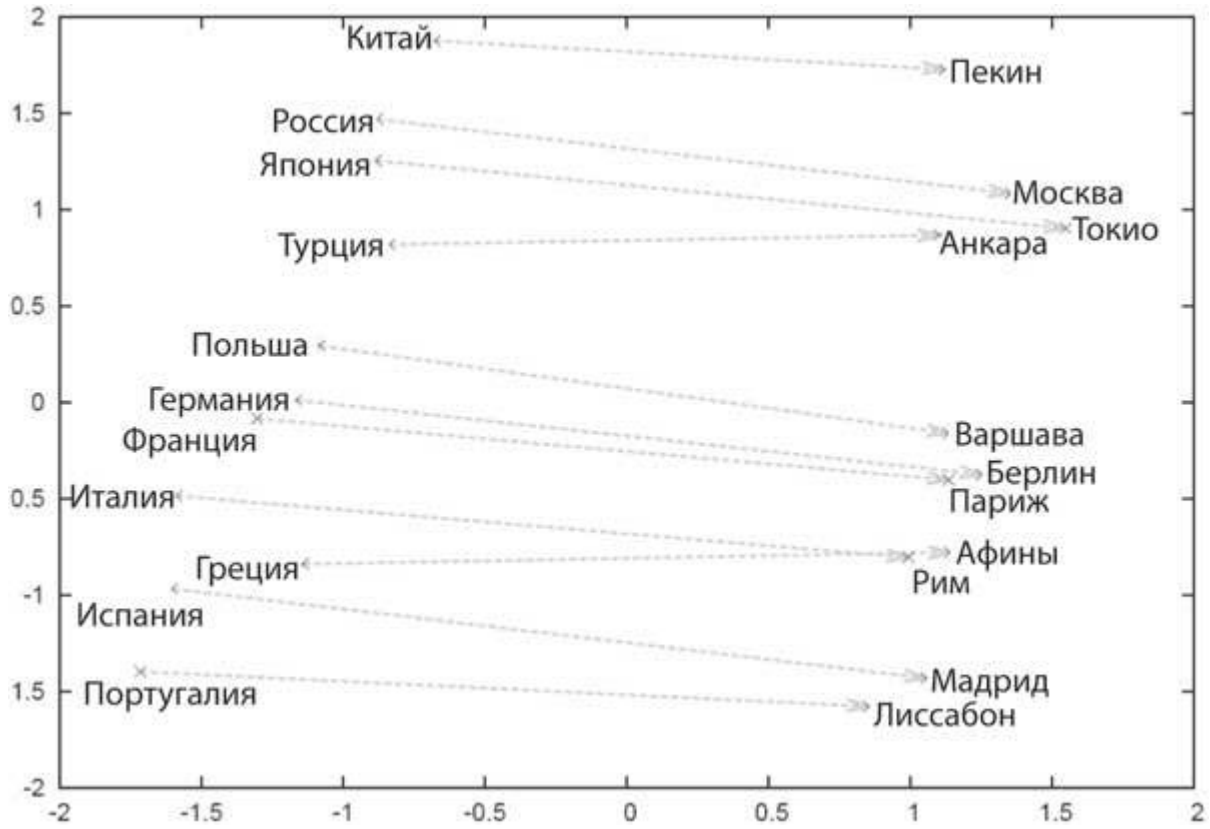


Рис. 17.4. Внутреннее представление слов в сети, обученной предсказывать следующее слово в предложении. Каждое слово – вектор активности в сети, который можно спроецировать вниз на двумерную плоскость, как показано выше. Стрелки соединяют страны со столицами. Поскольку все стрелки соответствуют друг другу и примерно одинаковой длины, пары слов тоже соответствуют. Например, если вы хотите узнать столицу другой страны, вы можете добавить стрелку в вектор страны и получить вектор ее столицы

Однажды я начал лекцию в МТИ с заявления, что «язык слишком важен, чтобы оставить его лингвистам»[453]. Мы не должны останавливаться на описании языка на поведенческом уровне, но должны стремиться понять биологию языка, лежащие в его основе биологические механизмы и то, как развивались языковые способности *Homo sapiens*. Это стало возможным благодаря неинвазивной визуализации мозга и записей, сделанных

непосредственно из мозга пациентов с эпилепсией. Не менее важно изучать человеческий мозг, сравнивая его с мозгом шимпанзе и других человекообразных обезьян, чтобы найти различия, давшие толчок к появлению речи. В масштабах эволюции способность использовать речь произошла моментально на фоне сенсомоторных навыков, которые были приобретены раньше и развивались намного медленнее. Мощные генетические инструменты позволят нам проанализировать развитие мозга и понять, как эволюция в ходе своих экспериментов породила нашу врожденную способность осваивать речь.

Речь можно использовать, чтобы сбивать с толку и манипулировать, внушая доверие и озвучивая аргументы, в основе которых лежит незнание, и печальные последствия выходят далеко за пределы науки. В истории полно демагогов с никуда не ведущими размышлениями, которых в конце концов отвергают, когда обнажается бедность их воображения. К счастью, мозг существуют намного дольше, чем речь, и мы будем лучше функционировать, полагаясь на те части нашего мозга, которые развились до появления речи[454].

Дело против «черных ящиков»

Оглядываясь назад, я понимаю, что бихевиоризм и когнитивная наука, которые в XX веке использовали противоположные подходы к поведению, совершили одну и ту же ошибку, проигнорировав мозг. Бихевиористы не хотели вводить себя в заблуждение самоанализом, поэтому взяли за правило не искать решений в мозге. Они считали, что можно открыть законы поведения, тщательно контролируя входы и выходы «черного ящика». Сторонники функционализма в когнитивистике отвергали бихевиоризм и полагали, что могут разобраться во внутреннем представлении разума, но они тоже отказались от изучения мозга, думая, что детали, как мозг реализует представления, не имеют значения[455]. Внутренние представления, разработанные функционалистами, основывались на интуиции и житейской психологии и были ненадежны. Природа оказалась умнее их.

Выявить внутренние представления и законы поведения «черного ящика» чрезвычайно сложно. Если когда-нибудь мы откроем законы поведения, вполне возможно, что мы сумеем дать им функциональное объяснение, хотя оно, вероятно, будет столь же парадоксальным, как квантовая механика для физиков. Чтобы открыть их, нам понадобится вся доступная помощь мозга. Глубокое обучение – хороший пример прогресса, которого можно достигнуть, обращая внимание на некоторые общие особенности архитектуры мозга и общие принципы его работы. Я не сомневаюсь, что ортодоксальные когнитивисты будут протестовать, но давайте двигаться вперед, а не оглядываться назад. На каждом этапе добавление новой функции из архитектуры мозга повышало функциональность глубоких сетей: иерархия корковых областей, соединение глубокого обучения с обучением с подкреплением, рабочая память в рекуррентных сетях, долговременная память о

фактах и событиях. Это только первые шаги, и в мозге много механизмов, которые мы еще не использовали[456].

Нейробиологи, изучающие восприятие, память и принятие решений, используют задачи, как правило, основанные на экспериментах, в которых животных обучают давать желаемый ответ на стимул. После нескольких месяцев тренировок эти простимулированные реакции становятся больше рефлекторными, чем осознанными, что может выявить механизмы, лежащие в основе нашего привычного поведения, но не нашего когнитивного поведения. Мышление – не рефлекс и может происходить без какого-либо сенсорного стимула. Но традиционный способ проведения экспериментов игнорирует текущую спонтанную активность, которая сохраняется и без внешних раздражителей. Необходимы новые методы для изучения внутренней деятельности, не связанной ни с сенсорными, ни с моторными функциями, включая сознательное мышление и неосознанную обработку информации. Эти методы уже появляются: эксперименты по визуализации мозга выявили состояния покоя, которые спонтанно возникают, когда испытуемого помещают в сканер и просят отдохнуть. Когда нечего делать, разум блуждает, и мысли проявляются как изменяющаяся модель активности мозга, которую мы уже видим, но еще не понимаем.

Визуализация мозга и особенно неинвазивная функциональная магнитная томография открыли новые способы изучения социальных взаимодействий и принятия решений, породив область, названную нейроэкономикой[457]. Люди не рациональны, как предполагала классическая экономика, и нам нужно построить поведенческую экономику, основанную на человеческих суждениях и мотивации, возникающей из сложных внутренних состояний мозга[458]. Как мы увидели в главе 10, дофаминовые нейроны оказывают мощное влияние на мотивацию, представляя ошибку предсказания вознаграждения. Нейровизуализация социальных взаимодействий исследовала человеческую мотивацию способами,

которые невозможны с чисто поведенческими экспериментами. Цель в том, чтобы заменить теорию рациональных решений, основанных на логике, теорией вероятностных решений, основанных на предыдущем опыте.

Дело против М. Л. Минского

История становления нейронных сетей – пример того, как небольшая, но влиятельная группа может помешать развитию конкурирующего направления исследований. В конце книги «Перцептроны» Марвин Минский и Сеймур Пейперт (рис. 17.5) выразили мнение, что алгоритм обучения перцептронов не может быть распространен на многослойные перцептроны:

«Проблема расширения не только техническая. Она также стратегическая. Перцептрон показал себя достойным изучения, несмотря на свои строгие ограничения и даже благодаря им. У него много особенностей, привлекающих внимание: линейность, интересная формула обучения, простая парадигма при схожести с параллельными вычислениями. Нет никаких оснований полагать, что какие-либо из этих достоинств будут присутствовать у многоуровневого варианта. Тем не менее мы считаем важной исследовательской задачей прояснить (или отвергнуть) наше интуитивное суждение, что расширение бесперспективно. Возможно, будет обнаружена какая-то мощная теорема сходимости или иная глубокая причина, по которой нельзя создать интересную «формулу обучения» для многослойной машины».

Рис. 17.5. Марвин Минский и Сеймур Пейперт примерно в то время, когда они писали «Перцептроны». Их книга была отличным математическим анализом простых сетей, но оказала сдерживающее воздействие на поколение исследователей, которые применяли подходы к ИИ, основанные на обучении в многослойных сетях

Действительно, бесперспективно. Негативное мнение в замечательной книге пагубно сказалось на развитии обучения в нейронных сетях и отодвинуло исследования на целое поколение. Лично я извлек выгоду из этой задержки, потому что она сделала мою карьеру возможной, хотя и непростой. Но Минский, вероятно, не был таким могущественным, как мы себе представляли. У меня был шанс «заглянуть за кулисы» на закате его карьеры.

В 2006 году меня пригласили в Дартмутский колледж на конференцию AI@50, посвященную годовщине известного летнего исследовательского проекта по ИИ, проведенного в 1956 году. Присутствовали пять из десяти участников конференции 1956 года: Джон Маккарти из Стэнфорда, Марвин Минский из МТИ, Тренчард Мор из IBM, Рэй Соломонофф, прибывший из Лондона, и Оливер Селфридж из МТИ. Это была увлекательная встреча как в научном, так и в социальном плане.

Такео Канадэ из Университета Карнеги – Меллона выступил с докладом «Зрительное восприятие ИИ: прогресс и отсутствие прогресса»[459]. В 1960-х компьютерная память была крошечной по сегодняшним меркам и могла хранить в памяти только одно

изображение за раз. В своей докторской диссертации в 1974 году Такео показал, что он может найти танк на одном изображении, но пришел к выводу, что это слишком сложно сделать на других, где танк иначе расположен или освещен. К тому времени когда его первые ученики получили научные степени, они могли распознавать танки в более общих условиях, потому что компьютеры стали мощнее. Сегодня его ученики могут распознавать танки на любом изображении. Разница в том, что сегодня у нас есть доступ к миллионам изображений под разными углами и с разным освещением, а компьютеры мощнее в миллионы раз.

В докладе «Разум и тела»[460] Род Брукс отталкивался от своего опыта создания роботов, умеющих ползать и передвигаться зигзагами. У деревьев нет мозга, потому что они не двигаются. Разум эволюционировал в мозге, чтобы контролировать движения, а тела эволюционировали, чтобы взаимодействовать с миром через разум. Брукс отошел от традиционных контроллеров, применяемых робототехниками, и использовал поведение, а не вычисления как модель при проектировании роботов. По мере того как мы узнаем больше о создании роботов, становится очевидно, что тело – часть разума.

Евгений Чарняк из Университета Брауна выступил с докладом «Почему обработка естественного языка стала статистической обработкой естественного языка»[461]. Основная роль грамматики – помечать части речи в предложении. Это то, чему людей можно научить лучше, чем программу. Компьютерная лингвистика первоначально пыталась применить генеративную грамматику, впервые предложенную Хомским в 1980-х годах, но результаты оказалась разочаровывающими. В конечном итоге пришлось привлечь студентов из Университета Брауна, чтобы они вручную пометили части речи в тысячах статей из Wall Street Journal, а затем применить статистические методы, чтобы определить, какой частью речи с большей вероятностью будет то или иное слово, если оно стоит рядом с другими конкретными словами. Необходимо много

примеров, потому что у большинства слов несколько значений, и каждое слово может оказаться в массе контекстов. Автоматическое помечание частей речи в предложениях – теперь решаемая задача, основанная на машинном обучении.

У этих историй успеха схожий путь: в прошлом компьютеры были медленными и позволяли исследовать только игрушечные модели с малочисленными параметрами, но эти игрушечные модели плохо масштабировались на данные из реального мира. Когда компьютеры стали быстрее, а данных – больше, появилась возможность собирать более сложные статистические модели и находить больше признаков и связей между ними. Глубокое обучение автоматизирует этот процесс. Вместо экспертов, вручную ищущих признаки для каждого приложения, глубокое обучение может само извлечь их из очень больших наборов данных.

Это снижает трудозатраты на вычисления, и по мере того, как вычисления продолжат дешеветь, все больше задач, которые научные работники кропотливо решали вручную, будут выполнятся с помощью мощных компьютеров. В конце встречи Марвин Минский подвел итоги. Он начал с того, что сказал, как был разочарован выступлениями и тем, куда движется вся область данной науки. Он пояснил это так: «Вы не работаете над проблемой общего интеллекта. Вы просто работаете над приложениями». Конференция знаменовала прогресс, которого мы достигли, и упрек был неприятен. Я читал доклад о достижениях в обучении с подкреплением и впечатляющих результатах TD-Gammon в обучении сетей игре в нарды на чемпионском уровне, которые, как я думал, впечатлят Минского. Но он сбросил их со счетов как простую игру.

Что Минский имел в виду под общим интеллектом? В своей книге «Общество разума» он исходит из того, что общий интеллект возникает из взаимодействия между более простыми программными агентами. Минский как-то сказал, что самым большим источником идей для его теории стала попытка создать

машину, которая использует роботизированную руку, видеокамеру и компьютер, чтобы строить домики из детских кубиков (см. рис. 2.1). Это подозрительно похоже на приложение. Конкретное приложение заставляет сосредоточиться и добраться до сути проблемы в тех случаях, когда не годится абстрактное теоретизирование. Успехи, о которых сообщили участники конференции в Дартмуте, стали результатом глубокого понимания реальных проблем, которое прокладывает путь к более общему теоретическому пониманию. Возможно, когда-нибудь из этих ограниченных успехов в сфере ИИ появится лучшая теория общего интеллекта.

Наш мозг не варится в своем котелке, побулькивая абстрактными мыслями. Мозг тесно связан со всеми частями тела, которые в свою очередь тесно связаны с миром через органы чувств и моторные реакции. Поэтому биологический интеллект телесен. Еще более важно то, что мозг развивается в течение длительного времени, взаимодействуя с окружающим миром. Обучение – процесс, который совпадает с процессом развития и продолжается после достижения зрелости, особенно у людей. Поэтому обучение занимает центральное место в развитии общего интеллекта. Интересно, что одна из самых сложных нерешенных проблем в ИИ – здравый смысл, который совершенно отсутствует у детей и медленно проявляется у большинства людей после продолжительного общения с миром. Эмоции и эмпатия, которые в ИИ часто игнорируются, также важная часть интеллекта[462]. Эмоции – глобальные сигналы для подготовки мозга к действиям, которые не могут быть решены локальными состояниями мозга.

В завершающий день конференции AI@50 состоялся банкет, на котором пять участников Дартмутского летнего исследовательского проекта по ИИ 1956 года сидели за высоким столом с видом на зал. В конце ужина они сделали краткие замечания о встрече и будущем ИИ. Когда после их выступления разрешили задать вопросы, я спросил Минского: «В сообществе исследователей нейронных сетей есть теория, что вы дьявол, который ответственен за упадок

нейронных сетей в 1970-х годах. Вы дьявол?» Минский начал тираду о том, как мы не понимаем математических ограничений наших сетей. Я перебил его: «Доктор Минский, я задал вам вопрос, на который нужно ответить «да» или «нет». Так вы дьявол или нет?» После недолгих колебаний он выкрикнул: «Да, я дьявол!»

Минский был не единственным, из-за кого в 1970-х годах замедлились исследования нейронных сетей. Фрэнк Розенблатт создал аналоговый компьютер, разработанный для имитации перцептрона, потому что цифровые компьютеры ужасно медленно воспроизводили сетевые модели, которым требовался большой объем вычислений. К 1980-м годам мощность компьютеров значительно возросла, и мы смогли исследовать алгоритмы обучения с помощью моделирования небольших сетей. Но только в 2010-х годах вычислительных мощностей стало достаточно, чтобы масштабировать сети до размеров, способных решать практические задачи.

Ирония моего диалога с Минским в том, что его диссертация по математике, которую он защитил в Принстонском университете в 1954 году, содержала теоретические и экспериментальные исследования вычислений в нейронных сетях. Он даже сконструировал небольшие сети из электронных компонентов, чтобы посмотреть, как они себя ведут. Когда я был аспирантом-физиком в Принстонском университете, я слышал байку, что на математическом факультете не было никого достаточно квалифицированного для оценки его диссертации на тему «Теория нейроаналоговых систем подкрепления и ее применение к проблеме модели мозга»[463], поэтому они отправили ее математикам в Институт перспективных исследований в Принстоне, члены которого, по слухам, говорили с Богом. Ответ пришел такой: «Если сегодня это не математика, то когда-нибудь ею будет», что оказалось достаточно для присуждения Минскому докторской степени. Нейронные сети действительно стали новым классом математических функций, которые стимулируют

исследования и становятся новой отраслью математики. Марвин Минский опередил свое время.

Шаги

До своей смерти в 2016 году Марвин Минский твердо верил, что нейронные сети – тупик на пути к достижению общего ИИ. Во вдумчивом эссе о своей дружбе с Минским[464] Стивен Вольфрам писал: «Хотя я не думаю, что кто-то мог знать об этом тогда. Теперь мы знаем, что нейронные сети, которые Марвин исследовал еще в 1951 году, на самом деле двигались по пути, который в конечном итоге приведет к впечатляющим возможностям ИИ, на которые надеялся Минский. Жаль, что это заняло столько времени, а Марвин едва успел их увидеть».

Вскоре после смерти Минского следующий шаг к общему искусственному интеллекту, основанному на глубоком обучении, сделали исследователи из проекта DeepMind, добавив динамическую внешнюю память[465]. В глубокой рекуррентной нейросети схемы деятельности могут храниться только временно, что затрудняет моделирование рассуждений и умозаключений. Добавляя в сеть стабильную память, которую можно записывать и считывать так же свободно, как и память цифрового компьютера, они продемонстрировали сеть, натренированную обучением с подкреплением и умеющую отвечать на вопросы, требующие рассуждений. Например, одна сеть рассуждала о путях в лондонском метро, а другая отвечала на вопросы о родственных отношениях в генеалогическом дереве. Сеть с динамической памятью также смогла справиться с задачей переноса объектов в Blocks World, которая занимала сотрудников Лаборатории искусственного интеллекта МТИ в 1960-х годах (см. рис. 2.1). Это возвращает нас к тому, с чего мы начали в главе 2.

Фрэнсис Крик умер в 2004 году, а Лесли Орджел – в 2007 году. Закончилась целая эпоха в Институте Солка. Этих научных гигантов больше нет с нами, и вперед продвигается новое поколение. Я проработал в Институте Солка 30 лет, половину его существования.

Он начинался в 1960 году в тесной, почти семейной атмосфере, когда преподаватели и сотрудники плыли в маленькой лодке и все знали друг друга. Сегодня в институте Солка работает тысяча человек, но, как ни странно, он все еще хранит семейную атмосферу. Это связано с тем, что в каждом учреждении есть культура, которая, как правило, переживает людей, вошедших в нее, как топор, у которого сначала заменили рукоять, а потом лезвие.

Мы – один вид в большой цепи бытия, начавшейся с бактерий. Это чудо, что мы подошли к грани понимания мозга и того, как он эволюционировал, что навсегда изменит наше представление о себе.

Глава 18. Глубокий интеллект

Франциск Крик в раю

Сидней Бреннер (рис. 18.1) получил образование в ЮАР и участвовал в становлении молекулярной генетики в Кембриджском университете. Он работал в Лаборатории молекулярной биологии вместе с Фрэнсисом Криком. Что бы вы сделали для своего следующего проекта после открытия структуры ДНК и генетического кода? Крик решил сосредоточиться на человеческом мозге, а Сидней приступил к новому модельному организму[466] – *C. elegans* – круглому червю, живущему в почве, длина которого всего один миллиметр и у которого лишь 302 нейрона. Этот червь, чью каждую клетку отслеживали в течение долгого времени, послужил отправной точкой для многих шагов к пониманию того, как существо развивается из эмбриона, за что Бреннер с коллегами и получил Нобелевскую премию в 2002 году.

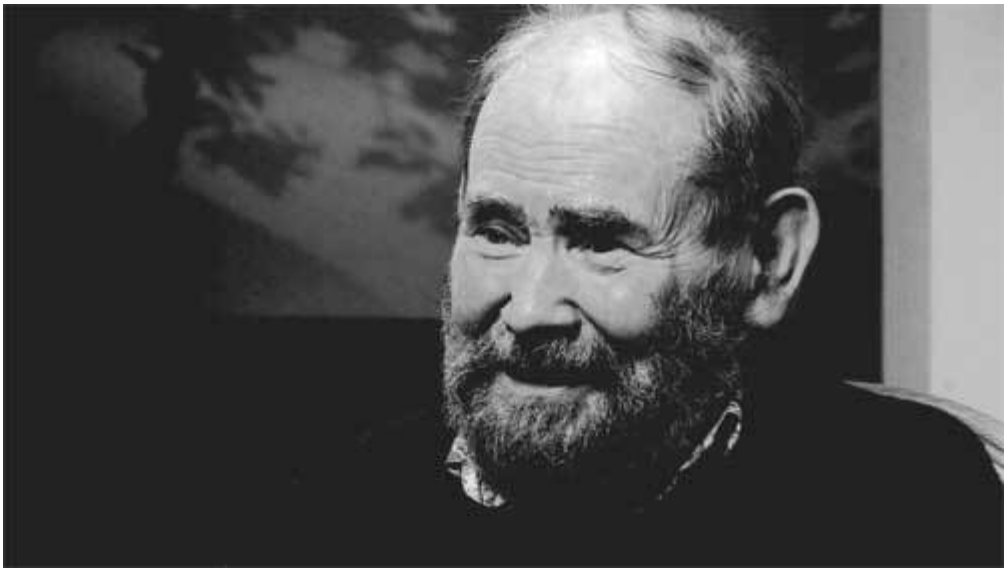


Рис. 18.1. Сидней Бреннер – легендарная личность в биологии. Он работал над генетическим кодом – способом, которым пары оснований в ДНК зашифрованы в белках, и получил Нобелевскую

премию за свою новаторскую работу над новой моделью организма Это фото из интервью журналу The Science Network: thesciencenetwork.org/programs/the-science-studio/sydney-brennerpart-1

Бреннер известен своим остроумием. В своей нобелевской речи он похвалил червя: «Название моей лекции – «Дар науке от природы». Это не лекция о том, как один научный журнал отдает дань уважения другому, а о том, как великое разнообразие живого мира может одновременно вдохновлять и помогать совершать открытия в биологических исследованиях»[467]. Бреннер будто присутствовал в момент Творения.

В 2009 году Сидней Бреннер провел в Институте Солка три лекции на тему «Чтение генома человека»[468] без опоры на какие-либо материалы и слайды – просто высший пилотаж. Он отметил, что на самом деле ни один человек не прочитал весь человеческий геном, пара оснований за парой оснований, – только компьютеры. Сидней решил сам сделать это и обнаружил многочисленные интересные сходства между участками ДНК в разных генах и у разных видов. Нам нужно создать машину для считывания ДНК столь же умную, как Сидней Бреннер.

Бреннер перекати-поле: у него экспериментальный проект в Сингапуре, он был президентом-основателем Окинавского института науки и технологий, он старший научный сотрудник Исследовательского городка Джанелия медицинского института Говарда Хьюза и в Центре теоретической и вычислительной биологии Крика – Джейкобса. Бреннер принял в Лабораторию молекулярной биологии Дэвида Марра для работы над вычислительной техникой после того, как Марр защитил докторскую диссертацию, а позже устроил Марра в Лабораторию искусственного интеллекта МТИ через своего друга Сеймура Пейперта. Связи между

молекулярной генетикой и нейрофизиологией были глубоки, и Бреннер находился в центре обеих областей.

Во время одного из его визитов в Ла-Хойя я рассказал Бреннеру анекдот, который слышал много лет назад, когда был научным сотрудником Гарвардской медицинской школы. По сюжету Фрэнсис Крик умер и попал на небеса. Святой Петр удивился, увидев ревностного атеиста Крика, но Крик был там, чтобы задать вопрос. Его направили в деревянную хижину, где Бог делал Божью работу, а все вокруг было усыпано всевозможными колесами и шестеренками – остатками неудачных экспериментов. Бог в кожаном фартуке сидел за верстаком и возился с очередным существом. «О, Фрэнсис, – воскликнул Бог, – рад тебя видеть! Что я могу для тебя сделать?» Фрэнсис Крик сказал: «Всю свою жизнь я хотел знать ответ на один вопрос: зачем мухам имагинальные диски?»[469] Бог ответил: «Фрэнсис, как неожиданно! Раньше никто никогда не задавал мне этот вопрос. Я вставлял имагинальные диски в организм мух в течение сотен миллионов лет и не получил ни единой жалобы».

Бреннер молчал, и я засомневался, была ли байка про его близкого друга хорошей идеей. «Терри, – сказал Бреннер, – Я могу рассказать о том моменте, когда мне в голову пришла эта история. Мы с Фрэнсисом сидели в нашем офисе, он читал книгу по биологии развития и вдруг всплеснул руками с возгласом: «Бог его знает, почему у мух есть имагинальные диски!»

Я был ошеломлен. Как часто удается услышать происхождение истории, которую знаешь десятилетиями и рассказывал бесчисленное количество раз? Я попросил Бреннера поделиться первоначальной версией. Он сказал, что она называлась «Франциск Крик в раю» – его история имеет ту же структуру, что и моя, но отличилась в мелочах[470]. Точно так же в процессе эволюции мутируют многие детали, но не сама суть.

Я посетил Бреннера в Сингапуре в январе 2017-го, чтобы отпраздновать его 90-летие[471]. Он больше не путешествовал из-за

проблем со здоровьем и был прикован к инвалидному креслу, но был таким оживленным, каким я его никогда не видел. Феодосий Добжанский однажды сказал, что в биологии ничто не имеет смысла, кроме как в свете эволюции[472]. 21 февраля 2017 года Сидней прочитал захватывающую лекцию об эволюции бактерий в рамках серии «10 из 10: Хроника эволюции» в Наньянском технологическом университете в Сингапуре[473]. Свое выступление об эволюции мозга, прошедшее 14 июля 2017 в рамках упомянутой серии, я начал с вариации на ту же тему: «Ничто в биологии не имеет смысла, кроме как в свете ДНК»[474].

Эволюция интеллекта

Интеллект у различных видов развивался ради решения проблем, с которыми они сталкивались, пытаясь выжить в своих экологических нишах. У животных, эволюционировавших в океане, были иные проблемы, чем у тех, кто жил на суше. Зрение позволяет нам воспринимать окружающий мир, и мы разработали визуальный интеллект для интерпретации визуальных сигналов. Этологи, изучающие поведение животных в их естественной среде, обнаружили способности и навыки, нехарактерные для людей, такие как эхолокация у летучих мышей. Летучие мыши активно посылают звуковые сигналы, чтобы исследовать окружающую их среду и анализировать возвращающееся эхо: это дает им внутреннее представление о внешнем мире так же четко, как нам – зрение. Они обладают слуховым интеллектом, который сортирует эти сигналы, отыскивая летающих насекомых для охоты и объекты, которых следует избегать. Летучие должны смотреть свысока на глупых людей, которые не видят в темноте и не могут летать.

Томас Нагель, философ из Нью-Йоркского университета, написал в 1974 году статью «Каково быть летучей мышью?», в которой пришел к выводу, что мы не можем себе представить, как выглядит мир летучих мышей без непосредственного опыта с эхолокацией[475]. Что, впрочем, не помешало нам изобрести радар и сонар – технологии, которые позволяют людям активно исследовать мир, да и слепые становятся чувствительнее к отраженному звуку. Мы можем не знать, каково это – быть летучей мышью, но мы можем создать интеллект как у летучей мыши, который помогает беспилотным автомобилям ориентироваться с помощью радара и лидара.

Люди – лучшие ученики в мире. Мы можем быстро изучать широкий диапазон тем, больше запоминать и передавать знания через бóльшее число поколений, чем любой другой вид. Мы создали технологию обучения, чтобы увеличить объем того, чему мы можем

научиться в жизни. Сейчас дети и подростки проводят годы своего взросления, сидя в классах и изучая явления, которые никогда не видели. Они учатся доказывать геометрические теоремы.

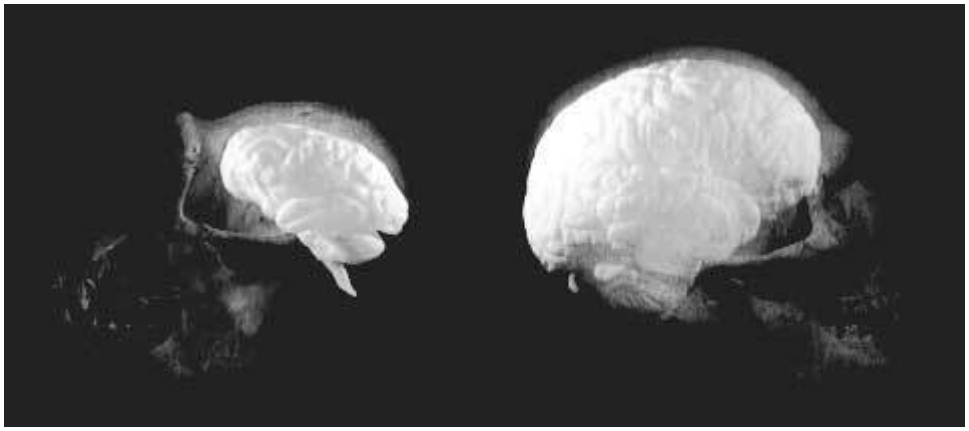


Рис. 18.2. Сравнение мозга шимпанзе с человеческим мозгом. Человеческий мозг гораздо крупнее, у него намного больше извилин и больше площадь коры. [John Allman (1999). Evolving Brains, New York: Scientific American Library.]

Чтение – относительно недавнее изобретение человека, освоение которого занимает много лет. Книги и чтение позволяют передать следующему поколению больше накопленных знаний, чем в устной традиции. Именно чтение и обучение, а не разговорный язык, сделали возможной современную цивилизацию.

Откуда мы вообще взялись?

Каковы эволюционное происхождение человека? В 1998 году я был одним из основателей исследовательской группы в Ла-Хойя, изучавшей происхождение человека. Первоначально небольшая группа проводила регулярные встречи, обсуждая многочисленные источники доказательств, которые могли бы помочь ответить на этот вопрос, начиная с палеонтологии, геофизики, антропологии, биохимии и генетики и заканчивая сравнительной нейробиологией. Линия, которая в конечном итоге породила род *Homo*, отделилась от

линии, ведущей к шимпанзе, около шести миллионов лет назад. Шимпанзе – очень умный вид, но интеллект шимпанзе значительно отличается от нашего. Попытки научить шимпанзе основам языка никогда не выходили за пределы нескольких сотен знаков, которые те используют для выражения простых потребностей. Однако это нечестный показатель их интеллекта. Как бы мы справились, если бы нам пришлось выживать в отряде шимпанзе? Все ли виды эгоцентричны так, как наш?

Группа в Ла-Хойя постепенно привлекала международных ученых и в 2008 году стала Центром академических исследований и обучения антропогенезу (Center for Academic Research and Training in Anthropogeny; CARTA)[476] Калифорнийского университета в Сан-Диего и Института Солка. В нем продолжают изучать, откуда появились мы, люди, и как мы сюда попали, а также обучать новое поколение тех, кто задумывается над этими извечными вопросами[477]. Данные вопросы требуют знаний из всех областей науки, точно так же, как NIPS собрала все области науки и техники, чтобы понять нейронные вычисления.

Одно из мест, где можно найти различия между людьми и шимпанзе, – в наших ДНК. С некоторых пор мы знали, что только 1,4 процента из трех миллиардов пар оснований ДНК у нас отличаются от таковых у шимпанзе. Когда геном шимпанзе был впервые секвенирован, считалось, что мы сможем прочитать книгу жизни и узнать, что отличает нас от шимпанзе. К сожалению, книга жизни написана на языке ДНК, 90 процентов которого мы еще не научились разбирать[478]. Наш мозг также удивительно похож на мозг шимпанзе. Нейроанатомы определили одинаковые области мозга у обоих видов (рис. 18.2). Большинство различий находятся на молекулярном уровне и едва заметны по сравнению с существенными различиями в нашем поведении. И снова природа оказалась умнее нас.

Логика жизни

Однажды я спросил Лесли Орджела, а каков первый закон Орджела? В нем, как ответил Лесли, говорится, что для всех основных реакций в клетках должен появиться фермент-катализатор. Это не только ускоряет реакцию, но и дает возможность управлять ею через взаимодействие с другими молекулами, так что клетка может быть как более эффективной, так и более адаптируемой. Природа начинается с продуманного хода реакции, а затем постепенно уточняет его, добавляя скорость и резервные пути. Филигранное наполнение клетки рано или поздно будет доведено до совершенства, но ничего не станет работать, если не выполняются четкие базовые требования – поддержание и репликация ДНК как ключевого звена всей цепи.

Одноклеточные приспособились к различным условиям и заняли множество ниш. Например, бактерии (рис. 18.3) адаптировались к экстремальным условиям от гидротермальных источников в океане до ледяных покровов Антарктиды и вашего кишечника, где обитают тысячи их видов. Бактерии, такие как кишечная палочка, разработали алгоритм, позволяющий им подплывать к источникам пищи, используя градиент концентрации. Поскольку, чтобы воспринять градиент непосредственно, бактерии слишком малы (несколько микрометров в поперечнике), они применяют хемотаксис, – периодически совершают кувырок и двигаются в случайном направлении[479]. Выглядит непродуктивно, но, увеличивая время движения при более высокой концентрации, они могут надежно подниматься вверх по градиенту. Это примитивная форма интеллекта. Более сложные формы интеллекта встречаются у многоклеточных животных.

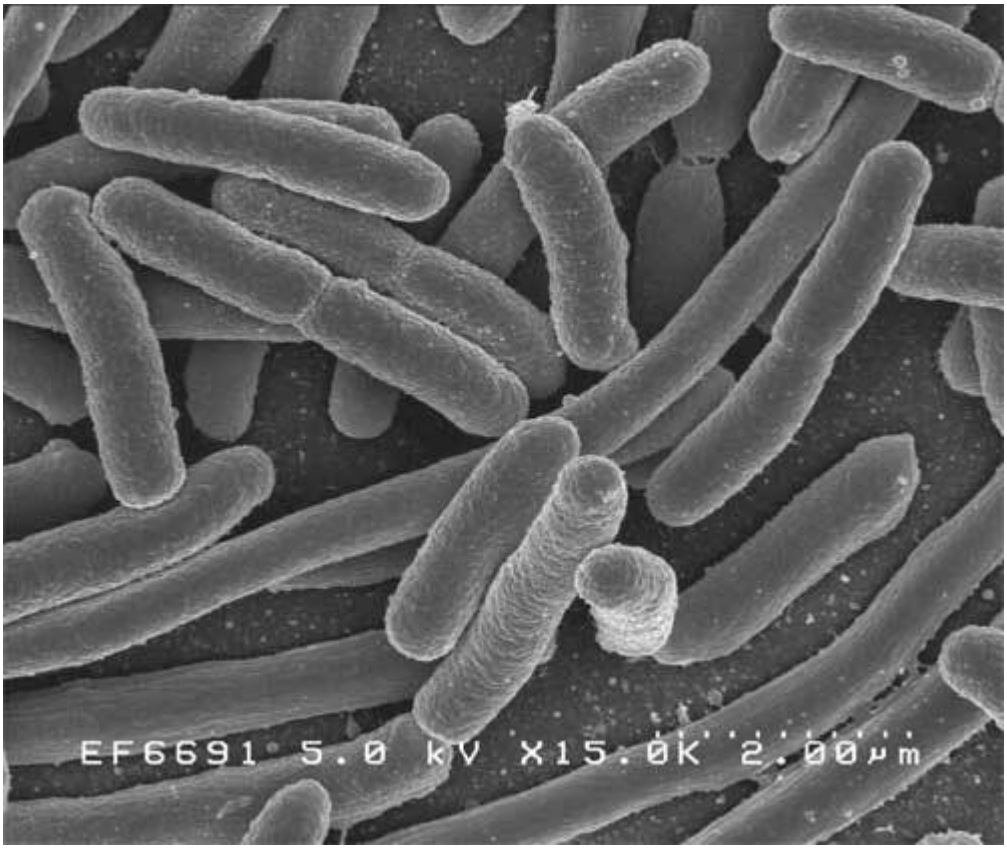


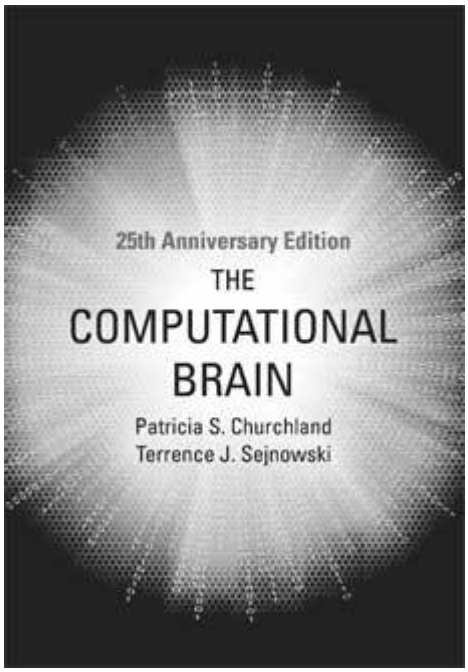
Рис. 18.3. Сканирующая электронная микрофотография кишечной палочки. Бактерии – самая разнообразная, жизнестойкая и успешная форма жизни на Земле. Изучая их, мы можем многое узнать об автономном ИИ

Мы видели, что алгоритм обучения с временной разницей, лежащий в основе обучения с подкреплением, может привести к очень сложному поведению. У людей это значительно усиливает глубокое обучение в коре головного мозга. В природе есть целый спектр интеллектуального поведения, которое могут перенять искусственные системы. Новая область науки, охватывающая информатику и биологию, направлена на выявление биологических алгоритмов с использованием математического анализа сетей[480]. Это край клина[481], который в конечном итоге может объяснить вложенные уровни сложности в биологических системах в пространственных и временных масштабах: генные сети, метаболические сети, иммунные сети, нейронные сети и социальные сети – сети на всех уровнях.

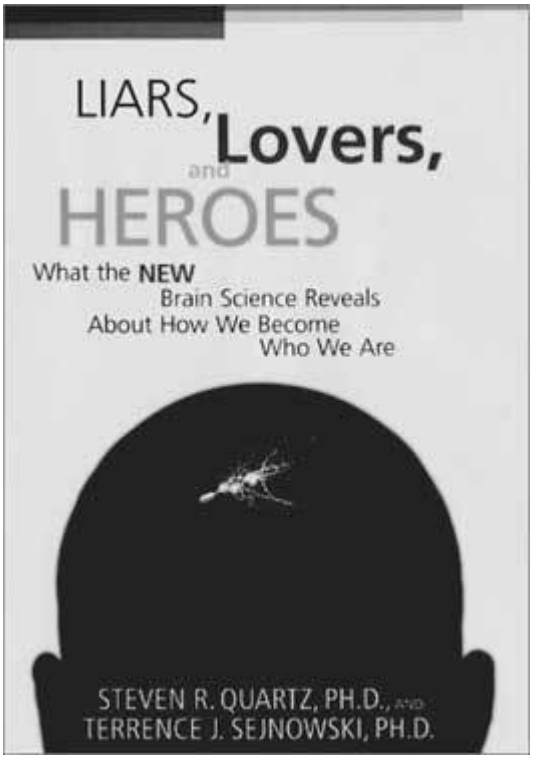
Мы все еще в поиске основных понятий, которые раскроют секрет высших форм интеллекта. Мы определили несколько ключевых принципов, но у нас нет ясной концептуальной основы, объясняющей, как работает мозг, – такой же элегантной, как ДНК, помогающая нам понять природу жизни. Алгоритмы обучения – хорошее место для поиска объединяющих понятий. Возможно, прогресс, к которому мы стремимся, чтобы разобраться, как сети глубокого обучения решают практические проблемы, даст больше подсказок. Возможно, мы откроем операционные системы в клетках и мозге, которые позволяют идти эволюции. Если мы разберемся в этом, то сложно вообразить последствия. Природа может быть умнее, чем каждый из нас, но я не вижу причин, почему мы как вид не можем рано или поздно раскрыть тайну интеллекта.

Дополнительная литература

Введение в нейробиологию



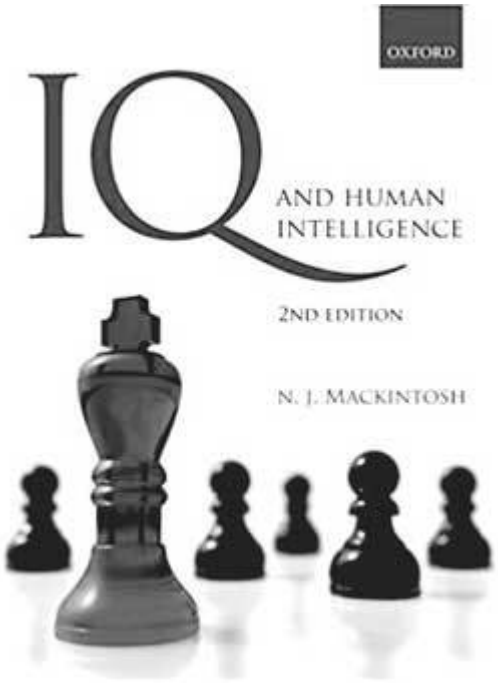
Эта книга лишь кратко коснулась нейробиологии, которая представляет собой обширную область с быстро расширяющимися границами. Наиболее актуальная для глубокого обучения часть нейробиологии называется системной нейробиологией. Если вы хотите узнать больше о мозге и нейронных сетях, хорошей отправной точкой станет книга **«Вычислительный мозг» («The Computational Brain»)**[482]. Она знакомит с основами нейробиологии и рассказывает, как нейронные сети применимы к широкому спектру структур мозга, таких как зрительная система, глазодвигательная система, управляющая движениями глаз, и способы представления пространства в коре.



Книга **«Лжецы, любовники и герои: Что новая наука о мозге говорит нам о том, как мы становимся теми, кто мы есть» («Liars, Lovers and Heroes: What the New Brain Science Has Revealed About How We Become Who We Are»)**[483], написанная для широкой аудитории, исследует, каким образом наши самые благородные и самые плохие черты коренятся в системах мозга, настолько древних, что мы разделяем их с насекомыми. Те самые алгоритмы, которые DeepMind использовала для обучения AlphaGo.

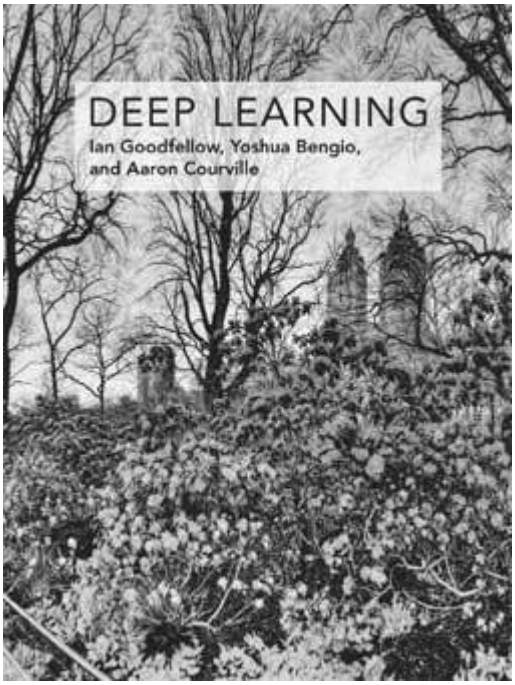
Общество нейробиологии поддерживает сайт **brainfacts.org**, где вы можете найти информацию о многих аспектах работы мозга и его расстройств.

Биологический интеллект



Книга **«IQ и человеческий интеллект» («IQ and human intelligence»)**[484] – заслуживающее доверия всестороннее введение в психологию интеллекта, включая социальный и эмоциональный интеллект. Биологическая основа интеллекта зависит от взаимодействия мозга с миром в процессе развития. Животный интеллект также был широко изучен, и «Животных разум» («Animal Minds»)[485] – хорошая книга для знакомства с ним.

Машинное обучение



Книга **«Распознавание образов и машинное обучение» («Neural Networks for Pattern Recognition»**)[486] хороша для изучения основ нейронных сетей. Между теорией информации и алгоритмами обучения есть глубинная связь, прекрасно изложенная в книге «**Теория информации, логический вывод и алгоритмы обучения» («Information Theory, Inference, and Learning Algorithms»)**[487]. Глубокое обучение быстро развивается: книга **«Глубокое обучение с точки зрения практика» («Deep Learning: A Practitioner's Approach»**)[488] подойдет для первого знакомства, а учебник **«Глубокое обучение» («Deep Learning»)**[489] в настоящее время доступен онлайн **www.deeplearningbook.org. «Машинное обучение: вероятностная перспектива» («Machine Learning: A Probabilistic Perspective»**)[490] – сборник, охватывающий более широкий спектр алгоритмов машинного обучения. Глубокое обучение с

подкреплением находится на переднем крае исследований, и удачная отправная точка – книга **«Обучение с подкреплением» (Reinforcement Learning: An Introduction)**[491].

Благодарности

Институт биологических исследований Солка (рис. 1), где я работаю[492], – особое место. Когда вы приближаетесь к институту снаружи, он выглядит как бетонная крепость, но когда вы входите в центральный двор, вид радикально меняется: широкое пространство из белого известняка простирается до Тихого океана, а башни по сторонам закрепляют ощущение нереальности. Моя лаборатория находится в южном корпусе со стороны внутреннего двора (на фото слева). Когда вы входите в лабораторию, слева вас встречает электронно-микроскопическая фотография гиппокампа размером со стену, которая выглядит как поперечное сечение тарелки со спагетти. Отсюда открывается вид на чайную комнату – сердце вычислительной лаборатории нейробиологии.

Рис. 1. Из Института биологических исследований Солка в Ла-Хойя в Сан-Диего открывается вид на Тихий океан. Это знаковое здание, спроектированное Луисом Каном, храм науки. Сюда я прихожу на работу каждый день

Круглый белый чайный стол был ареной для дискуссий с некоторыми из самых выдающихся ученых мира, в том числе Фрэнсисом Криком, который любил беседовать со студентами и коллегами на любые научные темы (рис. 2). Он даже упомянул чайную комнату в своей книге «Удивительная гипотеза»[493]:

Рис. 2. Чайная комната в Лаборатории вычислительной нейробиологии Института Солка в 2010 году. Ежедневные чаепития стали социальным инкубатором для развития многих алгоритмов обучения и научных открытий, описанных в этой книге

«Группа Терри Сейновски в Институте Солка собирается на неформальное чаепитие почти каждый вечер. Эти чаепития – идеальный повод обсудить результаты последних экспериментов, выдвинуть новые идеи или просто посплетничать о науке, политике и новостях в целом. Однажды я пришел на такое чаепитие и объявил Пэт Черчленд и Терри Сейновски, что нашел место, где находится воля! Она в передней поясной коре или где-то рядом. Когда я обсудил этот вопрос с Антонио Дамасио, выяснилось, что он тоже пришел к той же мысли».

Особенно мне памятен день в 1989 году, когда Фрэнсис Крик пришел на чай с Беатрис Голомб. Он сказал мне, что она хочет работать в области нейронных сетей, и я должен взять ее на работу. Беатрис получила докторскую степень по медицине и на тот момент

была аспиранткой в Калифорнийском университете в Сан-Диего и некоторое время занималась исследованиями под началом Крика. Она хотела работать над нейронными сетями для своей докторской диссертации, но биологический факультет не дал разрешения. Я последовал совету Крика и узнал от Беатрис столько же, сколько она узнала от меня, и продолжаю учиться у нее с тех пор, как мы поженились в 1990 году.

Стол для чаепитий приехал со мной из Университета Хопкинса: это был первый предмет, который я купил в 1981 году для новой лаборатории на моей первой работе на кафедре биофизики имени Томаса Дженкинса. Кафедра биофизики была похожа на старую семью, а я – на младшего сыночка, которого окружили заботой. Они вселили в меня уверенность двигаться в новом направлении, за что я бесконечно им благодарен. Я приобщился к традиции послеобеденного чаепития, будучи постдокторантом на кафедре нейробиологии Гарвардской медицинской школы. На крупной разносторонней кафедре это был способ поддерживать связь и узнавать о проводимых экспериментах. Моя лаборатория – миниатюрный университет, в котором обучаются студенты из разных областей естественных и технических наук, математики, и медицины, и чаепития – время, когда мы собираемся вместе как группа.

Я везунчик. Мои родители ценили образование и доверяли мне с раннего возраста. Я жил в период беспрецедентного экономического роста и возможности расширять свой кругозор. У меня были наставники и сотрудники, которые щедро делились со мной идеями и советами, и мне посчастливилось работать с поколением исключительно талантливых студентов. Я особенно благодарен Майку Стимаку, Джону Уилеру, Джону Хопфилду, Брюсу Найту, Штефану Куффлеру, Джеффри Хинтону и Соломону Голомбу, которые на развилках моей карьеры помогали мне повернуть в нужную сторону.

Я много кому благодарен за помощь в написании этой книги. Источником вдохновения послужили дискуссии с давней коллегой Патрицией Черчленд и Роджером Бингемом, основателем виртуального научного форума The Science Network. Выводы Джона Дойла из теории управления привели к обсуждению операционной системы мозга. Длительный поход с Кэри Сталлер в горы в Швейцарии помог мне разобраться со вселенной алгоритмов. Барбара Оакли научила меня обращаться к гораздо большей аудитории, чем класс. И Кэри, и Барбара помогли мне оформить эту историю. Многие помогли отзывами и идеями для книги, в том числе Марни Стюарт-Бартлетт, Йошуа Бенджио, Сидней Бреннер, Андреа Чиба, Патриция Черчленд, Гэри Коттрелл, Пол Экман, Микаэла Эннис, Джерри Фельдман, Адам Газзали, Джеффри Хинтон, Джонатан С. Ховард, Скотт Киркпатрик, Ли Тэ Вон, Марк и Джек Никрем, Джей Макклелланд, Барбара Оакли, Томми Поджио, Чарли Розенберг, Дэвид Сильвер, Джим Саймонс, Ричард Саттон, Джерри Тезауро, Паула Таллал, Себастьян Тран, Аджит Варки, Массимо Вергассола и Стивен Вольфрам.

На семинар по вычислительной нейробиологии в Вудс-Хоул, который проводится каждое лето с 1984 года с небольшим постоянным ядром и новыми участниками, всестороннее обсуждение проходило по утрам и вечерам, а послеобеденные часы оставались свободными для активного отдыха. Это было идеальное сочетание умственной и физической деятельности. Слушатели семинара сделали блестящую карьеру. Вудсхоулский семинар все еще существует, но в 1999 году он переехал в Теллурайд, чтобы совпасть с ежегодным семинаром по нейроморфной инженерии. Я благодарен всем тем, кто приходил на эти семинары за последние 30 лет, особенно Стиву Цукеру, Джону Оллману, Майку Страйкеру, Каталин Готард, Бэрри Ричмонду, Джону Дойлу, Дане Баллард, Джону Маунселлу, Бобу Десимону, Биллу Ньюсому и Кристофу Коху.

Мои коллеги из Института биологических исследований Солка и Калифорнийского университета в Сан-Диего – группа выдающихся

практиков и исследователей, создающих будущее биомедицинских наук. Преподаватели и студенты Института нейронных вычислений Калифорнийского университета в Сан-Диего интегрировали нейробиологию и вычисления таким образом, о котором я и не мечтал, когда основывал его в 1990 году.

Лаборатория вычислительной нейробиологии в Институте Солка была моим домом последние 30 лет, и многие мои «дети» сделали успешную карьеру по всему миру. Лаборатория походит на семью, и поколения аспирантов и постдокторантов, полных энтузиазма, значительно обогатили мою жизнь. О том, чтобы этот корабль держался на плаву, заботились Розмари Миллер и Мэри Эллен Перри.

Я благодарен издательству MIT Press, которое было надежным партнером 40 лет, опубликовав серию книг по вычислительной нейробиологии, которую я редактировал с Томми Поджио, и журналу Neural Computation, который я основал в 1989 году. Они выпустили в 1992 году мою книгу «Вычислительный мозг» и многие другие основополагающие книги по машинному обучению, включая книгу Саттона и Барто об обучении с подкреплением и передовой учебник по глубокому обучению. Боб Прайор из MIT Press помог довести эту книгу до момента публикации – это был долгий путь с неожиданными поворотами на дороге.

Благодарю сообщество NIPS, без которого я бы не писал эту книгу. Это не всеобъемлющая история данной области, и я сосредоточился только на нескольких темах и людях, которые участвовали в исследованиях нейронных сетей. Например, Международное общество нейронных сетей (International Neural Network Society; INNS) было основано в 1988 году и запустило новый журнал Neural Networks, который стойко ратовал за расширение охвата нейронных сетей. В партнерстве с Институтом инженеров электротехники и электроники (Institute of Electrical and Electronic Engineering; IEEE) они проводят ежегодную Международную совместную конференцию по нейронным сетям. Машинное обучение также породило много

конференций высокого уровня, включая Международную конференцию по машинному обучению (International Conference on Machine Learning; ICML), которая сродни Конференции NIPS. Область получила большую выгоду от всех подобных организаций и исследователей, внесших свой вклад.

Беатрис Голомб – критически мыслящий интеллектуал, и у нее я научился избегать шаблонного мышления. Я благодарен Джеффри Хинтону за то, что он делился со мной своим мнением о сетевых моделях на протяжении многих лет. Джефф был избран членом Королевского общества Англии и Канады, а я – членом Национальной академии наук, Национальной медицинской академии и Национальной инженерной академии: редкая честь – быть во всех трех национальных академиях США.

Соломон Голомб однажды сказал мне, что карьера происходит в ретроспективе, и я подтвердил его слова, когда писал эту книгу. Возвращаясь к своему прошлому, я анализировал события и решения, которые привели меня туда, где я сейчас, но, конечно, в то время я этого не знал. В 1964 году я написал три коротких текста для школьного литературного журнала. Первым было эссе о природе гравитации, которая приводила меня в восхищение. Вторым – рассказ о человеке, который обнаружил, что он компьютерная симуляция. Третий представлял собой размышления человека, продающего дом, в котором вырос. Много лет спустя, будучи аспирантом Принстонского университета, я продолжил исследование черных дыр и гравитационных волн в общей теории относительности, теории гравитации Эйнштейна. Получив докторскую степень по физике, я перешел в область биологии, и с тех пор объектом моего интереса стал мозг. Я пока не знаю, что делать с третьей историей, – возможно, она превратится в еще одну главу моей жизни.

Глоссарий

Адаптивная обработка сигналов – регулируемый фильтр, преобразующий сигналы. Примером может служить фильтр для уменьшения шума в регулируемой частотном диапазоне.

Алгоритм обучения – алгоритм изменения параметров функции на основе примеров. Алгоритм обучения может быть контролируемым, если заданы входные и желаемые выходные данные, или неконтролируемым, если заданы только входные данные. Обучение с подкреплением – частный случай контролируемого алгоритма обучения, когда единственная обратная связь – награда за хорошую работу.

Градиентный спуск – метод оптимизации, при котором параметры изменяются каждую эпоху, чтобы уменьшить функцию стоимости.

Логика – умозаключение, основанное на предположениях, которые могут быть только истинными или ложными. Математики используют логику для доказательства теорем.

Масштабирование – увеличение сложности алгоритма с увеличением размеров задачи.

Машина Тьюринга – гипотетический компьютер, изобретенный Аланом Тьюрингом в 1937 году в качестве простой модели для математических расчетов. Машина Тьюринга состоит из ленты, которую можно перемещать вперед и назад, головки записи-чтения, находящейся в одном из множества состояний, которая может изменять свойства активной ячейки под ней, и набора инструкций, как головка должна изменять активную ячейку и перемещать ленту. На каждом шаге машина может доработать свойство активной клетки и изменить положение головки, а после переместить ленту на одну ячейку.

MOOK – массовые открытые онлайн-курсы. Лекции по широкому кругу тем, находящиеся в свободном доступе в Интернете. Первый

МООС появился в 2006 году, к 2017 году было открыто 6850 онлайн-курсов, которые прослушали 59 миллионов человек.

Нейрон – специализированная клетка мозга, которая объединяет входные сигналы от нейронов и отправляет выходные данные другим нейронам.

Нормализация – поддержание амплитуды сигнала в заданных пределах. Например, если изменяющийся во времени положительный сигнал делится на его максимальное значение, то он будет ограничен 1.

Обратная связь – соединения, которые движутся в нейронной сети в обратном направлении от более высоких уровней к более низким, создавая в сети петлю, позволяющую сигналам циркулировать.

Обратное распространение ошибки – алгоритм обучения, который оптимизирует нейронную сеть с помощью градиентного спуска, чтобы минимизировать функцию затрат и повысить производительность.

Обучающие и тестовые наборы – производительность обучающего набора недостаточно точно оценивает, как нейронная сеть будет работать на новых входных данных. Тестовый набор, не используемый во время обучения, позволяет оценить, насколько хорошо обобщена сеть. Когда наборы данных малы, одну выборку можно убрать из обучающего набора и использовать для тестирования производительности сети, обученной на остальных примерах, повторяя процесс для каждой выборки, чтобы получить среднюю производительность теста. Это частный случай перекрестной проверки с n = 1, в которой удерживается n подвыборок.

Ограничения – условия задачи по оптимизации, которым должно удовлетворять решение. Например, решение может иметь только положительное значение.

Оптимизация – процесс максимизации или минимизации функции путем систематического поиска входных значений из допустимого

набора и вычисления значения функции.

Переобучение – состояние, когда количество настраиваемых параметров в сетевой модели сильно превышает количество обучающих данных, и большинство алгоритмов обучения просто запоминают примеры. Это значительно снижает способность обобщать новые примеры. Регуляризация – способ уменьшить переобучение.

Пластичность – изменения функций нейрона, проявляющиеся в усилении связей (синаптическая пластичность) или в том, как нейрон реагирует на его входные сигналы (внутренняя пластичность).

Правило Байеса – формула, обновляющая вероятность события на основе новых данных и уже известных условий, связанных с событием. В более общем случае байесовские вероятности – это представления о результатах, основанные на текущих и предыдущих данных.

Равновесие – состояние термодинамической системы, при котором отсутствуют чистые макроскопические потоки вещества или энергии. В машине Больцмана элементы вероятностны, и если входные сигналы остаются постоянными, то система приходит в равновесие.

Распределение вероятностей – функция, определяющая вероятность возникновения всех возможных состояний системы или результатов эксперимента.

Регуляризация – способ избежать переобучения модели с большим количеством параметров, когда данные ограничены. Распространенным методом является снижение веса, при котором все веса в сети уменьшаются в каждую эпоху обучения, и выживают только веса с большими положительными градиентами.

Рекуррентная сеть – нейронная сеть с обратными связями, позволяющими сигналам циркулировать внутри сети.

Свертка – способ смешивания одной функции с другой путем вычисления, в какой мере одна функция перекрывает другую при их наложении.

Сеть прямого распространения – многоуровневая нейронная сеть с односторонней связью между слоями, начиная с входного слоя и заканчивая выходным.

Синапс – особое соединение между двумя нейронами, где сигнал передается от пресинаптического нейрона к постсинаптическому нейрону.

Функция стоимости – функция, которая определяет цель сети и количественно оценивает ее производительность. Целью обучения является снижение функции стоимости.

Шипик – тонкий вырос на дендрите, способный образовать синаптическое соединение.

Эпоха – одно обновление весов во время обучения после того, как средний градиент вычислен на основе заданного количества примеров.

* * *


АНТОЛОГИЯ МАШИННОГО ОБУЧЕНИЯ
важнейшие исследования в области ИИ за последние 60 лет
Терренс Сейновски
Профессор в Институте биологических исследований Солка, директор Института нейронных вычислений Калифорнийского университета, Сан-Диего
БОМБОРА

Примечания

1

Астро Теллер, «капитан муншотов» Google (moonshot – полет на Луну, в переносном смысле – смелый и передовой проект), 6 февраля 2015 года сказал, что глубокое обучение помогло снизить потребление энергии на 15 % и тем самым ежегодно экономить компании Google X сотни миллионов долларов. – *Прим. авт.*

Вернуться

2

Реверсивная инженерия, или обратное проектирование, – исследование объекта с целью понять принцип его работы. – *Прим. ред.*

Вернуться

3

Перцептрон – математическая или компьютерная модель восприятия информации мозгом. – *Прим. ред.*

Вернуться

4

В переводе на русский книга вышла в 1965 году. – *Прим. ред.*

Вернуться

5

В переводе на русский книга вышла в 1971 году. – *Прим. ред.*
Вернуться

6

«О дивный новый мир, где обитают такие люди!» – «Буря», Шекспир, пер. О. Сорока, 1989. – *Прим. авт.*
Вернуться

7

Beer Bottle Pass – перевал в горной цепи Люси-Грей-Маунтинс в штате Невада. – *Прим. ред.*
Вернуться

8

Власич Б. General Motors хочет управлять беспилотными автомобилями будущего. New York Times, 4 июня 2017 года. www.nytimes.com/2017/06/04/business/general-motors-self-driving-cars-mary-barra.html
Вернуться

9

Ингрем К. Невероятный человеческий потенциал, потраченный на дорогу на работу. Washington Post, 24 февраля 2016 года. www.washingtonpost.com/news/wonk/wp/2016/02/25/how-much-of-your-life-youre-wasting-on-your-commute
Вернуться

10

Американские центры обработки данных потребляют все больше энергии. Совет по защите природных ресурсов, 2015. www.nrdc.org/resources/americas-data-centers-consuming-and-wasting-growing-amounts-energy

Вернуться

11

Льюис-Краус Гидеон. New York Times Magazine, 14 декабря 2016 года. www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html?_r=0 Hemingway is # 1

Вернуться

12

Цит. по: Хемингуэй Э. Снега Килиманджаро / Перевод с английского Н. А. Волжиной. М., 1968.

Вернуться

13

Перевод оригинала с английского языка на русский, выполненный Google Переводчиком в 2021 году. – *Прим. ред.*

Вернуться

14

Eugene Onegin. A Novel in Verse by Alexandr Pushkin / Translated from the Russian, with a Commentary, by Vladimir Nabokov. In four volumes. – NY: Pantheon Books, 1964.

Вернуться

15

Ранние попытки приведены в статье «Завышенная эффективность рекуррентных нейронных сетей» по ссылке: karpathy.github.io/2015/05/21/rnn-effectiveness/ – *Прим. авт.*

Вернуться

16

G. Hinton, L. Deng, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, B. Kingsbury, «Deep Neural Networks for Acoustic Modeling in Speech Recognition», IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82–97, Nov. 2012.

Вернуться

17

W. Xiong, Wayne Xiong, Jasha Droppo, Xuedong Huang, Frank Seide, Achieving Human Parity in Conversational Speech Recognition, arXiv:1610.0525.

Вернуться

18

Esteva A., Kuprel B., Novoa R. A., Ko J., Swetter S. M., Blau H. M., Thrun S. Dermatologist-Level Classification of Skin Cancer With Deep Neural Networks. Nature 542 (7639), 115–118. 2017.

Вернуться

19

Siddhartha Mukherjee, A. I. Versus M.D. What happens when diagnosis is automated? April 3, 2017 New Yorker. www.newyorker.com/magazine/2017/04/03/ai-versus-md

Вернуться

20

Kales A., Rechtschaffen A. (Eds.) A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Allan Rechtschaffen and Anthony Kales, editors. National Institutes of Health publication, no. 204, Bethesda, Md., U. S. National Institute of Neurological Diseases and Blindness, Neurological Information Network, 1968.

Вернуться

21

Sei Chong, Morning Agenda: Big Pay for Hedge Fund Chiefs Despite a Rough Year, New York Times May 16, 2017 www.nytimes.com/2017/05/16/business/dealbook/hedge-funds-amazon-bezos.html

Вернуться

22

За исключением Агентства Национальной Безопасности, в числе сотрудников которого – сотни математиков (Alfred W. Hales, personal communication). – *Прим. авт.*

Вернуться

23

Сарфраз Манзур. «Гении математики работают на Уолл-стрит. Биржевые маклеры остаются без работы, так как их заменят гении математики, использующие сверхсовременные компьютеры. Но во благо ли это или во вред?» The Telegraph, 23 июля 2013 года.

Вернуться

24

Shaw D. E.; Chao Jack C.; Eastwood Michael P.; Gagliardo Joseph; Grossman J. P.; Ho C. Richard; Ierardi Douglas J.; Kolossváry István; et al. (May 2007). «Anton – компьютер для моделирования молекулярной динамики». International Symposium on Computer Architecture: Proceedings of the 34th annual international symposium on Computer architecture. ACM. 35 (2).

Вернуться

25

Далее – МТИ.

Вернуться

26

Ян Эллисон. «Бывший физик-ядерщик Анри Вельбрук объясняет, как машинное обучение снижает риск высокочастотной торговли», International Business Times, 23 марта 2016 г. www.ibtimes.co.uk/formernuclear-physicist-henri-waelbroeck-explains-how-machine-learning-mitigates-high-frequency1551097

Вернуться

27

Такой подход уже получил имя «электронное расследование» (также «автоматизированный анализ», или technology-assisted review, TAR) – автоматизированный сбор и анализ цифровой информации, релевантной для конкретного судебного процесса. – *Прим. ред.*

Вернуться

28

Moravčík M., Schmid M., Burch N., Lisý V., Morrill D., Bard N., Davis T., Waugh K., Johanson M., Bowling M. «DeepStack: Expert-level artificial intelligence in heads-up no-limit poker». Science. 356: 508–513 2017. Стандартное отклонение – это половина ширины колоколообразной кривой. Только 16 процентов выборок больше одного стандартного отклонения от среднего. Только один из десяти миллионов образцов имеет более четырех стандартных отклонений от среднего значения. – *Прим. авт.*

Вернуться

29

Вспоминается сюжет американского фильма «Военные игры 1983 года». – *Прим. авт.*

Вернуться

30

Silver David; Huang Aja; Maddison Chris J.; Guez Arthur; Sifre Laurent; Driessche George van den; Schrittwieser Julian; Antonoglou, Ioannis; Panneershelvam Veda (2016). «Mastering the game of Go with deep neural networks and tree search». Журнал Nature. 529 (7587): 484–489.

Вернуться

31

«Surveyor-1» приземлился на поверхность Луны 2 июня 1966 года в 6:17:36 UT. Место посадки находилось на равнинной территории в 100-километровом кратере к северу от кратера Флемстид. – *Прим. авт.*

Вернуться

32

«Ужасное разочарование подростка, проигравшего AlphaGo», Quartz, May 27, 2017. qz.com/993147/the-awful-frustration-of-a-teenage-go-champion-playing-googles-alphago/

Вернуться

33

Paul Mozur, “Beijing Wants A. I. to Be Made in China by 2030,” New York Times, July 20, 2017. www.nytimes.com/2017/07/20/business/china-artificial-intelligence.html?_r=0

Вернуться

34

Выражение «состояние потока», «быть в потоке» означает психическое состояние, в котором человек полностью включен в то, чем он занимается. – *Прим. ред.*

Вернуться

35

Также текучий, флюидный. – *Прим. ред.*

Вернуться

36

Говард Гарднер. Структура разума: теория множественного интеллекта. – М.: Вильямс, 2007.

Вернуться

37

Flynn J. R. Massive IQ gains in 14 nations: What IQ tests really measure. Psychological Bulletin. 101, 171–191 (1987).

Вернуться

38

Douglas C. Engelbart, Augmented Intelligence: Smart Systems and the Future of Work and Learning, SRI Summary Report AFOSR-3223, www.dougengelbart.org/pubs/augment-3906.html

Вернуться

39

На русском языке доступно по ссылке https://ru.coursera.org/learn/learnhow. – *Прим. ред.*

Вернуться

40

Оакли Барбара. «Mindshift. Новая жизнь, профессия и карьера в любом возрасте». Издательство «Питер», 2020.

Вернуться

41

Рынок образования оценивается в 1,3 триллиона долларов и делится на следующие основные секторы: раннее детское образование – 70 миллиардов долларов; начальное и среднее образование – 670 миллиардов долларов; высшее образование – 475 миллиарда долларов. medium.com/students-for-the-future/how-big-is-the-education-market-in-the-us-report-from-whitehouse-91dc313257c5. – *Прим. авт.*

Вернуться

42

Британский ученый, известный своими работами в области глубинных нейронных сетей. – *Прим. ред.*

Вернуться

43

Algorithmic retailing: Automatic for the people, журнал The Economist, April 15, 2017, p. 56.

Вернуться

44

Нейронные системы обработки информации. Ранее была аббревиатура NIPS, в 2018 году ее сменили на NeurIPS. Так как данная книга была написана незадолго до переименования, в

оригинале и в библиографических ссылках употребляется аббревиатура NIPS, которую мы и будем использовать далее во избежание путаницы. – *Прим. ред.*

Вернуться

45

По словам Михаэлы Эннис, студентки МТИ в 2016 году, «историю о том, что старшекурсникам МТИ поручили создать компьютерную зрительную систему в качестве летнего проекта, ежегодно рассказывает профессор Винстон. Также он говорит, что его студентом был Сассман». – *Прим. авт.*

Вернуться

46

Roger Peterson, Guy Mountfort and P.A.D. Hollom, Field Guide to the Birds of Britain and Europe. Peterson Field Guides Series, 2001.

Вернуться

47

Бьюкенен; Шортлиф (1984). «Экспертные системы на основе правил. Эксперименты MYCIN Стэнфордского проекта эвристического программирования. Reading, MA: Addison-Wesley.

Вернуться

48

Сиддхартха Мукерджи. «Искусственный интеллект против медицины: Что случится, когда постановка диагноза станет

автоматизированной?» Нью-Йорк, 3 апреля 2017 года. www.newyorker.com/magazine/2017/04/03/ai-versus-md

Вернуться

49

Педро Домингос (2015 год). «Мастер-алгоритм: как поиски Ultimate Learning Machine переделают наш мир». Никто не знает, как измерить здравый смысл, который мы воспринимаем как данность. – *Прим. авт.*

Вернуться

50

«Коты легче людей и могут группироваться в воздухе, даже если падают спиной вниз». Sechzera Jeri A.; Folsteina Susan E.; Geigera Eric H.; Mervisa Ronald F.; Meehana Suzanne M. (December 1984). «Development and maturation of postural reflexes in normal kittens». Experimental Neurology. 86 (3): Pages 493–505.

Вернуться

51

Stefik M. Strategic computing at DARPA: Overview and assessment. Commun. ACM 28, 7 (July 1985). 690–704.

Вернуться

52

Находится в США, штат Массачусетс. – *Прим. ред.*

Вернуться

53

А также другие бытовые роботы и несколько линеек роботов-помощников для армии США. – *Прим. ред.*

Вернуться

54

Сейчас Национальный центр суперкомпьютерных приложений (NCSA) – один из ведущих центров по разработке высокопроизводительных вычислительных систем. – *Прим. ред.*

Вернуться

55

Tesauro G. Sejnowski T. J. A Parallel Network That Learns to Play Backgammon, Artificial Intelligence Journal, 39, 357–390, 1989.

Вернуться

56

На местном уровне обнаруживаются различия в клеточных свойствах и связях между отдельными частями коры, которые, по-видимому, отражают специализацию разных сенсорных систем и уровней иерархии. – *Прим. авт.*

Вернуться

57

Wason P. C. (1977). «Self-contradictions». In Johnson-Laird P. N.; Wason P. C. Thinking: Readings in cognitive science. Cambridge: Cambridge University Press.

Вернуться

58

Главная особенность такой архитектуры состоит в том, что память физически разделена и система строится из отдельных модулей, каждый из которых представляет собой полнофункциональный компьютер. – *Прим. ред.*

Вернуться

59

По Lindsay P. H, Norman D. A. Human Information Processing: An Introduction to Psychology. 2nd ed. New York: Academic Press; 1977.

Вернуться

60

Wiener Norbert (1948). Cybernetics, or Control and Communication in the Animal and the Machine. Cambridge: MIT Press.

Вернуться

61

O. G. Selfridge. «Pandemonium: A paradigm for learning». In D. V. Blake and A. M. Uttley, editors, Proceedings of the Symposium on Mechanisation of Thought Processes, pages 511–529, London, 1959.

Вернуться

62

Пандемониум (обучаемая классифицирующая машина) и демон (процедура, запускаемая автоматически при выполнении некоторых условий) давно стали терминами в сфере IT и утратили метафорическое значение. – *Прим. ред.*

Вернуться

63

Least Mean Squares. See B. Widrow and S. D. Stearns, Adaptive Signal Processing, PrenticeHall, 1985. LMS: Least Mean Squares.

Вернуться

64

Модель Samsung S6 была выпущена в 2015 году. На 2021 год у флагманских смартфонов производительность достигает триллиона операций в секунду. – *Прим. ред.*

Вернуться

65

Gray M. S., Lawrence D. T., Golomb B. A., Sejnowski T. J. A Perceptron Reveals the Face of Sex, Neural Computation, 7, 1160–1164, 1995.

Вернуться

66

Golomb B. A., Lawrence D. T., Sejnowski T. J. “SEXNET: A Neural Network Identifies Sex from Human Faces,” Touretzky, D. S. Lippmann, R. (Ed.), Advances in Neural Information Processing Systems, 3, San Mateo, CA: Morgan Kaufmann Publishers, 572–577, 1991.

Вернуться

67

Отсылка к популярному телешоу 1950-х годов Dragnet, показывающее преступников по материалам ФБР. – *Прим. авт.*

Вернуться

68

Mikel Olazaran (1996). «A Sociological Study of the Official History of the Perceptrons Controversy». Social Studies of Science. 26 (3): 611–659.

Вернуться

69

С 1990 года работает в США. – *Прим. ред.*

Вернуться

70

Vapnik V. The Nature of Statistical Learning Theory. Springer-Verlag, New York, 1995.

Вернуться

71

Liu W.; Principe J.; Haykin S. (2010). Kernel Adaptive Filtering: A Comprehensive Introduction. Wiley, New Jersey.

Вернуться

72

Christoph von der Malsburg. “The correlation theory of brain function.” Internal Report 81–2, MPI Biophysical Chemistry, 1981. See Reprint in Models of Neural Networks II, chapter 2, pages 95–119. Springer, Berlin, 1994. fias.uni-frankfurt.de/fileadmin/fias/malsburg/publications/vdM_correlation.pdf

Вернуться

73

P. Wolfrum, C. Wolff, J. Lücke, and C. von der Malsburg. “A recurrent dynamic model for correspondence-based face recognition.” Journal of Vision 8, 1–18, 2008.

Вернуться

74

Fukushima, Neocognitron (1980). «A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position». Biological Cybernetics. 36 (4): 93–202.

Вернуться

75

Kohonen Teuvo (1982). «Self-Organized Formation of Topologically Correct Feature Maps». Biological Cybernetics. 43 (1): 59–69.

Вернуться

76

Pearl J. 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Mateo, Calif.: Morgan Kaufmann.

Вернуться

77

Hinton G., Anderson J. A. Parallel Models of Associative Memory, Erlbaum (1981).

Вернуться

78

Временная (на 2–4 года) позиция кандидата или доктора наук, который занимается научно-исследовательской деятельностью. – *Прим. ред.*

Вернуться

79

Горнолыжный курорт в штате Вайоминг, США. – *Прим. ред.*

Вернуться

80

Sejnowski T. J. David Marr: A Pioneer in Computational Neuroscience Vaina, L. (Ed.), In: Selected Papers of David Marr, Boston, MA: Birkhäuser, 297–301, 1991 58.

Вернуться

81

Marr D., Poggio T. (1976), Cooperative computation of stereo disparity. Science 194: 283–287.

Вернуться

82

Julesz Béla (1971). Foundations of Cyclopean Perception. Chicago: The University of Chicago Press.

Вернуться

83

«Magic Eye» – автостереограммы со скрытой трехмерной структурой, создающей узор, который можно увидеть, расширяя глаза. Кристофер Тайлер создал первые черно-белые автостереограммы в 1979 году. – *Прим. авт.*

Вернуться

84

Марр Д. Зрение. Информационный подход к изучению представления и обработки зрительных образов / Изд-во «Радио и связь». М., 1987.

Вернуться

85

Теория вероятностей – раздел математики, изучающий закономерности случайных явлений. – *Прим. ред.*

Вернуться

86

Sejnowski T. J. A Stochastic Model of Nonlinearly Interacting Neurons, Princeton University, Thesis, 1978.

Вернуться

87

Sejnowski T. J. Vernon Mountcastle: Father of Neuroscience, Proceedings of the National Academy of Sciences U.S.A., 112, 6523–6524, 2015.

Вернуться

88

В Университете Хопкинса есть три отделения биофизики: на факультетах медицины и общественного здравоохранения и в Школе искусств и наук Кригера (я был на факультете биофизики Томаса Дженкинса в кампусе Homewood Школы искусств и наук). – *Прим. авт.*

Вернуться

89

(ориг.) Mind/Brain Institute.

Вернуться

90

Перевод С. Я. Маршака. – *Прим. ред.*

Вернуться

91

Электрорецепция – способность животных ощущать электрические сигналы окружающей среды. – *Прим. ред.*

Вернуться

92

Sejnowski T. J., Yodlowski M. L. A Freeze-Fracture Study of the Skate Electroreceptor, Journal of Neurocytology, 11, 897–912, 1982.

Вернуться

93

Sejnowski T. J., Reingold S. C., Kelley D. B. Gelperin, A. Localization of [3H]-2-Deoxyglucose in Single Molluscan Neurones, Nature, 287, 449–451, 1980.

Вернуться

94

Kuffler S. W., Sejnowski T. J. Peptidergic and Muscarinic Excitation At Amphibian Sympathetic Synapses, Journal of Physiology, 341, 257–278, 1983.

Вернуться

95

Разветвленный отросток нейрона, принимающий сигналы от других нейронов. – *Прим. ред.*

Вернуться

96

System Development Corporation – компания некоммерческих программ в Санта-Монике, работающая по контракту с американской армией. Была основана в 1969 году для

распределения доходов от продажи зданий в рамках программы грантов с 1980 по 1988 год. – *Прим. авт.*

Вернуться

97

Отсылка к английской пословице He is a good friend that speaks well of us behind our backs. – Тот хороший друг, который о нас за глаза хорошее говорит. – *Прим. ред.*

Вернуться

98

Лисп-машина – продукт компании Symbolics, предназначенный для написания программ обработки символов ИИ. Тем не менее он был недостаточно хорош для обработки чисел, что было необходимо для моделирования нейронных сетей. – *Прим. авт.*

Вернуться

99

Мне как участнику президентской программы «Молодой исследователь» предложили хорошую скидку от новой компьютерной компании Ridge, которая продавала компьютеры мощностью VAX-780, что в то время было очень хорошим вариантов для теоретических вычислений. – *Прим. авт.*

Вернуться

100

Далее – Калтех.

Вернуться

101

Клуб Гельмгольца был создан Фрэнсисом Криком, Вилейануром Рамачандраном и Гордоном Шоу в 1980-х годах и продолжало свою деятельность на протяжении 20 лет. Более подробная история: Aicardi C. Of the Helmholtz Club, SouthCalifornian seedbed for visual and cognitive neuroscience, and its patron Francis Crick, Stud Hist Philos Biol Biomed Sci. 45, 1–11 (2014). – *Прим. авт.*

Вернуться

102

«Я многому научился у всех, с кем сталкивался. Мне стыдно принимать идеи от людей… Мой самый интенсивный опыт – это Клуб Гельмгольца. Не знаю, слышали ли вы о нем. …Там было, может, человек двадцать. Я никогда не пропустил ни разу. Я находил, кто может меня подменить на занятии, если оно совпадало по времени со встречей в клубе. Но я бы ходил туда в любом случае, потому что это слишком важно, чтобы пропустить». Carver Mead. In: Anderson J. A., Rosenfeld E., editors. Talking nets: An oral history of neural networks. The MIT Press; Cambridge & London: 2000. P. 238.

Вернуться

103

Desimone R., Albright T. D., Gross C. G., Bruce C. Stimulus-selective properties of inferior temporal neurons in the macaque, J Neurosci. 4: 2051–2062. 1984. У многих исследователей в лаборатории Гросса были бороды, поэтому нейроны в зрительной коре, которые реагировали на ершики для туалета, так же реагировали на бороды. – *Прим. авт.*

Вернуться

104

Практически всегда один из глаз доминантный, то есть пропускает через себя чуть больший объем информации в визуальную область головного мозга и точнее передает информацию о местоположении объектов. – *Прим. ред.*

Вернуться

105

Hubel David (1988) Eye, Brain, and Vision. W. H. Freeman & Co., New York, 1988 стр. 191–216.

Вернуться

106

Конец критического периода может быть не таким резким, как считалось ранее, и стереовидение может быть достигнуто у взрослых с исправленным косоглазием после интенсивной практики. Более подробная информация: См.: Barry S. R. (2009). Fixing my gaze: A scientist's journey into seeing in three dimensions. New York: Basic Books. Я знал «стерео Сью», когда был аспирантом в Принстонском университете. – *Прим. авт.*

Вернуться

107

Есть несколько исключений из этого правила: гранулярные клетки в зубчатой извилине гиппокампа и нейроны в обонятельной луковице генерируются на протяжении всей жизни. Specter M., Rethinking the Brain, New Yorker, July 23, 2001, www.michaelspecter.com/2001/07/rethinking-the-brain/ – *Прим. авт.*

Вернуться

108

Sejnowski T. J. How Do We Remember the Past? Brockman, J. (Ed.), In: What We Believe But Cannot Prove: Today's Leading Thinkers on Science in the Age of Certainty, The Free Press, 97–99, 2005; Roger Y. Tsien, Very long-term memories may be stored in the pattern of holes in the perineuronal net, Proceedings of the National Academy of Sciences, USA, 110: 12456–12461 (2013).

Вернуться

109

При болезни Альцгеймера нарушается целостность внеклеточного матрикса, что может способствовать потере долговременной памяти (из личного разговора с нейробиологом Джоном Оллмэном). – *Прим. авт.*

Вернуться

110

Одна из знаковых черт главного героя юмористического романа «Жизнь и мнения Тристрама Шенди, джентльмена» – его любовь к пространным отступлениям от основной темы. – *Прим. ред.*

Вернуться

111

Интервью с Гери вы можете найти по ссылке scienceblogs.com/retrospectacle/2006/10/15/sfn-special-lecture-architect-1/ – *Прим. авт.*

Вернуться

112

Kunsberg B. S., Zucker S. W. Critical Contours: An Invariant Linking Image Flow with Salient Surface Organization, arXiv:1705.07329.

Вернуться

113

Связь между ними объясняется геометрией критических точек и градиентными переходами на поверхностях, называемых комплексом Морса – Смейла. – *Прим. авт.*

Вернуться

114

Lehky S. R., Sejnowski T. J. Network Model of Shape-from-Shading: Neural Function Arises from Both Receptive and Projective Fields, Nature, 333, 452–454, 1988.

Вернуться

115

Оптогенетика – методика исследования работы нервных клеток, основанная на внедрении в их мембрану специальных каналов – опсинов, реагирующих на возбуждение светом; разработана в 2005 году. – *Прим. ред.*

Вернуться

116

Sejnowski T. J. What Are the Projective Fields of Cortical Neurons? van Hemmen J. L., Sejnowski T. J. (Ed.), In: 23 Problems in Systems Neuroscience, Oxford University Press, 394–405, 2005.

Вернуться

117

В специальной литературе обычно фигурирует название «зона MT» от англ. *middle temporal area*. – *Прим. ред.*

Вернуться

118

Коннекто́м – полное описание структуры связей в нервной системе организма. – *Прим. ред.*

Вернуться

119

Geddes Linda (20 July 2016). «Human brain mapped in unprecedented detail; nearly 100 previously unidentified brain areas revealed by examination of the cerebral cortex». Nature doi:10.1038/nature.2016.20285.

Вернуться

120

Диффузионная спектральная томография отслеживает направление аксонов, которые составляют белое вещество в коре. – *Прим. авт.*

Вернуться

121

“Two Foundations Collaborate On Cognitive Neuroscience”, The Scientist, October 1989 www.the-scientist.com/?articles.view/articleNo/10719/title/Two-Foundations-Collaborate-OnCognitive-Neuroscience/

Вернуться

122

С электромагнитными волнами длиной менее одного миллиметра. – *Прим. ред.*

Вернуться

123

Hasson U., Yang E., Vallines I., Heeger D. J., Rubin N. A hierarchy of temporal receptive windows in human cortex. Journal of Neuroscience 28(10):2539–2550 (2008).

Вернуться

124

«The Organization of Behavior».

Вернуться

125

«Neural networks and physical systems with emergent collective computational abilities».

Вернуться

126

«A learning algorithm for Boltzmann Machines».
Вернуться

127

«Learning internal representations by error-propagation».
Вернуться

128

«Learning to predict by the methods of temporal differences».
Вернуться

129

«An information-maximization approach to blind separation and blind deconvolution».
Вернуться

130

«ImageNet Classification with Deep Convolutional Neural Networks».
Вернуться

131

J. Herault, C. Jutten. Space or time adaptive signal processing by neural network models. In JS Denker (Ed.) Neural Networks for Computing, AIP Conference Proceedings 151 (1), 206–211 (1986).

Вернуться

132

Синонимы: анализ независимых компонент, метод независимых компонент. В специальной литературе, как правило, употребляется аббревиатура ICA (англ. Independent Component Analysis). – *Прим. ред.*

Вернуться

133

System gets wired up during development: Linsker R (1988). «Self-organization in a perceptual network» IEEE Computer. 21 (3): 105–17.

Вернуться

134

Bell A. J., Sejnowski T. J. “An Information-Maximization Approach to Blind Separation and Blind Deconvolution”, Neural Computation, 7, 1129–1159, 1995.

Вернуться

135

Также важный вклад в разработку независимого компонентного анализа внесли Пьер Комон, Жан-Франсуа Кардозу, Эпо Хиваринен, Эркки Оя, Анджей Цихоцкий, Сюнъити Амари, Ли Тэ Вон, Майкл Левицки и многие другие. Почти все важные открытия совершаются разными способами несколькими исследователями почти одновременно. – *Прим. авт.*

Вернуться

136

Bell A. J., Sejnowski T. J. “The 'Independent Components' of natural scenes are edge filters,” Vision Research, 37, 3327–3338, 1997.

Вернуться

137

Бруно Ольсхаузен и Дэвид Филд пришли к такому же выводу с другим алгоритмом обучения, основанным на разреженности: BA Olshausen, DJ Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images.” Nature 38: 607, 1996. – *Прим. авт.*

Вернуться

138

Barlow H. (1961) «Possible principles underlying the transformation of sensory messages» in Sensory Communication, MIT Press.

Вернуться

139

Bell A. J., Sejnowski T. J. “Learning the Higher-Order Structure of a Natural Sound”, Network: Computation in Neural Systems, 7, 261–266, 1996.

Вернуться

140

A. Hyvarinen and P. Hoyer. “Emergence of phase- and shift invariant features by decomposition of natural images into independent feature

subspaces". Neural Computation, 12(7): 1705–1720, 2000.
Вернуться

141

Функция потерь – мера расхождения между истинным значением оцениваемого параметра и оценкой параметра. – *Прим. ред.*
Вернуться

142

Система с двумя противоположно заряженными полюсами. – *Прим. ред.*
Вернуться

143

McKeown M. J., Jung T.-P., Makeig S., Brown G. D., Kindermann S. S., Lee T.-W., Sejnowski T. J. "Spatially Independent Activity Patterns in Functional MRI Data During the Stroop Color-Naming Task," Proceedings of the National Academy of Sciences of the United States of America, U.S.A. 95, 803–810, (1998).
Вернуться

144

Mantini D., Perrucci M. G., Del Gratta C., Romani G. L., Corbetta M. (2007) "Electrophysiological signatures of resting state networks in the human brain." Proceedings of the National Academy of Sciences of the United States of America, U.S.A 104:13170–13175.
Вернуться

145

Donoho, D.L. (2006). «Compressed sensing». IEEE Transactions on Information Theory. 52 (4): 1289–1306.

Вернуться

146

Возможно, в мозжечке, на уровне мшистых волокнистых входов, сходящихся на дендритах гранулярных клеток: Eagleman D. M., Coenen O. J.-M. D., Mitsner V., Bartol T. M., Bell A. J., Sejnowski T. J. “Cerebellar Glomeruli: Does Limited Extracellular Calcium Implement a Sparse Encoding Strategy?”, Proceedings of the 8th Joint Symposium on Neural Computation, The Salk Institute, 2001.

Вернуться

147

Тони Белл изучает структуру воды с помощью ICA и спектроскопии в ближней инфракрасной области. Он пытается доказать, что вода образует связанные структуры, которые взаимодействуют через свет и образуют питательную среду для биомолекулярной жизни в невидимых до сих пор масштабах. Идея состоит в том, что решение возникает, когда «нейронные схемы» разряжаются в достаточной степени, чтобы распространить связанную информацию из более распределенных атомных сетей внутри клеток по всему телу. – *Прим. авт.*

Вернуться

148

Минимальное время, за которое большинство нейронов могут сформировать решение, – около 10 миллисекунд, а путь к решению за 1 секунду состоит не более чем из 100 временных шагов. – *Прим. авт.*

Вернуться

149

В физике электромагнетизма Майкл Фарадей применял «грязные» методы, а Джеймс Клерк Максвелл – «чистые». – *Прим. авт.*

Вернуться

150

Bullock T. H., and Horridge G. A. 1965. Structure and function in the nervous systems of invertebrates. W.H. Freeman and Co., San Francisco.

Вернуться

151

Chen E., Stiefel K. M., Sejnowski T. J., Bullock T. H. Model of Traveling Waves in a Coral Nerve Network, Journal of Comparative Physiology A, 194(2):195–200, 2008.

Вернуться

152

Levine D. S., Grossberg S. «Visual illusions in neural networks: line neutralization, tilt after effect, and angle expansion». J Theor Biol. 1976 Sep 21; 61(2):477–504.

Вернуться

153

Ermentrout G. B., Cowan J. D. «A mathematical theory of visual hallucination patterns». Biol Cybern. 1979 Oct; 34(3):137–50.

Вернуться

154

Hopfield J. J. «Neural networks and physical systems with emergent collective computational abilities», Proceedings of the National Academy of Sciences of the USA, vol. 79 no. 8 pp. 2554–2558, April 1982.

Вернуться

155

Алгоритм Марра – Поджио для стереозрения, упомянутый в главе 5, был симметричной сетью, но поскольку они использовали синхронные обновления всех единиц в сети, динамика была намного сложнее, чем с асинхронными обновлениями в сети Хопфилда. Marr D., Palm G., Poggio T., Analysis of a cooperative stereo algorithm. Biol Cybern. 28(4):223–39, 1978. – *Прим. авт.*

Вернуться

156

Colgin L. L., Leutgeb S., Jezek K., Leutgeb J. K., Moser E. I., McNaughton B.L., Moser M. B. Attractor-map versus autoassociation based attractor dynamics in the hippocampal network». J Neurophysiol. 2010 Jul; 104(1):35–50.

Вернуться

157

Также известна как Bell Laboratories или «Лаборатории Белла», ранее – AT&T Bell Laboratories, Bell Telephone Laboratories. Один из самых знаменитых научно-исследовательских центров по разработке современных технологий в мире. Компания существует с 1925 года. – *Прим. ред.*

Вернуться

158

Hopfield J. J., Tank D. W. «'Neural' computation of decisions in optimization problems». Biol Cybern.; 52(3):141–52. 1985.

Вернуться

159

Ballard D. H., Hinton G. E., Sejnowski T. J. «Parallel Visual Computation», Nature, 306, 21–26, 1983, R. A. Hummel, S. W. Zucker, «On the foundations of relaxation labeling processes», IEEE Trans. Pattern Anal. Machine Intell., 5, 267–287, 1983.

Вернуться

160

Kirkpatrick S.; Gelatt Jr C. D.; Vecchi M. P. (1983). «Optimization by Simulated Annealing». Science. 220 (4598): 671–680.

Вернуться

161

Kienker P. K., Sejnowski T. J., Hinton G. E., Schumacher L. E. “Separating Figure from Ground with a Parallel Network,” Perception, 15, 197–216, 1986.

Вернуться

162

Zhou H., Friedman H. S., von der Heydt R. “Coding of border ownership in monkey visual cortex.” J. Neurosci. 20: 6594–611. 2000.

Вернуться

163

Hebb D. O. (1949). The Organization of Behavior. New York: Wiley & Sons.

Вернуться

164

Ситуация, когда после длительного контакта больного с бредовым расстройством и здорового у здорового начинается аналогичный психоз. – *Прим. ред.*

Вернуться

165

Sejnowski T. J., Kienker P. K., Hinton G. E. «Learning Symmetry Groups with Hidden Units: Beyond the Perceptron», Physica, 22D, 260–275, 1986.

Вернуться

166

Cohen N. J., Abrams I., Harley W. S., Tabor L., Sejnowski T. J. Skill Learning and Repetition Priming in Symmetry Detection: Parallel Studies of Human Subjects and Connectionist Models, 8th Annual Conference of the Cognitive Science Society, Amherst, MA, 1986.

Вернуться

167

Изменение скорости реакции на воздействие раздражителей при постоянном повторе такого воздействия. – *Прим. ред.*

Вернуться

168

Yuhas B. P., Goldstein M. H. Jr., Sejnowski T. J., Jenkins R. E. «Neural network models of sensory integration for improved vowel recognition», Proceedings of the IEEE, 78, 1658–1668, 1990.

Вернуться

169

Hinton G. E., Osindero S. and Teh Y. (2006) «A fast learning algorithm for deep belief nets». Neural Computation 18, pp. 1527–1554.

Вернуться

170

Lettvin J. Y., Maturana H. R., McCulloch W. S., Pitts W. H. 1959, “What the frog's eye tells the frog's brain”, Proc. Inst. Rad. Eng., 47, 1940–1951. hearingbrain.org/docs/letvin_ieee_1959.pdf

Вернуться

171

R. R. Salakhutdinov and G. E. Hinton. “Deep Boltzmann machines”. In Proceedings of the International Conference on Artificial Intelligence and Statistics, volume 12, 2009.

Вернуться

172

B. Poole, S. Lahiri, M. Raghu, J. Sohl-Dickstein and S. Ganguli. “Exponential expressivity in deep neural networks through transient chaos.” In Advances In Neural Information Processing Systems, 3360–3368, 2016.

Вернуться

173

Elman J. L., Bates E. A., Johnson M. H., Karmiloff-Smith A., Parisi D. & Plunkett K. Rethinking innateness: A connectionist perspective on development. MIT Press. (1996).

Вернуться

174

Quartz S. R., Sejnowski T. J. Liars Lovers and Heroes: What the New Brain Science Has Revealed About How We Become Who We Are, New York: Harper-Collins, 2002.

Вернуться

175

Quartz S., Sejnowski T. J. “The neural basis of cognitive development: A constructivist manifesto”, Behavioral and Brain Sciences 20(4), 537–596 (1997).

Вернуться

176

Метилирование ДНК – модификация молекулы ДНК присоединением метильной группы (-CH3) без изменения самой нуклеотидной последовательности. Важный механизм адаптации организма к окружающим условиям. – *Прим. ред.*

Вернуться

177

This is called non-CG methylation. Lister R., Mukamel E. A., Nery J. R., Urich M., Puddifoot C. A., Johnson N. D., Lucero J., Huang Y., Dwork A., Schultz M. D., Tonti-Filippini J., Yu M., Heyn H., Hu S., Wu J. C., Rao A., Esteller M., He C., Haghighi F. G., Sejnowski T. J., Behrens M. M., Ecker J. R. “Global epigenomic reconfiguration during mammalian brain development”, Science, 341, 629, 2013.

Вернуться

178

Кафедра когнитивных наук была основана Доном Норманом, экспертом по человеческим факторам, и имела разнородный факультет. – *Прим. авт.*

Вернуться

179

Математика, используемая в алгоритме обратного распространения обучения, существовала начиная с 1960-х годов в литературе по теории управления, но наибольшее влияние оказало ее применение к многослойным перцептронам: Bryson A. E. and Y.-C. Ho. Applied Optimal Control, Hempshire Publishing Co., New York. (1975) – *Прим. авт.*

Вернуться

180

Michael Jordan gives a Here is a magisterial lecture on modern stochastic gradient descent: simons.berkeley.edu/talks/michael-jordan-2017–5-2

Вернуться

181

Rumelhart D. E., Hinton G. E., Williams R. J. “Learning representations by back-propagating errors”. Nature 323, 533 536. 1986.

Вернуться

182

Однажды Бертран Рассел прочитал публичную лекцию по астрономии. В конце лекции пожилая дама в конце зала встала и сказала: «То, что вы нам рассказали, – чушь. Мир на самом деле – это спина гигантской черепахи». Рассел улыбнулся и ответил: «А на чем стоит черепаха?» «Вы очень умный, молодой человек, очень умный, – сказала старушка, – но я знаю ответ на этот вопрос. На другой черепахе». Старушка решила свою проблему с рекурсией, правда, за счет бесконечной регрессии. На практике цикл должен завершиться. – *Прим. авт.*

Вернуться

183

Вероятно, здесь ошибка автора, потому что в английском языке 26 букв, а не звуков. – *Прим. ред.*

Вернуться

184

В английском слове «cat» буква «c» читается [k]. – *Прим. ред.*

Вернуться

185

W. N. Francis, H. Kucera, “A Standard Corpus of Present-Day Edited American English, for use with Digital Computers.” Brown University, 1964, Revised and Amplified, 1979 clu.uni.no/icame/manuals/BROWN/INDEX.HTM

Вернуться

186

Rosenberg C. R., Sejnowski T. J. “Parallel Networks That Learn to Pronounce English Text”, Complex Systems, 1, 145–168, 1987.

Вернуться

187

Запись того, как звучит сеть на разных этапах обучения, можно скачать по ссылке: papers.cnl.salk.edu/~terry/NETtalk/

Вернуться

188

Seidenberg M. S. & McClelland J. L. (1989). “A distributed developmental model of word recognition and naming”. Psychological Review, 96, 523–568.

Вернуться

189

Rumelhart D. E. & McClelland J. (1986). “On learning the past tense of English verbs”. In Parallel Distributed Processing, Volume 2 (eds. D. E. Rumelhart & J. L. McClelland) Cambridge, Mass: MIT Press; McClelland J. L. & Patterson K. (2002). “Rules or Connections in PastTense inflections: What does the evidence rule out?” Trends in Cognitive Sciences, 6(11), 465–472. Pinker S. & Ulman M. T. (2002). “The past and future of the past tense.” Trends in Cognitive Sciences, 6(11), 456–463.

Вернуться

190

Имеется в виду Past Simple Tense – простое прошедшее время. – *Прим. ред.*

Вернуться

191

Seidenberg M. S. & Plaut D. C. (2014). “Quasiregularity and its discontents: the legacy of the past tense debate.” Cognitive science, 38(6), 1190–1228.

Вернуться

192

Буква «j» обычно дает звук [G], который примерно произносится «дьж», буква «y» может давать несколько звуков – [ai, i, j], – и произносится как «ай», «и» или «й». По правилам фамилия автора должна читаться как Седжновский, но на деле – Сейновски. – *Прим. ред.*

Вернуться

193

Издательство МТИ находится в Кембридже и специализируется на выпуске изданий научного и технического профиля. – *Прим. ред.*

Вернуться

194

Zipser D., Andersen R. A. 1988. “A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons”. Nature 331:679–84.

Вернуться

195

Взрослые с повреждением мозга делают некоторые странные ошибки при чтении слов. Если сеть симулированных нейронов обучена читать, а затем повреждена, у нее поразительно похожее поведение. Geoffrey E. Hinton, David C. Plaut and Tim Shallice, “Simulating Brain Damage”, Scientific American, October 1, 1993. – *Прим. авт.*

Вернуться

196

Сейчас – Fair Isaac Corporation. – *Прим. ред.*

Вернуться

197

Srivastava N., Hinton G., Krizhevsky A., Sutskever I. & Salakhutdinov R., “Dropout: A simple way to prevent neural networks from overfitting”. Journal of Machine Learning Research, 15:1929–1958, 2014.

Вернуться

198

Эпоха – один подход, во время которого нейросети предъявляются все обучающие примеры. – *Прим. ред.*

Вернуться

199

en.wikipedia.org/wiki/Netflix_Prize

Вернуться

200

Carlos A. Gomez-Uribe, Neil Hunt, “The Netflix Recommender System: Algorithms”, Business Value, and Innovation, Journal ACM Transactions on Management Information Systems 6: #13 (2016).

Вернуться

201

Bartol T. M. Jr., Bromer C., Kinney J. P., Chirillo M. A., Bourne J. N., Harris K. M., Sejnowski T. J. “Nanoconnectomic upper bound on the variability of synaptic plasticity”, eLife, 4:e10778, 2015.

Вернуться

202

Если упрощенно, средний уровень «сотня» (0,100) в бейсболе означает 10 процентов успешных ударов. – *Прим. ред.*

Вернуться

203

Это следует из закона больших чисел в теории вероятностей. Вот почему казино всегда выигрывают в долгосрочной перспективе, даже если могут проиграть в краткосрочной. – *Прим. авт.*

Вернуться

204

Силу синапса, то есть его способность передавать возбуждение от одного нейрона к другому, можно измерить в битах. От силы синапсов и их числа зависит общая емкость памяти мозга и, соответственно, точность хранимой в ней информации. – *Прим. ред.*

Вернуться

205

Jasmine Collins, Jascha Sohl-Dickstein, David Sussillo, “Capacity and Trainability in Recurrent Neural Networks”, arXiv:1611.09913 (2016).

Вернуться

206

Опасно придавать слишком большой вес совпадению: «24 часа в сутки, 24 бутылки в ящике. Просто совпадение? Не думаю». Это совпадение отмечается ежегодно в Принстонском университете в день Пола Ньюмана, 24 апреля. – *Прим. авт.*

Вернуться

207

Приблизительную оценку размерности можно найти, взяв квадратный корень произведения нижней и верхней границ (Лоуренс Вайнштейн и Джон А. Адам. «Оценка: Решение мировых проблем на оборотной стороне салфетки для коктейля», Princeton University Press, Princeton, NJ, 2009). Примем за верхнюю границу общее количество синапсов в коре (100 триллионов), а за нижнюю – количество синапсов на одном нейроне (100 тысяч); получается, что оценочное число синапсов, необходимых для представления сложного объекта, примерно миллиард. Применим то же правило, чтобы найти количество необходимых нейронов: верхняя граница – десять миллиардов, число нейронов в коре головного мозга, а нижняя – один нейрон. Таким образом, количество нейронов, необходимых для представления сложного объекта, составляет 100 тысяч, что также верно для количества нейронов меньше квадратного миллиметра коры. Тем не менее они могут быть широко распространены в различных частях коры. Мы можем оценить количество кортикальных областей, которые должны быть связаны, чтобы это представить: верхняя граница – 100, общее число областей коры головного мозга, а нижняя граница – один, поэтому оценка – 10 кортикальных зон, каждая из которых содержит 10

тысяч нейронов. Новые методы, разрабатываемые в рамках правительственной программы BRAIN, будут закреплять эти цифры экспериментально. – *Прим. авт.*

Вернуться

208

Автор немного ошибся в подсчетах. По версии компании Global Language Monitor, которая фиксирует распространение слов английского языка по всему миру, число английских слов уже превысило миллиард. В наиболее современной версии Оксфордского словаря, доступной онлайн, уже более 600 тысяч слов. – *Прим. ред.*

Вернуться

209

Siddhartha Mukherjee A. I., Versus M. D.: “What happens when diagnosis is automated?” New Yorker, April 3, 2017 www.newyorker.com/magazine/2017/04/03/ai-versus-md

Вернуться

210

Daniel Kahneman (2011). Thinking, Fast and Slow. (Farrar, Straus & Giroux, New York).

Вернуться

211

Oakley B., Knafo A., Madhavan G., Wilson D. S., eds (2011) Pathological Altruism (Oxford Univ Press, Oxford).

Вернуться

212

T. S. Kuhn. The Structure of Scientific Revolutions Chicago, IL University of Chicago Press (1970).

Вернуться

213

Pearl J. Probabilistic Reasoning in Intelligent Systems. San Francisco CA: Morgan Kaufmann. 1988.

Вернуться

214

Bengio Y., Lamblin P., Popovici D. & Larochelle H. (2006). “Greedy layer-wise training of deep networks”. In B. Schölkopf, J. Platt & T. Hoffman (Eds.), Advances in neural information processing systems, 19 (pp. 153–160). Cambridge, MA: MIT Press.

Вернуться

215

S. Hochreiter, Y. Bengio, P. Frasconi and J. Schmidhuber. “Gradient flow in recurrent nets: the difficulty of learning long-term dependencies”. In S. C. Kremer and J. F. Kolen, editors, A Field Guide to Dynamical Recurrent Neural Networks. IEEE Press, 2001.

Вернуться

216

D. C. Ciresan, U. Meier, L. M. Gambardella, J. Schmidhuber, «Deep big simple neural nets for handwritten digit recognition», Neural Comput., vol. 22, no. 12, pp. 3207–3220, 2010.

Вернуться

217

G. Hinton, L. Deng, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, B. Kingsbury. «Deep Neural Networks for Acoustic Modeling in Speech Recognition», IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82–97, Nov. 2012.

Вернуться

218

Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”, Advances in Neural Information Processing Systems 25 (NIPS 2012) papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks

Вернуться

219

Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition” arxiv.org/abs/1512.03385.

Вернуться

220

Научно-фантастический фильм Стэнли Кубрика 1968 года по мотивам рассказа Артура Кларка «Часовой». – *Прим. ред.*

Вернуться

221

Yann LeCun, «Modeles connexionnistes de l'apprentissage» (Connecionist learning models), Universite P. et M. Curie (Paris 6) 1987.

Вернуться

222

База данных MNIST (сокращение от Modified National Institute of Standards and Technology) – объемная база данных образцов рукописного написания цифр, США. – *Прим. ред.*

Вернуться

223

Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, Advances in Neural Information Processing Systems 25 (2012) papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks

Вернуться

224

Matthew D. Zeiler, Rob Fergus, “Visualizing and Understanding Convolutional Networks”, arXiv:1311.2901 (2013).

Вернуться

225

Churchland Patricia Smith (1989). Neurophilosophy: Toward a Unified Science of the Mind-Brain. The MIT Press.

Вернуться

226

Churchland P. S. and Sejnowski T. J. The Computational Brain, 2nd edition, Cambridge, MA: MIT Press (2016).

Вернуться

227

Yamins D. L., DiCarlo J. J. “Using goal-driven deep learning models to understand sensory cortex”. Nat Neurosci. 19(3):356–65 (2016).

Вернуться

228

Funahashi S., Bruce C. J., Goldman-Rakic P. S. (1990) “Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms”. J Neurophysiol 63:814–831.

Вернуться

229

Elman J. L. (1990). «Finding Structure in Time». Cognitive Science. 14, 179–211. Jordan M. I. (1986). «Serial Order: A Parallel Distributed Processing Approach» Advances in Psychology 121, 471–495 (1997).

Вернуться

230

Hochreiter S., Schmidhuber J. «Long short-term memory». Neural Computation. 9 (8): 1735–1780 (1997).

Вернуться

231

John Markoff (27 November 2016). “When A. I. Matures, It May Call Jürgen Schmidhuber ‘Dad’ “. The New York Times.

Вернуться

232

Комик Родни Дейнджерфилд известен фразой «Я не получаю никакого уважения!» Редко можно найти кого-то, кто думает, что получает слишком много уважения. – *Прим. авт.*

Вернуться

233

Реже употребляется форма «генеративные состязательные сети». – *Прим. ред.*

Вернуться

234

Goodfellow Ian J., Pouget-Abadie Jean, Mirza Mehdi, Xu Bing, Warde-Farley David, Ozair Sherjil, Courville Aaron C. and Bengio Yoshua. “Generative adversarial nets”. Advances in Neural Information Processing Systems, 2014.

Вернуться

235

Schawinski Kevin; Zhang Ce; Zhang Hantian; Fowler Lucas; Santhanam Gokula Krishnan (201702-01). «Generative Adversarial Networks recover features in astrophysical images of galaxies beyond the deconvolution limit». arXiv:1702.00403, 2017.

Вернуться

236

Jonathan Chang, Stefan Scherer, “Learning Representations of Emotional Speech with Deep Convolutional Generative Adversarial Networks”, arXiv:1705.02394, 2017.

Вернуться

237

Визуальный эффект в компьютерном графике, при котором один объект на изображении плавно меняется на другой. – *Прим. ред.*

Вернуться

238

Nguyen Anh, Yosinski Jason, Bengio Yoshua, Dosovitskiy Alexey and Clune Jeff. “Plug & play generative networks: Conditional iterative generation of images in latent space”. arXiv:1612.00005, 2016.

Вернуться

239

Alec Radford, Luke Metz, Soumith Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks”, arXiv:1511.06434, 2016.

Вернуться

240

Guy Trebay, Miuccia, “Prada and Sylvia Fendi Grapple With the New World”, New York Times, June 19, 2017.

Вернуться

241

Группа данных с присвоенными справочными тегами или выходной информацией. – *Прим. ред.*

Вернуться

242

Poggio T., R. Rifkin S. Mukherjee and P. Niyogi. «General Conditions for Predictivity in Learning Theory», Nature, 428, 419–422, 2004.

Вернуться

243

Бенджио также является советником нескольких компаний, включая Microsoft, и соучредителем компании Element AI, однако его основные связи – в академических кругах, и он приверженец прогресса науки и общественного блага. – *Прим. авт.*

Вернуться

244

Сейчас это глобальная исследовательская организация, базирующаяся в Канаде. – *Прим. ред.*

Вернуться

245

Смотри предисловие в: Churchland P. S. and Sejnowski T. J. The Computational Brain, 2nd edition, Cambridge, MA: MIT Press (2016).

Вернуться

246

Восемнадцать квинтиллионов четыреста сорок шесть квадриллионов семьсот сорок четыре триллиона семьдесят три миллиарда семьсот девять миллионов пятьсот пятьдесят одна тысяча шестьсот пятнадцать. – *Прим. ред.*

Вернуться

247

Судьба изобретателя шахмат неизвестна. – *Прим. авт.*

Вернуться

248

Tesauro G., Sejnowski T. J. A Parallel Network That Learns to Play Backgammon, Artificial Intelligence Journal, 39, 357–390, 1989.

Вернуться

249

Richard Sutton (1988). «Learning to predict by the methods of temporal differences». Machine Learning. 3 (1): 9–44.

Вернуться

250

Richard Bellman's algorithm for dynamic programming. Richard Bellman (1961). Adaptive control processes: a guided tour. Princeton University Press.

Вернуться

251

Ричард Саттон погиб в апреле 2021 года в возрасте 83 лет. – *Прим. ред.*

Вернуться

252

Саттон Р. С., Барто Э. Г. Обучение с подкреплением / Перевод с английского А. А. Слинкина. М.: ДМК-Пресс, 2020.

Вернуться

253

Tesauro Gerald (1995). "Temporal Difference Learning and TD-Gammon". Communications of the ACM 38 (3) 58–68.

Вернуться

254

Полный перебор, или метод «грубой силы» (*англ.* brute force) – метод решения задачи путем перебора всех возможных вариантов. – *Прим. ред.*

Вернуться

255

Ученые, использующие систематический подход к изучению поведения людей и животных. – *Прим. ред.*

Вернуться

256

Garcia J., Kimeldorf D. J., Koelling R. A. “Conditioned aversion to saccharin resulting from exposure to gamma radiation.” Science 1955; 122: 157–8.

Вернуться

257

Montague P. R., Dayan P., Sejnowski T. J. “A Framework for Mesencephalic Dopamine Systems Based on Predictive Hebbian Learning”, Journal of Neuroscience, 16(5), 1936–1947, 1996.

Вернуться

258

Schultz W., Dayan P., Montague P. R. (1997) “A neural substrate of prediction and reward”. Science. 275: 1593–9.

Вернуться

259

Tobler P. N., O’Doherty J. P., Dolan R. J., Schultz W. “Human Neural Learning Depends on Reward Prediction Errors in the Blocking Paradigm”. Journal of Neurophysiology.95(1):301–310. 2006.

Вернуться

260

Hammer M., Menzel R. “Learning and memory in the honeybee”. J Neurosci. 15: 1617–30. 1995.

Вернуться

261

Real L. A. 1991. Animal choice behavior and the evolution of cognitive architecture. Science 253:980–86.

Вернуться

262

Montague P. R., Dayan P., Person C., Sejnowski T. J. Bee foraging in uncertain environments using predictive Hebbian learning. Nature 377: 725–728, 1995.

Вернуться

263

Mischel Walter; Ebbesen Ebbe B. (October 1970). «Attention in delay of gratification». Journal of Personality and Social Psychology. 16 (2): 329–337.

Вернуться

264

Atari – американская компания по производству и изданию компьютерных игр, существует с 1972 года. Pong – серия игровых приставок производства Atari, которая выпускалась с 1975 по 1977 год. – *Прим. ред.*

Вернуться

265

V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, D. Hassabis. “Human-level control through deep reinforcement learning”. Nature 518, 529–533 (2015)

Вернуться

266

Компьютерная игра в жанре стратегии в реальном времени, выпущенная в 1998 году. – *Прим. ред.*

Вернуться

267

Microsoft приобрела права на игру в 2014 году. В 2020 году Minecraft стала самой продаваемой игрой в истории. – *Прим. ред.*

Вернуться

268

Haykin S. Cognitive Dynamic Systems Perception Action Cycle, Radar and Radio; Cambridge University Press: New York, NY, USA, 2012.

Вернуться

269

Haykin S., Fuster J. M., Findlay D., Feng S. Cognitive Risk Control for Physical Systems, IEEE Access (в прессе).

Вернуться

270

Reddy G., Celani A., Sejnowski T. J., Vergassola M. Learning to soar in turbulent environments, Proceedings of the National Academy of Sciences of the United States of America, 113 (33), 2016.

Вернуться

271

Изменение высоты полета за единицу времени. – *Прим. ред.*

Вернуться

272

Doya K., Sejnowski T. J. “A Novel Reinforcement Model of Birdsong Vocalization Learning” In: Tesauro G., Touretzky D. S., Leen T. (Ed.), Advances in Neural Information Processing Systems, 7, MIT Press, Cambridge, MA 101–108, 1995.

Вернуться

273

Doupe A. J., Kuhl P. K. “Birdsong and human speech: common themes and mechanisms”. Annu Rev Neurosci. 22:567–631, 1999.

Вернуться

274

Turrigiano G. (2011). «Too many cooks? Intrinsic and synaptic homeostatic mechanisms in cortical circuit refinement». Annu Rev Neurosci. 34:89–103.

Вернуться

275

Wiskott L., Sejnowski T. J. “Constrained Optimization for Neural Map Formation: A Unifying Framework for Weight Growth and Normalization”, Neural Computation, 10(3), 671–716, 1998.

Вернуться

276

Тело нейрона, в котором находится ядро клетки. – *Прим. ред.*

Вернуться

277

Anthony J. Bell, “Self-organization in real neurons: Anti-Hebb in 'Channel Space'?” Advances in Neural Information Processing Systems 4, 1991; M. Siegel, E. Marder, L. F. Abbott, “Activity-dependent current distributions in model neurons”, Proc Natl Acad Sci U S A. 91: 11308–11312 (1994).

Вернуться

278

«Brains, Minds and Machines».

Вернуться

279

«Bits and Brains».
Вернуться

280

Темы, обсуждаемые на Конференции, доступны по ссылке: nips.cc
Вернуться

281

Чтобы показать особенности этого языка, вот случайное предложение из недавнего обзора в журнале: «Олигодендроциты представляют собой множество белков, ингибирующих рост аксонов, включая миелин-ассоциированный гликопротеин, ингибитор нейрита-выроста «Nogo», олигодендроцит-миелиновый гликопротеин и семафорины». Bireswar Laha, Ben K. Stafford, Andrew D. Huberman, Regenerating optic pathways from the eye to the brain, Science 356:.1031–1034, 2017. – *Прим. авт.*
Вернуться

282

Говард Вахтел, нейробиолог из Университета Колорадо в Боулдере, который изучал нервную систему аплизии. – *Прим. авт.*
Вернуться

283

На 2018 год. 1 Збайт = 10^{21} байт. – *Прим. ред.*
Вернуться

284

Чаще даже в русскоязычной профильной литературе употребляется термин «data science». – *Прим. ред.*

Вернуться

285

Официальный сайт: www.sdss.org

Вернуться

286

Orwell George (1949). Nineteen Eighty-Four. A novel. London: Secker & Warburg. Недавно эта книга обрела новый смысл. – *Прим. авт.*

Вернуться

287

Проект «WiML» был создан в 2006 году и открыл для женщин возможность представлять и продвигать свои исследования в области машинного обучения. Официальный сайт: http://wimlworkshop.org/ – *Прим. авт.*

Вернуться

288

Критический обзор работы психолога-бихевиориста Берреса Фредерика Скиннера «Вербальное поведение». – *Прим. ред.*

Вернуться

289

Здесь ошибка автора: указанная статья Клода Шеннона действительно основополагающая в данной области, но вышла она не в 1982, а в 1948 году. – *Прим. ред.*

Вернуться

290

Сверхбольшие интегральные схемы.

Вернуться

291

Каждый пиксель внутри камеры с таким датчиком работает независимо от других, сообщает об изменениях яркости по мере их возникновения и никак не реагирует, если изменений не было, тогда как в обычной цифровой камере вся картинка фиксируется одномоментно. – *Прим. ред.*

Вернуться

292

Изучение мозга путем развития инновационных нейротехнологий.

Вернуться

293

В марте 2017 года. – *Прим. ред.*

Вернуться

294

На сайте Kaggle миллион специалистов по анализу данных соперничают друг с другом, чтобы получить приз за лучший результат. Cade Metzjune, “Uncle Sam Wants Your Deep Neural Networks”, New York Times, June 22, 2017: https://www.nytimes.com/2017/06/22/technology/homeland-security-artificial-intelligence-neural-network.html?_r=0 – *Прим. авт.*

Вернуться

295

Запись моей лекции смотрите по ссылке: youtu.be/0BDMQuphd-Q – *Прим. авт.*

Вернуться

296

Проект IBM Watson назван в честь первого директора IBM Томаса Уотсона. – *Прим. ред.*

Вернуться

297

В России аналог выходит под названием «Своя игра». – *Прим. ред.*

Вернуться

298

Internet of Things – Интернет вещей – концепция сети передачи данных между физическими объектами («умными вещами»),

которым технологии позволяют взаимодействовать друг с другом или с внешней средой. – *Прим. ред.*

Вернуться

299

В докладе были даны рекомендации и обозначены приоритеты для инновационных технологий, которые помогут продвинуть наше понимание нейронных цепей и поведения: www.braininitiative.nih.gov/2025/ – *Прим. авт.*

Вернуться

300

Kosik K. S., Sejnowski T. J., Raichle M. E., Ciechanover A., Baltimore D. A path toward understanding neurodegeneration, Science, 353, 872–873, 2016.

Вернуться

301

В фильме «Особое мнение» 2002 года герой Тома Круза, скрываясь от правительства, смог избежать ареста при помощи пересадки глаз. – *Прим. авт.*

Вернуться

302

Nilekani N., Shah V. Rebooting India – Realizing a Billion Aspirations. Penguin Books India Pt. Ltd, Gurgaon, Haryana, 2015.

Вернуться

303

“Nandan Nilekani, Infosys, on rebooting India,” Financial Times, Jan 22, 2017 www.ft.com/content/058c4b48-d43c-11e6–9341-7393bb2e1b51?mhq5j=e1

Вернуться

304

Gymrek M.; McGuire A. L.; Golan D.; Halperin E.; Erlich Y. (2013). «Identifying Personal Genomes by Surname Inference». Science. 339 (6117): 321–324.

Вернуться

305

Моторный интеллект отвечает за координацию движений и манипуляции с объектами. Подвижный интеллект – способность с помощью логического мышления и анализа решать задачи, выходящие за пределы предыдущего опыта. – *Прим. ред.*

Вернуться

306

Бактерии прекрасно себя чувствуют в гидротермальных источниках на дне океана при температуре около 400°C. – *Прим. авт.*

Вернуться

307

Wilson Margaret (2002). «Six Views of Embodied Cognition», Psychonomic Bulletin & Review. 9 (4): 625–636

Вернуться

308

Ruvolo P., Messinger D., Movellan J. Infants Time Their Smiles to Make Their Moms Smile, PLoS One. 10(9):e0136492, 2015.

Вернуться

309

Tanaka F., Cicourel A., Movellan J. R. “Socialization between toddlers and robots at an early childhood education center”, Proc Natl Acad Sci U S A. 104:17954–8, 2007.

Вернуться

310

“Conserve elephants. They hold a scientific mirror up to humans”, Economist, Jun 17th 2017 pp. 72–74. www.economist.com/news/science-and-technology/21723394-biology-and-conservationelephants-conserve-elephants-they-hold

Вернуться

311

Brooks R. A. (1990). «Elephants don't play chess». Robotics and Autonomous Systems. 6: 139–159.

Вернуться

312

Робот был собран в Японии, где к его имени и добавили нейтрально-вежливый суффикс – сан. – *Прим. ред.*

Вернуться

313

www.youtube.com/watch?v=knRyDcnUc4U

Вернуться

314

papers.cnl.salk.edu/PDFs/Final%20Report%20To%20NSF%20of%20the%20Planning%20Workshop%20on%20Facial%20Expression%20Understanding%201992-4182.pdf

Вернуться

315

Gottman John, Robert Levenson and Erica Woodin. «Facial expressions during marital conflict». Journal of Family Communication 1.1 (2001): 37–57.

Вернуться

316

Donato F., Stewart-Bartlett M., Hager J. C., Ekman P., Sejnowski T. J. “Classifying Facial Actions”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(10), 974–989, 1999.

Вернуться

317

Набор вычислительных инструментов для распознавания эмоций. В русскоязычной литературе название системы употребляется без перевода. – *Прим. ред.*

Вернуться

318

Система, измеряющая количество аудитории телепрограмм в США для определения их популярности, созданная компанией Nielsen Media Research. – *Прим. ред.*

Вернуться

319

2006 год. – *Прим. ред.*

Вернуться

320

Meltzoff A. N., Kuhl P. K., Movellan J., Sejnowski T. J. “Foundations for a New Science of Learning,” Science, 325: 284–288, 2009.

Вернуться

321

J. Whitehill, Z. Serpell, Y. Lin, A. Foster, J. R. Movellan. “The faces of engagement: automatic recognition of student engagement from facial expressions.” Trans. Affect. Comput. 5(1), 86–98 (2014). Это была командная работа, в которой также принимали участие Гвен Литтлворт, Линда Саламанка, Айша Фостер и Джуди Райли. – *Прим. авт.*

Вернуться

322

Rogowsky B. A., Calhoun B. M. & Tallal P. (2015). Matching learning style to instructional method: Effects on comprehension. Journal of Educational Psychology, 107, 64–78. Вебинар по этой теме смотрите по ссылке: youtu.be/p-WEcSFdoMw – *Прим. авт.*

Вернуться

323

Bloom B. (1984). «The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring», Educational Researcher, 13:6(4–16).

Вернуться

324

Markoff J. Virtual and artificial, but 58,000 want course. New York Times (Aug 15, 2011). <www.nytimes.com/2011/08/16/science/16stanford.html?_r=0>

Вернуться

325

По разным подсчетам число участников онлайн-курсов за 2020–2021 год выросло в 4–10 раз, без учета студентов и школьников, переведенных на дистанционное обучение. – *Прим. ред.*

Вернуться

326

Под таким названием книга была выпущена на русском языке в 2015 году. В оригинале она называлась «A Mind for Numbers: How to Excel at Math and Science (Even If You Flunked Algebra)» – «Разум для чисел: как преуспеть в математике и естественных науках (даже если вы завалили алгебру)». – *Прим. ред.*

Вернуться

327

Курс доступен на официальном сайте: https://www.coursera.org/learn/learning-how-to-learn

Вернуться

328

На 2018 год. – *Прим. ред.*

Вернуться

329

От англ. Mind shift – смена мировоззрения.

Вернуться

330

Оакли Б., Сейновски Т., Макконвилл А. Уроки на отлично! Как научить ребенка заниматься самостоятельно и с удовольствием / перевод с английского А. Поповой. С.-П.: Питер, 2020.

Вернуться

331

Перевернутый класс – современная образовательная модель, при которой учитель дает материал для самостоятельного изучения дома, а на очном занятии происходит практическое закрепление материала. – *Прим. ред.*

Вернуться

332

www.coursera.org/learn/mindshift

Вернуться

333

Oakley B., Mindshift: Break Through Obstacles to Learning and Discover Your Hidden Potential, Penguin Publishing Group, 2017; Оакли Б. Mindshift. Новая жизнь, профессия и карьера в любом возрасте. М.: Прогресс-книга, 2020.

Вернуться

334

https://www.lumosity.com/en/

Вернуться

335

Bavelier D. & Green C. S. (2016). “The Brain Boosting Power of Video Games.” Scientific American, 315(1), 156–176.

Вернуться

336

Greg L. West, Kyoko Konishi, Veronique D. Bohbot, “Video Games and Hippocampus-Dependent Learning,” Current Directions in Psychological Science, 26 (2) 152–158 (2017).

Вернуться

337

Anguera J. A., Boccanfuso J., Rintoul J. L., Al-Hashimi O., Faraji F., Janowich J., Kong E., Larraburo Y., Rolle C., Johnston E. & Gazzaley A. (2013). “Video game training enhances cognitive control in older adults”. Nature, 501 (7465), 97–101.

Вернуться

338

В 2020 году созданная ими игра EndeavorRX была одобрена в США для дополнительной терапии СДВГ у детей в возрасте от 8 до 12 лет. Чтобы получить ее, необходим рецепт от врача. Официальный сайт игры: www.endeavorrx.com – *Прим. ред.*

Вернуться

339

Официальный сайт: www.fastfword.com

Вернуться

340

What Works Clearinghouse. “Beginning Reading intervention report: Fast ForWord”. U.S. Department of Education, Institute of Education Sciences.

Вернуться

341

Подробнее по ссылке: www.brainhq.com
Вернуться

342

Deveau Ozer and Seitz (2014). «Improved Vision and On Field Performance in Baseball through Perceptual Learning», Current Biology, 24(4), R146–7.
Вернуться

343

https://www.ftc.gov/news-events/press-releases/2015/09/ftc-charges-marketers-vision-improvement-app-deceptive-claims
Вернуться

344

Китайская компания, предоставляющая веб-сервисы, и в первую очередь – одноименный поисковик, четвертый по популярности в мире. – *Прим. ред.*
Вернуться

345

www.recode.net/2016/9/17/12943214/sebastian-thrun-self-driving-talent-pool
Вернуться

346

Джефф Дин – личность настолько легендарная, что его называют Чаком Норрисом от программирования и распространяют в сети шутливые факты о его всемогуществе. – *Прим. ред.*

Вернуться

347

В 2021 году США все еще остаются на первом месте, но их, обойдя другие страны, стремительно догоняет Китай. – *Прим. ред.*

Вернуться

348

Джеффри Хинтон – главный научный консультант Vector Institute. Официальный сайт центра: vectorinstitute.ai/ – *Прим. авт.*

Вернуться

349

P. Mozur and John Markoff, “Is China Outsmarting America in A.I.?” New York Times, May 27, 2017 www.nytimes.com/2017/05/27/technology/china-us-ai-artificial-intelligence.html?_r=0

Вернуться

350

Paul Mozur, “Beijing Wants A. I. to Be Made in China by 2030,” New York Times, July 20, 2017 www.nytimes.com/2017/07/20/business/china-artificial-intelligence.html?_r=0

Вернуться

351

С поправкой на инфляцию: www.space.com/17547-jfk-moon-speech-50years-anniversary.html. Выступление президента Джона Кеннеди «Мы решили отправиться на Луну» в Университете Райса в Хьюстоне 12 сентября 1962: www.youtube.com/watch?v=WZyRbnpGyzQ. Когда 20 июля 1969 года астронавт Нил Армстронг ступил на Луну, средний возраст инженеров НАСА составлял 26 лет,

инженеров, которых в школьные годы вдохновила речь Кеннеди. – *Прим. авт.*

Вернуться

352

«Hello, world! – тестовое сообщение учебной программы из классической книги Брайана Кернигана и Денниса Ритчи «Язык программирования C» (1-е издание) (1978) Englewood Cliffs, NJ: Prentice Hall. – *Прим. авт.*

Вернуться

353

Видеозаписи выступлений и дискуссий можно найти по адресу: www.youtube.com/channel/UCB65yIb3bl-cFwOB6j1s8Nw – *Прим. авт.*

Вернуться

354

W. B. Arthur, The Nature of Technology: What it is and How it Evolves, The Free Press, New York (2007).

Вернуться

355

Переменные, которые могут принимать любые значения. – *Прим. ред.*

Вернуться

356

В данном случае понятие «сложность» относится к общей теории систем и характеризует такое поведение системы, при котором у нее появляются свойства, не присущие ее элементам в отдельности. – *Прим. ред.*

Вернуться

357

George A. Cowan, Manhattan Project to the Santa Fe Institute: The Memoirs of George A. Cowan. University of New Mexico Press. (2010).

Вернуться

358

Алгоритм PageRank, изобретенный основателями Google Ларри Пейджем и Сергеем Брином, использует ссылки на веб-страницу для определения важности страниц в Интернете. За прошедшее время он был усовершенствован с помощью многоуровневых алгоритмов, чтобы устранить предвзятость при поиске. – *Прим. авт.*

Вернуться

359

Kramer Adam D. I.; Guillory Jamie E.; Hancock Jeffrey T. (June 17, 2014). «Experimental evidence of massive-scale emotional contagion through social networks». Proceedings of the National Academy of Sciences. 111 (24): 8788–8790.

Вернуться

360

Kauffman Stuart (1993). The Origins of Order: Self Organization and Selection in Evolution. Oxford University Press.

Вернуться

361

Christopher G. Langton. Artificial Life: An Overview. (Editor), MIT Press, 1995.

Вернуться

362

National Research Council. 2007. The Limits of Organic Life in Planetary Systems. Washington, DC: The National Academies Press. Chapter 5: Origin of Life. www.nap.edu/read/11919/chapter/7

Вернуться

363

Мозг человека содержит больше нервных клеток, чем требуется в действительности. Благодаря этому существует нейропластичность – свойство, позволяющее мозгу изменяться по мере накопления опыта и даже восстанавливать утраченные связи после травм. – *Прим. ред.*

Вернуться

364

McCulloch W. and Pitts W. (1943). “A logical calculus of the ideas immanent in nervous activity”. Bulletin of Mathematical Biophysics, 5:115–133.

Вернуться

365

Компьютер назывался JOHNNIAC, в подражание другому раннему цифровому компьютеру под названием ENIAC. – *Прим. авт.*

Вернуться

366

Силлимановская лекция – награда, присуждаемая Йельским университетом с 1901 года, дающая лауреату право провести цикл лекций по естественным наукам, истории или морали, которые затем выпускают в виде книги. – *Прим. ред.*

Вернуться

367

John von Neumann, The Computer and the Brain. New Haven/London: Yale University Press, 1958.

Джон фон Нейман «Вычислительная машина и мозг», перевод Чечина А. М.: АСТ, 2018.

Вернуться

368

Lyne John and Henry Howe «'Punctuated Equilibria': Rhetorical Dynamics of a Scientific Controversy». in Harris R. A. ed. (2007). Landmark Essays on Rhetoric of Science. Mahwah NJ: Hermagoras Press.

Вернуться

369

Holland John (1992). Adaptation in Natural and Artificial Systems. Cambridge, MA: MIT Press.

Вернуться

370

Stiefel K. M., Sejnowski T. J. “Mapping Function onto Neuronal Morphology”, Journal of Neurophysiology, 98: 513–526, 2007.

Вернуться

371

Язык для программирования с одновременным использованием множества парадигм – совокупности идей и понятий, определяющих стиль написания компьютерных программ. – *Прим. ред.*

Вернуться

372

www.wolframalpha.com

Вернуться

373

Интересное эссе о том, как развивалось мышление Вольфрама, можно найти тут: blog.stephenwolfram.com/2017/05/a-new-kind-of-science-a-15-year-view/ – *Прим. авт.*

Вернуться

374

Image identification: blog.stephenwolfram.com/2015/05/wolfram-language-artificial-intelligencethe-image-identification-project/
Вернуться

375

Athenaeum – частный клуб в кампусе Калтеха в Пасадене, основанный в 1930 году. Его членами могут стать преподаватели, аспиранты, старшекурсники, выпускники, попечители и сотрудники института. – *Прим. ред.*
Вернуться

376

Beckman Auditorium – небольшое здание в стиле нового формализма, созданное известным американским архитектором Эдвардом Дуреллом Стоуном в 1960 году. – *Прим. ред.*
Вернуться

377

Он покинул пост корпоративного вице-президента Intel в марте 2020 года. – *Прим. ред.*
Вернуться

378

Сейчас известен как Microsoft Cognitive Toolkit – стандартизированный инструментарий для работы с сетями глубокого обучения и ИИ. – *Прим. ред.*
Вернуться

379

Фонд, поддерживающий развитие науки и общественного понимания. Основан в 2000 году в Лос-Анджелесе норвежским бизнесменом и филантропом Фредом Калви. – *Прим. ред.*

Вернуться

380

Roger Blandford, Michael Roukes, Larry Abbott, Terrence Sejnowski (Chair), “Report on the Third Kavli Futures Symposium on Growing High Performance Computing in a Green Environment,” September 9–11, 2010, Tromø, Norway cnl.salk.edu/Media/Kavli-Futures.Final-Report.11.pdf

Вернуться

381

Экзафлопсными называют компьютеры, которые могут выполнять 10^{18} вычислительных операций в секунду. – *Прим. ред.*

Вернуться

382

10^{15} операций в секунду.

Вернуться

383

10^{15}.

Вернуться

384

Соучредитель корпорации Intel. Свой «закон» он сформулировал в 1965 году. – *Прим ред.*

Вернуться

385

Carver Mead, Lynn Conway, Introduction to VLSI systems, Addison-Wesley, 1980.

Вернуться

386

Mahowald M. A., Mead C. The silicon retina. Scientific American, 264:76–82 (1991).

Вернуться

387

Milton and Francis Clauser Doctoral Prize. Докторская премия имени братьев Клаузеров ежегодно присуждается в Калифорнийском технологическом институте кандидату, чьи исследования демонстрируют высочайшую степень оригинальности и потенциал для открытия новых направлений человеческой мысли. – *Прим. ред.*

Вернуться

388

Mahowald M., Douglas R. A silicon neuron. Nature 354:515–518 (1991).

Вернуться

389

Строительство продолжалось с 1911 по 1987 год. Девять электростанций имеют общую установленную мощность более 1000 мегаватт. – *Прим. ред.*

Вернуться

390

Carver Mead. Collective Electrodynamics: Quantum Foundations of Electromagnetism, MIT Press, 2002.

Вернуться

391

Отец Тоби, Макс Дельбрюк, был биофизиком и в 1950-х годах основал молекулярную биологию; в 1969 году он получил Нобелевскую премию. Это еще одно звено между микроэлектроникой и молекулярной биологией, которое будет рассмотрено в главе 18. – *Прим. авт.*

Вернуться

392

C. Posch, T. Serrano-Gotarredona, B. Linares-Barranco, and T. Delbruck, “Retinomorphic eventbased vision sensors: bioinspired cameras with spiking output,” Proc. IEEE, vol. 102, no. 10, pp. 1470–1484, 2014.

Вернуться

393

Потенциал действия, или спайк, – быстрое колебание мембранного потенциала, возникающее при возбуждении нервных, мышечных, некоторых железистых и растительных клеток. – *Прим. ред.*

Вернуться

394

Sejnowski T. J. The Book of Hebb, Neuron, 24, 773–776, 1999.

Вернуться

395

Обнаружение совпадений (англ. *coincidence detection*) в нейробилогии – процесс, позволяющий нейронам кодировать информацию, выявляя появление близких по времени, но распределенных в пространстве входных сигналов. – *Прим. ред.*

Вернуться

396

Hebb, D.O. (1949). The Organization of Behavior. New York: Wiley & Sons.

Вернуться

397

Robert F. Service. The brain chip. Science, 345: 614–616, 2014, science.sciencemag.org/content/345/6197/614.full

Вернуться

398

В 2017 году.
Вернуться

399

Recurrent spiking networks: Huh D., Sejnowski T. J. “Gradient Descent for Spiking Neural Networks,” arXiv:1706.04698v1.
Вернуться

400

Boahen K. A. “Neuromorph’s Prospectus”, IEEE Xplore: Computing in Science and Engineering, special issue on the topic “Beyond Moore’s Law” 19(2), 14–28, 2017.
Вернуться

401

Soni J., Goodman R. A Mind at Play: How Claude Shannon Invented the Information Age, (Simon & Schuster: New York) (2017).
Вернуться

402

Предел Шеннона (англ. *Shannon limit*) – максимальная скорость передачи, для которой имеется возможность исправить ошибки в канале с заданным соотношением сигнала и шумов. – *Прим. ред.*
Вернуться

403

Golomb S. W. Shift Register Sequences: Secure and Limited-Access Code Generators, Efficiency Code Generators, Prescribed Property Generators, Mathematical Models, World Scientific Publishing Co; 3rd Revised Edition (2017).

Вернуться

404

http://blog.stephenwolfram.com/2016/05/solomon-golomb-19322016/

Вернуться

405

Hedy's Folly: The Life and Breakthrough Inventions of Hedy Lamarr, the Most Beautiful Woman in the World' by Richard Rhodes, Doubleday, 2012.

Вернуться

406

Gödel, Escher, Bach: An Eternal Golden Braid.

Вернуться

407

Riggs L. A.; Ratliff F.; Cornsweet J. C.; Cornsweet T. M. N. (1953). «The Disappearance of Steadily Fixated Visual Test Objects». Journal of the Optical Society of America. 43 (6): 495.

Вернуться

408

Rajesh P. N. Rao & Dana H. Ballard. (1999). “Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects.” Nature Neuroscience 2, 79–87.

Вернуться

409

Осознание этого подвигло Мида на создание нейроморфных систем, которые, как и мозг, работают в режиме реального времени. Mead C. 1990. Neuromorphic electronic systems. Proc. IEEE 78:1629–36. – *Прим. авт.*

Вернуться

410

Так же употребляются названия предсказательное кодирование, предиктивное кодирование, прогностическая обработка. – *Прим. ред.*

Вернуться

411

von Helmholtz Hermann (1867). Handbuch der physiologischen Optik. 3. Leipzig: English translation: Optical Society of America (1924–25): Treatise on Physiological Optics.

Вернуться

412

McClelland J. L. & Rumelhart D. E. “An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic

Findings.” Psychological Review, 88(5) 401–436, 1981; “Part 2. The Contextual Enhancement Effect and Some Tests and Extensions of the Model,” Psychological Review, 89 (1) 60–94, 1982.

Вернуться

413

Muller L., Piantoni S., Koller D., Cash S. S., Halgren E., Sejnowski T. J. “Rotating waves during human sleep spindles organize global patterns of activity during the night,” eLife, e17267 (2016).

Вернуться

414

Соединяющая машина.

Вернуться

415

10^{15}.

Вернуться

416

Выражение означает, что для успеха в долгосрочной перспективе нужно предпринимать необходимые шаги прямо сейчас, а также не жертвовать долгосрочной выгодой ради сиюминутной прибыли. – *Прим. ред.*

Вернуться

417

Известное высказывание Артура Кларка: «Любая достаточно развитая технология неотличима от магии». – *Прим. авт.*

Вернуться

418

Глава адаптирована из: Sejnowski T. J. Consciousness, Daedalus, Winter, 144: 123–132, 2015.

Вернуться

419

Francis H. C. Crick, What Mad Pursuit: A Personal View of Scientific Discovery (New York: Basic Books, 1988).

Вернуться

420

Единого общепринятого научного определения сознания не существует. Однако оно включает в себя состояние бодрствования, способность воспринимать то, что находится или происходит вокруг, а также осознание себя и мира. – *Прим. авт.*

Вернуться

421

Crick F. “The function of the thalamic reticular complex: the searchlight hypothesis.” Proc. Natl. Acad. Sci. USA 81:4586–90, 1984.

Вернуться

422

Francis Crick and Christof Koch, “The Problem of Consciousness,” Scientific American 267 (3) (1992): 10–17; Francis Crick and Christof Koch, “Constraints on Cortical and Thalamic Projections: The NoStrong-Loops Hypothesis,” Nature 391 (1998): 245–250; Francis Crick and Christof Koch, “A Framework for Consciousness,” Nature Neuroscience 6 (2003): 119–126; and Francis Crick, Christof Koch, Gabriel Kreiman, and Itzhak Fried, “Consciousness and Neurosurgery,” Neurosurgery 55 (2) (2004): 273–281.

Вернуться

423

F. Crick and C. F. Koch. “Are We Aware of Neural Activity in Primary Visual Cortex?” Nature 375 (1995): 121–123; Koch C, Massimini M, Boly M & Tononi G (2016) The neural correlates of consciousness: Progress and problems. Nature Review Neuroscience 17: 307–21.

Вернуться

424

Quiroga R. Q., Reddy L., Kreiman G., Koch C. & Fried I. Invariant visual representation by single neurons in the human brain. Nature 435, 1102–1107 (2005).

Вернуться

425

Karl Deisseroth and Mark J. Schnitzer, “Engineering Approaches to Illuminating Brain Structure and Dynamics,” Neuron 80 (2013): 568–577.

Вернуться

426

Valerio Mante, David Sussillo, Krishna V. Shenoy, William T. Newsome, “Context-Dependent Computation by Recurrent Dynamics in Prefrontal Cortex,” Nature 503 (2013): 78–84.

Вернуться

427

Отчет BRAIN 2025 www.nih.gov/science/brain/2025/index.htm

Вернуться

428

Churchland P. S. and Sejnowski T. J. The Computational Brain, 2nd edition, Cambridge, MA: MIT Press (2016).

Вернуться

429

face cells – группа нейронов, которая активнее всего реагирует при распознании лиц. – *Прим. ред.*

Вернуться

430

L. Chang and D. Y. Tsao. “The code for facial identity in the primate brain.” Cell, 169:1013–1028.e14, 2017.

Вернуться

431

David A. Bulkin and Jennifer M. Groh, “Seeing Sounds: Visual and Auditory Interactions in the Brain,” Current Opinion in Neurobiology 16 (2006): 415–419.

Вернуться

432

David M. Eagleman and Terrence J. Sejnowski, “Motion Integration and Postdiction in Visual Awareness,” Science 287 (2000): 2036–2038.

Вернуться

433

Stephen L. Macknik, Susana Martinez-Conde, and Sandra Blakeslee, Sleights of Mind: What the Neuroscience of Magic Reveals About Our Everyday Deceptions (New York: Henry Holt, 2010).

Вернуться

434

Stanislas Dehaene and Jean-Pierre Changeux, “Experimental and Theoretical Approaches to Conscious Processing,” Neuron 70 (2011): 200–227.

Вернуться

435

Moeller S., Crapse T., Chang L. and Tsao D. Y. The effect of face patch microstimulation on perception of faces and objects Nat. Neurosci. 20, 743–752 (2017).

Вернуться

436

Parvizi J., Jacques C., Foster B. L., Withoft N., Rangarajan V., Weiner K. S. and Grill-Spector K. Electrical stimulation of human fusiform face-selective regions distorts face perception. J. Neurosci. 32, 14915–14920 (2012).

Вернуться

437

Sejnowski T. J. What Are the Projective Fields of Cortical Neurons? van Hemmen J. L. Sejnowski T. J. (Ed.), In: 23 Problems in Systems Neuroscience, Oxford University Press, 394–405, 2005.

Вернуться

438

Leanne Chukoskie, Joseph Snider, Michael C. Mozer, Richard J. Krauzlis, and Terrence J. Sejnowski, “Learning Where to Look for a Hidden Target,” Proceedings of the National Academy of Sciences of the United States of America 110 (2013): 10438–10445.

Вернуться

439

Terrence J. Sejnowski, Howard Poizner, Gary Lynch, Sergei Gepshtein, and Ralph J. Greenspan, “Prospective Optimization,” Proceedings of the Institute of Electrical and Electronic Engineering 102 (2014): 799–811.

Вернуться

440

F. C. Crick, C. Koch, “What is the function of the claustrum?” Philos. Trans. R. Soc. Lond. B Biol. Sci., 360 (2005), 1271–1279.

Вернуться

441

Lincoln T. A., Joyce G. F. (February 2009). «Self-sustained replication of an RNA enzyme». Science. 323 (5918): 1229–32.

Вернуться

442

Cech T. R. (Jul 2012). «The RNA worlds in context». Cold Spring Harbor Perspectives in Biology. 4 (7): a006742.

Вернуться

443

Feldman J. A. (2016). Mysteries of visual experience. arxiv.org/abs/1604.08612

Вернуться

444

Patricia S. Churchland, V. S. Ramachandran, and Terrence J. Sejnowski, “A Critique of Pure Vision,” in Large-Scale Neuronal Theories of the Brain, ed. Christof Koch and Joel D. Davis (Cambridge, Mass.: MIT Press, 1994), 23–60.

Вернуться

445

John Allman (1999). Evolving Brains, New York: Scientific American Library.

Вернуться

446

Skinner B. F. (1971). Beyond freedom and dignity. Hackett Publishing Co., Indianapolis, Indiana.

Вернуться

447

Chomsky N., (1971) The Case Against B. F. Skinner The New York Review of Books, 17: 18–24.

Вернуться

448

Специалист, разрабатывающий инструменты для обработки естественного языка, распознавания текста и речи, системы переводов и т. п. – *Прим. ред.*

Вернуться

449

Chomsky N. (1980). Rules and representations. Oxford: Basil Blackwell.

Вернуться

450

Gopnik A., Meltzoff A. & Kuhl P. (2000). The scientist in the crib: What early learning tells us about the mind. New York: HarperCollins. Goodman, E. (2003).

Вернуться

451

Мешок слов (англ. *Bag of Words*) – модель текстов на натуральном языке, в которой каждый документ или текст выглядит как неупорядоченный набор слов без сведений о связях между ними. – *Прим. ред.*

Вернуться

452

Mikolov T., Sutskever I., Chen K., Corrado G. & Dean J. “Distributed representations of words and phrases and their compositionality.” In Proc. Advances in Neural Information Processing Systems 26 3111–3119 (2013).

Вернуться

453

Это было обращение от МТИ к Институту Макговерна, чтобы убедить их сделать существенный вклад в понимание биологической основы языка и языковых нарушений. – *Прим. авт.*

Вернуться

454

«Да пребудет с вами сила», – как говорил мастер-джедай Оби-Ван Кеноби в «Звездных войнах». – *Прим. авт.*

Вернуться

455

Fodor Jerry. «The Mind/Body Problem». Scientific American (244): 124–132.

Вернуться

456

Hassabis D., Kumaran D., Summerfield C., Botvinick M., «Neuroscience-Inspired Artificial Intelligence», Neuron, 95, 245–258 (2017).

Вернуться

457

Neuroeconomics, Second Edition: Decision Making and the Brain: Paul W. Glimcher, Ernst Fehr, Academic Press: New York (2013).

Вернуться

458

Camerer C. Behavior game theory: Experiments in strategic interaction. Princeton University Press. (2003).

Вернуться

459

«Artificial Intelligence Vision: Progress and Non-progress».

Вернуться

460

«Intelligence and Bodies».
Вернуться

461

«Why Natural Language Processing is Now Statistical Natural Language Processing».
Вернуться

462

Синтия Бризил в МТИ и Хавьер Мовеллан разработали социальных роботов, которые взаимодействуют с людьми и используют выражение лица для общения, делая первые шаги к вычислительной теории эмоций. – *Прим. авт.*
Вернуться

463

«Theory of Neural-Analog Reinforcement Systems and Its Application to the Brain Model Problem».
Вернуться

464

Wolfram S., Farewell Marvin Minsky (1927–2016), Idea Makers Personal Perspectives on the Lives & Ideas of Some Notable People, blog.stephenwolfram.com/2016/01/farewell-marvin-minsky-19272016/
Вернуться

465

Graves A., Wayne G., Reynolds M., Harley T., Danihelka I., Grabska-Barwińska A., Colmenarejo S. G., Grefenstette E., Ramalho T., Agapiou J., Badia A. P., Hermann K. M., Zwols Y., Ostrovski G., Cain A., King H., Summerfield C., Blunsom P., Kavukcuoglu K., Hassabis D. Hybrid computing using a neural network with dynamic external memory. Nature. 2016 538: 471–476.

Вернуться

466

Организмы, используемые в качестве моделей для изучения тех или иных свойств, процессов или явлений живой природы. Полученные данные переносятся на другие организмы. – *Прим. ред.*

Вернуться

467

www.nobelprize.org/nobel_prizes/medicine/laureates/2002/brenner-lecture.html

Вернуться

468

thesciencenetwork.org/search?program=Reading+the+Human+Genome+with+Sydney+Brenner

Вернуться

469

Имагинальные диски – зачатки ног и усиков у мух. – *Прим. авт.*

Вернуться

470

Сидней Бреннер опубликовал свою оригинальную историю в Brenner S. «Francisco Crick in Paradiso», Curr Biol. 6(9):1202, 1996: «В Кембридже я делил кабинет с Фрэнсисом Криком в течение двадцати лет. Одно время он интересовался эмбриологией и много думал об имагинальных дисках у мушки-дрозофилы. Однажды он бросил на стол книгу, которую читал, с раздраженным криком: «Бог знает, как работают эти имагинальные диски!» В мгновение ока я увидел, как Крик прибывает на небеса, а Петр приветствует его словами: «О, доктор Крик, вы, должно быть, устали после долгого путешествия. Садитесь, выпейте и расслабьтесь». «Нет, – говорит Крик, – я должен увидеть Бога и задать ему вопрос». После некоторых уговоров ангел соглашается отвести Фрэнсиса к Богу. Они пересекают среднюю часть небес и наконец, перейдя железнодорожные пути, попадают в заваленный хламом сарай с гофрированной железной крышей, в глубине которого возился маленький человечек в комбинезоне с большим гаечным ключом в заднем кармане. «Бог, – произносит ангел, – это доктор Крик. Доктор Крик, это Бог». «Я так рад с вами познакомиться, – говорит Крик. – Я должен задать вам этот вопрос. Как работают имагинальные диски?» «Ну, – приходит ответ, – мы взяли немного этого материала и добавили к нему кое-что еще… на самом деле, мы не знаем, но я могу сказать вам, что мы создавали здесь мух на протяжении двухсот миллионов лет, и никто ни разу не жаловался». – *Прим. авт.*

Вернуться

471

Сидней Бреннер умер в апреле 2019 года. – *Прим. ред.*

Вернуться

472

Dobzhansky, Theodosius (March 1973), «Nothing in Biology Makes Sense Except in the Light of Evolution», American Biology Teacher, 35 (3): 125–129, JSTOR 4444260; reprinted in Zetterberg, J. Peter, ed. (1983), Evolution versus Creationism, Phoenix, Arizona: ORYX Press.

Вернуться

473

www.youtube.com/watch?v=C9M5h_tVlc8

Вернуться

474

www.youtube.com/watch?v=L9ITpz4OeOo

Вернуться

475

Nagel, Thomas (1974). «What Is It Like to Be a Bat?» The Philosophical Review. 83 (4): 435–450.

Вернуться

476

Антропогенез изучает происхождение человечества. Официальный сайт: carta.anthropogeny.org – *Прим. авт.*

Вернуться

477

Группа в Ла-Хойя и CARTA были основаны под руководством Аджита Варки, врача-ученого в Калифорнийском университете в Сан-Диего, увлеченного эволюцией. – *Прим. авт.*

Вернуться

478

Около 1 % ДНК – это последовательности, кодирующие белки, и 8 % – регуляторные последовательности, связывающие белки. – *Прим. авт.*

Вернуться

479

Berg Howard C. (2003). E. coli in motion. New York, NY: Springer.

Вернуться

480

Navlakha S. & Bar-Joseph Z. Algorithms in nature: the coka: nvergence of systems biology and computational thinking. Molecular Systems Biology 7, 546 (2011).

Вернуться

481

Отсылка к выражению «тонкий край клина» (англ. *thin edge of the wedge*) – незначительное изменение, с которого начинается большое развитие. – *Прим. ред.*

Вернуться

482

Churchland P. S. and Sejnowski T. J. The Computational Brain, 2nd edition, Cambridge, MA: MIT Press (2016). https://mitpress.mit.edu/books/computational-brain

Вернуться

483

Quartz S. R., Sejnowski T. J. Liars, Lovers and Heroes: What the New Brain Science Has Revealed About How We Become Who We Are, New York: Harper-Collins, 2002.

Вернуться

484

Mackintosh N. J. IQ and Human Intelligence (second Ed). Oxford: Oxford University Press. (2012).

Вернуться

485

Griffin D. R. Animal minds. Chicago: University of Chicago Press (1992).

Вернуться

486

Бишоп К. М. Распознавание образов и машинное обучение / Перевод Д. А. Клюшина. М.: Вильямс, 2020.

Вернуться

487

MacKay D. Information Theory, Inference, and Learning Algorithms, Cambridge: Cambridge University Press, 2003.

Вернуться

488

Паттерсон Дж., Гибсон А. Глубокое обучение с точки зрения практика / Перевод А. А. Слинкина. М.: ДМК Пресс, 2018.

Вернуться

489

Бенджио И., Курвилль А., Гудфеллоу Я. Глубокое обучение / Перевод А. А. Слинкина. М.: ДМК Пресс, 2018.

Вернуться

490

Murphy K. P. Machine Learning: A Probabilistic Perspective, Cambridge, MA: MIT Press (2014).

Вернуться

491

Саттон Р., Барто Э. Обучение с подкреплением / Перевод А. А. Слинкина. М.: ДМК Пресс, 2020.

Вернуться

492

Goldhagen Sarah Williams. Louis Kahn’s Situated Modernism. New Haven, CT: Yale University Press, 2001.

Вернуться

493

Francis Crick. The Astonishing Hypothesis: The Scientific Search for the Soul. Chicago: Charles Scribner's Sons, 1995. P. 146.

Вернуться