ЯН ЛЕКУН

КАК УЧИТСЯ МАШИНА

Революция в области
нейронных сетей
и глубокого обучения

PRO

Ян Лекун

Как учится машина. Революция в области нейронных сетей и глубокого обучения

«Альпина Диджитал»

2019

УДК 004.8
ББК 32.813

Лекун Я.

Как учится машина. Революция в области нейронных сетей и глубокого обучения / Я. Лекун — «Альпина Диджитал», 2019

ISBN 978-5-90-747055-2

Мы живем во время революции, еще 50 лет назад казавшейся невероятной, – революции в области умных машин, которые теперь обучаются самостоятельно, вместо того чтобы просто выполнять запрограммированные команды. И возможности таких машин огромны: распознавание изображений, лиц и голосов, переводы на сотни языков, беспилотное управление автомобилями, обнаружение опухолей на медицинских снимках и многое другое. Автор книги Ян Лекун стоит у истоков этой революции. Лауреат премии Тьюринга, профессор Нью-Йоркского университета и руководитель фундаментальными исследованиями в Facebook, он является одним из изобретателей глубокого обучения, применяемого к так называемым искусственным нейронным сетям, архитектура и функционирование которых вдохновлены устройством человеческого мозга. В своей книге он, не прибегая к метафорам, делится своим научным подходом на стыке компьютерных наук и нейробиологии, проливая свет на будущее искусственного интеллекта, связанные с ним проблемы и перспективы. Сегодня искусственный интеллект действительно меняет все наше общество. Эта понятная и доступная книга перенесет вас в самое сердце машины, открывая новый увлекательный мир, который уже является нашей реальностью.

УДК 004.8
ББК 32.813

ISBN 978-5-90-747055-2

Содержание

Ян Лекун
Как учится машина. Революция в области нейронных сетей и глубокого обучения

Перевод *Е. Арсеновой*
Редактор *В. Скворцов*
Научный редактор *М. Плец*
Руководители проекта *А. Марченкова, Ю. Семенова*
Дизайн *А. Маркович*
Корректор *Е. Жукова*
Компьютерная верстка *Б. Руссо*

* * *

Введение

«Открой дверь модульного отсека, Хэл!» В фильме «2001 год: Космическая одиссея» HAL 9000, сверхразумный компьютер, управляющий работой космического корабля, отказывается открыть дверь модульного отсека астронавту Дейву Боумену. В этой драматической сцене – вся трагедия искусственного интеллекта. Мыслящая машина оборачивается против человека, который ее сам же разработал. Что это: фантазия или обоснованные опасения? Стоит ли тревожиться о том, что однажды нашим миром будут управлять терминаторы – искусственные гуманоиды с почти неограниченными возможностями и темными замыслами? Этот вопрос люди задают все чаще и чаще сейчас, когда мы переживаем неслыханную революцию в интеллектуальных технологиях, которую никто не мог вообразить себе еще полвека назад. Искусственный интеллект, изучению которого я посвятил много лет, меняет все наше общество.

Я решил написать эту книгу, чтобы объяснить определенный набор методов и приемов в этой области, не скрывая всей ее сложности. Понять это не так просто, как научиться играть в шашки, но я думаю, что это необходимо для формирования аргументированного мнения по вопросам, связанным с искусственным интеллектом. Наше медиапространство пестрит такими терминами как «глубокое обучение», «машинное обучение» или «нейронные сети» … Я хочу, шаг за шагом, пролить свет на научный подход, который работает на стыке вычислительной техники и нейробиологии, не прибегая при этом к каким-либо метафорам.

Во время нашего погружения в основы работы вычислительных машин я буду использовать два способа изложения информации. Первый из них – традиционный: я рассказываю, описываю и анализирую. Время от времени для тех, кому интересно, я буду приводить более сложные примеры из математики и компьютерных наук.

Искусственный интеллект (ИИ) позволяет машине распознавать изображения, транскрибировать голос с одного языка на другой, переводить тексты, автоматизировать управление автомобилем или контролировать производственные процессы. Его широкое распространение в последние годы связано с методом, именуемым глубоким обучением, которое позволяет не просто программировать машину для выполнения определенной задачи, а обучать ее решению более широкого круга сходных задач. Глубокое обучение применяется к так называемым искусственным нейронным сетям, архитектура и функционирование которых вдохновлены устройством человеческого мозга.

Наш мозг состоит из 86 миллиардов нейронов, нервных клеток, связанных друг с другом. Искусственные нейронные сети также состоят из множества единиц, математических функций, подобных очень упрощенным нейронам. В мозгу обучение изменяет связи между нейронами; то же самое происходит и с искусственными нейронными сетями. Поскольку эти единицы часто организованы в несколько слоев, мы говорим о «сетях» и «глубоком» обучении.

Роль искусственных нейронов состоит в том, чтобы вычислить взвешенную сумму входных сигналов и создать выходной сигнал, если эта сумма превышает определенный порог. Но искусственный нейрон – это не больше и не меньше, чем математическая функция, рассчитанная компьютерной программой. Однако мы не случайно применяем к искусственным сетям те же термины, что и к реальным нейронам, – ведь именно открытия в области нейробиологии послужили стимулом исследованиям в области ИИ.

В этой книге я также хочу проследить свой интеллектуальный путь в рамках этого необычного научного приключения. Мое имя по-прежнему связано с так называемыми «сверточными» нейронными сетями, которые подняли распознавание объектов компьютером на небывалую высоту. Вдохновленные структурой и функцией зрительной коры головного мозга млекопитающих, они могут эффективно обрабатывать изображения, видео, звук, голос, текст и другие типы сигналов.

В чем состоит деятельность исследователя? Откуда берутся его идеи? Что касается меня, то я уделяю много внимания интуитивным догадкам. Дальше наступает очередь математики. Я знаю, что другие ученые действуют диаметрально противоположным образом. Я проецирую в свою голову пограничные случаи, которые Эйнштейн называл «мысленными экспериментами», благодаря которым вы сначала представляете ситуацию, а затем пытаетесь рассмотреть ее следствия для лучшего понимания проблемы.

Моя интуиция подпитывается чтением книг. Я просто пожираю книги. Я исследую работы тех, кто был до меня. Вы никогда ничего не создадите в одиночку. Идеи живут, дремлют, и они возникают в чьей-то голове, потому что пришло время. Так рождаются исследования. Они продвигаются неравномерно, то прыжками, то шажками, а порой – даже пятясь. Но деятельность эта всегда коллективна. Образ одинокого исследователя, делающего в своей лаборатории мировое открытия, – не более, чем романтическая фантастика.

Путь разработки глубокого обучения не был простым. Приходилось бороться со скептиками всех мастей. Сторонники «классического» искусственного интеллекта, основанного исключительно на логике и рукописных программах, пророчили нам провал. Люди, добившиеся успеха в традиционном машинном обучении, показывали на нас пальцами, хотя глубокое обучение, над которым мы работали, и было по существу набором определенных методов в более широкой области машинного обучения. Однако тот тип машинного обучения, который позволял машине решать задачу путем сравнения конкретных примеров внутри массива данных, а не прямым исполнением написанной программы, тоже имел свои пределы. Мы пытались их преодолеть. Средством для этого послужили глубокие нейронные сети. Они были очень эффективными, но при этом сложными в математическом анализе и в реализации. Поэтому мы прослыли чуть ли не алхимиками…

Сторонники традиционного машинного обучения перестали высмеивать нейронные сети в 2010 г., когда последние наконец продемонстрировали свою эффективность. Лично я никогда не сомневался в успехе. Я всегда был убежден, что человеческий интеллект настолько сложен, что для того, чтобы его скопировать, нужно стремиться построить самоорганизующуюся систему, способную учиться самостоятельно, через опыт.

Сегодня эта форма искусственного интеллекта так и осталась наиболее перспективной, благодаря доступности больших баз данных и прогрессу в разработке оборудования, например графических процессоров, намного увеличивших вычислительную мощность компьютеров.

По окончании учебы я планировал провести несколько лет в Северной Америке. И я все еще там! После некоторых жизненных перипетий я попал в компанию Facebook, владеющую сайтом с 2 миллиардами активных пользователей, чтобы вести фундаментальные исследования в области ИИ. Это – тоже часть моей публичной биографии. Я не хочу скрывать ничего из того, что происходит в компании Марка Цукерберга, которой в 2018 г. были предъявлены серьезные обвинения, и чье безграничное расширение вызывает опасение. В любом случае – я сторонник открытости.

В марте 2019 г. я был удостоен премии Тьюринга за 2018 г. от Ассоциации вычислительной техники – своего рода Нобелевской премии в компьютерной области. Я разделил эту награду с двумя другими специалистами по глубокому обучению, Йошуа Бенджио и Джеффри Хинтоном, моими партнерами, с которыми мы много спорили, но всегда сходились в главном.

Я многим обязан всем этим встречам, месту, которое я со временем занял в сообществе безумных наследников кибернетики 1950-х гг., не устававшим задавать друг другу «детские» на вид, но глубокие по сути вопросы, вроде: «Как получается, что нейроны, очень простые объекты, соединяясь друг с другом, производят новое свойство, которое называется интеллектом?»

Теперь эта научная авантюра порождает новые важные вопросы. Отличается ли работа машины, которая распознает автомобиль посредством выделения таких элементов, как колеса,

лобовое стекло и т. д. от работы нашей зрительной коры при идентификации той же самой машины? Что делать с наблюдаемым сходством между работой машины и мозгом человека или животного? Область исследования безгранична.

Посмотрим правде в глаза: машины, какими бы мощными и сложными они ни были, по-прежнему очень узкоспециализированы. Они учатся менее эффективно, чем люди и животные. По сей день у них нет ни здравого смысла, ни совести. По крайней мере, пока! Несомненно, они превосходят людей в определенных задачах: например, побеждают их в го и в шахматах; они переводят сотни языков, они узнают растения или насекомых, они обнаруживают опухоли на медицинских изображениях. Но человеческий мозг сохраняет значительное преимущество перед машинами в том, что он более универсален и гибок.

Смогут ли машины догнать нас, и если да – то как скоро?

Глава 1
Революция в искусственном интеллекте

Искусственный интеллект проникает во все секторы экономики, связи, здравоохранения и даже транспорта – благодаря созданию беспилотных автомобилей… Многие наблюдатели говорят уже не о технологической эволюции, а о революции.

Вездесущий искусственный интеллект

«Алекса, какая погода в Буэнос-Айресе?» Менее чем за секунду «умная» акустическая система записывает вопрос, передает его через домашний Wi-Fi на серверы Amazon, которые транскрибируют и интерпретируют его. Затем они получают информацию от метеорологической службы и возвращают ответ, который Алекса озвучивает приятным голосом: «В настоящее время в Буэнос-Айресе, Аргентина, температура воздуха 22 °C. Пасмурно».

В офисе ИИ – прилежный помощник. Он работает быстро, и его не пугают повторяющиеся задачи. Он может просмотреть миллионы записей в базе денных в поисках цитаты и найти нужную за долю секунды благодаря возможностям современных компьютеров, скорость вычислений у которых сделалась почти невероятной.

Один из первых программируемых электронных компьютеров, ENIAC, построенный в 1945 г. в Университете Пенсильвании для расчета траектории полета снарядов, выполнял приблизительно 360 умножений в секунду для десятизначных цифр. Сейчас эта машина выглядит неповоротливым доисторическим чудищем. Процессоры нынешних персональных компьютеров в миллиард раз быстрее. Они имеют производительность в сотни гигафлопс[1]. Графические процессоры, используемые нашими компьютерами для визуализации, имеют производительность в несколько десятков терафлопс. Гигантские числа с впечатляющими названиями.

Их уже не остановить! Современные суперкомпьютеры объединяют десятки тысяч этих графических процессоров и достигают скорости в сотни тысяч терафлопс, производя колоссальные объемы вычислений для всевозможных симуляций: прогнозирования погоды, моделирования климата, расчета воздушного потока вокруг самолета или конформации белка, моделирования таких головокружительных событий, как первые мгновения существования Вселенной, смерть звезды, эволюция галактик, столкновения элементарных частиц или ядерный взрыв.

Такие симуляции включают численное решение дифференциальных уравнений или уравнений в частных производных – это задача, которую в прошлом математикам приходилось решать вручную. И все же – так ли умны эти вычислительные гиганты, как математики прошлых лет? Нет, конечно… во всяком случае, пока. Одна из задач развития искусственного интеллекта заключается в том, чтобы когда-нибудь научить машины использовать их огромную вычислительную мощность для решения интеллектуальных задач, сейчас подвластных только животным и людям.

Не стоит судить только по внешности. Программы искусственного интеллекта умеют хорошо учиться, но лишь до определенного момента. В 2017 г. робот Норико Араи – специалиста Токийского университета, изучавшего влияние ИТ на общество, успешно сдал вступительный экзамен в один из японских университетов. Программа, получившая название Todai (название университета), сдала эссе, экзамены по математике и английскому языку лучше 80 % абитуриентов. Но это не значит, что система была умной: на самом деле робот вообще не понимал, что он пишет! Успех программы скорее говорит о несовершенстве как вступительных испытаний в высшие успешные заведения Японии, так и машинного интеллекта. Мы были бы рады, если, в конечном счете, Todai провалится на более продуманно организованных экзаменах.

[1] FLOPS (обозначается также как flops, flop/s, произносится по-русски как «флопс») – акроним от англ. Floating-point Operations Per Second (число операций с плавающей точкой в секунду). Представляет собой внесистемную единицу измерения производительности компьютеров. Правописание и склонение термин в русском языке еще не устоялось: иногда пишут «флоп», иногда «флопс». – *Прим. ред.*

Искусственный интеллект в искусстве

Искусственный интеллект может быть хорошим художником, точнее – копиистом. Он мастерски создает произведения «в стиле». Он способен превратить любую фотографию в картину Моне, сделать из зимнего пейзажа весенний[2] или заменить лошадь зеброй на видео. Остерегайтесь фальшивых видеороликов! Более того: в 2017 г. команда Ахмеда Эльгаммала из Университета Рутгерса (США) разработала систему, создающую картины, так похожие на оригинальные творения некоторых художников, что эксперты не могут отличить их от реальных картин тех же художников[3].

Музыка – тоже не исключение. Многие исследователи используют методы искусственного интеллекта для синтеза звука и музыкальной композиции. Так 21 марта 2019 г. в честь дня рождения Иоганна Себастьяна Баха связанный с Google[4] проект Magenta анонсировал новую систему, названную Google Doodle[5], которая позволяет любому пользователю сочинять и аранжировать мелодию в стиле известного композитора. Вспомним также, что 4 февраля 2019 г. в Лондонском концертном зале Cadogan Hall 66 музыкантов английского оркестра «English Session Orchestra» публично исполнили новую версию Симфонии № 8 Шуберта, известную как «неоконченная». В тот вечер она стала законченной. После анализа двух существующих частей система искусственного интеллекта, разработанная компанией Huawei, создала две недостающие части музыкального произведения (впрочем, композитору Лукасу Кантору все-таки пришлось написать партитуру для оркестра).

[2] https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix.

[3] https://arxiv.org/abs/1706.07068.

[4] https://ai.google/research/teams/brain/magenta/.

[5] https://www.google.com/doodles/celebrating-johann-sebastian-bach.

Гуманоиды? Ничего подобного!

София – красивая девушка со стрижкой «под ноль», загадочной улыбкой и кристально-чистыми глазами, стала настоящей звездой 2017 г. «Вы смотрите слишком много голливудских фильмов!» – усмехается она в ответ журналисту, который выразил беспокойство о том, что однажды роботы захватят Землю. Ее появление вызвало такой общественный резонанс, что в том же году Саудовская Аравия предоставила ей гражданство. На самом деле это прекрасное существо было всего лишь марионеткой, для которой программисты написали набор стандартных ответов на самые разные вопросы. То, что ей говорят, обрабатывается системой сопоставления, которая выбирает из каталога возможных ответов наиболее подходящий к данному случаю.

Софья обманывает свою аудиторию. Она не умнее, чем Todai в Токио. Просто у нее более выраженная и живая мимика, и мы, люди, впечатленные этой анимацией, верим в наличие у человекообразного робота настоящего ума.

«Старый добрый» искусственный интеллект…

Мир искусственного интеллекта меняется. Его границы постоянно расширяются. Когда та или иная проблема решена, она покидает сферу искусственного интеллекта и постепенно переходит в классический набор инструментов.

Например, преобразование математических формул в инструкции, выполняемые компьютером, было делом искусственного интеллекта лишь в 1950-х гг., когда информатика только зарождалась как наука. Сегодня – это обычная функция компиляторов – программ, которые преобразуют программы, написанные инженерами, в наборы инструкций, которые непосредственно исполняются машиной. А компиляция как предмет преподается всем студентам-компьютерщикам.

Или возьмем задачу поиска маршрута. В 1960-е она явно принадлежала к области ИИ. Сегодня – это опять всего лишь стандартный инструмент, каких много. Существуют эффективные алгоритмы поиска кратчайшего пути в графе (сети взаимосвязанных узлов), такие как алгоритм Дейкстры 1959 г.[6] или алгоритм A* («А звезда») Харта, Нильссона и Рафаэля 1969 г.[7] С повсеместным распространением GPS эта задача далеко отошла от переднего края науки.

Ядром искусственного интеллекта в 1970-х и 1980-х гг. был набор методов автоматического рассуждения, основанных на логике и манипулировании символами. С некоторой издевкой англоязычные сторонники этого подхода называют подобные системы GOFAI (аббревиатурой от Good Old-Fashioned Artificial Intelligence, то есть «старый добрый искусственный интеллект»).

Перейдем теперь к экспертным системам, где машина вывода применяет правила к фактам и выводит из них новые факты. В 1975 г. MYCIN, например, должен был помочь врачам выявлять острые инфекции, такие как менингит, и назначать лечение антибиотиками. В нем было около 600 правил вроде: «ЕСЛИ инфекционный организм грамотрицательный, И организм палочковидный, И организм анаэробный, ТО этот организм является (с вероятностью 60 %) бактерией».

Для своего времени MYCIN была инновационной системой. Ее правила включали в том числе и факторы достоверности, которые система объединяла для получения оценки общей достоверности результата. Она был оборудована так называемым механизмом вывода с «обратной связью», посредством которого система выдвигала одну или несколько диагностических гипотез и расспрашивала практикующего врача о симптомах пациента. В процессе этого система уточняла гипотезы в зависимости от ответов и ставила диагноз, и, наконец, назначала антибиотик и дозировку (с той или иной степенью уверенности).

Чтобы создать такую систему, врач-эксперт должен был сидеть рядом с инженером и подробно рассказывать инженеру о своих рассуждениях. Как он диагностирует аппендицит или менингит? Какие при этом бывают симптомы? Так у MYCIN появлялись правила. Иначе говоря, если у пациента имеется определенный симптом, то существует такая-то вероятность аппендицита, такая-то вероятность непроходимости кишечника и такая-то вероятность почечной колики. Инженер вручную записывал эти правила в базу.

Надежность MYCIN (и ее преемников) была довольно неплохой. Но они так и не вышли за рамки эксперимента. Компьютеризация в медицине тогда только зарождалась, а процесс ввода данных был утомительным. В конце концов, все эти экспертные системы, основанные на логике и деревьях поиска, оказались слишком тяжелыми и сложными в разработке. Они

[6] https://fr.wikipedia.org/wiki/Algorithme_de_Dijkstra.

[7] https://fr.wikipedia.org/wiki/Algorithme_A*.

вышли из употребления, но остаются эталоном и продолжают описываться в учебниках по искусственному интеллекту.

Тем не менее работа над логикой привела к появлению некоторых важных приложений: к символьному решению уравнений и исчислению интегралов в математике, а также к автоматической проверке программ. Например, с его помощью компания Airbus проверяет точность и надежность своего программного обеспечения для управления пассажирскими самолетами.

Часть исследовательского сообщества ИИ продолжает работать над этими темами. Другие специалисты, к которым принадлежу и я, посвятили себя совершенно другим подходам, основанным на машинном обучении.

… или же машинное обучение?

Рассуждения – это лишь малая часть человеческого разума. Мы часто думаем по аналогии, мы действуем интуитивно, опираясь на представления о мире, постепенно приобретаемые через опыт. Восприятие, интуиция, опыт, наборы усвоенных навыков – все это результат обучения.

В таких условиях, если мы хотим построить машину, интеллект которой будет приближен к человеческому, мы должны сделать ее тоже способной к обучению. Человеческий мозг состоит из 86 млрд взаимосвязанных нейронов (или нервных клеток), 16 млрд из которых находятся в коре головного мозга. В среднем каждый нейрон образует почти 2000 других соединений с другими нейронами – так называемых синапсов. Обучение происходит путем создания синапсов, удаления синапсов или изменения их эффективности. Поэтому, используя самый известный подход к машинному обучению, мы создаем искусственные нейронные сети, процесс обучения которых изменяет связи между нейронами. Приведем несколько общих принципов.

Машинное обучение включает первый этап обучения или тренировки, когда машина постепенно «учится» выполнять задачу, и второй этап – реализацию – когда машина закончила обучение.

Чтобы научить машину определять, содержит ли изображение автомобиль или самолет, мы должны начать с представления ей тысяч изображений, содержащих самолет или автомобиль. Каждый раз, когда на входе системе дается изображение, нейронная сеть (или «нейросеть»), состоящая из соединенных между собой искусственных нейронов (в действительности – это множество математических функций, вычисляемых компьютером), обрабатывает это изображение и выдает выходной ответ. Если ответ правильный, мы ничего не делаем и переходим к следующему изображению. Если ответ неверный, мы немного корректируем внутренние параметры машины, то есть силу связей между нейронами, чтобы ее выходной сигнал приближался к желаемому ответу. Со временем система настраивается и в конечном итоге сможет распознать любой объект, будь то изображение, которое она видела ранее, или любое другое. Это называется способностью к обобщению.

Сравнение этой способности нейросетей с соответствующими функциями мозга говорит о том, что машинам до нас еще далеко. Приведем некоторые цифры. В мозге $8{,}6 \cdot 10^{10}$ нейронов, связанных между собой примерно $1{,}5 \cdot 10^{14}$ синапсами. Каждый синапс может выполнять «вычисление» сотни раз в секунду. Данный синаптический расчет эквивалентен сотне цифровых операций на компьютере (умножение, сложение и т. д.) или $1{,}5 \cdot 10^{18}$ операций в секунду для всего мозга, хотя в действительности лишь часть нейронов активна в любой момент времени. Для сравнения, карта GPU (графического процессора) может выполнять 10^{13} операций в секунду. Чтобы приблизиться к мощности мозга, потребуется 100 000 видеокарт. И тут есть одна загвоздка: человеческий мозг потребляет мощность, эквивалентную 25 Вт. Карта с одним GPU (графического процессора) потребляет в десять раз больше, т. е. 250 Вт. Электронные цепи в миллион раз менее эффективны, чем биологические.

Коктейль из старого и нового

Сегодняшние приложения, как правило, используют сочетание машинного обучения, GOFAI и классических вычислений. Рассмотрим машину, способную управлять автомобилем без водителя. Бортовая система визуального распознавания, обученная распознавать визуальные объекты и сигналы, присутствующие на дороге, использует определенную архитектуру нейронной сети, называемую «сверточной сетью». Но решение, которое принимает автопилот автомобиля, когда он «видит» разметку полосы движения, тротуар, припаркованный автомобиль или велосипед, зависит от традиционных систем планирования траектории движения, с правилами, написанными вручную, или же систем, основанных на правилах, которые относятся к области GOFAI.

Полностью автономные транспортные средства находятся все еще на этапе тестирования, но ряд коммерческих автомобилей, подобных «Tesla» 2015 г., уже имеют системы помощи водителю с использованием сверточных сетей. Регуляторы скорости, оснащенные системами технического зрения, берут транспортное средство под автономное управление на автостраде, удерживают его в пределах полосы движения или автоматически меняют ее после того, как водитель включает сигнал поворота, и при этом следят за наличием других автомобилей вокруг.

Тест Тьюринга

Мы будем писать о возможностях и приложении искусственного интеллекта на протяжении всей этой книги, но сейчас пришло время сделать шаг назад. Как определить общие черты всех этих интеллектуальных машин?

Я бы сказал, что искусственный интеллект – это способность машины выполнять задачи, обычно выполняемые животными и людьми, то есть воспринимать, рассуждать и действовать. Эти свойства неотделимы от способности учиться, как это наблюдается и у живых существ. Системы искусственного интеллекта – это просто очень сложные электронные схемы и компьютерные программы. Но возможности хранения информации, доступ к памяти, скорость вычислений и возможности обучения позволяют им «абстрагироваться» от конкретных примеров, содержащейся в огромных объемах данных.

Воспринимать, рассуждать и действовать. Алана Тьюринга – английского математика, оказавшего существенное влияние на развитие информатики и расшифровавшего Enigma – систему шифрования сообщений немецкой армии времен Второй мировой войны, можно назвать первым «пророком» обучающихся машин. Он уже проникся важностью обучения, когда написал: «Вместо того чтобы пытаться создать программу, имитирующую сознание взрослого, почему бы не попытаться создать такую, которая имитирует ум ребенка. Ведь если ум ребенка получает соответствующее воспитание, он становится умом взрослого человека»[8].

Имя Алана Тьюринга связано, кроме того, со знаменитым тестом, суть которого сводится к диалогу между человеком и двумя собеседниками, которых он не видит: компьютером и еще одним человеком[9]. Если по истечении некоторого заданного времени человек не определяет, кто из двух «собеседников» является машиной, значит, машина успешно прошла тест. Но достижения в области ИИ сегодня таковы, что эксперты больше не считают тест Тьюринга эффективным. Способность вести осмысленный диалог является лишь одной из форм интеллекта, и здесь искусственный интеллект легко может обмануть даже опытного эксперта: для этого машине достаточно выдать себя за рассеянного и слегка аутичного подростка, плохо знающего английский язык, чтобы объяснить недостаточное понимание собеседника и ошибки в собственной речи.

[8] Alan Turing, Computing machinery and intelligence, *Mind*, october 1950, vol. 59, n236.

[9] То же.

Постоянное совершенствование

Я уверен, что глубокое обучение – это неотъемлемая часть будущего искусственного интеллекта. Однако на сегодняшний день эти системы не способны к логическим рассуждениям. В то же время подходы к ИИ, основанные на логике, в нынешнем их состоянии несовместимы с обучением. Наша важнейшая задача на ближайшие годы – сделать эти два подхода совместимыми друг с другом.

Таким образом, глубокое обучение пока остается очень мощным… и очень ограниченным инструментом. Речь не идет о том, чтобы заставить машину, обученную игре в шахматы, работать, и наоборот. Она выполняет действия, не имея ни малейшего представления о том, что делает, и не обладает здравым смыслом. Если бы системы искусственного интеллекта были помещены на шкалу интеллектуальных способностей от мыши до человека, то они оказались бы намного ближе к мыши, чем к человеку – и это несмотря на то, что производительность ИИ в точных и узкоспециализированных задачах является сверхчеловеческой.

Могущество алгоритма

Алгоритм – это последовательность инструкций. Вот и все. В этом нет ничего волшебного. Ничего непонятного. Приведем пример. Возьмем список цифр, которые я хочу расставить в порядке возрастания. Я пишу компьютерную программу, которая считывает первое число, сравнивает его со следующим и меняет их положение, если первое больше второго. Затем я сравниваю второе и третье и повторяю ту же операцию до последнего числа в списке. Затем я возвращаюсь к списку столько раз, сколько необходимо, пока при очередном проходе число произошедших замен не станет равным нулю.

Данный алгоритм сортировки списка чисел называется «сортировкой пузырьком». Я могу перевести его в серию точных инструкций на вымышленном языке программирования[10].

```
Сортировка пузырьком (Таблица T)
###для i в диапазоне от (значение T) –1 до 1
######для j в диапазоне от 0 до i –1
#########если T[j+1] < T[j]
############обменять (T, j+1, j)
```

Возьмите одно значение, сравните его с другим, прибавьте его к третьему, выполните такие-то и такие-то математические операции, циклы, проверьте, является ли условие истинным или ложным и т. д. Алгоритм – все равно, что кулинарный рецепт.

Мы обычно говорим об «алгоритме Фейсбука» или «алгоритме Гугла». Это неправильно. Скорее, алгоритмом (точнее, набором алгоритмов) является механизм, обеспечивающий работу поискового сайта, который создает список всех сайтов, содержащих поисковый текст. Таких сайтов может быть сотни, даже тысячи! Затем каждому из этих сайтов присваивается ряд баллов, полученных с помощью других алгоритмов, написанных вручную или выработанных самой машиной в процессе обучения. Эти баллы оценивают популярность сайта, его надежность, релевантность его содержания, наличие ответа, если поисковая фраза является вопросом, а также соответствие содержания интересам пользователя. Довольно сложное дело.

Однако, что касается обучаемых систем, то программный код, который заставляет их работать и вычисляет баллы, достаточно прост и мог бы уместиться в нескольких строках, если бы нас не интересовала скорость его выполнения (на самом деле требования к быстродействию приводят к его усложнению). Реальная сложность системы заключается не в коде, который вычисляет ее выходные данные, а в связях между нейронами сети, которые, в свою очередь, зависят от архитектуры этой сети и ее обучения.

Прежде чем мы с вами исследуем внутреннее устройство интеллектуальной машины, я хочу обрисовать историю искусственного интеллекта, начиная с середины XX века. Это – захватывающая история, в которой я принимаю участие уже довольно давно, и которая состоит из предвидений и дискуссий, скачков вперед и периодов застоя, где между собой столкнулись ученые, верящие в машинную логику, и те, кто, опираясь на нейробиологию и кибернетику, работают, как и я, над развитием способностей машин к обучению.

[10] https://fr.wikipedia.org/wiki/Tri_a_bulles.

Глава 2
Краткая история искусственного интеллекта… и моего карьерного пути

Вечный поиск

Американский автор Памела Маккордак заметила как-то, что история искусственного интеллекта начинается с «извечного желания играть в Бога». Издавна человек пытается сконструировать устройства, создающие иллюзию жизни. В XX веке достижения науки дали надежду на механизацию мыслительного процесса. С появлением первых роботов и компьютеров в 1950-х гг. некоторые утописты предсказывали, что вычислительные машины быстро достигнут уровня человеческого интеллекта. Фантасты описали такие компьютеры во всех подробностях, но на сегодняшний день мы еще далеки от их воплощения в реальность.

Прогресс на этом долгом пути зависит от технических инноваций: более быстрые процессоры, более емкие устройства памяти. В 1977 г. у суперкомпьютера Cray-1 вычислительная мощность составляла 160 MFLOPS (мегафлопс). Он весил 5 т, потреблял 115 кВт·ч и стоил 8 млн долларов. На сегодняшний день игровая видеокарта стоимостью 300 евро, которую можно найти в компьютере у каждого второго увлеченного видеоиграми школьника, обеспечивает скорость 10 TFLOPS (терафлопс), или в 60 000 раз больше. Скоро любой смартфон сможет похвастаться такой мощностью.

История искусственного интеллекта в современном понимании берет начало с Дартмутского семинара, на котором впервые и прозвучал сам термин «искусственный интеллект». Семинар проходил летом 1956 г. в Дартмутском колледже недалеко от Ганновера, штат Нью-Гэмпшир, и был организован двумя исследователями – Марвином Мински и Джоном МакКарти. Марвин Мински был увлечен концепцией самообучающейся машины. В 1951 г. он в компании еще одного студента из Принстона построил одну из первых подобных машин, SNARC, небольшую электронную нейронную сеть с 40 «синапсами», способную к элементарному обучению. Джон МакКарти, в свою очередь, изобрел LISP, язык программирования, широко используемый в разработке ИИ.

Джону МакКарти также приписывают применение графов для создания шахматных программ. В работе семинара приняло участие около 20 исследователей, в том числе Клод Шеннон, инженер-электротехник и математик из Bell Labs (лаборатории гигантской телефонной компании AT&T в Нью-Джерси), Натан Рочестер из IBM и Рэй Соломонофф – основоположник концепции машинного обучения.

Участники бурно обсуждали области, движимые зарождающимися автоматизированными вычислениями и кибернетикой: изучение правил в естественных и искусственных системах, сложная обработка информации, искусственные нейронные сети, теория автоматов… На этом небольшом семинаре были разработаны и декларированы основные принципы и подходы к созданию искусственного интеллекта. Сам термин предложил все тот же Джон МакКарти.

Логика превыше всего

В настоящее время часть ученых представляет себе интеллектуальные машины, работающие только на основе логики и использующие для этого деревья поиска и экспертные системы. Инженер вводит в систему данные и правила их обработки, а система извлекает из них результаты вычислений или анализа. Более широкая цель ученых заключается в создании машины, способной заменить человека в сложных рассуждениях. Аллан Ньюэлл и Герберт Саймон из Университета Карнеги-Меллона в Питтсбурге первыми разработали программу Logic Theorist («Логический теоретик»), которая умела доказывать простые математические теоремы, исследуя дерево поиска, составленное из преобразований математических формул. Это было прекрасное время.

Однако через некоторое время тематика ИИ надолго погрузилась в спячку. В 1970 г. Агентство Министерства обороны США ARPA[11] сократило бюджеты фундаментальных исследований в области ИИ. Три года спустя, после отчёта Лайтхилла, который дал крайне пессимистические прогнозы для будущих исследований в области искусственного интеллекта, Великобритания сделала то же самое. Нет денег – нет исследований…

Процесс сдвинулся с мертвой точки в начале 1980-х. В то время большие надежды подавали экспертные системы, и Япония запустила амбициозный проект по созданию «компьютера пятого поколения», который должен был интегрировать навыки логического мышления в саму его конструкцию. Он должен был уметь вести беседу, переводить тексты, интерпретировать изображения и, возможно, даже рассуждать, как человек. К сожалению, эта идея себя не оправдала. Разработка и коммерциализация экспертных систем, таких как MYCIN, о которой мы уже говорили, оказались намного сложнее, чем ожидалось.

Идея поставить рядом с врачами или инженерами когнитологов (специалистов по восприятию и познанию), пытавшихся записать ход их мыслей или рассуждений в процессе диагностики болезни или неисправности, не сработала. Это опять оказалось сложнее, дороже и вовсе не так надежно, как предполагалось вначале, и к тому же упрощало знания и опыт, накопленные специалистом, до примитивного набора правил.

С этим «классическим» интеллектом, который так трудно воспроизвести, связаны алгоритмы на графах, которые тоже имели в свое время оглушительный успех.

[11] Агентство перспективных исследовательских проектов, которое в 1972 г. переименовали в DARPA (Агентство перспективных оборонных исследовательских проектов), является агентством Министерства обороны по финансированию проектов исследований и разработок (НИОКР).

Игровой мир

В 1997 г. чемпиона мира по шахматам Гарри Каспарова пригласили в Нью-Йорк принять участие в матче из шести партий против Deep Blue, суперкомпьютера, созданного транснациональной корпорацией IBM, – монстра высотой почти 2 м и весом 1,4 т. В шестой, заключительной партии матча, которую предваряли три ничьи и по одной победе с каждой стороны, Гарри Каспаров сдался уже на двадцатом ходу. Он признался, что был поражен и побежден вычислительной мощью машины.

Давайте на минутку остановимся на устройстве Deep Blue. В компьютере было 30 процессоров, дополненных 480 схемами, специально разработанными для проверки позиций на шахматной доске. Обладая такой вычислительной мощностью, машина могла оценивать качество примерно 200 млн позиций на доске в секунду, используя классические алгоритмы деревьев поиска.

Несколько лет спустя, 14, 15 и 16 февраля 2011 г., после трех раундов игры компьютер IBM Watson одержал победу в американской игре-викторине – *Jeopardy!* Аватар компьютера, размещенный между двумя чемпионами, представлял собой земной шар, покрытый световыми лучами. Программа, написанная учеными-компьютерщиками, удаляла из вопроса ненужные слова (артикли, предлоги…), определяла значимые слова и искала эти слова в огромном количестве текстов, порядка 200 млн страниц, определяя там те предложения, в которых можно было найти ответ.

Эти тексты, а именно вся Википедия, энциклопедии, словари, тезаурусы, сообщения информационных агентств, литературные произведения, хранились в его оперативной памяти объемом 16 терабайт[12] (жесткие диски были бы слишком медленными для таких задач). Сотни тысяч объектов и имен собственных сформировали большой список записей, каждая из которых относилась к определенной статье Википедии, веб-странице или тексту, где она появилась… Система Watson проверяла, существует ли документ, в котором присутствует часть ключевых терминов вопроса. После этой проверки проблема сводилась к нахождению правильного ответа в статье.

Например, если бы вопрос звучал так: «Где родился Барак Обама?», система Watson знала, что ответом будет название места. В ее базе данных был список всех документов, в которых упоминался Барак Обама. Поэтому где-нибудь обязательно нашлась бы статья со словами «Обама», «родился» и «Гавайи». Затем машине просто нужно было выбрать слово «Гавайи», соответствующее ответу. Watson, по сути, состояла из системы быстрого поиска информации в сочетании с хорошей индексацией данных. Но эта система не понимала смысла вопроса. Она вела себя как школьник, выполняющий домашнее задание с открытой книгой (или открытой Википедией!). На заданный вопрос она может найти правильный ответ в учебнике и переписать его, ничего не понимая в том, что пишет.

В 2016 г. появляется еще одно достижение. В Сеуле южнокорейский чемпион по игре в го был побежден своим компьютерным противником AlphaGo, внушительной системой, разработанной DeepMind, дочерней компанией Google. Восемнадцатикратный чемпион мира Ли Седоль проиграл машине четыре игры из пяти. В отличие от Deep Blue, AlphaGo была «обучена». Она тренировалась, играя против самой себя, сочетая при этом несколько хорошо известных методов: сверточные сети, усиленное обучение и метод Монте-Карло для поиска в дереве, метод «рандомизированных деревьев поиска». Впрочем, не будем забегать вперед.

[12] Терабайт – это единица измерения количества цифровой информации, здесь используется в качестве единицы измерения объема памяти. Он составляет 240 байт. Один байт может иметь 256 различных значений.

Нейробиология и перцептрон

В 1950-х гг., когда «великие магистры» классического искусственного интеллекта, основанного на логике и графах, раздвинули границы его применения, пионеры машинного обучения сформулировали альтернативные идеи. Они были уверены, что логики недостаточно, чтобы компьютерные системы были способны решать сложные задачи. Необходимо было приблизиться к функционированию мозга и тем самым сделать системы способными программировать самих себя, опираясь на механизмы обучения мозга. Это направление основано на так называемом «глубоком обучении» (deep learning) и искусственных нейронных сетях, – именно в этой области я и работаю. На подобных механизмах основана работа большинства продвинутых современных приложений, включая автономные автомобили.

Происхождение метода относится к середине прошлого века. Еще в 1950-х гг. пионеры искусственного интеллекта поддерживали теории, разработанные Дональдом Хеббом, канадским психологом и нейробиологом, который, в частности, размышлял о роли нейронных связей в обучении. Вместо того чтобы воспроизводить логические цепочки человеческих рассуждений, почему бы не исследовать их носитель, этот потрясающий биологический процессор, которым является мозг?

Таким образом, исследователи вычислений сконцентрировались на нейронном способе обработки информации в отличие от ранее применявшейся логической, или «последовательной», обработки. Они нацелились на моделирование биологических нейронных цепей. Машинное обучение, на которое были направлены их усилия, основывалось на оригинальной архитектуре, сети математических функций, которые по аналогии называют «искусственными нейронами».

Они улавливают входной сигнал, и нейроны в сети обрабатывают его таким образом, что на выходе этот сигнал идентифицируется. Сложность операции, например, распознавание образов, поддерживается комбинированным взаимодействием очень простых элементов, а именно искусственных нейронов. Так и в нашем мозге взаимодействие основных функциональных единиц – нейронов – порождает сложные мысли.

Возникновение описываемой концепции датируется 1957 г.: в том же году в Корнельском университете психолог Фрэнк Розенблатт, вдохновленный когнитивной теорией Дональда Хебба, построил перцептрон – первую обучающуюся машину. Мы рассмотрим ее в следующей главе, так как она являются эталонной моделью машинного обучения. После обучения перцептрон способен, например, распознавать образы (геометрические фигуры, буквы и т. д.).

В 1970-х гг. два американца, Ричард Дуда, в то время профессор электротехники в Университете Сан-Хосе (Калифорния), и Питер Харт – ученый-компьютерщик из SRI (Стэнфордского исследовательского института) в Менло-Парке (Калифорния), обсуждали все эти так называемые методы «распознавания статистических форм[13]», примером которых является перцептрон. С самого начала их руководство стало мировым эталоном, Библией распознавания образов для всех студентов… и для меня тоже.

Но перцептрон далеко не всесилен. Система, состоящая из одного слоя искусственных нейронов, имеет «врожденные» ограничения. Исследователи пытались увеличить его эффективность, вводя несколько слоев нейронов вместо одного. Но без алгоритма обучения слоев, который к тому моменту еще не был известен, такие машины все еще оставались очень малоэффективными.

[13] Richard O. Duda, Peter E. Hart, Pattern Classification and Scene Analysis, Wiley, 1973.

Эпоха застоя

Перейдем к кому времени, когда в 1969 г. Сеймур Паперт и Марвин Мински – тот самый, который в 1950-х гг. увлекался искусственными нейронными сетями, прежде чем отречься от них, опубликовали книгу «Перцептроны: Введение в вычислительную геометрию»[14]. Они математически доказали пределы возможностей перцептрона, и некоторые из доказанных ими ограничений по сути поставили крест на использовании этой и подобных машин.

Казалось, развитие уперлось в непреодолимую стену. Эти два профессора Массачусетского технологического института пользовались большим авторитетом, так что их работа наделала много шума. Агентства, финансирующие исследования, прекратили поддержку исследовательских работ в этой области. Как и исследования в GOFAI, исследования нейронных сетей пережили серьезный застой.

В этот период большинство ученых перестали говорить о создании умных машин, способных к обучению. Они предпочитали ограничивать свои амбиции более приземленными проектами. Используя методы, унаследованные от нейронных сетей, они создали, например, «адаптивную фильтрацию» – процесс, лежащий в основе многих коммуникационных технологий в современном мире. Прежде физические свойства проводных линий связи сильно ограничивали передачу высокочастотных сигналов, приводя к их существенным искажениям уже на расстоянии нескольких километров. Теперь сигнал восстанавливается с помощью адаптивного фильтра. Используемый алгоритм называется Lucky's Algorithm в честь его изобретателя Боба Лаки, который в конце 1980-х руководил отделом Bell Labs, где тогда работало около 300 человек. в том числе и я.

Без адаптивной фильтрации у нас не было бы телефона с громкой связью, который позволяет вам говорить в микрофон без самовозбуждения, происходящего от усиления микрофоном звука громкоговорителя (когда это случается, мы слышим громкий вой или свист). В эхокомпенсаторах, кстати, используются алгоритмы, очень похожие на алгоритм перцептрона.

Не появился бы без этой технологии и модем[15]. Это устройство позволяет одному компьютеру коммуницировать с другим компьютером по телефонной линии или иной линии связи.

[14] Marvin L. Minsky, Seymour A. Papert, Perceptrons: An Introduction to Computional Geometry, The MIT Press, 1969.

[15] Устройство, состоящее из модулятора и демодулятора, предназначенное для передачи цифровых данных по телефону или по коаксиальному кабелю.

Преданные последователи

Тем не менее, и во времена застоя в 1970-х и 1980-х гг. некоторые ученые продолжали работать над нейронными сетями, хотя научное сообщество считало их сумасшедшими, чуть ли не фанатиками. Я имею в виду Теуво Кохонена, финна, который написал об «ассоциативных воспоминаниях» – теме, близкой к нейронным сетям. Я также говорю и о группе японцев – в Японии существует изолированная инженерная экосистема, отличная от западной, – и среди них о математике Сун-Ити Амари и исследователе искусственного интеллекта Кунихико Фукусима. Последний работал над машиной, которую он назвал «когнитроном», по аналогии с термином «перцептрон». Он создал две его версии: «когнитрон» 1970-х и «неокогнитрон» 1980-х. Как и Розенблатт в свое время, Фукусима был вдохновлен достижениями нейробиологии, особенно открытиями американца Дэвида Хьюбела и шведа Торстен Н. Визеля.

Эти два нейробиолога получили Нобелевскую премию по физиологии в 1981 г. за свою работу над зрительной системой кошек. Они обнаружили, что зрение возникает в результате прохождения визуального сигнала через несколько слоев нейронов, от сетчатки до первичной зрительной коры, затем в другие области зрительной коры, и, наконец – в нижневисочную кору. Нейроны в каждом из этих слоев выполняют особые функции. В первичной зрительной коре каждый нейрон связан только с небольшой областью поля зрения, а именно со своим рецепторным полем. Такие нейроны называются «простыми». В следующем слое другие нейроны включают активацию предыдущего слоя, что помогает поддерживать представление изображения, если объект немного перемещается в поле зрения. Такие нейроны называются «сложными».

Таким образом, Фукусима был вдохновлен идеей первого слоя простых нейронов, которые обнаруживают простые узоры в небольших рецепторных полях, выдающих изображение, и сложных нейронов в следующем слое. Всего в неокогнитроне было пять слоев: простые нейроны – сложные нейроны – простые нейроны – сложные нейроны, и затем «классификационный слой», подобный перцептрону. Он, очевидно, использовал для первых четырех уровней некоторый алгоритм обучения, но последний был «неконтролируемым», то есть он не принимал во внимание конечную задачу. Такие слои обучались «вслепую». Только последний слой обучался под наблюдением (как и перцептрон). У Фукусимы не было алгоритма обучения, который регулировал бы параметры всех слоев его неокогнитрона. Однако его сеть позволяла распознавать довольно простые формы, например, символы чисел.

В начале 1980-х гг. идеи Фукусимы поддерживали и другие ученые. Некоторые североамериканские исследовательские группы также работали в этой области: психологи Джей Макклелланд и Дэвид Румелхарт, биофизики Джон Хопфилд и Терри Сейновски, и ученые-компьютерщики, в частности Джеффри Хинтон – тот самый, с которым я впоследствии разделю Премию Тьюринга, присужденную в 2019 г.

Мой выход на сцену

Я начал интересоваться всеми этими темами в 1970-х гг. Возможно, любопытство к ним зародилось во мне еще, когда я наблюдал за моим отцом, авиационным инженером и мастером на все руки, который в свободное время занимался электроникой. Он строил модели самолетов с дистанционным управлением. Я помню, как он сделал свой первый пульт для управления небольшой машиной и лодкой во время забастовок в мае 1968 г., когда он проводил много времени дома. Я не единственный в семье, кому он передал свою страсть к любимому делу. Мой брат, который на шесть лет младше меня, тоже сделался ученым-компьютерщиком. После академической карьеры он стал исследователем в компании Google.

С самого раннего детства меня манили новые технологии, компьютеризация, покорение космоса… Еще я мечтал стать палеонтологом, потому что меня очень интриговал человеческий интеллект и его эволюция. Даже сегодня я по-прежнему верю, что работа нашего мозга остается самой загадочной вещью в мире. Я помню, как в Париже на большом экране я вместе с моими родителями, а также дядей и тетей – «фанатами» научной фантастики, смотрел фильм «2001: Космическая одиссея». Мне было тогда восемь лет. Фильм затронул все, что я любил: космические путешествия, будущее человечества и восстание суперкомпьютера «Хэл», который готов был убивать ради собственного выживания и успеха миссии. Уже тогда меня волновал вопрос о том, как воспроизвести человеческий интеллект в машине.

Неудивительно, что после школы я захотел воплотить эти мечты в жизнь. В 1978 г. я поступил в Парижскую высшую школу электронной инженерии (École Supérieure d'Ingénieurs en Électrotechnique et Électronique, ESIEE) в которую можно подавать заявление сразу после получения степени бакалавра, без затрат времени на дополнительную подготовку. (Откровенно говоря, длинная учеба – не единственный способ добиться успеха в науке. Я могу это подтвердить на своем примере.) А поскольку учеба в ESIEE предоставляет студенту некоторую свободу, я сумел воспользоваться этим.

Плодотворное чтение

Меня воодушевили новости о дебатах на конференции Cerisy о врожденном и приобретенном знании[16], прочитанные мною в 1980 г. Лингвист Ноам Хомски подтвердил, что в мозге существуют исходно заложенные структуры, позволяющие человеку научиться языку. Психолог Жан Пиаже защищал идею о том, что любое обучение задействует определенные, уже существующие структуры мозга, и что овладение языком осуществляется поэтапно по мере того, как формируется интеллект. Таким образом, интеллект будет результатом обучения, основанного на обмене информацией с внешним миром. Эта идея мне понравилась, и мне стало интересно, как ее можно применить в отношении машины. В этой дискуссии принимали участие именитые ученые, в том числе Сеймур Паперт. В ней он восхвалял перцептрон, который описывал как простую машину, способную обучаться сложным задачам.

Так я и узнал о существовании обучающейся машины. Эта тема меня просто очаровала! Поскольку я не учился по средам после обеда, я начал рыскать по полкам библиотеки Национального института компьютерных и автоматических исследований в Роккенкуре (National Institute for Research in Digital Science and Technology, сокращенно «Inria»). У этого учреждения самый богатый библиотечный фонд ИТ-литературы в Иль-де-Франс. Я вдруг понял, что на Западе больше никто не работает с нейронными сетями, и с еще большим удивлением обнаружил, что книга, положившая конец исследованиям перцептрона, принадлежит перу того же самого Сеймура Паперта!

Теория систем, которую в 1950-х гг. мы называли кибернетикой, и которая изучает естественные (биологические) и искусственные системы – еще одна моя страсть. Возьмем, например, систему регулирования температуры тела: организм человека поддерживает температуру 37 °C благодаря наличию своеобразного «термостата», который корректирует разницу между своей температурой и температурой снаружи.

Меня увлекла идея самоорганизации систем. Каким образом относительно простые молекулы или объекты могут спонтанно организовываться в сложные структуры? Как может появиться интеллект из большого набора простых взаимодействующих нейронов?

Я изучал математические работы по теории алгоритмической сложности Колмогорова, Соломонова и Чайтина. Книга Дуды и Харта[17], о которой я уже упоминал, стала для меня настольной. Я читал журнал «Биологическая кибернетика» («Biological Cybernetics. Advances in Computational Neuroscience and in Control and Information Theory for Biological Systems», издательство Springer), посвященный математическим моделям работы мозга или живых систем.

Все эти вопросы, оставленные без ответа в период застоя искусственного интеллекта, не выходили у меня из головы, и у меня постепенно стало формироваться убеждение: если мы хотим создавать интеллектуальные машины, недостаточно, чтобы они работали только логически, они должны быть способными учиться, совершенствоваться на собственном опыте.

Читая все эти труды, я понимал, что часть научного сообщества разделяет мое виденье проблемы. Вскоре я познакомился с работами Фукусимы и задумался о способах повышения эффективности нейронных сетей неокогнитрона. К счастью, ESIEE предоставлял студентам компьютеры, которые для того времени были очень мощными. Мы писали программы с Филиппе Метсу, школьным другом, любителем искусственного интеллекта, как и я, хотя его больше интересовала психология обучения детей. Преподаватели математики согласились

[16] Théories du langage, théories de l'apprentissage: le débat entre Jean Piaget et Noam Chomsky, d.bat recueilli par Maximo Piatelli-Palmarini, Centre Royaumont pour une science de l'homme, Seuil, Points, 1979.

[17] Richard O. Duda, Peter E. Hart, Pattern Classification and Scene Analysis, p. 6.

заниматься с нами дополнительно. Вместе мы пытались моделировать нейронные сети. Но эксперименты отнимали очень много сил: компьютеры не тянули наши эксперименты, а написание программ было сплошной головной болью.

На четвертый год обучения в ESIEE, одержимый этим исследованием, я догадался о не совсем математически обоснованном правиле обучения многослойных нейронных сетей. Я представил алгоритм, который будет распространять сигналы в обратном направлении по сети, начиная с выходного слоя, чтобы обучать сеть от начала до конца. Я назвал этот алгоритм HLM (от Hierarchical Learning Machine)[18].

Я очень гордился своей идеей… HLM является предшественником алгоритма «обратного распространения градиента», который сегодня повсеместно используется для обучения систем глубокого обучения. Вместо распространения обратных градиентов в сети, как это происходит сегодня, HLM распространял желаемые состояния для каждого нейрона. Это позволяло использовать бинарные нейроны, что являлось преимуществом, учитывая медлительность компьютеров того времени для выполнения умножения. HLM был первым шагом в обучении многоуровневых сетей.

[18] См. главу 5 «Мой HLM!».

Коннекционистские модели обучения

Летом 1983 г. я получил высшее образование по специальности «инженер». Тогда же я наткнулся на книгу, в которой рассказывалось о работе небольшой группы французов, интересующихся самоорганизующимися системами и сетями автоматов. Они экспериментировали в бывшем помещении Политехнической школы на холме Святой Женевьевы в Париже. Эта лаборатория сетевой динамики (Laboratoire de dynamique de réseau, или LDR) была независимой, хотя ее члены занимали должности в разных высших учебных заведениях. У них было мало денег, не было планового бюджета, а их компьютер нуждался в ремонте. Это означало, что исследования машинного обучения во Франции висят на волоске! Я решил примкнуть к ним. Я мог реально помочь им, потому что эти ученые не занимались изучением старых публикаций по нейронным сетям, как это делал я.

Я решил объяснить им, что меня интересует эта тема и что в своей инженерной школе я занимаюсь схожей тематикой. Я работал в их группе, продолжая учебу в аспирантуре в Университете Пьера и Марии Кюри. В 1984 г. мне нужно было подать заявление на защиту докторской диссертации. Я занимал должность младшего научного сотрудника ESIEE по гранту, но мне нужно было найти себе научного руководителя. Много времени я работал с Франсуазой Фогельман-Суле (сейчас Сули-Фогельман), которая в то время преподавала компьютерные науки в Университете Париж-V и, по логике вещей, именно она должна была бы курировать мою диссертацию, но у нее не было на это полномочий, поскольку она еще не прошла государственную сертификацию на право руководить аспирантами (необходимую во многих европейских странах).

Поэтому я обратился к единственному члену лаборатории, который мог курировать диссертацию по информатике, – Морису Милграму, профессору информатики и инженерии Технологического университета Компьена. Он согласился, но дал понять, что не сможет мне сильно помочь, потому что ничего не знает о нейронных сетях, но я и так был безмерно благодарен ему за эту помощь. Поэтому я посвятил свое время одновременно ESIEE (и ее мощным компьютерам) и LDR (и ее интеллектуальной среде). Я попал на ранее неизвестную мне территорию, и это было интересно.

За рубежом исследования, близкие к моим, набирали обороты. Летом 1984 г. я сопровождал Франсуазу Фогельман в Калифорнию, где прошел месячную стажировку в известной многим лаборатории Xerox PARC.

В то время, я помню, в мире было два человека, с которыми я мечтал встретиться: Терри Сейновски – биофизик и нейробиолог из Университета Джона Хопкинса в Балтиморе, и Джеффри Хинтон из Университета Карнеги-Меллон в Питтсбурге – тот самый, кто поделит с Йошуа Бенджио и мной Премию Тьюринга в 2019 г. В 1983 г. Хинтон и Сейновски опубликовали статью о машинах Больцмана[19], которая содержит процедуру обучения сетей со «скрытыми нейронами», то есть нейронами в промежуточных слоях между входом и выходом. Я увлекся этой статьей именно потому, что в ней говорилось об обучении многослойных нейронных сетей. «Главный» вопрос в моей работе! Эти люди сыграли важную роль в моей жизни!

[19] Машиной Больцмана называется один из видов нейронных сетей. – *Прим. ред.*

Лез-Уш

Моя профессиональная жизнь изменилась в феврале 1985 г. во время конференции в Лез-Уш, в Альпах. Там я встретился с лучшими представителями мировой науки, интересующимися нейронными сетями: физиками, инженерами, математиками, нейробиологами, психологами и, в частности, членами новой развивающейся исследовательской группы в области нейронных сетей, которая сформировалась внутри легендарной лаборатории Bell Labs. Через три года я попал в эту группу благодаря знакомствам, которые приобрел в Лез-Уш.

Встреча была организована теми французскими исследователями из LDR, с которыми я уже работал: Франсуазой, ее тогдашним мужем Жераром Вайсбухом, профессором физики ENS, и Эли Биненштоком – нейробиологом-теоретиком, работавшим в то время в CNRS. Конференция собрала вместе физиков, интересующихся «спиновыми стеклами», а также ведущих физиков и нейробиологов.

Спин – это свойство элементарных частиц и атомов, которое можно описать по аналогии с маленькими магнитами, с обращенными вверх или вниз полюсами. Эти два значения спина можно сравнить с состояниями искусственного нейрона: он либо активен, либо неактивен. Он подчиняется тем же уравнениям. Спиновые стекла представляют собой своего рода кристалл, в котором примесные атомы имеют магнитный момент. Каждый спин взаимодействует с другими спинами на основе связанных весовых показателей.

Если весовой коэффициент положительный, они, как правило, выстраиваются в одном направлении. Если вес отрицательный, они противопоставляются. Мы связываем значения +1 со спином «вверх», а –1 со спином «вниз». Каждый примесный атом принимает ориентацию, которая является функцией взвешенной суммы ориентаций соседних примесных атомов. Другими словами, функция, определяющая, будет ли спин идти вверх или вниз, аналогична функции, которая делает искусственный нейрон активным или неактивным.

После основополагающей статьи Джона Хопфилда[20], в которой были описаны аналогии между спиновыми стеклами и искусственными нейронными сетями, многие физики начали интересоваться и самими сетями, и их обучением – темами, по-прежнему не приветствовавшимися их коллегами – инженерами и компьютерщиками.

В Лез-Уш я был одним из самых молодых исследователей, и мне пришлось общаться на английском языке о многоуровневых сетях и алгоритме HLM, моем предшественнике алгоритмов обратного распространения. Я только начал подготовку своей диссертации, и нервничал, выступая перед столь именитой аудиторией.

Меня особенно привлекли две личности: Ларри Джекел, глава отдела Bell Labs (позже мне самому довелось работать в этом отделе) и Джон Денкер – настоящий ковбой из Аризоны: джинсовый костюм, большие бакенбарды, ковбойские сапоги… Этот не очень похожий на ученого человек, только что защитивший диссертацию, был невероятно уверен в себе! Когда на него находило вдохновение, он мог быть чертовски убедителен и изобретательно отстаивал свою точку зрения, причем без агрессии и часто вполне обоснованно. Франсуаза Фогельман говорила мне: «У ребят из Bell Labs огромное преимущество. Когда вы только хотите сделать что-то новое, то выясняется, что это либо уже было сделано в Bell Labs десять лет назад, либо это просто не работает». Черт возьми!

[20] John J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proceedings of the National Academy of Sciences, 1982, 79 (8), p. 2554–2558, DOI:10.1073/pnas.79.8.2554.

Конец ознакомительного фрагмента.