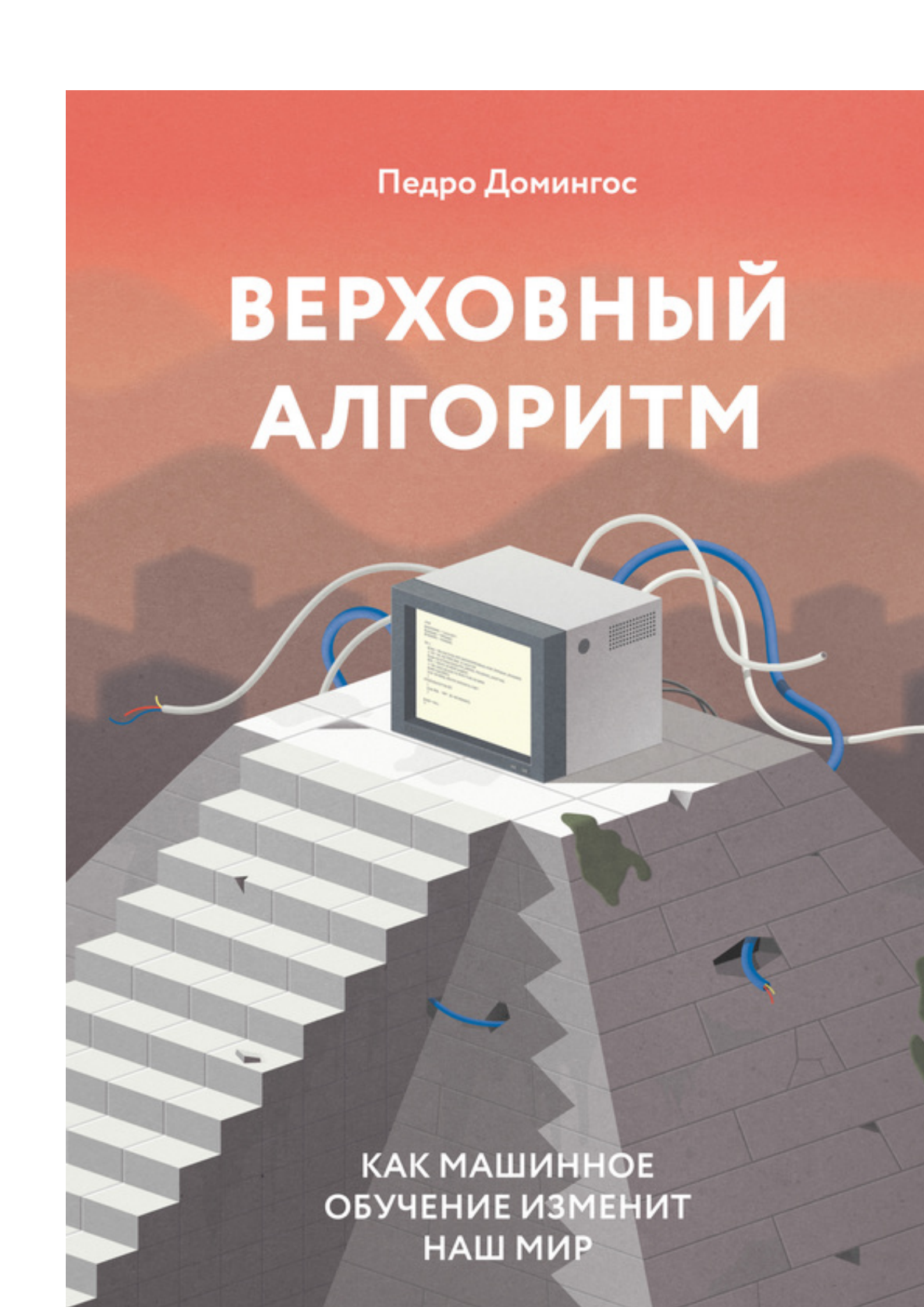


Педро Домингос

ВЕРХОВНЫЙ АЛГОРИТМ

The book cover features a stylized illustration of a computer monitor and tower unit perched atop a stone pyramid. A wide staircase of stone blocks leads up to the platform where the computer sits. Several blue and white cables snake across the scene, some connecting to the computer and others trailing off into the distance. The background is a hazy, orange-toned landscape with silhouettes of distant buildings. The overall aesthetic is a blend of ancient architecture and modern technology.

КАК МАШИННОЕ
ОБУЧЕНИЕ ИЗМЕНИТ
НАШ МИР

Педро Домингос

**Верховный алгоритм:
как машинное обучение
изменит наш мир**

«Манн, Иванов и Фербер»

2015

УДК 004.81:159.953.52
ББК 74.044.4

Домингос П.

Верховный алгоритм: как машинное обучение изменит наш мир / П. Домингос — «Манн, Иванов и Фербер», 2015

Машинное обучение преобразует науку, технологию, бизнес и позволяет глубже узнать природу и человеческое поведение. Программирующие сами себя компьютеры – одна из самых важных современных технологий, и она же – одна из самых таинственных. Ученый-практик Педро Домингос приоткрывает завесу и впервые доступно рассказывает о машинном обучении и о поиске универсального обучающегося алгоритма, который сможет выуживать любые знания из данных и решать любые задачи. Чтобы заглянуть в будущее и узнать, как машинное обучение изменит наш мир, не нужно специального технического образования – достаточно прочитать эту книгу. На русском языке публикуется впервые.

УДК 004.81:159.953.52

ББК 74.044.4

© Домингос П., 2015

© Манн, Иванов и Фербер, 2015

Содержание

Эту книгу хорошо дополняют:	6
Пролог	7
Глава 1	14
Познакомимся с обучающимся алгоритмом	18
Почему бизнес рад машинному обучению?	21
Турбоускорение для научного метода	24
Миллиард Клинтонов	27
Один сигнал, если сушей, два – если по интернету	29
Куда мы идем?	31
Конец ознакомительного фрагмента.	32

Педро Домингос

Верховный алгоритм: как машинное обучение изменит наш мир

Pedro Domingos

The Master Algorithm

How the Quest for the Ultimate Learning Machine Will Remake Our World

Научные редакторы Александр Сбоев, Алексей Серенко

*Издано с разрешения Pedro Domingos c/o Levine Greenberg Rostan Literary Agency
и литературного агентства Synopsis*

Правовую поддержку издательства обеспечивает юридическая фирма «Вегас-Лекс».

© Pedro Domingos, 2015

© Перевод на русский язык, издание на русском языке, оформление. ООО «Манн, Иванов и Фербер», 2016

* * *

Эту книгу хорошо дополняют:

[Искусственный интеллект](#)

Ник Бостром

[Красота в квадрате](#)

Алекс Беллос

[Теория игр](#)

Авинаш Диксит и Барри Нейлбафф

Светлой памяти моей сестры Риты, которая проиграла битву с раком, когда я писал эту книгу

Величайшая задача науки – объяснить как можно больше экспериментальных фактов логической дедукцией, исходящей из как можно меньшего количества гипотез и аксиом.

Альберт Эйнштейн

Прогресс цивилизации заключается в увеличении количества важных действий, которые мы выполняем не думая.

Альфред Норт Уайтхед

Пролог

Машинное обучение окружает вас повсюду, хотя, может быть, вы об этом и не подозреваете. Именно благодаря машинному обучению поисковая машина понимает, какие результаты (и рекламу) показывать в ответ на ваш запрос. Когда вы просматриваете почту, большая часть спама проходит мимо вас, потому что он был отфильтрован с помощью машинного обучения. Если вы решили что-нибудь купить на Amazon.com или заглянули на Netflix посмотреть фильм, система машинного обучения услужливо предложит варианты, которые могут прийтись вам по вкусу. С помощью машинного обучения Facebook решает, какие новости вам показывать, а Twitter подбирает подходящие твиты. Когда бы вы ни пользовались компьютером, очень вероятно, что где-то задействовано машинное обучение.

Единственным способом заставить компьютер что-то делать – от сложения двух чисел до управления самолетом – было составление некоего алгоритма, скрупулезно объясняющего машине, что именно от нее требуется. Однако алгоритмы машинного обучения – совсем другое дело: они угадывают все сами, делая выводы на основе данных, и чем больше данных, тем лучше у них получается. Это значит, что компьютеры не надо программировать: они программируют себя сами.

Это верно не только в киберпространстве: машинным обучением пронизана вся наша жизнь, начиная с пробуждения и заканчивая отходом ко сну.

Семь утра. Будильник включает радио. Играет незнакомая, но очень приятная музыка: благодаря сервису Pandora¹ радио познакомилось с вашими вкусами и превратилось в «персонального диджея». Не исключено, что сама песня тоже появилась на свет с помощью машинного обучения. За завтраком вы листаете утреннюю газету. Несколькими часами ранее она сошла с печатного станка, а тот был тщательно настроен с помощью обучающегося алгоритма, позволяющего устранить типографские дефекты. В комнате исключительно комфортная температура, а счета за электричество не кусаются, потому что вы поставили умный термостат Nest.

По дороге на работу автомобиль постоянно корректирует впрыск топлива и рециркуляцию выхлопных газов, чтобы свести к минимуму расходы на бензин. В часы пик Inrix, система прогнозирования трафика, экономит время, не говоря уже о нервах. На работе машинное обучение помогает вам бороться с информационной перегрузкой: вы используете куб данных², чтобы суммировать большой объем информации, смотрите на него под разными углами и подробно изучаете все самое важное. Если надо принять решение, какой макет сайта – *A* или *B* – привлечет больше перспективных клиентов, обученная система протестирует оба варианта и предоставит вам отчет. Надо заглянуть на сайт потенциального поставщика, а он на иностранном языке? Никаких проблем: Google автоматически его для вас переведет. Электронные письма удобно рассортированы по папкам, а во «Входящих» осталось только самое важное. Текстовый процессор проверяет за вас грамматику и орфографию. Вы нашли авиарейс для предстоящей командировки, но билет пока не покупаете, потому что, по прогнозу Bing Travel, цены вскоре станут ниже. Сами того не осознавая, вы ежечасно делаете намного больше работы, чем могли бы без помощи машинного обучения.

¹ Поток мультимедиа в интернете. Пользователь медиапроигрывателя Pandora выбирает музыкального исполнителя, после чего система ищет похожие композиции, используя около 400 музыкальных характеристик. С помощью функций «нравится» или «не нравится» слушатель часто может настроить «радиостанцию» по своему вкусу. *Здесь и далее, если не указано иное, прим. ред.*

² Действительная архитектура данных для быстрого агрегирования многомерной информации. Куб данных может быть реализован на основе универсальных реляционных баз данных или специализированным программным обеспечением.

В свободную минуту вы проверяете, как там ваши вклады в фонде взаимных инвестиций. Большинство таких фондов используют обучающиеся алгоритмы для выбора перспективных акций, а одним из них вообще полностью управляет система на основе машинного обучения. Во время обеда вы выходите на улицу и думаете, где бы перекусить. Обучающаяся система Yelp в смартфоне вам поможет. Мобильные телефоны вообще под завязку наполнены обучающимися алгоритмами, которые без устали исправляют опечатки, узнают голосовые команды, корректируют ошибки передачи данных, считывают штрихкоды и делают много других полезных дел. Смартфон даже научился угадывать ваше следующее действие и давать полезные советы. Например, он подскажет, что встреча начнется позже, потому что самолет, на котором должен прилететь ваш гость, задерживается.

Если вы закончите работать поздно вечером, машинное обучение поможет без приключений дойти до машины на парковке: алгоритмы отслеживают видео с камеры наблюдения и дистанционно предупреждают охрану, когда замечают что-то подозрительное. Предположим, по дороге домой вы тормозите у супермаркета. Товары на полках расположены согласно указаниям алгоритмов с обучением: именно они решают, какие товары лучше заказать, что поставить в конце ряда и где место салсе – в отделе соусов или рядом с чипсами тортильяс. На кассе вы расплачиваетесь кредитной карточкой. В свое время обученный алгоритм решил, что вам надо отправить предложение ее оформить, а затем одобрил вашу заявку. Другой алгоритм постоянно выискивает подозрительные операции и непременно предупредит вас, если ему покажется, что номер карточки украден. Третий алгоритм пытается понять, насколько вы удовлетворены. Если вы хороший клиент, но выглядите недовольным, вам отправят «подслащенное» предложение еще до того, как вы уйдете к конкурентам.

Вернувшись домой, вы подходите к почтовому ящику и находите там письмо от друга. Оно было доставлено благодаря алгоритму, который научился читать написанные от руки адреса. Кроме письма в ящике лежит обычная макулатура, тоже отобранная для вас алгоритмами с обучением (ничего не поделаешь). Вы на минутку останавливаетесь, чтобы подышать свежим вечерним воздухом. Преступников в городе сильно поубавилось, с тех пор как полиция начала использовать статистическое обучение для прогнозирования вероятности правонарушений и направлять в проблемные районы патрульных. Вы ужинаете в кругу семьи и смотрите телевизор. В новостях показывают мэра. Вы за него проголосовали, потому что в день выборов он лично вам позвонил. Ему на вас указал обучающийся алгоритм, увидевший в вас ключевого неопределившегося избирателя. После ужина можно посмотреть футбол. Обе команды подбирали игроков с помощью статистического обучения. Или лучше поиграть с детьми в Xbox? В таком случае обучающийся алгоритм в приставке Kinect будет отслеживать положение и движения вашего тела. Прежде чем отойти ко сну, вы принимаете лекарство, разработанное и протестированное с помощью алгоритмов с обучением. Не исключено, что даже ваш доктор пользовался машинным обучением при постановке диагноза, начиная с интерпретации рентгенограммы и заканчивая выводом на основе необычного набора симптомов.

Машинное обучение делает свое дело на всех этапах жизни человека. Если вы готовились к экзаменам в колледж с помощью интернета, специальный алгоритм оценивал ваши пробные сочинения. А если вы недавно поступали в бизнес-школу и сдавали GMAT³, обучающаяся система была одним из экзаменаторов, оценивающих эссе. Возможно, когда вы устраивались на работу, обученный алгоритм выудил ваше резюме из «виртуальной кучи» и сказал потенциальному работодателю: «Взгляни, вот сильная кандидатура». Вполне веро-

³ GMAT (Graduate Management Admission Test) – стандартизованный тест для определения способности успешно обучаться в бизнес-школах.

ятно, что недавним повышением зарплаты вы тоже обязаны какому-то обученному алгоритму. Если вы собираетесь купить дом, Zillow.com оценит, чего стоит каждый заинтересовавший вас вариант. Когда вы определитесь и решите взять ипотеку, алгоритм на основе машинного обучения рассмотрит заявку и порекомендует ее одобрить (или отклонить). И наверное, самое важное: если вас интересуют интернет-знакомства, машинное обучение поможет найти настоящую любовь.

Общество меняется с каждым новым алгоритмом. Машинное обучение преобразует науку, технологию, бизнес, политику, военное искусство. Спутники и ускорители частиц зондируют природу все более тщательно, а обучающиеся алгоритмы превращают реки данных в новое научное знание. Компании знают своих клиентов, как никогда раньше. На выборах побеждают кандидаты, умеющие лучше моделировать поведение избирателей (пример – Обама против Ромни⁴). Беспилотные транспортные средства завоевывают сушу, воду и воздушное пространство. В систему рекомендаций Amazon никто не вводит информацию о наших вкусах: обучающийся алгоритм определяет их самостоятельно, обобщая сведения о сделанных покупках. Беспилотный автомобиль Google сам научился не съезжать с дороги: никакой инженер не писал для него алгоритм, шаг за шагом объясняющий, как добраться из точки *A* в точку *B*. Никто не знает, как написать программу вождения автомобиля, да никому это и не надо, потому что машина, оборудованная обучающимся алгоритмом, посмотрит на действия водителя и разберется сама.

Машинное обучение – технология, которая строит саму себя. Это новое явление в нашем мире. С тех пор как наши далекие предки научились заострять камни и смастерили первые орудия труда, человечество разработало артефакты самостоятельно, вручную или массово. Обучающиеся алгоритмы – артефакты, которые создают другие артефакты. «От компьютеров никакой пользы, – говорил Пикассо. – Они умеют только давать ответы». Компьютеры не предназначены для творчества: они должны делать ровно то, что им говорят. Но если приказать им заняться творчеством, получится машинное обучение. Обучающийся алгоритм как искусный ремесленник: каждое из его творений уникально, и каждое создано именно таким, каким пожелал заказчик. Просто в отличие от мастеров обучающиеся алгоритмы превращают не камень в кладку и не золото в ювелирные изделия, а данные в алгоритмы. И чем больше у них данных, тем качественнее может получиться алгоритм.

Homo sapiens научился приспособлять мир под себя, вместо того чтобы самому приспосабливаться к существующим условиям. Машинное обучение открывает новую главу в долгой, растянувшейся на миллион лет эволюционной саге: с его помощью мир сам почувствует, чего вы хотите, и сам под вас подстроится. Не надо даже волшебной палочки: окружающий вас мир – сегодня виртуальный, а завтра физический – станет похож на волшебный лес. Если вы пойдете по тропинке в чаще, она станет дорогой. Если вы заблудитесь, из ниоткуда появятся стрелки, указывающие направление.

Эти волшебные технологии возможны потому, что глубинная суть машинного обучения – предсказание: предсказание наших желаний, результатов наших действий, путей достижения целей, изменений мира. Когда-то нам приходилось полагаться на шаманов и прорицателей, но это оказалось слишком ненадежно. Научные прогнозы более достойны доверия, однако они ограничены областями, которые мы можем систематически наблюдать и которые поддаются моделированию. Большие данные и машинное обучение заметно расширили эти границы. Иногда человек может предсказывать и без посторонней помощи, например, когда ловит мячик или ведет разговор. Бывает, что предсказать не получится,

⁴ Уиллард Митт Ромни (Willard Mitt Romney, род. 1947) – американский политик. Был кандидатом в президенты США на выборах 2012 года от Республиканской партии.

как бы мы ни старались. Но между этими крайностями лежит широкая область, для которой пригодится машинное обучение.

Хотя обучающиеся алгоритмы позволяют глубже узнать природу и человеческое поведение, сами они, как ни странно, окутаны пеленой тайны. Не проходит и дня, чтобы в СМИ не появилась новая история, связанная с машинным обучением, будь то запуск Apple личного помощника Siri, суперкомпьютер IBM Watson, победивший чемпиона в Jeopardy! (аналог «Своей игры»)⁵, торговая сеть Target, узнавшая о беременности подростка раньше родителей, или Агентство национальной безопасности, собирающее воедино разрозненные улики. Однако во всех этих случаях обучающиеся алгоритмы, сделавшие эти истории возможными, остаются для зрителей черным ящиком. Даже книги о больших данных обходят стороной вопрос, как именно компьютер, проглотив все эти терабайты, волшебным образом приходит к ценным выводам. В лучшем случае у нас остается впечатление, что обучающиеся алгоритмы просто находят корреляции между двумя событиями, например запросом «лекарство от простуды» в строке Google и самой простудой. Однако нахождение корреляций для машинного обучения – не более чем кирпичи для дома. В горе кирпичей жить не получится.

Если новая технология пронизывает нашу жизнь до такой степени, как машинное обучение, нельзя, чтобы она оставалась для нас загадкой. Неясности создают благодатную почву для ошибок и неправильного применения. Алгоритм Amazon лучше, чем любой человек, умеет определять, какие книги читают сегодня в мире. Алгоритмы Агентства национальной безопасности способны узнать в человеке потенциального террориста. Моделирование климата находит безопасный уровень углекислого газа в атмосфере, а модели подбора акций больше вкладывают в развитие экономики, чем большинство из нас. Но нельзя контролировать то, чего не понимаешь, и именно поэтому вы должны понимать машинное обучение – как гражданин, как специалист и как человек, стремящийся к счастью.

Первейшая задача этой книги – посвятить вас в секреты машинного обучения. Разбираться в автомобильном двигателе нужно только инженерам и механикам, однако любой водитель должен знать, что поворот руля меняет направление движения, а если нажать на тормоз, машина остановится. Сегодня лишь немногие имеют представление об обучающихся алгоритмах хотя бы на таком уровне, не говоря уже об умении ими пользоваться. Психолог Дональд Норман придумал термин «концептуальная модель»: это грубое знание какой-либо технологии, достаточное для того, чтобы эффективно ею пользоваться. Эта книга даст вам концептуальную модель машинного обучения.

Не все обучающиеся алгоритмы работают одинаково, и это имеет определенные последствия. Возьмем, например, системы рекомендаций Amazon и Netflix и прогуляемся с ними по обычному книжному магазину. Пытаясь найти книги, которые «точно вам понравятся», Amazon, скорее всего, подведет вас к полке, к которой вы в прошлом чаще подходили, а Netflix позовет вас в незнакомый и неочевидный на первый взгляд уголок, но то, что вы там найдете, обязательно вам понравится. Из этой книги вы узнаете, что у Amazon и Netflix просто разные типы алгоритмов. Алгоритм Netflix вникает в ваши вкусы глубже (хотя все еще довольно скромно), однако, как ни странно, это еще не значит, что Amazon выиграла бы от такого подхода. Дело в том, что для успешного развития бизнеса Netflix нужно направлять спрос к длинному шлейфу малоизвестных и поэтому недорогих фильмов и телешоу и отвлекать клиентов от блокбастеров, на оплату которых абонента просто не хватит. У мене-

⁵ Имеется в виду суперкомпьютер IBM, оснащенный системой искусственного интеллекта, который был создан группой исследователей под руководством Дэвида Феруччи. В феврале 2011 года для проверки возможностей Watson он принял участие в телешоу Jeopardy!. Его соперниками были Брэд Раттер – обладатель самого большого выигрыша в программе, – и Кен Дженингс – рекордсмен по длительности беспроигрышной серии. Watson одержал победу, получив миллион долларов, в то время как Дженингс и Раттер получили по 300 и 200 тысяч соответственно.

джерсов Amazon такой проблемы нет: им тоже выгодно сбыть неходовые товары, но продавать популярные и дорогие варианты не менее приятно (к тому же это упрощает логистику). Кроме того, клиенты с большей вероятностью посмотрят что-то необычное по подписке, чем купят специально.

Каждый год в мире появляются сотни новых алгоритмов с обучением, но все они основаны на небольшом наборе фундаментальных идей. Именно этим идеям и посвящена эта книга, и их вам будет вполне достаточно, чтобы понять, как машинное обучение меняет наш мир. Не уходя в дебри и даже не очень затрагивая применение алгоритмов в компьютерах, мы дадим ответы на важные для всех нас вопросы: «Как мы учимся?», «Можно ли учиться эффективнее?», «Что мы способны предсказать?», «Можно ли доверять полученному знанию?» Соперничающие школы машинного обучения отвечают на эти вопросы по-разному. Всего существует пять основных научных течений, каждому из которых мы посвятим отдельную главу. Символисты рассматривают обучение как процесс, обратный дедукции, и черпают идеи из философии, психологии и логики. Коннекционисты⁶ воссоздают мозг путем обратной инженерии и вдохновляются нейробиологией и физикой. Эволюционисты симулируют эволюцию на компьютерах и обращаются к генетике и эволюционной биологии. Сторонники байесовского подхода⁷ полагают, что обучение – это разновидность вероятностного вывода, и корни этой школы уходят в статистику. Аналогисты занимаются экстраполяцией на основе схожести суждений и находятся под влиянием психологии и математической оптимизации. Стремясь построить обучающиеся машины, мы пройдемся по истории мысли за последнюю сотню лет и увидим ее в новом свете.

У каждого из пяти «племен» машинного обучения есть собственный универсальный обучающийся – Верховный – алгоритм, который в принципе можно использовать для извлечения знания из данных в любой области. Для символистов это обратная дедукция, для коннекционистов – обратное распространение ошибки, для эволюционистов – генетическое программирование, для байесовцев – байесовский вывод, а для аналогистов – метод опорных векторов. Однако на практике каждый из этих алгоритмов хорош для одних задач, но не очень подходит для других. Хотелось бы, чтобы все их черты слились воедино в окончательном, совершенном Верховном алгоритме. Кто-то считает это несбыточной мечтой, но у многих из нас – людей, занимающихся машинным обучением, – при этих словах загораются глаза, и мечта заставляет нас работать до поздней ночи.

Верховный алгоритм сумеет извлечь из данных вообще все знание – знание прошлого, настоящего и будущего. Изобретение этого алгоритма станет одним из величайших прорывов в истории науки. Оно ускорит прогресс буквально во всем, изменит мир так, как мы едва можем себе сегодня представить. Верховный алгоритм для машинного обучения – это нечто вроде стандартной модели в физике элементарных частиц и центральной догмы молекулярной биологии: единая теория, объясняющая все, что мы сегодня знаем, и закладывающая фундамент десятилетий или целых веков будущего прогресса. Верховный алгоритм – ключ к решению стоящих перед человечеством сложнейших проблем – от создания домашних роботов до лечения рака.

Ведь рак так сложно лечить, потому что это не одно, а целый комплекс заболеваний. Опухоли бывают вызваны невообразимо широким спектром причин, к тому же они мутируют и дают метастазы. Самый надежный способ уничтожить опухоль – секвенировать⁸ ее

⁶ Один из подходов в области искусственного интеллекта, когнитивной науки (когнитивистики), нейробиологии, психологии и философии сознания.

⁷ Берет свои истоки из теоремы Байеса, одной из основных теорем элементарной теории вероятностей, и названа в честь Томаса Байеса (1702–1761) – английского математика и священника, который первым предложил использование теоремы для корректировки убеждений, основываясь на обновленных данных.

⁸ Секвенирование (от лат. sequentum – последовательность) – определение аминокислотной или нуклеотидной последовательности.

геном, определить, какие лекарства помогут без ущерба для пациента с учетом *конкретного* генома и истории болезни, и, возможно, даже разработать новое лекарство именно для данного случая. Ни один врач не в состоянии овладеть всеми необходимыми для этого знаниями, но решение таких задач – идеальная работа для машинного обучения. В сущности, это просто более сложная и серьезная версия поиска, которым каждый день занимаются Amazon и Netflix, только ищем мы не подходящую книгу или фильм, а подходящее лекарство. К сожалению, хотя обучающиеся алгоритмы уже умеют со сверхчеловеческой точностью диагностировать многие болезни, лечение рака выходит далеко за пределы их возможностей. Если нам удастся отыскать Верховный алгоритм, ситуация изменится. Поэтому вторая цель этой книги – помочь *вам самостоятельно* изобрести его. Можно подумать, что для этого нужны глубочайшие познания в математике и серьезная теоретическая работа. Отнюдь нет. Для этого нужно как раз отвлечься от тайн математики и посмотреть на всеобъемлющие механизмы обучения, и здесь неспециалист, подходящий к лесу издалека, во многом находится в более выгодном положении, чем профессионал, увязнувший в изучении отдельных деревьев. Концептуальное решение проблемы можно дополнить математикой, но это не самое главное, и к тому же не тема этой книги. Так что, когда мы будем заходить в гости к каждому «племени», надо будет собрать кусочки мозаики и сообразить, куда они подходят, не забывая при этом, что ни один слепец не может увидеть слона целиком. Мы увидим, какой вклад каждое из «племен» способно внести в лечение рака, чего ему не хватает, а затем шаг за шагом соберем кусочки в решение – вернее, *одно из* решений. Конечно, это не будет Верховным алгоритмом, но ближе к нему еще никто не подбирался. Будем надеяться, что результат станет удачной отправной точкой для вашего воображения. Потом мы посмотрим, как можно было бы использовать полученный алгоритм в качестве оружия в борьбе с раком. Читая эту книгу, не стесняйтесь пробегать глазами или пропускать сложные для понимания фрагменты. Важна общая картина, и, возможно, будет полезнее вернуться к этим местам уже после того, как мозаика сложится.

Я занимаюсь исследованиями машинного обучения более 20 лет. Интерес к этой теме во мне пробудила книга со странным названием, которую на последнем курсе колледжа я заметил в книжном магазине. Она называлась «Искусственный интеллект». Машинному обучению в ней была посвящена одна короткая глава, но, прочитав ее, я немедленно пришел к убеждению, что в ней – ключ к искусственному интеллекту, что знаем мы об этой области так немного и что, может быть, я смогу внести свой вклад. Поэтому я распрощался с планами получить степень MBA и поступил в аспирантуру в Калифорнийском университете в Ирвайне. Машинное обучение было тогда второстепенной дисциплиной, а в Калифорнийском университете работала одна из немногих приличных исследовательских групп. Некоторые из моих однокурсников ушли, сочтя тему бесперспективной, но я не сдался. Для меня не было ничего важнее, чем научить компьютеры учиться – ведь если удастся это сделать, мы получим фору в решении любой другой проблемы. Прошло пять лет. Я заканчивал аспирантуру, а на дворе царила революция добычи данных. Диссертацию я посвятил объединению подхода символистов и аналогистов, большую часть последнего десятилетия соединял символизм и байесовский подход, а в последнее время – оба этих метода с коннекционизмом. Теперь пора сделать следующий шаг и попытаться свести воедино все пять парадигм.

Работая над этой книгой, я представлял себе несколько разных, но пересекающихся групп читателей.

Если вам просто любопытно, откуда столько шума вокруг больших данных и машинного обучения, и вы подозреваете, что тут все не так просто, как пишут в газетах, вы правы! Эта книга станет для вас своеобразным путеводителем.

Если вы интересуетесь прежде всего применением машинного обучения в бизнесе, она поможет вам 1) стать более разборчивым потребителем аналитики; 2) получить максимальную отдачу от своих специалистов по обработке и анализу информации; 3) избежать ловушек, которые убили столь многие проекты по добыче данных; 4) узнать, какие области можно автоматизировать без затрат на ручное кодирование программ; 5) уменьшить жесткость своих информационных систем и 6) предвидеть появление ряда новых технологий, которые уже не за горами. Я много раз наблюдал, как деньги и время уходят впустую из-за того, что проблемы решаются неправильным обучающимся алгоритмом, и как неверно интерпретируют то, что алгоритм сообщает. Чтобы избежать фиаско, достаточно лишь прочитать эту книгу.

Если вы сознательный гражданин или если вы отвечаете за решение социальных, государственных вопросов, возникших после появления больших данных и машинного обучения, эта книга станет для вас азбукой новой технологии. Не углубляясь в скучные подробности, вы узнаете, что эта технология собой представляет, к чему может привести, что она умеет, а чего нет. Вы увидите, в чем состоят реальные проблемы – от защиты частной жизни до рабочих мест в будущем и этики боевых роботов – и как к ним подступиться.

Если вы ученый или инженер, машинное обучение – мощнейший инструмент, который очень вам пригодится. Старые, проверенные временем статистические подходы не спасут вас в век больших (и даже средних) объемов данных. Для точного моделирования большинства явлений понадобятся нелинейные подходы машинного обучения, и оно несет с собой новое научное мировоззрение. В наши дни любят злоупотреблять выражением «смена парадигмы», но я уверен, что тема моей книги именно так и звучит.

Даже если вы эксперт по машинному обучению и уже знакомы с большинством освещаемых мной тем, вы найдете в книге много свежих идей, экскурсов в историю, полезных примеров и аналогий. Я очень надеюсь, что это поможет вам по-новому взглянуть на машинное обучение и, может быть, даже по-новому направит ваши мысли. Полезно сорвать то, что висит на нижних ветках, однако не стоит терять из виду и то, что расположено чуть выше. (Кстати, прошу простить мне поэтическую вольность – эпитет «Верховный» в отношении универсального обучающегося алгоритма.)

Если вы учащийся любого возраста: старшеклассник, выбирающий, где учиться, студент старших курсов, размышляющий, идти ли в науку, или маститый ученый, планирующий изменение карьеры, моя книга, надеюсь, заронит в вас искорку интереса к этой захватывающей области знания. В мире остро не хватает специалистов по машинному обучению, и, если вы решите к нам присоединиться, можете быть уверены: вас ждут не только захватывающие мгновения и материальные блага, но и уникальный шанс послужить обществу. А если вы уже осваиваете машинное обучение, надеюсь, эта книга поможет вам лучше сориентироваться в теме. Если в своих поисках вы случайно наткнетесь на Верховный алгоритм, ради этого мне стоило браться за перо.

И последнее, но не менее важное. Если вы хотите ощутить вкус чуда, машинное обучение – настоящий пир для ума. Примите приглашение и угощайтесь!

Глава 1

Революция машинного обучения

Мы живем в эпоху алгоритмов. Всего поколение-другое назад слово «алгоритм» у большинства людей вызвало бы лишь непонимание. Сегодня алгоритмы проникли во все уголки нашей цивилизации. Они вшиты в ткань повседневной жизни и нашли себе место не только в мобильных телефонах и ноутбуках, но и в автомобилях, квартирах, бытовой технике и игрушках. Так, банк – гигантское хитросплетение алгоритмов, а люди просто слегка регулируют настройки то тут, то там. Алгоритмы составляют расписание полетов, а затем ведут самолеты. Алгоритмы управляют производством, торговлей, снабжением, подсчитывают выручку и занимаются бухгалтерией. Если все алгоритмы вдруг перестанут работать, настанет конец света – такого, каким мы его знаем.

Алгоритм – определенная последовательность инструкций, диктующая компьютеру его действия. Компьютеры состоят из миллиардов крохотных переключателей – транзисторов, и алгоритмы включают и выключают эти транзисторы миллиарды раз в секунду.

Самый простой алгоритм – «нажми переключатель». Положение одного транзистора – одна единица информации: «один», если транзистор включен, и «ноль», если выключен. Единичка где-то в компьютерах банка информирует, превысили ли вы кредит. Еще одна единичка в недрах Управления социального обеспечения сообщает, живы вы или уже умерли.

Второй простейший алгоритм – «соедини два бита». Клод Шеннон, признанный отец теории информации, первым осознал, что включение и выключение транзисторов в ответ на действия других транзисторов – это, в сущности, логический вывод. (Этой теме он посвятил свою дипломную работу в Массачусетском технологическом институте – самую важную дипломную работу в истории.) «Транзистор *A* включается, только если включены транзисторы *B* и *C*» – это крохотное логическое рассуждение. «*A* включается, когда включен либо *B*, либо *C*» – еще одна крупинка логики. «*A* включается всегда, когда выключен *B*, и наоборот» – третья операция. Хотите верить, хотите нет, любой алгоритм, как бы сложен он ни был, сводится всего к трем операциям: И, ИЛИ и НЕ. Используя для этих операций специальные символы, можно представить простые алгоритмы в виде диаграмм. Например, если у человека грипп или малярия и ему надо принять лекарство от температуры и головной боли, это можно выразить следующим образом:



Соединяя множество подобных операций, можно составлять очень сложные цепочки логических рассуждений. Люди часто думают, что вся суть компьютеров в вычислениях, но это не так. Сердце компьютеров – логика. Из логики в компьютере состоят и числа, и арифметика, и все остальное. Хотите сложить два числа? Есть комбинация транзисторов, которая это делает. Хотите победить чемпиона в «Своей игре»? Для этого тоже найдется комбинация (естественно, она будет намного больше).

Однако строить новый компьютер для каждой новой задачи, которая нам придет в голову, было бы невероятно дорого, поэтому современный компьютер представляет собой большую совокупность транзисторов, способных решать много разных задач в зависимости

от того, какие из них активны. Микеланджело говорил, что вся его работа – увидеть статую в глыбе мрамора и открыть ее миру, убрав лишнее. Аналогично алгоритмы «отсекают» избыточные транзисторы в компьютере, пока не обнажится нужная функция, будь то автопилот авиалайнера или новый мультфильм студии Pixar.

Алгоритм – не просто произвольный набор инструкций. Чтобы компьютер его выполнил, указания должны быть достаточно точными и однозначными. Например, кулинарный рецепт – это не алгоритм, потому что не задает однозначного порядка действий и не объясняет, что делать на каждом этапе. Например, сколько именно сахара уместится в столовую ложку? Любой человек, который хоть раз пробовал готовить по незнакомому рецепту, знает, что может получиться и восхитительное блюдо, и не пойми что. А алгоритмы всегда дают идентичный результат. При этом, даже если указать в рецепте ровно 15 граммов сахара, это по-прежнему не решает проблему, потому что компьютер не знает ни что такое сахар, ни что такое грамм. Если бы мы захотели запрограммировать робота-повара для выпечки тортов, пришлось бы научить его узнавать сахар на видео, научить брать ложку и так далее (ученые все еще над этим работают). Компьютер должен знать, как выполнять алгоритм – вплоть до включения и выключения конкретных транзисторов, поэтому рецепт готовки очень далек от алгоритма.

С другой стороны, вот вам алгоритм игры в крестики-нолики.

Если вы или ваш противник поставили две отметки на одной линии, ставьте отметку в оставшейся на этой линии клетке.

Если такой ход невозможен, но есть ход, который создаст две линии по две отметки, – делайте его.

Если такой ход невозможен, но центральная клетка свободна, ставьте отметку в ней.

Если такой ход невозможен, но противник поставил отметку в углу, ставьте отметку в противоположном углу.

Если такой ход невозможен, но одна из угловых клеток свободна, ставьте отметку в ней.

Если такой ход невозможен, ставьте отметку в любой пустой клетке.

У этого алгоритма есть одно приятное свойство: он беспробитный! Конечно, ему не хватает многих деталей – как доска отображается в памяти компьютера и как это отображение меняется после каждого хода. Например, каждой клетке могут соответствовать два бита: 00 – если клетка пуста, 01 – если в ней поставили нолик и 10 – если крестик. Тем не менее предложенный алгоритм достаточно точен и однозначен, и любой грамотный программист сможет его дописать. Еще полезно не конкретизировать алгоритмы вплоть до отдельных транзисторов, а пользоваться уже существующими алгоритмами как кирпичиками. Их огромное количество, так что есть из чего выбирать.

Алгоритмы предъявляют строгие требования: часто говорят, что по-настоящему понимаешь что-то только тогда, когда можешь выразить это в виде алгоритма (как заметил Ричард Фейнман⁹, «я не понимаю того, чего не могу создать»). Уравнения – хлеб насущный физиков и инженеров – на самом деле всего лишь особая разновидность алгоритмов. Например, второй закон Ньютона, который считают самым важным в мире уравнением, гласит, что для вычисления действующей на тело суммарной силы надо массу этого тела умножить на его ускорение. Он также подразумевает, что ускорение – это сила, разделенная на массу, но выводение этого следствия тоже алгоритм. Если теорию в любой научной дисциплине

⁹ Ричард Филлипс Фейнман (Richard Phillips Feynman, 1918–1988) – американский физик, лауреат Нобелевской премии, один из создателей современной квантовой электродинамики, внес существенный вклад в квантовую механику и квантовую теорию поля, его имя носит метод диаграмм Фейнмана.

не получается выразить в виде алгоритма, она недостаточно строгая, не говоря уже о том, что ее решение нельзя компьютеризировать, а это серьезно ограничивает сферу ее применения. Ученые строят теории, инженеры изобретают устройства, а специалисты в области информатики создают алгоритмы, которые представляют собой и теории, и устройства одновременно.

Написать алгоритм непросто: есть очень много ловушек, и ни в чем нельзя быть уверенным. Интуитивные предположения вполне могут оказаться ошибочными, и тогда придется искать другой подход. Затем алгоритм надо выразить на понятном компьютеру языке, например Java или Python, и с этого момента алгоритм начнет называться программой. Потом программу надо отладить: найти все до единой ошибки и исправить их, пока компьютер не начнет выполнять ее без запинки. Но когда у вас наконец появится программа, которая умеет делать то, что вам нужно, вы получите все козыри. Компьютер станет послушно выполнять ваши задания миллионы раз со сверхвысокой скоростью. Созданной вами программой сможет пользоваться любой человек в мире. Она даже сделает вас миллиардером, если решенная проблема достаточно важна. Программист – человек, пишущий алгоритмы и кодирующий их, – маленький бог, создающий вселенные по своему желанию. Можно даже сказать, что сам Господь тоже был программистом, ведь в Книге Бытия он творил с помощью слов, а не руками. Речения стали мирами. Сегодня, сидя в кресле перед ноутбуком, вы тоже можете почувствовать себя богом: нарисуйте в воображении Вселенную и сделайте ее реальной. Законы физики соблюдать необязательно.

Со временем информатики начинают опираться на уже проделанную работу и придумывают алгоритмы для все новых процессов. Одни алгоритмы соединяются с другими, чтобы использовать результаты третьих, производя, в свою очередь, еще больше алгоритмов. Каждую секунду миллиарды раз переключаются миллиарды транзисторов в миллиардах компьютеров. Алгоритмы образуют экосистему нового типа – непрерывно растущую и сопоставимую по богатству лишь с самой жизнью.

Однако, как это всегда бывает, в райском саду обитает змей – Монстр Сложности. У него, как у лернейской гидры, много голов. Одна из них – пространственная: количество битов информации, которое алгоритм должен хранить в памяти компьютера. Если алгоритму требуется больше памяти, чем есть в наличии, он бесполезен, и его приходится отбрасывать. У пространственной сложности есть злая сестрица: временная сложность. Сколько будет длиться выполнение алгоритма, то есть сколько раз нужно использовать транзисторы, прежде чем алгоритм даст желаемый результат? Если мы не можем столько ждать, алгоритм снова оказывается бесполезным. Но самая пугающая голова Монстра Сложности – сложность человеческая. Когда алгоритм становится слишком запутанным и непонятным для нашего скромного разума, а взаимодействия между его элементами – слишком многочисленными и обширными, в него начинают вкрадываться ошибки. Человек не в состоянии их отыскать и исправить, поэтому алгоритм не делает то, что от него требуется. Даже если каким-то образом заставить его работать, он окажется неоправданно сложным для пользователя, будет плохо взаимодействовать с другими алгоритмами и порождать все больше проблем.

Специалисты-информатики сражаются с Монстром Сложности каждый день. Когда они проигрывают, сложность прорывается в нашу жизнь. Вы, наверное, и сами замечали, как много было проиграно битв. Тем не менее башня алгоритмов продолжает расти, хотя строить ее все труднее: каждое новое поколение алгоритмов приходится возводить на вершине предшественников, их сложность суммируется. Башня растет и растет, алгоритмы опутывают весь мир, но конструкция становится все более шаткой – как картонный домик, который только и ждет толчка. Мизерная ошибка в алгоритме – и ракета, стоившая миллиард

долларов, взрывается на старте, или миллионы людей остаются без электричества. Непредвиденное взаимодействие алгоритмов – и рушится фондовый рынок.

Если программисты – маленькие боги, то Монстр Сложности – его величество Сатана. И мало-помалу он выигрывает войну.

Должен быть способ лучше.

Познакомимся с обучающимся алгоритмом

У любого алгоритма есть вход и выход: данные поступают в компьютер, алгоритм делает с ними то, что должен, и выдает результат. Машинное обучение переворачивает все задом наперед: имея в своем распоряжении данные и желаемый результат, оно выдает алгоритм, превращающий одно в другое. Обучающиеся алгоритмы – те, что создают другие алгоритмы, обученные на основе данных. С помощью машинного обучения компьютеры пишут себе программы, и нам не надо этим заниматься.

Здорово, правда?

Компьютеры сами пишут для себя программы. Эта мысль потрясает настолько, что даже страшно: если компьютеры начнут программировать сами себя, сможем ли мы их контролировать? Оказывается – и мы в этом убедимся, – людям вполне по силам с ними совладать. Но есть и другое возражение – все это слишком хорошо, чтобы быть правдой. Разве для написания алгоритмов не нужны ум, творческая жилка, умение решать проблемы – все те качества, которых у компьютеров просто нет? Чем машинное обучение отличается от магии? Все это правда: сегодня мы умеем писать много программ, которым компьютер научиться не может. Но еще удивительнее то, что и компьютеры могут научиться программам, которые не в состоянии написать человек. Мы умеем водить машину или читать написанный от руки текст, но эти навыки у нас подсознательные: рассказать компьютеру, как это делать, не получится. Однако если дать обучающемуся алгоритму достаточное количество примеров каждого из этих действий, он с легкостью во всем разберется и без нашей помощи, и тогда можно будет развязать ему руки. Именно так машины научились читать почтовые индексы, и именно поэтому на дорогах скоро появятся автомобили без водителей.

Мощь машинного обучения, наверное, лучше всего показать, сравнив технологию с сельским хозяйством. В индустриальном обществе товары делают на заводах, а это значит, что инженерам надо точно определить, как именно их собирать, как изготавливать все элементы и так далее, вплоть до сырья. Это требует больших усилий. Самые сложные устройства, которые человеку удалось изобрести, – компьютеры, и их разработка, производство и написание для них программ требуют колоссального труда. Но есть другой, намного более древний способ получить некоторые необходимые нам вещи: предоставить их изготовление самой природе. Посадить семечко, полить его, добавить удобрений, а потом сорвать спелый плод. Может ли технология выглядеть примерно так же? Может! Именно это сулит нам машинное обучение. Обучающиеся алгоритмы – как семена, почва – это данные, а обученные программы – это наша жатва. Эксперт по машинному обучению похож на крестьянина, сеющего, поливающего и удобряющего землю. Он присматривает за здоровьем растущего урожая, но в целом не вмешивается.

Если посмотреть на машинное обучение под этим углом, сразу бросаются в глаза два момента. Во-первых, чем больше у нас данных, тем больше мы можем узнать. Нет данных? Тогда и учиться нечему. Большой объем информации? Огромное поле для обучения. Вот почему машинное обучение заявляет о себе везде, где появляются экспоненциально растущие горы данных. Если бы в магазине продавали машинное обучение быстрого приготовления, на коробке было бы написано: «Просто добавь данных».

Второе наблюдение заключается в том, что машинное обучение – это меч-кладенец, которым можно обезглавить Монстра Сложности. Если дать обучающей программе длинной всего пару сотен строк достаточно данных, она не только с легкостью сгенерирует программу из миллионов строк кода, но и сможет делать это вновь и вновь для разных проблем. Уменьшение сложности для программиста просто феноменальное. Конечно, как и гидра,

Монстр Сложности будет отращивать все новые и новые головы, но они окажутся меньше и вырастут не сразу, так что у нас все равно будет большое преимущество.

Машинное обучение можно представить себе как вывернутое наизнанку программирование, точно так же как квадратный корень противоположен возведению во вторую степень, а интегрирование обратно дифференцированию. Если можно спросить, квадрат какого числа равен 16 или производной какой функции является $x + 1$, уместен и вопрос: «Какой алгоритм даст такой результат?» Вскоре мы увидим, как превратить оба наблюдения в конкретные обучающиеся алгоритмы.

Некоторые обучающиеся алгоритмы добывают знания, а некоторые – навыки. «Все люди смертны» – это знание. Езда на велосипеде – навык. В машинном обучении знание часто предстает в форме статистических моделей, потому что знание как таковое – это во многом статистика: смертны все люди, но только четыре процента людей американцы. Навыки зачастую представляют собой наборы процедур: если дорога сворачивает влево, поверни руль влево. Если перед тобой выскочил олень, дави на тормоз. (К сожалению, на момент написания этой книги беспилотная машина Google все еще путает оленей с полиэтиленовыми пакетами.) Часто процедура довольно проста, хотя заложенное в ней знание сложно. Спам надо отправить в корзину, однако сначала придется научиться отличать его от обычных писем. Если разобраться, какая позиция на шахматной доске удачна, станет ясно, какой сделать ход (тот, что приведет к лучшей позиции).

Машинное обучение принимает много разных форм и скрывается под разными именами: распознавание паттернов, статистическое моделирование, добыча данных, выявление знаний, предсказательная аналитика, наука о данных, адаптивные и самоорганизующиеся системы и так далее. Все они находят свое применение и имеют разные ассоциации. Некоторые живут долго, а некоторые не очень. Все это многообразие я буду называть просто – *машинное обучение*.

Машинное обучение иногда путают с искусственным интеллектом. С формальной точки зрения это действительно подраздел науки об искусственном интеллекте, однако он очень разросся и оказался настолько успешным, что затмил гордого родителя. Цель искусственного интеллекта – научить компьютеры делать то, что люди пока делают лучше, а умение учиться – наверное, самый важный из этих навыков, без которого компьютерам никогда не угнаться за человеком. Остальное приложится.

Если представить обработку данных в виде экосистемы, обучающиеся алгоритмы будут в ней суперхищниками. Базы данных, поисковые роботы, индексаторы и так далее – это травоядные, мирно пасущиеся на бескрайних лугах данных. Статистические алгоритмы, оперативная аналитическая обработка и так далее – просто хищники. Без травоядных не обойтись, потому что без них все остальное бы умерло, однако у суперхищника жизнь интереснее. Поисковый робот, как корова, пасется в интернете – поле мирового масштаба, а каждая страница в нем – травинка. Робот пощипывает травку, копии страниц оседают на его жестком диске. Затем индексатор создает список страниц, где встречается каждое слово, во многом как предметный указатель в конце книги. Базы данных похожи на слонов: они большие, тяжелые и никогда ни о чем не забывают. Среди этих степенных животных носятся статистические и аналитические алгоритмы, которые сжимают, выбирают и превращают данные в информацию. Обучающиеся алгоритмы поглощают эту информацию, переваривают ее и дают нам знание.

Эксперты по машинному обучению – элита, каста священников среди ученых-информатиков. Многие компьютерщики, особенно старшего поколения, понимают машинное обучение не так хорошо, как им хотелось бы. Дело в том, что компьютерные науки традиционно следовали в русле детерминизма, а в машинном обучении нужно мыслить в категориях статистики. Если какое-то правило, скажем, отмечать определенные письма как спам, срабаты-

вайт в 99, а не в 100 процентах случаев, это не значит, что в нем какая-то ошибка: может быть, это лучшее, что можно сделать, и даже такая точность очень полезна. Различия в стиле мышления во многом послужили причиной, по которой Microsoft оказалось намного сложнее нагнать Google, чем в свое время Netscape. В конце концов, браузер всего лишь стандартная программа, а вот поисковая система требует другого склада ума.

Еще одна причина, по которой эксперты по машинному обучению слывут сверхумниками, заключается в том, что в мире их намного меньше, чем надо, даже по меркам компьютерных наук. Тим О'Райли, гуру в области технологий, утверждает, что «специалист по обработке данных» – самая востребованная вакансия в Кремниевой долине. По оценке McKinsey Global Institute, в 2018 году в одних только Соединенных Штатах спрос на экспертов по машинному обучению будет превышать предложение на 140–190 тысяч человек. Кроме того, потребуется дополнительно полтора миллиона разбирающихся в данных управленцев. Поток программ, связанных с машинным обучением, оказался слишком внезапным и мощным – система образования просто не успевает за спросом, к тому же машинное обучение считается трудной специальностью, и учебники вполне могут вызвать неприятие математики. Однако сложность скорее мнимая: все важнейшие идеи машинного обучения можно выразить и без математики. Читая эту книгу, вы, может быть, даже поймаете себя на том, что изобретаете обучающиеся алгоритмы без всяких уравнений.

Промышленная революция автоматизировала ручной труд, информационная революция проделала то же с трудом умственным, а машинное обучение автоматизировало саму автоматизацию. Без него программирование стало бы узким горлом, сдерживающим прогресс. Если вы ленивый и не слишком сообразительный компьютерщик, машинное обучение для вас – идеальная специальность, потому что обучающиеся алгоритмы сделают всю работу сами, а вам достанутся только лавры. С другой стороны, обучающиеся алгоритмы могут оставить нас без работы, и поделом.

Подняв автоматизацию на невиданные высоты, революция машинного обучения вызовет огромные изменения в экономике и обществе, как в свое время интернет, персональные компьютеры, автомобили и паровой двигатель. Одна из областей, где изменения уже очевидны, – бизнес.

Почему бизнес рад машинному обучению?

Почему Google стоит намного дороже Yahoo? Обе компании зарабатывают на показе рекламы в интернете, и у той, и у другой прекрасная посещаемость, обе проводят аукционы по продаже рекламы и используют машинное обучение, чтобы предсказать, с какой вероятностью пользователь на нее кликнет (чем выше вероятность, тем ценнее реклама). Дело, однако, в том, что обучающиеся алгоритмы у Google намного совершеннее, чем у Yahoo. Конечно, это не единственная весьма серьезная причина разницы в капитализации. Каждый предсказанный, но не сделанный клик – упущенная возможность для рекламодателя и потерянная прибыль для поисковика. Учитывая, что годовая выручка Google составляет 50 миллиардов долларов, улучшение прогнозирования всего на один процент потенциально означает еще полмиллиарда долларов в год на банковском счету. Неудивительно, что Google – большая поклонница машинного обучения, а Yahoo и другие конкуренты из всех сил пытаются за ней угнаться.

Реклама в сети – всего лишь один из аспектов более широкого явления. На любом рынке производители и потребители перед тем, как заключить сделку, должны выйти друг на друга. До появления интернета основные препятствия между ними были физическими: книгу можно было купить только в книжном магазине поблизости, а полки там не безразмерные. Однако теперь, когда книги можно в любой момент скачать на «читалку», проблемой становится колоссальное число вариантов. Как тут искать, если на полках книжного магазина стоят миллионы томов? Это верно и для других информационных продуктов: видео, музыки, новостей, твитов, блогов, старых добрых сайтов. Это также касается продуктов и услуг, которые можно получить на расстоянии: обуви, цветов, гаджетов, гостиничных номеров, обучения, инвестиций и даже поисков работы и спутника жизни. Как найти друг друга? Это определяющая проблема информационной эры, и машинное обучение помогает ее решить.

В процессе развития компании можно выделить три фазы. Сначала все делается вручную: владельцы семейного магазинчика знают своих клиентов лично и в соответствии с этим заказывают, выставляют и рекомендуют товары. Это мило, но не позволяет увеличить масштаб. На втором, и самом неприятном, этапе компания вырастает настолько, что возникает необходимость пользоваться компьютерами. Появляются программисты, консультанты, менеджеры баз данных, пишутся миллионы строк кода, чтобы автоматизировать все, что только можно. Компания начинает обслуживать намного больше людей, однако качество падает: решения принимаются на основе грубой демографической классификации, а компьютерные программы недостаточно эластичны, чтобы подстроиться под бесконечную изменчивость человечества.

В какой-то момент программистов и консультантов начинает просто не хватать, и компания неизбежно обращается к машинному обучению. Amazon не может изящно заложить в компьютерную программу вкусы всех своих клиентов, а Facebook не смогла бы написать программу, чтобы выбрать обновления, которые понравятся каждому из пользователей. Walmart ежедневно продает миллионы продуктов. Если бы программисты этой торговой сети попытались создать программу, способную делать миллионы выборов, они бы работали целую вечность. Вместо этого компании спускают с цепи обучающиеся алгоритмы, науськивают их на уже накопленные горы данных и дают им предсказать, чего хотят клиенты.

Алгоритмы машинного обучения пробиваются через информационные завалы и, как свахи, находят производителей и потребителей друг для друга. Если алгоритмы достаточно умны, они объединяют лучшее из двух миров: широкий выбор, низкие затраты огромной корпорации и индивидуальный подход маленькой компании. Обучающиеся алгоритмы

не идеальны, и последний шаг в принятии решения все равно остается за человеком, но они разумно сужают выбор, чтобы человеку было под силу принять решение.

Сегодня очевидно, что переход от компьютеров к интернету, а затем к машинному обучению был неизбежен. Компьютеры сделали возможным интернет, тот породил поток данных и проблему безграничного выбора, а машинное обучение использует потоки данных, чтобы решить проблему безграничного выбора. Чтобы сдвинуть спрос от «одного размера на всех» до длинного, бесконечно разнообразного списка вариантов, одного интернета мало. У Netflix может быть хоть сто тысяч разных DVD-дисков, но, если клиент не знает, как найти то, что ему понравится, он будет по умолчанию выбирать хиты. И только когда Netflix обзавелся обучающимся алгоритмом, который угадывает ваши вкусы и советует музыку, длинный хвост менее популярных исполнителей «взлетел».

Когда-нибудь произойдет неизбежное: обучающиеся алгоритмы станут незаменимым посредником и в них сосредоточится власть. Алгоритмы Google во многом определяют, какую информацию вы видите, Amazon – какие продукты вы покупаете, а Match.com – с кем вы станете встречаться. Последний этап – выбрать из предложенных алгоритмом вариантов – все равно придется преодолеть вам, однако 99,9 процента отбора будет проходить без вашего участия. Успех или неудача компании станет зависеть от того, будут ли алгоритмы машинного обучения предпочитать ее продукцию. Успех экономики в целом, то есть получают ли все игроки нужные продукты по лучшей цене, будет зависеть от того, насколько хороши обучающиеся алгоритмы.

Лучший способ гарантировать, что алгоритмы машинного обучения станут отдавать предпочтение продукции вашей компании, – применять их. Победит тот, у кого лучше алгоритмы и больше данных. Здесь проявляется новый сетевой эффект: тот, у кого больше клиентов, собирает больше информации, лучше обучает модели, завоевывает новых клиентов и так далее по спирали (а с точки зрения конкурентов – по порочному кругу). Перейти с Google на Bing, может быть, даже проще, чем с Windows на Mac OS, но на практике вы этого не сделаете, потому что благодаря удачному старту и большей доле на рынке Google лучше знает, чего вы хотите, даже если непосредственно технологии у Bing не хуже. Новичкам на рынке поисковиков можно только посочувствовать: не имея данных, они вынуждены бороться против систем, которые обучают свои алгоритмы более десятка лет.

Можно подумать, что в какой-то момент данные просто начнут повторяться, однако точки насыщения не видно, и «длинный хвост» продолжает тянуться. Вы, конечно, и сами видите: рекомендации Amazon или Netflix пока еще очень грубы, а результаты, которые выдает Google, оставляют желать много лучшего. С помощью машинного обучения можно улучшить каждое свойство продукта, каждый уголок сайта. Ссылку внизу страницы лучше сделать красной или голубой? Попробуйте оба варианта и посмотрите, какой соберет больше кликов. А еще лучше вообще не выключать обучающиеся алгоритмы и постоянно корректировать все элементы сайта.

Та же динамика наблюдается на любом рынке, где имеется много вариантов и огромный объем данных. Гонка в разгаре, и побеждает тот, кто учится быстрее. Дело не только в лучшем понимании клиента: компании могут применять машинное обучение к каждому аспекту своей деятельности при условии, что на эту тему есть данные, а источники данных – компьютеры, устройства связи и все более дешевые и вездесущие сенсоры. Сейчас любят повторять, что «данные – это новая нефть» и, как и с нефтью, переработка – большой бизнес. IBM, как и все остальные корпорации, построила свою стратегию роста на предоставлении аналитических услуг компаниям. Бизнес видит в данных стратегический ресурс: что есть у нас, но отсутствует у конкурентов? Как воспользоваться этим преимуществом? А какие данные есть у конкурентов, но нет у нас?

Как банк, не располагающий базами данных, не может тягаться с банком, их имеющим, так и компания, не применяющая машинное обучение, не сможет соперничать с теми, кто его использует. Пока в первой компании будут писать тысячи правил для прогнозирования пожеланий покупателей, алгоритмы второй компании найдут миллиарды правил, по целому набору для каждого отдельного клиента. Такая конкуренция напоминает атаку с копиями на пулеметы. Конечно, машинное обучение – крутая новая технология, но для бизнеса дело даже не в этом: ее придется применять, потому что другого выбора просто нет.

Турбоускорение для научного метода

Машинное обучение – все равно что научный метод с допингом. Оно следует той же схеме обобщения, проверки, исключения и уточнения гипотез, однако ученый может за свою жизнь придумать и протестировать несколько сотен предположений, а система машинного обучения проделает то же самое в долю секунды. Машинное обучение ставит открытия на поток, поэтому неудивительно, что в науке оно производит революцию, во многом подобную революции в бизнесе.

Чтобы развиваться, любая область науки нуждается в данных, соизмеримых по сложности с явлениями, которые она изучает. Именно поэтому физика первой пошла вперед: записей Тихо Браге о положении планет и наблюдений Галилея за маятником и наклонными плоскостями оказалось достаточно, чтобы сформулировать законы Ньютона. По той же причине молекулярная биология обогнала более старую нейробиологию: ДНК-микрочипы и высокоэффективное секвенирование дают столько данных, сколько нейробиологам и не снилось. Социальные науки находятся в этом отношении в невыгодном положении: с выборкой всего лишь в сотню человек по десятку измерений на каждого смоделировать получается лишь очень узкие явления. Но даже такие небольшие феномены не существуют в изоляции: на них влияют мириады факторов, а это значит, что ученые очень далеки от того, чтобы их понять.

Хорошая новость: сегодня даже науки, некогда оперировавшие небольшими объемами информации, получили приток данных. Вместо того чтобы платить 50 студентам, которые будут клевать носом в лаборатории психолога, можно получить сколько угодно испытуемых, дав задание краудсорсинговой площадке Amazon Mechanical Turk (к тому же выборка окажется более разнообразной). Сейчас уже не все помнят, как немногим более десятилетия назад социологи, изучавшие социальные сети, жаловались, что не могут найти такую сеть, в которой было бы больше нескольких сотен участников. Теперь в их распоряжении весь Facebook, где больше миллиарда пользователей рассказывают о своей жизни во всех подробностях – чем не прямая трансляция общественной жизни на планете Земля? Коннектомика¹⁰ и функциональная магнитно-резонансная томография распахнули перед нейробиологами окно, через которое прекрасно виден головной мозг. В молекулярной биологии экспоненциально растут базы данных генов и белков. Даже «старые» дисциплины, например физика и астрономия, не стоят на месте благодаря потокам данных, льющимся из ускорителей частиц и цифрового исследования неба.

Однако от больших данных нет пользы, если их нельзя превратить в знание, и в мире слишком мало ученых, чтобы справиться с этой задачей. В свое время Эдвин Хаббл¹¹ открывал новые галактики, скрупулезно изучая фотографические пластинки, но можно ручаться, что таким способом не получилось бы найти полмиллиарда небесных тел, которые нам подарил проект Digital Sky Survey, – это было бы подобно ручному подсчету песчинок на пляже. Конечно, можно вручную написать правила, чтобы отличить галактики от звезд и шумов (например, птиц, самолетов или пролетающего мимо Супермена), но они будут не очень точными. Поэтому в проекте SKICAT, посвященном анализу и каталогизации изображений неба, был применен обучающийся алгоритм. Получив пластинки, где объектам уже были присвоены правильные категории, он разобрался, что характеризует каждую из них,

¹⁰ Область исследований, включающая в себя картографирование и анализ архитектуры нейрональных связей.

¹¹ Эдвин Пауэлл Хаббл (Edwin Powell Hubble, 1889–1953) – один из наиболее влиятельных астрономов и космологов XX века, внесший решающий вклад в понимание структуры космоса. Член Национальной академии наук в Вашингтоне с 1927 года.

а затем применил результаты ко всем необозначенным пластинкам. Эффективность превзошла все ожидания: алгоритм сумел классифицировать объекты настолько слабые, что человек не смог бы их выявить, и таких оказалось больше всего.

Благодаря большим данным и машинному обучению можно понять намного более сложные феномены, чем до появления этих факторов. В большинстве дисциплин ученые традиционно пользовались только очень скромными моделями, например линейной регрессией, где кривая, подобранная к данным, – всегда прямая линия. К сожалению (а может, и к счастью, потому что иначе жизнь была бы очень скучной – вообще говоря, никакой жизни бы и не было), большинство феноменов в мире нелинейны, и машинное обучение открывает перед нами огромный мир нелинейных моделей: это все равно что включить свет в комнате, которую до того освещала лишь Луна.

В биологии алгоритмы машинного обучения разбираются, где в молекуле ДНК расположены гены, какие фрагменты РНК вырезают при сплайсинге¹² перед синтезом белка, как белки принимают характерную для них форму и как заболевания влияют на экспрессию разных генов. Вместо того чтобы тестировать в лаборатории тысячи новых лекарств, обучающийся алгоритм прогнозирует, будут ли они эффективны, и допустит до этапа тестирования только самые перспективные. Алгоритмы будут отсеивать молекулы, которые, скорее всего, вызовут неприятные побочные эффекты, например рак. Это позволит избежать дорогих ошибок, к примеру, когда лекарство запрещают только после начала испытаний на человеке.

Однако самый большой вызов – это собрать всю эту информацию в единое целое. Какие факторы усугубляют риск сердечных заболеваний и как они между собой взаимодействуют? Все, что было нужно Ньютону, – это три закона движения и один гравитации, однако одиночке открыть полную модель клетки, организма и общества не под силу. По мере роста объема знаний ученые все больше специализируются на какой-то области, но никто не способен собрать все части воедино, потому что элементов просто слишком много. Они сотрудничают друг с другом, но язык – очень медленное средство общения. Ученые пытаются быть в курсе других исследований, однако объем публикаций настолько велик, что они все больше и больше отстают, и зачастую повторить эксперимент проще, чем найти статью, в которой он описан. Машинное обучение и здесь приходит на помощь: оно просеивает литературу в поисках соответствующей информации, переводит специальный язык одной дисциплины на язык другой и даже находит связи, о которых ученые и не подозревали. Машинное обучение все больше напоминает гигантский хаб¹³, через который методики моделирования, изобретенные в одной области, пробиваются в другие.

Если бы не изобрели компьютеры, наука застряла бы во второй половине XX столетия. Возможно, ученые заметили бы это не сразу и работали бы над все еще возможными небольшими успехами, но потолок прогресса был бы несравнимо ниже. Аналогично без машинного обучения многие науки в ближайшие десятилетия столкнулись бы с проблемой ослабевающей отдачи.

Чтобы увидеть будущее науки, загляните в лабораторию Манчестерского института биотехнологий, где трудится робот по имени Адам. Ему поручено определить, какие гены кодируют ферменты дрожжей. В распоряжении Адама есть модель метаболизма дрожжевой клетки и общие знания о белках и генах. Он выдвигает гипотезы, разрабатывает эксперименты для их проверки, сам проводит опыты, анализирует результаты и выдвигает новые гипотезы, пока не будет удовлетворен. Сегодня ученые все еще независимо прове-

¹² Процесс вырезания определенных нуклеотидных последовательностей из молекул РНК и соединения последовательностей, сохраняющихся в «зрелой» молекуле, в ходе процессинга РНК.

¹³ Хаб (англ. hub, буквально – ступица колеса, центр) – в общем смысле узел какой-то сети.

ряют выводы Адама, прежде чем ему поверить, но уже завтра проверкой этих гипотез займутся роботы.

Миллиард Клинтонов

На президентских выборах 2012 года судьбу Соединенных Штатов определило машинное обучение. Традиционные факторы: взгляды на экономику, харизма и так далее – у обоих кандидатов оказались очень схожи, и исход выборов должен был определиться в ключевых колеблющихся штатах. Кампания Митта Ромни шла по классической схеме: опросы, объединение избирателей в крупные категории и выбор важнейших целевых групп. Нил Ньюхауз, специалист по общественному мнению в штабе Ромни, утверждал: «Если мы сможем победить самовыдвиженцев в Огайо, то выиграем гонку». Ромни действительно победил с перевесом в семь процентов, но все равно проиграл и в штате, и на выборах.

Барак Обама назначил главным аналитиком своей кампании Раида Гани, эксперта по машинному обучению. Гани удалось провести величайшую аналитическую операцию в истории политики. Его команда свела всю информацию об избирателях в единую базу данных, дополнила ее сведениями из социальных сетей, маркетинга и других источников и приступила к прогнозированию четырех факторов для каждого отдельного избирателя: насколько вероятно, что он поддержит Обаму, придет на выборы, отзовется на напоминание это сделать и изменит мнение об этих выборах после бесед на определенные темы. На основе этих моделей каждый вечер проводилось 66 тысяч симуляций выборов, а результаты использовались, чтобы управлять армией волонтеров: кому звонить, в какие двери стучать, что говорить.

В политике, как в бизнесе и на войне, нет ничего хуже, чем смотреть, как противник делает что-то непонятное, и не знать, как на это ответить, пока не станет слишком поздно. Именно это произошло с Ромни. В его штабе видели, что соперники покупают рекламу на конкретных каналах кабельного телевидения в конкретных городах, но почему – было неясно. «Хрустальный шар» оказался слишком мутным. В результате Обама выиграл во всех ключевых штатах за исключением Северной Каролины, причем с большим перевесом, чем предсказывали даже самые авторитетные специалисты по общественному мнению. А наиболее авторитетные специалисты (например, Нейт Сильвер¹⁴), в свою очередь, использовали самые сложные методики прогнозирования. Их предсказания не сбылись, потому что у них было меньше ресурсов, чем у штаба Обамы, но и они оказались намного точнее, чем традиционные эксперты, чьи предсказания были основаны на собственных знаниях и опыте.

Вы можете возразить, что выборы 2012 года были просто случайностью: в большинстве избирательных кампаний шансы кандидатов не настолько одинаковы, и машинное обучение не может быть решающим фактором. Но дело в том, что машинное обучение будет *приводить* к тому, что в будущем все больше выборов окажутся уравновешенными. В политике, как и в других областях, использование обучения похоже на гонку вооружений. В дни Карла Роува, бывшего специалиста по прямому маркетингу и добыче данных, республиканцы лидировали. К 2012 году они отстали, но теперь вновь догоняют демократов. Неизвестно, кто вырвется вперед во время следующей избирательной кампании, но обе партии станут усердно работать над победой, а значит, лучше понимать избирателей и на основе этого знания точно наносить удары и даже подбирать кандидатов. То же самое касается общей политической платформы партий во время выборов и между ними: если основанная на достоверных данных подробная модель избирателя подсказывает, что программа у партии проигрышная, ее изменят. В результате разрыв между кандидатами на выборах будет менее

¹⁴ Натаниель (Нейт) Сильвер (Nathaniel (Nate) Silver, род. 1978) – аналитик, стал известен в 2000-х годах предсказаниями результатов соревнований по бейсболу, а затем и политических выборов.

значительным и устойчивым, и при прочих равных начнут побеждать кандидаты с лучшими моделями избирателей, а избиратели будут этому способствовать.

Один из величайших талантов политика – способность понимать людей, которые за него голосуют по отдельности и в небольших группах, и апеллировать к их нуждам (или делать вид). Образцовый пример из недавней истории – Билл Клинтон. Машинное обучение действует так, будто к каждому избирателю приставлен персональный, преданный ему Клинтон. Каждый из этих мини-Клинтонов и близко не сравним с настоящим, но они берут числом, ведь даже сам Билл Клинтон не может знать, о чем думает каждый американский избиратель, хотя ему бы, конечно, хотелось. Обучающиеся алгоритмы – это агитаторы высшего класса.

Конечно, политики, как и коммерческие организации, могут использовать знание, полученное благодаря машинному обучению, и во благо, и во вред, например, давать разным избирателям противоречащие друг другу обещания. Однако избиратели, средства массовой информации и организации, следящие за выборами, могут провести собственный анализ данных и указать на политиков, переходящих черту. Гонка вооружений будет происходить не только между кандидатами, но и между всеми участниками демократического процесса.

В целом это приведет к лучшему функционированию демократических институтов, потому что канал связи между избирателями и политиками очень сильно расширится. Даже в век высокоскоростного интернета объем информации, которую получают от нас наши представители, все еще ближе к XIX веку: примерно сотня бит раз в два года – столько умещается в бюллетене. К этому прибавляются опросы общественного мнения и, может быть, периодические электронные письма и встречи в городской администрации. Это практически ничто. Большие данные и машинное обучение изменят ситуацию. Учитывая, что в будущем модели избирателей станут точнее, выборные чиновники смогут хоть тысячу раз на дню узнавать, чего хотят люди, и поступать в соответствии с этими пожеланиями, не надоедая при этом настоящим, живым гражданам.

Один сигнал, если сушей, два – если по интернету

В киберпространстве алгоритмы машинного обучения крепят национальную оборону. Каждый день иностранные хакеры пытаются взломать компьютеры Пентагона, стратегических предприятий, других организаций и государственных учреждений. Их тактика постоянно меняется, поэтому меры, работавшие против вчерашних атак, сегодня уже бессильны. Вручную написанные программы для выявления и блокировки таких атак были бы очередной линией Мажино¹⁵, и киберкоманда Пентагона это понимает. А если это атака совершенно нового типа и научиться на прошлых примерах нельзя? Для этого обучающиеся алгоритмы строят модели нормального поведения, примеров которого хватает, и отмечают аномалии. Еще они могут вызвать кавалерию – системных администраторов. Если когда-нибудь разразится кибервойна, генералами в ней будут люди, а пехотой – алгоритмы. Люди слишком медлительны, и их слишком мало, поэтому армия ботов их быстро сметет. Нам нужна собственная армия ботов, и машинное обучение для них – как Военная академия в Вест-Пойнте¹⁶.

Кибервойна – это частный случай асимметричного конфликта, где одна из сторон не может сравниться с другой по мощи обычного вооружения, но тем не менее способна нанести противнику тяжелый урон. Небольшой отряд террористов, вооруженных канцелярскими ножами, смог обрушить башни-близнецы и убить тысячи невинных людей. Сегодня все наиболее серьезные угрозы безопасности США – асимметричные, и от них есть эффективное противоядие: информация. Если враг не сможет скрыться, он не выживет. Информации у нас предостаточно, и это хорошо, но есть и плохие новости.

Агентство национальной безопасности США печально известно своим неуемным аппетитом к данным: по некоторым оценкам, оно перехватывает более миллиарда телефонных звонков и других сообщений по всему земному шару. Не будем сейчас рассуждать об этических вопросах защиты частной жизни. Важно, что у агентства нет столько сотрудников, чтобы прослушать все эти звонки, прочитать электронные письма и даже отследить, кто с кем разговаривает. Большинство звонков вполне безобидны, поэтому написать программу, которая выловит из этого моря несколько подозрительных, очень сложно. Когда-то для этой цели использовались ключевые слова, но этот метод легко обвести вокруг пальца: достаточно назвать теракт свадьбой, а бомбу – свадебным тортом. В XXI веке за эту работу взялось машинное обучение. Конечно, работа агентства овеяна тайной, но в выступлении перед Конгрессом его директор признал, что анализ телефонных разговоров уже предотвратил десятки террористических уговоров.

Если террористы смешаются с толпой футбольных фанатов, то обучающиеся алгоритмы смогут распознать их лица. Если террористы изобретут необычные взрывные устройства, алгоритмы обнаружат их. Алгоритмы могут решать и более тонкие задачи: связывать между собой события, которые по отдельности выглядят безобидными, но вместе складываются в зловещую схему. Такой подход мог бы предотвратить теракты 11 сентября 2001 года. Есть и еще один аспект. В ответ на действия обученной программы злоумышленники будут менять поведение, чтобы обвести ее вокруг пальца, и станут выделяться на фоне обычных людей, которые ведут себя по-прежнему. Чтобы этим воспользоваться, машинное обучение нужно объединить с теорией игр. В прошлом я работал над этой темой: надо не просто уметь

¹⁵ Система французских укреплений длиной около 400 км на границе с Германией от Бельфора до Лонгйона. Была построена в 1929–1934 годах (затем совершенствовалась вплоть до 1940 года). Французские военные стратеги считали линию Мажино неприступной, однако 14 июня 1940 года она была прорвана за несколько часов в результате наступления германской пехоты даже без танковой поддержки.

¹⁶ Военная академия Соединенных Штатов Америки (United States Military Academy), известная также как Вест-Пойнт (West Point) – высшее федеральное военное учебное заведение армии США.

побеждать сегодняшнего противника, но учиться парировать действия, которые он может предпринять против твоего алгоритма. К тому же учет плюсов и минусов различных действий, который возможен благодаря теории игр, может помочь найти правильный баланс между частной жизнью и безопасностью.

Во время битвы за Британию¹⁷ Королевские ВВС выстояли, несмотря на значительный перевес люфтваффе. Немецкие летчики недоумевали: куда бы они ни летели, их всегда поджидали британские самолеты. У Великобритании было секретное оружие: радар, который замечал самолеты противника задолго до того, как тот входил в ее воздушное пространство. Машинное обучение – как радар, который сканирует будущее. Он позволяет не просто реагировать на ходы неприятеля, а превосходить их и рушить его планы.

Близкий каждому пример – так называемая полицейская профилактика. Благодаря прогнозированию тенденций в преступном мире, стратегическому распределению патрулей в наиболее опасных районах города и другим мерам правоохранительные органы эффективно выполняют задачи, которые без этих технологий потребовали бы больших сил. Работа полиции – будь то выявление мошенничества, раскрытие преступных сетей или старая добрая патрульная служба – во многом схожа с асимметричными боевыми действиями, и здесь находят применение многие из соответствующих методик обучения.

Машинное обучение играет все большую роль в военном деле. Обучающиеся алгоритмы могут развеять «туман войны»: анализ изображений, полученных при рекогносцировке, обработка рапортов после боя, составление картины положения для командира. Обучение усилит интеллект боевых роботов, поможет им ориентироваться, приспосабливаться к местности, отличать вражескую технику от гражданской, правильно целиться. Робот AlphaDog, разработанный Агентством по перспективным оборонным проектам, может нести солдату снаряжение. С помощью обучающихся алгоритмов дроны смогут летать автономно. Пока они отчасти контролируются людьми, но все идет к тому, что один пилот станет управлять все большим и большим роем летательных аппаратов. В армии будущего обучающихся алгоритмов будет значительно больше, чем солдат, а это спасет множество жизней.

¹⁷ Авиационное сражение Второй мировой войны, продолжавшееся с 10 июля по 30 октября 1940 года. Термин «битва за Британию» впервые использовал Уинстон Черчилль, назвав так попытку Третьего рейха завоевать господство в воздухе над югом Англии и подорвать боевой дух британского народа.

Куда мы идем?

Тенденции в мире технологий приходят и уходят, но в машинном обучении необычно то, что, несмотря на все трудности, оно продолжает развиваться. Первым крупным всплеском популярности стало прогнозирование взлетов и падений на рынках ценных бумаг, появившееся в конце 1980#х годов. Следующей волной стал анализ корпоративных баз данных, который начал довольно активно внедряться в середине 1990#х годов, а также такие области, как прямой маркетинг, управление работой с клиентами, оценка кредитоспособности и выявление мошенничества. Затем пришел черед интернета и электронной коммерции, где автоматизированная персонализация быстро стала нормой. Когда лопнувший пузырь доткомов нанес удар по этому бизнесу, приобрело популярность использование машинного обучения для поиска в интернете и размещения рекламы. События 11 сентября бросили машинное обучение на передовую войны с террором. Web 2.0 принес с собой целый спектр новых применений – от анализа социальных сетей до определения, что блогеры пишут о продукции данной компании. Параллельно ученые всех мастей все чаще обращались к масштабному моделированию. В первых рядах шли молекулярные биологи и астрономы. Едва наметился кризис на рынке недвижимости, как таланты стали перетекать с Уолл-стрит в Кремниевую долину. На 2011 год пришелся пик популярности мема¹⁸ о больших данных, и машинное обучение оказалось прямо в центре глобального экономического кризиса. Сегодня, кажется, сложно найти область приложения человеческих усилий, не затронутую машинным обучением, включая неочевидные на первый взгляд сферы, например музыку, спорт и дегустацию вин.

¹⁸ Единица культурной информации. Мемом может считаться любая идея, символ, манера или образ действия, осознанно или неосознанно передаваемые от человека к человеку посредством речи, письма, видео, ритуалов, жестов и так далее.

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.