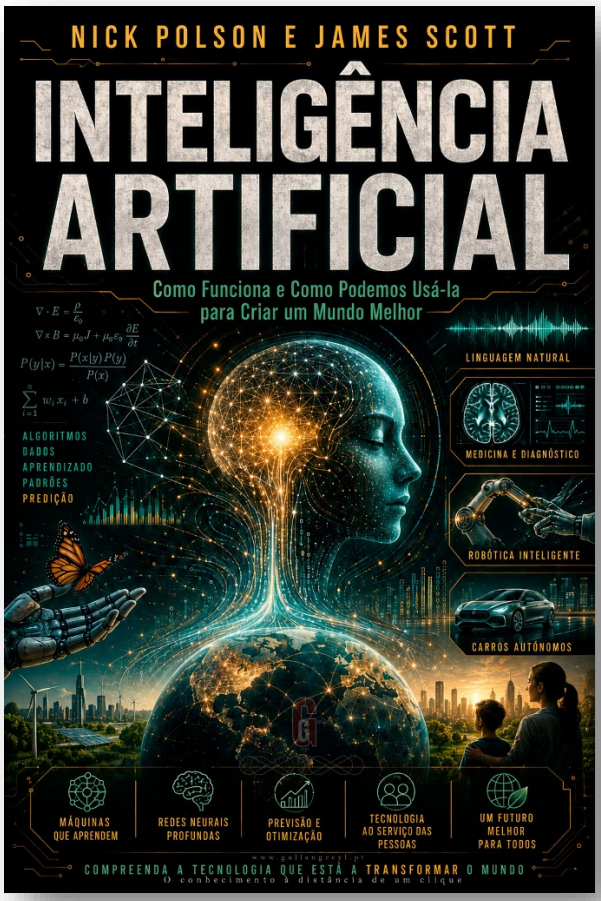
INTELIGÊNCIA ARTIFICIAL

Como Funciona e Como Podemos Usá-la para Criar um Mundo Melhor



Autor: Nick Polson e James Scott

Edição e Preparação Digital: Gullan Greyl

Edição original: AIQ: How People and Machines Are Smarter Together, 2018.

Edição portuguesa concluída em 31-12-2023

Livroteca Gullan Greyl

Sinopse

Em Inteligência Artificial, Nick Polson e James Scott abrem a caixa negra da tecnologia que já atravessa silenciosamente o quotidiano: recomendações da Netflix, carros autónomos, reconhecimento facial, diagnósticos médicos, tradução automática, assistentes digitais, deteção de fraudes e sistemas capazes de aprender padrões a partir de oceanos de dados. A IA deixa de aparecer como ficção científica distante e revela-se como uma força presente, prática e disseminada, a refazer o século XXI com impacto comparável ao da Revolução Industrial no século XIX.

Mas este livro não explica a inteligência artificial a partir de equações frias ou jargão técnico. Explica-a através de histórias humanas: Abraham Wald e os aviões da Segunda Guerra Mundial, Henrietta Leavitt e a medição do Universo, Thomas Bayes e a probabilidade, Grace Hopper e a linguagem dos computadores, Isaac Newton e a deteção de anomalias, Florence Nightingale e a medicina baseada em dados. Cada figura revela uma peça do mecanismo: dados em falta, padrões, previsão, linguagem natural, variabilidade, decisão, enviesamento e modelos que envelhecem.

A grande força da obra está em mostrar que a IA não é magia nem consciência artificial: é uma arquitetura de probabilidades, algoritmos, dados e aprendizagem. As máquinas não "pensam" como nós; aprendem a reconhecer relações, estimar possibilidades e produzir decisões úteis dentro de domínios específicos. E, precisamente por isso, o livro também abre a zona crítica: privacidade, manipulação, marketing direcionado, viés, maus pressupostos, modelos obsoletos e decisões automatizadas que podem amplificar tanto a inteligência como as fragilidades humanas.

Inteligência Artificial é uma viagem ao presente oculto da civilização digital. Um livro que transforma a IA de ameaça nebulosa em linguagem compreensível, mostrando que o futuro não será construído apenas por máquinas inteligentes, mas por seres humanos suficientemente lúcidos para compreenderem como elas funcionam, onde falham e como podem ser usadas para criar um mundo melhor.

Esta edição foi preparada para formato digital, com intervenção ao nível da organização, revisão e uniformização do texto, de forma a assegurar clareza, consistência e legibilidade em ambiente de leitura eletrónica.

Foram realizadas correções pontuais sempre que necessário, respeitando integralmente o conteúdo e a estrutura da obra original.

O objetivo desta preparação é preservar a fidelidade ao texto, garantindo simultaneamente uma experiência de leitura fluida e acessível.

Gullan Greyl

Não se trata de saber mais,
mas de ver melhor

Índice

INTELIGÊNCIA ARTIFICIAL

Como Funciona e Como Podemos Usá-la para Criar um Mundo Melhor

Nick Polson e James Scott

Introdução

Todos os anos, ensinamos a ciência de dados a centenas de alunos, e todos eles estão fascinados pela inteligência artificial (IA). E colocam *grandes* questões. Como é que um carro aprende a conduzir-se a si mesmo? Como é que a *Alexa* percebe o que eu estou a dizer? Como é que o *Spotify* escolhe listas de reprodução tão boas para mim? Como é que o *Facebook* reconhece os meus amigos nas fotos que eu publico? Estes alunos têm noção de que a IA não é um androide de ficção científica do futuro; já está presente aqui e agora, e está a mudar o mundo ao ritmo de um smartphone de cada vez. Todos eles querem compreendê-la, e todos querem fazer parte dela.

Mas os nossos estudantes não são os únicos entusiasmados com a IA. A sua exaltação é partilhada pelas maiores empresas do mundo — desde a Amazon, Google e Facebook na América, até à Baidu, Tencent e Alibaba na China. Como poderá ter ouvido, estas grandes empresas tecnológicas estão a travar uma competição dispendiosa e global pelos talentos da IA, os quais elas consideram essenciais para o seu futuro. Há anos que as vemos a seduzir recém-licenciados com ofertas de salários superiores a 300 mil dólares anuais e com muito melhor café do que o que temos na universidade. Agora vemos muitas mais companhias a envolverem-se na disputa do recrutamento para IA — empresas com quantidades enormes de dados relativos a, por exemplo, seguros ou ao setor petrolífero, que estão a avançar com as suas próprias ofertas de enormes salários e requintadas máquinas de café expresso.

No entanto, ainda que esta competição seja real, achamos que há uma propensão muito mais poderosa a atuar no mundo da IA atual — uma tendência de difusão e disseminação, em vez de concentração. Sim, todas as grandes empresas tecnológicas estão a tentar arrebanhar talento matemático e de codificação. Mas, ao mesmo tempo, as ideias e as tecnologias subjacentes à IA estão a alastrar-se a uma velocidade extraordinária: às empresas mais pequenas, a outras áreas da economia, a entusiastas, a codificadores e a investigadores em todo o mundo. Essa tendência de democratização, mais do que qualquer outra coisa, é o que hoje entusiasma os nossos alunos enquanto contemplam uma vasta gama de problemas que praticamente imploram boas soluções de IA.

Quem teria imaginado, por exemplo, que tantos universitários ficariam tão entusiasmados com a matemática dos pepinos? Bem, eles ficaram quando ouviram falar de Makoto Koike,

um engenheiro de automóveis japonês cujos pais possuem uma quinta de pepinos. No Japão, os pepinos apresentam uma vertiginosa variedade de tamanho, forma, cor, e espinhos — e com base nestes aspetos visuais têm de ser divididos em nove classes diferentes, a que correspondem diferentes preços de mercado. A mãe de Koike costumava passar *oito horas por dia* na triagem manual dos pepinos. Foi então que Koike descobriu que podia utilizar um *software* livre de IA da Google, chamado *TensorFlow*, para realizar a mesma tarefa. Para tal, tinha de codificar um algoritmo de aprendizagem profunda que conseguisse classificar um pepino com base numa fotografia. Koike nunca antes usara IA ou o *TensorFlow*, mas com todos os recursos gratuitos disponíveis, não achou difícil aprender a fazê-lo sozinho. Quando um vídeo da sua máquina de triagem operada com IA chegou ao *YouTube*, Koike tornou-se uma celebridade internacional de aprendizagem profunda dos pepinos. Não se tratou apenas de ter partilhado com as pessoas uma história feliz, poupando horas de trabalho duro à sua mãe. Ele também enviou uma mensagem de esperança a estudantes e codificadores de todo o mundo: se a IA pode resolver problemas na produção de pepinos, então pode fazê-lo praticamente em todo o lado.

E essa mensagem está agora a espalhar-se rapidamente. Atualmente, os médicos estão a usar a IA para diagnosticar e tratar o cancro. As empresas elétricas usam a IA para aperfeiçoar a eficiência na produção de energia. Os investidores usam-na para gerir o risco financeiro. As empresas petrolíferas utilizam-na para melhorar a segurança nas plataformas em alto-mar. As agências de segurança usam-na para caçar terroristas. Os cientistas usam-na para fazerem novas descobertas na astronomia, na física e nas neurociências. Empresas, investigadores e entusiastas em todo o mundo estão a usar a IA de milhares de maneiras diferentes, quer seja para detetar fugas de gás, minerar ferro, prever surtos de doenças, salvar abelhas da extinção, ou quantificar a discriminação sexual em filmes de Hollywood. E isto é apenas o começo.

Nós vemos a *verdadeira* história da IA como sendo a crónica desta disseminação: de um punhado de conceitos-base matemáticos que remontam a várias décadas, ou mesmo a séculos, até aos supercomputadores e máquinas que falam/pensam/separam pepinos dos dias de hoje, e até às novas e omnipresentes maravilhas digitais do amanhã. O nosso objetivo neste livro é contar-lhe essa história. É em parte uma história de tecnologia, mas é principalmente uma história de ideias, e das pessoas por detrás dessas ideias — pessoas de uma época muito anterior, discretas, que apenas estavam a tentar resolver os seus próprios

problemas que envolviam matemática e dados, e que não tinham *qualquer ideia* acerca do papel que as suas soluções desempenhariam na conceção do mundo moderno. No final dessa história irá perceber o que é a IA, de onde ela vem, como funciona, e por que é importante na sua vida.

1. O Que Significa, Verdadeiramente, «IA»?

Quando ouvir «IA», não pense num androide. Pense num *algoritmo*. Um algoritmo é um conjunto de instruções passo a passo, tão explícitas que até algo com uma «mente» tão literal como um computador pode segui-las. (Talvez tenha ouvido a anedota do robot que ficou preso para sempre no duche por causa do algoritmo na embalagem de champô: «Ensaboar. Enxaguar. Repetir.») Por si só, um algoritmo não é mais inteligente do que um berbequim elétrico; apenas executa uma coisa muito bem, como ordenar uma lista de números ou pesquisar fotografias de animais fofinhos na Internet. Mas se conseguirmos juntar muitos algoritmos, e encadeá-los de uma maneira inteligente, podemos produzir IA: uma ilusão de comportamento inteligente de um domínio específico. Por exemplo: considere um assistente digital como o *Google Home*, ao qual pode colocar uma questão como «Onde é que posso comer os melhores tacos ao pequeno-almoço em Austin». Esta pergunta espoleta uma reação em cadeia de algoritmos:

- Um algoritmo converte a onda sonora inicial num sinal digital.
- Outro algoritmo traduz esse sinal numa cadeia de fonemas portugueses, ou sons percetualmente distintos, como «ta-kus-aw pekenwal'mosu.»
- O algoritmo seguinte segmenta esses fonemas em palavras: «tacos ao pequeno-almoço».
- Essas palavras são enviadas para um motor de busca — ele próprio um imenso oleoduto de algoritmos que processa a pergunta e devolve uma resposta.
- Outro algoritmo formata a resposta numa frase coerente.

- Um algoritmo final verbaliza essa frase de maneira a que não soe a um robot: «Os melhores tacos ao pequeno-almoço em Austin são no Julio's, na Duval Street. Deseja obter direções?»

Isso é IA. Praticamente todos os sistemas de IA — quer se trate de um carro autónomo, de um separador de pepinos automático ou de um *software* que monitoriza a sua conta do cartão de crédito alertando para fraudes — seguem um modelo semelhante ao deste «oleoduto de algoritmos». O oleoduto recebe dados de um domínio específico, executa uma cadeia de cálculos e emite uma previsão ou uma decisão.

Os algoritmos utilizados em IA têm duas caraterísticas distintivas. A primeira é estes algoritmos tipicamente lidarem com probabilidades em vez de certezas. Um algoritmo de IA, por exemplo, não dirá de forma absoluta que uma dada transação com o cartão de crédito é fraudulenta. Ao invés, transmitirá que a probabilidade de fraude é de 92 por cento — ou qualquer outro valor que ele considere face aos dados. A segunda é a forma como os algoritmos «sabem» quais as instruções que devem seguir. Nos algoritmos tradicionais, como os que operam websites ou processadores de texto, essas instruções são fixadas previamente por um programador. Na IA, contudo, as instruções são aprendidas pelo próprio algoritmo, diretamente a partir de «dados de treino». Ninguém diz a uma IA como classificar transações com cartão de crédito em fraudulentas ou legítimas. Em vez disso, o algoritmo observa muitos exemplos de cada categoria (fraudulenta, não fraudulenta) e descobre o padrão que distingue uma da outra. Na IA, o papel do programador não é dizer ao algoritmo o que deve fazer. É dizer ao algoritmo como treinar-se a si próprio para o que deve fazer, recorrendo aos dados e às leis das probabilidades.

2. Como é que Chegámos Até Aqui?

Os sistemas modernos de IA, como um carro autónomo ou um assistente digital doméstico, entraram em cena muito recentemente. Mas talvez fique surpreendido por saber que a maioria das grandes ideias na IA são, de facto, *antigas* — algumas têm centenas de anos — e que os nossos antepassados têm-nas usado para resolver problemas há gerações. Por exemplo: considere os carros que se conduzem a si próprios. A Google lançou um carro deste tipo em 2009. Mas irá descobrir no Capítulo 3 que uma das ideias principais para o funcionamento desses carros foi descoberta por um pastor presbiteriano na década de 1750

— e que esta ideia foi usada há mais de 50 anos por uma equipa de matemáticos para resolver um dos mistérios mais espetaculares da Guerra Fria.

Ou então considere a classificação de imagens, como o *software* que automaticamente marca os seus amigos nas fotos no *Facebook*. Os algoritmos para processamento de imagens têm-se tornado radicalmente melhores nos últimos cinco anos. Mas no Capítulo 2 aprenderá que as ideias-chave neste domínio remontam a 1805 — e que estas ideias foram usadas há um século, por uma astrónoma pouco conhecida chamada Henrietta Leavitt, para ajudar a responder a uma das questões mais profundas que os humanos alguma vez colocaram: quão grande é o Universo?

Considere ainda o reconhecimento do discurso, um dos grandes triunfos da IA dos últimos anos. Os assistentes digitais como a *Alexa* ou o *Google Home* são notavelmente fluentes na linguagem e apenas irão melhorar. Mas a primeira pessoa a conseguir que um computador percebesse inglês foi uma contra-almirante na Marinha dos Estados Unidos, e ela fê-lo há quase 70 anos. (Veja no Capítulo 4.)

Estes são apenas três exemplos de um facto extraordinário: para onde quer que olhe na IA, irá encontrar uma ideia com a qual as pessoas se debatem há já muito tempo. Assim, em muitos aspetos, o grande quebra-cabeças histórico não é por que está a IA a acontecer agora, mas por que não aconteceu há muito mais tempo. Para explicar este enigma temos de olhar para três forças tecnológicas facilitadoras que trouxeram estas ideias veneráveis para uma nova Era.

O primeiro facilitador da IA é o crescimento exponencial ao longo de décadas da velocidade dos computadores, habitualmente conhecido por Lei de Moore. É difícil transmitir intuitivamente quão rápidos os computadores se tornaram. O clichê costumava ser que os astronautas da *Apollo* aterraram na Lua munidos de computadores com menos capacidade do que uma calculadora de bolso. Mas isto já não é bem assim, porque... o que é uma calculadora de bolso? Por isso vamos antes tentar uma analogia com carros. Em 1951, um dos computadores mais rápidos era o *UNIVAC*, que efetuava dois mil cálculos por segundo, enquanto um dos carros mais rápidos era o *Alfa Romeo 6C*, que fazia 177 quilómetros por hora. Tanto os carros como os computadores foram melhorados desde 1951 — mas se os veículos tivessem melhorado ao mesmo ritmo dos computadores, então um *Alfa Romeo* moderno faria oito milhões de vezes a velocidade da luz.

O segundo facilitador de IA é a *nova* Lei de Moore: o crescimento explosivo na quantidade de dados disponíveis, dado que toda a informação da humanidade tornou-se digitalizada. A Biblioteca do Congresso dos EUA consome 10 terabytes de armazenamento, mas só em 2013, os dados que as quatro grandes empresas tecnológicas — Google, Apple, Facebook e Amazon — recolheram atingiram cerca de 120 mil vezes este valor. E isso foi há uma vida, em anos de Internet. O ritmo de acumulação de dados está a acelerar mais depressa do que um foguetão *Apollo*; em 2017, mais de 300 horas de vídeo foram carregadas para o *YouTube* em cada minuto, e mais de um milhão de imagens foram postadas no *Instagram* quotidianamente. Mais dados significa algoritmos mais inteligentes.

O terceiro facilitador é a computação em nuvem. Esta tendência é praticamente invisível para os consumidores, mas teve um enorme efeito democratizador na IA. Para ilustrar isto vamos recorrer a uma analogia entre dados e petróleo. Imagine que todas as empresas do início do século XX possuíam algum petróleo, mas tinham elas próprias de construir as infraestruturas para extrair, transportar e refinar esse petróleo. Para começar, qualquer empresa que tivesse uma ideia nova para dar bom uso ao seu petróleo ter-se-ia deparado com enormes custos fixos; como consequência, a maior parte do petróleo teria permanecido no solo. Bem, a mesma lógica aplica-se aos dados, o petróleo do século XXI. A maioria dos entusiastas ou das pequenas empresas enfrentaria custos proibitivos se tivesse de comprar todo o equipamento e deter o conhecimento necessário para construir um sistema de IA a partir dos seus dados. Mas os recursos de computação em nuvem fornecidos por consórcios como a Microsoft Azure, a IBM e a Amazon Web Services transformaram esses custos fixos em custos variáveis, alterando radicalmente o cálculo económico para o armazenamento e análise de dados em larga escala. Hoje, qualquer pessoa que queira utilizar o seu «petróleo» pode fazê-lo com custos reduzidos, alugando a infraestrutura de outrem.

Quando estas quatro tendências se juntam — processadores mais rápidos, conjuntos gigantescos de dados, computação em nuvem e, acima de tudo, boas ideias — obtém-se uma explosão do tipo supernova tanto na procura como na capacidade de usar a IA para resolver problemas reais.

3. Ansiedades com a IA

Já lhe contámos quão entusiasmados os nossos alunos estão com a IA e como as maiores empresas do mundo se apressam a abraçá-la. Mas estaríamos a mentir se disséssemos que *toda a gente* foi tão otimista em relação a estas novas tecnologias. Na verdade, muitas pessoas estão ansiosas, quer seja por causa dos seus empregos, da privacidade dos dados, da concentração de riqueza ou de russos com contas-*bots* de notícias falsas no *Twitter* (atualmente *X*). Algumas pessoas — entre as quais a mais famosa é Elon Musk, o empreendedor no setor da tecnologia por trás da Tesla e da SpaceX — pintam um quadro ainda mais negro, um em que os robots se tornam autoconscientes, decidem que não gostam de ser governados pelas pessoas e começam a *governar-nos* com um punho de silicone.

Falemos primeiro da preocupação de Musk. As opiniões dele receberam bastante atenção, presumivelmente porque as pessoas estão atentas quando um membro da classe multimilionária disruptiva fala acerca da inteligência artificial. Musk tem afirmado que a humanidade, ao desenvolver a tecnologia da IA, está a «invocar um demónio», e que as máquinas inteligentes são «a maior ameaça à nossa existência» enquanto espécie.

Depois de ler este livro, poderá decidir por si mesmo se considera que estas preocupações são credíveis. Contudo, queremos alertá-lo antecipadamente de que é muito fácil cair numa ratoeira a que os cientistas cognitivos chamam «heurística da disponibilidade»: o atalho mental em que as pessoas avaliam a plausibilidade de uma afirmação ao basearem-se nos exemplos que nesse instante surgem nas suas mentes. No caso da IA, a maior parte dos exemplos advêm da ficção científica e são essencialmente maléficos — do *Exterminador* ao *Borg* e ao *Hal 9000*. Nós pensamos que estes exemplos de ficção científica têm um forte efeito de ancoragem, que faz com que muitas pessoas vejam a narrativa da «maléfica IA» de maneira menos cética do que deviam. Afinal de contas, só porque conseguimos sonhar e fazer um filme acerca dela não quer dizer que a consigamos desenvolver. Atualmente ninguém tem qualquer ideia de como criar um robot com inteligência geral, à imagem de um humano ou de um *Exterminador*. Talvez os seus descendentes distantes descubram como fazê-lo; talvez eles até programem a sua criação para que aterrorize os descendentes distantes de Elon Musk. Mas isso será um problema e uma escolha deles, porque nenhuma das opções disponíveis hoje em dia nem sequer remotamente predestina tal possibilidade.

Agora, e no futuro próximo, as máquinas «inteligentes» apenas são espertas nos seus domínios específicos:

- A Alexa consegue ler-lhe uma receita para esparguete à bolonhesa, mas não consegue picar a cebola e certamente não se virará contra si com uma faca de cozinha.

- Um carro autónomo pode levá-lo até ao campo de futebol, mas não consegue arbitrar o jogo, quanto mais decidir por si só atá-lo a um dos postes da baliza e chutar a bola contra as suas partes sensíveis.

Além disso, considere o custo de oportunidade de nos preocuparmos com a eventualidade de em breve sermos conquistados por robots autoconscientes. Focarmo-nos agora nesta possibilidade é equivalente à de a Havilland Aircraft Company, tendo realizado o primeiro voo com um jato comercial em 1952, preocupar-se com as implicações das viagens à velocidade *warp* a galáxias distantes. Talvez um dia, mas neste momento há coisas muito mais importantes com que nos inquietarmos — como, para levar a analogia com o jato um pouco mais longe, estabelecer uma política inteligente para todos os aviões *presentemente* no ar.

O tema da política conduz-nos a todo um outro conjunto de ansiedades acerca da IA, muito mais plausível e imediato. Irá a IA criar um mundo sem empregos? Irão as máquinas tomar decisões sobre a nossa vida, sem nunca prestarem contas? Irão as pessoas que detêm os robots mais inteligentes acabar por dominar o futuro?

Estas questões são profundamente importantes, e ouvimo-las a serem debatidas a toda a hora — em conferências de tecnologia, nas páginas dos jornais mais importantes do mundo e ao almoço com os nossos colegas. Convém que saiba antecipadamente que não encontrará as respostas a estas perguntas neste livro, porque nós não as conhecemos. Tal como os nossos alunos, estamos definitivamente otimistas em relação ao futuro da IA e esperamos que no final do livro partilhe esse otimismo. Mas não somos economistas do trabalho, especialistas em política ou adivinhos. Somos cientistas de dados e também somos académicos, o que significa que o nosso instinto é permanecer firmemente no nosso caminho, onde confiamos nas nossas competências. Podemos ensinar-lhe acerca da IA, mas não podemos ao certo dizer o que o futuro trará.

No entanto, *podemos* dizer-lhe que encontrámos um conjunto de narrativas que as pessoas usam para enquadrar este assunto e que as consideramos todas incompletas. Estas narrativas enfatizam a riqueza e o poder das grandes empresas tecnológicas, mas negligenciam a enorme democratização e difusão da IA que já está a acontecer. Elas destacam os perigos de as máquinas tomarem decisões importantes com base em dados distorcidos, mas não reconhecem o enviesamento ou a flagrante malícia na tomada de decisões humanas com que temos vivido desde sempre. Acima de tudo focam-se intensamente no que as máquinas poderão retirar, mas perdem de vista o que poderemos receber em troca: empregos diferentes e melhores, novas conveniências, fim do trabalho duro, locais de trabalho mais seguros, melhores cuidados de saúde, menos barreiras linguísticas, novas ferramentas de aprendizagem e de tomada de decisões, que nos ajudarão a sermos pessoas melhores e mais inteligentes.

Considere a problemática do emprego. Nos Estados Unidos, os pedidos de subsídios de desemprego continuaram a atingir novos mínimos entre 2010 e 2017, mesmo com a IA e a automatização a ganharem corpo enquanto forças económicas. O ritmo da automatização robótica foi ainda mais implacável na China; todavia, aí os salários têm vindo a aumentar há anos. Isso não quer dizer que a IA não tenha ameaçado os empregos das *pessoas individualmente*. Tem-no feito e continuará a fazê-lo, tal como o tear mecânico ameaçou os empregos das tecedeiras, e como de igual forma o carro pôs em perigo os empregos dos condutores de charretes. As novas tecnologias sempre alteraram a composição dos empregos necessários à economia, colocando pressão descendente no salário de algumas áreas e ascendente no de outras. Com a IA não será diferente, e nós apoiamos fortemente programas de formação profissional e de apoio social para providenciarem ajuda significativa aos afetados pela tecnologia. Neste caso, um rendimento básico universal até poderá ser a solução, como parecem pensar muitos patrões de Silicon Valley — nós não nos arrogamos saber. Mas os argumentos de que a IA criará um futuro sem empregos não são, até ao momento, totalmente apoiados pelos exemplos concretos.

Depois há a questão do domínio do mercado. A Amazon, a Google, o Facebook e a Apple são empresas gigantescas com tremendo poder. É crucial que sejamos vigilantes perante esse poder, de modo a que não seja usado para sufocar a concorrência ou minar as regras democráticas. Mas não se esqueça de que estas empresas são bem-sucedidas porque criaram produtos e serviços que as pessoas adoram. E só manterão o sucesso se continuarem

a inovar, o que não é fácil para grandes organizações. Além disso, lemos muitas previsões de que as atuais grandes empresas tecnológicas permanecerão sempre dominantes, e descobrimos que habitualmente estas previsões nem sequer explicam o passado e muito menos o futuro. Lembra-se de quando a Dell e a Microsoft dominavam na computação? Ou de quando a Nokia e a Motorola eram dominantes nos telemóveis — tão dominantes que era difícil imaginar outro cenário? Lembra-se de quando todos os advogados tinham um *BlackBerry*, quando todas as bandas estavam no *Myspace*, ou quando todos os servidores pertenciam à Sun Microsystems? Lembra-se da AOL, da Blockbuster Video, da Yahoo!, da Kodak, ou do *Walkman* da Sony? As empresas vêm e as empresas vão, mas o tempo continua a avançar, e os *gadgets* só ficam cada vez mais fixes.

Nós assumimos uma perspetiva prática sobre a emergência da IA: ela está aqui hoje, e haverá mais amanhã, quer qualquer um de nós goste ou não. Estas tecnologias irão trazer imensos benefícios, mas também refletirão, inevitavelmente, os nossos pontos fracos enquanto civilização. Como resultado haverá ameaças a que devemos estar atentos, quer seja à privacidade, à igualdade, às instituições existentes ou a algo em que ninguém ainda sequer pensou. Temos de enfrentar estas ameaças com políticas inteligentes — e se temos a esperança de elaborar políticas inteligentes num mundo de comentários superficiais de 140 carateres, é essencial que, enquanto sociedade, alcancemos um ponto em que possamos discutir estes assuntos de uma maneira equilibrada, que reflita tanto a sua importância como a sua complexidade. O nosso livro não irá apresentar essa discussão. Mas irá mostrar-lhe o que necessita de compreender caso queira desempenhar um papel devidamente informado nessa discussão.

4. Uma Nota Sobre Matemática

Antes de começarmos, devemos fazer um último aviso: irá encontrar alguma matemática neste livro. Mesmo que nunca se tenha considerado uma pessoa com jeito para a matemática, por favor não se preocupe. A matemática da IA é *surpreendentemente* simples e prometemos-lhe que estará à altura. Também prometemos que valerá o esforço: se entender um bocadinho da matemática subjacente à IA, descobrirá que ela se torna muito menos misteriosa.

Podíamos ter escrito um livro acerca da IA sem qualquer tipo de matemática, segundo a teoria que ouvimos ao longo da nossa vida de que podemos escolher matemática ou amigos,

mas não as duas coisas. No início o nosso editor *implorou*-nos que usássemos esta abordagem, resmungando algo entre dentes acerca de «perdermos três mil leitores por cada símbolo matemático», ou talvez fossem «cinco mil leitores por cada letra grega». O que quer que tenha sido, dissemos não, que fosse para o $\delta\iota\alpha\beta o$, porque a nossa experiência mostra-nos que podemos ter muito mais fé em si do que isso. Entre nós os dois, já ensinamos ciência de dados e probabilidade há 40 anos, incluindo a muitos universitários e alunos de MBA que chegaram até nós pré-instalados com este vírus horrível que faz com que realmente *não gostem de matemática*. Contudo, vimos como os olhos desses mesmos alunos se iluminam quando aprendem como é que todas as fantásticas aplicações de IA de que ouviram falar, da *Alexa* até ao reconhecimento de imagem, realmente funcionam — como é que, quando mergulhamos no essencial, tudo se reduz a probabilidades em grandes volumes de dados com esteroides. Eles acabam por perceber que as equações nunca foram difíceis. No final, eles até se sentem *empoderados* pela matemática e apercebem-se de que, nas circunstâncias certas, pensar um pouco mais como uma máquina — isto é, tomar decisões usando dados e as leis da probabilidade — até pode ajudá-los a serem pessoas mais espertas.

Assim, junte-se a nós ao longo dos próximos sete capítulos, nos quais apresentaremos sete figuras históricas fascinantes, cada uma com uma importante lição que lhe explicará por que necessitam as máquinas inteligentes de pessoas inteligentes, e vice-versa. Ficará com um quociente de inteligência artificial maior e com um novo entendimento sobre quão brilhantes os seres humanos podem ser quando aliam as suas mentes às tecnologias.

CAPÍTULO 1

O Refugiado

Sobre personalização: como um emigrante húngaro utilizou a probabilidade condicionada para proteger aviões do fogo inimigo durante a Segunda Guerra Mundial, e como as empresas tecnológicas atuais estão a usar a mesma matemática para fazerem sugestões personalizadas para filmes, músicas, notícias — até mesmo medicamentos contra o cancro.

A Netflix chegou tão longe, tão rapidamente, que é difícil lembrarmo-nos de que começou como uma empresa de «máquina a aprender por e-mail». Ainda em 2010, o negócio principal da empresa envolvia encher envelopes vermelhos com DVD que jamais incorreriam «em multas por atraso!». Cada envelope regressaria alguns dias após o envio, a par com uma avaliação do assinante acerca do filme usando uma escala de 1 a 5. À medida que os dados dessas avaliações se acumulavam, os algoritmos da Netflix procuravam padrões e, com o passar do tempo, os assinantes iriam receber melhores recomendações para filmes. (Este tipo de IA é geralmente chamado «sistema de recomendação»; nós também gostamos do termo «motor de sugestão».) O *Netflix 1.0* estava tão focado em melhorar o seu sistema de recomendação que em 2007, para grande entusiasmo de matemáticos *geeks* em todo o mundo, anunciou um concurso público para uma máquina de aprendizagem automática, com um

prémio de um milhão de dólares. A empresa colocou num servidor público alguns dos seus dados com as classificações, e desafiou todos os visitantes a melhorar o próprio sistema da Netflix, chamado *Cinematch*, em pelo menos 10 por cento — ou seja, prever como iríamos classificar um filme com uma precisão 10 por cento superior à da Netflix. A primeira equipa a ultrapassar o limiar dos 10 por cento receberia o dinheiro.

Ao longo dos meses seguintes, a empresa foi inundada por milhares de respostas. Algumas ficaram tentadoramente perto do limiar dos 10 por cento, mas ninguém os ultrapassou. Então, em 2009, após dois anos a refinarem o seu algoritmo, uma equipa autodenominada BellKor's Pragmatic Chaos finalmente submeteu o código de um milhão de dólares, derrotando o motor da Netflix com 10,06 por cento. E ainda bem que não fizeram uma pausa para verem um episódio extra de *A Teoria do Big Bang* antes de premirem o botão «submeter». A BellKor chegou à meta da corrida de dois anos apenas 19 minutos e 54 segundos à frente de uma segunda equipa, The Ensemble, que submeteu um algoritmo que também alcançava uma melhoria de 10,06 por cento — mas não foi suficientemente rápida.

Em retrospetiva, o Prémio Netflix foi um símbolo perfeito da dependência inicial da empresa de uma função central de aprendizagem automática: prever através de um algoritmo como é que um assinante iria classificar um filme. Depois, em março de 2011, três pequenas palavras alteraram para sempre o futuro da Netflix: *House of Cards*.

House of Cards foi a primeira «série original Netflix», a primeira tentativa de a empresa produzir televisão em vez de simplesmente distribuí-la. Inicialmente, a equipa de produção por trás de *House of Cards* bateu à porta de todas as grandes cadeias de televisão com a sua ideia, e todas, sem exceção, estavam interessadas. Mas foram cautelosas — todas queriam ver primeiro um episódio-piloto. O programa, no fim de contas, é um conto de mentiras, traição e assassínio. Quase que podemos imaginar as grandes cadeias de televisão a perguntarem a si próprias: «Como podemos ter a certeza de que alguém irá ver algo tão sinistro?» Bem, a Netflix teve. De acordo com os produtores do programa, a Netflix foi a única estação com coragem para dizer «Acreditamos em vocês. Processámos os nossos dados, e os resultados dizem-nos que a nossa audiência iria assistir à série. Não precisamos que façam um episódio-piloto. Quantos episódios querem fazer?»[1]

Processámos os nossos dados e não precisamos que façam um episódio-piloto. Pense nas implicações económicas dessa afirmação para a indústria televisiva. No ano anterior ao da estreia de *House of Cards*, as grandes cadeias de televisão encomendaram 113 episódios-pilotos, com um custo total de perto de 400 milhões de dólares. Desses, apenas 35 foram para o ar, e tão-só 13 — um programa em *nove* — chegaram à segunda temporada. Claramente, as estações não tinham ideia alguma do que teria sucesso.

Então, o que sabia a Netflix em março de 2011 que as grandes estações desconheciam? O que fez com que o seu pessoal confiasse tanto na sua avaliação que esteve disposto a ir além de recomendar

televisão personalizada e começar a *produzir* televisão personalizada?

A resposta fácil é que a Netflix tinha dados sobre a sua base de assinantes. Mas ainda que os dados fossem importantes, esta explicação é demasiado simples. As estações de televisão também tinham bastantes dados, sob a forma de classificações Nielsen, grupos focais e milhares de inquéritos — e grandes orçamentos para recolher mais dados, caso acreditassem na sua importância.

Os cientistas de dados na Netflix, contudo, tinham duas coisas de que as estações televisivas não dispunham, coisas que eram tão importantes quanto os próprios dados: 1) um conhecimento profundo das probabilidades, necessário para colocar as questões certas aos seus dados, e 2) a coragem para reconstruir todo o seu negócio em torno das respostas obtidas. O resultado foi uma transformação surpreendente para a Netflix: de uma rede de distribuição, impulsionada por uma máquina de aprendizagem, para uma nova estirpe de empresa de produção, em que cientistas de dados e artistas se uniam para criar televisão fantástica. Ficou célebre a frase de Ted Sarandos, diretor de conteúdos da Netflix, numa entrevista à *GQ*: «O objetivo é transformarmo-nos na HBO mais depressa do que a HBO consegue transformar-se em nós.»[2]

Atualmente, poucas empresas utilizam a IA para personalização melhor do que a Netflix, e a abordagem em que ela foi pioneira agora domina a economia online. O nosso rasto digital gera sugestões personalizadas para músicas no *Spotify*, vídeos no *YouTube*, produtos na *Amazon*, notícias do *The New York Times*,

amigos no *Facebook*, anúncios no *Google* e empregos no *LinkedIn*. Os médicos, com base nos nossos genes, até podem usar a mesma abordagem para nos fornecer sugestões personalizadas para terapias contra o cancro.

Anteriormente, o algoritmo mais importante na nossa vida digital era a pesquisa, o que para a maioria de nós significava *Google*. Mas os algoritmos-chave do futuro dizem respeito a sugestões, não a pesquisa. A pesquisa é estreita e circunscrita; temos de saber o que procurar e estamos limitados pelo nosso próprio conhecimento e experiência. As sugestões, por outro lado, são ricas e ilimitadas; tiram partido do conhecimento acumulado e da experiência de biliões de outras pessoas. Os motores de sugestões são como um «software sósia», que um dia poderá conhecer as suas preferências melhor do que você mesmo, pelo menos conscientemente. Quanto tempo faltará para que, por exemplo, possa dizer à *Alexa* «Estou a sentir-me aventureiro; marca-me uma semana de férias», esperando um ótimo resultado?

Obviamente, há imensa matemática sofisticada por trás destes motores de sugestões. Mas caso tenha uma fobia à matemática, também há algumas notícias muito boas. Acontece que existe apenas um conceito-chave que precisa de compreender, que é este: para uma máquina de aprendizagem, «personalização» significa «probabilidade condicionada».

Na matemática, uma probabilidade condicionada é a possibilidade de uma coisa acontecer, tendo em conta que outra já aconteceu. Um ótimo exemplo é a previsão meteorológica. Se esta

manhã tivesse olhado pela janela e visto nuvens a acumularem-se, poderia assumir que provavelmente iria chover e levaria um guarda-chuva para o trabalho. Na IA expressamos este julgamento como uma probabilidade condicionada — por exemplo, «a probabilidade condicionada de chover esta tarde, dadas as nuvens desta manhã, é de 50 por cento». Os cientistas de dados escrevem isto de modo um pouco mais compacto: *P* (chuva esta tarde | nuvens esta manhã) = 60%. *P* significa «probabilidade» e a barra vertical significa «dado» ou «condicionada por». O que se encontra à esquerda da barra é o acontecimento em que estamos interessados. O que está à direita é o nosso conhecimento, também chamado «acontecimento condicionante»: aquilo em que acreditamos ou que assumimos ser verdadeiro.

A probabilidade condicionada é como os sistemas de IA expressam julgamentos de uma maneira que reflete o seu conhecimento parcial:

- Acabou de dar ao *Sherlock* uma pontuação elevada. Qual é a probabilidade condicionada de gostar de *O Jogo da Imitação* ou de *A Toupeira*?

- Ontem ouviu o Pharrell Williams no *Spotify*. Hoje, qual é a probabilidade de querer ouvir o Bruno Mars?

- Acaba de comprar comida biológica para cães. Qual é a probabilidade condicionada de também comprar uma coleira para cães com GPS?

- Segue o Cristiano Ronaldo (@cristiano) no *Instagram*. Qual é a probabilidade condicionada de responder a uma sugestão para seguir o Lionel Messi (@leomessi) ou o Gareth Bale (@garethbale11)?

A personalização funciona através das probabilidades condicionadas, que terão de ser estimadas a partir de conjuntos maciços de dados nos quais você é o acontecimento condicionante. Neste capítulo irá aprender um pouco da magia por trás de como isto funciona.

1. Abraham Wald, Herói da Segunda Guerra Mundial

A ideia central subjacente à personalização é muito mais velha do que a *Netflix*, mesmo mais velha do que a própria televisão. De facto, se quer compreender a revolução da última década na maneira como as pessoas interagem com a cultura popular, então o melhor sítio para começar não é em Silicon Valley, em Brooklyn ou em Shoreditch na sala de estar de um *millennial* que cortou o acesso ao telefone fixo e à televisão por cabo. Em vez disso é em 1944, nos céus sobre a Europa ocupada, onde o domínio de um homem sobre a probabilidade condicionada salvou as vidas de um número inaudito de tripulações de bombardeiros aliados na maior campanha aérea da história: o bombardeamento do Terceiro Reich (Império).

Durante a Segunda Guerra Mundial, a dimensão da guerra aérea sobre a Europa foi verdadeiramente impressionante. Todas as

manhãs, vastos esquadrões de *Lancaster* britânicos e *B-17* americanos descolavam de bases em Inglaterra e voavam até aos seus alvos para lá do canal da Mancha. Em 1944, o conjunto das forças aliadas estava a lançar do ar mais de 15 milhões de quilos de bombas *por semana*. Mas, à medida que a campanha aérea se intensificou, o mesmo aconteceu às perdas. Numa mesma missão em agosto de 1943, os aliados enviaram 376 bombardeiros a partir de 16 bases aéreas diferentes, num raide aéreo conjunto para bombardear fábricas em Scheinfurt e Ratisbona, na Alemanha. Sessenta aviões nunca regressaram — uma perda diária de 16 por cento. Nesse dia, o 381.º Grupo de Bombardeiros, que partiu da base da RAF em Ridgwell, perdeu nove dos seus 20 bombardeiros.[3]

Os aviadores na Segunda Guerra Mundial estavam dolorosamente conscientes de que cada missão era uma jogada de risco. Mas perante estas probabilidades sombrias, as tripulações dos bombardeiros tinham pelo menos três defesas.

1. Os seus próprios artilheiros na torre e na cauda, para repelir atacantes.

2. Os seus caças de escolta: os *Spitfire* e *P-51 Mustang* que os acompanhavam para os defender dos ataques da *Luftwaffe*.

3. Um estatístico húngaro-americano chamado Abraham Wald.

Abraham Wald nunca abateu um *Messerschmidt* ou sequer viu o interior de um avião de combate. Não obstante, ele deu um contributo descomunal para o esforço de guerra aliado utilizando uma arma igualmente potente: a probabilidade condicionada. Especificamente, Wald construiu um sistema de recomendação que podia produzir sugestões personalizadas para a capacidade de sobrevivência de diferentes tipos de aviões. No seu cerne, era tal e qual um sistema de recomendação moderno de IA para programas de televisão. E quando compreender como ele o construiu também compreenderá muito mais acerca da *Netflix*, *Hulu*, *Instagram*, *Amazon*, *YouTube*, e praticamente todas as empresas tecnológicas que alguma vez lhe fizeram uma sugestão automática que valeu a pena seguir.

1.1 Os Primeiros Anos de Wald

Abraham Wald nasceu em 1902 numa vasta família de judeus ortodoxos em Kolozsvár, na Hungria, que passou a pertencer à Roménia e que mudou o nome para Cluj a seguir à Primeira Guerra Mundial. O pai dele, que trabalhava numa padaria na cidade, criou um ambiente caseiro de aprendizagem e curiosidade intelectual para os seis filhos. O jovem Wald e os seus irmãos cresceram a tocar violino, a resolver quebra-cabeças matemáticos e a escutarem histórias aos pés do avô, um famoso e adorado rabino. Wald frequentou a universidade local, licenciando-se em 1926. Depois foi para a Universidade de Viena, onde estudou matemática com um distinto académico chamado Karl Menger.[4]

Em 1931, quando terminou o seu doutoramento, Wald despontara como um talento raro. Menger apelidou a dissertação do seu pupilo como «obra-prima de matemática pura», descrevendo-a como «profunda, bela, e de importância fundamental». Mas nenhuma universidade austríaca contrataria um judeu, não importa quão talentoso, ou quanto veementemente o seu famoso conselheiro o recomendasse. Assim, Wald procurou outras opções. Na verdade, ele disse a Menger que ficaria feliz por encontrar qualquer trabalho que lhe permitisse pagar as contas; tudo o que ele queria era continuar a demonstrar teoremas e a assistir a seminários de matemática.

No início, Wald trabalhou como professor particular de matemática de um banqueiro austríaco rico chamado Karl Schlesinger, a quem Wald ficou grato para sempre. Depois, em 1933, foi contratado como investigador pelo Instituto Austríaco para a Investigação de Ciclos Económicos, onde um outro académico famoso ficou impressionado com Wald: era o economista Oskar Morgenstern, o coinventor da teoria dos jogos. Wald trabalhou lado a lado com Morgenstern durante cinco anos, analisando a variação sazonal em dados económicos. Foi nesse instituto que pela primeira vez Wald se deparou com a estatística, um assunto que em breve viria a definir a sua vida profissional.

Mas sobre a Áustria acumulavam-se nuvens escuras. Como disse Menger, o conselheiro de Wald, «a cultura vienense assemelhava-se a uma cama de flores delicadas à qual o seu dono recusou solo e luz, enquanto um vizinho diabólico esperava uma oportunidade para destruir todo o jardim». A primavera de 1938 trouxe o

desastre: o *Anschluss*. A 11 de março, o líder eleito da Áustria, Kurt Schuschnigg, foi deposto por Hitler e substituído por um fantoche nazi. Numa questão de horas, 100 mil tropas da Wehrmatch alemã marcharam sem oposição através da fronteira, e a 15 de março desfilavam em Viena. Num amargo presságio, Karl Schlesinger, o benfeitor de Wald nos anos difíceis de 1931-1932, suicidou-se nesse mesmo dia.

Felizmente para Wald, o seu trabalho em estatísticas económicas recebera atenção no estrangeiro. No verão anterior, em 1937, ele tinha sido convidado para ir à América por um instituto de investigação económica de Colorado Springs. Embora estivesse satisfeito com o reconhecimento, inicialmente Wald estava hesitante em abandonar Viena. Mas o *Anschluss* fê-lo mudar de ideias, ao testemunhar os judeus da Áustria tornarem-se vítimas de uma orgia de assassínio, roubo e traição. As suas lojas foram saqueadas, as suas casas vandalizadas, os seus cargos na vida pública esbulhados pelas Leis de Nuremberga — incluindo o cargo de Wald, no Instituto para a Investigação de Ciclos Económicos. Wald ficou triste por dizer adeus a Viena, a sua segunda casa, mas ele podia ver os ventos de loucura a soprarem mais fortes a cada dia.

Assim, no verão de 1938, correndo grande perigo, atravessou sorrateiramente a fronteira com a Roménia e viajou em seguida para os EUA, iludindo os guardas que estavam sempre atentos a judeus que tentassem fugir do país. A decisão de partir provavelmente salvou-lhe a vida. Os pais de Wald, os avós e os cinco irmãos e irmãs ficaram na Europa — e todos, à exceção do

seu irmão Hermann, foram assassinados durante o Holocausto. Nessa época, Wald vivia na América. Ele estava em segurança e a trabalhar arduamente, casado e com dois filhos, e encontrou consolo nas alegrias da sua nova vida. No entanto, permaneceria tão devastado pela dor face ao destino da sua família que nunca mais tocou violino.

1.2 Wald na América

Contudo, Abraham Wald iria fazer mais do que a sua quota-parte para garantir que Hitler sofresse as consequências.

Wald tinha 35 anos quando chegou à América no verão de 1938. Ainda que tivesse saudades de Viena, de imediato gostou da sua nova casa. Colorado Springs evocava as encostas dos Cárpatos da sua juventude, e os seus novos colegas receberam-no com carinho e simpatia. Todavia, ele não permaneceu muito tempo em Colorado. Oskar Morgenstern, que tinha ele próprio fugido para a América, encontrava-se agora em Princeton, e estava a contar aos seus amigos matemáticos por toda a Costa Leste acerca do seu antigo colega Wald, que ele descrevia como um «cavalheiro» com «dons excecionais e grande poder matemático». A reputação de Wald continuou a crescer, e logo captou a atenção de um eminente professor de estatística em Nova Iorque chamado Harold Hotelling. No outono de 1938, Wald aceitou uma oferta para integrar o grupo de Hotelling na Universidade de Colúmbia. Começou como investigador associado, mas vingou tão rapidamente tanto como professor quanto como investigador que logo lhe ofereceram um cargo permanente na faculdade.

Nos finais de 1941, Wald já estava há três anos em Nova lorque, e os riscos do que se passava do outro lado do oceano eram claros para todos exceto para os obstinadamente cegos. Durante dois desses anos, a Grã-Bretanha tinha estado a combater sozinha contra os nazis, a lutar, como disse Churchill, «para salvar não só a Europa como também a humanidade». No entanto, durante esses dois longos anos, a América tinha permanecido à margem. Foi preciso o bombardeamento de Pearl Harbor para despertar o povo americano do seu torpor, mas finalmente acordaram. Os homens mais jovens logo avançaram para se alistarem. As mulheres foram trabalhar em fábricas e unidades de enfermagem. E os cientistas apressaram-se para os seus laboratórios e quadros de ardósia, principalmente os muitos emigrantes que aterrorizados fugiram dos nazis: Albert Einstein, John von Neumann, Edward Teller, Stanislaw Ulam e centenas de outros refugiados brilhantes que deram um impulso decisivo à ciência americana durante a guerra.

Abraham Wald também estava desejoso de responder à chamada. A sua oportunidade chegou prontamente, quando o seu colega W. Allen Wallis o convidou para se juntar ao Grupo de Investigação Estatística (SRG — Statistical Research Group) de Colúmbia. O SRG tinha sido inaugurado em 1942 por quatro estatísticos que se reuniam periodicamente numa sala velha no Rockefeller Center, no centro de Manhattan, para prestar consultoria estatística aos militares. Enquanto académicos, eles inicialmente não estavam acostumados a fornecer conselhos sob pressão. Por vezes isto originava episódios que comicamente revelaram uma fraca perspetiva das exigências da guerra. Nos primeiros tempos do SRG, um matemático queixou-se

ressentidamente acerca de ser forçado por uma secretária a poupar papel, tendo de escrever as equações em ambos os lados da folha.

Mas os dias de aprendizagem não duraram muito. Em 1944, o Grupo de Investigação Estatística tinha amadurecido e transformou-se numa equipa de 16 estatísticos e 30 jovens mulheres das universidades de Hunter e Vassar que tratavam do trabalho computacional. A equipa tornou-se uma fonte indispensável de aconselhamento técnico para os militares do Departamento de Pesquisa Científica e Desenvolvimento, e a sua orientação era procurada pelos mais altos níveis de comando — e obtinham resultados. Os estatísticos de Colúmbia não desenvolveram algo tão temível ou famoso como as equipas reunidas na mesma época em Los Alamos ou em Bletchley Park, mas a sua missão era mais ampla, e o seu efeito na guerra foi profundo. Eles estudaram combustíveis para foguetes, torpedos, fusíveis de proximidade, a geometria do combate aéreo ou a vulnerabilidade dos navios mercantes — qualquer coisa que envolvesse matemática e que fizesse avançar o esforço de guerra. Tal como o diretor do grupo, Wallis, mais tarde recordou:

> Durante a Batalha das Ardenas em dezembro de 1944, vários oficiais de alta patente do Exército voaram da batalha até Washington, passaram um dia a debater as melhores configurações dos fusíveis de proximidade para rajadas aéreas de projéteis de artilharia contra tropas terrestres e voaram de volta para a batalha [...] Este tipo de responsabilidade, embora raras vezes mencionado, estava sempre presente na atmosfera e exercia uma pressão poderosa, generalizada e ininterrupta.[5]

Felizmente, era uma equipa de investigadores com algumas das melhores mentes matemáticas do país, muitas das quais iriam

liderar as suas matérias de eleição: dois vieram a ser reitores de universidades; quatro exerceram as funções de presidente da Associação Americana de Estatística; Mina Rees tornou-se a primeira mulher presidente da Associação Americana para o Desenvolvimento da Ciência; Milton Friedman e George Stigler receberam o Prémio Nobel da Economia. E nesta equipa de estrelas, Abraham Wald era uma espécie de LeBron James: o homem que fazia tudo. Apenas os problemas mais difíceis faziam o caminho até à sua secretária, pois até mesmo os seus colegas génios reconheciam que, nas palavras do diretor do grupo, «o tempo do Wald era demasiado valioso para ser desperdiçado».

2. Wald e os Aviões Desaparecidos

O contributo mais célebre de Wald para o trabalho do grupo foi um artigo em que inventou um ramo da análise de dados, conhecido por amostragem sequencial. A sua perspicácia matemática mostrou às fábricas como poderiam produzir menos peças defeituosas para tanques e aviões simplesmente através da implementação de protocolos de inspeção mais inteligentes. Quando este artigo foi desclassificado pelos militares, transformou Wald numa celebridade académica e alterou o rumo das estatísticas do século XX, pois investigadores de todo o lado apressaram-se a aplicar as ideias matemáticas de Wald em novas áreas — especialmente em ensaios clínicos, onde ainda hoje esse conhecimento é aplicado.

Mas aqui a nossa história, acerca do crescimento exponencial da personalização ao estilo *Netflix*, está relacionada com uma outra contribuição de Abraham Wald universalmente incompreendida: o

seu método para elaborar recomendações personalizadas da capacidade de sobrevivência para aeronaves.

Todos os dias, as forças aéreas aliadas enviavam imensos esquadrões de aviões para atacarem alvos nazis e muitos aviões regressavam tendo sofrido danos de fogo inimigo. A dada altura, alguém na Marinha teve a ideia inteligente de analisar a distribuição dos danos nestes aviões regressados. O pensamento era simples: se se conseguisse encontrar padrões de onde os aviões estavam a ser atingidos, então podia-se recomendar o local para os reforçar com blindagem adicional. Estas recomendações, ademais, podiam ser personalizadas para cada avião, dado que os riscos para um ágil caça *P-51* eram muito diferentes dos que seriam para um pesado bombardeiro *B-17*.

A estratégia simplista seria colocar mais blindagem onde quer que se observasse mais buracos de balas nos aviões regressados. Mas isto seria uma má ideia, porque a Marinha não tinha quaisquer dados sobre os aviões que eram abatidos. Para ter uma ideia de como isto é tão importante, considere um exemplo extremo. Suponha que um bombardeiro podia ser abatido com um único tiro no motor, mas que era invulnerável a tiros na fuselagem. Se isso fosse verdade, então os analistas de dados da Marinha veriam centenas de aviões a regressar com buracos de balas inócuos na fuselagem — mas não veriam um único a regressar com buracos em volta do motor, dado que todos esses aviões ter-se-iam despenhado. Sob este cenário, se simplesmente se adicionasse blindagem onde se vissem buracos de balas — na fuselagem —,

então na verdade estar-se-ia a afetar os bombardeiros, adicionando peso que os «protegeria» de um perigo inexistente.

Este exemplo ilustra um caso extremo de viés de sobrevivência. Embora o mundo real seja muito menos extremo — balas no motor não são cem por cento letais, nem na fuselagem são cem por cento inócuas — mantém-se o ponto estatístico: o padrão de danos nos aviões regressados tinha de ser cuidadosamente analisado.

Nesta encruzilhada, temos de fazer uma pausa para referir dois apartes importantes. Primeiro, a Internet gosta imenso desta história. Segundo, praticamente todas as pessoas que já a contaram — com a notável exceção de um artigo obscuro e altamente técnico publicado no *Journal of the American Statistical Association* em 1984 — interpretam-na erradamente.[6]

Experimente pesquisar «Abraham Wald» e «Segunda Guerra Mundial» no *Google* e veja o que encontra: publicações em blogues, umas atrás das outras, acerca de como um intrépido matemático impediu que aqueles palermas da Marinha fizessem uma asneira terrível e colocassem um monte de blindagem desnecessária na fuselagem dos aviões. Nós lemos dúzias de coisas destas e poupámos-lhe a mesma triste tarefa ao criarmos o seguinte retrato falado.

> Durante a Segunda Guerra Mundial, a Marinha descobriu um padrão evidente de danos nos aviões que regressavam de missões de bombardeamento na Alemanha, no qual a maioria dos buracos de bala encontrava-se na fuselagem. Os tipos da Marinha chegaram à conclusão óbvia: colocar mais blindagem na fuselagem. Todavia, eles deram os seus

> dados a Abraham Wald, só para verificar novamente. As pequenas células cinzentas de Wald começaram a trabalhar. E então surge um relâmpago. «Esperem!», exclama Wald. «Isso está errado. Nós não vemos quaisquer danos nos motores porque os aviões que são atingidos no motor nunca regressam. Vocês têm de adicionar mais blindagem ao motor, não à fuselagem.» Wald tinha assinalado a falha crucial no raciocínio da Marinha: viés de sobrevivência. O seu conselho final capaz de salvar vidas foi exatamente em sentido contrário ao dos outros supostos especialistas: *coloquem a blindagem onde não veem buracos de balas*.

Conseguimos perceber por que esta versão da história é tão irresistível: o percurso da contraintuição acaba por fazer uma volta completa de 360 graus. Imagine perguntar a qualquer pessoa na rua, «Onde é que devemos colocar blindagem extra em aviões para os ajudar a sobreviver ao fogo inimigo?». Embora ainda não tenhamos feito esta sondagem, suspeitamos de que «no motor» seria uma resposta popular. Mas numa primeira fase uma interpretação simplista dos dados parece sugerir outra possibilidade: se os aviões regressados sofreram danos na fuselagem, então, por amor de Deus, vamos antes colocar aí a blindagem. Só um génio como Wald, prossegue a história, consegue ver o cerne da questão, levando-nos de volta à nossa conclusão intuitiva inicial.

Lamentavelmente, tanto quanto conseguimos perceber através dos registos históricos, este relato tem muito pouco fundamento. Pior ainda, esta versão embelezada, em que a moral da história se prende com o viés de sobrevivência, passa ao lado do que é verdadeiramente importante acerca do contributo de Abraham Wald para o esforço de guerra dos Aliados. O problema obviamente era o viés de sobrevivência nos dados, e toda a gente sabia isso, de outro modo não teria havido qualquer razão para à partida chamar

o Grupo de Investigação Estatística — a Marinha não precisava de um monte de professores de matemática só para contarem buracos de balas. A pergunta deles era mais específica: como estimar a probabilidade condicionada de uma aeronave sobreviver ao fogo inimigo num local específico, apesar do facto de grande parte dos dados relevantes não estarem disponíveis? O pessoal da Marinha não sabia como fazer isto. Eles eram realmente inteligentes, mas não é insulto dizer que não eram tão inteligentes quanto Abraham Wald.

A verdadeira contribuição de Wald foi muito mais subtil e mais interessante do que apresentar com bazófia o viés de sobrevivência a uma caricatura pateta de um comandante da Marinha. O seu toque de mestre não foi identificar o problema, mas inventar uma solução: um «sistema de recomendação sobre a capacidade de sobrevivência» ou um método que poderia fornecer aos comandantes militares recomendações personalizadas acerca de como melhorar a capacidade de sobrevivência de qualquer modelo de aeronave, utilizando os dados de danos em combate. O algoritmo de Wald era, segundo o diretor do Grupo de Investigação Estatística, uma «obra engenhosa da autoria de uma das maiores figuras na história da estatística americana».

Embora o algoritmo de Wald só tenha sido publicado na década de 1980, ele foi usado nos bastidores da Segunda Guerra Mundial e durante muitos anos depois dela.[7] Na Guerra do Vietname, a Marinha usou o algoritmo de Wald no *A-4 Skyhawk*; anos mais tarde, a Força Aérea empregou-o para aperfeiçoar a blindagem do

B-52 Stratofortress, a aeronave com mais tempo de serviço na história militar americana.

3. Dados em Falta: O Que Não Conheces Pode Enganar-te

Como agora pode entender, o problema de melhorar a capacidade de sobrevivência das aeronaves com que Abraham Wald se deparou era bastante semelhante ao da *Netflix* de fazer sugestões personalizadas de filmes. Mas há um porém, e é bem grande.

> Marinha dos Estados Unidos, 1943: «Queremos estimar a probabilidade condicional de um avião ser abatido, dado que sofre fogo inimigo num dado local, com base em dados de danos de todos os outros aviões. Isto irá permitir-nos personalizar recomendações sobre a capacidade de sobrevivência para cada modelo de avião. Mas faltam muitos dados, pois os aviões abatidos nunca regressam.»
>
> *Netflix*, 70 anos mais tarde: «Queremos estimar a probabilidade condicional de um subscritor gostar de um filme, dado o historial pessoal de visionamento dele ou dela, com base nas avaliações de todos os outros subscritores. Isto irá permitir-nos personalizar as recomendações de filmes para cada espetador. Mas faltam muitos dados: grande parte dos subscritores não viu a maioria dos filmes.»

O dilema é que tanto Abraham Wald como a *Netflix* precisaram de estimar uma probabilidade condicional, mas ambos se depararam com o problema dos dados em falta. E por vezes o que está em falta pode ser muito informativo.

Considere, por exemplo, algo que aconteceu quando um dos seus autores (Polson, britânico) visitou o outro (Scott, texano) em Austin pela primeira vez. Num passeio até uma cafetaria local, reparámos numa enorme carrinha branca estacionada na rua onde se lia:

ARMADILLO*

PET CARE†

Imagine a perplexidade de Polson face à ideia de um próspero negócio local dedicado às necessidades destas criaturas pouco britânicas. Como seriam os tatus enquanto animais de estimação? Será que reconheceriam os seus nomes? E porquê uma carrinha tão grande? Mas depois um estafeta moveu um carrinho empilhado com embalagens que estava em frente da carrinha, e a verdade banal foi revelada:

ARMADILLO

CARPET CARE‡

Às vezes, a parte dos dados em falta altera toda a história. Aconteceu o mesmo com os dados de Abraham Wald sobre a capacidade de sobrevivência de aviões. Embora os valores em bruto estejam perdidos para a história, podemos usar o seu relatório para a Marinha para conjeturar o que ele terá visto. Vamos imaginar que seguimos os passos de Wald enquanto ele analisa os dados acerca do raide de Schweinfurt-Ratisbona de agosto de 1943, em que os

Aliados num único dia perderam 60 dos seus 376 aviões. Os relatórios originais vindos do terreno teriam tido um aspeto parecido com este, onde um ponto de interrogação significa «dado em falta»:

Avião	Tido de dano	Resultado da missão
1) Hellcat Agnes	Fuselagem	Regressou a casa
2) The Bronx Bomber	?	Abatido
3) Pistol Pack'n Papa	Motor	Regressou a casa
…	…	…
375) Homesick Angel	?	Abatido
376) Calamity Jane	Nenhum	Regressou a casa

A partir deste relatório, Wald poderia ter feito uma tabulação cruzada com os aviões, o tipo de danos e o resultado da missão.* Isto teria produzido a tabela seguinte:

Tipo de dano sofrido	Regressados (total 316)	Abatidos (total 60)
Motor	29	?
Cockpit	36	?
Fuselagem	105	?
Nenhum	146	0

Dos 316 aviões que regressaram, 105 sofreram danos na fuselagem. Este facto teria permitido a Wald estimar a probabilidade condicionada de um avião sofrer danos na fuselagem, dado que regressa em segurança:

P (dano na fuselagem | regressa em segurança) = 105/316 = 32%

Mas essa é a resposta correta à pergunta errada. Em vez disso, o que queremos saber é exatamente o inverso: a probabilidade condicionada de um avião regressar em segurança, dado que sofreu danos na fuselagem. Este número poderá ser muito diferente.

Isto conduz-nos a uma regra importante acerca das probabilidades condicionadas: elas não são simétricas. Só porque Wald conhecia *P* (dano na fuselagem | regressa em segurança), não conhecia necessariamente a probabilidade inversa, *P* (regressa em segurança | dano na fuselagem). Para ilustrar porque não, considere um exemplo simples:

- Todos os jogadores da NBA praticam basquetebol. Isso significa que *P* (pratica basquetebol | joga na NBA) é quase 100%.

- Uma fração infinitamente baixa daqueles que praticam basquetebol chegará à NBA, o que significa que *P* (joga na NBA | pratica basquetebol) é praticamente 0%.

Assim, *P* (pratica basquetebol | joga na NBA) não é equivalente a *P* (joga na NBA | pratica basquetebol). Quando se pensa em probabilidades, é muito importante ser-se claro acerca de qual o evento que está do lado esquerdo da barra e qual o evento que está do lado direito.

Wald sabia isto. Ele sabia que para calcular uma probabilidade como *P* (avião regressa em segurança | dano na fuselagem) precisava de estimar quantos aviões tinham sofrido danos na fuselagem e *nunca regressaram a casa*. A sua tarefa era colocar números reais no lugar daqueles pontos de interrogação na tabela atrás: ou seja, preencher os dados em falta através da reconstrução da assinatura estatística dos aviões abatidos. Os cientistas de dados chamam a este processo «imputação». Em geral é muito melhor do que «amputação», que significa simplesmente cortar os dados em falta.

As incursões de Wald na imputação dependiam das suas premissas de modelação. Ele tinha de recriar o encontro típico de um *B-17* com o inimigo, usando apenas o testemunho mudo dos buracos de bala nos aviões que regressaram, conjugado com o modelo hipotético de uma batalha aérea. Para garantir que as suas premissas de modelação eram tão realistas quanto possível, Wald começou a trabalhar como se fosse um cientista forense. Analisou o ângulo de ataque provável dos caças inimigos. Conversou com engenheiros. Estudou as propriedades de uma nuvem de estilhaços resultantes de uma bateria antiaérea. Até sugeriu que o Exército disparasse milhares de balas falsas contra um avião, para que ele pudesse tabular os pontos de contacto.

Depois de tudo dito e feito, Wald tinha inventado um método para reconstruir toda a tabela. Com base no seu modelo de batalhas aéreas, as estimativas teriam mais ou menos este aspeto:

Tipo de dano sofrido	Regressados (total 316)	Abatidos (total 60)
Motor	29	31
Cockpit	36	21
Fuselagem	105	8
Nenhum	146	0

A partir de um conjunto completo de dados como este, é agora relativamente simples estimar as probabilidades condicionadas de que Wald necessitava. Por exemplo: dos 113 aviões atingidos na fuselagem, 105 regressaram a casa, e estima-se que oito não o fizeram. Consequentemente, a probabilidade condicionada de regressar em segurança, dado ter sofrido danos na fuselagem, é

$$P(\text{avião regressa em segurança} \mid \text{dano na fuselagem}) = \frac{105}{(105+8)} \approx 93\%$$

De acordo com esta estimativa, era muito provável que um *B-17* sobrevivesse a um tiro na fuselagem.

Por outro lado, dos 60 aviões que sofreram danos no motor, apenas 29 regressaram em segurança. Assim:

$$P(\text{regressa em segurança} \mid \text{dano no motor}) = \frac{29}{(29+31)} \approx 48\%$$

Os bombardeiros tinham uma probabilidade muito mais elevada de serem abatidos se sofressem danos no motor.

Este, finalmente, era o tipo de dados que a Marinha podia usar. Mas mais do que apenas os dados para um avião específico,

também podia usar a abordagem de Wald para personalizar as recomendações de capacidade de sobrevivência para *qualquer* avião. A probabilidade condicionada mais a criteriosa modelação dos dados em falta provaram ser uma combinação capaz de salvar vidas.

4. Bombardeiros Desaparecidos, Avaliações Desaparecidas

Setenta anos mais tarde, estas mesmas ideias iriam desempenhar um papel fundamental na maneira como a Netflix se reinventou enquanto empresa.

Tudo começou a partir do sistema de recomendações da *Netflix 1.0*, que iremos explicar aqui em linhas gerais. Imagine que enfrenta a tarefa assustadora de conceber este sistema. Como *input*, o sistema tem de aceitar o histórico de visualizações de um subscritor, e como *output* tem de produzir uma previsão acerca de esse subscritor ir ou não gostar de um determinado programa. Decide então começar com um exemplo simples inspirado por Wald: aferir quão provável é que um subscritor goste de *O Resgate do Soldado Ryan*, dado que ele ou ela gostou da série da HBO *Irmãos de Armas*. Isto parece uma boa aposta: ambos são dramas épicos acerca da invasão da Normandia e as suas consequências.

Para este emparelhamento particular de programas, força: recomende à vontade. Tenha em mente, no entanto, que quer ser capaz de fazer isto automaticamente. Certamente não seria

rentável colocar uma enorme equipa de anotadores humanos no círculo de sugestões, a identificar laboriosamente todos os possíveis pares de filmes através das similaridades. Mas agora recorde que tem acesso a toda a base de dados da *Netflix* mostrando quais os clientes que gostaram de que filmes. O seu objetivo é potenciar este vasto recurso de dados para automatizar o sistema de recomendações.

A ideia-chave é enquadrar o problema em termos de probabilidade condicionada. Suponha que, para um qualquer par de filmes A e B, a probabilidade *P* (subscritor aleatório gosta do filme A | o mesmo subscritor aleatório gosta do filme B) é elevada — digamos, 80 por cento. Agora sabemos através do histórico de visualizações da Linda que ela gostou do filme B, mas ainda não viu o filme A. Não seria o filme A uma boa recomendação? Com base no facto de ela ter gostado do filme B, há uma probabilidade de 80 por cento de gostar do A.

Mas como é que podemos descobrir um número como *P* (subscritor gosta de *O Resgate do Soldado Ryan* | subscritor gosta de *Irmãos de Armas*)? É nesta situação que a sua base de dados se torna útil. Para manter os números simples, digamos que existem 100 pessoas na sua base de dados, e que cada uma delas viu os dois filmes. Os históricos de visualização delas apresentam-se sob a forma de uma enorme «matriz de avaliações», onde as linhas correspondem a subscritores e as colunas a filmes:

Subscrito	**Gostou de *O Resgate do Saldado Ryan*?**	**Gostou de *Irmãos de Armas*?**
1. Aaron	Sim	Sim

2. Alice	Sim	Sim
…	…	…
99. Wendy	Não	Não
100. Zack	Sim	Não

A seguir, fará uma tabulação cruzada com os dados da matriz de avaliações, contando o número de subscritores que tiveram uma combinação específica de preferências para estes dois filmes:

	Gostou de *Irmãos de Armas*	**Não Gostou**
Gostou de *O Resgate do Soldado Ryan*	56 subscritores	6 subscritores
Não Gostou	14 subscritores	24 subscritores

A partir desta tabela podemos facilmente calcular a probabilidade condicionada de que o seu sistema de recomendações necessita:

- 70 subscritores gostaram de *Irmãos de Armas* (56 + 14).
- Destes 70 subscritores, 56 gostaram de *O Resgate do Soldado Ryan*, e 14 não gostaram.

Isto permite-lhe calcular a probabilidade condicionada de alguém que gostou de *Irmãos de Armas* gostar também de *O Resgate do Soldado Ryan*:

$$\text{P (gosta de O Resgate do Soldado Ryan | gosta de Irmãos de Armas)} = \frac{56}{(56 + 14)} \approx 80\%$$

O elemento-chave que faz com que esta abordagem funcione tão bem é ela ser automática. Os computadores não são muito bons (por enquanto) a analisarem automaticamente os filmes para determinarem o conteúdo temático. Mas são brilhantes na contabilização — ou seja, a efetuarem tabulações cruzadas de vastas bases de dados de históricos de visualização de filmes de subscritores a partir de matrizes de avaliação para determinarem probabilidades condicionadas.

O verdadeiro problema que a Netflix enfrenta é muito mais complicado do que este exemplo de faz de conta, pelo menos por três razões. A primeira é a escala. A *Netflix* não tem 100 subscritores, tem 100 milhões, e não dispõe de avaliações sobre dois programas, mas sobre mais de 10 mil. Como resultado, a matriz de avaliações tem mais de um *bilião* de entradas possíveis.

O segundo problema é o «que está em falta». A maioria dos subscritores não viu a maioria dos filmes, pelo que a maioria daqueles mais de um bilião de entradas está em falta. Além disso, tal como no caso dos bombardeiros da Segunda Guerra Mundial, esse padrão de ausência é informativo. Se ainda não viu o *Clube de Combate*, talvez ainda não tenha tido tempo para o ver — mas, por outro lado, talvez os filmes acerca de niilismo não sejam o seu género.

O último problema é a explosão combinatória. Ou, se preferir continuar com o *Clube de Combate* e com a filosofia em vez da matemática: cada subscritor da *Netflix* é um belo e único floco de neve fenomenológico. Numa base de dados com apenas dois filmes,

milhões de utilizadores partilharão as mesmas experiências de gostar/não gostar, dado que apenas são possíveis quatro dessas experiências: gostar dos dois, gostar de nenhum, ou gostar de um e não do outro. O mesmo não acontece numa base de dados com 10 mil filmes. Considere o seu próprio histórico de visualização cinematográfica. Nenhuma outra pessoa tem um histórico exatamente igual ao seu, e nunca terá, pois existem demasiadas maneiras de divergir. Mesmo numa base de dados com apenas 300 filmes, haveria imensamente mais combinações de gostar ou não gostar desses filmes (2^{300}) do que há átomos no Universo (cerca de 2^{272}). Muito antes de chegar aos $2^{10.000}$, mais vale parar de contar — as variedades de experiências de gostar de filmes são, para todos os efeitos práticos, infinitas.

Isto suscita uma questão importante: como pode a Netflix fazer uma recomendação com base no seu histórico de visualização, utilizando os históricos de visualização de outras pessoas, quando o seu é inédito e os delas nunca serão repetidos?

A solução para estes três problemas é a modelação criteriosa. Tal como Wald resolveu o problema de dados em falta ao construir um modelo de um embate entre um *B-17* e um caça inimigo, a Netflix resolveu o seu problema construindo um modelo de encontros de subscritores com um filme. E enquanto o modelo atual da *Netflix* é confidencial, o modelo de um milhão de dólares construído pela equipa BellKor's Pragmatic Chaos, vencedora do Prémio Netflix, está disponível gratuitamente na Internet.[8] Eis um resumo de como funciona. (Lembre-se, a *Netflix* prediz avaliações numa escala de 1 a 5, a partir das quais se pode efetuar uma

previsão de gosto/não gosto utilizando um simples ponto de corte, por exemplo quatro estrelas.)

Neste caso, a equação fundamental é:

Avaliação Prevista = Média Global + Ponderação Filme
+ Ponderação Utilizador + Interação Utilizador/Filme.

Os três primeiros elementos desta equação são fáceis de explicar.

- A avaliação média geral considerando todos os filmes é de 3,7 estrelas.

- Cada filme tem a sua própria ponderação. *A Lista de Schindler* e *A Paixão de Shakespeare* têm ponderações positivas porque são populares, enquanto *O Guarda-Fraldas* e *A Lei de Dredd* têm ponderações negativas porque não o são.

- Cada utilizador tem uma ponderação, porque certos utilizadores são mais ou menos críticos do que a média. Talvez o Vladimir seja impiedoso e avalie severamente todos os filmes (ponderação negativa), enquanto o Donald acha que todos os filmes são fantásticos e atribui-lhes uma pontuação muito elevada (ponderação positiva).

Estes três termos fornecem uma avaliação de base para qualquer par utilizador/filme. Por exemplo: imagine recomendar *007 — Agente Irresistível* (ponderação do filme = 0,4) ao Vlad, o severo (ponderação do utilizador = -0,2). A avaliação-base do Vlad seria 3,7 + 0,4 - 0,2 =3,9.

Mas esse é apenas um ponto de partida. Ele ignora a interação utilizador-filme, que é onde quase toda a ciência dos dados acontece. Para estimar esta interação, a equipa vencedora do prémio construiu algo chamado modelo de «caraterística latente». («Caraterística latente» significa simplesmente algo não medido de forma direta.) A ideia aqui é que as avaliações feitas por uma dada pessoa a filmes semelhantes revelam um padrão porque todas essas avaliações estão associadas a caraterísticas latentes dessa pessoa. As caraterísticas latentes de cada pessoa podem ser estimadas a partir de avaliações anteriores e utilizadas para efetuar previsões sobre o que ainda não viram. Esta mesma ideia surge em todo o lado, sob muitos e diferentes nomes:

- Os participantes em sondagens dão respostas semelhantes a perguntas acerca do seu trabalho e formação. Ambas estão associadas a uma caraterística latente, «estatuto socioeconómico», que também pode ser usada para prever a resposta de um inquirido a uma questão sobre rendimento. Os cientistas sociais chamam a isto «análise fatorial».

- Os senadores votam de forma semelhante em políticas de impostos e de cuidados de saúde. Ambas estão

relacionadas com uma caraterística latente, «ideologia», que também pode ser usada para prever o voto de um senador face a gastos com a defesa. Os cientistas políticos chamam a isto «modelo do ponto ideal».

- Os alunos que fazem exames nacionais têm padrões semelhantes de respostas acerca de geometria e álgebra. Ambas estão relacionadas com uma caraterística latente, «competência matemática», que também pode ser usada para prever a resposta de um estudante a uma questão sobre trigonometria. Os autores de testes chamam a isto «teoria da resposta ao item».

- Os subscritores da *Netflix* avaliam *Rockefeller 30 e Arrested Development* de formas semelhantes. Ambas estão relacionadas com uma caraterística latente — vamos chamar-lhe «afinidade com comédias excêntricas e inteligentes» — que também pode ser usada para prever a avaliação que um utilizador fará de *Parks and Recreation*. Os cientistas de dados chamam a isto «filtragem colaborativa baseada no utilizador».

Obviamente, não existe apenas uma caraterística latente para descrever os subscritores da Netflix, mas dezenas ou mesmo centenas. Há uma caraterística para «mistério de homicídio britânico», outra para «drama criminal com uma personagem sombria», uma terceira para «programa de culinária», uma ainda para «filmes de comédia moderna», e assim por diante. Estas caraterísticas formam os eixos coordenados de um gigantesco

espaço multidimensional no qual cada utilizador ocupa uma posição exclusiva, correspondente à combinação única de preferências do utilizador. Adora o *Poirot* mas não consegue lidar com a violência do *Narcos*? Talvez no eixo mistério de homicídio britânico seja +2,5 e no eixo drama-criminal -2,1. Adora *Os Tenembaus — Uma Comédia Genial* mas acha que o *The Great British Baking Show* é aborrecido? Talvez seja 3,1 em comédias modernas e -1,9 em programas de culinária.

A parte mais fascinante de todo este processo é que as caraterísticas latentes que definem estes eixos não são escolhidas previamente. Em vez disso, elas são descobertas de maneira orgânica pela IA, usando os padrões de correlação em dezenas de milhões de avaliações de utilizadores. Os dados — não um crítico ou um anotador humano — determinam que programas ficam juntos.

5. As Caraterísticas Escondidas Contam a História

Agora podemos, finalmente, concluir a nossa história sobre a personalização na IA. É a história de como as caraterísticas latentes ao nível do subscritor, descobertas através de conjuntos maciços de dados usando a probabilidade condicionada, constituíram a força escondida por trás da transformação estratégica da Netflix de distribuidora para produtora. É também a história de como estas caraterísticas latentes são o elixir mágico da economia digital — uma mistura especial de dados, algoritmos e conhecimento da realidade humana que representa a ferramenta mais perfeita alguma vez concebida para o marketing direcionado. As pessoas

que gerem a Netflix perceberam isto. Elas decidiram usar essa ferramenta para começarem a produzir programas de televisão e nunca mais olharam para trás. Pense no que torna a Netflix diferente enquanto produtora de conteúdos. Ao contrário das grandes cadeias de televisão, a Netflix não se importa com a sua idade, a etnicidade ou o lugar onde vive. Não se importa com o seu trabalho, a formação, o ordenado ou o seu género. E certamente não se importa com o que os anunciantes pensam, porque não tem. A única coisa com que a Netflix se importa é com quais são os programas de televisão de que você gosta — algo que ela compreende com um detalhe extraordinário, com base nas estimativas dela sobre as suas caraterísticas latentes.

Essas caraterísticas permitem que a Netflix segmente a sua base de subscritores em função de centenas de critérios diferentes. Gosta de dramas ou de comédias? É um fã de desporto? Aprecia programas de culinária? Adora musicais? Gosta de filmes com um elenco mais diverso? Assiste a todos os segundos dos filmes de ação, ou avança rapidamente quando surgem cenas violentas? Vê desenhos animados? Os padrões no seu próprio histórico de visualização, em conjunto com os aprendidos com base no histórico de todas as outras pessoas, fornecem uma resposta matemática precisa a cada uma destas perguntas e a centenas de outras. A sua combinação específica de caraterísticas latentes — o seu minúsculo cantinho de um espaço euclidiano gigantesco e multidimensional — transforma-o num público-alvo individual.

E é assim que a Netflix inventou um novo modelo de negócio de encomendar histórias fantásticas a artistas de classe mundial —

algumas dirigidas a uma miniaudiência e determinadas a outra. Um ótimo exemplo é a série *The Crown*, um drama complexo e opulento acerca dos primeiros anos da vida da rainha Isabel II. Desde 2017, *The Crown* é a série de televisão mais cara de sempre: 130 milhões de dólares por 10 episódios. Nesse orçamento estavam incluídos sete mil trajes de época, sendo o mais famoso o vestido de noiva de 35 mil dólares para o casamento real. Pode parecer que a Netflix está a gastar dinheiro em programas novos como se fosse um marinheiro bêbedo, mas lembre-se daquelas estatísticas macabras de um único ano das cadeias de televisão: 400 milhões de dólares para encomendar 113 episódios-pilotos, dos quais apenas 13 chegaram a uma segunda temporada. Quando a prática corrente da indústria é esbanjar centenas de milhões de dólares em programas destinados à irrelevância, até um vestido de noiva que custa 300 subscrições anuais da *Netflix* começa a parecer uma bagatela. De modo que em vez de um marinheiro bêbedo, uma melhor metáfora será uma vidente com uma bola de cristal — uma bola de cristal alimentada a dados, probabilística, capaz de dizer aos tipos da Netflix por que género de programas os seus subscritores pagariam 130 milhões de dólares. Quando o sabem, confiam que os artistas façam o resto.

Os números começam a mostrar que esta abordagem resulta. A Netflix não divulga estatísticas de visionamento, mas pelo menos temos uma métrica, que são os prémios. Em 2015, a Netflix estava em sexto lugar entre as cadeias de televisão em nomeações para os Emmy. Em 2017, estava em segundo lugar — as suas 91 nomeações apenas ficaram atrás do rol de 110 da HBO, e a HBO estava justificadamente preocupada com o que aconteceria quando

a imensamente popular *A Guerra dos Tronos* chegasse ao fim. Parece ser apenas uma questão de tempo até que serviços de *streaming* como a Netflix dominem o circuito de prémios.

De qualquer forma, a abordagem da Netflix à personalização já domina a economia digital. Se o futuro da vida digital tem a ver com sugestões e não com pesquisas, como acreditamos, então o futuro também tem a ver, inevitavelmente, com a probabilidade condicionada.

6. O Legado Misto dos Motores de Sugestão

Os motores de sugestão têm sido uma área de investigação importante na IA há uma década ou mais, tanto nas universidades como na indústria. Apesar de esse legado ainda se estar a desenvolver, vale a pena refletir sobre o ponto em que estamos. As novidades são variadas.

6.1 O Lado Negro do Marketing Direcionado

Primeiro as más notícias: estas tecnologias não têm sido apenas utilizadas para fazer sugestões acerca de coisas divertidas, como programas de televisão e música. Os motores de sugestão também têm um lado negro, que tem sido explorado de formas cínicas e divisionistas. Não há melhor exemplo do que o uso do *Facebook* por agentes russos nos meses que antecederam as eleições presidenciais americanas de 2016.

O *Facebook* é popular entre os anunciantes pela mesma razão que a *Netflix* é popular entre os espetadores de televisão: ele dominou a arte do marketing direcionado com base no seu rasto digital. Em épocas passadas, quando as empresas queriam alcançar uma certa demografia — estudantes universitários, por exemplo, ou pais com crianças em idade escolar —, essas empresas compravam anúncios em locais onde o seu público-alvo pudesse prestar atenção. Os profissionais de marketing tomavam estas decisões de comprar anúncios usando dados agregados acerca de que pessoas tendem a ver este programa ou a ler aquela revista. Mas não conseguiam visar indivíduos específicos. Como diz o velho ditado dos profissionais do marketing, metade de todos os dólares investidos em publicidade é perdida; só não sabemos qual é a metade em causa.

Mas se um anúncio costumava ser um instrumento pouco preciso, hoje é um raio laser. Os profissionais de marketing podem agora conceber um anúncio online para qualquer público-alvo que consigam imaginar, definido a um nível de demografia e de psicografia tão detalhado que o deixaria com a cabeça a andar à roda. Se fosse ter com a equipa de vendas do *Facebook* com o objetivo de alcançar um grupo vago como «jovens profissionais», por exemplo, eles provavelmente ririam nas suas costas. Diga-nos quem *realmente* quer alcançar, dir-lhe-iam. Quer advogados ou banqueiros? Democratas ou republicanos? Fãs de desporto ou conhecedores de ópera? Negro ou branco, homem ou mulher, Norte ou Sul, bife ou salada — e se salada, alface-icebergue ou couve? A lista não tem fim. Assim que tiver decidido acerca da sua audiência, os algoritmos do *Facebook* podem selecionar exatamente os

utilizadores que devem visar, e podem exibir um anúncio ou uma publicação patrocinada a esses utilizadores no exato momento em que é mais provável estarem recetivos à sua mensagem. Isto deixa os profissionais do marketing inebriados — e é a razão para o *Facebook* valer, quando estamos a escrever, mais de meio bilião de dólares, mais do que o PIB da Suécia.

Este tipo de marketing direcionado já acontece há algum tempo, e a julgar pelo comportamento dos utilizadores do *Facebook*, a maioria deles está disposta a aceitar a troca de «dados por mexericos» que está implícita no seu uso continuado da plataforma. Mas para muitas pessoas os sinais de alarme começaram a disparar no seguimento da eleição presidencial de 2016, quando ficou claro como a Rússia tinha explorado inteligentemente o sistema de direcionamento de anúncios para semear a discórdia entre os eleitores americanos. Os agentes russos, por exemplo, focaram-se num grupo de utilizadores que tinham ido ao *Facebook* para expressar solidariedade com os agentes da polícia no rescaldo de protestos do movimento Black Lives Matter (BLM — Vidas Negras Importam). Eles visaram estes utilizadores com um anúncio que continha uma imagem de um caixão coberto com uma bandeira no funeral de um polícia, a par da legenda: «Outro ataque infame à polícia por parte de um ativista do movimento BLM. Os nossos corações estão com esses 11 heróis.» Eles visaram um grupo de cristãos conservadores com um anúncio diferente: uma fotografia de Hillary Clinton a apertar a mão de uma mulher com um lenço na cabeça, a par de uma legenda produzida numa escrita pseudoarábica: «Apoie Hillary. Muçulmanos Americanos.» Os russos produziram anúncios diferentes para nova-iorquinos e para

texanos, para defensores da comunidade LGBTQ e para apoiantes da NRA (Associação Nacional de Armas), para veteranos de guerra e para ativistas dos direitos civis — todos visados com uma implacável eficiência algorítmica.[9]

Não conhecemos ninguém, de qualquer partido ou qualquer profissão, que não esteja horrorizado com a ideia de uma potência estrangeira hostil transformar as redes sociais em armas para influenciar uma eleição americana. E é evidente que a tecnologia por trás dos motores de sugestão foi pelo menos um ingrediente neste cocktail tóxico de dinheiro russo e políticas identitárias. Assim que nos afastamos desses pontos de consenso quase universal, contudo, as questões tornam-se muito mais complexas. Por exemplo:

1. Estas atividades alteraram o resultado da eleição presidencial? Provavelmente nunca saberemos, dado que não podemos aplicar a engenharia inversa aos processos de tomada de decisão das 138,8 milhões de pessoas que foram às urnas, ou das dezenas de milhões que ficaram em casa.

2. Seria diferente se um interveniente sediado nos Estados Unidos — digamos, os irmãos Koch, à direita, ou a Coligação Blue Dog, à esquerda — fizesse algo semelhante? Se acha que isso continuaria a ser censurável, e que o *Facebook* ou alguém deveria acabar com isso, confiaria nos seus opositores políticos para decidir exatamente onde fixar os limites?

3. Serão estas técnicas da Era digital qualitativamente mais eficazes a influenciar pessoas do que as técnicas que outros propagandistas, de Leni Riefenstahl aos programas na rádio, têm utilizado há anos? Esta é uma questão empírica simples: se as pessoas são visadas com anúncios digitais hiperespecíficos baseados no que os dados dizem acerca delas, e se os dados são bastante fidedignos, quantas mudam de opinião ou se comportam de maneira diferente? Se a resposta é que os anúncios não alteram quaisquer opiniões, mas apenas fazem com que as pessoas votem nas posições que já defendem — mais uma vez assumindo um comprador de anúncios americano —, isso é bom ou mau para a democracia?

4. O que, especificamente, deve ser feito neste momento? É certo que os algoritmos tiveram um papel no fiasco Rússia/*Facebook*, mas também o teve a nossa cultura política preexistente, tal como as nossas leis sobre publicidade, particularmente a publicidade política paga. Qual será a combinação certa de respostas legais e políticas para impedir que este tipo de coisas volte a acontecer? Indo mais longe, deve isto ser uma chamada de atenção para a *Era* do marketing digital direcionado, não só na política?

Não sabemos as respostas a estas perguntas, mas acreditamos que é possível ter uma conversa esclarecida sobre elas. Assim, em prol dessa conversa, eis duas coisas a considerar.

Em primeiro lugar, não nos ocorre exemplo mais claro do que o abuso do *Facebook* pela Rússia como razão para a sociedade não poder confiar na inteligência das máquinas sem supervisão humana e ainda assim esperar um futuro melhor. Os motores de sugestão não vão desaparecer, e não há outra opção senão criar um enquadramento cultural e legal de supervisão no qual possam ser usados de forma responsável. Estamos otimistas em que, dada a oportunidade, as pessoas possam ser suficientemente espertas para prevenir os piores abusos tecnológicos sem simplesmente destruírem todas a máquinas.

Em segundo lugar, não nos ocorre melhor razão do que esta conversa como motivo para que todos os cidadãos do século XXI necessitem de entender alguns factos básicos sobre inteligência artificial e ciência de dados. Se a educação, como disse Thomas Jefferson, é a pedra angular da democracia, então quando se trata de tecnologia digital, as nossas paredes democráticas estão a ruir. Os americanos têm debatido os limites da liberdade de expressão comercial quase desde o nosso nascimento como país. Mas hoje estamos muito para lá de uma conversa acerca de anúncios para cereais açucarados nos desenhos animados de sábado de manhã — longe do único exemplo de uma prática dúbia de marketing tornada totalmente antiquada pela nossa nova tecnologia. Existem *imensas* incertezas à nossa espera ao longo do caminho. No mínimo, tribunais e legislaturas deviam saber mais acerca dos seus próprios ângulos mortos e parar de desvalorizar detalhes que não compreendem considerando-os «algaraviada». E os cidadãos deviam participar nestas discussões a partir de uma posição de conhecimento, e não de receio, acerca dos detalhes técnicos

básicos. Em termos simples, pessoas inteligentes que se importam com o mundo simplesmente *devem* saber mais acerca da IA. Esta é uma das razões por que escrevemos este livro.

6.2 O Lado Positivo Para a Ciência

Agora algumas boas notícias acerca de motores de sugestão. Os conhecimentos matemáticos e algorítmicos produzidos na última década de trabalho sobre personalização começam presentemente a verter para outras áreas da ciência e da tecnologia. À medida que isso acontece, aguardam-nos muitas coisas boas.

Veja-se o caso das redes sociais centradas nos pacientes — como a *Crohnology*, para pessoas com distúrbios gastrointestinais; a *Tiatros*, para soldados com perturbação de stress pós-traumático; ou a *PatientsLikeMe*, para praticamente tudo. Estas redes também funcionam com algoritmos de personalização, tal como o *Facebook*. Os pacientes veem-nas como um recurso importante para sugestões sobre tratamentos e alterações no estilo de vida, enquanto os investigadores as encaram como um valioso repositório de dados médicos do mundo real, que podem ser usados para tornar ainda melhores aquelas sugestões.

Ou considere o crescente conjunto de ferramentas estatísticas dos neurocientistas, que agora, de forma rotineira, podem monitorizar a atividade de centenas de neurónios ao mesmo tempo, enquanto tentam compreender como é que o cérebro processa informação. Em breve, os avanços no *hardware* irão permitir-lhes monitorizar milhares de neurónios ou ainda mais. À medida que os

seus conjuntos de dados ficam cada vez maiores, os neurocientistas viram-se cada vez mais para modelos do tipo caraterística latente, ao jeito da *Netflix* para encontrar grupos de neurónios que tendem a disparar em conjunto como resposta a um certo estimulo — o equivalente neurofisiológico a gostar do mesmo programa de televisão. Este trabalho pode levar a novas descobertas e a tratamentos inovadores para alguns dos nossos males mais comuns, do autismo à Alzheimer.

Talvez o trabalho mais empolgante esteja a acontecer na investigação oncológica, especificamente em algo chamado «terapia direcionada». Ainda que o cancro possa ser identificado de acordo com a parte do corpo, fundamentalmente é uma doença do seu genoma. Os genomas tumorais, além disso, variam bastante. Mesmo os pacientes com o mesmo tipo de cancro podem ter tumores com subtipos genéticos diferentes, e os investigadores descobriram que frequentemente estes subtipos respondem de maneira muito diversa aos fármacos. Agora já é comum os médicos testarem a presença de proteínas e genes específicos numa amostra do tumor do paciente e, conforme o resultado, escolherem um fármaco anticancerígeno.

Ao longo dos anos, os investigadores na área da oncologia construíram extensas bases de dados com informação genética sobre diferentes tipos de tumores e uniram esforços com cientistas de dados para perscrutarem essas bases de dados à procura de padrões que possam ser explorados pelas terapias direcionadas. Por exemplo: cerca de 60 por cento dos tumores colorretais têm a versão selvagem (sem mutação) do gene KRAS. Um fármaco

anticancerígeno em particular, o *Cetuximab*, é eficaz contra estes tumores mas ineficaz contra os 40 por cento que têm a mutação KRAS.

Esse é um padrão simples, que envolve apenas um gene. Outros padrões, por sua vez, são bastante complexos, e envolvem dezenas ou centenas de genes relacionados com uma das muitas intrincadas vias de sinalização molecular que é danificada nas células cancerígenas. Para lidar com essa complexidade, os investigadores cada vez mais estão a virar-se para modelos de grandes volumes de dados de caraterística latente, do tipo que abriu caminho em Silicon Valley na última década para potencializar sistemas de recomendação em grande escala. Estes modelos estão a ser utilizados para analisar dados genómicos em busca de uma explicação para o que leva alguns doentes com cancro a responderem a um fármaco e a outros não. Tal como a *Netflix* usa as caraterísticas dos perfis de visualização dos subscritores para os visar com programas de televisão, os investigadores na área da oncologia esperam usar as caraterísticas dos «perfis genómicos» dos pacientes para os visar com terapias — e até talvez desenvolver terapias novas, ao estilo de *House of Cards*. Esta ideia começa a pegar. Por exemplo: em 2015, cientistas do Instituto Nacional do Cancro anunciaram a descoberta de dois subtipos distintos de linfomas difusos de grandes células B, com base em caraterísticas genómicas latentes. Os cientistas conjeturaram que os dois subtipos, ABC e GCB, poderão reagir de forma diferente a um fármaco particular, o *Ibrutinib*. Então, eles inscreveram 80 pacientes com linfoma num ensaio clínico, recolheram amostras dos seus tumores para determinar se eram do subtipo ABC ou do GCB,

administraram *Ibrutinib* a todos, e acompanharam os seus progressos ao longo dos meses e anos subsequentes. Os resultados foram impressionantes: a probabilidade de o *Ibrutinib* ser eficaz era sete vezes maior com o subtipo ABC.[10]

Dado o longo horizonte temporal e os milhares de milhões de dólares necessários para desenvolver e testar um novo fármaco anticancerígeno, esta estratégia de caraterização genómica ainda se encontra longe de estar consolidada. Mas, como o *Ibrutinib* demonstra, a probabilidade condicionada começa a pagar dividendos na pesquisa oncológica, e laboratórios de todo o mundo estão a trabalhar arduamente na próxima geração de terapias direcionadas.

7. Pós-Escrito

Esperamos que este capítulo o tenha ajudado a compreender um pouco mais acerca da ideia-chave por trás de empresas como a Netflix, o Spotify e o Facebook: que, para uma máquina, «personalização» significa «probabilidade condicionada». Também esperamos que tenha entendido que estes sistemas de IA modernos representam apenas um passo no longo e sinuoso percurso histórico do engenho humano — uma trajetória que certamente conduzirá a novas maravilhas, mas que também estará repleta de novos desafios.

Para encerrar este capítulo, deixamo-lo com uma última história sobre motores de sugestão que aconteceu connosco. Durante o

verão de 2014, Scott (um dos autores deste livro) visitou Ypres, uma cidade na zona ocidental da Bélgica, cuja posição estratégica foi motivo de grande preocupação no início da Primeira Guerra Mundial. O exército alemão e o dos Aliados encontraram-se às portas de Ypres em outubro de 1914. Ambos os lados cavaram trincheiras, e seguiu-se um brutal impasse que durou anos:

> «Homens marchavam adormecidos. Muitos haviam perdido as suas botas mas lá continuavam, calçados de sangue. Todos estavam mancos, todos cegos, bêbedos de fadiga; não escutavam nem mesmo o rugido dos obuses de gás que discretamente explodiam lá atrás.»
>
> *Wilfred Owen*

No final da terceira batalha de Ypres, em 1917, perto de meio milhão de soldados tinham morrido, e a cidade era um amontoado de ruínas.

Um século mais tarde, visitar a Ypres reconstruída é uma ocasião solene e, na sua visita em 2014, o Scott encontrou aquele sentido de solenidade reforçado por uma rede de altifalantes exteriores que debitavam música clássica pelo centro da cidade. Era um detalhe agradável, e todas as escolhas eram convencionalmente de bom gosto... até à intrusão inesperada de uma batida moderna de baixo. Ao início não foi fácil identificar a música, mas logo a letra dissipou quaisquer dúvidas: o responsável pela música em Ypres tinha escolhido tocar o êxito de 2006 «SexyBack», de Justin Timberlake, por toda a cidade.

Talvez tenha sido intencional. Ypres tinha de facto trazido o *sexy* de volta às suas ruas medievais, reconstruindo cada maravilhoso

tijolo após a Grande Guerra. Apesar disso, tendo em conta todas as músicas clássicas, pareceu uma escolha peculiar. De modo que, quando ele foi ao posto de turismo à procura de um mapa com os memoriais das batalhas circundantes, inocentemente perguntou à simpática senhora flamenga atrás do balcão se ela tinha alguma música preferida para os altifalantes da cidade.

«Ah, não», disse ela. «Na verdade, nós usamos simplesmente o *Spotify*.»

Por vezes, mesmo o melhor sistema de recomendação faz uma má sugestão.

---

[1] Citação de Kevin Spacey na conferência em memória de James Mac-Taggart no Fringe Festival de Edimburgo, 2013. Vídeo disponível no *YouTube* em https://www.youtube.com/watch?v=oheDqofa5NM;

[2] Nancy Hass, «And the Award for the Next HBO Goes to...», *GQ*, 29 de janeiro de 2013. https://www.gq.com/story/netflix-founder-reed-hastings-house-of-cards-arrested-development.

[3] Números retirados de *Statistical Abstract of the United States* (*Washington*, D. C.: USGPO, 1944, 1947, 1950), Departamento de Censos dos Estados, e de *Statistical Digest* (World War II), Forças Armadas do Exército, disponível em https://archive.org/details/ArmyAirForcesSatisticalDigestWorldWarII.

[4] O material sobre a vida de Abraham Wald foi obtido nas seguintes fontes: W. Allen Wallis, «The Statistical Research Group, 1942-1945», *Journal of the American Statistical Association 75, n.º 370* (junho de 1980): 320-30; Marc Mangel e Francisco J. Samaniego, «Abraham Wald's Work on Aircraft Survivability», *Journal of the American Statistical Association 79, n.º 386* (junho de 1984): 259-67, e veja também «Comment», de James O. Berger (267-69), e «Rejoinder» dos autores (270-71); J. Wolfowitz, «Abraham Wald, 1902-1950», *Annals of Mathematical Statistics 23, n.º 1* (1952): 1-13; Oskar Morgenstern, «Abraham Wald, 1902-1950», *Econometrica 19, n.º 4* (outubro de 1951): 361-67; Karl Menger, «The Formative Years of Abraham Wald and His Work in Geometry», *Annals of Mathematical Statistics 23, n.º 1* (1952): 14-20; L. Weiss, «Wald, Abraham», em *Leading Personalities in Statistical Sciences: From the Seventeenth Century to the Present*, ed. Norman L. Johnson e Samuel Kotz (Nova Iorque: John Wiley & Sons, 1997), 164-67; «Abraham Wald», *MacTutor History of Mathematics*, http://www-history.mcs.st-andrews.ac.uk/Biographies/Wald.html.

[5] W. Allen Wallis, «The Statistical Research Group, 1942-1945», *Journal of the American Statistical Association 75, n.º 370* (junho de 1980): 320-30.

[6] A nossa apresentação da abordagem de Wald usa anotações e termos modernos e é por isso intencionalmente anacrónica. Wald não definiu o problema exatamente nestes termos. Também deixámos de fora muitos detalhes técnicos. Encorajamos o leitor interessado a consultar «Abraham Wald's Work on Aircraft Survivability», de Marc Mangel e Francisco J. Samaniego, em *Journal of the American Statistical Association 79, n.º 386* (junho de 1984): 259-67; veja também «Comment», de James O. Berger, 267-69, e «Rejoinder» dos autores, 270-71.

[7] W. Allen Wallis, «The Statistical Research Group, 1942-1945», *Journal of the American Statistical Association 75, n.º 370* (junho de 1980): 320-30; Mangel e Samaniego, «Rejoinder».

* Em português, *tatu*. [N. T.]

† Em português, *cuidados para animais*. [N. T]

‡ Em português, cuidados para tapetes. Tratava-se, na verdade, de um serviço de limpeza de carpetes e não de cuidados com animais, confusão gerada pela parte da palavra «carpet» escondida pelo carrinho. [N. T]

* Para si, especialista em folhas de cálculo, isto é como construir uma tabela dinâmica.

8 https://www.netflixprize.com/community/topic1537.html (Já não disponível)

9 Dan Keating, Kevin Schaul e Leslie Shapiro, «The Facebook Ads Russians Targeted at Different Groups», *The Washington Post*, 1 de novembro de 2017, https://www.washingtonpost.com/graphics/2017/business/russian-ads-facebook-targeting/.

10 National Cancer Institute, «Study Shows Promise of Precision Medicine for Most Common Type of Lymphoma», 20 de julho de 2015, https://www.cancer.gov/news-events/press-releases/2015/ibrutinib-lymphoma-subtype.

CAPÍTULO 2

A Fabricante de Castiçais

O que é que medir o tamanho do Universo tem a ver com salvar as abelhas? A resposta reside no modo como os computadores aprendem a reconhecer padrões nos dados e em como utilizam esses padrões para tomar decisões surpreendentemente inteligentes.

Em 2017, funcionários públicos em Pequim perceberam que tinham um problema. Um importante elemento criminoso estava a operar no seio deles.

Felizmente, uma análise minuciosa a estas ações criminosas revelou um padrão: o seu alvo principal parecia ser o Parque do Templo do Céu, lar de uma das grandes obras-primas arquitetónicas da dinastia Ming, e local de grande significado espiritual para o povo chinês. Eles seguiam um modo de atuação muito específico, chegando cedo ao parque, misturando-se com os idosos que ali se reuniam para fazer exercício ou entoar músicas, e esperando pacientemente até terem a certeza de que os cofres dos seus alvos estavam cheios. A meio da manhã, poriam o seu plano em ação, esgueirando-se indolentemente até uma casa de banho pública próxima e permanecendo no seu interior um ou dois minutos, de modo a que o *timing* não levantasse suspeitas. Então

atacariam como um relâmpago, agarrando em todos os rolos de papel higiénico que conseguissem encontrar, enfiando-os numa mochila e saindo da casa de banho como se nada fosse. Estavam mais expostos durante aqueles poucos primeiros passos de regresso à luz, mas assim que alcançavam a multidão, estavam a salvo.

Estes ladrões tinham-se tornado excecionalmente competentes e ousados. Estavam a roubar grandes quantidades de papel higiénico e as autoridades de Pequim prepararam-se para os capturar.

O primeiro passo foi instalar dispensadores automáticos de papel higiénico em todas as casas de banho públicas perto do Templo do Céu, de modo a que cada pessoa recebesse exatamente 60 centímetros de papel, o equivalente a seis quadrados. No entanto, rapidamente se tornou claro que se estes ladrões não pudessem furtar papel higiénico em rolos, estavam dispostos a roubar seis quadrados de cada vez. Eles simplesmente deram uma volta pelo parque, recolhendo as suas porções de papel em cada casa de banho ao longo do caminho, até regressarem ao ponto de partida. Depois repetiam a volta vezes sem conta, como se fosse uma viagem num carrossel clepto-escatológico, até terem recolhido todo o papel higiénico disponível. A operação toda demorava muito mais do que naqueles dias descontraídos dos rolos desguarnecidos, mas o resultado era igualmente ruinoso para o orçamento para papel higiénico da cidade.

Claramente, estes ladrões não iriam ser parados por um dispensador de seis quadrados. Por momentos as autoridades consideraram a óbvia abordagem de inteligência humana: contratar seguranças para vigiar as casas de banho. Tendo em conta que estamos em 2017, contudo, em vez disso eles decidiram adotar a abordagem de inteligência artificial: instalaram câmaras e software de reconhecimento facial, potenciados por algo a que chamaram algoritmo de «aprendizagem profunda», em todas as casas de banho públicas do parque.

Atualmente, se espera utilizar uma casa de banho perto do Templo do Céu, deverá: 1) tirar o seu chapéu, óculos, máscara de Guy Fawkes, etc., e 2) olhar diretamente para a câmara, a qual, graças a Deus pelas pequenas dignidades, se encontra no exterior. Se o *software* detetar que nos últimos 10 minutos a sua cara surgiu numa casa de banho próxima, então azar: não tem direito aos seis quadrados.

Equipar casas de banho com inteligência artificial pode parecer uma solução extrema, ou talvez simplesmente sinistra, para o problema do roubo de papel higiénico. Muitas pessoas manifestaram preocupação quanto à privacidade — e, como era previsível, surgiram muitos problemas logísticos, desde longas filas a câmaras partidas e a identidades trocadas. O nosso objetivo ao relatar este exemplo certamente não é apoiar a abordagem de Pequim mas salientar um facto simples da vida: atualmente o reconhecimento de padrões com base em IA está por todo o lado — até em casas de banho. Portanto, se pretende compreender o

mundo moderno, é importante que entenda como funcionam estes sistemas e porque dependem tanto de dados.

1. Input/Output: Como as Máquinas Reconhecem Padrões

As pessoas são fantásticas a reconhecer padrões. Desde tenra idade, por exemplo, aprendemos a associar rostos a pessoas e muita da nossa educação subsequente consiste em aprender a aplicar os padrões certos:

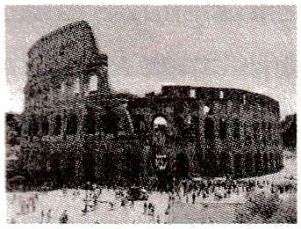
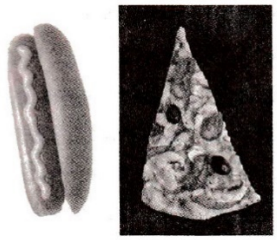
- Para falar, associamos um som ao significado certo.
- Para ler, associamos uma sequência de símbolos escritos à palavra certa.
- Para seguir a etiqueta, associamos uma pista social ao comportamento adequado.
- Para exercer medicina, associamos sintomas ao diagnóstico e à cura certa.
- Para ser um cientista de dados, associamos um conjunto de dados à maneira correta de os analisar.

Qualquer que seja a área do conhecimento, ser inteligente significa entender imensos padrões — como associar um *input* ao

output adequado. As pessoas não são os únicos seres capazes de reconhecer padrões. Por exemplo: o Scott tem um querido gatinho bicolor que detesta viagens de carro. O gato aprendeu que quando os seus donos começam a fazer as malas, ele vai ter de passar algum tempo no carro. Agora, de cada vez que alguém retira um saco de desporto do armário, o *Markov*, à cautela, esconde-se de imediato debaixo da cama.

Nos dias de hoje, os computadores também podem aprender padrões, tal como os gatos e as pessoas. Talvez se recorde da história de Makoto Koike, que construiu um separador de pepinos que tirava partido das capacidades de reconhecimento de padrões da IA. Ali, o *input* era uma imagem, o *output* era uma decisão de colocar o pepino numa das nove classes diferentes, e o padrão era a relação entre as caraterísticas visuais do pepino e a sua classe. Em IA chama-se a isso «classificação de imagens», e é usada em todo o lado — nas casas de banho em Pequim; pelo *Facebook*, para identificar os seus amigos em fotografias não marcadas; e pelo CERN, o monumental laboratório de física em Genebra, para detetar colisões entre partículas subatómicas em imagens de experiências de física de alta energia. Mas o *input* não precisa de ser uma imagem. Em última instância, os computadores são agnósticos em relação ao tipo de *input* que lhes damos, porque para um computador tudo é apenas números. O *input* pode ser uma onda sonora (para interpretar um pedido a um assistente digital doméstico), uma sequência de genes (para prever a suscetibilidade de alguém a uma doença), ou uma frase em inglês (para ser traduzida para castelhano). Como indica a tabela em baixo, tudo o que conseguir representar como um conjunto de números pode ser

usado como *input* num destes sistemas de reconhecimento de padrões. Como discutiremos mais tarde, no entanto, por vezes a maneira de representar um *input* como um número é óbvia, e outras vezes não.

Input	Output
	Geolocalização: «Roma, Itália.»
	Discurso para texto: «Ta-kus pekenwall'mosu aus-tin.»
 © National Cancer Institute; foto de Renee Comet	Classificação de imagem: «Cachorro-quente» / «Não cachorro quente».
20 ºC, 70% de humidade, céu praticamente limpo	Previsão numérica: «O consumo de energia em Londres será de 25.500 megawatts/hora.»
«Buenos días!»	Tradução: «Bom dia!»
«Ser, ou não ser…»	Atribuição a autor: »Shakespeare.»

Neste capítulo ficará a conhecer as duas ideias-chave por trás do funcionamento destes sistemas de reconhecimento de padrões:

1. Na IA, um «padrão» é uma regra de previsão que mapeia um *input* a um *output* esperado.

2. «Aprender um padrão» significa adequar uma boa regra de previsão a um conjunto de dados.

Isto envolve um pouco de matemática, mas não tenha medo: estas ideias tornam-se, na verdade, bastante simples e elegantes assim que as conhecemos e vamos passar o resto do capítulo a ajudá-lo a fazer justamente isso.

Vejamos primeiro um exemplo rápido do tipo de coisa de que estamos a falar. Talvez tenha ouvido a regra seguinte num website ou dita por um guru do exercício: para estimar a sua frequência cardíaca máxima, subtraia a sua idade a 220. Esta regra pode ser expressa como uma equação: FCM = 220 - Idade. Esta equação fornece uma descrição matemática de um padrão num conjunto de dados: a frequência cardíaca máxima (o *output*) tende a ser mais lenta com a idade (o *input*). Também lhe proporciona uma maneira de fazer previsões. Por exemplo: se tiver 35 anos predirá a sua frequência cardíaca ao introduzir na equação Idade = 35, o que produz FCM = 220 - 35, ou seja 185 batimentos por minuto.

Uma regra preditiva em IA é exatamente assim: uma equação que descreve um padrão na relação entre o *input* e o *output*. Se já tiver usado o conjunto de dados para descobrir uma boa regra de previsão, então cada vez que encontrar um novo *input* poderá introduzi-lo para prever o *output* correspondente — tal como pode introduzir a sua idade na equação «FCM = 220 - Idade» e ler a previsão para a sua frequência cardíaca máxima.

Eis um pouco de jargão. Na IA, as regras preditivas são geralmente apelidadas «modelos» — por exemplo, um «modelo de reconhecimento facial» para receber *input* de uma imagem e gerar a identidade de uma pessoa como *output*, ou um «modelo máquina de tradução» para receber uma frase em inglês como *input* e gerar uma tradução em castelhano como *output*. O processo de usar dados para encontrar uma boa regra de previsão é frequentemente chamado «treinar o modelo». Gostamos da palavra «treinar» neste contexto, porque evoca os ganhos progressivos na condição física, que se acumulam com cada nova sessão de treino no ginásio — ou, no caso de um modelo na IA, os melhoramentos progressivos na previsão que se acumulam com cada novo ponto de dados. Se não podemos ir nós ao ginásio, então pelo menos os nossos modelos podem.

Mas isto suscita um grande número de questões. O que significa «treinar um modelo» num conjunto de dados? O que torna um modelo melhor do que outro? Como é que explicaria isso a um computador — como é que ensinaria um algoritmo a encontrar o padrão certo num conjunto de dados? Ademais, os computadores não *pensam* apenas em termos de números? Como é que tudo isto funciona quando o *input* é algo complexo, como uma imagem ou uma onda sonora, em vez de um único número, como a idade de alguém? Talvez seja mais premente para aqueles que procuram um entendimento mais profundo da IA: Inicialmente, de onde surgiu esta ideia de treinar modelos com dados? E como é que esta ideia veio a desempenhar um papel tão central, ainda que invisível, nas nossas vidas, estando ali sentada ao fundo atrás de todas as coisas,

desde as redes sociais à terapia oncológica, do cultivo de pepinos à tradução de castelhano e de casas de banho a redes elétricas?

2. Uma Descoberta Brilhante

Para responder a estas questões começaremos por guiá-lo através de um exemplo alargado de como uma regra preditiva pode representar um padrão. Este padrão é exatamente igual aos que surgem constantemente na IA e é muito mais interessante do que o padrão idade *versus* frequência cardíaca. Este padrão, de facto, levou a um dos maiores triunfos intelectuais de todos os tempos, ao ajudar os cientistas a responder a uma pergunta que eles colocavam há milénios: qual é o tamanho do Universo?

Hoje, qualquer pessoa curiosa pode abrir um navegador de Internet e encontrar milhares de imagens captadas pelo telescópio espacial Hubble capazes de provocar um arrepio na espinha: galáxias a colidirem, vestígios de estrelas que explodiram, quasares distantes com a energia de um milhão de sóis. Os astrónomos de apenas há um século, contudo, dificilmente reconheceriam estas maravilhas. Para eles, o Universo era um lugar muito mais pequeno. Em 1924, a opinião científica esclarecida sustentava que a nossa galáxia, a Via Láctea, era a única galáxia no Universo — e para lá do seu horizonte havia apenas um vazio. Foi só no início do século XX que as pessoas finalmente descobriram a fantástica verdade: nós habitamos um vasto Universo com um bilião ou mais de galáxias.

Aqui iremos focar-nos em três caraterísticas essenciais na história desta grande descoberta:

1. Uma inexplicável mancha de luz no céu noturno, até visível aos antigos;

2. Um princípio matemático secular para reconhecimento de padrões que hoje potencia os nossos sistemas mais sofisticados de IA;

3. Uma astrónoma pouco conhecida do início do século XX chamada Henrietta Leavitt que, ao utilizar esse princípio, ensinou-nos a medir o tamanho do Universo.

Quando vir como estas três vertentes encaixam umas nas outras, ficará com um conhecimento muito mais rico de como as máquinas aprendem padrões no mundo à sua volta, e de como os usam para produzir predições surpreendentemente corretas — quer isso envolva classificar pepinos, reconhecer os seus amigos em fotos ou erradicar o roubo de papel higiénico em Pequim.

2.1 Uma «Mancha Enevoada» no Céu do Norte

Há mais de mil anos, observadores perspicazes repararam pela primeira vez em algumas pequenas mechas no céu — não estrelas, exatamente, talvez mais como nuvens turvas de luz. A maior, visível a olho nu numa noite escura, era um ponto luminoso na cintura de Andrómeda, uma constelação no céu do hemisfério norte.

No século X, o astrónomo persa Abd al-Rahman al-Sufi (também conhecido por Azofi) referiu-se a este objeto como uma «mancha enevoada»[1]. Nem AL-Sufi nem qualquer outra pessoa conseguiram perceber o que era aquela mancha. Quando o telescópio surgiu na primeira década do século XVII, o mistério da «Grande Nebulosa de Andrómeda» apenas aumentou, uma vez que os astrónomos começaram a descobrir ainda mais pequenas manchas como ela — muitas, tal como Andrómeda, com a forma inconfundível de uma espiral.

No século XX os astrónomos chamavam-lhes «nebulosas espirais», a partir da palavra latina para «névoa». Estas nebulosas clamavam por uma explicação. Seriam estrelas recém-nascidas? Seriam nuvens de gás luminoso nos confins da Via Láctea? Ou seria cada uma delas, como algumas pessoas afirmaram, uma galáxia distante tal como a nossa própria galáxia?[2]

Figura 1. Uma imagem contemporânea de Andrómeda obtida através do *Galaxy Evolution Explorer* da NASA. Cortesia NASA/JPL-Caltech

Esta última interpretação — de que as nebulosas espirais eram «universos-ilhas», o termo então usado para uma galáxia — tinha sido popular durante grande parte dos séculos XVIII e XIX. O seu defensor mais afamado havia sido o filósofo alemão Immanuel Kant. No início do século XX, todavia, a teoria do universo-ilha caíra em desuso. Não havia provas concretas que a apoiassem, de modo que a maioria dos astrónomos optou pela hipótese mais simples de «uma-galáxia». Concluíram que as espirais se situavam na periferia da Via Láctea e provavelmente eram nuvens de estrelas recém-formadas. A ideia de que elas eram galáxias independentes passou a ser vista como «pomposa» e «enganadora», uma noção que o manual de um astrónomo desse tempo descrevia como tão pateta que «praticamente dispensava qualquer discussão».[3]

Contudo, à medida que os telescópios ficavam melhores e se acumulavam novas provas, alguns astrónomos começaram a interrogar-se se teriam sido demasiadamente apressados a descartar a velha teoria da galáxia-independente. Um dos pontos a favor deles era o ritmo a que os astrónomos estavam a descobrir novas: «estrelas novas» que apareciam subitamente no céu noturno e em seguida desapareciam gradualmente durante semanas ou meses. As pessoas já andavam a ver as novas há centenas de anos, mas os poderosos novos telescópios do início do século XX apresentavam um facto curioso aos astrónomos: a Grande Nebulosa de Andrómeda parecia ter uma quantidade surpreendente de novas no seu interior — mais, na verdade, do que o resto da galáxia em conjunto. Se Andrómeda fosse apenas uma nuvem de pó na margem exterior da Via Láctea, como é que isto

poderia ser possível? Como é que um cantinho da nossa galáxia podia ser incomparavelmente mais rico em novas?

Depois havia a questão da enorme velocidade a que Andrómeda se deslocava. Em 1913, o astrónomo Vesto Slipher tinha medido cuidadosamente a velocidade dela usando um espetrómetro, uma «pistola-radar» cósmica que funciona explorando o efeito Doppler — o mesmo princípio que faz a sirene de uma ambulância soar mais alto à medida que se aproxima de si, e depois mais baixo enquanto se afasta. Os resultados de Slipher foram tão surpreendentes que ele próprio mal acreditou: Andrómeda estava a mover-se em relação à Terra à velocidade de 300 quilómetros *por segundo*, cerca de 20 vezes mais rápido do que qualquer outra coisa na Via Láctea. Ainda mais chocante era o facto de a maioria das outras nebulosas espirais estarem a deslocar-se ainda mais velozmente do que Andrómeda — muitas tão velozmente como mil quilómetros por segundo. Para muitos astrónomos, os resultados de Slipher encerravam o assunto: as espirais estavam a deslocar-se depressa demais para estarem dentro da nossa própria galáxia[4].

Os céticos da teoria da galáxia-independente, todavia, tinham uma réplica pronta. Se a nebulosa Andrómeda fosse uma galáxia do tamanho da nossa, seguir-se-iam duas conclusões aparentemente impossíveis. Por um lado, Andrómeda teria de estar a milhões de anos-luz, caso contrário seria muito mais brilhante no céu noturno. E se isso fosse verdade, então cada uma das novas de Andrómeda teria de estar a arder com a energia de milhões de sóis, senão nunca as veríamos de tão longe. Em retrospetiva, sabemos agora que ambas as «impossibilidades» são efetivamente

verdadeiras. Mas, do ponto de vista de muitos astrónomos do início do século XX, elas reduziam a teoria da galáxia independente a um absurdo.

Então, como é que os astrónomos deveriam interpretar todas aquelas nebulosas? Eram pequenas ou grandes? Nuvens de pó na nossa própria galáxia ou galáxias inteiras individuais? Ninguém possui provas decisivas num ou noutro sentido — o que deixou os astrónomos numa confusão horrível, pois sobre estas questões surgiu uma outra ainda mais profunda: qual é o tamanho, realmente, do universo? A dada altura, Copérnico tornara-nos mais humildes refutando a ideia de que a Terra era o centro da criação. Galileu tornara-nos mais humildes uma segunda vez, mostrando que a Via Láctea era um enorme amontoado de estrelas semelhantes à nossa galáxia. Será que iríamos receber uma terceira lição de humildade, descobrindo que a nossa galáxia não estava sozinha? Este foi o «grande debate» da astronomia, e prolongou-se durante a década de 1910 e início da década de 1920. E a única razão pela qual se prolongou foi porque ninguém conseguia responder a uma simples questão: a que distância se encontra a Grande Nebulosa de Andrómeda?

3. Como Podem Medir-se as Estrelas?

Imagine que está a conduzir numa estrada rural mal iluminada numa noite escura. Sobe uma colina e uma luz surge ao longe a piscar. A que distância é que ela se encontra? O que está a ver, será a luz ténue de um alpendre numa casa a cem metros de distância? Serão os faróis de outro carro a um quilómetro mais à

frente? Ou talvez algo mais distante mas muito mais brilhante, como o luzir de uma pequena cidade no vale a dez quilómetros de distância?

Nesta situação encontra-se perante o problema fundamental da astronomia. Os seus olhos podem dizer-lhe apenas quão brilhante um objeto *aparenta* ser, não quão brilhante efetivamente é na sua origem. Vénus, por exemplo, parece o objeto mais brilhante no céu noturno com exceção da Lua, mas apenas porque está tão perto. A estrela Alfa de Centauro, entretanto, parece cem vezes menos brilhante do que Vénus, mas apenas porque está a 40 biliões de quilómetros de distância. Ao perto, é mais brilhante do que o nosso Sol.

Um telescópio enfrenta o mesmo problema. Ele pode medir a aparente luminosidade de uma estrela, ou quão brilhante ela nos parece vista da Terra, mas não a sua verdadeira luminosidade ou quanta luz está efetivamente a emitir. Isto faz com que os astrónomos coloquem a mesma questão face a cada ponto de luz no céu: a luz é fraca e está perto, ou é forte e está longe?

Poderá perguntar-se: se isso é verdade, então como é que sabemos que a Alfa de Centauro está a 40 biliões de quilómetros de distância? A resposta é que os astrónomos utilizam um padrão muito útil chamado «paralaxe». Pode ver a paralaxe por si próprio jogando ao olho esquerdo/olho direito: levante o seu dedo indicador, ou talvez um dedo diferente se estiver a amaldiçoar-nos por tanta matemática, e mantenha-o alguns centímetros à frente do seu nariz. Olhe para ele primeiro com um olho fechado, depois

com o outro. Continue a alternar os olhos. Deve reparar que o seu dedo parece mover-se de um olho para o outro. Esse movimento aparente é a paralaxe.

E agora eis o padrão: quanto mais longe do nariz colocar o seu dedo, menos ele parecerá mover-se enquanto alterna entre os olhos. Pode descrever esse padrão matematicamente, como uma equação que relaciona o movimento aparente do seu dedo, ou paralaxe, com a distância ao seu nariz. A equação enuncia que, à medida que a paralaxe diminui, a distância aumenta: distância = 1/paralaxe. Isto pode ser deduzido através da trigonometria; vamos poupá-lo aos detalhes, porque o importante é que se trata novamente apenas de um caso de «*output* = função do *input*», como a regra preditiva que relaciona a sua frequência cardíaca máxima com a sua idade.

Os astrónomos medem a distância até estrelas próximas através do mesmo jogo de olho esquerdo/olho direito. Especificamente, eles usam duas imagens telescópicas da estrela obtidas com seis meses de intervalo. Isto permite que a Terra complete meia órbita em torno do Sol, maximizando a separação entre os «olhos» direito e esquerdo do astrónomo. Eles comparam estas duas imagens para medir a paralaxe da estrela, cujo valor introduzem na equação atrás para obter uma previsão do distanciamento da estrela.

Mas o grande inconveniente do padrão paralaxe/distância é que, enquanto fita métrica cósmica, não chega muito longe. Se um objeto se encontra a uma distância de mais de cerca de 300 anos-luz, a sua paralaxe será demasiado pequena para ser medida com

fiabilidade — e 300 anos-luz é pouco mais do que um centímetro em termos galáticos. Mesmo no início do século XX, quando o grande debate acerca da nebulosa espiral estava no auge da sua ferocidade, todos aceitaram que a Via Láctea se encontrava a pelo menos dezenas de milhares de anos-luz de distância, senão mais. Portanto, ambos os lados reconheceram que independentemente de quem estava certo, a distância a Andrómeda era demasiadamente longínqua para ser medida através da paralaxe. Os astrónomos desesperavam por uma melhor maneira de calcular a distância, mas ninguém tinha alguma.

Ninguém tinha até uma astrónoma pouco conhecida, chamada Henrietta Leavitt, ter feito uma descoberta maravilhosa — uma regra preditiva nova que iria permitir aos astrónomos medirem distâncias superiores a milhões de anos-luz, muito mais longe do que eles alguma vez pensaram ser possível. Ela não descobriu a regra recorrendo à trigonometria, que havia sido a forma como a regra da paralaxe fora estabelecida. Em vez disso, descobriu-a usando dados, aplicando o mesmo princípio que o *Google*, a *Apple* e o *Facebook* usam hoje para construir os seus sistemas de reconhecimento de padrões.

4. A Grande Descoberta de Henrietta Leavitt

Henrietta Leavitt tornou-se astrónoma quase por acaso. Nascida em 1868, numa família numerosa em Lancaster, no Massachusetts, entrou no Radcliffe College em 1888 para estudar humanidades. Foi apenas no seu último ano que ela acabou por frequentar um curso de astronomia. Felizmente para a ciência, ela gostou tanto de

astronomia que acabou por ali ficar depois de concluir a sua licenciatura, para tirar cursos de pós-graduação e para ser voluntária no observatório do Harvard College. As suas capacidades excecionais depressa chamaram a atenção do diretor do observatório, Edward C. Pickering, que a convidou a juntar-se às «Computadores de Harvard», uma equipa de prodígios matemáticos — todas mulheres — contratada para analisar dados provenientes de telescópios. Muito antes de um «computador» ser um dispositivo, significava uma pessoa que fazia cálculos[5].

A principal tarefa de Henrietta foi estimar e catalogar a luminosidade das estrelas para o gigantesco «mapeamento do céu» que estava a decorrer, o que implicava comparar o tamanho de pequenos pontos de luz em milhares de imagens de arquivo provenientes dos grandes telescópios do mundo. Era um trabalho repetitivo, meticuloso; os humanos nem sempre tiveram a sorte de ter algoritmos que automaticamente conseguem extrair padrões a partir de imagens.

Mas houve uma coisa que a manteve focada durante todo este trabalho árduo, e isso foi a busca de *estrelas pulsantes*. Uma estrela pulsante é aquela cujo brilho varia ao longo do tempo de uma forma notavelmente regular: ela oscila entre luminosidade intensa e fraca e intensa outra vez, repetidamente como um relógio (Ver Figura 2). Sabemos agora que estas estrelas pulsantes são milhares de vezes maiores do que o nosso Sol e que a sua luminosidade oscila porque as atmosferas estelares estão sucessivamente a inchar e a encolher, tal como os seus pulmões quando respira. No tempo de Henrietta, no entanto, sabia-se pouquíssimo acerca destas estrelas

curiosas. Os astrónomos estavam fascinados por elas, e Henrietta tinha instruções permanentes para estar atenta a todas as que pudesse encontrar.

Para tal, ela reuniu imagens de uma única estrela obtida em muitas noites diferentes. Debruçou-se sobre elas com uma lupa e uma régua pequenina para procurar o sinal indicativo de uma estrela pulsante, ou seja se o seu ponto de luz ficava repetidamente maior e menor ao longo do tempo. Ela fez isto imagem a imagem, estrela a estrela, durante anos a fio, e acabou por encontrar 1777 estrelas pulsantes anteriormente desconhecidas da ciência.

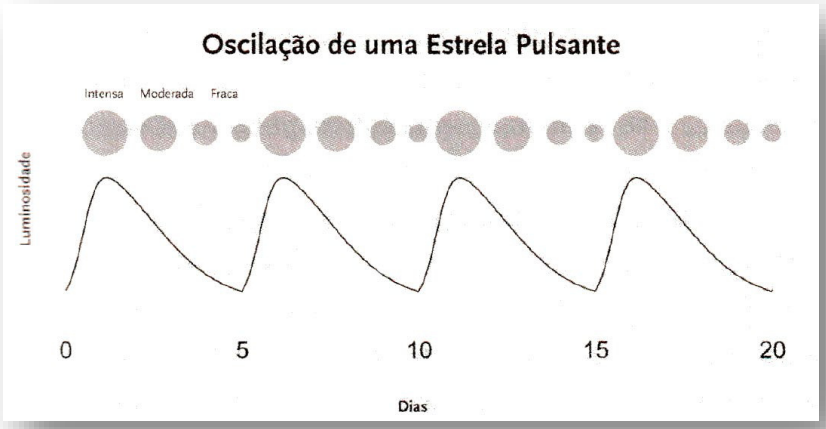


Figura 2. A oscilação da luminosidade de uma estrela pulsante. Esta estrela em particular completa um ciclo de intensa a fraca, e a intensa outra vez, a cada 5,4 dias. Henrietta Leavitt descobriu que o período de uma estrela pulsante estava relacionado com a sua luminosidade: estrelas pulsantes mais intensas oscilam mais lentamente do que as mais fracas, de uma maneira matematicamente previsível.

Em 1912, Henrietta Leavitt tinha-se focado num grupo de 25 estrelas pulsantes na Pequena Nuvem de Magalhães*. Uma vez que todas estas estrelas faziam parte do mesmo aglomerado, Henrietta sentiu-se segura ao assumir que todas estavam aproximadamente à mesma distância da Terra. Portanto, se uma estrela parecia mais

brilhante, efetivamente *era* mais brilhante na origem. Para cada estrela, ela tabulou dois pontos de dados. O primeiro era o seu período de pulsação, ou o tempo que uma estrela demorava a completar um ciclo completo de brilho intenso a fraco e a intenso outra vez. Cada estrela tinha um período específico, variando entre 1,25 dias e um máximo de 127 dias. O segundo era a luminosidade da estrela, ou quanta luz ela estava a emitir.

Em seguida, construiu um gráfico com os dados. Na Figura 3 fizemos a nossa própria versão desse gráfico usando os dados originais dela. Cada ponto é uma das 25 estrelas pulsantes de Henrietta. A coordenada horizontal (X) representa o período de pulsação da estrela, e a coordenada vertical (Y) representa a sua luminosidade. Gráficos como este são ótimos para revelar padrões em conjuntos de dados — e Henrietta revelou um padrão extraordinário referente à luminosidade e ao período de uma estrela pulsante. Os seus pontos de dados, como se verificou, assentavam de forma quase perfeita ao longo de uma linha reta. As estrelas menos brilhantes do seu conjunto de dados tinham períodos medidos em dias, enquanto as mais brilhantes apresentavam períodos medidos em meses. Quanto maior fosse o período, mais brilhante seria a estrela. O padrão era tão regular que podia descrever-se com uma equação: uma linha reta diretamente pelo meio dos pontos de dados.

Esta acabou por ser uma das linhas retas mais importantes da história da ciência. Para perceber porquê, imagine que está outra vez naquela estrada rural escura, onde vê uma luz ao longe, mas não sabe a que distância ela está. Agora imagine que alguém lhe

dá uma pista, ao dizer-lhe exatamente quão brilhante a luz é na sua origem.

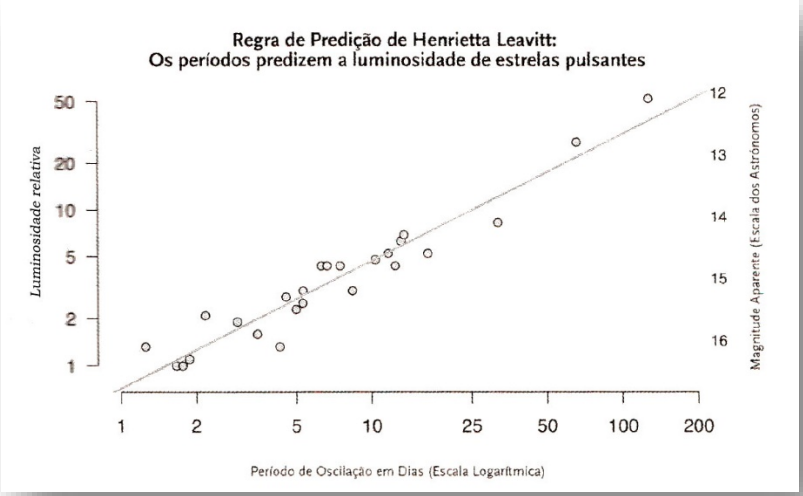


Figura 3. Dados de 1912 de Henrietta Leavitt sobre 25 estrelas pulsantes. Este padrão matemático — uma linha reta que relaciona o período de oscilação com a luminosidade — permitiu aos astrónomos medirem distâncias cósmicas em escalas anteriormente inimagináveis.

A partir deste indício pode determinar a distância à luz: se lhe disserem que a luz pertence a uma única lâmpada de 60 watts, sabe que a luz tem de estar perto, mas se lhe disserem que é a luz de uma cidade inteira, sabe que tem de estar longe. O princípio geral é que se se consegue medir a luminosidade *aparente* do objeto, e se alguém o informa da verdadeira luminosidade do objeto — a quantidade de luz que efetivamente está a emitir —, então pode-se trabalhar de trás para a frente, usando as leis da física, para determinar a que distância está o objeto. O processo de trabalhar de trás para a frente é matematicamente entediante mas ainda assim concetualmente simples. A pista que realmente desvenda o mistério da distância é o *conhecimento da verdadeira luminosidade de um objeto*.

A descoberta de Henrietta Leavitt proporcionou aos astrónomos exatamente esse tipo de pista. Eles podiam apontar os seus telescópios a uma estrela pulsante e medir tanto a sua luminosidade aparente como o seu período. Depois podiam pegar no período da estrela, inseri-lo na equação de Leavitt e obter a previsão correspondente para a sua *verdadeira* luminosidade*. Isto imediatamente gerava a distância da estrela — e, portanto, a distância a quaisquer outras estrelas na sua vizinhança. Na gíria da astronomia, Henrietta descobrira que as estrelas pulsantes eram «velas-padrão»: objetos de luminosidade conhecida, cuja distância podia ser fiavelmente medida. Na gíria da IA, ela tinha descoberto uma regra preditiva. A equação dela era mais uma vez apenas «*output* = função do input».

Henrietta Leavitt publicou os seus resultados em 1912, num artigo com apenas três páginas. Os seus colegas imediatamente reconheceram que a descoberta dela proporcionava a fita métrica cósmica que tinham procurado avidamente, e começaram a usá-la assim que os seus instrumentos o permitiram.

O primeiro resultado significativo veio de um astrónomo chamado Harlow Shapley. Ele mediu o período de várias estrelas pulsantes da Via Láctea, aplicou a regra preditiva de Leavitt para determinar as suas verdadeiras luminosidades e em seguida usou o resultado para calcular as distâncias delas. A localização destas estrelas acabou por se revelar surpreendentemente longínqua. Os resultados de Shapley implicavam que a nossa galáxia tinha pelo menos cem mil anos-luz de diâmetro, muito mais do que alguém

imaginara — e, reproduzindo as palavras de Copérnico, que o nosso próprio sol não estava sequer perto do centro galático[6].

Mas o verdadeiro triunfo chegou com outro astrónomo, que agora é um dos cientistas mais famosos de todos os tempos: Edwin Hubble.

Em 1919, Hubble começou a trabalhar no observatório do monte Wilson, em Pasadena, na Califórnia, tendo chegado mesmo a tempo da inauguração do novo telescópio Hooker de 2,54 metros de diâmetro. Com a regra preditiva de Leavitt em mente, ele começou a procurar estrelas pulsantes em nebulosas espirais — e como tinha o maior telescópio do mundo à sua disposição, dispunha de boas hipóteses de as encontrar. Cada estrela pulsante funcionaria como um farol, uma vela-padrão que Hubble podia usar para calcular a distância à sua espiral hospedeira.

Levou anos, mas a pesquisa lenta e cuidadosa de Hubble deu frutos. Em outubro de 1923, Hubble viveu finalmente o seu momento eureca: encontrou uma estrela pulsante em Andrómeda, a mesma «mancha enevoada» que chamara a atenção de Abd al-Rahman al-Sufi mais de mil anos antes e que tinha desconcertado todos os astrónomos desde então. Ele mediu a luminosidade aparente da estrela e calculou que o seu período era de 31,4 dias. Introduziu o valor na regra preditiva de Henrietta Leavitt para obter a luminosidade real. Em seguida procedeu de trás para a frente, usando tanto a luminosidade real como a aparente para calcular a distância a Andrómeda.

O resultado foi uma revelação: a Grande Nebulosa de Andrómeda estava a mais de um milhão de anos-luz da Terra — muito para lá da Via Láctea. Andrómeda era portanto *gigantesca*, dado que podíamos vê-la da Terra a uma distância tão vasta. Só podia ser uma coisa: uma galáxia em si mesma. De uma assentada, Hubble tinha resolvido uma questão acerca do nosso lugar no Cosmos que tinha estado em aberto durante mais de um milénio.

Mais tarde, Hubble usaria a técnica da estrela pulsante para descobrir muitas mais galáxias — ou, como ele disse, «mundos inteiros, cada um deles um grandioso Universo, espalhados por todo o céu [...] como os proverbiais grãos de areia na praia»[7]. No entanto, foi aquela primeira estrela pulsante que ele encontrou em Andrómeda — conhecida atualmente por «variável Hubble número um», ou V1 — que ficou para a história. Décadas mais tarde, em 1990, quando o vaivém espacial *Discovery* transportou o telescópio espacial Hubble até à órbita terrestre baixa, também transportou algo cujo valor era inteiramente sentimental: uma cópia da fotografia original da V1 obtida por Hubble em 1923.[8] Foi uma fotografia que transformou Hubble num nome conhecido e que para sempre alterou o rumo da astronomia. Mas também era uma fotografia cuja importância Hubble só pôde ver por ter estado sobre os ombros* de Henrietta Leavitt — pois foi ela que mostrou a Hubble, e a todos os outros, como medir o tamanho do Universo.

5. Ajustar as Regras Preditivas aos Dados

Regressaremos à história de Henrietta Leavitt mesmo no final do capítulo. Para já, contudo, vamos manter presente a grande

descoberta dela enquanto revisitamos as duas ideias-chave acerca de reconhecimento de padrões que mencionámos no início do capítulo.

1. Na IA, um «padrão» é uma regra preditiva que mapeia um *input* a um *output*.

2. «Aprender um padrão» significa ajustar uma boa regra preditiva a um conjunto de dados.

Esperamos que as estrelas pulsantes de Henrietta Leavitt lhe tenham ensinado o valor de uma boa regra preditiva para descrever um padrão. Mas também esperamos que ainda tenha algumas questões. Por exemplo: o que faz com que uma regra preditiva seja melhor do que outra? E como é que algo com uma mente tão literal como um computador aprende a regra preditiva certa?

Na IA, o critério para avaliar regras preditivas é simples: em média, qual a dimensão dos erros que a regra produz? Nenhuma regra preditiva pode ser perfeita, mapeando com exatidão cada *input* ao *output* certo, pois todas as regras cometem erros. Mas quanto mais pequeno for o erro médio melhor será a regra.

Para compreender isto vamos olhar uma segunda vez para a regra preditiva de Henrietta Leavitt para estrelas pulsantes. No painel do lado esquerdo da Figura 4 vê os dados de Leavitt, a par da sua equação original: a linha reta que relaciona a luminosidade de uma estrela pulsante com o seu período. A escala de

luminosidade exibida é a que os astrónomos usam, chamada «magnitude»; por razões históricas, os astrónomos registam os valores como os jogadores de golfe, com os números mais baixos a indicar estrelas mais luminosas.

Concentre-se na estrela que destacámos com uma seta, com um período de aproximadamente dois dias. Podemos medir o quanto a regra de Leavitt erra aqui calculando a distância vertical entre o ponto e a linha; a esta distância chama-se «erro residual» ou «de reconstrução». Para esta estrela, a regra de Leavitt previu que a magnitude estelar seria cerca de 16,1, enquanto na realidade era cerca de 15.6 — um erro de reconstrução de 0,5 unidades.

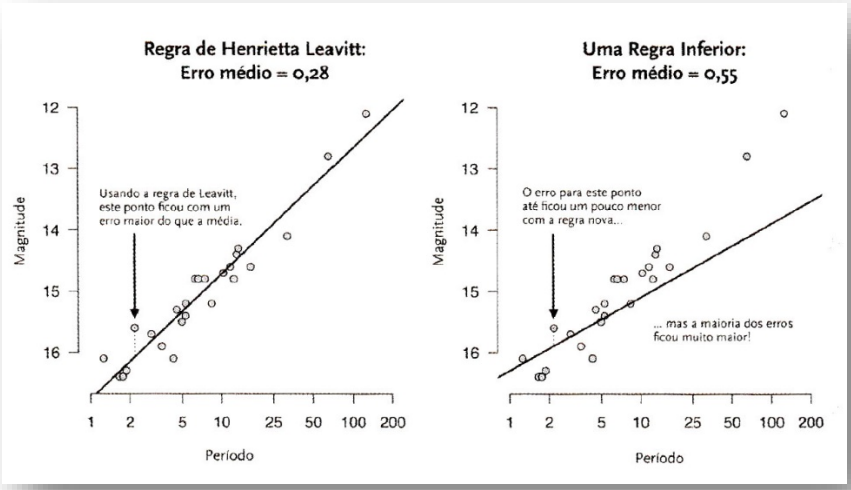


Figura 4. A equação original de Leavitt (esquerda) versus uma equação modificada (direita), que não se ajusta tão bem aos dados.

Agora considere uma regra ligeiramente diferente, como a que se encontra no painel do lado direito. Aqui ajustámos a linha de Leavitt tornando-a um pouco menos inclinada. Para aquela estrela que destacámos, o erro tornou-se efetivamente mais pequeno. Mas, para a maioria das outras estrelas, os erros tornaram-se maiores.

A regra de Leavitt é portanto melhor do que a nossa — em média, é quase duas vezes mais exata.

De facto, verifica-se que a regra preditiva de Leavitt é a melhor regra: entre todas as linhas retas, é a que tem o menor erro de reconstrução médio possível. Pode ajustar a linha de todas as maneiras que quiser, e alguns dos erros individuais poderão diminuir. Mas os seus ajustes, como o nosso no painel da direita, garantidamente farão aumentar o erro médio. Sabemos isto porque Leavitt seguiu uma receita matemática chamada «princípio dos mínimos quadrados» para ajustar uma regra preditiva aos dados dela. Publicado pela primeira vez em 1805 pelo matemático francês Adrien-Marie Legendre, o princípio proporcionou uma fórmula explícita para ajustar a linha reta «ideal» a um conjunto de dados — ou seja, a linha que produz o menor erro de reconstrução médio possível[*]. Os cientistas têm usado esta fórmula desde então[9]. Além disso, o mesmo princípio básico que Legendre enunciou há mais de dois séculos ainda é usado hoje, para construir alguns dos mais sofisticados sistemas de inteligência artificial. As regras preditivas na IA simplesmente são versões mais requintadas da regra preditiva descoberta por Henrietta Leavitt: são equações que mapeiam *inputs* a *outputs* e são escolhidas para minimizar o erro de reconstrução médio num conjunto de dados, tal como Legendre sugeriu há mais de dois séculos[†].

Antes de analisar esta ideia em pormenor, dar-lhe-emos três exemplos breves, que pode usar diretamente a partir do seu telefone. Primeiro, considere o *software* de reconhecimento de imagens, como o do tipo que identifica os seus amigos nas fotos

que carrega para o *Facebook*. O reconhecimento de imagem é apenas uma regra preditiva: o *input* é uma imagem do rosto de uma pessoa e o *output* é a identidade dessa pessoa. O mapeamento de *inputs* a *outputs* é uma equação complicada que descreve um padrão complexo nos dados de treino: quais as caraterísticas faciais que tendem a acompanhar que nomes em fotos carregadas anteriormente.

Segundo, considere o *Google Tradutor*. Isto também é apenas uma regra preditiva: mapeia frases de *input* numa língua (digamos, o inglês) a frases de *output* noutra língua (por exemplo o castelhano). O modelo subjacente é, mais uma vez, uma equação complexa que descreve um padrão complicado: que frases inglesas tendem a acompanhar frases castelhanas através de uma enorme base de dados de frases apresentadas lado a lado em ambas as línguas.

Finalmente, considere uma nova aplicação para smartphone desenvolvida pela Dra. Elina Berglund Scherwitzl, uma física sueca que ajudou a descobrir o bosão de Higgs — e que agora está bem a meio de uma segunda carreira como empresária, tendo inventado uma nova tecnologia contracetiva baseada na IA. Há muito que Berglund Scherwitzl procurava uma alternativa para a contraceção hormonal, porém nunca conseguira encontrar uma solução de que gostasse. Ela e o marido, Raoul Scherwitzl, despediram-se dos seus trabalhos como físicos e empenharam-se em usar as suas competências em ciência de dados para construírem uma nova variação sobre uma velha ideia: o «método do ritmo», que envolve

acompanharmos a nossa história menstrual para prevermos quando será mais provável a ovulação.

O problema com o método do ritmo tradicional é que, para utilizá-lo com sucesso, necessita de uma abordagem invulgarmente meticulosa ao registo de dados. A versão de Berglund Scherwitzl assenta em algo muito mais fiável: a temperatura corporal, cujo ciclo mensal se correlaciona fortemente com a fertilidade. Para usar o método introduzimos dois tipos de informação na aplicação para smartphone, *Natural Cycles*: a nossa temperatura corporal diária e a data da nossa menstruação. Com o passar do tempo, à medida que fornecemos mais dados de treino à aplicação, ela ajusta uma regra preditiva personalizada aos padrões no nosso próprio ciclo. O *input* é a nossa temperatura, enquanto o *output* é uma previsão acerca da nossa fertilidade nesse dia, sob a forma de um pequeno semáforo no smartphone (verde significa avançar.) As aplicações que nos ajudam a acompanhar o ciclo menstrual são muito populares, mas esta é a primeira a ter sido certificada por reguladores na União Europeia como um método contracetivo eficaz. Em ensaios clínicos, a aplicação mostrou ser praticamente tão eficaz quanto a pílula para prevenir a gravidez sob padrões de utilização típicos*. Em meados de 2017 estava a ajudar mais de 300 mil subscritores a assumirem o controlo das suas escolhas reprodutivas, usando a IA.

6. Para Lá das Linhas Retas

Nesta altura talvez esteja a interrogar-se acerca de uma coisa. Explicámos que na IA, reconhecer um padrão significa ajustar uma

equação a dados. E também explicámos que esta ideia remonta a 1805. Então, o que explica a recente revolução? Como é que todos estes sistemas de reconhecimento de padrões, do reconhecimento facial à tradução automática e aos anticoncecionais baseados em IA, só entraram em cena nos últimos anos?

Eis a questão fundamental: os padrões a serem encontrados em bases de dados enormes de imagens, texto e vídeo são radicalmente mais complicados do que qualquer padrão que se possa visualizar com um gráfico de dispersão, como o gráfico de Henrietta Leavitt das estrelas pulsantes. E estes padrões complicados têm de ser descritos através de equações complicadas — muito mais complicadas, pelo menos, do que a equação de uma linha reta. Equações assim tão complicadas são altamente exigentes. Como iremos explicar, é necessário muito poder computacional para trabalhar com elas, e são necessários muitos dados para se poder estimá-las de maneira fidedigna. A nossa tecnologia só recentemente tornou isto viável e barato.

O grande avanço na IA envolve a utilização de *redes neurais* para estimar regras preditivas a partir de dados. O termo «rede neural» soa terrivelmente a algo como cérebro, mas isso não é mais do que uma brilhante jogada de marketing. Na realidade, uma rede neural é apenas uma equação bastante complicada capaz de descrever padrões bastante complexos em dados — ou seja, mapeamentos muito complicados de *inputs* a *outputs*. A razão por que usamos redes neurais não é porque elas replicam o que os cérebros humanos fazem mas porque funcionam incrivelmente bem numa

gama surpreendente de tarefas preditivas, da linguagem às imagens e ao vídeo.

Vejamos com mais atenção os quatro fatores que impulsionam este avanço.

6.1 Fator 1: Modelos Maciços

No passado, costumávamos construir regras preditivas para descrever padrões simples usando modelos reduzidos — uma espécie de mineração de dados com pás e picaretas. Hoje, descrevemos padrões complicados usando modelos maciços — mais como um daqueles camiões gigantescos de mineração com pneus do tamanho de uma casa pequena. É a mesma ideia, só que com uma pá muito maior.

Para perceber o que queremos dizer com modelo «maciço» tem de perceber o conceito de parâmetro. Um parâmetro é um número na sua equação que pode escolher livremente com o objetivo de obter o melhor ajuste possível aos dados. Modelos pequenos têm poucos parâmetros, enquanto modelos maciços têm muitos parâmetros. Por exemplo: talvez se lembre desta equação usada atrás: Frequência Cardíaca Máxima = 220 - Idade. Este é um modelo pequeno, porque a equação tem apenas um parâmetro: o valor basal 220, ao qual subtrai a sua idade. Podíamos ter escolhido um valor basal de 210, 230 ou qualquer outro — mas a opção 220 ajusta-se melhor aos dados.

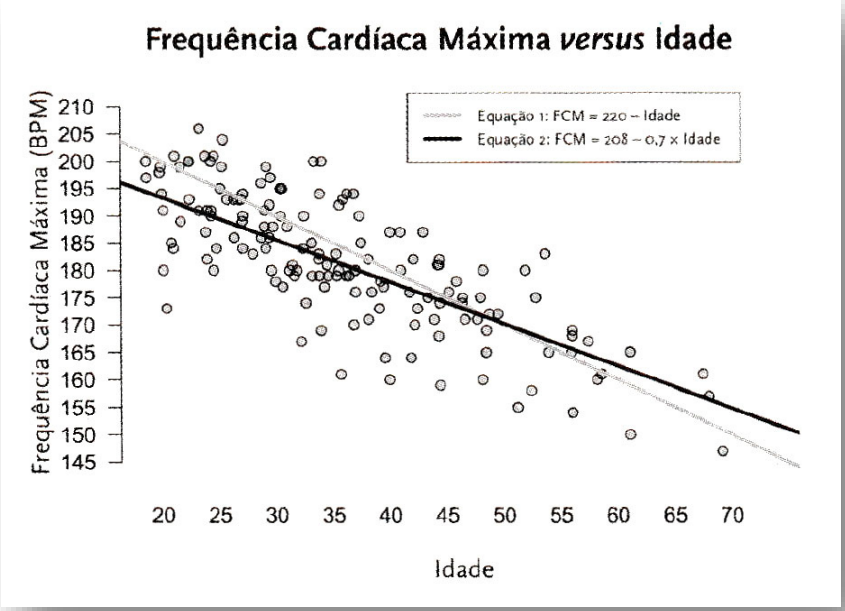


Figura 5. Duas equações para prever a frequência cardíaca máxima a partir da sua idade. A linha cinzenta tem um parâmetro; a linha preta tem dois parâmetros e consequentemente ajusta-se melhor aos dados.

Mas na verdade existe um modelo ligeiramente maior que funciona ainda melhor: Frequência Cardíaca Máxima = 208 - 0,7 x Idade. Por outras palavras, multiplique a sua idade por 0,7 e subtraia o resultado a 208 para prever a sua frequência cardíaca máxima. A regra anterior tinha apenas um parâmetro, enquanto esta regra nova tem dois: o valor basal 208 e o «multiplicador de idade» 0,7, os quais podem ser afinados para se ajustarem aos dados. Pode ver estas duas regras na Figura 5, que mostra um gráfico de dispersão de 151 adultos cuja frequência cardíaca máxima foi medida em laboratório. O gráfico mostra ambas as regras preditivas: a regra velha de «220 - Idade» a cinzento e a regra nova de «208 - 0,7 x Idade» a preto. Os cientistas que estudam o exercício preferem a linha preta; tendo dois parâmetros em vez de um dá-lhes uma flexibilidade suplementar para afinar a equação de modo a que se ajuste aos dados tão bem quanto possível*. Algumas pessoas ainda repetem a velha regra de

«subtraia a sua idade a 220» porque é mais simples — um parâmetro em vez de dois, mas paga-se um preço pela simplicidade. Essa previsão acerca da sua frequência cardíaca máxima não será tão exata quanto a feita pelo modelo com dois parâmetros, pelo menos em média.

Vejamos agora qual o aspeto de uma regra preditiva com *três* parâmetros. Imagine que é um cientista de dados na agência imobiliária online *Zillow* e que tem de construir uma regra para prever o preço de uma casa[†]. Talvez comece com duas caraterísticas óbvias das casas, como os metros quadrados e o número de casas de banho, a par de *multiplicadores* para cada caraterística. Por exemplo:

Preço = 10.000 + 125 x (metros quadrados)

+ 26.000 x (número de casas de banho)

Por outras palavras, isto enuncia que para prever o preço de uma casa deverá seguir três passos:

1. Multiplique os metros quadrados da casa por 125 (parâmetro 1).

2. Multiplique o número de casas de banho por 26.000 (parâmetro 2).

3. Adicione os resultados de 1 e 2 ao valor basal 10.000 (parâmetro 3). Isto produz o preço previsto.

Mas porquê ficarmos por apenas dois multiplicadores de caraterísticas? As casas têm muitas outras caraterísticas que têm impacto no seu preço — por exemplo, a distância ao centro da cidade, o tamanho do jardim, a idade do telhado e o número de lareiras. Usando o princípio dos mínimos quadrados que Legendre enunciou em 1805, os cientistas de dados podem facilmente ajustar uma equação que incorpore todas estas caraterísticas e centenas mais. É basicamente assim que a *Zillow* constrói a sua regra preditiva para o preço de uma casa. Cada caraterística recebe o seu próprio multiplicador e caraterísticas mais importantes acabam por receber multiplicadores maiores, porque os dados mostram que elas têm maior influência no preço. Obviamente, se tentar descrever tal regra preditiva por palavras — «some isto», «multiplique isto», como fizemos para a regra anterior de duas caraterísticas — ela começará a parecer os impressos do IRS que são engendrados no Inferno. Mas um computador avança simplesmente através dos cálculos sem qualquer problema, mesmo em modelos com centenas de parâmetros.

Na IA, contudo, sonhamos ainda mais alto do que isso, ajustando modelos com *muito* mais do que algumas centenas de parâmetros. Considere, por exemplo, um modelo para rotular imagens. Para uma máquina, uma imagem é apenas píxeis, e píxeis são apenas números: intensidades de cor de 0 a 100%. Uma imagem não comprimida de um megapíxel, por exemplo, tem três milhões de números associados: uma intensidade vermelha, verde e azul para cada um de um milhão de píxeis. Logo aí são três milhões de características. E precisamos de muitos parâmetros para sermos capazes de usar os três milhões de caraterísticas da melhor forma

— especialmente se vamos combinar essas caraterísticas de maneiras interessantes, em vez de simplesmente atribuir a cada uma um único multiplicador, tal como anteriormente fizemos no nosso modelo imaginário *Zillow*.*

É aqui que entram as redes neurais. Em 2014, por exemplo, engenheiros da Google publicaram um artigo sobre um modelo de rede neural — apelidado de *Inception*, por causa do filme de Leonardo DiCaprio [*A Origem*, em Portugal] — que podia reconhecer e rotular automaticamente uma imagem. E era assustadoramente eficaz. Os modelos antigos de reconhecimento de imagens podiam ter sido capazes de dizer se uma dada fotografia era de um cão ou de um não cão. O *Inception* conseguia dizer se um certo cão era um *husky*-siberiano ou um *malamute-do-alasca*. O modelo incluía 388.736 parâmetros, e usá-lo para efetuar uma previsão requeria 1,5 mil milhões de operações matemáticas — 1,5 milhões de pequenos passos de «some isto» ou «multiplique aquilo» — *para uma única imagem de input*. É um longo formulário de IRS. Ainda bem que uma placa gráfica da Nvidia em 2018 consegue fazer 1,5 mil milhões de cálculos em menos de 0,0001 segundos.

6.2 Fator 2: Dados Maciços

Mas há um senão: para ajustar um modelo maciço, é necessário um conjunto maciço de dados.

Um modelo como o *Inception* da Google, com 388.736 parâmetros, tende a dar a volta à cabeça de cientistas e engenheiros da velha guarda, que olham para esses modelos

maciços com desprezo. O grande matemático John von Neumann, por exemplo, certa vez criticou um modelo complicado com o seguinte gracejo enigmático: «Com quatro parâmetros consigo ajustar um elefante, e com cinco consigo fazer com que mexa a tromba.» Von Neumann quis dizer que um modelo com muitos parâmetros corre o risco de «sobreajustar», o que acontece quando um modelo apenas memoriza o ruído aleatório nos dados de treino em vez de aprender o padrão subjacente. Um modelo sobreajustado pode descrever o passado com uma precisão perfeita, mas ainda assim ser mau a prever o futuro.

Se quer entender o sobreajustamento, basta olhar para todos aqueles analistas políticos que vê na televisão e que são pagos para criar algumas pérolas de «sabedoria» acerca das eleições presidenciais — por exemplo, «Jamais um democrata em exercício de funções, sem experiência de combate, derrotou alguém cujo primeiro nome valha mais no *Scrabble*». Na verdade, esta regra preditiva *de facto* manteve-se durante 208 anos na história americana, até Bill Clinton derrotar Bob Dole em 1996*. Mas nunca teve qualquer valor para prever o futuro. A regra é um exemplo clássico de sobreajustamento: retrospetivamente, vasculhar em milhares de detalhes complicados sobre eleições passadas e escolher a dedo o único detalhe complicado que se revelou verdadeiro.

Então, se estiver a ajustar uma regra preditiva aos dados, como é que evita o sobreajustamento? Existem apenas duas formas. A primeira é que pode recusar explicações complicadas. O seu modelo não pode memorizar factos complicados e não generalizáveis acerca

do passado se o obriga a ignorar todos os factos exceto os mais simples. Esta solução funciona bem nas ciências puras. Na verdade, era exatamente o que Von Neumann estava a defender com a sua perspetiva de «ajustar um elefante»; ele estava interessado em encontrar teorias simples que conseguissem explicar leis físicas universais da matéria e da energia, em vez de particularidades mundanas pontuais como elefantes ou trombas a mexer. Mas a abordagem de «ignorar teorias complicadas» não funciona de todo na IA. Os padrões que queremos encontrar nos dados — por exemplo, que combinações de píxeis são rotuladas como «husky» e quais são classificados como «malamute» — são na verdade complicados, mundanos e particulares. Modelos pequenos com dois ou três parâmetros, ou mesmo dois ou três mil, simplesmente não conseguem explicar com precisão estes padrões*.

Assim, isto força-nos a recorrer à segunda estratégia: recolher quantidades maciças de dados. Uma enorme quantidade de dados significa imensa experiência — e com experiência suficiente podemos descartar as explicações complicadas e más, ficando apenas com as complicadas e boas. Esta solução não funciona nas eleições presidenciais; até ao momento só se realizaram 56, de modo que não há maneira de se saber, apenas a partir dos dados, se uma explicação complicada *post hoc* de quem vence a presidência tem algum valor na previsão do futuro. Mas funciona magnificamente em modelos que extraem padrões de imagens, textos e vídeos, que é algo que temos em abundância.

John von Neumann certamente ficaria fascinado com o resultado. Ele pensava que se podia «ajustar um elefante» com apenas quatro

parâmetros, mas na verdade são precisos 388.736 — ou pelo menos precisamos de um número tão elevado de parâmetros para *identificar* um elefante nas fotos que tirámos no safari que fizemos em África. Aqui não há qualquer magia, apenas conjuntos de dados maciços com milhões ou milhares de milhões de pontos de dados. Isto permite-nos usar modelos complicados para descrever padrões complexos, sem sobreajustar. E, sendo justos para com Von Neumann, ele certamente nunca imaginou um mundo em que as pessoas carregariam diariamente 100 milhões de imagens para o *Instagram*, muitas delas com rótulos úteis como #safari ou #elefante.

6.3 Fator 3: Tentativa e Erro, um Milhão de Vezes por Segundo

No início da década de 1900, Henrietta Leavitt ajustou a sua regra preditiva usando papel e caneta, ao aplicar a fórmula matemática de Legendre, de 1805, para linhas retas otimizadas. Mesmo tão recentemente como o início da década de 2000, a maioria dos cientistas ainda ajustava as regras preditivas usando a mesma fórmula com variações mínimas. A única diferença é que o pessoal moderno acomodara-se, deixando que uma máquina fizesse os cálculos.

Mas não existe qualquer fórmula matemática para ajustar as regras preditivas atuais. Na verdade, existe apenas uma boa maneira para ajustar um modelo maciço como o *Inception* da Google, que é fazê-lo incrementalmente, através de tentativa e erro. Começamos com uma qualquer conjetura inicial para uma

regra preditiva — por exemplo, que todas as imagens com formas cinzentas sobre fundos verdes são elefantes a relaxar na savana. Esta regra inicial por certo será péssima. À medida que os dados começarem a chegar, contudo, refinamos a nossa regra. Por cada novo ponto de dados colocamos duas questões: qual a dimensão do erro que o meu modelo atual produz neste ponto de dados, e como posso afinar o modelo para obter um erro mais pequeno? Os computadores modernos podem colocar estas duas questões e responder milhares ou mesmo milhões de vezes a cada segundo. Quando um conjunto de dados maciço é sujeito a este tipo de investida computacional implacável, não demora muito até que a nossa regra preditiva melhore drasticamente — por exemplo, ao aprender que algumas formas cinzentas num fundo verde são elefantes, enquanto outras são rinocerontes.

Atualmente, esta estratégia de tentativa e erro para ajustar modelos é utilizada em todo o lado. É o que permite que os grandes retalhistas, por exemplo, vaticinem o que nós queremos comprar online, antes mesmo de sabermos o que queremos comprar. Considere o *Alibaba*, o gigante chinês do comércio eletrónico que contabilizou 24 mil milhões de dólares de receita em 2017. Tal como a *Amazon*, o *Alibaba* promete entregar as suas coisas rapidamente — tão rápido que é impossível enviá-las todas de um armazém central. Em vez disso, os cientistas de dados na *Cainiao*, o braço logístico do *Alibaba*, têm de ser mestres da IA, prevendo exatamente o que os clientes irão querer durante os próximos dias e semanas, de modo a que a empresa possa enviar os itens certos para os centros locais de distribuição específicos antes mesmo de alguém clicar em «comprar». Mais: eles têm de fazer isto para todos

os produtos que o *Alibaba* vende e para todos os mercados que abastece, quer seja um bairro de Xangai ou um distrito do Cantão. E eles fazem-no por tentativa e erro: usando conjuntos maciços de dados para treinar modelos maciços que se tornam um pouquinho melhores com cada nova compra.

No negócio da IA, este processo de refinar um modelo através da tentativa e erro inteligente tem muitos nomes diferentes, como «aprendizagem online» e «gradiente descendente estocástico». Estamos aqui a omitir muitos detalhes que são essenciais para esta estratégia ser bem-sucedida, mas são todos apenas minudências, o tipo de coisa que se aprende se estudar IA na pós-graduação. Se apenas pensar em «tentativa e erro», já terá feito 90 por cento do caminho.

6.4 Fator 4: Aprendizagem Profunda

Além da riqueza dos nossos modelos, da dimensão dos conjuntos de dados e da velocidade dos computadores, há uma quarta maneira significativa em que as regras preditivas melhoraram profundamente: as pessoas aprenderam a forma de extrair informação útil a partir de *inputs* bastante mais complicados. Se já ouviu o termo «aprendizagem profunda» e interrogou-se acerca do que significa, estamos prestes a explicar.

No início do capítulo dissemos que os computadores são agnósticos em relação ao tipo de *inputs* que recebem. Mas isso é apenas mais ou menos verdade. Ficou célebre a frase de Henry Ford de que os clientes da Ford Motor Company podiam comprar um

carro com a cor que quisessem, desde que fosse preto. O mesmo acontece com os computadores: podemos dar-lhes um *input* sob a forma que quisermos, desde que seja um número. A parte mais difícil na maioria das aplicações da IA não é treinar o modelo mas responder à questão que surge primeiro: Como é que represento o *input* para o meu modelo como um conjunto de números? Os cientistas de dados chamam a isto «engenharia de recursos», que apenas significa extrair caraterísticas numéricas de um determinado *input* que claramente não é um número, como uma imagem ou uma sequência de palavras.

Ao longo da última década, os especialistas em IA tornaram-se exponencialmente melhores a extrair caraterísticas numéricas automaticamente, usando um tipo específico de regra preditiva chamada «rede neural profunda». Já aprendeu anteriormente que uma rede neural é apenas uma equação complicada com muitos parâmetros. Uma rede neural profunda é uma variação sobre esta ideia, na qual a equação é estruturada de maneira a extrair tanta informação quanto possível de um tipo específico de *input*.

Considere as imagens. Aqui, as redes neurais profundas resolvem um desafio concetual basilar na engenharia de recursos: muitos dos diferentes arranjos possíveis dos píxeis de uma imagem podem em última análise significar a mesma coisa: rotações, translações, alterações de cor — todas estas coisas podem mudar drasticamente os píxeis numa imagem, sem alterar o conteúdo. O símbolo de um coração vermelho, por exemplo, significa o mesmo quer o coloquemos no lado esquerdo da imagem ou no direito, ou quer o rode alguns graus num ou noutro sentido. Até significa a

mesma coisa se alterarmos a cor — como na famosa música *country* da década de 1990 da autoria de Joe Diffie quando um jovem trabalhador rural chamado Billy Bob trepa ao reservatório de água local para pintar um coração com três metros de altura, a par de uma mensagem de amor para a sua amada, Charlene, usando a única cor de tinta de que dispunha: verde John Deere.[10] Tal como a Charlene compreende, um coração é um coração, quer esteja pintado de vermelho ou de uma outra cor mais agrícola. Mas um computador que tenha sido programado para interpretar píxeis de forma demasiado literal pode facilmente ficar baralhado. É por isso que precisamos da engenharia de recursos: para transformar píxeis não editados em factos úteis e generalizáveis acerca de uma imagem.

As redes neurais profundas resolvem este problema de uma maneira muito inteligente. Para explicar isto vamos regressar a Henrietta Leavitt, à sua busca das estrelas pulsantes e à sua subsequente descoberta de que elas podiam ser usadas para medir distâncias a cantos remotos do Universo. Vamos refrescar a sua memória acerca do que Henrietta teve de fazer nesta situação. Ela precisou de seguir uma única estrela em várias fotografias. Teve de medir a luminosidade da estrela em cada fotografia, para verificar se aumentava e diminuía da maneira indicativa de uma estrela pulsante. Se o fizesse, tinha de calcular o período da estrela, ou quanto tempo demorava a completar uma pulsação.

Ao fazer tudo isto, Henrietta teve de recorrer a pelo menos cinco conceitos visuais, cada um a um nível de abstração sucessivamente mais elevado.

- Nível 1: As partes brilhantes da foto representam luz vinda do céu.

- Nível 2: Uma estrela é um ponto de luz rodeado por escuridão.

- Nível 3: A luminosidade de uma estrela é o tamanho e a intensidade do seu ponto.

- Nível 4: Uma estrela pulsante é uma estrela cuja luminosidade varia de forma regular em diversas fotografias.

- Nível 5: O período de uma estrela pulsante é o intervalo de tempo que ela demora a passar de luminosidade intensa a fraca e a intensa outra vez.

A engenharia de recursos é parecida com isto. Se seguirmos a hierarquia do nível 1 ao nível 5 lá aparece um número: o período de uma estrela pulsante, que subsequentemente podemos usar como *input* na nossa regra preditiva. (Lembre-se, a regra preditiva de Leavitt para estrelas pulsantes tinha o período como *input* e a luminosidade real como *output*.)

Podemos dizer que Henrietta estava aqui a usar uma «rede neural profunda de cinco camadas». Ela estava a aplicar uma série de conceitos visuais, encadeados numa hierarquia com cinco camadas de profundidade, para extrair uma caraterística útil de

uma imagem. E é exatamente isso que uma rede neural profunda faz*. Cada nova camada da hierarquia tira partido dos conceitos dos níveis mais baixos — como aqui, onde o conceito de «estrela pulsante» do nível 4 é definido em termos de um conceito do nível 2 (estrela) e de um conceito do nível 3 (luminosidade). No nível mais alto da hierarquia acabamos com uma caraterística — período — que pode ser usada como *input* numa regra preditiva.

Leavitt sabia aplicar esta hierarquia de conceitos visuais devido à sua formação em astronomia. Mas, ao longo da última década, os especialistas em IA descobriram que podem ensinar os computadores a extrair essas hierarquias conceituais diretamente a partir das imagens brutas, e que esta abordagem é na realidade bastante mais eficaz do que programar esses mesmos computadores com os conhecimentos específicos desses domínios, quer se trate de astronomia ou de pepinos.

Toda esta abordagem é conhecida por «aprendizagem profunda», e até há pouco tempo era uma curiosidade académica. Mas já não é; em certas tarefas de identificação de imagens, as redes neurais profundas já suplantam as pessoas. Os especialistas em IA utilizam um conjunto de dados chamado Concurso de Reconhecimento Visual ImageNet para aferir os seus modelos. O ImageNet é uma base de dados online com milhões de fotos distribuídas por milhares de categorias, como «veleiro» e «malamute-do-alasca», e o objetivo é treinar um modelo para identificar automaticamente as imagens. Nesta tarefa os humanos têm uma taxa de erro média de cerca de cinco por cento, enquanto em 2011 os melhores modelos de IA tinham uma taxa de erro de

25 por cento. Em 2014, contudo, o modelo *Inception* da Google tinha reduzido o recorde mundial da taxa de erro de uma máquina para 6,7 por cento. O *Inception* era uma rede neural profunda de 22 camadas, com cada uma delas a representar um nível superior de abstração visual, de conceitos como «círculo» e «borda» na base até conceitos como «veleiro» e «malamute» no topo — todos aprendidos de forma orgânica a partir dos dados. E, em 2016, modelos subsequentes tinham alcançado uma taxa de erro inferior a três por cento, melhor do que o humano médio.

6.5 Os Benefícios Potenciais...

A aprendizagem profunda representa nada menos que uma revolução nas competências visuais das máquinas — e as ideias centrais e a tecnologia estão a disseminar-se por todo o lado. Aqui, no passado os limites foram a disponibilidade de dados, a velocidade dos computadores e a riqueza dos nossos modelos. Hoje parece existir apenas um limite, que é a imaginação das pessoas:

- Um apicultor sueco, Björn Lagerman, está a tentar salvar as abelhas melíferas com um modelo de aprendizagem profunda, treinado com 40 mil imagens de colónias de abelhas, que pode alertar os apicultores para a presença do ácaro da varroa, o inimigo mais perigoso da abelha-europeia.[11]

- Mark Johnson e a sua *startup*, Descartes Labs, têm treinado redes neurais profundas em quatro *petabytes* de imagens de satélite, em conjunto com relatórios de

colheitas do Departamento de Agricultura dos Estados Unidos, para prever a produção de milho. Estas previsões são cruciais para os incontáveis comerciantes ao longo da cadeia de abastecimento agrícola, dos proprietários de silos para grão aos produtores de etanol. Desde 2014, a Descartes Labs tem consistentemente suplantado as predições do Departamento de Agricultura.[12]

- Os fornecedores de eletricidade estão atualmente a treinar modelos que usam dados sobre água para preverem a procura por eletricidade ao nível da rede. Em Inglaterra, estão a ser conjugados com dados de imagens de satélite para prever a produção de energia a partir de fontes renováveis, como o sol e o vento. A National Grid estima que este tipo de modelo de aprendizagem profunda pode, em última análise, fazer baixar em 10 por cento a conta inglesa da luz, apenas por ajustar a oferta e a procura de forma mais eficiente.[13]

Depois há o estudo recente publicado pelo Geena Davis Institute on Gender in Media. Investigadores do instituto começaram a recolher dados em 2007 acerca de como os homens e as mulheres são retratados de maneira diferente nos filmes. De início fizeram a análise dos dados manualmente, vendo milhares de horas de filmes e procurando padrões cena a cena. Mas há pouco tempo estabeleceram uma parceria com a Google para automatizar esta tarefa, usando redes neurais profundas para classificação de imagens. Eles usaram uma versão atualizada do modelo *Inception* para analisar os cem filmes de Hollywood mais rentáveis ao longo

de vários anos. Este modelo classificou automaticamente o género de cada pessoa no ecrã, assim como quem estava a falar a cada momento.

Os resultados foram surpreendentes. Há apenas um tipo de filme em que as mulheres têm mais tempo de ecrã do que os homens: filmes de terror, onde habitualmente elas são as vítimas. Em todos os outros géneros, as mulheres estão sub-representadas. Em média, elas recebem 36 por cento do tempo de ecrã e 35 por cento do tempo de diálogo — e apenas 27 por cento do tempo de diálogo em filmes nomeados para um Óscar da Academia. Estes resultados mostram como a IA pode ajudar a enriquecer o debate acerca dos estereótipos de género e dos preconceitos inconscientes.[14]

6.6 ...e a Ameaça à Privacidade

Embora tenhamos salientado os muitos benefícios potenciais destes novos algoritmos de reconhecimento de padrões, é igualmente importante reconhecer as preocupações acerca da privacidade que eles suscitam. A mesma tecnologia de aprendizagem profunda que é usada para identificar preconceitos de género em filmes de Hollywood também pode ser usada pela polícia, por exemplo, para monitorizar gravações de câmaras de vigilância em locais públicos. Com câmaras e dados de treino suficientes, não há qualquer razão tecnológica para que um sistema de IA não possa ser programado para seguir uma única pessoa, passo a passo, por toda uma grande cidade. De facto, a polícia sempre procurou vigiar as pessoas suspeitas de crime, quer abrindo com vapor as suas cartas, colocando escutas nos telefones ou ao

monitorizar os metadados dos telemóveis. O que difere na Era da IA é o potencial teórico para monitorizarem *toda a gente* ao mesmo tempo — algo que, constrangimentos leais à parte, é logisticamente impossível recorrendo apenas à inteligência humana. E o potencial para abuso não se limita à polícia. Uma empresa privada com acesso a todas as imagens poderia usá-las para construir uma base de dados extremamente detalhada com as coisas para que olhamos e durante quanto tempo, ou um funcionário público poderia usá-las para retrospetivamente procurar pormenores embaraçosos que poderiam ser usados para ameaçar ou intimidar alguém — como um jornalista ou um opositor político, por exemplo.

Se lê bastante acerca da IA irá encontrar duas narrativas que se sobrepõem acerca da questão da vigilância. A versão menos radical é mais ou menos assim: a tecnologia de IA de reconhecimento facial é extremamente poderosa e precisa de ser regulada com cuidado, da mesma maneira que regulamos outras tecnologias de grande potencial para abuso. Nós apoiamos inteiramente esta visão. Há uma enorme discrepância entre a nossa tecnologia de IA e as nossas leis, e a sociedade deveria ter lidado com este problema ontem, não fazê-lo amanhã. Precisamos de regras inteligentes, elaboradas por pessoas que entendam aquilo que estão a legislar — tanto os benefícios como as ameaças potenciais.

A narrativa mais radical, por outro lado, defende que há algo único e inevitavelmente autocrático acerca das capacidades de vigilância da IA moderna. Claramente não somos especialistas na sociologia da tecnologia, mas ainda não vimos provas válidas que sustentem esta afirmação. A Gestapo certamente não precisou da

IA para aperfeiçoar a arte de espiar — e, já agora, também o FBI na década de 1950, quando esquadrinhou as organizações dos direitos civis, não precisou. Além disso, se olharmos para o mundo atual, não há uma correlação óbvia entre o compromisso de um país com a tecnologia digital e o seu respeito pela privacidade e direitos humanos fundamentais. Na Escandinávia, por exemplo, podemos encontrar as leis de privacidade digital mais rigorosas do mundo *e* paralelamente as economias digitais mais avançadas. (Boa sorte para pagar um café com dinheiro em Estocolmo.) Estes factos complicam qualquer esforço para estabelecer uma associação simplista entre tecnologia de IA e tirania.

Portanto, tragam os advogados e os decisores políticos. Os receios acerca da privacidade têm fundamento, mas estamos confiantes em que eles também são solúveis.

7. Pós-Escrito

Vamos terminar com uma última história sobre estereótipos — uma particularmente relevante num mundo em que apenas 17 por cento dos finalistas em ciências informáticas em universidades americanas são mulheres, uma fração que está a diminuir há décadas.

Lembrar-se-á de que Edwin Hubble usou a regra preditiva de Henrietta Leavitt para estrelas pulsantes — as velas-padrão do Universo — para provar categoricamente que a Via Láctea não era a única galáxia lá fora. Ao fazê-lo, resolveu uma questão que os

astrónomos debateram durante séculos. Quando anunciou a sua descoberta ao mundo, Hubble tornou-se instantaneamente uma celebridade. Cientistas e jornalistas disputavam a sua atenção. Ele veio a receber medalhas e prémios, a caminhar entre estrelas de cinema e chefes de Estado, com Einstein a telefonar-lhe para casa e a ter um enorme telescópio que orbita a Terra batizado em sua honra.

Nenhum destes louvores foi para Henrietta Leavitt. Ela morreu de cancro em 1921, quatro anos antes de Hubble anunciar a sua descoberta. Astrónomos profissionais, todos homens, certamente sabiam da sua equação inovadora, que lhes mostrou como usar as estrelas pulsantes para medir o tamanho do Universo. Enquanto grupo, no entanto, deram-lhe bastante menos crédito do que o que ela merecia. Para demasiados deles, Henrietta era *apenas* um computador: uma mulher impedida de entrar num observatório, e que precisou de um patrocinador masculino para que confiassem nela para publicar o seu artigo numa revista científica conceituada. Para o público, ela simplesmente era desconhecida — tal como permanece, mais ou menos, atualmente. Podemos prever com confiança que o centésimo aniversário da descoberta de Hubble, quando acontecer em 2025, fará as manchetes dos principais jornais do mundo. Mas o centésimo aniversário da descoberta de Henrietta, em 2012, nem sequer fez as manchetes das principais revistas de astronomia mundial.

Nós devemos-lhe muito mais do que isto. Porque se as estrelas pulsantes são as velas do Universo, então Henrietta Leavitt foi quem

lhes concebeu o castiçal, dando-nos uma equação que podíamos apontar aos céus para levar a luz à escuridão.

[1] Ihsan Hafez, «Abd al-Rahman al-Sufi and His Book of the Fixed Stars: A Journey of Re-discovery», dissertação de doutoramento, James Cook University, 2010.

[2] Marcia Bartusiak, *The Day We Found the Universe* (Nova Iorque: Vintage Books, 2010), 52.

[3] Agnes Clerke em *The System of the Stars*, 1890, em *ibid*., 53.

[4] K. C. Freeman, «Slipher and the Nature of the Nebulae», *Astronomical Society of the Pacific Conference Series*, vol. 471 (2013), http://arxiv.org//abs/1301.7509.

[5] As nossas duas referências principais sobre Henrietta Leavitt foram Nina Byers e Gary Williams, *Out of the Shadows: Contributions of Twentieth-Century Women to Physics* (Cambridge: Cambridge University Press, 2006), e Marcia Bartusiak, *The Day We Found the Universe* (Nova Iorque: Vintage Books, 2010).

[*] Mais tarde descobriu-se que estas pertencem a uma classe diferente de estrelas pulsantes chamadas «variáveis Cefeidas».

[*] Isto é uma simplificação ligeiramente excessiva. A linha de Leavitt permitia-nos determinar a luminosidade verdadeira de uma estrela pulsante, relativamente à verdadeira luminosidade de qualquer outra estrela pulsante. Assim, a afirmação tecnicamente correta é: uma vez conhecida a verdadeira luminosidade de uma estrela pulsante — só uma —, então, a partir do padrão de Leavitt, de imediato podemos determinar a verdadeira luminosidade de qualquer outra estrela pulsante no universo. Na gíria da astronomia, o padrão de Leavitt ainda tinha de ser «calibrado», sendo para isso necessário estimar a verdadeira luminosidade de uma única estrela pulsante por um qualquer outro meio. Os astrónomos demoraram alguns anos até descobrirem como fazer isto, razão por que o padrão de Leavitt não pôde de imediato ser usado para medir distâncias estelares. Mas deixaremos essa parte da história para outros; veja, por exemplo, o capítulo 8 de *The Day We Found the Universe* (Nova Iorque, Vintage Books, 2010), de Marcia Bartusiak.

[6] Na verdade, Shapley estimou que a Via Láctea tinha uma extensão de 300 mil anos-luz. Aperfeiçoamentos posteriores produziram uma estimativa de 100 mil anos-luz de extensão.

[7] Citado em Bartusiak, *The Day We Found the Universe*, 218.

[8] Clara Moskowitz, «Star That Changed the Universe Shines in Hubble Photo», Space.com, 23 de maio de 2011, https://www.space.com/11761-historic-star-variable-hubble-telescope-photo-aas218.html.

[*] Referência à frase «Se vi mais longe, foi por estar sobre os ombros de gigantes», atribuída a Isaac Newton, significando que quando alguém faz um avanço científico foi porque outros foram fazendo descobertas ao longo do tempo que permitiram a acumulação de conhecimento. [N. T.]

[*] Nota técnica: se alguma vez frequentou uma disciplina de cálculo, talvez se lembre de que aprendeu a minimizar funções. Legendre fez exatamente este tipo de coisa de modo a encontrar a regra preditiva que minimizasse o erro médio. Na verdade, a solução de Legendre minimiza o erro *quadrado* médio (daí «mínimos quadrados»). Isto é um aspeto técnico importante, mas não é de modo algum importante para entender a ideia básica.

[9] Há uma disputa de prioridade acerca dos mínimos quadrados. Gauss, o grande matemático alemão, pode tê-los inventado primeiro, embora Legendre tivesse a primeira demonstração publicada do método. Os que estiverem interessados na história deverão consultar Stephen Stigler, «Gauss and the Invention of Least Squares», *Annals of Statistics* 9, n.º 3 (1981): 465-74.

[†] Isto é uma simplificação ligeiramente excessiva. Também temos de nos preocupar com algo chamado «sobreajuste», que iremos discutir algumas páginas mais à frente.

[*] A pílula é muito mais eficaz sob «utilização perfeita», ou utilização pretendida pelo fabricante, com uma taxa de insucesso de apenas 0,3% ao longo de um único ano.

[*] A regra «FCM = 208 - 0,7 x Idade» é o ajuste dos mínimos quadrados aos dados usando a fórmula de Legendre de 1805.

[†] No Reino Unido, a *Zillow* é conhecida por *Zoopla*.

[*] Em linguagem de IA, este modelo seria «não linear», ao contrário do nosso modelo imaginário da *Zillow* com um único multiplicador para cada caraterística.

[*] Este exemplo é extraído das brilhantes tiras de cartoon *XKCD*: https://xkcd.com/1122/.

[*] Nota técnica: os criadores de modelos na IA na verdade tentam construir modelos mais simples, usando uma técnica matemática chamada «regularização». Isto também ajuda a evitar o sobreajustamento, e é verdadeiramente importante para se poder ajustar boas regras preditivas. Se está interessado nesta matéria, encorajamo-lo a ler mais acerca de sobreajustamento.

[10] «John Deere Green», escrita por Dennis Linde, interpretada por Joe Diffie.

[*] Isto aplica-se apenas a imagens. Existem arquiteturas de redes neurais profundas para todos os tipos de *inputs*, dos vídeos ao texto, e todos são estruturados de forma diferente.

[11] John Mannes, «This Beekeeper Is Rescuing Honeybees with Deep Learning and an iPhone», *TechCrunch*, 2 de maio de 2017, https://techcrunch.com/2017/05/02/beekeepers/.

[12] Alex Brokaw, «This Startup Uses Machine Learning and Satellite Imagery to Predict Crop Yields», *The Verge*, 4 de agosto de 2016, https://www.theverge.com/2016/8/4/12369494/descartes-artificial-intelligence-crop-predictions-usda.

[13] Sam Shead, «Google's DeepMind Wants to Cut 10% Off the Entire UK's Energy Bill», *Business Insider*, 13 de março de 2017, http://www.businessinsider.com/google-deepmind-wants-to-cut-ten-percent-off-entire-uk-energy-bill-using-artificial-intelligence-2017-3.

[14] «The Women Missing from the Silver Screen and the Technology Used to Find Them», Google.com, https://www.google.com/intl/en/about/main/gender-equality-films/.

CAPÍTULO 3

O Reverendo e o Submarino

P: O que é que têm em comum uma bicicleta, a neve, um canguru e um submarino?
R: Todos são importantes para construir um carro autónomo.

Vejamos as bicicletas. As bicicletas são um problema. Os sensores num carro autónomo são bastante bons a identificar coisas como peões e esquilos, que se movem muito mais devagar do que os carros e que têm basicamente o mesmo aspeto vistos de qualquer ângulo. Outros carros é muito fácil: são bolhas gigantes refletivas de metal que num radar brilham como uma árvore de Natal. Mas... as bicicletas? As bicicletas podem ser rápidas ou lentas, grandes ou pequenas, de metal ou de fibra de carbono — e, dependendo do nosso ângulo de visão, podem ser tão largas como um carro ou tão estreitas como um livro. Se não repararmos na bicicleta propriamente dita, como é que podemos saber que um ciclista não é apenas um peão com uma postura excêntrica? Já para não falar de todas aquelas guinadas, tão erráticas e impulsivas. São o suficiente para dar uma forte dor de cabeça a um robot.

A neve também é um problema, e não por causa da tração — os robots são suficientemente espertos para usar pneus de inverno e

conhecem os seus próprios limites. Mas a neve cobre as marcações da faixa de rodagem. A neve oculta os sinais de stop. A neve interfere com os raios laser que o carro usa para medir a distância a objetos próximos. Para um carro-robot, a neve significa privação sensorial.

Quanto aos cangurus, embora as outras criaturas possam ser imprevisíveis, pelo menos permanecem no chão. Mas um canguru salta nove metros de cada vez. À medida que salta para cima e para baixo, aparenta ficar maior e mais pequeno, maior e mais pequeno, no campo de visão da câmara, como um coelho gigantesco num caleidoscópio. Isto é deveras confuso para um robot. Com todas estas alterações repentinas do tamanho aparente, como é que podemos saber a que distância se encontra? É quase como se precisássemos de um laser específico para detetar a variedade de cangurus — talvez bastantes lasers, dado que os cangurus viajam em... bem, há uma razão para os zoólogos lhes chamarem «turbas».

Depois temos um submarino. Prometemos tratar disso em seguida.

Mas primeiro convidamo-lo a refletir em algo. Não é fantástico que estejamos a falar de *turbas de cangurus* como um dos grandes problemas tecnológicos, ao invés de, digamos, sair da garagem ou não chocar contra a sala de estar do seu vizinho? Coloque a si mesmo uma questão simples. Se tivesse de colocar uma pessoa querida num táxi, preferia que ela fosse conduzida por um adolescente de 16 anos com carta de condução, escolhido

aleatoriamente, ou por um dos carros da Waymo? (Waymo é a empresa de carros autónomos que emergiu da Google.) Se tem de pensar nesta questão, encorajamo-lo a considerar alguns factos.[1]

Enquanto conduzem, 56 por cento dos adolescentes americanos falam ao telemóvel.

Em 2015, 2715 adolescentes americanos morreram e 221.313 foram parar às urgências, devido a ferimentos resultantes de acidentes de viação.

Metade de todos os acidentes que envolveram condutores adolescentes foram colisões de um só veículo.

Em contrapartida, os carros da Waymo nunca se distraem. Eles nunca bebem. Nunca ficam cansados e nunca enviam mensagens de texto aos amigos quando deviam estar atentos à estrada. Desde 2009, conduziram mais de três milhões de quilómetros em estradas públicas, e nesse período causaram apenas um acidente: um pequeno toque num autocarro urbano na Califórnia, enquanto circulavam a três quilómetros por hora. Globalmente, a taxa de acidentes por quilómetro, em que os Waymo foram os culpados, ao longo de nove anos, é 40 vezes inferior à de condutores com 16—19 anos, e é 10 vezes inferior à de condutores com 50-59 anos. E isso é o protótipo.

Estes números preveem uma trajetória nítida nas normas culturais do futuro: a noção de deixar um adolescente de 16 anos

conduzir um carro parecerá absurdamente irresponsável. Quando os nossos descendentes descobrirem que isto costumava ser normal, reagirão da mesma maneira como as pessoas o fazem hoje após descobrirem que os seus avós costumavam conduzir para casa sem cintos de segurança depois de beberem quatro martínis. E quanto às bicicletas, à neve e aos cangurus? Esses são apenas problemas de engenharia, que serão resolvidos no futuro próximo, talvez mesmo no momento em que está a ler isto, e quase de certeza com a mesma solução: melhores dados. Em IA, os dados são como a água — são o solvente universal.

Aliás, se passar tempo suficiente com cientistas de dados que trabalham com carros autónomos, rapidamente será confrontado com uma questão surpreendente: será que já nasceu o último californiano a ter uma carta de condução?

1. A Revolução Robótica

Os robots percorreram um longo caminho num curto período de tempo.

Na década de 1950, o topo de gama era o *Theseus*, um rato autónomo em tamanho real construído por Claude Shannon nos laboratórios Bell e alimentado por um módulo de relés de telefone. O antigo herói grego Theseus entrou num labirinto para matar o Minotauro. O rato *Theseus* tinha ambições mais modestas: entrou num labirinto de mesa com 25 quadrados para procurar um pedaço de queijo. Primeiro navegaria por tentativa e erro até encontrar o

queijo. Após este primeiro triunfo, conseguia encontrar o caminho até ao queijo a partir de qualquer ponto do labirinto sem se enganar.[2]

Nas décadas de 1960 e 1970 houve o *Cart* de Stanford: um veículo do tamanho de uma carroça com quatro rodas de bicicleta pequenas, um motor elétrico e uma única câmara de filmar. O *Cart* começou por ser uma viatura de teste para estudar como poderiam os cientistas controlar um veículo lunar remotamente a partir da Terra. Mas rapidamente se transformou numa plataforma para investigar a navegação autónoma para uma geração inteira de estudantes de robótica, no Laboratório de Inteligência Artificial de Stanford. Em 1979, após anos de aperfeiçoamento, o *Cart* conseguia conduzir-se a si próprio através de uma sala cheia de cadeiras em cinco horas, sem intervenção humana — um enorme feito para a época.[3]

E hoje? Os carros autónomos são apenas o princípio. Não se esqueça dos táxis autónomos voadores, como os que o governo do Dubai anda a testar desde setembro de 2017. Ou a mina de ferro autónoma gerida pela Rio Tinto, no meio do interior australiano. Ou o terminal de embarque autónomo no porto de Quingdao, na China — seis ancoradouros enormes que se estendem ao longo de dois quilómetros de costa, 5,2 milhões de contentores marítimos por ano, centenas de gruas e camiões robóticos, e ninguém ao volante.[4]

Uma das perguntas mais frequentes que ouvimos da parte dos estudantes nas nossas aulas de ciência de dados é: «Como é que estes robots funcionam?» Adoraríamos responder a essa questão

em todo o seu esplendor. Lamentavelmente, não podemos. Desde logo existem tantos detalhes que seria necessário um livro muito maior, com montes e montes de equações. Além disso, muitos dos detalhes são confidenciais. Talvez tenha ouvido, por exemplo, que a Waymo processou a Uber em 1,86 milhões de dólares pelo alegado roubo de algumas dessas particularidades — um processo cujo desfecho, quando este livro foi escrito, era desconhecido[5*].

Detalhes à parte, todavia, vamos pensar no panorama global. Eis uma analogia. Podemos certamente aprender os fundamentos do modo como um avião permanece no ar, mesmo que não saibamos construir um *Boeing 787*. Da mesma maneira, podemos aprender como é que um carro autónomo «lê» o seu ambiente, mesmo que nós próprios não consigamos criar um tal carro. É exatamente esse o nível de compreensão que adquirirá no final deste capítulo, tendo como base aquilo que já aprendeu acerca da probabilidade condicionada.

Para lá chegar começaremos com uma questão simples e quase infantil, que é fundamental para qualquer robot autónomo — quer ele ande, circule ou voe; quer ele desenterre ferro ou nos conduza até à mercearia; quer seja do tamanho de um rato ou de um navio porta-contentores. Esta questão, na verdade, é tão importante que tem de ser perguntada e respondida dezenas ou mesmo centenas de vezes a cada segundo.

Essa questão é: Onde é que eu estou?

Em IA, chama-se a isto um problema SLAM[†], de «localização e mapeamento simultâneos». Aqui a palavra «simultâneo» é determinante. Quer sejamos uma pessoa ou um robot, saber onde estamos significa fazer duas coisas ao mesmo tempo: 1) construir um mapa mental de um ambiente desconhecido, e 2) inferir a nossa própria localização desconhecida nesse ambiente. Isto é um problema do tipo «o ovo ou a galinha». As nossas crenças acerca do ambiente dependem da nossa localização, mas aquelas acerca da nossa localização dependem do ambiente. Nenhuma delas pode ser conhecida independentemente da outra, de modo que parece logicamente impossível inferir ambas ao mesmo tempo. Imagine, por exemplo, que está a tentar chegar a Times Square nunca tendo estado antes em Nova Iorque, e damos-lhe direções dizendo-lhe que fica a uma paragem de metro a norte de Penn Station — e depois, quando perguntasse onde fica Penn Station, nós dir-lhe-íamos que fica a uma paragem a sul de Times Square. Agora espera-se que encontre ambos os locais sem um mapa. Isso é o problema SLAM.

Embora possa não lhe parecer, a informação que obtém através dos seus sentidos acerca da sua localização no mundo é quase tão circular quanto as nossas direções para Times Square. O milagre cognitivo é que conseguirá resolver o problema SLAM de forma rotineira, sem qualquer esforço consciente, sempre que entrar numa sala desconhecida. Os neurocientistas não compreendem inteiramente como é que o faz, mas sabem que estão envolvidos bastantes circuitos cerebrais muito especializados e antigos do ponto de vista filogenético, particularmente no hipocampo. E, tal como muitas capacidades aprimoradas pela evolução, esta revela-

se bastante difícil de alcançar através da engenharia inversa. Em IA isto é frequentemente chamado «paradoxo de Moravec»*: as coisas fáceis para uma criança de cinco anos são as coisas difíceis para uma máquina, e vice-versa.

A revolução atual nos robots autónomos tornou-se possível apenas porque toda a investigação em torno dos sistemas SLAM finalmente deu frutos. Os robots passaram de evitar cadeiras a evitar outros condutores; de cinco horas a atravessar uma sala a cinco *gigabytes* de dados sensoriais por segundo; de um rato autónomo que pode percorrer uma grelha de 25 quadrados a um carro autónomo que pode circular por milhões de quilómetros de via pública. A SLAM é uma das maiores histórias de estrondoso sucesso da IA. Daí que, neste capítulo, gostaríamos de abordar duas questões relacionadas com SLAM — uma óbvia e outra um pouco mais inesperada.

1. Como é que um carro-robot sabe onde está?

2. Como é que pode tornar-se uma pessoa mais inteligente ao pensar um pouco mais como um carro robotizado?

A resposta a ambas as questões implica algo chamado *regra de Bayes*. A regra de Bayes define como os carros autónomos sabem a sua localização na estrada — mas é muito mais do que isso. É uma ideia matemática profunda usada todos os dias, em quase todas as áreas da ciência e da indústria. Além disso, é um princípio extremamente útil para que viva o seu quotidiano de maneira mais

inteligente — como, por exemplo, se quiser investir de maneira mais prudente, ou tomar decisões mais informadas acerca de cuidados de saúde. A regra de Bayes proporciona o melhor exemplo de como treinar-se a si próprio para pensar um pouco mais como uma máquina pode ajudá-lo a ser uma pessoa mais sábia e mais saudável.

2. Como é que Encontrar um Submarino É Semelhante a Encontrar-nos a Nós Próprios na Estrada?

Podemos agora, finalmente, regressar à nossa promessa anterior. Dissemos-lhe que compreender um submarino seria importante para conseguir que um carro se conduzisse a si mesmo. Agora vamos mostrar-lhe porquê.

Aqui a ligação é a regra de Bayes, a qual iremos explicar contando-lhe uma história acerca de um submarino. Não um submarino autónomo ou qualquer coisa desse tipo — somente um banal submarino nuclear de ataque chamado *USS Scorpion*. O *Scorpion* é famoso porque certo dia, em 1968, desapareceu, algures numa expansão de mar aberto que se estendia por milhares de quilómetros. Isto deixou os militares dos Estados Unidos num frenesim — é um grande problema quando um submarino nuclear desaparece. Apesar das hipóteses remotas, os oficiais da Marinha utilizaram na busca todos os meios que tinham à disposição. Vasculharam o oceano durante meses a fio, ainda assim não conseguiram encontrar o *Scorpion*. Desencorajados e sem esperança, estavam prestes a dar a busca por terminada.

Mas havia um homem demasiado teimoso para desistir. O nome dele era John Craven, e foi teimoso porque estava convencido de que tinha a probabilidade do seu lado — e o incrível é que estava certo. John Craven e a sua equipa de busca usaram a regra de Bayes para responder à questão «Onde é que o *Scorpion* se encontra neste oceano enorme e vazio?». Assim que souber como eles o fizeram, irá compreender como um carro autónomo utiliza a mesma matemática para responder a uma questão muito parecida: «Onde é que eu me encontro nesta enorme estrada?»

2.1 A Busca do *Scorpion*

Em fevereiro de 1968, o *USS Scorpion* zarpou de Norfolk, na Virgínia, sob o olhar atento do comandante Francis A. Slattery. O *Scorpion* era um submarino de ataque de alta velocidade da classe *Skipjack*, a mais veloz da frota americana. Tal como outros submarinos da sua categoria, desempenhou um papel fundamental na estratégia militar dos Estados Unidos: este foi o auge da Guerra Fria, e tanto os americanos como os soviéticos destacaram grandes frotas destes submarinos de ataque para localizar, seguir e — caso o impensável acontecesse — destruir os submarinos de mísseis balísticos do inimigo.

Neste destacamento, o *Scorpion* navegou para leste, com destino ao Mediterrâneo, onde durante três meses participou em exercícios de treino a par da 6ª Frota da Marinha. Então, a meio de maio, o *Scorpion* foi enviado de volta para oeste, passando Gibraltar e entrando no Atlântico. Aí foi-lhe ordenado que observasse embarcações navais soviéticas a operar perto dos Açores — uma

cadeia de ilhas longínqua no meio do Atlântico Norte, a cerca de 1500 quilómetros da costa de Portugal — e depois continuasse para oeste, em direção a casa. O submarino deveria chegar a Norfolk às 13h de segunda-feira, 27 de maio de 1968.

Nesse dia, as famílias dos 99 membros da tripulação do *Scorpion* tinham-se reunido nas docas para dar as boas-vindas a casa aos seus entes queridos. As 13h chegaram e o tempo continuou a avançar sem que o submarino surgisse à superfície. Os minutos transformaram-se em horas, o dia deu lugar à noite. As famílias permaneceram à espera. Mas não havia qualquer sinal do *Scorpion*.

Com a preocupação a aumentar, a Marinha ordenou uma busca. Às 22h a operação já envolvia 18 navios e, na manhã seguinte, 37 navios e 16 aviões de patrulha de longo alcance.[6] Mas as probabilidades de um desfecho favorável eram reduzidas. O último contato do *Scorpion* tinha sido ao largo dos Açores, há seis dias. Ele podia estar em qualquer parte ao longo daquela faixa de oceano com 4900 quilómetros entre os Açores e a costa leste. A cada hora que passava, as possibilidades de se poder localizar o submarino e enviar equipamento de resgate a tempo dissipavam-se rapidamente. Numa tensa conferência de imprensa no dia 28 de maio, o presidente Lyndon Johnson sintetizou o ânimo de uma nação: «Estamos todos bastante angustiados [...] Não temos qualquer informação que seja animadora.»[7]

Passados oito dias, a Marinha foi forçada a admitir o óbvio: os 99 homens da tripulação foram declarados perdidos no mar, presumivelmente mortos. Em seguida, a Marinha entregou-se à

tarefa sombria de localizar a última morada do *Scorpion* — uma agulha minúscula num vasto palheiro que se estendia ao longo de três quartos da travessia do Atlântico Norte. Embora as esperanças de salvar a tripulação estivessem desfeitas, ainda era essencial encontrar o submarino, e não apenas para as famílias dos desaparecidos: o *Scorpion* transportava dois torpedos nucleares, cada um capaz de afundar um porta-aviões com um só embate. Estas ogivas perigosas estavam agora algures no fundo do mar.

2.2 John Craven, Guru da Busca Bayesiana

Para liderar a busca, o Pentágono recorreu ao Dr. John Craven, cientista chefe do Gabinete de Projetos Especiais da Marinha e o maior guru no que respeita a encontrar objetos em águas profundas.

Surpreendentemente, Craven já fizera este tipo de coisas. Dois anos antes, em 1966, um bombardeiro *B-52* colidira em pleno ar com um avião de abastecimento sobre a costa espanhola, perto da vila costeira de Palomares. Ambos os aviões despenharam-se, e as quatro bombas de hidrogénio do *B-52*, cada uma 50 vezes mais poderosa do que a explosão de Hiroxima, foram espalhadas ao longo de quilómetros. Felizmente nenhuma das ogivas detonou, e três bombas foram prontamente encontradas. Mas a quarta bomba estava desaparecida e presumia-se que tivesse caído no mar.

Craven e a sua equipa tiveram de equacionar muitas variáveis desconhecidas acerca do acidente. Teria a bomba permanecido no avião ou caíra para o exterior? Se tivesse caído, os seus paraquedas

ter-se-iam aberto? Se os paraquedas se tivessem aberto, os ventos teriam levado a bomba para o mar alto? Se sim, em que direção, e exatamente a que distância? Para abrir caminho por este matagal de incertezas, Craven recorreu à sua estratégia preferida: a busca bayesiana. Esta metodologia tinha sido apresentada na Segunda Guerra Mundial, quando os Aliados a usaram para ajudar a localizar submarinos alemães. Mas as suas origens remontam muito mais atrás, até ao princípio matemático chamado regra de Bayes, descoberta pela primeira vez na década de 1750.[8]

A busca bayesiana tem quatro passos fundamentais. Primeiro, devemos criar um mapa de probabilidades prévias sobre a nossa rede de pesquisa. Estas probabilidades são «prévias» na medida em que representam as nossas crenças antes de termos qualquer dado. Elas combinam duas fontes de informação:

- As opiniões anteriores à busca de vários especialistas. No caso da bomba de hidrogénio desaparecida, alguns destes especialistas estariam familiarizados com choques em pleno voo, outros com bombas nucleares, alguns com correntes oceânicas e por aí fora.

- A capacidade dos nossos instrumentos de busca. Por exemplo: suponha que o cenário mais plausível coloca a bomba perdida no fundo de uma fossa oceânica profunda. Apesar da plausibilidade, podemos, no entanto, não querer começar a busca aí: a fossa é tão escura e remota que mesmo que a bomba ali esteja, seria pouco provável encontrá-la. Para recorrer a uma metáfora familiar, uma

> busca bayesiana põe-nos à procura da chave que perdemos usando uma combinação matemática precisa de dois fatores: onde achamos que a perdemos e onde a luz da rua brilha com mais intensidade.

Pode ver um exemplo de um mapa de probabilidades prévias no quadro superior da Figura 6.

O segundo passo é procurar no local de probabilidade prévia mais elevada, que é o quadrado C5 na Figura 6. Se encontrarmos o que procuramos, então estamos despachados. Se não, prosseguimos para o terceiro passo: rever as nossas crenças. Suponha que a busca à volta do quadrado C5 nada encontrou. Então, reduzimos a probabilidade em volta do quadrado C5 e em conformidade aumentamos a probabilidade de outras regiões. As nossas probabilidades prévias, à luz dos novos dados, tornaram-se agora probabilidades posteriores.

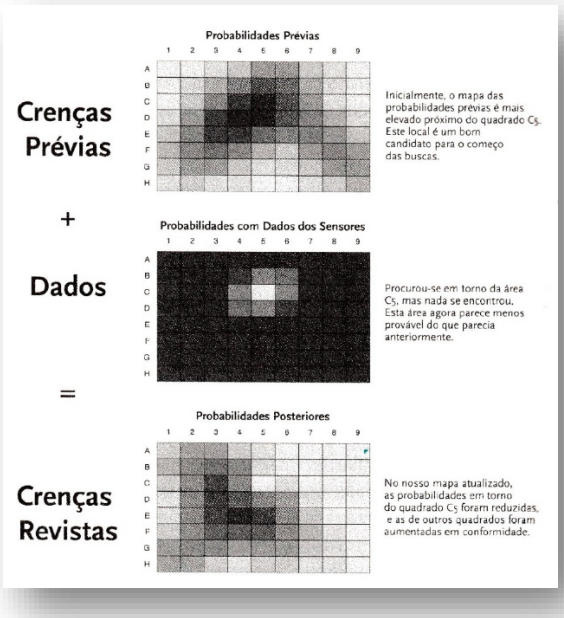


Figura 6. Na busca bayesiana, as crenças prévias são combinadas com dados de sensores de busca para produzir um conjunto de crenças revistas.

Podemos visualizar isto sobrepondo dois mapas:

- O mapa original de probabilidades prévias (quadro superior).

- O mapa de probabilidades com dados dos sensores (quadro intermédio). Estas probabilidades são baixas nas regiões onde procurámos e nada encontrámos, mas permanecem altas nas regiões em que não procurámos de todo, pelo que não podemos descartá-las.

Isto é a essência da regra de Bayes: crença prévia + factos = crença revista.

Quarto, e por fim repetimos. Reiteramos os passos 2 e 3, procurando sempre na região de maior probabilidade desse dia. Se continuarmos de mãos vazias, revemos as nossas crenças. A posterior de hoje transforma-se na prévia de amanhã, dia após dia, até encontrarmos o que procuramos.

2.3 Craven Impedido

Infelizmente, Craven e a sua equipa nunca chegaram a ter a oportunidade de aplicar estes princípios bayesianos à busca de 1966 da bomba H desaparecida ao largo de Palomares. Numa jogada militar clássica, o Pentágono pedira a Craven para fazer algo

importante, e depois nomeou outra pessoa com uma patente mais alta para lhe dificultar o mais possível a vida. O comandante no local, o contra-almirante William «Buldogue» Guest, tinha uma visão consideravelmente diferente de como a busca deveria ser conduzida. Ele tinha pouca paciência para probabilidades, regras de Bayes ou para doutorados em matemática com vinte e poucos anos, vestidos com bombazina e tecido Oxford. As suas ordens iniciais para Craven foram para ele provar que a bomba caíra em terra e não no mar, para que o trabalho de encontrar a maldita coisa recaísse sobre outra pessoa. Em resultado, a procura pela bomba H de Palomares consistia na verdade em duas pesquisas. Havia a busca bayesiana «sombra» de Craven, com as suas réguas de cálculo e mapas de probabilidades, e com números atualizados a zunir constantemente na teleimpressora à medida que os matemáticos alimentavam um computador central na Pensilvânia com cálculos feitos à distância. Mas as pistas que emergiam destes cálculos eram ignoradas a favor do «plano dos quadrados» do almirante Guest, que orientava a verdadeira busca.

Por fim, a bomba H de Palomares foi encontrada, depois de se descobrir que um pescador local a tinha visto cair no mar sob um paraquedas e podia conduzir a Marinha ao ponto da queda. Assim, ainda que a busca tenha sido um sucesso, a parte bayesiana dela foi um fracasso, pela simples razão de nunca ter tido uma oportunidade. Ainda assim, o incidente de Palomares ensinou algumas lições valiosas a John Craven — tanto sobre os aspetos práticos de conduzir uma busca bayesiana como sobre a maneira de reunir o apoio das chefias militares para essa procura.[9]

Dois anos mais tarde, quando foi chamado para encontrar o *Scorpion*, Craven estava preparado.

2.4 A Busca do *Scorpion* Continua

Quando o *Scorpion* desapareceu, em maio de 1968, Craven e a sua equipa de especialistas em busca bayesiana reuniram-se rapidamente. De início, a sua tarefa parecia bastante mais assustadora do que a busca da bomba de Palomares. Naquela época, eles sabiam que a investigação se confinava a uma área relativamente pequena nos mares pouco profundos ao largo da costa sul de Espanha. Agora, a equipa tinha de encontrar um submarino três quilómetros debaixo de água, algures entre a Virgínia e os Açores, sem ter sequer um único indício.

Por sorte, descobriram uma pista. Com começo no início da década de 1960, as Forças Armadas dos Estados Unidos tinham gastado 17 mil milhões de dólares a instalar uma rede enorme e altamente secreta de microfones subaquáticos por todo o Atlântico Norte, de modo a poderem seguir as movimentações da Marinha soviética. Técnicos altamente treinados em pontos de escuta secretos monitorizavam estes microfones 24 horas por dia. Depois de investigar, Craven descobriu que um destes postos de escuta nas ilhas Canárias tinha, um dia no final de maio, gravado uma série muito invulgar de 18 sons subaquáticos. Em seguida descobriu que outros dois pontos de escuta — ambos a milhares de quilómetros de distância, ao largo da Terra Nova — tinham registado esses mesmos sons por volta da mesma ocasião. A equipa de Craven comparou estas três leituras e, por triangulação, concluiu que os

sons deviam ter emanado de uma parte muito profunda do oceano Atlântico, cerca de 740 quilómetros a sudoeste dos Açores. Esta localização caía sobre a rota prevista do *Scorpion* de regresso a casa. Além disso, os próprios sons eram altamente sugestivos: uma explosão subaquática abafada, seguida de 91 segundos de silêncio, e depois mais 17 eventos sonoros em sucessão rápida que, a Craven, soavam à implosão de vários compartimentos de um submarino à medida que ele se afundava, entrando na profundidade de rutura do Casco.[10]

Esta revelação acústica estreitou drasticamente o tamanho da área de busca. Ainda assim, a equipa tinha de cobrir cerca de 360 quilómetros quadrados de fundo oceânico, todo ele três mil metros abaixo da superfície, e por isso somente acessível aos submersíveis mais avançados.

A busca bayesiana desenvolveu-se então a alta velocidade. Craven e a sua equipa entrevistaram tripulantes experientes de submarinos, que elaboraram nove cenários possíveis — um incêndio a bordo, um torpedo a explodir no seu compartimento, um ataque russo clandestino, etc. — para o afundamento do submarino. Avaliaram a probabilidade prévia de cada cenário e realizaram simulações em computador para compreenderem como se teriam desenrolado os movimentos prováveis do submarino sob cada um. Até rebentaram cargas de profundidade em localizações precisas, para calibrar os dados acústicos originais dos postos de escuta nas ilhas Canárias e na Terra Nova.

Por fim, juntaram toda esta informação para obter uma única probabilidade de eficiência de busca para cada célula da sua rede. Este mapa cristalizou muitos milhares de horas de entrevistas, cálculos, experiências e reflexão profunda. O seu aspeto seria parecido com a Figura 7.

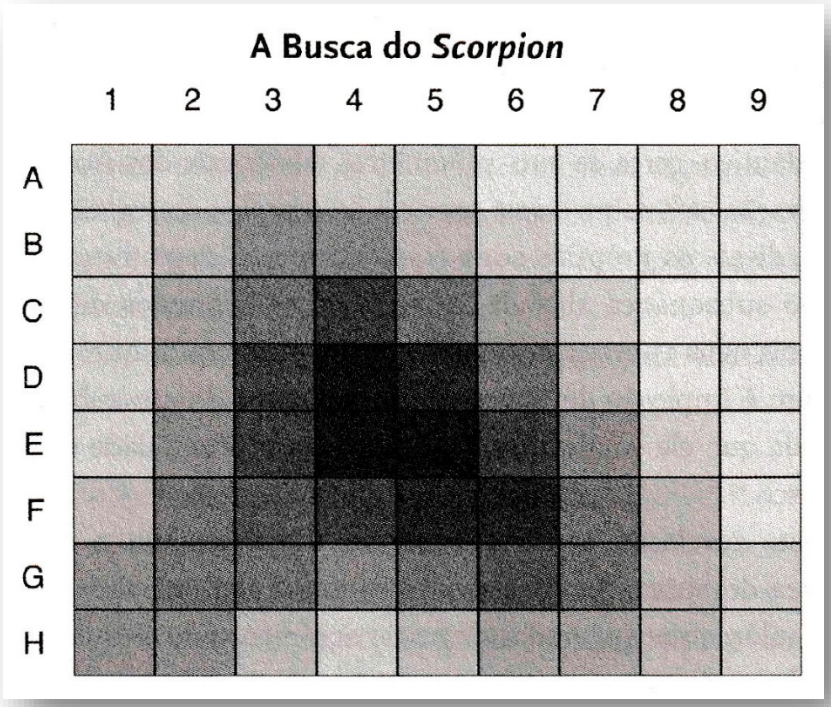


Figura 7. Uma reconstrução do mapa de probabilidades prévias utilizado na busca do *Scorpion*.

Como era previsível, Craven deparou-se com dificuldades logísticas e burocráticas ao tentar que o Pentágono prestasse atenção ao seu mapa de probabilidades. O verão chegou e partiu. Por esta altura, a busca do *Scorpion* já decorria há três meses, sem resultados.

Finalmente a sua insistência compensou e as chefias militares ordenaram que o mapa de Craven fosse usado para orientar a busca. Deste modo, no início de outubro, quando os comandantes

que lideravam a busca a bordo do *USS Mizar* finalmente receberam o mapa, a operação tornou-se verdadeiramente bayesiana. Dia após dia, a equipa investigou a região de maior probabilidade, verificou os números e atualizou o mapa para o dia seguinte. E, dia após dia, esses números encaminhavam-se lentamente para o retângulo F6.

A 28 de outubro, Bayes finalmente deu frutos.

O Mizar estava a meio da sua quinta viagem e da septuagésima quarta passagem individual pelo fundo oceânico. Subitamente o magnetómetro do navio disparou, sugerindo uma anomalia no fundo do mar. De imediato foram enviadas câmaras para investigar — e, de facto, lá estava ele. Parcialmente enterrado na areia, a 740 quilómetros da costa e três quilómetros abaixo da superfície do mar, o *USS Scorpion* fora finalmente encontrado.[11]

Até à data, ninguém sabe ao certo o que aconteceu ao *Scorpion* — ou se sabem não dizem. A versão oficial da Marinha sobre os acontecimentos refere a explosão acidental de um torpedo ou o mau funcionamento de uma unidade de eliminação de lixo como sendo as duas causas mais prováveis. Muitas outras explicações foram propostas ao longo dos anos — e, como acontece com qualquer mistério famoso, abundam as teorias da conspiração.[12]

Mas o incidente produziu pelo menos uma conclusão definitiva: a busca bayesiana funcionou de forma brilhante. Como se constatou, o local do descanso final do submarino ficava a meros

238 metros do retângulo E5, a região inicial mais promissora no mapa de probabilidades prévias de Craven. Na verdade, a equipa de busca até tinha passado por esse local numa viagem anterior mas não detetara os indícios devido a um sonar avariado.[13]

Figura 8. Uma foto da secção da proa do *USS Scorpion,* tirada em 1968 pela tripulação do batiscafo *Trieste II*.

Medite sobre isso mais um pouco. Pense em como é difícil encontrar algo que perdeu numa faixa de 30 metros de praia, ou até mesmo na sua própria sala de estar. No entanto, quando um submarino solitário desapareceu algures numa faixa de 4800 quilómetros de mar aberto, uma busca bayesiana tinha identificado a sua localização com um erro de 238 metros, apenas três vezes o comprimento do próprio submarino. Foi um triunfo notável para a equipa de Craven — e para a regra de Bayes, a fórmula matemática com 250 anos que tinha servido como o princípio orientador da busca.

3. Regra de Bayes, de Reverendo a Robô

Eis o mantra principal que retiramos da história do *Scorpion*: todas as probabilidades são na verdade probabilidades *condicionadas*. Por outras palavras, todas as probabilidades estão dependentes daquilo que sabemos. Quando o nosso conhecimento se altera, as nossas probabilidades também têm de se alterar — e a regra de Bayes diz-nos como fazê-lo.

A regra de Bayes foi descoberta por um clérigo inglês obscuro chamado Thomas Bayes. Nascido em 1701 numa família presbiteriana em Londres, Bayes revelou um talento precoce para a matemática, mas atingiu a maioridade numa época em que os dissidentes religiosos em Inglaterra eram impedidos de entrar nas universidades. Ao ser-lhe negada a possibilidade de estudar matemática em Oxford ou Cambridge, acabou por estudar teologia na Universidade de Edimburgo. Bayes, como tantos outros da sua época, terá encarado isto como uma barreira cruel. Mas esta discriminação teve um efeito colateral peculiar: a Inglaterra, devido às suas políticas religiosas intolerantes, acolhia um número surpreendente de sociedades amadoras de matemática formadas por presbiterianos talentosos que, tal como Bayes, viram vedado o acesso às universidades inglesas e por isso criaram localmente, como alternativa, as suas próprias comunidades intelectuais. Já na casa dos 40, Bayes tornou-se membro de uma dessas sociedades, numa cidade termal em Kent chamada Tunbridge Wells, onde aceitara um trabalho como pastor — e onde algures na década de 1750 inventou a regra que agora tem o seu nome.

Surpreendentemente, de início a sua descoberta não causou grande impacto. Bayes nem sequer a publicou durante a sua vida; morreu em 1761 e o seu manuscrito foi lido à Sociedade Real postumamente em 1763, pelo seu amigo Richard Price. Houve um período breve ao virar do século XX em que as ideias bayesianas floresceram, sobretudo pela mão do matemático francês Pierre-Simon Laplace. Mas após a morte de Laplace, em 1827, a regra de Bayes caiu no esquecimento e na irrelevância por mais de um século.

3.1 Atualização Bayesiana e Carros-Robôs

Hoje, no entanto, a regra de Bayes está de volta e melhor do que nunca, sentada ao volante de todos os carros robóticos que andam por aí.

A regra de Bayes é uma equação que nos diz como atualizar as nossas crenças à luz de informação nova, transformando probabilidades prévias em probabilidades posteriores. Ela oferece a solução perfeita para o problema da robótica que discutimos anteriormente: SLAM, ou localização e mapeamento simultâneos. A SLAM é um problema inerentemente bayesiano. À medida que chegam novos dados provenientes dos sensores, um carro-robot tem de atualizar o seu «mapa mental» do ambiente ao seu redor — as marcações na faixa de rodagem, os cruzamentos, os semáforos, os sinais de STOP e todos os outros veículos na estrada — ao mesmo tempo que infere a sua própria localização incerta *dentro* desse ambiente. Na sua essência, um carro-robot «pensa» em si

próprio como uma bolha de probabilidade, a viajar ao longo de uma estrada bayesiana.

Antes de descrevermos como isto funciona vamos abordar primeiro uma questão óbvia: Por que é que simplesmente não se navega usando a tecnologia de GPS, como a que tem no seu smartphone? O problema é que mesmo sob condições ideais, a precisão dos sistemas de GPS de classe civil tem uma margem de erro de cinco metros — e eles são muito piores em túneis ou junto de prédios altos, onde podem errar por 30 ou 40 metros. Tentar percorrer o trânsito da cidade usando apenas um GPS seria como realizar uma cirurgia vascular vendado e com luvas para o forno.

Assim, para suplementar a informação que obtém através do seu recetor GPS, um carro-robot tem de recorrer a uma panóplia de outros sensores. Alguns destes são simples câmaras de vídeo, enquanto outros são exatamente iguais aos dispositivos de segurança dos atuais carros novos — por exemplo, radar montado no para-choques, como aquele que apita sempre que há perigo de embater em algo ao fazer marcha-atrás.

O sensor mais fantástico e útil de um carro-robot chama-se LIDAR, uma amálgama das palavras «light» (luz) e «radar» (radar), que significa «Light Detection and Ranging» (deteção e telemetria por feixe de luz). Imagine que estava vendado e que o mandavam atravessar uma sala desconhecida apenas com a ajuda de uma bengala. Provavelmente fá-lo-ia através do toque; ou seja, usando a bengala para examinar o espaço à sua volta, medindo de maneira informal as distâncias às coisas na sua vizinhança imediata. Se o

fizesse vezes suficientes, em todas as diversas direções, então poderia criar um bom mapa mental do espaço em redor.

Um conjunto LIDAR funciona com o mesmo princípio: dispara um raio laser e mede a distância cronometrando quanto tempo a luz demora a ser refletida. Um conjunto LIDAR típico poderá ter 64 lasers individuais, cada um a emitir centenas de milhares de impulsos por segundo. Cada raio laser fornece informação detalhada sobre uma direção muito específica. Assim, para permitir que o carro veja em todas as direções, o LIDAR é montado no tejadilho, numa armação rotativa que gira aproximadamente 300 vezes por minuto, como se fosse uma versão mais rápida do feixe rotativo nos ecrãs do radar no *Top Gun*. Desta forma, os lasers irão apontar em qualquer das direções cerca de cinco vezes a cada segundo, dando ao carro atualizações posicionais discretas em vez de *contínuas*. Por outras palavras, o carro vê um mundo iluminado não por uma luminosidade constante mas por uma luz estroboscópica — por *flashes* curtos de dados do seu LIDAR e de outros sensores, cada um fornecendo ao carro uma nova perspetiva do ambiente em volta, como na Figura 9.

Cada vez que o carro recebe uma nova descarga de dados, usa a regra de Bayes para atualizar as suas «crenças» acerca da sua localização. Podemos visualizar este processo de atualização bayesiano usando um mapa em que a estrada está dividida em pequenas células de uma grelha, cada uma com a sua própria probabilidade. Imagine que saiu da garagem no seu carro autónomo, e que já está a viajar há 60 segundos, conduzindo a cerca de 48 km/h. Com base nos dados até ao momento, o carro

tem um conjunto de crenças acerca da sua posição. Isto está representado como um mapa de probabilidades no quadro superior esquerdo da Figura 10. Vamos agora conferir o que se passa com o carro um quinto de segundo mais tarde, após uma passagem do conjunto LIDAR, 60,2 segundos após o início da viagem. De que forma é que essas crenças se alteraram?

Figura 9. Uma imagem LIDAR de uma autoestrada, cortesia da Universidade Estadual do Oregon.

O raciocínio do carro tem três passos. O primeiro é aquilo a que os peritos em navegação chamam «navegação estimada», ou, como gostamos de chamar, «introspeção e extrapolação». *Introspeção* significa recolher informação sobre o «estado interno», como velocidade, ângulo das rodas e aceleração; *extrapolação* significa usar esta informação, em conjunto com as leis da física, para prever o movimento provável do carro durante a próxima fração de segundo. O resultado é um mapa de probabilidades prévias para a localização do carro ao fim de 60,2 segundos de viagem, representado no quadro superior direito da Figura 10. Estas

probabilidades são «prévias» porque ainda não incorporam dados atualizados dos sensores; estamos no instante imediatamente anterior ao próximo *flash* do estroboscópio.

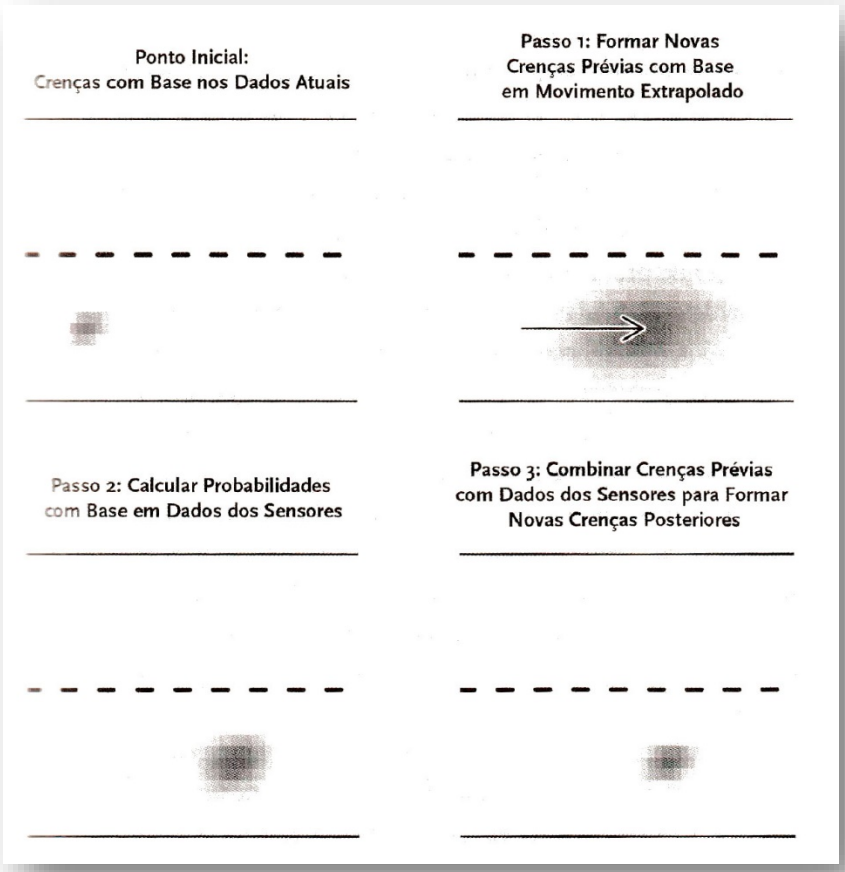


Figura 10. Como um carro autónomo usa a regra de Bayes para atualizar as suas crenças acerca da sua localização.

Neste ponto irá reparar em duas coisas: a bolha de probabilidade avançou um pouco ao longo da estrada, e a sua mancha «espalhou-se» para cobrir uma área maior. Este espalhar representa a incerteza adicional introduzida pela extrapolação. Por exemplo: se estivermos a viajar a 48 quilómetros por hora, esperamos percorrer 2,7 metros em 0,2 segundos. Mas na verdade poderemos percorrer um pouco mais ou um pouco menos em função de mudanças de direção, travagens ou acelerações imprevistas.

O segundo passo do carro é recolher dados através dos sensores externos, tais como as suas câmaras e o LIDAR. Isto proporciona uma verificação da posição real do carro, ajudando a corrigir os erros introduzidos pela extrapolação. Esta informação surge no quadro inferior esquerdo da Figura 10. Pode pensar neste mapa como um conjunto de probabilidades «apenas dos sensores» — isto é, na ausência de qualquer informação prévia, o que pensaria o carro acerca da sua posição ao basear-se apenas em sensores externos?

Contudo, o carro tem *mesmo* informação prévia, por isso o terceiro e último passo é a síntese. Usando a regra de Bayes, as probabilidades prévias com base na extrapolação (do passo 1) são combinadas com os dados dos sensores (do passo 2). No quadro inferior direito pode ver este novo mapa de probabilidades posteriores, que fornece uma resposta revista à questão fundamental: «Onde é que eu estou?» Crucialmente, a mancha da bolha das probabilidades posteriores é menos dispersa do que a das probabilidades prévias ou a dos dados do sensor por si só.

Deixámos aqui de fora bastantes detalhes. Eis o maior: no exemplo anterior fizemos de conta que a estrada era um quadro de referência fixo e que a única variável desconhecida era a localização do carro nesse quadro. Isso é o *L* em SLAM, de «localização». Mas não se esqueça do *M* de «mapeamento». Na realidade, a própria estrada é desconhecida e todas as suas caraterísticas são submetidas ao mesmo tratamento bayesiano. Divisões na estrada, linhas das faixas de rodagem, peões, outros carros, até cangurus — todos são representados como bolhas de probabilidade cujas

localizações estão constantemente a ser atualizadas com cada *flash* de dados provenientes dos sensores.

4. Como a Regra de Bayes Pode Torná-lo Mais Esperto

Do ponto de vista da regra de Bayes, encontrar um submarino perdido e encontrarmo-nos a nós próprios na estrada acabam por ser problemas muito similares. Mas a regra de Bayes é muito maior do que isso. Na verdade, em termos de aplicabilidade à vida quotidiana, é uma das equações mais úteis alguma vez descobertas — uma dose matemática perfeita de antidogmatismo que nos diz quando devemos ser céticos e quando devemos ter uma mente aberta. Pense em toda a informação nova com que se depara todos os dias. A regra de Bayes responde a uma questão muito importante: Quando é que essa informação deve fazê-lo mudar de opinião, e em que medida?

Na sua vida provavelmente nunca se sentou com papel e lápis para de facto desvendar a matemática da regra de Bayes, e isso não tem mal nenhum. A questão é que, mesmo que não o faça, aprender a pensar acerca do mundo um pouco como um carro bayesiano — em termos de antecedentes, dados e como combiná-los — pode ajudá-lo a ser uma pessoa mais sábia. Eis dois exemplos significativos.

4.1 Regra de Bayes em Diagnósticos Médicos

Vamos começar com um exemplo com alguns números evidentes acoplados — e no qual até especialistas altamente treinados tendem a dar a resposta errada, por não terem aplicado a regra de Bayes.

Imagine que é médico e que uma mulher de 40 anos chamada Alice entra no seu consultório para uma mamografia de rastreio de rotina. Infelizmente, o resultado da mamografia é positivo, indicando que ela poderá ter cancro da mama. Mas, devido ao seu treino médico, sabe que nenhum teste é perfeito e que Alice poderá ter obtido um falso positivo no resultado. O que deverá dizer-lhe acerca da probabilidade de ela ter cancro, considerando a mamografia positiva? Eis alguns factos para o ajudar a decidir.

- A prevalência de cancro da mama em pessoas como a Alice é de um por cento. Ou seja, por cada 1000 mulheres de 40 anos que realizam uma mamografia de rotina, cerca de 10 têm cancro da mama.

- O exame tem uma taxa de deteção de 80 por cento: se o realizarmos em 10 mulheres com cancro, irá detetar cerca de 8 desses casos, em média.

- O exame tem uma taxa de falsos positivos de 10 por cento: se o realizarmos em 100 mulheres com cancro da mama, irá sinalizar erradamente cerca de 10, em média.

À luz destes números, qual é a probabilidade posterior *P* (cancro | mamografia positiva)?

A resposta, de acordo com a regra de Bayes, é bastante baixa: apenas 7,4 por cento. Este número poderá surpreendê-lo. Se for esse o caso, não está sozinho: um número escandalosamente elevado de médicos imagina um número muito maior. Num estudo famoso deram a cem médicos a mesma informação que acabou de receber e 95 por cento estimaram que a *P* (cancro | mamografia positiva) era um valor algures entre 70 e 80 por cento.[14] Eles não deram apenas a resposta errada — enganaram-se por um fator de 10.

Este exemplo suscita duas questões. Primeira: por que é que a probabilidade posterior *P* (cancro | mamografia positiva) é de apenas 7,4 por cento apesar do facto de a mamografia ser 80 por cento precisa? Segunda: como é que tantos médicos podem dar uma resposta tão errada?

A explicação à primeira pergunta é esta: a maioria das mulheres que obtém um resultado positivo numa mamografia é saudável, porque a grande maioria das mulheres que realiza mamografias à partida é saudável. Em termos simples, o *cancro tem uma probabilidade prévia baixa*. Podemos visualizar isto usando um *diagrama em cascata*, que é como uma versão «quotidiana» do mapa de probabilidade que um carro autónomo utiliza para circular na estrada. Na Figura 11, acompanhamos uma coorte hipotética de mil mulheres, todas com 40 anos, enquanto efetuam mamografias de rotina. O ramo esquerdo mostra 10 mulheres (um por cento de

mil) que de facto têm cancro da mama. Dado que o exame é 80 por cento preciso, contamos que destes 10 casos dois sejam ignorados e oito sejam detetados. Entretanto, o ramo direito mostra 990 pacientes sem cancro. Como o exame tem uma taxa de falsos positivos de 10 por cento, contamos que cerca de 890 estejam fora de perigo e 100 sejam sinalizados incorretamente, arredondando um pouco.[15]

Assim, no fundo da cascata ficamos com mil casos, distribuídos da seguinte maneira.

- 108 mamografias positivas. Destas, oito são verdadeiros positivos, ou casos de cancro que foram detetados. Os restantes 100 são falsos positivos, ou mulheres saudáveis incorretamente sinalizadas pelo exame.

- 892 mamografias negativas. Destas, duas são falsos negativos, ou casos de cancro ignorados. As outras 890 são verdadeiros negativos, ou mulheres que corretamente receberam um atestado de saúde.

Cada um destes mil casos é igualmente provável, por isso pintámos os respetivos quadrados com o mesmo cinzento-claro. O facto de o cancro ser relativamente improvável reflete-se não pelo sombreado mas pelos números absolutos: apenas 10 destes mil quadrados correspondem a casos de cancro.

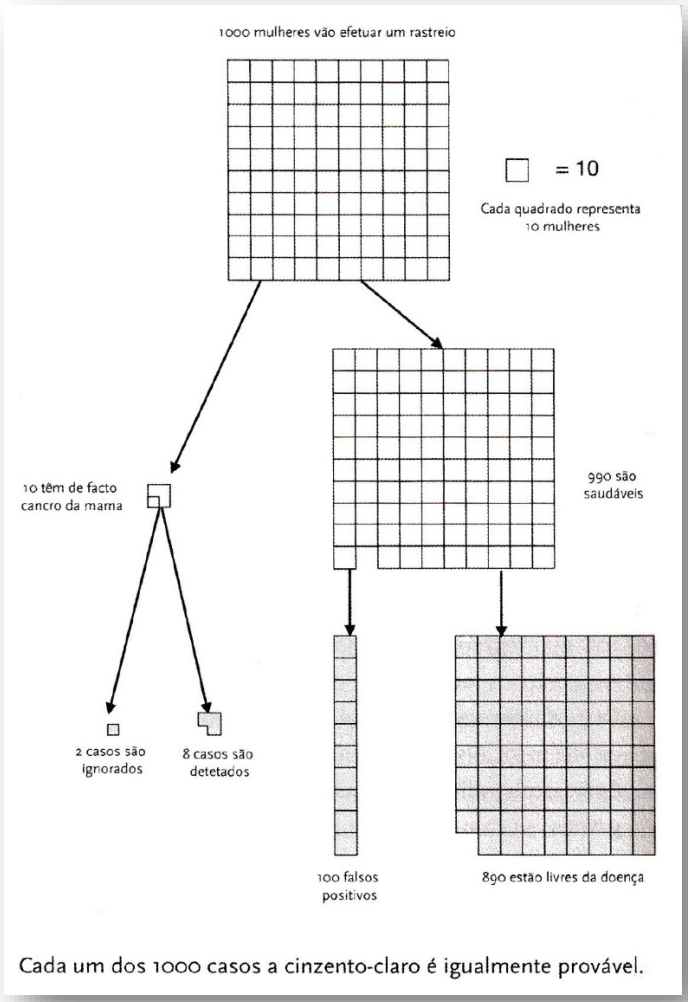


Figura 11. Um diagrama em cascata que acompanha uma coorte hipotética de mil mulheres com 40 anos enquanto realizam mamografias de rastreio.

Agora vamos usar este diagrama para refletir sobre a situação da sua hipotética paciente, a Alice. Quando entra pela primeira vez no consultório do médico, sabemos que Alice será como uma das mil mulheres na parte de baixo do diagrama em cascata. Só não sabemos qual será. Quando a mamografia dela se revelar positiva, então saberemos que Alice terá de ser como uma das 108 mulheres com o mesmo resultado. Então revisitemos o diagrama e vamos pintar esses 108 casos de cinzento-escuro, ao mesmo tempo que «excluímos» os outros 892 casos pintando-os de branco, na Figura 12.

Destas 108 mamografias positivas, oito são verdadeiros casos de cancro, enquanto 100 são falsos positivos. Portanto, a probabilidade posterior de Alice ter cancro, *P* (cancro | mamografia positiva), é cerca de 8/108 ≈ 7,4 por cento.

É isso a regra de Bayes. A probabilidade prévia de cancro era um por cento. Depois de vermos os dados, a probabilidade posterior de cancro passa a ser 7,4 por cento — bastante mais elevada do que a prévia, mas ainda assim a léguas de distância do valor de 70-80 por cento estimado pela maioria dos médicos. (Se gostaria de ver isto calculado usando uma equação verdadeira, veja a matéria adicional no final do capítulo.)

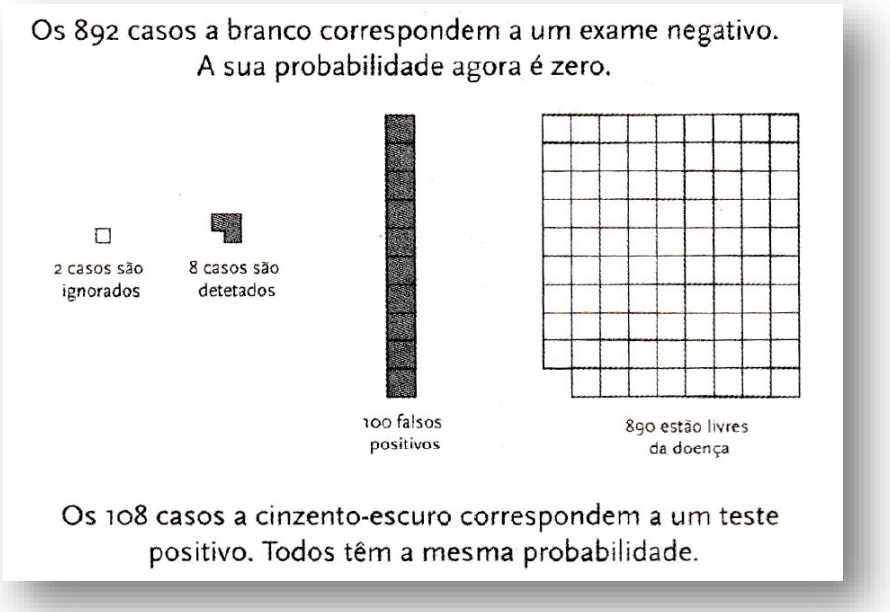


Figura 12.

Vamos agora abordar a segunda questão que colocámos anteriormente. Quando lhes pediram para estimar a probabilidade posterior *P* (cancro | mamografia positiva), por que é que tantos

médicos surgiram com um valor 10 vezes demasiado alto? Basicamente, os médicos estavam a ignorar a probabilidade prévia, algo a que se chama «falácia da taxa-base». As estimativas dos médicos de 70-80 por cento não tiveram em conta a taxa baixa de cancro na população (um por cento), o que implica que a maioria dos exames positivos serão falsos positivos. Em vez disso, os médicos estavam focados apenas num número: o facto de o teste ser «80 por cento preciso», querendo dizer que deteta 80 por cento de verdadeiros casos de cancro. Eles estavam a dar demasiado crédito aos dados e crédito insuficiente ao antecedente.

Deste exemplo, podem inferir-se três morais. Primeiro, nunca pergunte ao seu médico «O exame é exato?». Na melhor das hipóteses receberá a resposta correta à pergunta errada. Ao invés, pergunte: «Qual a probabilidade posterior de eu ter a doença?» (Contudo, esteja preparado para um olhar carrancudo, pois o seu médico poderá não saber o que isto é.)

Segundo, embora a regra de Bayes seja convencionalmente expressa como uma equação, raramente necessitará dessa equação para calcular a probabilidade posterior. Em vez disso, pode simplesmente fazer um diagrama em cascata como o da Figura 11 seguindo uma coorte hipotética de sujeitos através de um determinado processo de recolha de dados. Irá desfrutar da omeleta bayesiana sem ter de partir o ovo matemático.

Finalmente, nunca negligencie a taxa-base, também conhecida como antecedente, quando estiver a interpretar os dados. A regra de Bayes diz que a probabilidade posterior certa alcança-se sempre

ao combinar dados e antecedentes — tal como um carro-robot circula na estrada.

4.2 Regra de Bayes e Investimento

Na verdade, uma vez que fique sensibilizado para o fenómeno da falácia da taxa-base, começará a vê-lo em todo o lado. Como mostra o nosso próximo exemplo, é uma falácia especialmente importante para se ter em atenção quando contemplar uma das decisões financeiras mais importantes da sua vida: como investir para a sua reforma.

Em termos gerais existem duas estratégias financeiras populares para um portefólio de reforma: indexar e apostar. «Apostar» significa que tenta escolher um vencedor, confiando o seu dinheiro a um gestor de fundos ativo que tenta suplantar o mercado. «Indexar» expressa que desiste de tentar vencer o mercado e em vez disso simplesmente *compra* o mercado, sob a forma de um índice de ações de base alargada, como o S&P 500.

Os defensores da estratégia de jogo argumentam que é inteiramente possível derrotar o mercado a longo prazo. O melhor argumento deles para essa afirmação não tem mais do que duas palavras: Warren Buffett. Também conhecido como o «Oráculo de Omaha», Warren Buffett é uma figura ímpar na história do investimento. Os números do seu desempenho são espantosos: de 1964 até 2014, um investidor na empresa-mãe de Buffett, a Berkshire Hathaway, teria transformado 10 mil dólares em 182 milhões. Igualmente notável é a consistência de Buffett: as ações

que escolheu suplantaram o S&P 500 ao longo de quase todos os períodos contíguos de cinco anos, desde meados da década de 1960. E embora Buffett seja a história de sucesso mais famosa de Wall Street, também tem havido outras — uma mão-cheia de verdadeiros génios do mercado, de Joel Greenblatt a Peter Lynch, cuja trajetória é demasiado impressionante para ser atribuível à pura sorte. Os investidores que têm identificado e confiado nestes extraordinários gestores de fundos têm sido recompensados generosamente.

Porém, em contraponto ao exemplo destes génios raros, encontramos uma realidade numérica agreste: a maioria dos gestores de fundos em nada se parecem com Warren Buffett. Os números do desempenho deles foram particularmente devastadores no período de dez anos entre 2007 e 2016, uma década que incluiu um colapso histórico seguido de um crescimento rápido do mercado. Estas condições eram ideais para *stock pickers*[*] inteligentes. Contudo, segundo a Standard and Poor's, 86 por cento dos fundos de ações geridos ativamente tiveram um desempenho inferior ao dos seus índices de referência durante este período. Na Europa, o cenário foi pior: 98,9 por cento dos fundos de ações nacionais, 97 por cento de fundos de mercado emergentes e 97,8 por cento de fundos de ações internacionais tiveram um desempenho abaixo do índice de referência, após comissões. Os gestores ativos de fundos nos Países Baixos conquistaram o prémio global: 100 por cento deles não conseguiram superar os indicadores de referência.[16]

A conclusão é que há por aí verdadeiro talento para selecionar ações, mas é difícil encontrá-lo. Então, como é que estes factos devem influenciar a sua estratégia de investimento? Deve contentar-se com um fundo de índice ou apostar na grandeza, na esperança de que conseguirá encontrar um daqueles raros gestores de fundos que de facto conseguem vencer o mercado?

Se decidir apostar, então deverá ser honesto consigo próprio acerca do seu objetivo: conduzir a sua própria busca bayesiana do próximo Warren Buffett. Os «locais de busca» possíveis são todos os diferentes gestores de fundos que competem pelo seu capital e os seus «dados de busca» são as informações sobre o registo de desempenho de cada gestor. Quais são as hipóteses de a sua busca conseguir localizar uma das raras exceções, num oceano de gestores de fundos que não conseguem de forma consistente vencer o mercado?

Infelizmente, a regra de Bayes dá uma resposta bastante clara: as hipóteses são péssimas.

Para explicar porquê, iremos recorrer a uma metáfora que torna o assunto fácil de abordar usando probabilidades: a maioria dos gestores de fundos mutualistas está apenas a atirar uma moeda ao ar. Há anos em que sai cara, e eles vencem o mercado; noutros sai coroa, e é o mercado que os vence. (Claro que, independentemente de obterem cara ou coroa, eles continuarão a cobrar-lhe comissões.)

Mas um investidor raro como Warren Buffett está, metaforicamente, a atirar uma moeda ao ar com cara de ambos os lados. Ele vence o mercado ano após ano, sem exceção.

De acordo com esta metáfora, se fôssemos comparar o desempenho de Warren Buffett ao longo de dez anos com o de cinco *stock pickers* comuns, poderíamos ver algo parecido com a tabela abaixo.

	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5	Ano 6	Ano 7	Ano 8	Ano 9	Ano 10	Total Caras
Jane Doe	C	C	C	c	C	c	c	C	c	c	5
John Bull	c	c	C	C	C	C	c	c	c	C	5
Jean Dupont	C	C	c	C	C	C	C	c	c	C	7
Jan Jansen	c	c	C	C	C	c	c	c	C	c	4
Max Mustermann	c	c	C	C	c	c	c	c	c	C	3
Warren Buffett	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**10**

C: Cara; c: coroa

No caso dos nossos cinco investidores médios, o desempenho é simplesmente aleatório. Mas, no caso de Buffett, o desempenho dele é guiado pela sua capacidade superior de selecionar ações — aquela moeda de duas caras que ele mantém escondida num cofre em Omaha, Nebrasca. Como pode ver na tabela, o seu notável desempenho destaca-se dos demais.

Mas eis o problema: em Wall Street temos de destacar-nos entre uma multidão muito maior. Lá não há apenas cinco gestores

medíocres de fundos, a atirarem moedas ao ar e a cobrarem comissões. Há milhares e milhares deles — e há boas hipóteses de que pelo menos alguns irão desfrutar de longas séries de vitórias, por mero acaso.

É aí que entra a regra de Bayes. Imagine um frasco contendo 1024 moedas normais. No interior deste frasco um amigo coloca uma única moeda com duas caras. Em seguida o seu amigo agita bem o frasco, e o leitor retira uma única moeda ao acaso. Quer saber se a sua moeda tem duas caras, mas é contra as regras espreitar para os dois lados: não conseguiria fazer isso no mundo real, uma vez que todos os gestores de fundos que andam por aí têm um fantástico discurso de vendas que os faz parecer uma moeda de duas caras. Sendo assim, está obrigado a realizar um teste *estatístico* para duplo encabeçamento, que consiste em atirar a moeda ao ar 10 vezes.

Agora suponha que nas 10 tentativas saiu sempre cara. Perante as evidências, estará a segurar a moeda de duas caras ou uma das 1024 moedas normais? Para responder com a regra de Bayes a esta questão, consideremos os seguintes factos:

- Há 1025 moedas no frasco; 1024 são normais e uma tem duas caras.

- A moeda única de duas caras garante que sairá cara 10 vezes seguidas.

- Qualquer moeda normal escolhida ao acaso tem uma probabilidade de 1/1024 de sair cara 10 vezes seguidas. (Isto é calculado multiplicando 1/2 por si próprio 10 vezes.) Desta forma, das 1024 moedas normais no frasco esperamos que nunca saia cara 10 vezes seguidas.

Podemos tabular toda esta informação da seguinte maneira:

	Pelo menos 1 coroa	10 caras seguidas
Moedas Normais	1023 (verdadeiros negativos)	1 (falsos positivos)
Moeda de duas caras	0 (falsos negativos)	1 (verdadeiros positivos)

Esta matriz diz-nos que, das 1025 moedas no frasco, esperamos que em duas delas saia cara 10 vezes seguidas. Apenas uma delas é de facto a moeda de duas caras. A probabilidade de estar a segurar nela é de apenas 50 por cento, mesmo depois de a atirar ao ar 10 vezes.

Vamos agora comparar este cenário das moedas num frasco com o tipo de linguagem de marketing que poderá ouvir da parte de um *stock picker* de um fundo mutualista, gerido de forma ativa com um desempenho acima da média:

> Veja o meu desempenho anterior. Eu tenho gerido o meu fundo nos últimos dez anos e bati o mercado em todos os anos. Se eu fosse apenas um desses *stock pickers* medianos, com um fundo inferior, isto seria bastante improvável: menos de uma hipótese em mil.

A matemática deste cenário é exatamente igual à do grande frasco de moedas. Metaforicamente, o gestor de fundos está a alegar que é uma moeda de duas caras, com base em ter atirado a moeda ao ar 10 vezes e ter sempre saído cara, batendo o mercado todos os anos durante dez anos. Mas, da sua perspetiva, as coisas não são tão claras. Deve reconhecer que o astuto discurso de marketing está implicitamente a misturar duas probabilidades diferentes: *P* (série de vitórias de 10 anos | *stock picker* bom) com *P* (*stock picker* bom | série de vitórias de 10 anos). Mas lembre-se da nossa lição principal da história de Abraham Wald: as probabilidades condicionadas não são assim simétricas.

Então, o gestor de fundos é bom ou tem sorte? Vamos efetuar o cálculo bayesiano sob duas assunções prévias diferentes. Primeiro, suponha que acredita que um por cento de todos os *stock pickers* são efetivamente capazes de bater o mercado e que os outros 99 por cento apenas estão a atirar moedas ao ar. Sob esta assunção, vamos imaginar seguir uma coorte de 10 mil *stock pickers* durante 10 anos.

- Todos os 100 *stock pickers* excelentes (1% de 10 mil) vão bater o mercado todos os anos.

- Um *stock picker* medíocre tem uma hipótese em 1000 de bater o mercado 10 vezes seguidas.* Como existem 9900 *stock pickers* medíocres, esperamos que cerca de 10 deles batam o mercado em todos os 10 anos, por mera sorte.

Portanto, são 110 investidores a bater o mercado, dos quais 100 são bons e 10 tiveram sorte. Assim, a probabilidade posterior *P* (investidor que bate o mercado | série de vitórias de 10 anos) é 100/110, ou cerca de 91 por cento.

E se, contudo, acreditasse que a excelência é muito mais rara, tal que *P* (investidor que realmente bate o mercado) = 1/10.000? Sob esta assunção, a probabilidade posterior acaba por ser muito mais baixa:

- Agora há apenas um *stock picker* que bate o mercado todos os anos.

- Dos 9999 *stock pickers* medíocres, voltamos a esperar que 10 deles batam o mercado 10 vezes seguidas, apenas por mero acaso.

Assim, temos *P* (investidor que bate o mercado | série de vitórias de 10 anos) = 1/11, ou cerca de nove por cento.[17]

A regra de Bayes implica que a reação certa ao desempenho de um investidor depende fortemente da probabilidade prévia, ou seja, se os excelentes gestores de fundos são comuns ou raros. Contudo, todas as evidências disponíveis sugerem que os investidores que verdadeiramente batem o mercado são extremamente raros. Lembre-se de todas aquelas estatísticas pavorosas em que apenas alguns poucos fundos superavam o seu indicador num único ano, quanto mais em dez anos seguidos.

Para o investidor típico, isto tem uma consequência muito importante. Podem de facto existir muitos bons *stock pickers* por aí. Mas a regra de Bayes implica que, sem um muito longo registo de desempenho, não se consegue distinguir de forma fiável estes génios do muito maior grupo de medíocres que está apenas a ter sorte. Até mesmo a genialidade de Warren Buffett só se tornou evidente apenas após um período de algumas décadas. De modo que, no que toca à questão de procurar um talentoso gestor de fundos, a lição da regra de Bayes é: não vale a pena arriscar. É ainda mais difícil do que encontrar um submarino perdido em 4800 quilómetros de mar aberto. Certamente ficaríamos mais bem servidos investindo num índice alargado de ações e obrigações, em vez de tentar selecionar vencedores.

Ainda assim, a esperança é a última a morrer. Por isso, no caso de o seu otimismo não ter sido afetado pela dura realidade da regra de Bayes, deixamo-lo com este pensamento. Se está à espera de encontrar o próximo Warren Buffett, só terá de reger-se por um único discurso de vendas — no início da carreira de um *stock picker*, os dados de desempenho são praticamente inúteis. Portanto, negoceie com cuidado ou acabará por se deixar enrolar pelo gestor com língua de prata, em vez de obter proveitos de ouro.

5. Pós-Escrito

Conhecemos a regra de Bayes como um princípio para encontrar um submarino perdido, e hoje a busca bayesiana é uma pequena indústria, com empresas inteiramente dedicadas à consultoria sobre operações de busca e salvamento.[18] Por exemplo: talvez se lembre

da tragédia do voo 447 da Air France, que se despenhou no oceano Atlântico durante a sua viagem do Rio de Janeiro para Paris, em junho de 2009. Nos finais de 2011, a busca dos destroços decorria há dois infrutíferos anos. Então contrataram uma firma de busca bayesiana, desenharam um mapa de probabilidades e encontraram o avião ao fim de uma semana de buscas subaquáticas.[19] Mais: a ideia principal da regra de Bayes — atualizar o conhecimento prévio à luz de novas evidências — aplica-se em todo o lado, nomeadamente ao volante de um carro autónomo. Os biólogos usam-na para ajudar a compreender o papel dos nossos genes na explicação do cancro. Os astrónomos usam-na para encontrar planetas que orbitam outras estrelas nos confins da nossa galáxia. Tem sido usada para detetar *doping* nos Jogos Olímpicos, para filtrar o *spam* da sua caixa de correio e para ajudar os tetraplégicos a controlar braços robóticos diretamente com a mente, tal como o Luke Skywalker.[20] E, como viu, é essencial para navegar nos terrenos traiçoeiros dos cuidados de saúde e das finanças.

Portanto a regra de Bayes é muito mais do que um princípio para encontrar o que se tiver perdido. Sim, ajudou a encontrar o *Scorpion*, e ajuda os carros autónomos a encontrarem-se a si próprios na estrada. Mas também pode ajudá-lo a descobrir sabedoria para enfrentar a torrente de informação com que se depara diariamente.

6. Matéria Adicional: Regra de Bayes como Equação

Na maioria das situações quotidianas não necessita de saber a própria equação da regra de Bayes para poder aplicar a lógica

subjacente. Mapas e diagramas em cascata, como os apresentados neste capítulo, podem levá-lo bastante longe com muito pouca matemática. Mas se está interessado numa carreira em ciência de dados, ou se simplesmente gosta de perceber os pormenores, é bom ver essa equação. Por isso aqui fica a regra de Bayes da maneira que é ensinada num curso universitário de IA ou estatística.

Iremos usar a letra H para representar uma hipótese que pode ser verdadeira ou falsa, e o D para representar certos dados relevantes. A regra de Bayes mostra-nos como usar os dados para transformar a probabilidade prévia da hipótese, P (H), numa probabilidade posterior, P (H | D):

$$P(H \mid D) = \frac{P(H) \ * P(D \mid H)}{\mathrm{P(D)}}$$

No nosso exemplo do exame médico, H é a hipótese de uma dada paciente ter cancro da mama e D é o dado de o resultado da mamografia dela ter sido positivo. Sabemos que um por cento dos pacientes têm cancro da mama: P (H) = 0,01. De igual modo, sabemos que o exame é 80 por cento fiável a detetar cancro da mama, se ele estiver presente: *P* (D | H) = 0,8. A última coisa de que precisamos é P (D), a probabilidade geral de um exame positivo. A partir do nosso diagrama em cascata sabemos que em 1000 exames, cerca de 108 serão positivos: oito verdadeiros positivos e 100 falsos positivos. Assim, a *P* (D) ≈108/1000 = 0,108.

E isto é tudo aquilo de que precisamos. Vamos introduzir estes três números na regra de Bayes para calcular a probabilidade posterior de cancro, dado o exame positivo:

$$P\,(H \mid D) \;=\; \frac{0{,}01 * \;0{,}8}{0{,}108} \;=\; 0{,}074$$

Esta é a mesma probabilidade (7,4%) que calculámos a partir do diagrama em cascata.

---

[1] Estatísticas do Insurance Institute for Highway Safety, http://www.rmiia.org/auto/teens/TeenDrivingStatistics.asp.

[2] «Claude E. Shannon: A Goliath Amongst Giants», https://www.belllabs.com/claude-shannon/.

[3] Les Earnest, «Stanford Cart», dezembro de 2012, https://web.stanford.edu/learnest/cart.htm.

[4] Thuy Ong, «Dubai Starts Testing Crewless Two-Person "Flying Taxis"», *The Verge*, 26 de setembro de 2017, https://www.theverge.com/2017/9/26/16365614/dubai-testing-uncrewed-two-person-flying-taxis-volocopter; Tom Simonite, «Mining 24 Hours a Day with Robots», *MIT Technology Review*, dezembro de 2016, https://www.technologyreview.com/s/603170/mining-24-hours-a-day-with-robots/; «Asia's First Automated Container Terminal, at Port of Qingdao, China», reportagem ao vivo na New China TV, 11 de maio de 2017, https://www.youtube.com/watch?v=bn2GPNJmR7A.

[5] Peter Henderson, «U.S. Judge Deals Setback to Waymo Damage Claim in Uber Lawsuit», Reuters, 3 de novembro de 2017, https://www.reuters.com/article/us-alphabet-uber-lawsuit/u-s-judge-deals-setback-to-waymo-damage-claim-in-uber-lawsuit-idUSKBN1D32Jo.

[*] Depois de um ano de batalha judicial, e quando o julgamento estava na fase inicial, a Uber chegou a acordo com a Google e aceitou pagar cerca de 245 milhões de dólares e comprometer-se a não utilizar nenhum do *hardware* ou *software* ligado à condução autónoma e pertença da Google. [N. T.]

[†] Sigla inglesa para as palavras Simultaneous Localization and Mapping. [N. T.]

[*] Batizado em homenagem a Hans Moravec, um pioneiro da robótica.

[6] William Beecher, «Vast Search Fails to Find Submarine», *The New York Times*, 29 de maio de 1968, A1.

[7] «The President's News Conference of May 28, 1968», em *Public Papers of the Presidents of the United States: Lyndon B. Johnson, 1968-1969* (Washington, D. C.: USGPO, 1970), 656.

[8] Estes detalhes e outros subsequentes acerca do incidente de Palomares foram retirados de Sharon Bertsch McGrayne, *The Theory That Would Not Die* (New Haven: Yale University Press, 2011), 182-94.

[9] *Ibid*., 192-94.

[10] Documentário da PBS Nova, «Submarines, Secrets, and Spies», transmitido inicialmente a 19 de janeiro de 1999, disponível em https://www.youtube.com/watch?v=RvjTAMQQQUY (Já não disponível).

[11] McGrayne, The Theory That Would Not Die, 202.

[12] Documentário da PBS Nova, «Submarines, Secrets, and Spies».

[13] McGrayne, The Theory That Would Not Die, 202.

[14] David M. Eddy, «Probabilistic Reasoning in Clinical Medicine: Problems and Opportunities», em *Judgment Under Uncertainty: Heuristics and Biases*, ed. Daniel Kahneman, Paul Slovic, e Amos Tversky (Cambridge: Cambridge University Press, 1982), 249-67.

[15] Na verdade são 99 falsos positivos, mas arredondámos para 100 para ser mais fácil trabalhar com os números. Se corrigir o nosso modesto erro de arredondamento, a verdadeira probabilidade posterior *P* (cancro | teste positivo) é de 7,5%, não 7,4%.

[*] *Stock pickers* — investidores que gerem e selecionam ações de forma ativa. [N. T.)

[16] Madison Marriage, «86% of Active Equity Funds Underperform», *The Financial Times*, 20 de março de 2016, https://www.ft.com/content/e555d83a-ed28-11e5-888e-2eadd5fbc4a4 (Já não funciona).

[*] Aqui arredondámos todos os números.

[17] Veja a matéria técnica adicional na página 89.

[18] Lawrence D. Stone, *The Theory of Optimal Search* (Nova Iorque: Academic Press, 1975).

[19] Lawrence D. Stone, Colleen M. Keller, Thomas M. Kratzke e Johan P. Strumpfer, «Search for the Wreckage of Air France Flight AF 447», *Statistical Science* 29, n.º 1 (2014): 69-80.

[20] «Breakthrough: Robotic Limbs Moved by the Mind», *60 Minutes*, transmitido inicialmente a 30 de dezembro de 2012, disponível em https://www.cbsnews.com/news/breakthrough-robotic-limbs-moved-by-the-mind-30-12-2012/.

CAPÍTULO 4

«Amazing Grace»

De Babel aos bits: como as máquinas aprenderam a falar a nossa língua.

Ao longo do tempo em que os humanos têm tentado que as máquinas compreendam a linguagem, essas máquinas têm vindo a cometer erros infantis. Por certo já se debateu com a função de autocorreção do seu telemóvel. Ou talvez tenha viajado para o estrangeiro e visto como os serviços de tradução da Internet podem induzir as pessoas em erro, quer seja no jardim zoológico («Não alimente os animais; dê toda a comida ao guarda de serviço») ou na lavandaria («Baixe as suas calças aqui»). E a piada recorrente entre especialistas em IA é que se Stanley Kubrick tivesse realizado hoje o seu filme *2001: Odisseia no Espaço*, então a conversa entre Dave e o malévolo supercomputador *Hal-9000* poderia ter decorrido assim:

> Dave: HAL, podes abrir as portas do módulo do hangar.
> HAL: Dave, eu procurei na Internet e encontrei alguns resultados para iPods. Gostarias de os ver?

As máquinas também cometem erros subtis. Em tempos, o supercomputador *Watson* da IBM foi confrontado com um teste de

rimas antes do seu grande confronto com concorrentes humanos no *Jeopardy*! Uma das dicas no teste era «um termo de boxe para bater abaixo da cintura». A resposta em rima correta era «golpe baixo» — mas Watson respondeu «*wang bang*», uma expressão que não figurava na sua base de dados e que ele terá cunhado por iniciativa própria*.

Por isso vá em frente, acrescente mais uns insultos. No entanto, encorajamo-lo a ter presentes dois factos. Primeiro, as pessoas também cometem erros com a linguagem. Espalham barbaridades como «para todos os propósitos intensivos» ou «ao seu aceno chamativo»†. E interpretam erradamente letras de músicas, por exemplo do Billy Joel («We didn't start the fire, it was always burning, said the worst attorney») ou da Madonna («Like a virgin, touched for the thirty-first time»).‡ As pessoas também cometem erros de tradução. Em 2009, por exemplo, a secretária de Estado Hillary Clinton transformou uma oferta ao ministro dos Negócios Estrangeiros russo num espetáculo elaborado. O presente era um enorme botão vermelho que pretendia dizer «Reiniciar» tanto em inglês como em russo, para simbolizar a política da administração Obama de «premir o botão de reiniciar» as relações com a Rússia. A política não resultou lá muito bem, todavia — e o mesmo aconteceu com o presente, que afinal em russo não significava «Reiniciar» mas «Sobrecarregar».

O segundo facto a ter em mente é que as máquinas estão a ficar melhores no que respeita à linguagem — e rapidamente. (Tem de admitir que «*wang bang*» é uma definição criativa em comentários de boxe.) Os especialistas em IA usam o termo «processamento de

linguagem natural», ou PLN, para descrever como conseguimos que os computadores trabalhem com a linguagem. Ao longo dos últimos anos temos vivido num período de tremendo crescimento de sistemas de PLN bem-sucedidos:

- Assistentes digitais como o *Echo* da Amazon e o *Google Home* são imensamente melhores do que os toscos programas de «discurso para texto» de apenas há alguns anos. Eles podem agendar reuniões, criar uma lista de compras, escolher uma canção ou acumular os gastos no seu cartão de crédito — tudo através da voz e com um nível de precisão na transcrição que até há pouco tempo pareceria ficção científica.

- A versão do *Google Tradutor* que ficou operacional em 2016 representou um extraordinário melhoramento em relação a esforços anteriores de tradução automática. Agora o *software* pode gerar traduções aceitáveis para mais de cem línguas — muitas delas diretamente a partir da câmara do seu smartphone, como com o menu de um restaurante ou uma tabuleta numa estação de comboios. O *Skype* consegue fazer algo parecido durante um *chat* de vídeo, em tempo real.

- Os *chatbots*— software concebido para estimular a conversação humana — estão a tornar-se uma funcionalidade pervasiva do mundo digital. Eles são especialmente populares no *Facebook Messenger*, onde pode pedir a um *bot* para marcar uma viagem através do

> *Kayak* ou importunar um comerciante para que verifique a situação de uma encomenda que está atrasada. Os *chatbots* são ainda mais populares na China, onde a maioria das *startups* cria um *bot* oficial no *WeChat* — base de utilizadores: 930 milhões de pessoas — antes mesmo de criar uma página na Internet.

Atualmente as máquinas até estão a aprender a escrever. A Associated Press começou a usar um algoritmo que é capaz de escrever um resumo razoável de um jogo de basebol a partir da tabela de estatísticas, e ao qual recorrem nos jogos universitários em localidades distantes sem a presença de jornalistas. O sistema até aprendeu a incluir os lugares-comuns da linguagem jornalística; ele simplesmente analisa os dados de um jogo de cada vez. Recentemente, os cientistas de dados da Salesforce desenvolveram um programa similar que consegue resumir corretamente artigos extensos para ajudar os empregados da empresa a digerirem as notícias mais rapidamente. E enquanto académicos que passaram as passas do Algarve com a revisão pelos seus pares, não ficámos de todo surpreendidos quando soubemos de um algoritmo criado por investigadores da Universidade de Trieste — um que criava falsas revisões pelos pares suficientemente boas para enganar editores de revistas verdadeiros.[1]

Depois há o projeto paralelo do programador de software Andy Heard, que treinou uma rede neural com um monte de guiões do *Friends*, a famosa série da década de 1990, para ver que tipo de novos episódios ela escreveria. Claro que os resultados são absurdos, mas um absurdo surpreendentemente do tipo do *Friends*.

A Monica é estranhamente agressiva, o Chandler queixa-se bastante e até há participações especiais de estrelas de cinema aleatórias dos anos 90:

> Van Damme: Eu alinho nessa merda.
>
> Monica: Continua a falar!
>
> Phoebe: Calma aí, menina! Vais simplesmente chegar-te a ele toda saltitante...
>
> Chandler: Então, a Phoebe gosta das minhas calças.
>
> Monica: Chicken Bob!
>
> Chandler (Dentro de um queque)
>
> (Corre até às raparigas para chorar):
>
> Podem dar-me alguns presentes?[2]

Imagine o que poderia obter se houvesse mais do que 236 episódios — uma grande longevidade para uma série, mas uma pequena quantidade de dados de treino segundo padrões de redes neurais.

Assim, se quer compreender um futuro com sistemas de IA que reconhecem a linguagem, então a questão interessante não é acerca dos erros por vezes cómicos que estas máquinas produzem. Em vez disso, é sobre como é que elas aprenderam a ouvir, falar e até a escrever de forma tão eficaz.

1. Um Conto de Duas Revoluções

Há na realidade duas revoluções que devemos mencionar aqui. Há a Revolução da Linguagem de Programação, que teve o seu auge na década de 1950, e há a Revolução da Linguagem Natural, que vivemos atualmente. Estas duas revoluções diferem de maneiras importantes, mas há uma grande ideia que as une: para que uma máquina entenda palavras, temos de representar essas palavras numa linguagem com a qual a máquina possa trabalhar. Isso significa que temos de transformar palavras em números.

Durante décadas, a única maneira eficaz de fazer isto era através de uma abordagem descendente (do topo para a base) alicerçada em regras pré-especificadas. Pense nestas regras como um contrato que descreve como duas partes, a «máquina» e o «humano», podem usar a linguagem para interagir. Pense no contrato legal mais detalhado que conseguir imaginar, escrito pelos advogados mais bem pagos que por aí andam. Agora torne-o cem vezes mais detalhado.

- Há um conjunto de regras para o humano, a que se chama linguagem de programação. (Exemplos famosos são Python, Java e Pearl.) Uma linguagem de programação tem símbolos matemáticos como + e =, a par de um vocabulário limitado de palavras inglesas, habitualmente apresentadas numa fonte de largura fixa para intimidar as pessoas: IF, THEN, WHILE, etc. A linguagem também possui uma gramática: regras para combinar as palavras

de modo a formarem «frases» legais que dão instruções à máquina para que faça algo específico.

- Depois há um conjunto de regras para a máquina, codificadas em algo chamado «compilador».[3] Estas regras atuam nos bastidores e são invisíveis para o programador humano. Elas fornecem às máquinas instruções detalhadas passo a passo para traduzir qualquer frase imaginável da linguagem de programação para a sua própria «linguagem de máquina» de *bits* e vetores.

Este contrato é interpretado da forma mais literal possível. Se escrevermos uma frase gramatical na linguagem de programação, então a máquina tem de fazer *exatamente* o que nós dissermos. E se nos desviarmos um bocadinho que seja da gramática, tal como darmos um erro ortográfico numa palavra ou esquecermo-nos de um ponto e vírgula, então a máquina basicamente mostra-nos o dedo médio, ou como gostamos de o escrever, 00100.

Ao longo de décadas, estes eram os únicos termos sob os quais pessoas e computadores podiam ter uma conversa bem-sucedida. Como irá descobrir neste capítulo, eles são uma enorme melhoria face à maneira como as coisas eram no início da Era dos computadores, quando as pessoas eram forçadas a falar com eles na linguagem nativa «binária» deles de zeros (0) e uns (1). Mas estes termos dificilmente nos deixam usar os nossos poderes de linguagem para passarmos a nossa mensagem. É claro que também podemos fazer com que os computadores realizem algumas coisas triviais ao apontar, clicar, arrastar, etc. Mas isso simplesmente é

tão grosseiro...! Imagine que tinha de comunicar com outras pessoas apenas apontando para as coisas, ou usar menus que lhes caíssem das sobrancelhas. A linguagem é muito mais eficaz — e, desde a década de 1950, se queríamos usar a linguagem para realmente dar ordens a um computador estávamos presos a uma linguagem de programação.

Mas isso já não acontece. Desde cerca de 2010, as mentes mais brilhantes em IA forjaram um segundo conjunto de termos contratuais — um «New Deal» (novo acordo) para a interação linguística entre o humano e a máquina. Este novo acordo é ascendente em vez de descendente. Começamos por deitar fora o enorme livro de regras gramaticais predeterminadas. Ao invés, podemos falar com as máquinas na nossa própria língua natural: inglês, chinês, coreano, o que for. E as máquinas têm de interpretar o que pretendemos dizer, e responder na linguagem que escolhermos, sem terem um qualquer advogado a dizer-lhes ao ouvido que elas podem ignorar-nos se nós falharmos um ponto e vírgula.

Para forjar este *New Deal* demos três coisas às máquinas, cujo significado explicaremos neste capítulo.

1. Brinquedos: GPU (unidades de processamento gráfico) rápidas e imensa memória.

2. Software sofisticado, sob a forma de redes neurais baseadas em «vetores-palavras», uma ideia realmente

fantástica na interseção entre linguagem e matemática, que nos permite transformar palavras em números, de modo a podermos usá-los para construir regras preditivas.

3. Acima de tudo, o precioso manancial de dados que ficou disponível ao longo das duas últimas décadas, quando a produção linguística da humanidade se tornou maioritariamente digital.

Esta última é a mais importante. As pessoas dependem de milhares de milhões de factos linguísticos, a maioria dos quais tomam como garantidos — como o conhecimento de que «deixe as suas calças» e «*deixe cair* as suas calças»* são usados em situações muito diferentes, e apenas um deles tem lugar na lavandaria. Conhecimento como este é difícil de codificar em regras explícitas, porque existe em demasia. Acredite ou não, a melhor maneira que conhecemos de ensiná-lo às máquinas é dar-lhes um disco rígido enorme, cheio de exemplos de como as pessoas dizem as coisas, e deixar que a máquina descubra por si própria com um modelo estatístico.

Esta abordagem impulsionada meramente pelos dados poderá parecer ingénua, e até há pouco tempo não tínhamos dados suficientes ou computadores suficientemente rápidos para fazê-la funcionar. Atualmente, no entanto, funciona surpreendentemente bem. Por exemplo: a Google, na sua conferência de tecnologia, anunciou ousadamente que as máquinas já estavam a par dos humanos em relação ao reconhecimento do discurso, com uma taxa de erro por palavra nos ditados de 4,9 por cento — drasticamente

melhor do que as taxas de erro de 20-30 por cento que eram habituais ainda em 2013. Este salto quântico no desempenho linguístico é uma das principais razões por que as máquinas agora parecem tão inteligentes.

Podemos até argumentar, de facto, que na última década o reconhecimento do discurso ao nível humano é o avanço mais importante em inteligência artificial.

Então, quando foi o momento de viragem, e como é que chegámos lá? O que são «vetores-palavra», e por que são tão úteis? Por que é que os dados são tão importantes aqui — por que é que não podemos fazer com que uma máquina simplesmente siga regras linguísticas que escrevemos explicitamente, do mesmo modo que ensinamos um aluno do terceiro ano a perceber a gramática ou uma máquina a perceber Python?

Para responder a estas questões, gostaríamos de contar-lhe a história de Grace Hopper. Ela foi apelidada de «Amazing Grace», e não apenas por ser a única pessoa deste livro a ter ido ao programa de David Letterman*. Hopper doutorou-se em matemática em Yale em 1934, ingressou na Marinha dos Estados Unidos durante a Segunda Guerra Mundial e serviu o seu país em uniforme durante mais de 42 anos. Pelo meio, tornou-se a primeira pessoa na história a fazer com que um computador entendesse inglês. De modo que a história das máquinas capazes de falar, ouvir e escrever — a história do *Watson*, da *Alexa*, dos *chatbots*, do *Google Tradutor* e de todas as outras maravilhas do mundo digital —, na verdade tudo começa com a Amazing Grace».

2. Grace Hopper, Rainha do Software

Grace Hopper nasceu em Nova Iorque em 1906. Quando era miúda depressa aprendeu que a sua família tinha um apreço particularmente elevado por dois valores: autossuficiência e servir o país. Num passeio de verão que Grace fez com os seus pais a New Hampshire, ela estava a remar sozinha numa canoa. De repente, uma rajada de vento virou a canoa, atirando Grace para a água. Porém, a sua mãe, que da margem via Grace a esbracejar na água, parecia estranhamente despreocupada. Ela simplesmente pegou num megafone e gritou: «Lembra-te do teu bisavô, o almirante!» De imediato Grace nadou de regresso à margem, com a canoa a reboque.[4]

O bisavô em questão era o contra-almirante Alexander Wilson Russell, que quando era jovem combatera os piratas da Berbéria e que mais tarde serviu na Marinha da União. Mas a linhagem militar de Grace era ainda mais antiga. Durante toda a sua vida ela iria contar a história de Samuel Lemuel Fowler, um antepassado que em 1775 marchara com o seu mosquete até Concord, no Massachusetts, para lutar pelo seu país. Grace faria o mesmo 168 anos mais tarde, embora ela lutasse pelo seu país ao lado dos britânicos, em vez de contra eles.[5]

No outono de 1924, Grace Hopper rumou ao Vassar College decidida a preparar-se para o mundo do trabalho. Nesse mesmo ano, Vassar apresentou três cursos novos no seu currículo: «Maternidade», «Marido e Mulher» e «A Família Enquanto Unidade Económica». Hopper não frequentou esses cursos. Pelo contrário,

optou por «Eletromagnetismo», «Probabilidade e Estatística» e «A Teoria das Variáveis Complexas». Com o grande encorajamento da sua mãe, sempre estivera interessada em matemática e nunca nos caminhos tradicionais reservados para as mulheres dessa época. Ela prosperou em Vassar, formando-se com distinção em 1928 nos cursos de Matemática e Física, e pouco depois seguiu para Yale para entrar no programa de doutoramento em Matemática.

Em 1931, com a sua dissertação ainda por completar, Hopper regressou a Vassar para integrar a equipa de docentes de matemática, onde a sua paixão e curiosidade pela matéria se revelaram contagiantes. Ela tornou-se numa professora extremamente popular; fez com que um curso que tinha dez inscritos passasse a contar com 75, a par de uma lista de espera. Ela tinha uma maneira peculiar de ensinar matemática, com pouca abstração e muita demonstração prática. Para ensinar a matemática do deslocamento, por exemplo, levou toda a sua turma até à casa de banho e colocou uma pessoa dentro de uma banheira.[6] Ela terminou a sua dissertação em Yale à distância, doutorando-se em 1934, e continuou a ensinar em Vassar durante a década seguinte.

2.1 Grace Hopper na Guerra: o *Harvard Mark I*

O eclodir da Segunda Guerra Mundial iria para sempre mudar a vida de Grace Hopper. Em 1942, com a lembrança de Pearl Harbor — e o seu bisavô — bem viva na sua mente, ela tentou alistar-se na Reserva Naval Feminina, uma das poucas funções militares disponíveis para as mulheres. Aos 35 anos, no entanto, era

considerada demasiado velha e com 1,52 metros e 48 quilos estava sete quilos abaixo do peso exigido. Foi rejeitada. Mas na família de Grace a determinação feroz em servir era simplesmente um ponto assente. Ela tentou novamente, apresentando documentação especial para obter uma revogação do peso exigido. Desta vez foi aceite e em dezembro de 1943 entrou para as Reservas da Marinha dos Estados Unidos.[7] O curso de aspirante passou a voar, como Grace referiu: «Trinta dias para aprender a acatar ordens, 30 dias para aprender a dar ordens e era-se oficial da Marinha.»[8] Ela graduou-se com as melhores notas da sua turma de 800 alunos e foi nomeada tenente (2ª classe) em unho de 1944.

Devido à sua formação em matemática, Hopper pressupôs que seria colocada numa unidade de criptografia para ajudar a decifrar os códigos de rádio das forças do Eixo. Na realidade, porém, havia algo ainda mais adequado a alguém com a sua formação. Foi-lhe dito para se apresentar em Cambridge, no Massachusetts, onde ela se tornaria a terceira pessoa apenas a aprender a operar o *Harvard Mark I*, o primeiro computador digital programável do país. Anos mais tarde, quando Hopper já era famosa e um entrevistador lhe perguntou como é que ela entrara para o mundo da computação, ela simplesmente respondeu: «A Marinha ordenou-me que fosse para o primeiro computador dos Estados Unidos, e eu apresentei-me ao serviço».[9]

O *Mark I* havia sido concebido por um homem chamado Howard Aiken, fora construído pela IBM, e doado a Harvard — onde Aiken era professor assim como comandante da Marinha — em prol do esforço de guerra. Era uma monstruosidade, mais comprido do que

um semirreboque e mais pesado do que dois rinocerontes: cinco toneladas, 15,5 metros de comprimento, 2,4 metros de altura e um metro de largura. Tinha 853 quilómetros de cabos, 765 mil interruptores eletromecânicos e uma caixa elegante e modernista desenhada por Norman Bel Geddes. O *Mark I* distinguia-se dos outros primeiros computadores na medida em que se destinava verdadeiramente a múltiplos fins. Ele podia lidar com equações diferenciais, álgebra linear, análise harmónica e estatística; podia ser programado para simular um foguete, um submarino, uma onda de radar, tudo o que possa pensar. Aiken, o seu cocriador, chamava-lhe «uma máquina de aritmética geral», mas os jornais preferiam termos como «cérebro robótico» ou «supercérebro algébrico».[10] Quando os mandachuvas militares foram visitá-lo, Aiken gabou-se de o *Mark I* ser tão rápido que podia adicionar três números a cada segundo, ou efetuar uma divisão longa a cada 14,7 segundos.[11] Como aparte, é interessante comparar estes números com um iPhone X de 2017:

	Tamanho (centímetros)	Peso (gramas)	Adições por segundo
Harvard Mark I (1944)	1550x250x90	4,284,180	3
Apple iPhone X (2017)	14x7x1	138	350.000.000.000

Medido por computações por segundo por unidade de volume, um *iPhone X* é quatro quatrilhões (4×10^{15}) de vezes mais capaz. Ainda assim, o *Mark I* era muito mais rápido a efetuar cálculos do que uma pessoa. Além disso, a fila para o novo *iPhone* tinha 73 anos de espera.

Quando Hopper chegou a Harvard, o seu comandante deu-lhe uma semana para aprender a programar o *Mark I*. Como irá aprender daqui a algumas páginas, era um trabalho lento e frustrante — contudo, não havia qualquer manual de instruções, nenhum *chatbot* de apoio técnico nem tempo a perder. Foi no verão de 1944. Os soldados aliados estavam a tomar de assalto as praias da Normandia e a equipa do *Mark I* tinha a missão de calcular as tabelas balísticas que mostravam aos soldados como apontar a sua nova artilharia de longo alcance. Além disso, esse estava longe de ser o único projeto importante da equipa. O de maior dimensão era o «Problema K», em agosto de 1944: um conjunto de cálculos imensamente complexo e altamente secreto, pedido por um laboratório de Los Alamos, no Novo México. Quando este pedido chegou, o *Mark I* foi reservado durante semanas. Mas a Marinha disse para darem ao matemático de Los Alamos — John von Neumann, que trabalhava em algo chamado Projeto Manhattan — todo o tempo que ele pedisse.

2.2 Como é que se Falava com um Computador em 1944?

Depois da guerra, Hopper facilmente podia ter voltado para Vassar e vivido como professora catedrática. No entanto, ela gostava mais de computadores do que de estabilidade, de modo que permaneceu nas Reservas da Marinha para iniciar a sua nova vida como uma das únicas pessoas peritas em computadores. Nesse papel, Hopper rapidamente daria um passo histórico, nascido da frustração que experienciou ao programar o *Mark I*: tornar-se-ia a primeira pessoa a falar em inglês com um computador.

Depois da guerra, Hopper trabalhou no Laboratório de Computação de Harvard durante quatro anos. Em seguida, em 1949, aceitou um emprego na Eckert-Mauchly Computer Corporation, que fabricou um computador chamado *UNIVAC*. Foi uma decisão crucial, tanto para Grace como para o futuro da computação. Antes do *UNIVAC*, os computadores eram vistos como ferramentas ótimas para realizar cálculos matemáticos e científicos, mas para pouco mais. Os especialistas achavam que talvez houvesse procura para 20 computadores em todo o país, principalmente em laboratórios de investigação governamentais. Mas o trabalho de Hopper no *UNIVAC* alterou esse cenário. Ela demonstrou que os computadores também eram úteis para resolver problemas empresariais relativos a bases de dados — uma complicação que simplesmente nunca surgira nos problemas de matemática pura que o *Mark I* resolvera durante a guerra. Com a ajuda de Hopper, as empresas em todo o lado começaram a ver o potencial destas novas máquinas. A US Steel acabou por comprar um *UNIVAC* para gerir a sua folha de pagamentos. A Metlife comprou um para calcular prémios de seguros. A Dupont, a General Electric, o Departamento do Censo e a Westinghouse, todos compraram um *UNIVAC* para processar os seus dados, tornando-o o primeiro computador com êxito comercial.[12]

Para explicar a grande inovação de Hopper nesta área temos de voltar a uma questão de 1944: como é que Hopper deu instruções ao *Harvard Mark I*? Como é que ela disse a 765 mil interruptores eletromecânicos para dançarem uma batida que calculava uma tabela balística?

Certamente não o fez em inglês. Hopper descreveu-o desta maneira: «Simplesmente decompúnhamos todos os nossos processos matemáticos numa série de passos muito pequenos de adição, multiplicação, divisão [...] e púnhamo-los em sequência.»[13] Ela fazia com que parecesse tão simples. Mas não era. A parte difícil era frasear estas instruções na própria «linguagem de máquina» do Mark I, a única que ele podia compreender.

Para perceber o que é uma linguagem de máquina, imagine um programa de computador para fazer chá. Numa linguagem de computação moderna de alto nível como a *Python*, o programa poderia ser lido desta maneira: 1) Ponha duas colheres de chá com chá num bule. 2) Ferva 470 mililitros de água. 3) Despeje a água a ferver sobre o chá e deixe a infusão repousar durante quatro minutos. Mas em linguagem de máquina, teríamos de decompor essas instruções em tarefas muito mais pequenas e específicas. Em vez de dizer «ferver água», primeiro descreveríamos como caminhar até ao lava-loiças: mova o seu pé esquerdo, mova o seu pé direito, mova o seu pé esquerdo, uma e outra vez. Em seguida descreveríamos como encher a chaleira: levante a sua mão esquerda, agarre a torneira, rode o manípulo no sentido contrário ao dos ponteiros do relógio, e assim por diante. Em seguida descreveríamos como aquecer a água, infundir o chá e servi-lo, tudo com um idêntico e entediante detalhe biomecânico. Além disso, teríamos de emitir cada instrução não em inglês mas usando códigos numéricos introduzidos no computador através de furos feitos numa longa fita de papel perfurado. Estes códigos diziam ao *Mark I* exatamente como manipular os *bits* («dígitos binários», ou 0 e 1) nos seus circuitos internos. Enquanto programadores,

teríamos de saber que códigos realizavam quais tarefas. No nosso exemplo de preparação de chá, o código 72 04 poderia significar «mova o pé esquerdo», 61 07 poderia significar «agarre a torneira com a mão esquerda», e assim por diante.

Portanto: 72 O4, 61 O7... isto é típica linguagem de máquina. Está muito longe de «Ser, ou não ser», muito longe até de «*Alexa*, toca música dos anos oitenta». Como o autor Douglas Hofstadter disse: «Olhar para um programa escrito em linguagem de máquina é vagamente comparável a olhar para uma molécula de ADN átomo a átomo.»[14] Mas é assim que os computadores «pensam», mesmo nos nossos dias. E nos primórdios da Era digital, simplesmente não havia outra maneira de dizer-lhes o que fazer. Enquanto programadores dessa Era, éramos basicamente canalizadores binários, enviando *bits* através dos circuitos de um computador com a ajuda de um livro de códigos que ensinava a traduzir problemas matemáticos em linguagem de máquina. Procuraríamos itens no livro de códigos, faríamos os furos certos na fita, inseriríamos a fita no computador e faríamos figas para que tudo corresse bem.

Falar desta maneira com um computador era entediante e propenso a erros. Pior ainda: alguns dos primeiros computadores nem sequer usavam números decimais comuns (base 10). Em vez disso utilizavam números octais (base 8), o que originava uma matemática alucinante do tipo 7 + 1= 10 ou 5 x 5 = 31.* Isto deixava os programadores loucos, e Grace Hopper não era exceção. Certa vez, após passar semanas a programar um computador chamado *BINAC* em octal, Hopper percebeu que não conseguia fazer o balanço do seu livro de cheques. Ela refez as contas

inúmeras vezes, mas não conseguia encontrar o seu erro. Finalmente percebeu. Os números dela não coincidiam com os do banco porque, sem dar por isso, ela geria o livro de cheques em octal.[15]

2.3 Grace Hopper Inventa o Compilador

Para Hopper, o episódio do livro de cheques demonstrou realmente o problema com os computadores: eles simplesmente não falavam a nossa língua. Mas ela também entreviu a possibilidade de haver uma maneira melhor. Tudo começou com a ideia de observar e registar padrões comuns, ou o que mais tarde viria a ser conhecido como «sub-rotina».

Talvez conheça a história acerca dos prisioneiros que ouviram cada anedota tantas vezes que atribuíram um número a cada uma, só para tornar as coisas mais fáceis. Assim, um tipo grita «31!» e o outro prisioneiro desata a rir. Outro tipo grita «17!» e há ainda mais gargalhadas. Então um terceiro tipo grita «104!», mas desta vez toda a gente fica em silêncio, porque a piada está toda na forma de contar a anedota.

Na computação, a sub-rotina é como uma piada numerada: um bocado de código genericamente útil para resolver um problema que surge repetidamente, tal como fatorizar uma equação de segundo grau ou classificar uma lista de números. Sempre que Hopper escrevia uma sub-rotina para o *Mark I* ela copiava-la à mão para um bloco de notas, para que na próxima vez não tivesse de reinventar a roda. Em pouco tempo ela passou a ter um extenso

conjunto de sub-rotinas escritas em linguagem de máquina. Quando queria usar uma de novo voltava a copiá-la do seu bloco de notas para a fita. Isto demorava muito tempo e bastava um só erro durante a cópia para arruinar todo o programa. Hopper percebeu, no entanto, que o Mark I era muito melhor a copiar do que qualquer humano. Isto deu-lhe uma ideia: por que não armazenar uma biblioteca destas sub-rotinas — indexadas por palavras de código matemáticas, tal como os números nas anedotas dos presos — e programar o computador para copiar e compilar quaisquer sub-rotinas que fossem necessárias para uma dada tarefa? Por outras palavras: por que não programar o computador para se programar a si próprio?

Esta ideia de um «compilador» é porventura a inovação de *software* mais importante na história da computação. Para voltar à metáfora da preparação do chá, os programadores já não tinham de codificar instruções como «levante a mão esquerda, agarre a torneira, rode a torneira» sempre que queriam encher a chaleira, ferver a água, etc. A própria máquina iria então compilar todas as sub-rotinas certas para um programa de preparação de chá. Programas que antes levavam uma semana a escrever passariam a demorar cinco minutos. Além disso, cada sub-rotina era previamente depurada e sabia-se que funcionava. Era impossível contar a anedota de modo errado.

De início, quando Hopper explicou a ideia aos seus chefes, eles acharam que ela estava louca. Os computadores apenas podem fazer contas, disseram-lhe. Eles não poderiam escrever os seus próprios programas, apenas um humano poderia fazê-lo. Foi por

esta altura que Grace adotou a sua observação favorita, a qual viria a repetir muitas vezes ao longo dos anos: «A frase mais perigosa na linguagem é "Sempre fizemos isto dessa maneira".» É claro que os chefes de Hopper estavam errados, como ela viria a provar.[16]

Hopper não ficou por aí. A sua visão acerca dos compiladores convenceu-a de uma verdade essencial: que o futuro dos computadores dependia de tornar fácil o diálogo com eles. Isso exigia muito mais do que apenas substituir o velhinho manual binário de canalização por palavras de código matemáticas. A notação matemática era adequada para um cientista ou para um investigador da Marinha, mas a maioria dos compradores potenciais de computadores, notou Hopper, «não reconheceria um cosseno mesmo que se cruzasse com ele na rua»[17]. Os profissionais das empresas não precisavam de uma linguagem para dizerem a um computador como calcular a trajetória de um foguete, precisavam de uma linguagem para trabalhar com bases de dados: contas, preços, vendas, salários, horas trabalhadas e por aí fora. Havia apenas uma linguagem comum para conversar acerca de dados de uma maneira que abrangesse os diferentes setores do comércio: o inglês. Para Hopper, a conclusão era óbvia: os computadores tinham de ser programados para trabalharem com *inputs* em inglês.

Mais uma vez, os chefes dela acharam que era descabido e nem valia a pena tentar, e em 1953 recusaram o financiamento para a sua proposta. É claro que não podemos fazer com que um computador perceba inglês, disseram-lhe. A ideia é completamente absurda. Temos de programar os computadores usando símbolos e matemática. Sempre o fizemos dessa maneira.[18]

Não foi a primeira vez que ela enfrentou a cultura machista da computação, e não seria a última. Mas quando o vento a derrubou da canoa, Grace simplesmente nadou para terra. Ela continuou a trabalhar na sua ideia por conta própria e em janeiro de 1955 tinha um protótipo funcional. Falando para uma sala cheia de quadros superiores de empresas, ela demonstrou que o seu «compilador de processamento de dados» podia fazer com que o *UNIVAC* compreendesse um programa em língua inglesa cujas primeiras linhas tinham este aspeto.[19]

> Input Inventário Ficheiro A; Preço Ficheiro B;
> Comparar Produto #A com Produto #B.
> Se Maior, Ir para Operação 10;
> Se Igual, Ir para Operação 5.

E assim por diante. Hopper tinha programado o computador para traduzir estas frases nos bastidores, permitindo que os utilizadores se focassem no que conheciam (o fluxo de dados) em vez de no que não conheciam (os pormenores da matemática).

Então, Hopper cometeu um erro. Para enfatizar que o computador estava apenas a aplicar regras para fazer corresponder frases a padrões de *bits*, ela demonstrou que as frases equivalentes em francês podiam produzir o mesmo programa: «Lisez-paquet A; Si Finde Donnes Allez en Operation 14.» Isto deixou os seus chefes em polvorosa, como Grace descreveu: «Isso "atingiu a ventoinha"! Era absolutamente óbvio que um computador americano respeitável, construído em Filadélfia, na Pensilvânia, não poderia de maneira alguma entender francês.»[20] Com umas poucas frases

suspeitamente não americanas, ela fez o projeto retroceder quatro meses.

Hopper viria a prevalecer e recebeu fundos empresariais para desenvolver o seu compilador de processamento de dados, chamado *FLOW-MATIC*. Um estudo-piloto mostrou que o *FLOW-MATIC* estava a permitir que os clientes realizassem as mesmas tarefas que o velho método de «matemática e símbolos», mas num quarto do tempo. Com os seus clientes tão entusiasmados com a nova abordagem de Hopper, os chefes não tiveram outra opção senão ceder. Eles tiveram a sorte de Grace ter sido tão determinada.

Com isso, a Revolução da Linguagem de Programação tinha começado. De meados de 1950 em diante, praticamente toda a gente que falou com um computador fê-lo usando o modelo que Grace Hopper concebeu. As linguagens de máquina continuaram a ser importantes, mas tornaram-se o domínio de uns quantos especialistas altamente qualificados. Todas as outras pessoas usaram linguagens de programação de alto nível cujos conjuntos de instruções eram muito mais parecidos com «encha a chaleira» do que com «agarre a torneira, rode a torneira». Ao instituir este modelo, Hopper teve um impacto enorme naquilo que acreditamos ser a mais importante tendência na história humana desde 1945: a disseminação da tecnologia digital a todas as áreas da vida.

3. De Grace a Alexa: A Revolução na Linguagem Natural

Isto leva-nos até cerca de 1960. Então, como é que subsequentemente chegámos ao ponto em que podemos ter qualquer bem de consumo entregue à nossa porta, bastando para isso verbalizar em voz alta o pedido a um computador?

Para começar, vamos resumir o modelo descendente baseado em regras que Grace Hopper concebeu na década de 1950 para a interação linguística entre humanos e máquinas.

- As pessoas dizem às máquinas o que devem fazer usando uma linguagem de programação, com uma gramática bastante restrita e um vocabulário reduzido de palavras inglesas.

- As máquinas interpretam estes comandos na sua própria linguagem usando vastos livros de regras de tradução pré-programadas, operando nos bastidores.

- Tanto as regras de programação para os humanos como as de tradução para as máquinas têm de ser definidas a partir do zero, uma a uma, por programadores humanos.

Entre a década de 1950 e a de 1970, os peritos tentaram fazer com que as máquinas entendessem a linguagem natural utilizando esta mesma abordagem descendente: 1) colocando limitações aos

utilizadores humanos, restringindo a gramática e o vocabulário que eles podem usar, e 2) programando as máquinas enchendo-as com regras de tradução: sintaxe, pronúncia, escolha de palavras... basicamente, todas as normas que enquanto crianças aprendemos sem esforço, a par de todas as regras gramaticais que aprendemos com a Sra. Thistlebum na escola primária.

Esta filosofia baseada em regras tinha dado ótimos resultados com as linguagens de programação. Mas nunca funcionou muito bem com as línguas naturais.

Um excelente exemplo de como correu mal é o reconhecimento do discurso pelo computador. Os primeiros sistemas de reconhecimento de discurso eram essencialmente brinquedos. Na Exposição Mundial de 1962, por exemplo, a IBM apresentou uma máquina que podia reconhecer palavras faladas em inglês — exatamente 16 palavras, e apenas se pronunciadas com uma clareza dramática. Em 1970 houve uma falsa alvorada, sob a forma de um programa chamado *Harpy*, criado por investigadores em Carnegie Mellon. O *Harpy* reconhecia exatamente 1011 palavras, mais ou menos as mesmas que uma criança pequena. Foi construído segundo os princípios de Grace Hopper: gramática e vocabulário limitados para os humanos e um conjunto de normas diabolicamente complicadas para a máquina, para transcrever a fala em texto. A equipa de cinco programadores do *Harpy* demorou dois anos a codificar estas regras — regras para a acústica, a fonética, a estrutura frásica, o limite das palavras, etc. Em condições laboratoriais altamente idealizadas, o sistema até alcançou um nível de correção na transcrição de 70 por cento. Isto provocou bastante

entusiasmo entre os investigadores de IA. O *Harpy* parecia indicar que, com melhores regras e computadores mais rápidos, o desempenho de nível humano estaria mesmo ao virar da esquina.[21]

No entanto, estas melhorias esperadas no reconhecimento do discurso nunca se materializaram. Em testes posteriores que envolveram condições reais, a correção do *Harpy* ao nível das palavras baixou para 37 por cento. Passados cinco anos, o governo dos Estados Unidos retirou o financiamento ao projeto. Hoje, os sistemas puros baseados em regras para processamento de linguagem natural tornaram-se bastante raros. Em última instância, nunca foram capazes de ultrapassar três problemas básicos: acumulação de regras, robustez e ambiguidade.

3.1 Problema 1: Acumulação de Regras

Primeiro, é mesmo muito difícil escrever todas as regras para línguas naturais. Elas são demasiadas, consideravelmente mais do que qualquer linguagem de programação. Embora talvez não o saiba, mas na verdade pode aprender bastante *Python* num dia. Já não conseguirá aprender bastante coreano no mesmo tempo. Parte do problema é que todas as regras têm exceções — ou, como o famoso linguista Edward Sapir disse, «todas as gramáticas vazam». Em inglês, por exemplo:

- *Adjectives come before nouns, but don't tell that to the attorney general's heir apparent* [Os adjetivos vêm antes dos nomes, mas não diga isso ao herdeiro provável do procurador-geral].

- *«I before e, except after c», except for weirdly presciente scientists who drink protein shakes with caffeine* [«I antes do *e*, exceto depois do *c*», exceto para estranhos cientistas prescientes que bebem batidos de proteína com cafeína].

- *Two positives don't make a negative? Yeah, right* [Dois positivos não fazem um negativo? Sim, certo].

E assim por diante. As exceções criam problemas, pois as máquinas são consistentemente tirânicas em relação às regras. A única forma de lidar com isto é escrever uma regra para cada exceção.

Isto parece suficientemente penoso, mas o problema das «demasiadas regras» vai ainda mais longe. Para muitos aspetos da linguagem, nós simplesmente não sabemos quais são as regras. Considere o que os linguistas referem como o problema da «segmentação do discurso». Tente ler a frase seguinte em voz alta: «O boletim meteorológico diz que vai chover amanhã.» Irá percecionar isto como uma série descontínua de palavras: «meteorológico», «chover», «amanhã», etc. Mas esta descontinuidade é uma ilusão; apenas os robots da ficção científica... falam... com... pausas. Aquilo que realmente está a ouvir é um fluxo contínuo de sons, sem intervalos acústicos evidentes entre as palavras. Perceber onde uma palavra acaba e começa outra é um problema realmente difícil. Os linguistas descobriram todo o tipo de regras auditivas ocultas das quais dependemos para fazer isto, com nomes como «fonotática» e «variação alofónica». Mas os linguistas também sabem que não

encontraram as regras todas, porque aquelas que eles *encontraram* não conseguem explicar como somos bons na segmentação do discurso.

O problema é óbvio: se nem sequer conseguimos identificar as regras todas, então certamente não as conseguimos ensinar a um computador.

3.2 Problema 2: Robustez

O segundo problema com regras descendentes é que elas geralmente falham quando em contato com o mundo real. Em termos simples, elas não são robustas.

Considere, por exemplo, o problema de distinguir entre discurso e ruído de fundo. O seu cérebro é *incrível* nesta tarefa — habitualmente consegue compreender os seus amigos mesmo num barbarulhento, apesar da algazarra. Os neurocientistas não entendem a cem por cento como é que consegue fazê-lo, e é por isso que o barulho de fundo é uma queixa constante das pessoas que usam aparelhos auditivos.

Outra fonte de variação irrelevante é aquilo a que de forma pouco caridosa poderá chamar «erros». Imagine o que poderia acontecer se descartasse uma regra de inglês em cada frase que escrevesse ou dissesse hoje. Poderia receber alguns olhares estranhos, mas as pessoas percebê-lo-iam facilmente. Mesmo que usasse fragmentos de frases. Mesmo se «como Yoda você falasse».

Mesmo se deixasse os seus modificadores pendentes, impassíveis face ao desprezo da Sra. Thistlebum para com o seu abuso grosseiro de particípios. A compreensão da linguagem pelas pessoas é bastante robusta perante este tipo de variações, mas essa robustez é difícil de replicar usando regras descendentes.

Outra questão importante é a pronúncia. Peça a alguém de Derry, em New Hampshire, para dizer «caramelo». Depois peça a alguém de Derry, na Irlanda do Norte. As respostas soarão bastante dissemelhantes; não terão as mesmas vogais, nem sequer o mesmo número de sílabas. Poderá responder: OK, vamos simplesmente fazer duas regras para «caramelo». Mas isso são apenas os irlandeses e os ianques. Não se esqueça dos texanos e dos londrinos e dos californianos, e... bem, consegue ver o problema? É novamente a acumulação de regras. Imagine tentar programar um computador com um conjunto de regras que mapeiem de forma robusta todas estas diferentes pronúncias — «care-a-mell», «crrr-mul», e todos os tons de caramelo pelo meio — da mesma palavra subjacente. Parabéns, acaba de resolver o reconhecimento do discurso para «caramelo». Já só faltam 171 475 palavras no *Oxford English Dictionary*. A seguir pode começar com a gíria, e depois talvez com o mandarim.

3.3 Problema 3: Ambiguidade

Finalmente, é difícil conceber regras que lidem bem com a ambiguidade — e as línguas estão cheias de ambiguidade. Os exemplos mais óbvios envolvem homófonas: nós/noz, coser/cozer, conselho/concelho, etc. Depois há o que os linguistas chamam

«ambiguidade sintática», ou frases que podem ser interpretadas de múltiplas maneiras. As manchetes dos jornais são frequentemente culpadas disto, no entanto a maioria das vezes conseguimos percebê-las:

- ESQUERDA BRITÂNICA ÀS ARANHAS NAS FALKLANDS é sobre um partido trabalhista indeciso, não sobre uma expedição entomológica.

- INCLUA AS SUAS CRIANÇAS QUANDO ESTIVER A FAZER BOLOS NO FORNO não é uma sugestão de receita.

- RESTAURANTE LONDRINO TEM ESQUILOS AO JANTAR é sobre uma praga e não acerca de um novo menu.

A questão essencial aqui é que, no que toca à linguagem, nós somos um incrível motor de inferência probabilística, programado durante milénios de evolução para troçar da ambiguidade. Nós mal notamos nas vgs dsprcds em mensagens de texto. Avançamos através de analogias como uma faca quente na manteiga. Sabemos que «temos de fazer uma pausa» significa algo diferente no local de trabalho ou numa discussão com o nosso parceiro. Usamos informação contextual para interpretar o que alguém pretende dizer, mesmo que exista outra frase plausível que soe exatamente da mesma maneira:

- Presidente muda de direção

- Presidente muda de ereção

Este tipo de ambiguidade é eliminado por qualquer linguagem concebida para computadores, precisamente por causa dos problemas que criam para o artífice das regras. Contudo, nós parecemos lidar com elas quase sem problemas. Como?

4. 1980-2010: O Crescimento do Processamento Estatístico da Linguagem Natural

Os filósofos gostam de distinguir entre dois tipos de conhecimento: saber como *versus* saber que. «Saber como» significa conhecimento intuitivo e prático. Por exemplo: sabemos como caminhar e como andar de bicicleta sem esforço consciente. «Saber que», por outro lado, significa conhecimento factual, clássico. Ou seja: ao lermos uma página aleatória na *Wikipedia* começada com a letra *N*, podemos aprender *que Nike* é uma marca de sapatos, ou que *Napoleão* invadiu a Rússia em 1812 mas achou-a bastante fria.

Para as pessoas, a linguagem falada assemelha-se ao exemplo máximo de «saber como». O milagre cognitivo é como tudo parece fácil — como podemos falar de maneira inteligível, e descodificar as ondas sonoras ambíguas que emanam da boca das outras pessoas, sem sequer pensarmos nisso. Para o fazer extraímos ilações automaticamente usando todo o tipo de informação lateral: a nossa experiência do mundo, a nossa compreensão implícita do que as

outras pessoas provavelmente estarão a pensar e muitas pistas auditivas subtis.

Como vimos, os especialistas em PLN passam bastante tempo a tentar que os computadores entendam linguagem natural fornecendo-lhes inúmeras regras explícitas. Estas regras devem supostamente mimetizar o conhecimento que as crianças adquirem naturalmente quando aprendem a falar. Contudo, mesmo no seu melhor, estas abordagens baseadas em regras cometem erros crassos. Elas não conseguiriam igualar a competência linguística de uma criança de 5 anos, e muito menos a de um adulto.

Após três décadas de experiências, para os especialistas em processamento de linguagem natural tornou-se evidente que era necessário uma nova abordagem. Esta abordagem teria de ser flexível em vez de rígida. Probabilística em vez de determinística. Ascendente, baseada em dados reais, em vez de descendente, alicerçada numa profusão de regras. Acima de tudo teria de lidar com a maneira como as pessoas verdadeiramente falam, em vez de como um gramático pensa que elas deveriam falar.

Assim, na década de 1980, os investigadores tentaram algo diferente: levantaram os braços, livraram-se das regras e disseram «vamos usar apenas dados». Inventaram algoritmos novos que funcionavam com uma premissa muito diferente: o conhecimento linguístico humano pode ser demasiado complicado para aplicar a engenharia inversa, mas este conhecimento *tem* de facto uma sombra estatística clara, discernível na maneira como falamos e escrevemos. Por exemplo: se «weather report» (boletim

meteorológico) faz mais sentido do que «whether report» (se relatório), então um conjunto considerável de frases reais deve conter muitos mais exemplos da primeira do que da última. Efetivamente, é exatamente isso que observamos. Utilizámos uma ferramenta online chamada *Google Ngram Viewer*[*], que nos permite verificar a popularidade de qualquer palavra ou frase curta em todos os livros publicados em inglês. Descobrimos que entre 1950 e 2000, cerca de 150 em cada mil milhões de frases de duas palavras em livros publicados eram «weather report» (0,0000155797%). Isto é cerca de 250 vezes mais comum do que «whether report» (0,0000000652%), que é principalmente usada como um mau trocadilho ou um exemplo de ambiguidade fonética.

Da década de 1980 em diante, os investigadores de PLN começaram a reconhecer o valor desta informação puramente estatística. Antes, eles tinham estado a criar manualmente regras capazes de *descrever como* é que uma dada tarefa linguística deveria ser realizada. Agora, estes especialistas começaram a treinar modelos estatísticos capazes de prever que uma pessoa iria realizar uma tarefa de uma dada maneira. Enquanto área de estudo, o PLN mudou o seu foco da compreensão para o mimetismo — de *saber como*, para *saber que*.

Estes modelos novos precisavam de bastantes dados. A máquina era alimentada com tantos exemplos quantos se conseguissem encontrar de como os humanos usam a linguagem, e programava-se a máquina para usar as regras da probabilidade para encontrar padrões nesses exemplos. A linguagem tornou-se um problema de regras preditivas baseadas em pares de *input/output*, similar aos

problemas resolvidos por Henrietta Leavitt ou por aquele agricultor no Japão que usa a aprendizagem profunda para classificar pepinos:

- Para reconhecimento de discurso emparelhamos uma gravação de voz (input = «takusawpekenwal'mosu») com a transcrição correta (output = «tacos ao pequeno-almoço»).

- Para traduzir de inglês para russo emparelhamos uma palavra ou frase inglesa («reset») com a tradução correta em russo («perezagruska»).

- Para prever sentimento emparelhamos uma frase («Que manhã tão agradável passada na fila da DGV») com uma anotação humana (:().

E assim por diante. Em cada caso, a máquina tem de usar os dados para aprender uma regra preditiva que mapeia corretamente *inputs a outputs*.

Na década de 1980, o *software* de reconhecimento de discurso baseado neste princípio começou a chegar ao mercado. Estes sistemas conseguiam reconhecer alguns milhares de palavras, mas apenas se falássemos... como... um... robot. A década de 1990 e o início da de 2000 viram a chegada de modelos mais ricos com um desempenho cada vez melhor e que nos deixavam falar a um ritmo natural. Mas aqui o grande obstáculo era a disponibilidade de dados. Talvez se recorde do problema de «sobreajustamento» que

descrevemos no Capítulo 2, no qual um modelo complicado simplesmente memoriza o ruído aleatório num pequeno conjunto de dados, sem aprender o padrão subjacente. Aqui ocorreu o mesmo problema. Os investigadores de PLN simplesmente não possuíam a quantidade de dados necessária para construir modelos que fossem suficientemente complicados para descrever a linguagem humana sem sobreajustar os poucos dados de que dispunham. Em resultado, na década de 2000, o reconhecimento de discurso voltou a estabilizar em cerca de 75-80 por cento de nível de correção de palavras. Durante quase uma década, o avanço foi desanimadoramente lento — e não apenas para o reconhecimento do discurso como também para outras tarefas no processamento da linguagem natural que foram prejudicadas pela falta de dados, da tradução automática à análise de sentimentos.

4.1 Pós-2010: A Revolução da Linguagem Natural

Por volta de 2010, tudo começou a mudar — de início lentamente, depois a um ritmo alarmante. O que motivou esta mudança foi uma instilação maciça de dados.

Em tempos Jorge Luis Borges escreveu uma história intitulada *A Biblioteca de Babel*, acerca de uma biblioteca cujos livros continham todas as obras possíveis de prosa, isto é, todos os ordenamentos possíveis das letras do alfabeto e os principais sinais de pontuação. A maioria dos livros desta biblioteca eram um perfeito disparate, como se produzidos por um macaco numa máquina de escrever, mas algures, num certo livro, poderíamos encontrar todas as frases possíveis — todas as histórias de amor, todas as aventuras, todas

as obras geniais alguma vez escritas ou que possam vir a ser escritas.

A nossa Biblioteca de Babel da vida real chama-se Internet, e apesar de ainda não estarmos propriamente em terreno de Borges, estamos a aproximar-nos. Pense nos imensos conjuntos de frases inglesas faladas e escritas que se encontram nos servidores das maiores empresas tecnológicas do mundo. Pense numa biblioteca com todos os livros, revistas, jornais, diários, canções, filmes e peças alguma vez escritas. Agora pense numa escala ainda maior. Todas as páginas na Internet. Todos os e-mails na história. Todas as buscas no *Google* ou avaliações de produtos, todas as mensagens de texto já enviadas, todos os *chats* no *Slack* ou no *Skype*, todas as publicações no *Facebook* ou no *Twitter* (agora X), todos os comentários no *YouTube* ou no *Instagram*. Em termos de volume, este conjunto de frases faz a Biblioteca do Congresso parecer uma livraria itinerante de terceira categoria. E por volta de 2010, as melhores mentes em IA tinham finalmente desenvolvido as ferramentas certas para usar todos os dados no seu pleno efeito.

Alguns destes dados foram simplesmente parar ao colo das grandes empresas tecnológicas, mas estas empresas também se esforçaram para recolher mais dados. Um exemplo foi o *Google 411*, que foi lançado em 2007. Talvez se lembre de uma altura em que as pessoas ligavam o 411* para procurar um número de telefone de uma empresa local, ao custo de cerca de um dólar por chamada. O *Google 411* permitia-lhe fazer a mesma coisa de forma gratuita, ao ligar 1-800-GOOG-411. Este serviço foi útil numa época anterior aos omnipresentes smartphones — e também uma ótima

maneira para a Google construir uma enorme base de dados de consultas de voz que ajudariam a treinar os seus modelos estatísticos para reconhecimento de discurso. O sistema encerrou discretamente em 2010, presumivelmente porque a Google já tinha todos os dados de que necessitava.

É claro que, desde 2007, tem havido uma imensa quantidade de codificação ao estilo de Grace Hopper para transformar todos os dados em boas regras preditivas. Assim, mais de uma década depois, qual é o resultado? Vamos tentar uma experiência simples. Abra um e-mail em branco no seu telefone e tente ditar uma frase de teste: «The weather report calls for rain, whether or not the reigning queen has an umbrella.»[†] Se for um nativo de língua inglesa e o seu telefone funcionar com iOS ou Android, ele quase de certeza escreverá a frase corretamente, sem confundir *weather/whether* ou *rain/reign*.

Este é um pequeno exemplo do que os dados lhe proporcionam. O software sabe que «*whether*» e «*reign*» são estatisticamente mais prováveis em certos contextos, enquanto «*weather*» e «*rain*» são mais plausíveis noutros. Isto não acontece porque o seu telefone de alguma maneira compreende o significado das palavras. Aqui não há qualquer significado envolvido, apenas um vasto conjunto de probabilidades específicas para um contexto[‡], para basicamente todas as palavras e frases inglesas alguma vez proferidas na Internet. Quando os dados acústicos são ambíguos, o seu telemóvel resolve o impasse usando estas probabilidades. Ainda que algumas coisas continuem a criar-lhe problemas, pelo menos em 2018, o software melhora a cada dia que passa.

Outros sistemas de PLN também progrediram rapidamente, e pela mesma razão. Considere por exemplo a tradução automática. Durante muitos anos houve a moda dos memes na Internet dedicados aos erros cometidos pelo *Google Tradutor*, e pode encontrar centenas deles espalhados pela Web. Por exemplo: em 2011, um engraçadinho descobriu que, ao traduzir de inglês para vietnamita «Irá o Justin Bieber alguma vez chegar à puberdade.»[22] transformava-se em «O Justin Bieber nunca chegará à puberdade.» Este tipo de erro sintático era uma anomalia clássica dos algoritmos de tradução automática mais antigos. Eles acertavam na maioria das palavras, mas frequentemente confundiam a ordem na língua-alvo, produzindo algo que era errado ou disparatado.

À medida que ficam disponíveis mais dados de treino, e à medida que os modelos estatísticos de linguagem se tornam melhores, estes erros sintáticos grosseiros têm-se tornado muito menos frequentes.[23] Além disso, toda essa quantidade de dados fez com que as regras de tradução explícitas fossem muito menos importantes. Por exemplo: ninguém disse explicitamente ao *Google Tradutor* que o inglês usa a ordem sujeito-verbo-objeto, como em «programadores adoram café», mas que o japonês usa a ordem sujeito-objeto-verbo, como em «programadores café adoram». O algoritmo simplesmente aprendeu a sintaxe através da estatística — a partir de milhões de frases de dados de treino, apresentadas lado a lado em inglês e japonês.

Atualmente, a grande maioria dos sistemas de PLN bem-sucedidos representa o exemplo máximo do segundo tipo de conhecimento dos filósofos: *saber que*, em vez de *saber como*.[24]

Para o software, tudo são apenas factos. Mas estes factos habitualmente são suficientes porque os próprios conjuntos de dados são tão grandes e os algoritmos tão sofisticados.

5. Como as Palavras se Tornam Números

Vamos então virar-nos para a questão dos algoritmos. Imagine que tem uma gigantesca base de dados de frases — uma Biblioteca de Babel para mais de cem línguas naturais, de inglês e chinês a farsi. Como é que se constrói um sistema de IA para uma tarefa relacionada com linguagem e se o põe efetivamente a funcionar?

Como pode imaginar, há imensos detalhes que não podemos abordar aqui porque eles simplesmente são demasiado complexos. Esses detalhes são a razão por que uma companhia como a *Google* tem 70 mil empregados, um pequeno exército de doutorados, e mais computadores do que alguma vez viu na sua vida. Mas podemos dar-lhe uma explicação de alto nível de algo realmente importante, chamado «vetor-palavra». Especificamente, vamos explicar-lhe o famoso modelo *word2vec* da Google, que providencia uma descrição numérica («vetor») para cada palavra em inglês. Se compreender o *word2vec*, então estará a compreender uma das ideias mais importantes da IA da última década. Até os sistemas que não usam este algoritmo diretamente, empregam ainda assim a mesma abordagem subjacente.

O *word2vec* responde a uma questão simples: como é que transformamos palavras em números, de modo a que palavras com

um significado semelhante tenham números semelhantes? Isto pode parecer estranho ou mesmo impossível. Em que medida é que o significado de palavras como «torradeira» ou «coragem», ou uma frase como «*Toronto maple leafs*» [folhas de ácer de Toronto] pode ser descrito em números? Mas nós defendemos que de modo algum é tão difícil quanto parece. De facto, as crianças fazem-no a toda a hora.

5.1 A Matemática das 20 Perguntas

Há uma cena de *Um Conto de Natal* que decorre na sala de estar de Fred, o sobrinho de Ebenezer Scrooge. Fred convidara o seu tio rico e avarento para o jantar de Natal, e acabou por ouvir que «todo o idiota que anda por aí com "Feliz Natal" nos lábios devia ser fervido com o seu próprio pudim, e enterrado com um galho de azevinho a trespassar-lhe o coração». Mas, entretanto, Scrooge foi visitado por espíritos que lhe mostraram o erro da sua conduta avarenta, e o terceiro espírito, o Fantasma do Natal Presente, levara Scrooge até à casa de Fred no Dia de Natal. Eles observam, sem serem vistos, enquanto Fred e a sua família jogam ao Sim e Não. O sobrinho de Scrooge, Fred, tem de pensar em algo, e todas as outras pessoas na sala têm de adivinhar o que é, apenas fazendo perguntas de resposta *sim* ou *não*:

> A rajada de perguntas a que foi submetido permitiu concluir que [Fred] estava a pensar num animal, um animal vivo, um animal bastante desagradável, um animal bravio, um animal que por vezes rosnava e grunhia, e por vezes falava, e vivia em Londres, e percorria as ruas, e não fazia parte de nenhum espetáculo, e não pertencia a ninguém, e não vivia num zoológico.

O que poderia ser? As crianças agora estão com ataques de riso. Elas dão vários palpites, e aprendem que o Fred não estava a pensar num urso ou num cavalo, nem num tigre ou num burro. Então, finalmente, a cunhada de Fred descobre a resposta: «Fred, eu sei o que é! Eu sei o que é!... É o teu Tio Scro-o-o-o-oge!» E era mesmo.

As crianças americanas cresceram a chamar a este jogo 20 Perguntas, e embora possa não parecer, é um jogo muito matemático. Na verdade, o 20 Perguntas mostra-nos como transformar palavras em números, tal como o faz a IA. Peguemos na palavra «Scrooge» do jogo na sala de estar do Fred. A sua representação numérica tem este aspeto:

	Animal	Agradável	Rosna ou grunhe	Fala	Vive em Londres	É um urso
Scrooge	1	0	1	1	1	0

A isto chama-se um «vetor-palavra».* Especificamente, é um vetor «binário» ou 0/1: 1 significa sim, 0 significa não. Palavras ou frases diferentes, como «Pequeno Tim» ou «Urso Paddington», produziriam respostas diferentes às mesmas questões, pelo que teriam vetores-palavras diferentes. Se dispusermos todos estes vetores numa matriz, onde cada linha é uma palavra e cada coluna é uma questão, obtemos algo parecido com isto:

	Animal	Agradável	Rosna ou grunhe	Fala	Vive em Londres	É um urso
Scrooge	1	0	1	1	1	0
Rafael Nadal	1	1	1	1	0	0

Pequeno Tim	1	1	0	1	1	0
Urso Paddington	1	1	0	1	1	1
Árvore de Natal de Trafalgar Square	0	1	0	0	1	0

5.2 Como a IA Joga às 20 Perguntas

De modo que é realmente muito fácil transformar palavras em números, usando o jogo de 20 Perguntas. Agora vamos mudar as regras de três maneiras — tanto para torná-lo num jogo mais parecido com o que um sistema de IA tem de jogar, como para fazer com que as palavras-vetor que obtivermos estejam tão carregadas de significado quanto possível.

Primeira alteração das regras: o resultado não é meramente ganhar ou perder. Em vez disso somos pontuados por «proximidade semântica», ou por quão perto ficámos do verdadeiro significado da palavra. Não vamos entrar em detalhes sobre o que «perto» significa. Em IA há uma resposta matemática pormenorizada, mas imagine apenas a pessoa mais honesta que conhece a servir de juiz. Por exemplo: suponha que a resposta é «urso»:

- Se o seu palpite final for um urso, receberá 100 pontos.

- Se o seu palpite for um cão ou um glutão, talvez receba 90 pontos. Ficou bastante perto, em termos filogenéticos.

- Se o seu palpite for um mosquito, talvez receba 50 pontos. Pelo menos adivinhou que era um animal.

- Se o seu palpite for xarope para a tosse, receberá dois pontos. Está completamente enganado, mas a tosse pode de algum modo remoto lembrar-lhe o rosnido de um urso.

Esta maneira de pontuar equipara-se aos requisitos reais de conceção da maioria dos sistemas de PLN. Por exemplo: se traduzirmos JFK a dizer «Ich bin ein Berliner» como «Eu sou um alemão» estará errado, mas muito mais perto do que se traduzir como «Eu sou um *cronut*».*

A segunda alteração das regras é que em vez de apenas um sim ou um não, cada resposta é um número entre 0 (completamente não) e 1 (completamente sim). Tomemos como exemplo «Será um urso?»

- Para um urso verdadeiro e vivo, responderia com um 1.

- Para um urso como o Paddington, talvez respondesse com um 0,9. Ele continua a ser um urso, mas não é um urso propriamente dito.

- Para o Scrooge, a sua resposta poderia ser 0,65. Ele não é verdadeiramente um urso, mas de facto tem tendências parecidas com as de um urso. (Em *Um Conto de Natal*, a família de Fred até se queixou «de que a resposta a "Será

um urso?" deveria ter sido "Sim", na medida em que uma resposta na negativa seria suficiente para desviar os seus pensamentos do Sr. Scrooge».)

- Para Rafael Nadal, a sua resposta poderia ser 0,2. Na televisão ele parece ser uma ótima pessoa não ursina, mas ele está vivo e de facto grunhe bastante.

Com esta alteração às regras ficamos com vetores-palavra feitos de números contínuos — de 0 *até* 1 para cada questão, em vez de 0 ou 1; cinzento, em vez de preto e branco.

Agora a maior alteração de regras: temos de colocar as mesmas questões em todos os jogos. Isto certamente acabaria com o divertimento na sua próxima festa de jogos de salão vitorianos, dado que reduz todos os jogos das 20 Perguntas ao Vamos Jogar aos Censos, onde preencheria o mesmo deprimente formulário de inquérito. Apesar disso, deixe as suas dúvidas de lado. Imagine tentar obter questões que são ao mesmo tempo suficientemente abrangentes e ricas para distinguir todas as palavras possíveis, de «Scrooge» a «*Scrabble*», de «basáltico» a «basquetebol», de «eritrócito» a «epistemologia».

Não é fácil, certo? Mas é realmente assim que funcionam os modelos de linguagem natural em IA. Uma ressalva: vamos precisar de muito mais que 20 perguntas, dado que já não podemos adaptar as nossas últimas respostas às primeiras perguntas. Assim, em IA, jogamos antes às 300 Perguntas.

Em relação a quais as perguntas a fazer, não iremos abordar isso diretamente. Ao invés, falemos acerca do processo. A sua primeira ideia poderá ser fazer isto através de um conselho: colocamos algumas pessoas inteligentes numa sala e dizemos-lhes que não podem sair até encontrarem 300 questões que em conjunto proporcionem uma codificação única para cada palavra ou frase em inglês. Isto poderá ser uma interessante experiência sociológica. Mas se pensa que provavelmente resultará, então a sua fé em conselhos é muito maior do que a nossa. No mínimo, demoraria imenso tempo. E nós queremos dizer aos nossos telefones para encomendarem uma pizza agora, não quando um qualquer conselho tomar uma decisão.

Isto deixa-nos com apenas uma maneira certa de escrever perguntas: deixar um algoritmo escolher.

Que tipo de perguntas poderá um algoritmo colocar? Perguntas sobre significado não resultarão, porque as máquinas não compreendem o significado. Mas as máquinas compreendem *estatísticas de colocalização de palavras*, ou seja, que palavras tendem a surgir com que outras palavras em frases reais escritas pelos humanos. Estas estatísticas são substitutos surpreendentemente bons para o significado. Eis o exemplo de uma pergunta dentro desta linha: «Pegue em todas as frases que incluam “batatas fritas”, “ketchup” ou “pão”. Nessas frases, esta palavra também aparece frequentemente? » É certo que este é o tipo de pergunta que uma rede neural poderia atribuir ao Ross num novo episódio de *Friends*. Crucialmente, no entanto, é também uma

pergunta que uma máquina pode não só colocar como responder, porque não requer compreender, apenas contar.

É claro que, como só temos direito a 300 perguntas, esta questão em particular é demasiado fechada. Mas a premissa básica — colocar questões acerca de estatística de colocalização de palavras — é válida. Ainda que estejamos a omitir aqui imensos detalhes, basicamente isto é o que o *word2vec* faz. Através de tentativa e erro, ele aprende 300 conjuntos adequados de palavras-sonda (análogas a «pão» e «ketchup» no exemplo acima).[25] Em seguida ele joga às 300 Perguntas repetidas vezes, aprendendo um vetor-palavra por cada palavra ou frase em inglês, com base nas suas estatísticas de colocalização com as palavras-sonda.

Os vetores-palavra resultantes podem ser dispostos numa matriz, uma palavra por linha, tal como fizemos para «Scrooge» e «Pequeno Tim» algumas páginas atrás. Esta matriz é enorme, com 300 colunas e milhares de linhas. Na próxima página, damos-lhe um subconjunto de quatro colunas e 40 linhas, para que tenha uma noção do tipo de questões que o algoritmo aprende a colocar.

Por exemplo: na primeira coluna, vemos palavras como «*router*», «*píxeis*» e «*firewall*», cujas respostas à pergunta 1 são todas muito próximas de 1. Fica claro que o algoritmo aprendeu a colocar uma questão na mesma linha de «Tende a palavra a aparecer em frases com palavras de computador?»* (Lembre-se que, de acordo com a nossa regra revista, 1 significa «completamente sim» e 0 significa «completamente não».) Do mesmo modo, na terceira coluna, vemos palavras como «assado»,

«fumado», «carne» e «grelhado», todas com respostas perto de 1. O algoritmo terá aprendido a formular uma pergunta acerca de cozinhar com fogo. Ele também aprendeu a colocar questões acerca de universidades e direito criminal — e, em outras colunas não apresentadas, questões acerca de animais, política, desporto, saúde e centenas de outros tópicos.

5.3 Pôr os Vetores-Palavra a Trabalhar

Esta abordagem oferece uma imagem surpreendentemente matizada da linguagem. Os investigadores de IA até têm um truque de magia favorito para exibir a riqueza dos seus vetores-palavra: responder a perguntas de analogia do tipo das dos testes de aptidão escolar (SAT — *Scholastic Aptitude Test*), apenas através de adição e subtração. Por exemplo: considere a analogia «homem está para rei, como mulher está para _____?» Como é que podemos formular esta pergunta como um problema de matemática, adequado para ser descrito em termos de aritmética em vetores-palavra?

COMO UM ALGORITMO JOGA ÀS 20 PERGUNTAS

	Pergunta 1: «computadores»	Pergunta 2: «universidades»	Pergunta 3: «cozinhar»	Pergunta 4: «direito»
Nvidia	1	0,045	0,156	0,083
servidores	0,999	0,944	0,214	0,184
username	0,999	0,468	0,842	0,963
ethernet	0,999	0,587	0,617	0,072
interface	0,999	0,355	0,831	0,032
router	0,998	0,697	0,986	0,911
ecrãs	0,998	0,693	0,111	0,174
porta	0,997	0,646	0,583	0,184
píxeis	0,997	0,253	0,017	0,21
firewall	0,995	0,729	0,957	0,636
universitário	0,089	0,999	0,107	0,627
docentes	0,365	0,999	0,114	0,944
bolsas	0,063	0,999	0,291	0,398
candidatos	0,153	0,999	0,22	0,77
faculdades	0,206	0,997	0,132	0,514
catedráticos	0,216	0,997	0,035	0,688
comissão	0,32	0,996	0,912	0,824
departamentos	0,42	0,994	0,502	0,77
residencial	0,145	0,993	0,569	0,801
publicações	0,173	0,993	0,534	0,938
assado	0,778	0	1	0,767
fumado	0,596	0,012	1	0,799
cervejas	0,815	0,043	1	0,613
churrasco	0,182	0,077	1	0,039
milho	0,827	0,044	1	0,122
carne	0,471	0,015	0,999	0,699
malagueta	0,403	0,002	0,999	0,425
pimentos	0,398	0	0,999	0,572
grelhado	0,531	0,001	0,999	0,46
sabor	0,281	0,026	0,997	0,248
fiança	0,221	0,63	0,923	1
custódia	0,509	0,536	0,943	1
prender	0,149	0,444	0,839	1
acusações	0,002	0,157	0,57	1
sanções	0,44	0,105	0,413	0,999
posse	0,123	0,304	0,73	0,999
ilegal	0,045	0,406	0,478	0,999
condenação	0,015	0,121	0,928	0,999
processo	0,175	0,147	0,735	0,999
polícia	0,275	0,305	0,882	0,997

Eis como. Pegue no vetor para a palavra «rei» e *subtraia* o vetor para a palavra «homem». (Podemos adicionar e subtrair vetores tal

como com os números, porque os vetores são números.) Intuitivamente, ao subtrair «homem» de «rei», despojámos a palavra «rei» do seu componente de género masculino, resultando num novo vetor que presumivelmente representa um conceito de realeza de género neutro. A este novo vetor *adicione* agora o vetor para a palavra «mulher», reintroduzindo assim uma componente de género — desta feita uma componente feminina. Por outras palavras, *pegue na palavra «rei» e torne-a feminina*, a qual expressamos em termos de aritmética como «rei — homem + mulher». A resposta do *word2avec* é certeira: se de facto efetuarmos a aritmética, obtemos o vetor-palavra para «rainha».

Outros tipos de analogias funcionam da mesma maneira, usando a adição e a subtração de vetores.

- Capitais do mundo: Londres — Inglaterra + Itália. Resposta do *word2vec*: «Roma».

- Tempos de palavras: capturou — captura + vai. Resposta do *word2vec*: «Foi.»

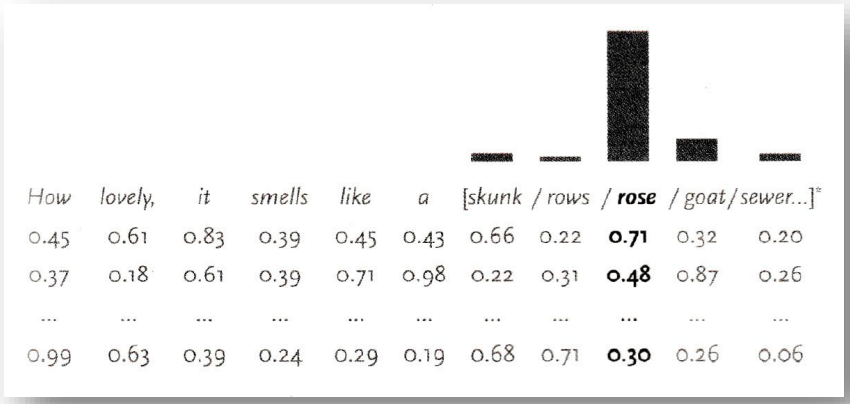
- Que equipas de hóquei jogam em que cidades: Canadianas — Montreal + Toronto. Resposta do *word2avec*: «Maple Leafs.»

Com efeito, o *word2vec* aprendeu a realizar o teste de aptidão escolar verbal usando apenas competências do teste de aptidão escolar de matemática. O modelo subjacente nada sabe, em

absoluto, acerca de monarquia, género, gramática, hóquei, ou qualquer coisa real acerca do mundo. Ele só sabe acerca das propriedades estatísticas do uso de palavras recolhidas através de dados de treino, em conjunto com as regras de probabilidades.[26]

Isto pode ser um truque de magia, ou uma diversão deliciosamente *geek* para codificadores, mas também realça uma questão importante: assim que se transformam palavras em vetores podemos realizar matemática com eles. Isto é essencial para construir sistemas de IA para a linguagem. Os computadores não apreendem palavras, mas compreendem matemática.

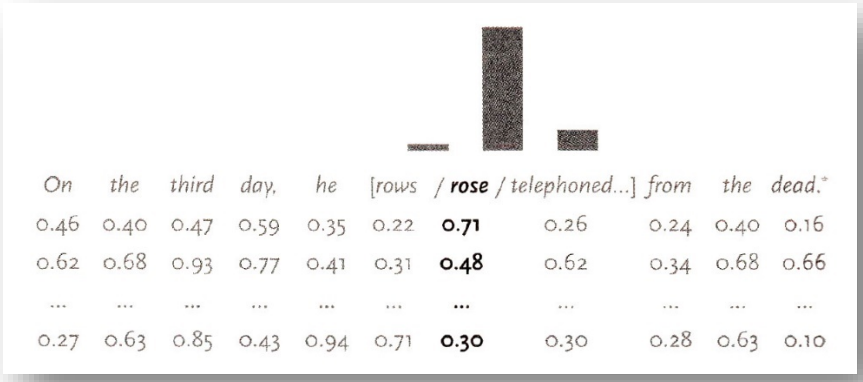
Considere o software de reconhecimento de discurso, do tipo que faz funcionar a *Alexa* ou o *Google Voice*. Os vetores-palavra são aqui uma tremenda ajuda, porque codificam o contexto de uma frase numa linguagem matemática com a qual um computador consegue trabalhar. Entre outras coisas, isto fornece uma informação crucial de desempate para homófonas, como «*rows*» (filas) *versus* «*rose*» (rosa). Ainda que estas palavras soem da mesma maneira, elas têm vetores-palavra diferentes — respostas diferentes no grande jogo das 20 Perguntas da IA — e normalmente um destes vetores encaixará melhor nos vetores das palavras circundantes numa dada frase. Evidentemente, é muito complicado saber o que significa «encaixará melhor» e não vale a pena aprofundar o assunto aqui. Resume-se a um cálculo sofisticado que envolve aritmética de vetores, que gera probabilidades que desfazem o empate quando a informação acústica é ambígua, como aqui:



How	*lovely,*	*it*	*smells*	*like*	*a*	[*skunk*	*/ rows*	*/* ***rose***	*/ goat/ sewer...*]*
0.45	0.61	0.83	0.39	0.45	0.43	0.66	0.22	**0.71**	0.32	0.20
0.37	0.18	0.61	0.39	0.71	0.98	0.22	0.31	**0.48**	0.87	0.26
...	...	...	...	...	...	...	...	...	...	...
0.99	0.63	0.39	0.24	0.29	0.19	0.68	0.71	**0.30**	0.26	0.06

How lovely, it smells like a* [Skunk / rows / rose / goat / sewer…]

Ou aqui:



On	*the*	*third*	*day,*	*he*	[*rows*	*/* ***rose*** */ telephoned...*]	*from*	*the*	*dead.**
0.46	0.40	0.47	0.59	0.35	0.22	**0.71**	0.26	0.24	0.40	0.16
0.62	0.68	0.93	0.77	0.41	0.31	**0.48**	0.62	0.34	0.68	0.66
...	...	...	...	...	...	...	...	...	...	...
0.27	0.63	0.85	0.43	0.94	0.71	**0.30**	0.30	0.28	0.63	0.10

On the third day, he [rows / rose / telephoned…] from the dead.*

Ou aqui:

He	planted	100	[ears	/ rows	/ rose...]	of	corn.†
0.35	0.75	0.37	0.19	**0.22**	0.71	0.83	0.45
0.41	0.75	0.23	0.22	**0.31**	0.48	0.75	0.15
...	...	...	...	...	...	...	...
0.94	0.25	0.80	0.96	**0.71**	0.30	0.04	0.99

He planted 100* [ears / rows / rose...] *of corn.†

Os vetores-palavra proporcionam uma descrição matemática clara de algo que para qualquer ouvinte humano parece simples: às vezes é uma palavra que encaixa melhor, outras vezes é outra. E com algumas modificações importantes específicas para tarefas, a mesma matemática básica potencia a tradução automática, os *chatbots*, os sistemas de busca por voz... e até redes neurais que podem escrever sobre basebol.

6. Humanos e Máquinas Conversando Juntos

Esperamos que agora compreenda algumas das ideias-chave que nos conduziram até ao momento presente, quando as máquinas claramente passaram um ponto crítico na sua capacidade para usar linguagem. Este melhoramento foi em parte impulsionado por computadores rápidos e algoritmos inteligentes, como as redes neurais e o *word2avec*. Mais fundamentalmente foi impulsionado pelos dados — os nossos dados. As máquinas falantes não nos proporcionam uma janela para um qualquer novo tipo de génio

linguístico. Elas apenas seguram um espelho virado para o nosso próprio génio.

O que nos reserva o futuro? É impossível saber, claro, mas destacam-se algumas tendências prováveis.

Primeiro: os modelos de linguagem tornar-se-ão personalizados; as máquinas à sua volta irão adaptar-se à sua maneira de falar, tal como se adaptaram às suas preferências cinematográficas. Em resultado, elas tornar-se-ão muito melhores a compreendê-lo. Considere, por exemplo, a trajetória histórica do *iPhone*. Para usar o *iPhone 6* tinha de ensiná-lo a reconhecer a impressão digital do seu polegar. Para usar o *iPhone X* tinha de ensiná-lo a reconhecer a sua cara. Não é difícil imaginar um *iPhone* futuro em que primeiro terá de ler-lhe uma história para dormir, a fim de lhe ensinar a sua voz.

Segundo: políticas adequadas e uma regulamentação ponderada serão altamente importantes se quisermos colher os benefícios destas ferramentas de PLN, sem que elas sejam utilizadas de maneiras destrutivas. Um algoritmo que consegue escrever um episódio de *Friends* parece engraçado, ainda que algo inútil. Esse mesmo algoritmo parecerá muito mais pernicioso quando alguém o puder programar para inundar a Internet com notícias falsas em tempo de eleições. Ainda que em última análise estejamos otimistas, não somos peritos em política e não sabemos a resposta certa para este tipo de problema. Mas sabemos, contudo, que os próprios problemas têm de fazer parte do debate. No desenvolvimento histórico de todas as novas tecnologias, do fogo à

manipulação genética, sempre houve um momento em que «andar depressa e partir coisas» deixou de ser uma posição moralmente sustentável para um adulto. Com os computadores e a linguagem, esse momento chegou.

Mas, mesmo ao reconhecermos o potencial aspeto negativo, também não nos esqueçamos do aspeto positivo. Se acha que hoje temos algoritmos de PLN mais inteligentes e maiores conjuntos de dados, ainda não viu nada. Pense nas centenas de milhões de pessoas por aí a ditarem e-mails para os seus telefones, a usarem o *Google Tradutor* ou a falarem com um *bot* no *Facebook* ou no *WeChat*. Cada uma destas interações origina modelos mais ricos e melhores desempenhos, dado que estas máquinas estão apenas a andar às cavalitas do rasto de dados que deixamos para trás. À medida que melhoram — e ainda há *bastantes* melhorias a fazer — esperamos que estas máquinas se tornem ferramentas de rotina em todas as profissões e em todos os segmentos da vida.

Se quiser, chame-nos otimistas tecnológicos despreocupados, mas achamos que isto é maravilhoso. O trabalho árduo não é algo que devamos desejar a alguém — e pelo menos no mundo rico, muito do trabalho árduo é do tipo eletrónico, do tipo que nos torna sedentários e doentes. Por que é que os médicos devem todos os dias passar horas a fazer inserção de dados? Por que tem uma pessoa cega de ser forçada a usar um teclado em braille? Por que precisa um advogado de gastar centenas de horas/pessoa a vasculhar milhões de páginas de documentos? Por que devem os funcionários de escritórios passar décadas das suas vidas a digitar e-mails? Por que é que a UE deve gastar centenas de milhões de

euros por ano a traduzir tudo para 23 línguas oficiais? Por que tem de confiar nos seus polegares meio burros para dizerem ao seu telefone o que ele deve fazer? E já viu aqueles computadores que eles usam nos balcões de *check-in* das companhias aéreas?

Nós não estamos à espera que os seres humanos suguem o lixo do chão ou filtrem o *spam* da sua caixa de correio; nós temos aspiradores e algoritmos para isso. Então, por que é que digitar tem de ser diferente?

7. Pós-Escrito

Grace Hopper poderá ter sido a primeira pessoa a falar com um computador em inglês, mas ela decididamente não parou aí.

Após o seu trabalho pioneiro com o *UNIVAC*, Hopper teve uma longa carreira na indústria privada e na Reserva da Marinha, antes de se reformar em 1966, com 60 anos. Depois, em 1967, ela foi inesperadamente convocada para o serviço ativo pela Marinha, onde serviu durante mais 19 anos — muito para lá da idade de reforma obrigatória, devido a uma autorização especial do Congresso. Ela ajudou a convencer o Departamento de Defesa a atualizar a sua infraestrutura informática, e até se tornou uma das primeiras mulheres na história da Marinha a receber uma patente de oficial-general. Na sua promoção a comodoro, em 1983, as suas palavras enquanto apertava a mão ao presidente Ronald Reagan foram: «Sou mais velha do que o senhor.» Ela finalmente reformou-se definitivamente em 1986, com 79 anos.

Hopper morreu em 1992, mas o seu legado mantém-se vivo. Ao longo dos anos ela viu o seu nome ser atribuído a muitas coisas, incluindo um navio da Marinha, um supercomputador *Cray*, e ao Grace Hopper College, na Universidade de Yale. Ela foi homenageada postumamente com um *Google Doodle*[*] em dezembro de 2013, e com a Medalha Presidencial da Liberdade em novembro de 2016. Não há dúvida de que o seu bisavô, o almirante, teria ficado orgulhoso. Através dos seus esforços para por meio da linguagem aproximar um pouco mais as pessoas e as máquinas, Grace Hopper desempenhou um papel grandioso na invenção do mundo moderno.

---

[*] A rima original é entre as palavras «bellow» (abaixo) e «low blow» (golpe baixo). [N. T.]

[†] O original alude à reprodução incorreta de duas expressões idiomáticas. «For all intents and purposes» (para todos os efeitos) reproduzida como «for all intensive purposes» (para todos os propósitos intensivos) e «at his beck and call» (às suas ordens) reproduzida como «at his beckon call» (ao seu aceno chamativo). [N. T.]

[‡] Interpretações incorretas de «for the thirty-first time» (pela 31.ª vez) em vez de «for the very first time» (pela 1.ª vez) e de «said the worst attorney» (disse o pior advogado) em vez de «since the world's been turning» (desde que o mundo anda à volta). [N. T.]

[1] A. Bartoli, A. De Lorenzo, E. Medvet e F. Tarlao, «Your Paper Has Been Accepted, Rejected, or Whatever: Automatic Generation of Scientific Paper Reviews», em *Availability, Reliability, and Security in Information Systems*, ed. F. Buccafurri *et al.* (Nova Iorque: Springer Berlin Heidelberg, 2016), 19-28.

[2] Andy Pandy, 18 de janeiro de 2016, https://twitter.com/Pandy/status/689209034143084547 (Inexistente).

[3] Um ponto técnico menor: para os nossos objetivos não é necessário distinguir entre um «compilador» (para uma linguagem como C++ ou Java) e um «intérprete» (para uma linguagem como Python). Neste caso utilizámos «compilador» para abarcar ambos os conceitos.

[*] No original, «drop your trousers» e «drop off your trousers» [N. T.]

[*] Pode encontrar facilmente no YouTube a entrevista dela com David Letterman.

[4] Kathleen Broome Williams, *Grace Hopper: Admiral of the Cyber Sea* (Annapolis, Md.: Naval Institute Press, 2004), 1.

[5] *Ibid*., 2.

[6] *Ibid*., 11.

[7] *Ibid*., 18-20.

[8] *Ibid*., 22.

[9] *Ibid*., 26.

[10] *Ibid*., 29.

[11] *Ibid*., 27-28.

[12] *Ibid*., 82.

[13] Kurt W. Beyer, *Grace Hopper and the Invention of the Information Age* (Cambridge, Mass.: MIT Press, 2009), 53.

[14] Douglas Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid* (Nova Iorque: Vintage, 1980), 290.

[*] Os números decimais comuns 8 e 25 são expressos em octais como 10 e 31, respetivamente.

[15] Williams, *Grace Hopper*, 70.

[16] *Ibid*., 8o.

[17] *Ibid*., 85.

[18] *Ibid*., 86.

[19] *Ibid*.

[20] Veja *ibid*., 87. Referência original em Richard L. Wexelblat, ed., *History of Programming Languages I* (Nova Iorque: ACM, 1978), 17.

[21] Bruce T. Lowerre, «The HARPY Speech Recognition System», tese de doutoramento, Departamento de Ciências da Computação, Universidade de Carnegie Mellon, 1976.

[*] https://books.google.com/ngrams. Um «n-grama» é um termo *geek* de linguística para uma frase que contém n itens, como palavras ou símbolos. Por exemplo, «weather report» é um 2-grama, dado que contém duas palavras.

[*] Em Portugal, ligava-se para o 118.

[†] Em português: o boletim meteorológico prevê chuva, quer a rainha reinante tenha ou não um guarda-chuva. [N. T.]

[‡] Aqui «contexto» apenas significa «as outras palavras da frase».

[22] «10 Inexplicable Google Translate Fails», https://www.searchenginepeople.com/blog/10-google-translate-fails.html.

[23] Para detalhes sobre o método, bem como avaliações extensivas da precisão, veja Yonghui Wu *et al.*, «Google's Neural Machine Translation System: Bridging the Gap Between Human and Machine Translation», 8 de outubro de 2016, https://arxiv.org/abs/1609.08144.

[24] Peter Norvig, «On Chomsky and the Two Cultures of Statistical Learning», http://norvig.com/chomsky.html.

[*] Em matemática, um vetor é simplesmente um conjunto de números, todos associados à mesma coisa.

[*] *Cronut* — espécie de bolo que resulta da mistura de croissant com dónute. [N. T.]

[25] Se quer mesmo entrar em aspetos técnicos, estas «palavras-sonda» são na realidade chamadas «vetores de contexto». Veja Tomas Mikolov *et a*l., «Distributed Representations of Words and Phrases and Their Compositionality», *Advances in Neural Information Processing Systems 26* (NIPS, 2013), https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.

[*] É claro que o algoritmo não sabe que a pergunta é «acerca de computadores». Ele simplesmente sabe que a pergunta é acerca de estatística de colocalização que envolvem outras palavras que nós, enquanto humanos, podemos subsequentemente interpretar como sendo acerca de computadores.

[26] Tomas Mikolov, Wen-tau Yih e Geoffrey Zweig, «Linguistic Regularities in Continuous Space Word Representations», em *Proceedings of NAACL-HLT*, 2013 (Stroudsburg, PA: Association for Computational Linguistics, 2013), 746-51.

[*] Em português: Que bonito, cheira como uma [doninha / filas / ***rosa*** / cabra / esgoto ...][N. T.]

[*] Em português: No terceiro dia, ele [filas / ***rosa*** / telefonou...] dentre os mortos, sendo que a palavra «rose» surge aqui como o pretérito indicativo do verbo «to rise», que significa erguer. [N. T.]

[†] Em português: Ele plantou 100 [orelhas / *filas* / rosa...] de milho. [N. T.]

[*] Um *Doodle* do *Google* é uma alteração especial e temporária do logótipo nas páginas iniciais do *Google* com o objetivo de comemorar algo ou alguém. [N.T.]

CAPÍTULO 5

O Génio na Casa da Moeda

Monitorização em tempo real, do desporto ao policiamento e à fraude financeira: o que o maior erro matemático de Isaac Newton pode ensinar-lhe acerca da procura de anomalias em conjuntos maciços de dados.

Se é um fã da NFL, e se vive junto a uma estreita faixa de terreno entre o Connecticut médio e o Maine, então provavelmente olha para os New England Patriots — a mais bem-sucedida equipa de futebol americano dos últimos 15 anos — com um misto de irritação e desconfiança. Primeiro há todas as vitórias, que garantidamente irritam os fãs das outras 31 equipas da NFL. Depois há o Bill Belichick, o austero treinador principal dos Patriots, cujo capuz e ar carrancudo fazem-no ter uma parecença incrível com o imperador maléfico de *Star Wars*. Mas mesmo que não seja um fã de futebol americano, pode ainda assim ser um adepto do *fair play* — caso em que os Patriots poderão continuar a irritá-lo devido a todos os episódios altamente publicitados de falcatruas, tal como espiar as práticas das outras equipas ou (alegadamente) esvaziar bolas de futebol americano de modo a ganhar vantagem em tempo frio.

Mas será que os Patriots seriam até capazes de fazer batota no atirar de moeda ao ar que antecede o jogo? Acredite ou não, muitas pessoas acham que sim: durante um período de 25 jogos na NFL que se estenderam pelas temporadas de 2014 e 2015, os Patriots venceram 19 dos 25 lançamentos de moeda ao ar, para uma percentagem de vitórias questionavelmente elevada de 76 por cento. Um comentador televisivo salientou, quando foi alertado para este mais recente «escândalo»: «Isto simplesmente mostra que ou Deus ou o Diabo é um fã dos Patriots, e Deus não é certamente.»[1]

Antes de invocarmos a religião ou *a Força* para elucidar esta anomalia, vamos primeiro considerar a explicação inocente: pura sorte. Cada vez que escolhe cara ou coroa quando atira uma moeda ao ar tem 50 por cento de hipóteses de ganhar. Mas também há que ter em conta a variabilidade. Se atirar uma moeda ao ar repetidas vezes, facilmente terá períodos de sorte em que acerta mais vezes do que erra, apenas por mero acaso. Será plausível que os Patriots apenas tenham tido um período de sorte de 25 jogos?

Aqui o raciocínio complica-se por um facto: se os Patriots tivessem tido uma série suspeita de 25 jogos de cara ou coroa em qualquer ocasião desde 2007, quando rebentou o seu primeiro escândalo de fraude, as pessoas teriam notado. Portanto, não podemos simplesmente selecionar esta série de 25 jogos em particular e perguntar quão atípica ela é isoladamente. A questão certa a colocar é: até que ponto é improvável que os Patriots tenham ganhado pelo menos 19 lançamentos de moeda ao ar em qualquer período de 25 jogos ao longo das últimas 11 épocas?

Para responder a esta questão usámos um computador para simular o lançamento de moeda ao ar — mais de 17 milhões de vezes. Especificamente, escrevemos um programa para simular um jogo de cara ou coroa imparcial em todos os 176 jogos dos Patriots na temporada regular, nas épocas de 2007 a 2017 da NFL.* Repetimos esta simulação 100 mil vezes, verificando de cada vez se os Patriots tinham um período contínuo de 25 jogos com pelo menos 19 vitórias no lançamento da moeda ao ar. Pode ver nove destas 100 mil simulações na Figura 13. Na maior parte do tempo, o recorde dos Patriots no jogo de cara ou coroa ronda as 12 ou 13 vitórias, que é o que se esperaria. Mas há muita variabilidade. Às vezes os Patriots entram numa maré de sorte — como nas simulações 3, 6 e 9, em que tiveram séries de 25 jogos com pelo menos 19 vitórias no lançamento da moeda. Na simulação 3, até tiveram uma série com 22 vitórias.

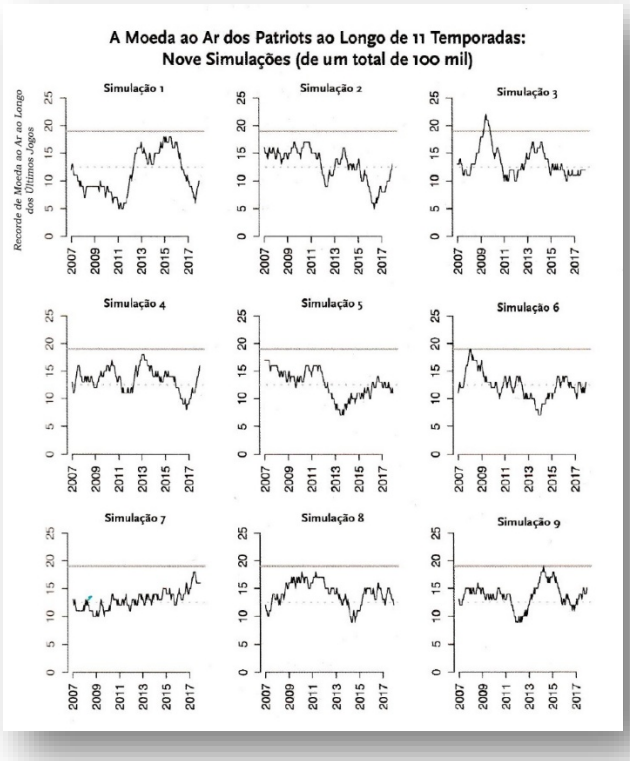


Figura 13. Cada painel mostra o recorde de moeda ao ar dos Patriots ao longo de 11 temporadas simuladas, entre 2007 e 2017. O eixo vertical indica o número de vitórias ao longo de cada série consecutiva de 25 lançamentos de moeda. A barra cinzenta horizontal está em 19 vitórias, enquanto a linha tracejada é a média esperada a longo prazo de 12,5 vitórias.

No geral, os Patriots atingiram o patamar de 19 vitórias em 23 por cento das nossas 100 mil simulações, o que não é suficientemente pequeno para descartar a sorte como explicação*. Assim, embora não possamos falar sobre as câmaras de espionagem ou as bolas de futebol vazias, não há qualquer prova que indique que os Patriots estavam a fazer batota no jogo de cara ou coroa. Eles simplesmente tiveram sorte durante uma série de 25 jogos em 2014-15.

1.1 A Importância da Variabilidade

O recorde de moeda ao ar dos New England Patriots ilustra um princípio importante. Para decidir se algo é uma anomalia tem de saber duas coisas: 1) o que esperar em média, e 2) os limites normais de variabilidade em torno da média. Se não compreende a variabilidade — por exemplo, como a percentagem de vitórias dos Patriots nos 25 jogos de moeda ao ar varia em torno da média a longo prazo de 50 por cento — então nunca será capaz de distinguir uma verdadeira anomalia de flutuações aleatórias inocentes.

Isto leva-nos diretamente ao nosso assunto principal, que é sobre monitorização em tempo real para detetar anomalias.† Em IA, significa analisar um fluxo de pontos de dados e identificar os que não correspondem ao padrão típico. Isto pode salvar vidas e dinheiro e levar a novas perspetivas acerca dos seus dados:

- Os bancos utilizam software que procura padrões de despesas anómalas, para detetar se o seu cartão de crédito foi roubado.

- As grandes empresas monitorizam as suas redes em busca de tráfego anómalo, de modo a detetarem falhas na cibersegurança.

- Os analistas de dados nas «cidades inteligentes» procuram concentrações anómalas de crimes numa determinada área, para melhorarem as estratégias de policiamento.

- Os investigadores procuram anomalias em dados de pedidos de reembolsos médicos, para detetarem fraudes em seguros de saúde.

- As equipas desportivas monitorizam os dados provenientes dos aparelhos utilizáveis, procurando anomalias que sugiram risco de lesão.

Em todas estas aplicações da IA, e em milhares de outras, detetar anomalias tem tanto que ver com compreender a variabilidade dos dados como com compreender o que é normal.

1.2 O Mais Antigo Sistema de Deteção de Anomalias da História: Uma Lição Acerca do que Não Fazer

Para ilustrar a importância deste princípio para a IA iremos começar com um exemplo de alguém que não o compreende de todo — e não uma pessoa qualquer, mas um dos maiores génios de todos os tempos. Especificamente, levá-lo-emos até à Inglaterra de 1696, quando o sistema de deteção de anomalias mais duradouro de sempre estava já em pleno funcionamento. Este sistema, chamado o «Julgamento da Píxide», foi concebido para prevenir fraudes na Casa da Moeda, onde o dinheiro inglês era produzido. É fascinante precisamente porque fracassou: não detetou as mais variadas anomalias, ao longo de séculos, desempenhando um importante, ainda que subvalorizado, papel numa crise económica que causou revolta e sofrimento generalizados.

E em 1696, a pessoa no centro de tudo isto era Isaac Newton. Sim, *esse* Isaac Newton — o inventor do cálculo, o explicador da gravidade, e o homem imortalizado no famoso verso de Alexander Pope: «A natureza e as leis da natureza estavam imersas em trevas; Deus disse "Haja Newton" e tudo se iluminou.» Em 1696, Newton era uma estrela de rock científica com 54 anos e uma cátedra vitalícia em Cambridge. Ele não era obrigado a ensinar e podia trabalhar no que quisesse — física, alquimia, malabarismo com maçãs, qualquer coisa. Contudo, em 1696 ele abandonou a vida de professor, mudou-se para Londres e aceitou uma sinecura oferecida pelos seus poderosos amigos do governo: supervisor da Casa da Moeda.

Em geral, Newton teve um desempenho notável no seu novo trabalho — exceto num ponto crucial, onde errou seriamente, ao não compreender um princípio estatístico importante que esteve mesmo à frente do seu nariz durante quase cinco anos. Atualmente, esse princípio está no centro de todos os sistemas de monitorização em tempo real alimentados por IA: em Silicon Valley, em cidades inteligentes, nos gabinetes de análise de todas as equipas desportivas e nos departamentos de prevenção de fraude de todos os bancos. Assim, se quer compreender qualquer uma destas coisas, então tem de abarcar três elementos principais na história de Newton na Casa da Moeda:

1. Uma crise na economia inglesa no final do século XVII, na qual a Casa da Moeda desempenhou um papel subtil, mas fundamental.

2. A Grande Recunhagem de 1696, uma medida drástica na política monetária inglesa concebida para estancar a crise, a qual Newton tinha de salvar do desastre.

3. A importância da variabilidade estatística na deteção de anomalias — a matéria do pior erro matemático que Newton alguma vez cometeu.

2. A Segunda Carreira de Isaac Newton

Newton chegou à Casa da Moeda em 1696, no meio de uma verdadeira crise monetária, que ameaçava deixar a economia

inglesa paralisada. Para compreender a experiência de Newton na Casa da Moeda tem de compreender as raízes dessa crise.

O problema era este: em 1696, o dinheiro inglês já estava a desaparecer de circulação há pelo menos três décadas. Naquela época, a Inglaterra utilizava o padrão-prata, em que o peso e o conteúdo de prata das moedas determinavam o seu valor. Devido à Guerra dos Nove Anos, contudo, a procura por prata no continente subiu em flecha, até ao ponto em que as moedas inglesas valiam menos como dinheiro em Inglaterra do que como metal precioso na Europa. Em resposta, as pessoas em Inglaterra fizeram exatamente o que seria de esperar: levaram moedas para França ou para os Países Baixos, fundiram-nas, trocaram a prata pura por ouro, depois de novo em Inglaterra venderam esse ouro por mais moedas de prata — e pronto, estavam um pouco mais ricas do que quando tinham começado. À medida que toda aquela prata escorria lentamente para o outro lado do Canal, a Inglaterra começou, literalmente, a ficar sem dinheiro.[2]

Também havia uma segunda razão para a prata estar a desaparecer, que amplificou maciçamente o efeito da primeira: o cerceio de moedas, o flagelo da massa monetária inglesa durante o século XVII. Para cercear uma moeda procurava-se um pedaço de prata saliente ao longo da borda. Depois bastava cortar essa saliência e limar a parte danificada, deixando-a lisa. Não se obtinha muito de uma qualquer moeda individual, mas se se cerceasse moedas suficientes podia obter-se uma boa pilha de prata. O cerceio era punível com enforcamento desde o reinado da rainha Isabel I. No entanto, isso não parecia deter os cerceadores. Numa

investigação parlamentar em 1690, três ourives recolheram cada um 100 libras de moedas em circulação; juntas, estas moedas deviam ter pesado 1200 onças *troy*, mas na verdade pesaram 624. De modo que o cerceio era óbvio no conjunto, mas a menos que se apanhasse alguém em flagrante era um crime praticamente impossível de provar.[3]

2.1 Os Cerceadores de Moedas Adoravam a Variabilidade

O cerceio tornou-se muito mais fácil pelo facto de, antes de 1662, todas as moedas inglesas serem batidas à mão — isto é, por um ourives numa bigorna, a martelar uma bolha de prata fundida transformando-a num disco. A principal caraterística a destacar acerca destas moedas batidas à mão é a sua variabilidade, tanto na forma como no peso. Esta variabilidade era essencial para o êxito do cerceio de moedas. A variabilidade da forma garantia que as moedas tinham à partida pequenos altos para cercear. A variabilidade de peso garantia que um cerceador podia sempre encontrar uma moeda ligeiramente mais pesada — e que, uma vez cerceada, poderia ainda gastá-la com um encolher de ombros inocente, como se ela sempre tivesse sido um pouco mais leve.

Em 1662, o parlamento finalmente deu atenção ao problema do cerceio, dando à Casa da Moeda os fundos de que necessitava para começar a produzir moedas com uma máquina. O objetivo era simples: tirar o sustento aos cerceadores ao eliminar a variabilidade na forma e no peso das moedas.

Vale a pena descrever em pormenor este novo processo de produção mecânico, para lhe dar uma ideia da operação que mais tarde Newton viria a herdar. Começava com prata fundida a 1000 ºC, a fervilhar em caldeirões gigantescos. A prata era vertida para moldes de lingotes e, à medida que estes lingotes arrefeciam, eram achatados em feitio de folha por uma enorme engenhoca em forma de rolo da massa, movida por uma equipa de quatro cavalos. Uma segunda máquina do tipo de cortador de bolos recortava discos na placa de prata. Estes discos alimentavam uma prensa de rosca, que fazia moedas em branco. Então uma quarta máquina, muito perigosa, estampava uma imagem em cada uma das faces da moeda. Um homem introduzia uma moeda em branco numa pequena câmara no meio da montagem. Em seguida, outros quatro homens puxavam cordas para virar uma roda 180 graus, fazendo com que uma prensa enorme estampasse na moeda uma imagem profunda e indelével da face do rei. Outra meia volta da roda fazia a prensa voltar a subir — e, durante esse intervalo, o primeiro homem tinha de tirar a moeda estampada para fora da câmara e inserir uma nova moeda em branco. Os quatro homens a girar a roda ficavam exaustos ao fim de 15 minutos, e o que inseria as moedas em branco trabalhava sob o medo permanente de perder os dedos.[4]

Por último veio a máquina que gravava dois elementos especiais nos bordos. Primeiro havia um padrão serrilhado à volta da circunferência; a isto chama-se «borda serrilhada», que ainda se pode encontrar em muitas moedas modernas, incluindo a americana de 25 cêntimos e a britânica de duas libras. Em segundo lugar havia uma inscrição em latim: *Decus et Tutamen*, uma frase

da Eneida de Virgílio, que significa «Um ornamento e uma salvaguarda». Tal como o latim sugeria, era uma salvaguarda contra o cerceio de moedas — dado que seria difícil cortar um bocado, ainda que pequeno, da moeda sem que o dano fosse óbvio.*

Poderá pensar que a introdução destas máquinas de serrilhar moedas teria ajudado a resolver o problema que Inglaterra tinha com o dinheiro, ao eliminar a variabilidade da massa monetária. Na verdade, após 1662, o problema piorou. Tal aconteceu porque as moedas mais antigas batidas à mão ainda circulavam e os comerciantes continuavam a aceitá-las pelo valor facial. No final do século XVII, a Inglaterra efetivamente tinha duas moedas paralelas. Havia as moedas batidas à mão anteriores a 1662, até então depreciadas por não poderem ser derretidas com lucro, e havia as serrilhadas à máquina, que não podiam ser cerceadas com facilidade. Assim, as moedas serrilhadas à máquina foram fundidas e levadas clandestinamente para a Europa, enquanto apenas circulavam as moedas produzidas à mão.[5]

Os economistas referem-se a este fenómeno, em que o dinheiro mau expulsa o bom, como a Lei de Gresham. Mas este é um exemplo típico de economistas a batizarem coisas óbvias com o nome de outros economistas. Na verdade, esta «lei» foi observada 17 séculos antes de Gresham, por Aristófanes, na sua peça *As Rãs*: «As únicas moedas cunhadas corretamente e testadas na Grécia e no estrangeiro não nos servem mais para nada; ao invés, usamos esse cobre barato, gravado ontem ou anteontem em péssima cunhagem.» Isto é senso comum. Se estivermos a pagar produtos

alimentares e tivermos a opção de entregar uma moeda de prata ou uma de cobre, ficaremos com a de prata para nós. O dono da loja raciocinará da mesma maneira quando nos der o troco. Apenas as moedas más irão circular.

Foi exatamente isso que aconteceu com as moedas inglesas. Em resultado, quando Newton ingressou na Casa da Moeda em 1696, a vida comercial quotidiana de Inglaterra estava quase em ruínas. Um bocado de humor de cadafalso tornou-se viral: ainda que os impostos tenham aumentado durante o reinado do rei Jaime, pelo menos tinha havido dinheiro para os pagar. Muitas pessoas não tinham qualquer dinheiro — e aqueles que tinham guardavam-no em vez de o gastar, pensando corretamente que valeria mais amanhã. Em resultado, escreveu o historiador Charles Macaulay, «todo o comércio, toda a indústria, foram afetados como que por uma paralisia. O flagelo sentia-se diariamente e a toda a hora em praticamente todos os lugares e por quase todas as classes». Uma testemunha contemporânea explicou a um amigo que «nenhum negócio é feito senão por confiança. Os nossos inquilinos não conseguem pagar a renda. Os nossos comerciantes de milho nada podem pagar por aquilo que tiveram, e não negociarão mais, pelo que tudo isso está num impasse.»[6] Como Macaulay resumiu, «toda a miséria que fora infligida à nação inglesa num quarto de século por maus reis, maus ministros, maus parlamentares e maus juízes» perdem importância quando comparados com «a miséria causada num único ano por más coroas e maus xelins».[7]

3. O Julgamento da Píxide

Até agora aprendemos três factos importantes acerca da situação do dinheiro inglês em 1696.

1. Todas aquelas moedas de prata produzidas à mão antes de 1662 ainda eram um enorme problema. Elas levavam as pessoas a discussões diárias acerca do seu valor, e a sua depreciação conduziu as moedas de peso completo e produzidas por máquinas para fora de circulação e em última instância para fora de Inglaterra.

2. Estas moedas batidas à mão tinham-se desvalorizado devido ao cerceio.

3. A chave para um cerceio bem-sucedido era a variabilidade: disparidades na forma e no peso das moedas que podiam ser exploradas pelos criminosos, que implacavelmente cortavam as moedas ligeiramente pesadas demais e passavam-nas como moedas ligeiramente mais leves.

Assim sendo, uma pergunta crucial para entender a crise na economia inglesa é: por que se permitiu que as moedas fossem à partida tão variadas?

Alguma desta variabilidade era inevitável, particularmente para as moedas batidas à mão. A lei inglesa tinha até evoluído para

reconhecer isto, ao estabelecer um limite legal de variabilidade aceitável e ao colocar em prática medidas de segurança para garantir que estes limites eram respeitados. Mas o sistema fracassou — e para explicar por que estava isto a acontecer chegamos ao Julgamento da Píxide.

O Julgamento da Píxide* é um sistema de deteção de anomalias que tem funcionado continuamente desde a década de 1150. Embora os pormenores tenham mudado com o tempo, o seu objetivo sempre tem sido mais ou menos o mesmo: verificar se a Casa da Moeda tem sido fraudulenta, inepta ou ambas. Os oficiais da Casa da Moeda, por exemplo, podiam trapacear fazendo as moedas sistematicamente demasiado leves, permitindo-lhes meter ao bolso a prata restante. Ou podiam simplesmente ser ineptos no controlo de qualidade, fazendo algumas moedas excessivamente pesadas e outras demasiado leves. Se isto acontecesse, então as acidentalmente mais pesadas podiam ser fundidas para um mercador perspicaz obter lucro.

O Julgamento da Píxide foi concebido para prevenir tal artimanha. Por cada 27,2 quilos de prata batidos pela Casa da Moeda colocava-se de lado uma única moeda. Quando se tinham acumulado alguns milhares de moedas, geralmente ao fim de poucos anos, eram testadas em busca de anomalias por um júri de prateiros, para garantir que iam ao encontro dos padrões legais de peso e pureza metálica.[8]

Mas lembre-se da lição do exemplo dos New England Patriots: quando procuramos anomalias, temos de ter em conta a

variabilidade. Mesmo que não tenha havido qualquer intrujice, não se podia esperar que as moedas pesassem *exatamente* o que era esperado, devido às inevitáveis imperfeições do processo de manufatura da Casa da Moeda. Desde pelo menos 1345, a lei inglesa havia reconhecido esta variabilidade ao especificar limites aceitáveis para o peso de uma moeda. Estes limites eram chamados «remédio» e eram estabelecidos em cerca de aproximadamente um por cento do peso-alvo*. Se as moedas ficassem fora destes limites, então os oficiais da Casa da Moeda teriam de «remediar» a Coroa por qualquer falta — e poderiam sujeitar-se a algo muito mais nefasto, dado que o contrato de 1280 da Casa da Moeda colocava-os «à mercê do príncipe [...] a vida e os membros», caso alguma irregularidade fosse descoberta.[9]

3.1 Por Que Era o Julgamento da Píxide Tão Ineficaz?

Ao início, da perspetiva da ciência de dados, o Julgamento da Píxide parece maravilhoso. Envolve um processo de amostragem bem definido sem qualquer preconceito óbvio, a par de um problema que qualquer professor de estatística do século XXI podia passar como TPC: os funcionários da Coroa calcularam o peso médio de uma amostra de moedas e queriam testar se a média estava suficientemente próxima do peso esperado.

Mas o que poderia contar como «suficientemente próxima»? Foi aí que o julgamento se deparou com problemas. Os funcionários da Coroa assumiram que a resposta a esta questão era óbvia: se a lei diz que uma moeda individual deve ter cerca de um por cento do esperado, então o peso médio também deveria estar dentro dessa

margem. Esta resposta «óbvia», contudo, estava bastante errada. Originou limites para declarar uma anomalia que eram *demasiado amplos* — e, portanto, involuntariamente favoráveis aos cerceadores de moedas.

Para compreender o erro aqui, imagine que tem uma amostra de 2500 xelins à sua frente. Suponha que cada xelim deverá pesar 100 gramas — o «valor esperado» — com uma margem de erro permitida de cerca de um grama. A sua tarefa é determinar se estas moedas estão dentro do intervalo de cerca de um grama do seu valor esperado. A abordagem óbvia é pesar todas as 2500 moedas individualmente. Mas consegue imaginar quão aborrecido isso seria? Afinal, estamos no século XVII; tem muitas outras maneiras de passar o seu tempo, como tocar alaúde ou assistir a uma execução pública. Assim, decide poupar todo aquele trabalho calculando o peso médio das moedas: pesa todas ao mesmo tempo, na mesma balança, e divide o resultado por 2500. Isto parece oferecer um retrato bastante bom do trabalho da Casa da Moeda. Se a média estiver próximo de 100 gramas, então a maioria das moedas individuais também deve estar muito próximo de 100. Se a média estiver muito afastada de 100 gramas, então pelo menos algumas moedas individuais também devem estar afastadas de 100.

Este procedimento de determinação da média — que é quase exatamente como o Julgamento da Píxide de facto funcionava — irá por certo detetar as anomalias mais óbvias. Suponha, por exemplo, que o peso médio era de apenas 50 gramas, contra os 100 gramas esperados. Isto é mais evidente que o nariz do Pinóquio: deve

aconselhar os funcionários da Casa da Moeda a contratarem um bom advogado com toda a prata que roubaram, pelo menos se derem valor às suas vidas e membros. Agora, e se for um caso em que não é tão óbvio — digamos, se o peso médio fosse de 99,5 gramas? Isto inicialmente também poderá parecer uma prova de batota, tal como o recorde de 19 vitórias em 25 jogos dos Patriots, no lançamento da moeda ao ar. Mas lembre-se, algumas «anomalias» resultam apenas da sorte. Como é que podemos decidir se uma média de 99,5 gramas é de facto suspeita?

Eis a pergunta-chave. Se os limites permitidos para uma única moeda são cerca de um grama do valor esperado, então quais deverão ser os limites para o peso médio de *muitas* moedas? Há aqui um princípio da Caracóis Dourados. Suponha que estabelece limites bastante estreitos, permitindo que as moedas apenas passem no julgamento se o seu peso médio estiver dentro do intervalo de ± 0,0001 gramas do valor esperado de 100 gramas. Então o seu sistema de deteção de anomalias estará a funcionar de forma demasiado sensível: irá detetar anomalias em todo o lado, como aqueles primeiros alarmes para carros da década de 1980, que disparavam se tocássemos no carro com uma pena. Por outro lado, suponha que estabelece limites bastante alargados, como cerca de 10 gramas — o peso aproximado de uma rodela de limão. Então o seu sistema estará demasiadamente rombo, fazendo com que não detete anomalias reais. A pergunta de um milhão de xelins é: se ± 0,0001 é demasiadamente apurado, e cerca de 10 é muito embotado, então o que é adequado?

3.2 Matéria Adicional: A Regra da Raiz Quadrada, Vulgo Equaçãod e De Moivre

Há uma equação muito importante em estatística, chamada regra da raiz quadrada, que indica exatamente quão estreitos deveriam ser os limites para declarar uma anomalia no Julgamento da Píxide. Esta equação foi descoberta por Abraham de Moivre, um matemático suíço, em 1718. Na nossa opinião, representa um dos triunfos mais subestimados do raciocínio humano de sempre. A maioria das pessoas, por exemplo, já ouviu falar da equação de Einstein: $E = mc^2$. A equação de, De Moivre, é igualmente profunda — representa uma verdade igualmente universal e é identicamente útil para fazer predições científicas precisas. Todavia, muito poucas pessoas fora da estatística e da IA a conhecem. Isto é uma pena, dado ser tão central nesta época de máquinas inteligentes. A equação de, De Moivre, estabelece uma relação inversa entre a variabilidade da média de uma amostra e a raiz quadrada do tamanho da mesma. Ela diz o seguinte:

$$\text{Variabilidade de uma Média} = \frac{\text{Variabilidade de uma Medição Única}}{\sqrt{\text{Tamanho da Amostra}}} = \frac{\sigma}{\sqrt{N}}$$

Em geral os cientistas de dados usam a letra grega o (sigma) para representar a variabilidade de uma medição única, e N para representar o tamanho da amostra. É por isso que expressámos esta equação tanto por palavras como de forma mais compacta por símbolos, como σ/√N. Os cientistas de dados referem-se à «variabilidade de uma média» usando um termo que soa ligeiramente mais técnico: o «erro-padrão da média».

Vejamos um exemplo com alguns números reais associados. Imagine que está a pesar 2500 xelins (*N* = 2500). A lei permite que cada moeda varie no máximo cerca de um grama do peso esperado de 100, em média. Assim, se o controlo de qualidade da Casa da Moeda tiver competência, então σ = 1. A regra da raiz quadrada diz que, se isto for verdade, então o peso médio das 2500 moedas deve cair na vizinhança de $1\sqrt{2500}$ = 0,02 do valor esperado de 100. Deste modo os limites devem ser 100 ± 0,02. Qualquer coisa fora desses limites sugere uma de duas anomalias possíveis: ou um «enviesamento», o que significa que o peso médio das moedas não é realmente 100, ou «sobredispersão», o que revela que a variabilidade de uma medição única é na verdade superior a 1.

Como dissemos, as pessoas que geriam o Julgamento da Píxide acreditavam que se os limites permitidos para uma única moeda eram cerca de um grama, então os limites para o peso médio de

muitas moedas também deviam ser idênticos. Mas, de acordo com a estatística moderna, isto era um erro crasso. Em vez disso, os limites deviam depender de quantas moedas existem na amostra: quanto maior for a amostra, mais estreitos serão os limites. Isto é uma consequência de uma equação muito importante chamada «regra da raiz quadrada», também conhecida como «equação de, De Moivre». Esta regra diz que a variabilidade da média de uma amostra diminui à medida que a raiz quadrada do tamanho da amostra aumenta. A matemática aqui é um bocado complicada, mas a intuição é simples. Numa amostra pequena, uma só moeda que seja demasiado leve pode baixar bastante a média. Mas numa amostra grande, uma moeda leve será provavelmente compensada por uma moeda pesada, pelo que a média deverá ficar próxima do valor esperado. De modo que se fizermos a média de milhares de medidas e o resultado não for muito próximo do esperado, então há algo que não está bem.* Esta matemática é a mesma que os casinos usam quando decidem se enviam ou não uns tipos corpulentos até à mesa do vinte-e-um para uma conversinha com os miúdos do MIT.

Para lhe mostrar quão importante é a regra da raiz quadrada para raciocinar acerca de anomalias, vamos ver como ela se comporta, através da comparação de dois conjuntos de limites lado a lado.

Tamanho da amostra	Limites usados no Julgamento da Píxide	Limites corretos com base na estatística moderna
1	100 ± 1	100 ± 1
100	100 ± 1	100 ± 010
2500	100 ± 1	100 ± 0,02

10.000	100 ± 1	100 ± 0,01

O Julgamento da Píxide estava a usar limites que eram decididamente *demasiado* largos. Este erro teria permitido dois tipos possíveis de anomalias não detetadas, ambas más para Inglaterra.

Primeiro — e isto provavelmente não aconteceu, mas é a possibilidade que desperta o interesse de praticamente todos os cientistas de dados que ouvem falar do Julgamento da Píxide —, os funcionários da Casa da Moeda podiam ter estado a roubar prata. Para ver como isto poderia funcionar, imagine que a Casa da Moeda era capaz de cunhar moedas dentro do padrão legal de variabilidade (± 1 por cento do peso), mas os funcionários sorrateiramente apontavam para um valor esperado de 99,5 gramas por xelim, em vez de 100. (A este tipo de anomalia chama-se um «enviesamento» de 0,5 gramas.) Vamos também imaginar que o Julgamento da Píxide estava a tentar detetar esta fraude pesando uma amostra de 2500 moedas. Se mergulhar na matemática da regra da raiz quadrada, descobrirá que o peso médio destas 2500 moedas provavelmente cairá entre 99,48 e 99,52 gramas. Isto fica bastante fora dos limites estatisticamente corretos de 100 ± 0,02. No entanto, o Julgamento da Píxide teria falhado em fazer soar o alarme, dado que o Júri teria aceitado qualquer média entre 99 e 101 gramas. Os oficiais intrépidos da Casa da Moeda podiam, em teoria, ter surripiado 0,5 por cento de toda a prata em Inglaterra sem terem sido apanhados. Mas apenas alguém que conhecesse a regra da raiz quadrada poderia ter tirado partido do Julgamento da

Píxide dessa maneira inteligente, e não há provas que indiquem que uma fraude tão espetacular tenha alguma vez ocorrido.

Contudo, *há* provas de que de facto aconteceu um segundo tipo de anomalia mais subtil: a Casa da Moeda era honesta mas desleixada, cunhando dinheiro de peso bastante díspar e assim dando aos cerceadores de moedas a benesse da variabilidade excessiva. Suponha, por exemplo, que a Casa da Moeda de facto apontou para o peso esperado de 100 gramas em média. Agora imagine que em resultado do fraco controlo de qualidade, as moedas tinham um peso 10 vezes mais variado do que a lei permitia: 100 ± 10 gramas, em vez de 100 ± 1. Este tipo de anomalia chama-se «sobredispersão». Não é necessariamente prova de fraude, é apenas negligência. Mas o Julgamento da Píxide continuaria a ser incapaz de a detetar. Se as moedas individualmente caem nos limites de 100 ± 10, então a regra da raiz quadrada sugere que o peso *médio* de 2500 dessas moedas quase certamente cairia entre 99,8 e 100,2. Mais uma vez, isto ficaria, muito provavelmente, fora dos limites corretos de 100 ± 0,02, e por isso seria detetado por um procedimento moderno de controlo de qualidade. Mas o Julgamento teria aceitado qualquer coisa entre 99 e 101.

Numa enorme bênção para os cerceadores de moedas, este tipo de anomalia de sobredispersão por certo persistiu durante décadas, se não séculos. Baseamos esta conclusão em dois factos. Primeiro, Isaac Newton comentou especificamente os baixos padrões de manufaturação que encontrou quando chegou à Casa da Moeda. Ele até prestou uma atenção especial à variabilidade das moedas:

«Quando cheguei à Casa da Moeda, *e em muitos anos anteriores*», escreveu ele, «o dinheiro era cunhado desigualmente, peças sendo dois ou três grãos demasiado pesadas e outras igualmente demasiado leves».[10] Newton também escreveu que as moedas pesadas «eram chamadas "guinéus-que-voltam" porque eram cunhadas e trazidas de volta para a Casa da Moeda para serem recunhadas», com lucro para outra pessoa.

Ele estimou que a fração de «guinéus-que-voltam» era tão alta quanto *uma moeda em quatro* — uma prova clara de variabilidade excessiva. Isto não era um problema novo na Casa da Moeda. No Julgamento da Píxide, em 1534, por exemplo, o júri tinha comentado especificamente que «as moedas eram muito desiguais, pelo que era rentável escolher exemplares pesados».[11]

Segundo, na época de Newton terá havido duas ocasiões históricas em que as moedas reprovaram no Julgamento da Píxide.[12] Dois fracassos pode não parecer muito, mas se as moedas individuais tivessem de facto respeitado o padrão legal de variabilidade de ± 1 por cento, então à luz dos limites demasiado largos do julgamento, até mesmo um fracasso teria sido muito mais improvável do que o leitor ganhar a lotaria. À luz dos comentários de Newton acerca do fraco controlo de qualidade, a explicação mais simples para esta taxa elevada de fracasso é que as moedas eram muito mais variadas do que a lei permitia.

3.3 Newton na Casa da Moeda

Os funcionários da Casa da Moeda, como veio a constatar-se, eram mesmo muito maus polícias da variabilidade. Durante décadas deixaram os prateiros realizar impunemente um trabalho descuidado, produzindo moedas que eram muito mais variadas do que o padrão legal de cerca de um por cento por peso. Todavia, o Julgamento da Píxide *nunca os chamou à responsabilidade*, dando assim aos cerceadores de moedas um aliado imensamente poderoso: as leis da probabilidade.

A matemática por trás deste erro terrível, contudo, era algo que nenhum funcionário da Casa da Moeda ao longo dos séculos podia ter compreendido — com a única muito notável exceção de Isaac Newton.

Temos de ter pena do pobre Newton aquando da sua chegada à Casa da Moeda. O seu novo trabalho não era exatamente a prenda que lhe fora prometida. Prometeram-lhe que pagavam 600 libras por ano, mas isto fora um exagero deliberado da parte do chanceler do erário público, pois pagavam-lhe apenas 400 libras. Disseram-lhe que os seus novos colegas eram uma equipa de profissionais de mente astuta, quando na realidade eram um bando de incompetentes; o diretor-adjunto em Norwich acabou na prisão com a sua propriedade confiscada, enquanto o auditor-adjunto foi rapidamente exonerado das suas funções e subsequentemente reconduzido como embaixador de Sua Majestade junto dos piratas de Madagáscar.[13] Finalmente, também foi dito a Newton que ele não teria de trabalhar arduamente no seu novo emprego — e esta

foi a maior mentira de todas. Ele chegou à Casa da Moeda durante a Grande Recunhagem de 1696: a solução drástica e de emergência do Parlamento para o problema do cerceio de moedas, na qual todos os milhões de moedas batidas à mão foram chamadas à Casa da Moeda, fundidas e reformuladas por uma máquina.

A Grande Recunhagem estava em pleno andamento quando Newton chegou e, segundo consta, caminhava para o abismo, devido à má liderança. Contudo, Newton não tratou o seu novo emprego como a sinecura que era, o que só abona a seu favor. Em vez disso, passou à ação. Assumiu trabalho suplementar quando os seus colegas não faziam o deles. Dominava todos os detalhes do complexo sistema de contabilidade da Casa da Moeda. Sugeriu melhorias com base no seu conhecimento de metalurgia, os quais ele aprimorara durante os seus muitos anos em persecução da alquimia. Este conhecimento nunca ajudou a transmutar chumbo em ouro, mas certamente ajudou Newton a transformar lingotes de prata em moedas.[14]

Depois havia o ritmo da Grande Recunhagem. Lembra-se daquela operação infernal de cunhagem mecanizada, com o rolo da massa de dois andares e com a máquina que comia os dedos dos homens? Os trabalhadores tinham assumido que três ou quatro moedas por minuto era um ritmo aceitável, mas isto claramente era demasiado lento para completar a Grande Recunhagem a tempo de evitar o desastre. De modo que Newton empreendeu pessoalmente um estudo detalhado de tempo e movimento dos trabalhadores na linha de montagem, e por fim as suas alterações aumentaram o

ritmo para 50 moedas por minuto. Esta cunhagem foi mantida das 4h à meia-noite, sete dias por semana, durante quase dois anos.[15]

Em 1701, a Grande Recunhagem estava completa e já não havia moedas batidas à mão em Inglaterra.[16] Newton tinha sido promovido ao cargo muito mais prestigiante de diretor da Casa da Moeda, ocasionando um Julgamento da Píxide. O júri reuniu, as moedas passaram o teste, toda a gente teve um grande jantar às custas do novo diretor, e foi tudo. Newton queixou-se amargamente acerca do custo do jantar: duas libras por cada membro do júri, ou mais de 200 libras por cabeça na moeda atual.[17]

Portanto, o Julgamento da Píxide de Newton é impressionante precisamente porque aconteceu com lamúria em vez de estrondo. Aqui estava Isaac Newton, o melhor polícia da variabilidade da história da Casa da Moeda. Tinha passado cinco anos a analisar cada detalhe do processo de cunhagem. Tinha especificamente comentado que as moedas eram demasiado variadas para satisfazerem o padrão legal. Constatara que o excesso de variabilidade era há muito um problema da Casa da Moeda e ficara obcecado com reduzir essa variabilidade. Finalmente, ele era o maior matemático do mundo, a enfrentar um julgamento público com consequências sérias, no qual a variabilidade das moedas era precisamente o assunto em questão.

Se houve alguma vez um caso da pessoa certa, no local certo, na ocasião certa para fazer uma descoberta fundamental em estatística, foi este. Contudo, nada aconteceu. Newton nem sequer deu a entender que havia um problema que precisava de solução —

e o Julgamento da Píxide continuou a cometer o mesmo erro por mais um século. Por que é que Newton não descobriu a regra da raiz quadrada? Isto é um mistério. É difícil acreditar que não ocorreu a Newton uma questão empírica simples: se as moedas individuais eram drasticamente mais variadas do que o padrão legal, como ele explicitamente declarara, então por que passaram por tantos Julgamentos da Píxide ao longo de centenas de anos?

Isto é particularmente surpreendente à luz do apetite permanente de Newton por problemas matemáticos, o qual permaneceu insaciável até durante o frenesim da Grande Recunhagem. Numa tarde em 1696, por exemplo, ele chegou a casa às 16h vindo da Casa da Moeda e sentou-se para trabalhar num problema famosamente difícil, chamado «curva braquistócrona», que fora colocado pelo inimigo mais irritante de Newton, Johann Bernoulli.* Nesse dia, Newton estava muito cansado devido ao trabalho na Casa da Moeda — mas, como escreveu no seu diário, estava ainda mais cansado de ser «importunado e gozado [...] acerca de coisas matemáticas». Assim, nesse dia, saltou o jantar e recusou-se a descansar até às 4h da manhã seguinte, quando já resolvera o problema e mostrara a Bernoulli quem é que mandava. Este tipo de coisa era normal para Newton, mesmo na sua pretensa reforma.[18]

De modo que não foi por falta de oportunidade, genialidade, tenacidade ou motivação proporcionada por um problema matemático real mesmo à sua frente, que Newton não conseguiu ver o seu erro. Isto são tudo coisas que indicamos como ingredientes essenciais numa descoberta científica. O caso de Isaac

Newton na Casa da Moeda tinha todos estes ingredientes, mas nenhuma descoberta. A ironia é que, em comparação com a curva braquistócrona, a matemática por trás da regra da raiz quadrada teria sido uma brincadeira de crianças para Newton — se ele ao menos se tivesse lembrado de logo no início colocar a questão certa. Mas não o fez, e já agora tão-pouco outra pessoa o fez durante bastante tempo. Probabilidade e estatística nem sequer viriam a ser formalizadas como disciplinas durante quase outro século, e coube a dois grandes matemáticos de idade mais avançada — Gauss e Laplace — articular todas as implicações da regra da raiz quadrada.

4. Deteção de Anomalias na Era da IA

O período de Newton na Casa da Moeda é um episódio histórico fascinante e surpreendentemente pouco conhecido, que tem igualmente grandes implicações para a inteligência artificial. Calcular a média de muitas medições juntas é a ideia mais importante na história da ciência de dados. Um número enorme de aplicações da IA depende desta ideia, da prevenção de fraudes ao policiamento inteligente, e todas funcionam ao longo das mesmas linhas fundamentais que o Julgamento da Píxide.

- Recolha de dados: obtêm-se várias medições de um processo subjacente.

- Estabelecer a média: calcula-se a média do conjunto dessas medições para obter um «retrato numérico» do processo.*

- Tomada de decisão: está a média «suficientemente próxima» do que esperávamos, ou está fora dos limites normais de variabilidade?

Existem três grandes diferenças em relação ao tempo de Newton. A decisão de sinalizar uma anomalia é geralmente tomada por uma máquina e não por uma pessoa. Esta decisão ocorre numa escala temporal de poucos milissegundos em vez de poucos anos. Por último, ao contrário das pessoas que geriam o Julgamento da Píxide, estas máquinas fazem bem as contas.

Estes sistemas de IA tornaram-se omnipresentes. As equipas de Fórmula 1 monitorizam fluxos de dados de centenas de sensores nos seus carros em busca de anomalias — temperatura do motor, desgaste dos pneus, aerodinâmica, qualquer coisa que possa afetar a tática da corrida. As empresas de cartões de crédito escrutinam todas as transações que fazemos, em busca de anomalias que indiciem fraude. Os polícias nas grandes cidades transportam sensores de radiação programados para procurar anomalias que possam assinalar uma «bomba suja» deixada por um terrorista. *Facebook* e *Google*, armazéns e mercearias, companhias aéreas e plataformas petrolíferas, senadores e corretores da bolsa, os Cleveland Cavaliers e os desportistas de fim de semana, todos efetuam medições e calculam a média do conjunto, procurando algoritmicamente por anomalias em conjuntos maciços de dados.

Embora a velocidade e a escala destes sistemas tenham mudado bastante nos últimos três séculos, o princípio fundamental não se

alterou: para detetarmos uma anomalia temos de entender a variabilidade.

4.1 Cidades Inteligentes: *N* Grande, *D* Grande

Basta perguntar às pessoas que trabalham no Departamento de Análise de Dados do presidente da câmara, ou MODA*, em Nova lorque. O MODA foi criado em 2013 pelo então presidente da câmara, Michael Bloomberg, para analisar a vasta riqueza de dados recolhidos pelo governo municipal — desde chamadas para o 112 a formulários para fiscalização de obras e relatórios hortícolas acerca dos 5,2 milhões de árvores da cidade.

A riqueza e a escala das diferentes fontes de dados do MODA revelam um facto importante acerca da interseção dos conjuntos de dados maciços com a inteligência artificial. Os conjuntos desses dados não ficam com tal dimensão simplesmente por causa de um «*N* grande»: o número de pontos de dados que têm. Eles também implicam um «*D* grande»: o número de detalhes registado acerca de cada ponto de dados. Por exemplo: no caso de um conjunto de dados sobre apartamentos em Nova Iorque, os detalhes poderão ser tamanho, localização e conveniências; num conjunto de dados sobre pacientes cirúrgicos, os detalhes poderão incluir um conjunto de indicadores de saúde. «*N* grande» significa imensos pontos de dados — imensos apartamentos, imensos pacientes cirúrgicos, etc. «*D* grande» significa imenso detalhe.

Podemos pensar num conjunto de dados «*N* grande, *D* grande» como um conjunto de muitos subconjuntos mais pequenos que

coletivamente exibem uma amplitude vertiginosa (*N* grande) e uma combinação hiperespecífica de detalhes (*D* grande). Em resultado, usar a inteligência artificial neste tipo de conjunto de dados raramente tem que ver com procurar uma anomalia num mar de dados; em vez disso, relaciona-se com procurar milhares de possíveis anomalias em milhões de lagos diferentes. Quanto maior e mais rico for o conjunto de dados, mais lagos haverá e mais detalhadas serão as anomalias que seremos capazes de encontrar.

Por exemplo: a cidade de Nova Iorque tem apenas cerca de 200 inspetores de edifícios para investigar mais de 20 mil queixas por ano respeitantes a conversões ilegais de apartamentos, como quando um senhorio transforma um espaço industrial numa residência ou fraciona um apartamento, já de si pequeno, em minúsculas subunidades.[19] Estes inspetores têm de ser espertos em relação à maneira como alocam os seus recursos, pelo que recorrem ao Departamento de Análise de Dados do presidente da câmara para que os ajude a determinar quais as caraterísticas de uma propriedade mais suscetíveis de produzir um «êxito»: uma inspeção que encontra uma conversão ilegal.

Para ver como isto funciona, imagine que a taxa de êxito dos inspetores é de 10 por cento em todos os apartamentos. Agora considere os seguintes subconjuntos de apartamentos que, com base em inspeções anteriores, parecem ter taxas de êxito elevadas.

- Subconjunto A: prédios de cinco andares sem elevador abaixo da Fourteenth Street construídos antes de 1940,

com lojas no rés do chão. Taxa de êxito = 2 em cada 10 (20%).

- Subconjunto B: condomínios de construção nova com dois quartos em Queens. Taxa de êxito = 17 em cada 100 (17%).

- Subconjunto C: fábricas de vestuário desativadas com mais de cinco novos pedidos de licenciamento para restaurantes num raio de cinco quarteirões. Taxa de êxito = 2 em cada 5 (40%).

Todos estes subgrupos têm taxas de êxito que excedem 10 por cento, mas apenas um é uma anomalia — com uma taxa de êxito que é difícil de explicar como resultado do acaso. Qual deles é a anomalia? Antes de passarmos à resposta, vamos enfatizar a ideia principal: analisar qualquer um destes subconjuntos à procura de uma taxa de êxito elevada é como conduzir uma miniatura de um Julgamento da Píxide avulso. O objetivo é testar para detetar uma anomalia, ou uma diferença da taxa de êxito geral de 10 por cento que seja demasiado grande para ser explicada pelo acaso. (Dica: preste atenção à dimensão das amostras.)

Poderá assumir que a anomalia será o que tem a taxa de êxito mais alta: o subconjunto C, com 40 por cento. Mas, de acordo com a regra da raiz quadrada, na verdade é o que tem a taxa de êxito mais baixa: o subconjunto B, com 17 por cento. A razão é que o subconjunto B também tem a amostra de maior dimensão (100), o

que significa que estamos bastante seguros de que a taxa de êxito elevada é real. As taxas de êxito nos subconjuntos A e C, por outro lado, podem estar elevadas devido à variabilidade da amostragem — ou seja, que apartamentos em particular é que já terá inspecionado. Isto relembra a lição do exemplo da moeda ao ar dos Patriots: a variabilidade é verdadeiramente importante para detetar anomalias, e as amostras pequenas podem ter uma elevada variabilidade.*

É claro que no mundo real há muito mais do que três subconjuntos de apartamentos. É por isso que recorremos à IA, que pode procurar anomalias em milhares ou milhões de combinações possíveis de caraterísticas de apartamentos, incluindo as que um humano nunca imaginou serem importantes. A conceção de algoritmos capazes de fazer isto com precisão e eficácia continua a ser uma área fundamental de investigação. (Não o vamos maçar com os detalhes matemáticos escabrosos.)

Quando o pessoal do MODA começou a aplicar estes algoritmos para correlacionar os relatórios de inspeções de obras com o vasto conjunto de outras fontes de dados de Nova Iorque, produziram resultados assombrosos. A taxa de êxito dos inspetores multiplicou-se por cinco e encontraram dois fatores que se correlacionavam fortemente com conversões ilegais de apartamentos: incrementos súbitos em faturas de serviços públicos e aumento de denúncias de problemas de saneamento. Uma outra equipa de inspetores usou as mesmas técnicas para procurar lojas com venda ilegal de tabaco e álcool, e também melhorou a sua taxa de êxito: de 30 por cento para 82 por cento. Uma terceira equipa conseguiu desferir um golpe

na fraude com opioides, usando dados de reembolsos de seguros de saúde para identificar um pequeno conjunto de farmácias — cerca de um por cento do total — que contabilizava 60 por cento das prescrições de oxicodona na cidade.[20]

E isso são apenas os inspetores. Imagine o potencial para melhorias similares à medida que outras agências municipais comecem a recolher e monitorizar novos tipos de dados — da polícia às brigadas que tapam buracos, do Departamento de Parques aos Bombeiros. Imagine, por exemplo, quantas vidas poderiam ser salvas se se pudesse identificar não apenas onde e quando as pessoas tendem a ser atropeladas por carros, mas também *porquê*. Então começará a compreender a razão para os governos municipais em todo o mundo estarem a brindar ao poder da IA.

4.2 Farejar à Procura de Raios Gama e Fugas de Gás

Uma das pessoas a que brindarão em breve poderá ser Alex Reinhart, um estudante de doutoramento em estatística na Universidade de Carnegie Mellon, que está a trabalhar num novo sistema de deteção de anomalias que pode um dia ajudar as forças de segurança a farejar uma das mais hediondas de todas as ameaças terroristas: a bomba suja.

Uma bomba suja é uma arma cruel e mortífera que utiliza um explosivo convencional para dispersar material radioativo através do ar. A explosão inicial destruiria uma pequena zona e envenenaria uma área muito maior — talvez dezenas de quarteirões de uma cidade. Mas a boa notícia potencial para os agentes da lei é que

qualquer isótopo radioativo emite raios gama em energias previsíveis, de uma maneira relacionada com a estrutura atómica desse isótopo. Os raios gama anómalos emitidos por uma bomba suja podiam assim, em teoria, ser detetados por um «farejador» de radiação antes de a bomba explodir.

Há, no entanto, três problemas — três fontes de variabilidade que tornam difícil sinalizar uma anomalia. Primeiro, não se pode lançar um alerta cada vez que se deteta radiação, porque a radiação natural está em todo o lado. A maioria dos materiais de construção, como tijolos e pedra, tem quantidades minúsculas de urânio e tório radioativos. As bananas e os solos de jardim têm vestígios de potássio radioativo — já para não falar dos raios gama que chegam permanentemente do espaço sideral. Os investigadores usam o termo «NORM» para «materiais radioativos que ocorrem naturalmente», quando se referem a gastas fontes benignas de raios gama. Eles são inofensivos, mas implicam que não podemos simplesmente sinalizar uma anomalia quando detetamos radiação.

Segundo, esta radiação natural varia de local para local, principalmente numa cidade grande. Se atravessarmos a rua ou virarmos uma esquina durante o nosso patrulhamento diário à procura de bombas — uma triste necessidade para os agentes da lei em muitas cidades — encontrar-nos-emos junto a diferentes edifícios construídos com materiais diversos, cada um com um perfil de NORM ligeiramente distinto.

Por último, a radiação é estatisticamente ruidosa, por razões que em última análise têm a ver com a mecânica quântica. Um isótopo

radioativo emite um número variável de raios gama em energias variáveis durante um dado período de tempo. Portanto, nunca podemos saber ao certo se um determinado raio gama veio da natureza ou de uma anomalia.

O resultado final é que procurar anomalias de radiação é um problema surpreendentemente traiçoeiro da ciência de dados. Temos de comparar a radiação observada, que é variável, com a radiação natural normal, que também é variável, e que se altera de local para local. Para fazer isto, precisamos de um mapa detalhado da radiação natural em toda a cidade, a par de um bom algoritmo para detetar anomalias pequenas em dados ruidosos.

Atualmente, a melhor forma de proceder implica inteligência humana — basicamente, contratar alguém com um doutoramento em física nuclear para monitorizar as leituras em tempo real. Mas isto dificilmente será uma solução ajustável ao tipo de operações antiterroristas que se realizam em Londres, Nova Iorque ou Paris. onde é preciso um pequeno exército de pessoas com esta competência de alto nível.

Em vez disso, Reinhart e os seus colaboradores propuseram usar a inteligência artificial. Imaginaram um agente equipado com um pequeno farejador de raios gama, conetado a um smartphone com um sensor de GPS. A cada dois segundos, o smartphone carrega, para um servidor central, a leitura obtida com o farejador de raios gama, a par das coordenadas GPS do agente. O servidor consulta uma base de dados geoespacial da radiação natural da cidade, compilada, usando sensores móveis baratos ao longo de muitos

meses. A leitura atual é comparada com a radiação natural típica na localização do agente, usando a regra da raiz quadrada para determinar os limites para declarar uma anomalia. Se esses limites forem excedidos, o sistema de IA alerta o agente para que investigue.

As aplicações desta tecnologia de farejamento com reconhecimento geoespacial não se ficam pela deteção de bombas. Um dos mentores da investigação de Reinhart, o Dr. Alex Athey, salienta que todas as grandes cidades têm uma vasta infraestrutura de condutas de gás natural, suscetíveis a fugas potencialmente perigosas. Por exemplo: a cidade de Nova Iorque tem mais de 9600 quilómetros de gasodutos sob as ruas, e só em 2012 descobriu 9906 fugas.[21] Em março de 2014, uma dessas fugas provocou uma explosão que matou oito pessoas em East Harlem.

Embora uma cidade pudesse instalar uma nova rede de «condutas inteligentes» capazes de emitir um alarme se ocorresse uma fuga, isso seria não só disruptivo como bastante dispendioso. A solução proposta por Athey é muito mais barata. Imagine colocar sensores de metano em veículos municipais comuns, como camiões do lixo, autocarros ou ambulâncias. Com o passar do tempo, estes veículos iriam percorrer em conjunto grande parte da cidade, permitindo a criação de um mapa com níveis de metano em condições «padrão» de baixas concentrações do mesmo. Se ocorresse algures uma fuga numa conduta de gás, esses mesmos sensores móveis provavelmente iriam registar a anomalia mais depressa do que a companhia de gás conseguiria — e de forma

muito mais barata do que readaptar milhares de quilómetros de condutas enterradas a vários metros da superfície.

4.3 Deteção de Fraudes na Atualidade

Os inspetores municipais e os polícias não são os únicos à procura de quem infringe a lei, através da deteção das anomalias que eles deixam em conjuntos maciços de dados. Nesse esforço também têm a companhia dos maiores bancos do mundo, que cada vez mais recorrem à ajuda da inteligência artificial para afastar a maldição da economia digital moderna: a fraude.

É possível que a fraude seja a segunda profissão mais antiga do mundo. Os gregos da antiguidade acreditavam em Apate, a deusa do engano e um dos espíritos malignos que se encontrava na Caixa de Pandora. Os egípcios contrataram uma classe inteira de escribas para monitorizarem transações que envolviam o inventário de grão do faraó. E alguém deve ter feito algo há 3000 anos para provocar o rei Salomão, que disse em Provérbios 11: «A balança fraudulenta é abominável aos olhos do Senhor; mas o peso justo é-lhe agradável.»

Até há bem pouco tempo, a batalha contra a fraude acontecia pessoalmente, usando a inteligência humana. Em 1685, as suas moedas de prata seriam inspecionadas e pesadas. Em 1885, a sua nota promissória era tão boa quanto a sua reputação. Em 1985, o seu cheque apenas seria aceite se correspondesse aos detalhes pessoais na sua carta de condução. Hoje, contudo, estamos num mundo de *chips* e PIN, e este tipo de vigilância cara a cara já não é

exequível. Por exemplo: em 2015 os bancos americanos processaram 178 biliões de dólares de transações sem dinheiro vivo. Esse valor inclui 70 mil milhões de pagamentos individuais com uma passagem de cartão de débito, 34 mil milhões de pagamentos com uma passagem de cartão de crédito, e 24 mil milhões de transferências bancárias individuais.[22] Infelizmente, também inclui milhares de milhões de dólares de transações fraudulentas — a maioria delas no comércio a retalho normal, que passa esses custos para si. Por cada dólar que gastamos na mercearia, 1,3 cêntimos vão para os bandidos eletrónicos. Eles são os cerceadores de moedas dos tempos modernos.

Felizmente, os cientistas de dados estão a trabalhar arduamente em sistemas de IA capazes de reagir. A solução, tal como acontece com todas as formas de deteção de anomalias, é medir a variabilidade. Pense nos seus próprios hábitos de despesa, os quais variam de maneira previsível de dia para dia e de semana para semana. Essa variabilidade forma a base estatística em relação à qual a fraude pode ser detetada

Durante anos, todos os bancos principais analisaram as suas transações em tempo real com cartões de débito e de crédito em busca de fraudes, e é por isso que o seu cartão por vezes é recusado. Mas a maioria destes sistemas antigos depende de regras fixas simples, como o montante e localização da transação. Isto ignora muita da importante variabilidade de pessoa para pessoa. Para um professor uma série de pagamentos com cartão em três países diferentes no meio de uma semana de aulas poderá ser um claro sinal de fraude. Para um representante comercial externo,

esse mesmo padrão poderá ser normal — e a diferença entre estes dois clientes deveria ser óbvia, tendo em conta os respetivos históricos de transações.

Poderá pensar que as empresas de cartões de crédito andam há já bastante tempo a minerar o seu histórico de transações para aprenderem sobre essas diferenças. E de facto andam — mas só de certa maneira e principalmente para fins comerciais pouco claros. Infelizmente, tem-se revelado bastante mais difícil potenciar todos esses dados num sistema de pagamento em tempo real que aceite ou recuse o seu cartão apenas num décimo de segundo.

A razão é simples: o colossal desafio de engenharia colocado por trabalhar com conjuntos de dados numa tão grande escala. As companhias de cartões de crédito geram *petabytes* de dados de transações, e um *petabyte* corresponde a cerca de 220 mil DVDs. Até há pouco tempo, nenhum sistema de IA de ponta a ponta era suficientemente rápido para potenciar todos esses dados para auxiliar a deteção sofisticada de fraude em tempo real. Todos padeciam de alguma fraqueza determinante — era o desempenho do próprio algoritmo de deteção de fraude, a velocidade da rede ou o processo surpreendentemente lento de ler a partir de um disco físico todos aqueles incontáveis biliões de 1s e 0s.

Em resultado, os bancos enfrentaram um compromisso: se quisessem analisar 100 mil milhões de transações, cada uma em milissegundos, estavam presos a regras relativamente básicas de «*D* pequeno» para deteção de anomalias baseadas no momento, localização ou valor em dólares. E se quisessem explorar o nível de

detalhe fantástico que se encontra no histórico de transações único de cada cliente iriam precisar de meses, não milissegundos, para procurar anomalias. Podiam escolher um *N* grande ou um *D* grande, mas não ambos.

A Paypal, contudo, é apenas uma das muitas empresas de sistemas de pagamento que finalmente resolveram este problema, com a ajuda de algoritmos modernos e de uma infraestrutura contemporânea de supercomputação. O sistema deles de deteção de fraude usa a aprendizagem profunda para comparar todas as transações com o seu próprio comportamento no passado, bem como com o comportamento de utilizadores semelhantes a si. Com base nessa comparação, que usa milhares de caraterísticas possíveis, o sistema produz uma pontuação de probabilidade de fraude que pode ser utilizada para aceitar ou rejeitar a transação — tudo numa fração de segundo.

Com este novo sistema em vigor, a Paypal tem agora um muito melhor entendimento dos limites normais da variabilidade nos seus dados, até ao nível de um utilizador individual. O investimento da Paypal em IA compensou amplamente: em 2016 a taxa de fraude caiu para 0,32 por cento da sua receita, menos de um quarto da média global da indústria.[23] Outras empresas de sistemas de pagamento, como a Alipay na China ou a Stripe nos Estados Unidos, investiram em tecnologias similares. E estes sistemas continuam a melhorar, dado que com cada novo ponto de dados eles aprendem um pouco mais sobre a fraude de dados.

Tanto o rei Salomão como Isaac Newton estariam orgulhosos.

5. *Moneyball* para a Era Digital

Se é um fã de desporto, provavelmente já ouviu falar de «Moneyball»*, o termo do autor Michael Lewis para uma abordagem particular centrada nos dados para construir e treinar uma equipa desportiva. No final da década de 1990, os Oakland A's chegaram à conclusão de que os olheiros tradicionais do basebol não eram na verdade muito eficazes a avaliar o que constituía um bom jogador. Muito do que estes olheiros atribuíam à competência era na realidade sorte, e vice-versa; eles andavam a confundir sistematicamente sinal com ruído. Sob o aconselhamento dos seus olheiros, a maioria das equipas de basebol pagou milhões de dólares por jogadores que na verdade não ajudaram a ganhar qualquer jogo, ou cujos sucessos do passado não eram reproduzíveis. Entretanto, muitos outros jogadores passavam despercebidos, apesar de ajudarem as suas equipas de maneiras que eram importantes e reproduzíveis, mas não óbvias. Esta ineficiência criou uma oportunidade para a primeira equipa que encontrasse um caminho melhor. Os Oakland foram triplamente inovadores. Usaram dados para determinar quais as caraterísticas e hábitos dos jogadores que efetivamente ganhavam jogos. Depois aproveitaram essas descobertas para procurar anomalias no mercado — caraterísticas e hábitos vencedores que eram sistematicamente subvalorizados por outras equipas. Seguidamente contrataram jogadores com essas caraterísticas e treinaram-nos para terem esses hábitos. Em resultado, foram capazes de competir — e vencer — contra equipas como os Red Sox e os Yankees, que podiam gastar em jogadores três vezes mais do que os Oakland.

Se avançar 25 anos, estas inovações alteraram todas as modalidades mais importantes do mundo. Hoje, contudo, há uma grande diferença: na década de 1990, *Moneyball* era algo que podíamos jogar com uma folha de cálculo e um estagiário inteligente. Agora precisamos de um supercomputador com base na nuvem e uma equipa dedicada de cientistas de dados — tudo por causa dos novos conjuntos de dados maciços que as equipas desportivas começaram a recolher assim que perceberam a vantagem que lhes proporcionaria.

5.1 Fórmula 1

O exemplo perfeito desta revolução é a Fórmula 1, a competição de corridas de carros mais popular em todo o mundo. Numa corrida de Fórmula 1, os dados fluem mais rápido do que o champanhe nos camarotes de luxo. Todos os aspetos relacionados com o desempenho de um carro são monitorizados em tempo real e com um detalhe microscópico. Um carro de Fórmula 1 gera vários *gigabytes* de dados por volta, mais ou menos a quantidade necessária para emitir 30 horas de música ou descarregar 6000 livros eletrónicos. Estes dados são transmitidos sem fios de volta para a equipa nas *boxes*, que usa algoritmos sofisticados para procurar anomalias que possam afetar as táticas de corrida: potência do motor, temperatura dos travões, consumo de combustível, desgaste dos pneus, forças G laterais, força descendente na asa traseira e centenas de outras variáveis. As equipas já não têm de esperar até que uma peça falhe inesperadamente, arruinando a sua corrida. Agora podem prever essas falhas antes que aconteçam.

Na verdade, a pesquisa de dados não termina na pista. A Fórmula 1 envolve uma dispendiosa corrida ao armamento entre as equipas pela tecnologia de ponta, e para restringir isto as regras limitam o número de colaboradores junto à pista que cada equipa pode ter em dia de corrida. Sem isto, o dinheiro gasto pelas equipas grandes conduziria ao desaparecimento das pequenas; afinal de contas, esta é uma modalidade na qual as equipas pagam cem milhões de dólares por ano pelos motores e empregam três pessoas por pneu para as paragens na box. As equipas mais ricas decidiram, todavia, que precisavam de uma capacidade ainda maior de tratamento de dados, pelo que se viraram para engenheiros fora da pista. A *Red Bull Racing*, por exemplo, associou-se recentemente com a AT&T a fim de construir uma rede global para transmitir dados de corridas de qualquer circuito de Fórmula 1 no mundo para a sede da sua equipa em Milton Keynes, em Inglaterra. Ali, uma segunda equipa de cientistas de dados monitoriza o carro da *Red Bull* em tempo real. Ou melhor, quase tempo real: o fator limitador do desempenho do sistema é a velocidade da luz, que pode dar a volta ao mundo apenas 7,5 vezes num segundo. Esse tipo de investimento deve dar-lhe uma boa noção de quanto significa para os engenheiros a deteção em tempo real de anomalias.

As equipas de Fórmula 1 tornaram-se tão boas em monitorização em tempo real que algumas das melhores começaram a vender os seus serviços a outras grandes empresas. A equipa da McLaren, por exemplo, deslocou recentemente a sua equipa de análise de dados para uma outra empresa, chamada Mclaren Applied Technologies, que de imediato assinou um contrato com a empresa de consultoria KPMG. Entre outros projetos, está agora a ajudar clientes da

indústria petrolífera a monitorizarem em tempo real dados dos sensores de plataformas petrolíferas, à procura de anomalias que possam indiciar problemas.

5.2 Para Lá da Pista de Corridas

Estas inovações difundiram-se por outras modalidades. Em 2016, por exemplo, os Brooklyn Nets assinaram um contrato de patrocínio com uma empresa chamada Infor, muito pouco conhecida por quem está fora do círculo do software empresarial. A Infor desenvolve software para análise de dados maciços — incluindo para a equipa de Fórmula 1 da *Ferrari* — e, ainda que tenha pagado milhões de dólares pelo direito de exibir o seu logótipo na camisola dos Nets, também trouxe muito mais para a mesa de negociações do que simplesmente um livro de cheques em branco.

Brett Yormark, o presidente executivo dos Nets, explicou que ao vender património imobiliário na camisola da sua equipa, queria identificar um parceiro estratégico «que fosse suficientemente substancial para nos ajudar com o desempenho tanto dentro como fora do campo». O acordo que assinou com a Infor é emblemático da nova Era de *Moneyball* da NBA, na qual algumas das maiores estrelas da liga irão usar o logótipo de uma empresa de dados maciços nas suas camisolas.[24]

Na NBA, grande parte desta revolução foi alimentada por novas fontes de dados, como sensores de movimento em todos os jogadores e câmaras que cobrem todos os ângulos do campo. Mas

também tem sido alimentada por uma mudança generalizada na filosofia de constituição de equipas e por um grande investimento em talento analítico. Os Sacramento Kings, por exemplo, recentemente contrataram Luke Bornn, anteriormente um professor assistente de estatística em Harvard, para explorar todos aqueles dados de vídeo e seguimento de jogadores. Como Bornn disse numa entrevista à NBC Sports:

> Muito do que acontece no campo realmente não é captado pelo quadro de estatísticas. Muitos jogadores que têm grandes contribuições fazem-no de maneiras que não aparecem. Não é uma assistência, não é um ressalto, não é um desarme.[25]

Ao invés, é outra coisa — algo anteriormente não reconhecido pelos treinadores, mas escondido à vista de todos nos dados e à espera para ser descoberto. Bornn está convencido de que usar a inteligência artificial para explorar todos os dados em busca de anomalias interessantes irá ajudar os Kings a encontrarem jogadores subvalorizados e a treiná-los de maneiras inovadoras. Ele e um grupo de coautores, por exemplo, recentemente publicaram um artigo sobre medidas avançadas de competência defensiva no basquetebol. Usando dados obtidos através de câmaras montadas nas vigas de suporte em todas as arenas da NBA, eles foram capazes de responder a duas questões simples que nunca antes desempenharam um papel nas estatísticas do basquetebol: quem estava a marcar quem em todos os momentos, e como se saíram determinados defensores contra determinados oponentes?

Bornn e os seus colegas descobriram que a seleção do lançamento (onde e quando um jogador lança) e a *eficiência* do

mesmo (se o lançamento é ou não convertido) são duas componentes distintas da competência defensiva no basquetebol. Estas competências também têm uma estrutura espacial clara: dependem de onde um defensor está posicionado no campo. Perto do cesto, por exemplo, o poste Dwight Howard, dos Charlotte Hornets, é melhor do que a média a diminuir a frequência de lançamento, mas pior do que a média a diminuir a eficácia de lançamento — e ele está abaixo da média em ambas quando está afastado do cesto. Estas descobertas permitiram a Bornn e aos seus colegas preverem os resultados de emparelhamentos defensivos específicos. Por exemplo: o modelo deles inferiu que seria expetável que LeBron James marcasse menos pontos contra Kawhi Leonard, dos San Antonio Spurs, do que contra qualquer outro defensor da NBA.[26] Não se trata apenas de Leonard ser em geral um grande defensor, é que a sua *mistura* específica de competências defensivas constitui um emparelhamento especialmente favorável contra as competências ofensivas de James.

Os hábitos quotidianos dos jogadores da NBA também são perscrutados em busca de anomalias, e esta iteração «comportamental» do *Moneyball* é algo completamente novo. Jeremy Lin, base dos Brooklyn Nets, acha que a parceria da sua equipa com a Infor já começou a pagar dividendos, ao ajudá-lo a cuidar do corpo do mesmo modo que uma equipa de Fórmula 1 trata dos seus carros. Em particular, atribui à análise avançada o mérito de melhorar o seu sono e de o ajudar a recuperar mais rapidamente de uma lesão irritante num tendão.[27]

As equipas de outras ligas desportivas profissionais também abraçaram a IA, pela mesma razão por que abraçaram a publicidade nas camisolas: há muito dinheiro em jogo. Por exemplo: o Leicester City Football Club, na Premier League inglesa, fez um uso muito inteligente dos dados de movimentação dos seus jogadores durante a época de 2015-16, em que conquistaram o título. A equipa tinha acesso a dados de um sistema chamado *Prozone3*, que combina câmaras e sensores utilizáveis. Tal como todas as equipas da Premier League, ela usou esses dados para adaptar a cada oponente as táticas durante o jogo. Mas o Leicester City também explorou esses dados em busca de outra coisa: anomalias na movimentação e na carga de treino de um jogador que indicassem um risco elevado de lesão. Estes esforços conduziram a equipa à mais baixa taxa de lesões e ao 11 inicial mais consistente da Premier League.

6. Pós-Escrito

Vamos deixá-lo com um último detalhe sobre o Julgamento da Píxide. Veio a verificar-se que, apesar de as moedas inglesas já não serem feitas de prata, esse julgamento ainda ocorre nos nossos dias. Todos os anos, na segunda terça-feira de fevereiro, um júri de ourives reúne em Londres para pesar uma amostra de moedas e testar a sua pureza. Felizmente, eles aprenderam com os erros do passado: os limites para declarar uma anomalia têm sido calculados com uma base estatística sólida desde meados do século XIX.

Há ainda uma outra diferença: cerca dos últimos 75 anos, o júri também testou a largura e o diâmetro das moedas. Essas medidas

particulares não eram importantes no tempo de Newton. Curiosamente, também não são importantes hoje em dia, pois foram acrescentadas para dar resposta às exigências de um período fugaz na longa história de Inglaterra, quando os londrinos usavam as moedas para efetuarem chamadas telefónicas a partir de caixas vermelhas nas esquinas das ruas.

---

[1] Lembramo-nos distintamente de ouvir este comentário num programa de televisão, no rescaldo do incidente da moeda ao ar, mas não conseguimos encontrar online uma transcrição do programa. As nossas desculpas ao espirituoso e infelizmente anónimo comentador.

[*] Também simulámos os 24 jogos que antecederam o jogo 1 da temporada de 2007, para que a média de 25 jogos estivesse bem definida no início da série dos 176 jogos em questão. Isto implica que a percentagem de vitórias da série de 25 jogos que começou com aquele primeiro jogo em 2007 na verdade remonta a meados de 2005.

[*] Se teve uma disciplina de estatística, talvez reconheça este número como o valor-p (p = 0,23) sob a hipótese nula de não haver batota.

[†] Dois sinónimos de anomalias que talvez tenha encontrado são «sinais no ruído» ou «violações da hipótese nula».

[2] Stephen Quinn, «Gold, Silver, and the Glorious Revolution: Arbitrage Between Bills of Exchange and Bullion», *The Economic History Review* 49, n.º 3 (1996): 479-90.

[3] Thomas Levenson, *Newton and the Counterfeiter* (Boston: Mariner Books, 2010), 626-63.

[4] John Craig, *Newton at the Mint* (Cambridge: Cambridge University Press, 1946), 6—7.

[*] Esta inscrição, *Decus et Tutamen*, permaneceu nas moedas inglesas até 2017, quando infelizmente foi removida da última versão da moeda de uma libra.

[5] Ming-hsun Li, *The Great Recoinage* of 1696 to 1699 (Londres: Weidenfeld and Nicolson, 1963), 47.

[6] Levenson, Newton and the Counterfeiter, 137-38.

[7] Thomas Babington Macaulay, *The History of England from the Accession of James II*, Volume 1 (Nova Iorque: Harper & Brothers, 1856), 187.

[*] O Julgamento deve o seu nome a uma divisão na Abadia de Westminster, a Capela da Píxide. «Píxide» é uma palavra que vem do grego antigo que significa uma caixa com pão para a comunhão. As moedas que aguardavam o julgamento eram colocadas numa caixa, e a caixa era guardada na capela — daí o Julgamento da Píxide.

[8] Os pormenores do processo de inspeção usado no Julgamento da Píxide são descritos em Stephen M. Stigler, *Statistics on the Table* (Cambridge, Mass.: Harvard University Press, 1999), 386-89.

[*] Decidimos arredondar para manter os números simples. Na verdade, o remédio era estabelecido a 48 grãos por libra *troy* de prata, o que é cerca de sete gramas por quilograma, ou 0,7 por cento por peso. Veja Stigler, *Statistics on the Table*, Capítulo 23.

[9] *Ibid*., 389-90.

[*] Se quiser ver a matemática por trás da regra da raiz quadrada, veja a matéria adicional técnica na página 133. Mas não é necessário acompanhar aqui a matemática para entender a lição principal deste capítulo.

[10] John Craig, *The Mint: A History of the London Mint from A.D. 287 to 1948* (Cambridge: Cambridge University Press, 2011), 212, ênfase acrescentada.

[11] *Ibid*., 104.

[12] Stigler, Statistics on the Table, 391.

[13] Craig, Newton at the Mint, 12-14.

[14] Levenson, Newton and the Counterfeiter, 139-41.

[15] *Ibid*., 141-44.

[16] Os historiadores de economia consideram que a Grande Recunhagem foi um fracasso enquanto política monetária. Estamos somente a salientar que foi um sucesso enquanto empreendimento industrial, independentemente dos seus efeitos económicos.

[17] Craig, Newton at the Mint, 48-49.

[*] Bernoulli achava que a teoria da gravidade de Newton era um disparate, e também era amigo de Leibniz, que tinha uma disputa de prioridade com Newton acerca de cálculo.

[18] *Ibid*., 23.

[*] Muitos destes sistemas não dependem de calcular a média propriamente dita, mas de um outro qualquer retrato numérico dos dados. Um exemplo simples seria a mediana, enquanto exemplos complicados têm nomes como «pontuações de componentes principais» ou «estatística de Kolmogorov-Smirnov». Este detalhe não é importante; a necessidade de entender a variabilidade permanece a mesma, quer estejamos a usar uma média ou qualquer outro retrato numérico mais vistoso.

[*] MODA — Iniciais de Mayor's Office of Data Analytics. [N. T.]

[19] David A. Schweidel, Profiting from the Data Economy: Understanding the Roles of Consumers, Innovators and Regulators in a Data-Driven World (Upper Saddle River, N.J.: Pearson FT Press, 2014), 81.

[*] Estes outros subconjuntos podem vir a revelar-se anomalias, simplesmente precisamos de mais dados para ter a certeza. Na prática, há um compromisso entre explorar subconjuntos com poucos dados e obter «vitórias fáceis» através de inspecionar subconjuntos com anomalias claras

[20] *Ibid*., 82; Livro Branco da Accenture, «City of New York: Using Data Analytics to Achieve Greater Efficiency and Cost Savings», 2013, https://www.accenture.com/t20150624T211456Zw/usen/acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Technology. 7/Accenture-Data-Analytics-Helps-New-York-City-Boost-Efficiency-Spend-Wisely.pdf. (Página não encontrada.)

[21] Patrick McGeehan, Russ Buettner e David W. Chen, «Beneath Cities, a Decaying Tangle of Gas Pipes», *The New York Times*, 23 de março de 2014, A1.

[22] Estudo de Pagamentos da Reserva Federal, 2016, https://www.federalreserve.gov/paymentsystems/2016-payment-study.htm.

[23] Michael Morisy, «How PayPal Boosts Security with Artificial Intelligence», *MIT Technology Review*, 25 de janeiro de 2016, https://www.technologyreview.com/s/545631/how-paypal-boosts-security-with-artificial-intelligence/.

[*] O livro de Michael Lewis, *Moneyball: The Art of Winning an Unfair Game*, de 2002. foi adaptado para o cinema em 2011 e estreou em Portugal com o título Moneyball — Jogada de Risco. [N. T.]

[24] «Brooklyn Nets' Jeremy Lin on New Partnership», entrevista televisiva no *Squawk Box*, CNBC, 8 de fevereiro de 2017, http://video.cnbc.com/gallery/?video=3000591640.

[25] James Ham, «Kings Add New Stat Guru Luke Bornn to Front Office», NBC Sports, 20 de abril de 2017, http://www.csnbayarea.com/kings/kings-add-new-stat-guru-luke-bornn-front-office.

[26] Alexander Franks, Andrew Miller, Luke Bornn, e Kirk Goldsberry, «Counterpoints: Advanced Defensive Metrics for NBA Basketball», artigo apresentado na 9.ª MIT Sloan Sports Analytics Conference, 2015, http://www.lukebornn.com/papers/franksssac2015.pdf. (Não funciona).

[27] «Brooklyn Nets' Jeremy Lin on New Partnership», entrevista televisiva no *Squawk Box*, CNBC, 8 de fevereiro de 2017, http://video.cnbc.com/gallery /?video=3000591640.

CAPÍTULO 6

A Dama da Lâmpada

O que a Guerra da Crimeia nos pode ensinar acerca das perspetivas de uma revolução da IA nos cuidados de saúde — e acerca da cultura e das instituições que ajudaram a inovação a criar raízes.

Se ler as notícias atuais acerca dos cuidados de saúde irá encontrar duas narrativas muito diferentes.

Primeiro, as más notícias: os sistemas de cuidados de saúde em todo o mundo rico estão a agonizar sob o peso de populações doentes e envelhecidas. A obesidade e a doença cardíaca estão a aumentar e os custos estão descontrolados. Em 2016, dois terços de todos os fundos hospitalares britânicos apresentaram um défice, e o serviço de saúde francês ultrapassou o seu orçamento em 3,4 mil milhões de euros. Entretanto, a percentagem do PIB que os americanos gastam em saúde é superior à de qualquer outro povo, mas isso não faz com que sejam mais saudáveis. Os médicos passam os seus dias a combater as companhias de seguros, a temer processos judiciais e a digitar dados para um sistema eletrónico de registos de saúde; em comparação com os não médicos, eles têm uma probabilidade 40 por cento maior de abusarem de álcool ou drogas, e duas vezes superior de cometerem suicídio.[1]

Talvez como um antídoto a todas estas histórias deprimentes também nos dizem que a inteligência artificial está pronta para revolucionar os cuidados de saúde. Os evangelistas da IA descrevem um mundo futurístico onde o seu cirurgião é assistido por um robot guiado por *laser*, tal como o carro da Google; onde os seus sinais vitais são monitorizados algoritmicamente para detetar anomalias, tal como o seu cartão de crédito; e onde os seus tratamentos são personalizados, tal como a sua conta *Netflix*. É um mundo onde o seu *Fitbit* pode dizer-lhe se vai entrar em trabalho de parto, onde pode tirar uma fotografia de uma lesão na pele e obter um diagnóstico instantâneo a partir do seu telefone, e onde o seu relógio inteligente sabe o incentivo certo para que coma mais vegetais ou para que opte pelas escadas.

Neste mundo, os médicos já não gastam um terço do seu tempo a efetuar o registo manual de dados. Em vez disso dizem tudo a uma espécie de *Amazon Echo* em esteroides, que de imediato atualiza o seu processo clínico — que então é analisado utilizando regras preditivas sofisticadas, treinadas em bases de dados enormes, que ajudam os médicos a procurar sinais de alerta escondidos. É um mundo de interação perfeita entre humano e inteligência automática, já que os sensores utilizáveis baratos, acoplados a tecnologia de diagnóstico e monitorização baseada em IA e acessível através de smartphone, proporcionam uma melhoria significativa no cuidado das comunidades carenciadas — primeiro no mundo rico e depois no mundo em desenvolvimento. O parto torna-se mais seguro; as doenças são detetadas mais cedo; navios carregados de potencial humano chegam a bom porto.

Esperamos que concorde que este mundo soa muito bem — assumindo que podemos responder às suas preocupações acerca da privacidade de dados, o que tentaremos fazer até ao final deste capítulo. Assim, a nossa pergunta é: por que não estamos já a viver neste mundo? Todas as tecnologias de IA que acabámos de enunciar já existem numa qualquer etapa de investigação ou desenvolvimento e é manifestamente evidente aquilo que é necessário para agilizar a sua adoção generalizada: melhores dados, colaboração mais profunda entre prestadores de cuidados de saúde e cientistas de dados, leis mais inteligentes que possam fomentar a inovação ao mesmo tempo que salvaguardam os pacientes e a sua privacidade. Mas, como irá descobrir neste capítulo, só porque uma coisa boa *pode ser* feita com os dados não significa que *venha a ser* feita.

Até ao momento destacámos exemplos de progresso tecnológico tremendo na IA. Iremos agora mudar o nosso foco para a interação entre tecnologia e cultura — os valores, incentivos e hábitos que determinam como as pessoas se comportam. Para produzir o tipo de revolução que todos desejamos nos cuidados de saúde, por certo precisamos de recursos, dados e pessoas. Acima de tudo precisamos de um compromisso cultural — reunir esses recursos, dados e pessoas. O *Google*, o *Facebook*, a *Amazon*, o *PayPal*, o *Baidu*, o *Alibaba*, a *Fórmula 1*, o Departamento de Análise de Dados do presidente da câmara em Nova Iorque, a quinta de pepinos de Makoto Koike no Japão... todos assumiram este compromisso nas suas áreas respetivas, com resultados impressionantes. O que torna ainda mais trágico que tal compromisso esteja a faltar nos cuidados de saúde, onde a IA poderia ajudar mais pessoas do que em

qualquer outro lado. Provavelmente ainda estaremos a anos de ver as nossas tecnologias mais avançadas de IA a ajudarem um número significativo de pacientes reais, e as razões nada têm a ver com ciência ou capacidade de computação e tudo a ver com cultura, incentivos e burocracia. E isto não é um problema apenas dos Estados Unidos. Os sistemas de cuidados de saúde na América, Europa e Ásia diferem em aspetos importantes, mas partilham algumas semelhanças em termos de como a IA poderia estar a ajudar, e por que ainda não está. O cancro e as doenças renais não têm nacionalidade, mas há uma palavra para burocracia em todos os idiomas.

Num momento como este, ajuda procurar um exemplo histórico de alguém que enfrentou um problema similar e que o superou — alguém que possuía o conhecimento, a estatura e a determinação para enfrentar as pessoas poderosas que geriam os sistemas de cuidados de saúde e em nome de todos nós dizer: Parem, por favor. Por que não o fazem desta maneira? Não conseguem ver como as coisas podem ser tão melhores?

Felizmente conhecemos a pessoa certa: Florence Nightingale.

Dependendo da sua idade, poderá ou não conhecer Florence Nightingale como a enfermeira mais famosa de sempre — a «dama da lâmpada», que se tornou um símbolo vivo de compaixão enquanto cuidava dos soldados britânicos feridos na Guerra da Crimeia. Acontece que quando não estava a cuidar dos soldados, Nightingale era também uma competente cientista de dados que conseguiu convencer os hospitais de que podiam melhorar os

cuidados de saúde usando estatísticas. Na verdade, nenhum outro cientista de dados da história pode reclamar ter salvado tantas vidas quanto Florence Nightingale. Em 1859, para homenagear estes triunfos, ela tornou-se a primeira mulher de sempre eleita para a Sociedade Real de Estatística do Reino Unido.

O percurso de Nightingale para desvendar o poder dos dados nos cuidados de saúde oferece atualmente três lições distintas.

Primeiro, ilustra o tipo de compromisso institucional necessário para que uma revolução de ciência de dados se instale numa determinada área. De facto, se tiver qualquer interesse profissional em saber como a IA poderá mudar o seu setor, não encontrará melhor lição.

Segundo, mostra aquilo que enfrentamos enquanto pacientes que querem os melhores cuidados. Na sua demanda para trazer uma melhor análise de dados para os cuidados de saúde na década de 1850, Florence lutou contra interesses instalados que defendiam o status quo contra mudanças que iriam ajudar os pacientes. Hoje. a luta para fazer a mesma coisa está a desenrolar-se de maneira escandalosamente similar e, se a primeira vez foi uma tragédia, esta segunda parece uma farsa.

Por último, a história de Florence é inspiradora. O sistema de cuidados de saúde atual necessita sem dúvida de pessoas com a tenacidade, a inteligência e a coragem moral que Florence

Nightingale demonstrou há 160 anos. Talvez o leitor venha a ser uma dessas pessoas.

1. O Anjo da Crimeia

Florence Nightingale nasceu em 1820, num ambiente de conforto e privilégio. Todos os anos a sua família arrendava uma suíte de hotel na cidade para uma temporada em Londres, antes de se retirarem para uma das suas duas propriedades no campo. Quando passavam férias na Europa viajavam numa carruagem grandiosa com espaço para doze pessoas e desfrutavam de entretenimentos sumptuosos: óperas todas as semanas, «bailes até ao infinito», banquetes oferecidos pelo grão-duque da Toscana.[2]

Mas Florence experienciava esta vida como uma gaiola dourada. Isto porque ela tinha dois amores verdadeiros, e nenhum deles podia ser encontrado nos prazeres ociosos das salas de visitas.

O seu primeiro amor era a matemática. Mesmo em criança, Florence mergulhara no seu livro de matemática resolvendo problemas de uma Era desaparecida: «Se existem 600 milhões de pagãos no mundo, quantos missionários são precisos para atribuir um a cada 20 mil?»[3] Ela entretinha-se com jogos de palavras com matemática — «Eu peguei em "respiração" e construí 40 palavras», escreveu ela com 7 anos.[4] Na adolescência aprendeu geometria com Euclides e logaritmos com o seu primo, Henry, e implorou aos pais que a deixassem fazer uma visita prolongada ao seu tio Octavius, que tinha uma fantástica biblioteca de matemática.[5]

Ainda mais do que a matemática, Florence adorava a enfermagem. Em criança tratava de cães feridos, escreveu um epitáfio para uma carriça e lamentou a situação de uma vaca com uma tosse horrível; na adolescência visitava os doentes e os pobres da aldeia local quase todos os dias. Quando Florence desaparecia durante a noite, a sua mãe sabia que tinha de ir bater às portas na aldeia, onde ela podia ser encontrada «sentada à cabeceira de alguém que estava doente, a dizer que não podia ir sentar-se à mesa para o grandioso jantar das 19h».[6] Se ela de facto se sentasse para jantar, tenderia a colocar questões constrangedoras a qualquer convidado, não importava quão distinto, que parecesse fechar os olhos ao sofrimento dos outros.

Ela cedo estabeleceu como objetivo ter uma carreira como enfermeira profissional, escrevendo no seu diário: «A minha mente está absorta com a ideia do sofrimento do homem [...]. Tudo o que os poetas cantam sobre as glórias deste mundo me parece falso. Todas as pessoas que vejo estão consumidas pela preocupação, a pobreza ou a doença.» Ela chegava a acordar tão cedo quanto as 3h para ler alguma coisa que conseguisse encontrar sobre assistência social: estatísticas dos censos, minutas do Parlamento ou um «Relatório sobre as Condições Sanitárias das Classes Trabalhadoras da Grã-Bretanha».

Os seus pais, infelizmente, achavam que as ambições profissionais dela eram frustrantes e bizarras, completamente impróprias para uma senhora com a sua posição, e recusaram o desejo dela de entrar para um programa de formação de enfermeiras. Florence respondeu ao rejeitar o ideal deles de

feminilidade como «alguém alimentado a rebuçados», como a vida «da cotovia a cantar sob o sol brilhante», que nunca «desce como o resto de nós à toca de coelho lotada que até arranha, onde os seus habitantes estão a cavar a abrir buracos e a fazer pó».[7] Sentia-se culpada por estar tão infeliz, quando ela mesma desfrutava de tantos privilégios enquanto tantos outros levavam vidas de sofrimento e sem dinheiro. Porém, quando completou o seu trigésimo aniversário, e como a sua família barrara os seus desejos uma e outra vez, os seus pensamentos tornaram-se depressivos e suicidas.

Em última instância, contudo, a força de vontade de Florence — aquilo a que a sua irmã Parthenope chamou «a coisa mais resoluta e férrea que alguma vez conheci» — triunfou: aos 31 anos finalmente obteve a permissão dos pais para estudar enfermagem em Kaiserwerth, um famoso hospital de beneficência na Alemanha. Foi um ponto de viragem. O seu período em Kaiserwerth envolveu longas horas com os doentes e os desesperados. Fez curativos, tratou pacientes com febre tifoide, cuidou de amputados, e sentou-se em vigília à cabeceira dos moribundos. A experiência fê-la sentir-se como uma nova mulher. Ela estava finalmente a responder ao chamamento que ouvira durante toda a sua vida e, como lhe escreveu uma amiga, «irás achar o tagarelar da vida comum mais insuportável que nunca, após teres saboreado a escolha do teu próprio coração».[8]

De facto, quando Florence finalmente regressou a Inglaterra, nem a sua família nem as corteses convenções da sua classe iriam impedi-la de concretizar a sua ambição de sempre. Ela começou a

trabalhar num pequeno hospital para mulheres na Harley Street, em Londres, ganhando rapidamente uma reputação de competência e compaixão, e em 1854 foi-lhe oferecido o emprego dos seus sonhos: superintendente das enfermeiras no hospital King's College. Mas a história tinha outros planos para ela, pois em outubro, numa época em que se agravava a guerra entre a Grã-Bretanha e a Rússia, Florence Nightingale seria chamada para servir o seu país.

1.1 «Ar Viciado e Males Evitáveis»

A centelha da Guerra da Crimeia fora acesa em 1853, quando a Rússia invadiu os Balcãs, ameaçando dessa forma a Turquia, aliada da Grã-Bretanha. Em resposta, os britânicos declararam guerra à Rússia em março de 1854, enviando tropas para a península da Crimeia para fazerem o cerco a Sebastopol, o principal porto para a Frota do mar Negro russa. As pessoas em Londres, cheias de fervor nacionalista, tinham assumido que a guerra acabaria num ápice. Estas expetativas de uma vitória célere foram rapidamente liquidadas, quando se tornou claro que o exército britânico, com uma geração removida na sua última grande guerra — contra Napoleão, em 1815 — estava totalmente impreparado para enfrentar a Rússia.[9]

Isto foi sobretudo evidente no decadente sistemas médico do Exército, onde assuntos básicos de saneamento e redes de abastecimento eram considerados abaixo da dignidade dos homens da medicina no comando. O resultado de todo este planeamento deficiente foi uma catástrofe logística e humanitária. Um soldado

ferido na Crimeia geralmente dava por si acondicionado num navio encardido a ser transportado ao longo de 480 quilómetros até ao hospital Barrack, em Scutari, do outro lado de Constantinopla, no Bósforo. Ali chegado, ele poderia esperar até três dias para ser levado para terra, antes de ser colocado numa maca, ou talvez amarrado a uma mula, para uma subida trepidante ao longo de uma colina íngreme até ao imundo hospital. Ali ele iria encontrar carnificina, sob a forma de colegas soldados espalhados em esteiras finas no meio de ratos, sangue, fedor e imundície. A cólera e a disenteria eram galopantes: os esgotos estavam entupidos, as sanitas vazavam excrementos para o pátio, e uma conduta da água estava bloqueada pela carcaça de um cavalo em decomposição.[10] O hospital tinha uma escassez grave de material médico, roupas limpas, comida saudável, e clorofórmio — muitas amputações eram realizadas sem ele.[11] Também faltavam médicos e todos os que ali estavam corriam de uma emergência para a seguinte pelos corredores, desviando-se de homens e cadáveres.

À chegada do outono de 1854, as condições em Scutari estavam sob forte escrutínio. Um artigo de 30 de setembro do *The Times* canalizou a crescente indignação do público:

> Eles não só são deixados a falecer em agonia, ignorados e abandonados, ainda que desesperadamente agarrando o cirurgião sempre que ele faz as suas rondas pelo navio fétido, como agora, quando são colocados no [hospital], onde fomos levados a crer que tudo o que podia aliviar-lhes a dor ou facilitar-lhes a recuperação estava pronto, descobre-se que faltam os equipamentos mais vulgares necessários ao funcionamento de uma enfermaria.[12]

Sidney Herbert, secretário da Guerra e um amigo próximo da família Nightingale, estava sob enorme pressão. Ele tinha assistido

à ascensão rápida de Florence no campo da enfermagem e abordou-a com uma proposta. Consideraria ela liderar um grupo de enfermeiras em Scutari financiadas pelo governo, para auxiliarem os médicos e cuidarem dos homens em sofrimento?

Florence aceitou prontamente e preparou-se para o pior. Mas nada podia tê-la preparado para as condições que encontrou quando chegou: 6,4 quilómetros de corredores cheios de homens grotescamente feridos a dormirem a 45 centímetros uns dos outros, as suas vidas pauperizadas por «ar viciado e males evitáveis». Além disso, a rede de abastecimento do hospital tinha colapsado completamente. Florence não conseguia encontrar tecido para fazer ligaduras, nem camisas lavadas para substituir as ensopadas em sangue. Havia bastante «gangrena, piolhos, insetos e pulgas», mas não havia «nenhuma esfregona, pratos, tabuleiros de madeira, chinelos [...] facas ou garfos, tesouras (para cortar o cabelo dos homens, que está literalmente vivo), bacias, toalhas, cloreto de cálcio». Ela depressa percebeu que os requerimentos para suprimentos tinham de passar por oito diferentes departamentos governamentais em Londres — e que quando estes requerimentos eram finalmente processados, frequentemente enviavam os suprimentos errados, ou então os fornecimentos certos eram enviados para o sítio errado. Na própria Scutari, Florence encontrou apenas preguiça e obstrução por parte do provedor-geral. A situação era tão má que ela pediu ao *The Times* para que lhe confiasse os donativos que tinha recolhido para um fundo de soldados, de modo a que pudesse contornar o provedor e comprar os produtos essenciais no Grande Bazar de Constantinopla.[13] Em seguida, ela efetivamente tornou-se o provedor-sombra do

hospital, como o principal canal para a enorme variedade de ofertas que cidadãos comuns enviavam para Scutari — alimentos, dinheiro, tecidos, chinelos, lavandaria portátil... até geleias de framboesa e biscoitos de gengibre, de uma Sra. Gollop, de Buckinghamshire.[14]

Embora Florence fosse uma enfermeira dotada, os seus talentos brilharam ainda mais enquanto administradora. Ela começou modestamente, implementando novos padrões de limpeza, mas rapidamente deu por si encarregada de reorganizar praticamente todas as funções não médicas do hospital. Florence descreveu o seu papel como «cozinheira, governanta, coletora, lavadeira, revendedora-geral, guarda de loja».[15] O esforço esgotou-a até aos ossos. Trabalhava 20 horas por dia e comia as refeições em pé. Estava exausta pela «quantidade de escrita, a quantidade de conversa [...) o lidar com os egoístas, os maldosos». Ela sentia-se «como Prometeu», amarrada «ao rochedo da ignorância [e] incompetência».[16]

Contudo, durante todo esse tempo ela estava a fazer a diferença. Apenas dois meses após a sua chegada, o capelão do hospital notou uma surpreendente «atmosfera de conforto e satisfação». Havia um fogão em cada enfermaria e banheiras de metal em cada esquina. Cada homem tinha uma cama, um colchão limpo e uma muda de camisa duas vezes por semana.[17] E a mortalidade estava a diminuir: depois de ter atingido um pico de 52 por cento das admissões no inverno de 1855, caiu para 20 por cento em março e em seguida continuou a baixar durante o inverno seguinte, quando não estava mais elevada do que a taxa entre civis numa grande cidade.[18]

Florence dificilmente assumiria ela própria todo o crédito por isto, e nunca o tentou.[19] Ainda assim, durante mais de um ano, a operação médica em Scutari fora um navio que mal conseguia sobreviver à tempestade — e, nas palavras do coronel do exército que assistira à situação em primeira mão, «a menina Nightingale [era] a sua única âncora». Os seus colegas recordaram a sua energia, o seu exemplo, a sua maneira de cortar a burocracia com um machado. Eles recordaram os dias mais sombrios de inverno, quando tropas feridas chegavam às centenas e «os oficiais perderam a cabeça gritando para Flo» pedindo isto e aquilo.[20] E recordaram o caos que reinava durante as suas curtas ausências — como aquele dia em 1854 em que ela fez uma pequena pausa para descansar dos seus deveres enquanto fornecedora não oficial, quando todos os homens do corredor C ficaram bêbados, tendo emborcado o vinho diretamente da garrafa, dado que ninguém lhes tinha fornecido copos.[21]

2. O Legado de Ciência de Dados de Florence Nightingale

De volta à Grã-Bretanha, um jornalista do *The Times* veiculou a imagem de Florence Nightingale que iria perdurar para sempre: «Quando à noite já todo o pessoal médico se retirou e o silêncio e a escuridão assentam sobre aqueles quilómetros de doentes prostrados, ela pode ser observada sozinha, com uma pequena lâmpada na mão, a fazer as suas rondas solitárias.»[22] Com o tempo, a sua lenda apenas cresceu. Canções e poemas sentimentais foram escritos acerca dela. Os diários privados dos soldados registavam devaneios de saltar em auxílio dela perante o perigo. Navios,

cavalos de corrida e bebés de todas as classes sociais foram batizados em honra dela.[23]

Mas para Florence, esta reputação não era senão «falsa popularidade, baseada na ignorância».[24] Ela acreditava que o seu trabalho em Inglaterra, muito depois de a guerra ter terminado, em última análise fez uma diferença muito maior — e os historiadores modernos concordam amplamente com ela. Desse período destacam-se três legados significativos, tornados possíveis pela experiência e a fama que ganhou durante a Guerra da Crimeia.

2.1 A Dama da Lâmpada

O primeiro legado de Florence foi um símbolo vivo de reforma na enfermagem. Antes dela, o símbolo da enfermagem vitoriana tinha sido a Sra. Sara Gamp, a caricatura selvagem de uma cuidadora doméstica de Martin Chuzzlewit, de Dickens. Sem treino, malcriada e perpetuamente bêbada, a Sra. Gamp exalava «uma fragrância peculiar [...] como se uma fada de passagem tivesse soluçado e previamente tivesse estado numa adega de vinhos». O olhar habitual dela, de acordo com Dickens, era «um olhar de soslaio resultante de uma mistura de doçura e malícia [...] parcialmente espiritual, parcialmente espirituoso e totalmente profissional».

A Sra. Gamp era um estereótipo, mas a sua imagem tornou-se tão icónica que deve ter ressoado nos contemporâneos de Dickens, para quem a Sra. Gamp se tornou um símbolo do estado vergonhoso da enfermagem. Um proeminente médico chamado Edward Henry Sieveking escreveu em 1852: «Que os termos de

enfermeira e bebedor de gin não mais sejam convertíveis; vamos banir as Sras. Gamps até ao limite das nossas forças; e substituí-las por cuidadores de doentes que sejam limpos, inteligentes, bem-falantes.»[25]

No seguimento dos feitos de Florence, a imagem pública de uma enfermeira foi transformada em mais ou menos a noção moderna que temos hoje. Isto apenas poderia ter acontecido em resultado de um período de várias décadas de reforma na formação e certificação de enfermeiras, e Florence não foi de todo a primeira defensora destas reformas. Ela inspirou-se em muitos pioneiros anteriores — especialmente as enfermeiras que geriam Kaiserswerth, onde ela recebeu formação no início da década de 1850. Não obstante, para o público britânico, Florence passou a ser o símbolo da enfermeira vitoriana moderna. Ela fez mais do que qualquer outra pessoa para transformar a enfermagem numa via respeitável para mulheres da classe média, ajudando dessa forma a criar um círculo virtuoso em que melhores enfermeiras contribuem para uma melhor profissão, atraindo enfermeiras ainda melhores.

2.2 A Estatística Entusiasmada

O segundo legado de Florence foi a sua análise pessoal de estatísticas médicas da Guerra da Crimeia. Florence regressou a Inglaterra cheia de justa indignação contra o escândalo de Scutari. No seu diário, ela escreveu: «Eu permaneço no altar dos homens assassinados e enquanto viver lutarei pela sua causa.»[26] Foi uma luta contra aqueles, no Exército e na comunidade médica, que

permaneciam implacáveis contra a mudança — como o médico do exército John Hall, por exemplo, que desvalorizou Florence considerando-a um «saiote autoritário».[27] E ela trouxe todas as suas armas para usar nessa luta: o seu intelecto, a sua rede de amigos, a sua escrita ácida... e acima de tudo, matemática e estatística, que ela via como as flechas mais poderosas na sua aljava.

O primeiro biógrafo de Florence, E. T. Cook, apelidou-a como «a estatística entusiasmada» — o que, por razões óbvias, não captou a imaginação do público como «a dama da lâmpada», mas proporcionou uma descrição muito melhor de como ela mudou o mundo para melhor. Florence era particularmente exímia no uso de representações gráficas de dados — «visualização de dados», no linguajar moderno — para chamar a atenção da pátria para as condições vergonhosas que tinham prevalecido nos hospitais militares. Como referiu um dos seus colegas, as imagens de dados de Florence conseguiam «influenciar através dos olhos aquilo que podíamos não conseguir transmitir aos cérebros do público através dos seus ouvidos à prova de palavras». Ela até inventou um novo tipo de figura estatística: a área polar ou diagrama «coxcomb», que mostrava alterações na mortalidade ao longo do tempo usando uma série de cunhas coloridas. O seu diagrama de área polar da Guerra da Crimeia, representando o aumento e a descida da mortalidade devidos a doença, é apresentado na Figura 14.

A análise dela revelou que nos primeiros sete meses da campanha na Crimeia, os soldados ingleses sofreram uma taxa de mortalidade de 60 por cento somente devida a doenças. Isto foi

mais elevada do que a que os londrinos experienciaram durante a Grande Praga de 1665, mais elevada ainda do que a probabilidade de em 1850 um civil morrer após contrair cólera.[28] Sim, era literalmente mais seguro ter cólera no próprio país do que aventurar-se na Crimeia — e isso antes mesmo de enfrentar uma única bala inimiga. Florence referiu-se a isto como «a maior experiência que a história moderna alguma vez vira [...] em relação a qual o número dos que serão mortos arbitrariamente pela entidade única de má alimentação e ar viciado» — uma experiência que condenou à morte 16 mil homens.[29]

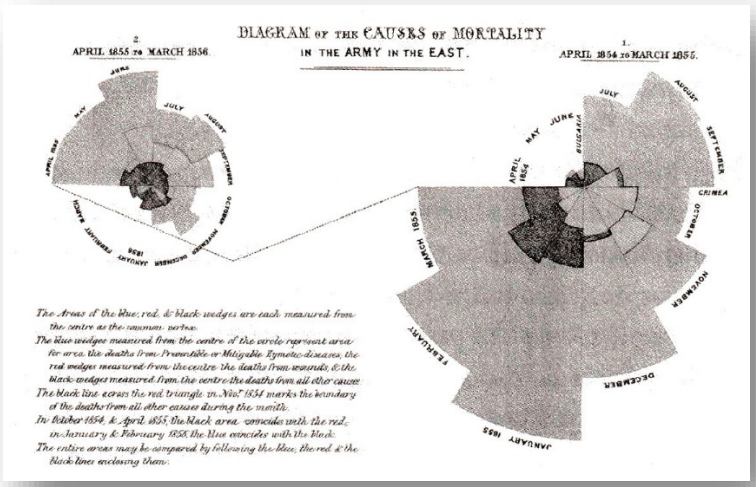


Figura 14. Diagrama de área polar de Nightingale de 1858. No círculo da direita, as 12 cunhas à volta do centro representam as mortes mês a mês na Guerra da Crimeia devidas a «doenças zimóticas evitáveis ou mitigáveis» de abril de 1854 a março de 1855. A linha tracejada leva-o depois ao círculo do lado esquerdo, que mostra os dados do ano seguinte: de abril de 1855 até março de 1856. Em cada círculo, os dois conjuntos de cunhas mais interiores representam mortes devidas a ferimentos de combate (a preto) e a todas as outras causas (cinzento-claro).

Ela também analisou estatísticas do tempo de paz e descobriu que, devido a más condições sanitárias, a taxa de mortalidade do Exército no seu país era o dobro da mortandade da população civil equivalente. Ela considerou esta situação «criminosa», nada

diferente de «levar 1100 homens para a planície de Salisbury e fuzilá-los».[30] Isto envergonhou o Exército, levando-os a reapetrechar os quartéis e a redesenhar os hospitais, o que originou uma queda imediata na mortalidade relacionada com doenças.[31] As recomendações dela depressa chamaram a atenção do mundo civil. A incansável argumentação de Florence contribuiu, e não de forma despicienda, para que os hospitais com corredores longos e quartos abafados passassem a ser vistos como incubadoras de infeções. O seu modelo preferido de construção de hospitais logo se tornou a norma: o hospital ao estilo pavilhão, com luz e ventilação abundantes, e com alas separadas para controlar a propagação de doenças. Estas «enfermarias Nightingale» permaneceram populares durante uma boa parte do século XX.[32]

2.3 A Mãe da Medicina Baseada em Evidências

O terceiro legado de Florence talvez seja o menos conhecido: o seu papel em criar um novo padrão de profissionalismo na recolha e análise dos dados médicos.

Costuma-se dizer acerca dos generais que estão sempre a travar a última guerra. Mas um médico do Exército à procura de lições na enorme variedade de experiência médica da Guerra da Crimeia não teria sido sequer capaz de fazer isso. Não foi recolhida qualquer estatística, poucas histórias clínicas foram preservadas e quase nenhum exame *post mortem* foi realizado. Em muitos casos, os homens doentes eram carregados de um lado do navio na Crimeia, apenas para serem atirados borda fora pelo outro lado quando chegavam a Scutari. Florence desesperava com o destino dos

homens, mas ela também achava «desencorajante e dececionante ao extremo» que um tal «tesouro científico» tenha sido perdido devido a má gestão.[33]

Depois de regressar a Inglaterra após a guerra, Florence descobriu que estas lacunas estavam reproduzidas no mundo civil. O país não tinha qualquer sistema para a recolha das estatísticas médicas, até mesmo as mais básicas, como restabelecimentos, duração de internamentos, ou a mortalidade de diferentes doenças. Mesmo que tivesse existido tal sistema, não haveria forma alguma de comparar resultados entre hospitais diferentes, os quais todos usavam sistemas de classificação idiossincráticos para as doenças.[34]

Florence via esta falta de atenção aos dados como uma emergência de saúde pública. Ela viu como a nova disciplina da estatística estava a transformar outros campos, como a astronomia e as ciências da Terra. Também reparou como os estatísticos continentais — nomeadamente o eminente belga Adolphe Quetelet, um dos seus ídolos — estavam a usar ferramentas novas para lidar com questões complexas da ciência social acerca de crime e alterações demográficas. Florence viu um potencial incrível em aplicar essas mesmas técnicas estatísticas aos cuidados de saúde. Isso «iria permitir-nos poupar vidas e sofrimento, e melhorar o tratamento e a gestão dos doentes»[35], o que exigia dados muito melhores por parte do sistema de cuidados de saúde. Para esse efeito, ela concebeu um conjunto padronizado de formulários médicos, obteve o apoio de muitos dos melhores estatísticos do mundo e instou os grandes hospitais de Londres a começarem a

usá-los. Também pressionou o governo para que passasse a recolher dados sobre doença e qualidade da habitação como parte do censo, argumentando que «a relação entre a saúde e os alojamentos da população é uma das mais importantes que existem»[36]. No seu conjunto, o trabalho de Florence claramente prenunciou os 160 anos seguintes de cuidados de saúde baseados em evidências. As suas ideias formaram um modelo claro para o sistema internacional de classificação de doenças usado atualmente, que serve de pedra angular para toda a epidemiologia moderna e ciência de dados médicos.[37]

3. Males Evitáveis na Era da IA

Os três legados de Florence têm paralelismos atuais evidentes. E também levantam questões pertinentes. Ela falou do «ar viciado e males evitáveis» que mataram os soldados da Crimeia e, embora o ar nos hospitais modernos possa ser menos viciado, ainda existem males em abundância.

Uma questão importante é como preencher e treinar uma equipa moderna de cuidados de saúde. Depois de Florence, nenhum hospital podia funcionar sem enfermeiras. Quando é que o mesmo será verdade em relação aos cientistas de dados e aos peritos em inteligência artificial, cujo papel no dia a dia dos cuidados de saúde é quase nulo?

Uma segunda questão é como planear um hospital para esta nova Era. Florence ajudou a estabelecer novos padrões sanitários,

e em resposta os hospitais foram reformulados de raiz. Quando é que os hospitais passarão por outra época de reconceção, para acomodar o que é agora possível usando a inteligência artificial? Quando é que a higiene dos dados será levada tão a sério como a higiene dos doentes?

Por último, a questão mais importante de todas é como devem as estatísticas médicas ser recolhidas, partilhadas, analisadas e usadas. Nos últimos 160 anos melhorámos muito nesse segmento, em grande parte devido aos esforços de Florence. Contudo, como em breve irá aprender, apenas melhorámos em certos aspetos — e podíamos estar a fazer muito mais. À luz do que aconteceu fora dos cuidados de saúde, isto começa a parecer-se com um embaraço moral. Vivemos numa época em que os carros de Fórmula 1 são monitorizados em tempo real por algoritmos e equipas de engenheiros, em que as suas preferências cinematográficas são alvo da atenção de operações multimilionárias de IA, e em que a sua propensão para clicar num anúncio de comida para cão é analisada em supercomputadores, usando milhões de variáveis e milhares de milhões de pontos de dados. Todavia, para a maioria das coisas, continuamos a depender de números que Florence Nightingale podia ter processado com papel e caneta para quantificar o risco de os seus rins falharem. E nalguns aspetos não melhorámos de todo: um artigo de 2017 no *Journal of the Royal Statistical Society* referiu-se ao protocolo de Florence de 1860 para recolha de dados hospitalares como «concetualmente mais completo» do que muitos sistemas atuais.[38] O que nos deve deixar a todos a pensar: quando irá a ciência de dados médicos entrar no século XXI?

3.1 Razão para Precisarmos da IA nos Hospitais

Queremos desde logo ser claros que isto não é culpa de médicos ou enfermeiros individuais. É culpa de todo o sistema de cuidados de saúde, que durante demasiado tempo tem sido a Sra. Gamp da ciência de dados: atrasado nas suas normas estatísticas, embriagado com burocracia e ignorante do que a Inteligência Artificial transformou em banalidade.

Para ilustrar este ponto, gostaríamos de contar-lhe a história de um homem algures na Costa Leste dos Estados Unidos — vamos chamar-lhe Joe — que morreu aos 62 anos com doença crónica dos rins. A história do Joe explica bastante como a abordagem contemporânea à ciência de dados médicos está a falhar aos doentes e por que uma combinação de uma melhor curadoria dos dados com a IA podia evitar tanto sofrimento.

Com 40 e poucos anos, o Joe já sofria de diabetes tipo 2 e insuficiência cardíaca congestiva. Talvez o seu trabalho fosse stressante, ou a sua dieta e os hábitos de exercício fossem insuficientes. Qualquer que fosse a combinação de causas, elas finalmente tiveram o seu efeito. Poucas semanas antes do seu quadragésimo sétimo aniversário, o Joe sentiu uma dormência súbita no braço direito, tropeçou e caiu pesadamente no chão. Foi levado para as urgências, e de imediato foi-lhe diagnosticado um acidente vascular cerebral isquémico, o que significava que um coágulo tinha bloqueado o fluxo de sangue para o cérebro.

Felizmente, o Joe sobreviveu ao AVC. Embora a sua pressão arterial elevada e a diabetes o sinalizassem como tendo um risco mais alto de doença renal em algum momento no futuro, por agora o exame aos seus rins estava normal. A medida-padrão da função renal é a TFG, ou taxa de filtração glomerular. A TFG do Joe foi estimada em 99, bem acima da zona de perigo: uma TFG de 60 ou menos indica uma perda de ligeira a moderada da função renal, enquanto 30 ou menos significa perda severa.[39]

Ao longo do ano seguinte, o Joe fez mais nove visitas às urgências devido a enfermidades de um tipo ou de outro, nenhuma das quais relacionada expressamente com os seus rins. Em duas dessas ocasiões, ele foi internado no hospital, e a sua função renal foi medida: primeiro a sua TFG foi 96, e depois 95 cerca de um mês mais tarde. Este declínio foi um pouco mais abrupto do que a taxa de um a dois por cento ao ano que seria de se esperar numa pessoa saudável. Ainda assim, cada leitura individual estava acima do limiar clínico de 60 que habitualmente preocupa os médicos.

Cerca de um ano após o seu AVC, o Joe começou a ir regularmente a uma clínica em regime de ambulatório: oito visitas em 14 meses. Em cada visita, o médico pedia uma série rotineira de testes, e os funcionários da clínica introduziam corretamente os dados da função renal do Joe numa base de dados eletrónica — a mesma utilizada pelos médicos no hospital. Os seus valores de TFG alternaram ligeiramente um pouco entre 60 e 75 — ainda acima do limiar de 60, mas bastante abaixo da leitura de 99 no ano anterior, e numa tendência descendente inequívoca.

Aos 49 anos, o Joe foi novamente internado no hospital, e a sua TFG era de 54. Ao longo dos meses seguintes fez mais 10 visitas às urgências, bem como mais 12 visitas à clínica. O Joe estava agora muito doente. Um mês antes do seu quinquagésimo aniversário, a sua TFG era 40, bem dentro da zona de perigo. No entanto ele não recebeu qualquer tratamento que pudesse evitar a sua derrapagem para a falência renal. Só podemos especular sobre o motivo, mas uma razão talvez seja que por vezes os resultados dos testes demoram a regressar do laboratório — e nessa altura o doente já poderá estar em casa, fora do cuidado direto do médico que inicialmente pediu o exame.

Ao longo dos três anos seguintes, o Joe teve mais 20 consultas com um médico. Em muitas destas ocasiões a sua função renal foi medida, e estava a cair a um ritmo assustador: abaixo de 30 aos 51 anos, e abaixo de 20 aos 52 anos, após o que o Joe foi finalmente encaminhado para um especialista em rins, mais de um ano depois de os seus valores de TFG terem caído abaixo do nível que tipicamente desencadeia esse encaminhamento.

Mas a falência renal era agora inevitável. Três meses após a sua consulta com o especialista, os rins de Joe finalmente cederam. Ele foi prontamente levado para as urgências, a sua vigésima quinta visita desde o primeiro AVC. A sua TFG foi de 12; a função renal tinha decaído 34 por cento ao ano, em cada um dos últimos cinco anos, a partir da leitura inicial de 99 após o AVC. Os médicos das urgências colocaram-no em diálise de emergência, um dos procedimentos previstos mais traumático e dispendioso.

Durante a década seguinte, o Joe tornou-se naquilo ao que a indústria dos seguros se refere como um «superutilizador», o que é na linguagem de gestão um ser humano terrivelmente doente — um dos cinco por cento de doentes que nos Estados Unidos são responsáveis por mais de 50 por cento de todos os gastos em cuidados de saúde. No caso do Joe isto significava diabetes severa, doença renal de estádio 5, angina, doença vascular e doença inflamatória do tecido conjuntivo, a par de uma série de ataques cardíacos. Os rins do Joe foram examinados 124 vezes neste período, o que incluiu mais 26 visitas às urgências e nove a um nefrologista. A sua TFG andou para cima e para baixo, mas nunca mais subiu além de 20.

Ele morreu a uma semana de completar 63 anos, cerca de dez anos após ter começado a diálise.

O Joe morreu de quê? Num certo sentido, a resposta é clara: os rins falharam. Mas para isso ter acontecido, outra coisa deve ter falhado primeiro, de maneira tão completa que custa a acreditar. Pois se pegarmos em todas as leituras da TFG do Joe nos oito anos após o seu AVC e as representarmos graficamente ao longo do tempo, a tendência parece bastante óbvia (veja a Figura 15).[40]

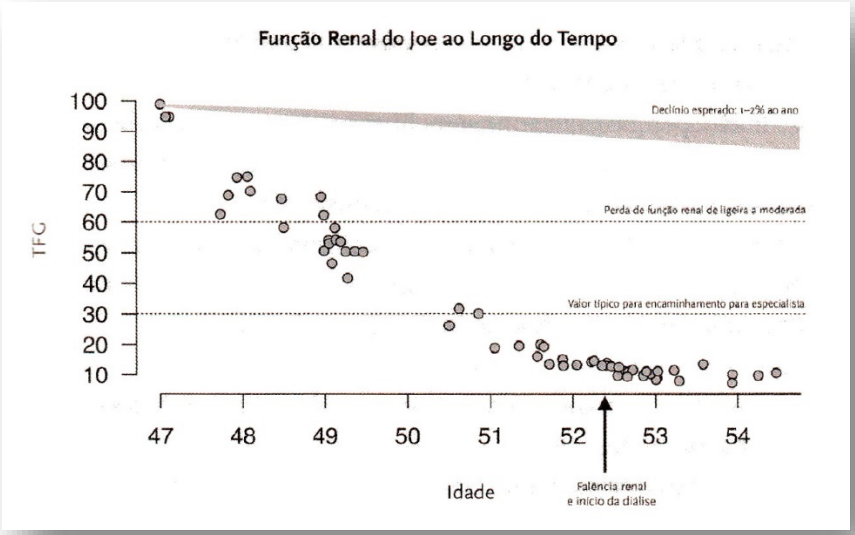


Figura 15

Naqueles três anos de declínio abrupto entre os 47 e os 50 anos, nenhum dos prestadores de cuidados de saúde do Joe olhou para uma representação simples das suas leituras de TFG ao longo do tempo. Era uma questão, muito literalmente, de ligar os pontos. Fazê-lo teria produzido uma previsão simples e óbvia: a função renal deste sujeito está a decair tão depressa que provavelmente irá continuar a baixar, e se isso acontecer o resultado será doloroso e dispendioso. O Joe morreu certamente por falta de um rim. Mas, a nível mais fundamental, ele morreu por falta de um gráfico de dispersão.

3.2 Pensamento Limiar

Como pode isto ter acontecido? Colocámos esta questão a uma professora de estatística e aprendizagem automática, Dra. Katherine Heller, da Universidade de Duke, que analisou estes dados e inicialmente nos chamou a atenção para o caso do Joe. «Em retrospetiva», disse Heller, «esse declínio abrupto entre os 47 e os

50 anos representa uma óbvia oportunidade perdida. Tudo o que temos de fazer é traçar uma linha reta através da nuvem de pontos de dados, e podemos ver para onde as coisas se encaminham.»

Então, por que é que ninguém, humano ou máquina, desenhou a tal linha reta? Esta é a questão essencial nos cuidados de saúde modernos. Para entender a resposta temos de revisitar duas questões anteriores que Florence Nightingale colocou há 160 anos, quando refletiu sobre como as novas ferramentas matemáticas da década de 1850 poderiam ser usadas em hospitais:

1. Atualmente, como é que o sistema de cuidados de saúde utiliza os dados?

2. Em alternativa, à luz das novas tecnologias de análise de dados, o que poderia ter sido feito?

Hoje, a maneira principal como o sistema de cuidados de saúde usa os dados é para criar listas de verificação. Estas listas codificam os «padrões de cuidados de saúde» recomendados por entidades nacionais, como a Associação Médica Americana ou o Conselho Médico Geral do Reino Unido. Estes padrões de cuidados, por sua vez, são em última análise guiados por dados de descobertas de investigações publicadas: acerca de quais os sinais de alerta a ter em atenção, quais os tratamentos que realmente funcionam e quais os protocolos de diagnóstico que ajudam mais pessoas. Por exemplo: talvez se lembre de que a Sociedade Americana contra o Cancro criou alguma controvérsia em 2015 quando atualizou a sua

lista de verificação recomendada em relação a mamografias. De acordo com a nova lista de verificação, mulheres com risco médio de cancro da mama deviam começar a efetuar mamografias anuais aos 45 anos, em vez de aos 40. Essa alteração foi apenas feita depois de uma equipa de 19 especialistas ter realizado uma síntese maciça de todos os dados disponíveis e concluído que as novas recomendações possivelmente evitariam 11 falsos positivos em cada 500 mulheres rastreadas, sem qualquer impacto discernível no número de mortes por cancro da mama.[41]

As listas de verificação médicas são fantásticas e a maneira como são criadas e atualizadas representam um triunfo dos dados sobre as historietas — algo que deixaria Florence Nightingale, se ainda fosse viva, imensamente orgulhosa. As listas de verificação salvam vidas ao ajudarem os médicos a detetarem sinais subtis quando tomam decisões complexas. Inspirado pelas suas experiências enquanto cirurgião, o escritor e médico Atul Gawande até escreveu *O Efeito Checklist: Como Aumentar a Eficácia*, acerca de como as listas de verificação podem ajudar a tomar decisões complexas em todo o lado, não apenas na medicina. E apresenta bons argumentos.

Mas as listas de verificação podem falhar — especialmente quando dependem daquilo a que Katherine Heller chama «pensamento limiar». Para ver isto vamos voltar à tendência que é tão óbvia no diagrama de dispersão das leituras do rim do Joe que lhe mostrámos anteriormente. Heller supõe que cada médico ao longo daquele triste caminho de pontos estava a pensar no caso do Joe em termos de um limiar binário numa lista de verificação. Está

a TFG do paciente acima de 30? Sim. Estão os seus níveis de potássio no sangue abaixo de 5,5 milimoles por litro? Sim. Os seus níveis de albumina na urina são normais? Os outros indicadores relacionados com os rins estão nos intervalos esperados? Terei seguido o protocolo? Sim, sim, sim.

Todos aqueles «sins» dizem-nos algo acerca da função renal do Joe naquela visita isolada e são realmente importantes para prestar bons cuidados de saúde. Mas aquelas verificações nada dizem acerca da tendência a longo prazo. Assim, mesmo que o Joe tenha estado a encaminhar-se durante anos para aquele terrível limiar de 30 da TFG, ele ainda não o tinha ultrapassado — e ninguém acionou o alarme até ser demasiado tarde. Em retrospetiva, isto não devia ser surpreendente. Em termos de IA, as listas de verificação são apenas regras preditivas: procedimentos que utilizam os dados do paciente como *input* e produzem uma decisão clínica como *output*. Enquanto regras preditivas, contudo, elas são concebidas para ajudar os médicos a compreenderem e a responderem ao que está a acontecer no *imediato*, não ao que é provável acontecer no futuro. Aliás, essa é uma caraterística de design inerente às listas de verificação: elas centram a mente do médico nos detalhes do presente. Mas num mundo onde os maiores e mais dispendiosos problemas médicos são as doenças crónicas que se desenvolvem num período de anos, essa caraterística começa a parecer uma falha.

Poderá perguntar: por que não corrigir simplesmente a falha com uma lista de verificação mais longa, adicionando um item que encoraje os médicos a procurar tendências a longo prazo? Nós

tínhamos a mesma questão. De modo que perguntámos a Katherine Heller se teria sido de algum modo possível, para alguém à cabeceira de Joe, colocar as suas leituras de TFG num ecrã e dispô-las ao longo do tempo para procurar uma tendência. «Talvez pudesse consultar a base de dados dessa maneira, se soubesse como fazê-lo». disse ela após alguns momentos de reflexão. «Mas certamente não seria uma maneira natural e óbvia para um médico usar o sistema.» Para ver a tendência, prosseguiu, «teria mesmo de recuar no registo manualmente, uma leitura de cada vez». Ironicamente, isto provavelmente teria sido mais fácil nos tempos dos relatórios em papel.

Além disso, salientou Katherine, não se trata apenas de um conjunto de leituras para verificar mas centenas ou até milhares: análises ao sangue, análises de urina, eletrocardiogramas, frequência cardíaca, pressão arterial, sintomas clínicos, fatores sociais — e em breve informações acerca da expressão genética do paciente e perfil epigenético. Simplesmente... *há imensos dados*. Para um ser humano é difícil assimilar tudo mesmo como um retrato único, muito menos como uma história que se desenvolve ao longo do tempo.

Por último há a questão de como é que, na lista de verificação, um tal hipotético item de «procurar tendências» encaixaria no habitual processo de trabalho do médico. Quando o leitor aparece nas urgências, a principal preocupação do seu médico é: qual a gravidade da sua situação neste momento? Deve ser tratado e enviado para casa, ou está suficientemente doente para ser internado? Os médicos enfrentam riscos elevados e enorme pressão

quando tomam essas decisões — e mesmo fora das urgências, numa clínica normal, têm de tomá-las *depressa*, porque há dezenas de outras pessoas na sala de espera que também precisam de ajuda. Será razoável esperar que esses médicos interrompam o que estão a fazer, iniciem um pacote de *software* estatístico e explorem um vasto conjunto de dados eletrónicos de saúde, tudo para encontrar uma ou duas tendências no historial que possam ser relevantes para prevenir algo daqui a uns meses ou anos?

O Dr. Mark Sendak, do Instituto para a Inovação na Saúde, da Universidade de Duke, explica que embora os médicos possam fazer este tipo de coisa em programas de televisão como o Dr. House, não o fazem em hospitais normais. «Os médicos dizem sempre que querem os dados», diz Sendak. Mas, continua ele...

> O problema é que não existe qualquer processo de trabalho para eles acederem ou usarem os dados. Da maneira como os registos estão estruturados é preciso tempo e competência. Tem de se escrever uma pesquisa, tem de se transferir os dados para uma folha de cálculo, e depois tem de efetivamente fazer-se coisas com isso. Mas os médicos já estão sob imenso stress. Eles têm consultas de 15 minutos na clínica. Quando, exatamente, é que vão estar a brincar com os dados da sua clínica e perceber o que precisam de fazer pelos seus pacientes?

Isso traz-nos a um assunto ainda mais profundo: todo o sistema de ciência de dados médicos foi concebido apenas para lidar com questões ao nível de uma população. Por exemplo: quantas vidas salvaríamos se usássemos o limiar A em vez do limiar B para detetar a doença renal? Deve haver centenas de estudos relacionados com uma qualquer questão desse tipo. Mas a ciência de dados médicos está praticamente silenciosa em resposta a questões estatísticas básicas ao nível do paciente individual. Como é que as leituras da

TFG do Joe variam a longo prazo? Para onde é que é provável que vão a partir daqui? O que prediz isso para a saúde do Joe no próximo mês, ou no próximo ano? Estas questões teriam sido fáceis para uma pessoa ou algoritmo responder usando o registo do historial médico do Joe — contudo, todos aqueles pontos de dados nunca tiveram uma hipótese de falar. Não havia qualquer rotina implantada para esmiuçar os registos de saúde do Joe em busca de sinais de uma doença crónica subjacente: nenhuma equipa de cientistas de dados, nenhum algoritmo, nenhum médico com formação interdisciplinar em estatística.

Com algumas exceções aqui e ali, o mesmo é válido para a maioria dos hospitais e clínicas. Ao conversarmos com amigos e colegas sobre este tópico, reparámos que muitas pessoas têm a ideia de que deve haver algum tipo de «carro-robot» nos bastidores de um hospital moderno — algum conjunto sofisticado de algoritmos que analisa registos ao nível do paciente e ajuda os médicos a efetuarem recomendações personalizadas e a tomarem decisões. Talvez eles fiquem com esta ideia por verem os seus próprios médicos a efetuarem tanta inserção de dados ou por verem a IA a transformar tantas outras indústrias. Qualquer que seja a razão, geralmente ficam chocados quando lhes dizemos a verdade: que no que diz respeito à análise de dados ao nível do paciente, na maioria dos hospitais atuais não só não existe carro-robot algum, como não há literalmente *alguém ao volante*.

Quando falámos com Katherine Heller, a frustração dela acerca deste assunto era óbvia. «Acontece que não é suficiente recolher simplesmente todos aqueles dados», ironizou ela. «Temos

efetivamente de fazer algo com eles.» Aqui ela inconscientemente canaliza Florence Nightingale, que em 1859 escreveu sobre o hospital St. Thomas em Londres dizendo que «parece manter as suas estatísticas mais com o intuito de controlar pacientes irascíveis, o que é certamente um objetivo, mas não para um fim científico».[42]

A história do Joe, pelos vistos, é muito mais do que a saga de um homem com doença nos rins. É uma história do vasto desfiladeiro entre o que os dados *podiam fazer* por nós e o que o nosso sistema de cuidados de saúde deixa que façam.

4. A IA Vai em Socorro?

Se lhe parece que os profissionais de cuidados de saúde estão a afogar-se em dados e precisam mesmo de um colete salva-vidas — que uma combinação de inteligência humana e automática podia melhorar radicalmente os cuidados de saúde —, então não está sozinho nesse pensamento. Empresas e investigadores estão a trabalhar arduamente numa nova geração de tecnologias baseadas em IA que aguardam uma oportunidade, prontas para ajudar médicos e enfermeiros a realizarem os seus trabalhos de maneiras mais eficientes.

A equipa da Dra. Katherine Heller em Duke, por exemplo, aliou-se a médicos para desenvolver um sistema de IA que pode sinalizar sinais de iminente doença renal crónica.[43] No centro do sistema deles está uma regra preditiva, tal como as que encontrámos no

Capítulo 2: ela analisa o historial de leituras de TFG do paciente, combina-o com os dados provenientes de outros testes laboratoriais e sinais vitais, e faz uma previsão para a trajetória futura da função renal desse paciente. Essa previsão é apresentada numa aplicação móvel à qual os médicos podem aceder enquanto tratam o doente. Com este tipo de IA, os médicos *efetivamente* podem perceber a parte de tendência a longo prazo da sua lista de verificação, sem terem eles próprios de vasculhar os dados.

Outros grupos de investigação inventaram sistemas de alerta precoce similares para outras afeções — para a paragem cardíaca, a depressão, o sofrimento fetal durante o parto e as infeções adquiridas em ambiente hospitalar, apenas para citar algumas. Em breve, outros avanços igualmente impressionantes na tecnologia de IA podem revolucionar todas as áreas da medicina, da radiologia ao tratamento do cancro e à dermatologia. Antes de regressarmos à questão de quais as alterações culturais que devem ocorrer antes que vejamos uma aplicação generalizada da IA nos cuidados de saúde, iremos mergulhar um pouco mais fundo na linha da frente do setor.

4.1 Equipamentos Médicos Inteligentes

A lâmina eletrocirúrgica, que utiliza ondas de rádio de alta frequência para aquecer o tecido até ao ponto em que é vaporizado, é uma enorme evolução em relação aos bisturis de antigamente. A lâmina permite cortes drasticamente mais precisos e, como cauteriza o tecido circundante quase de imediato, também minimiza a perda de sangue. Mas não se pode esperar que mesmo a lâmina

mais requintada consiga ajudar o cirurgião a saber *onde* cortar. Quando os cirurgiões oncológicos removem um tumor, por exemplo, eles frequentemente consideram ser impossível determinar a olho nu com exatidão onde o tumor acaba e o tecido saudável começa.

Espantosamente, uma nova lâmina inteligente baseada em IA, desenvolvida pelo Dr. Zoltan Takats e pela sua equipa no Imperial College London, pode em breve ajudá-los. Quando o tecido é vaporizado por uma lâmina eletrocirúrgica cria fumo, que habitualmente é sugado por um extrator. Takats, contudo, teve uma perceção inteligente acerca desse fumo: ele deve conter metabolitos provenientes do tecido vaporizado, que podem ser usados para inferir se o tecido era ou não canceroso. Então, ele construiu uma lâmina eletrocirúrgica em que, ao invés, o fumo é redirecionado para um espetrómetro de massa que constrói um perfil químico. Este perfil é então introduzido numa regra preditiva que classifica o fumo como proveniente de células saudáveis ou de células cancerígenas. Além disso, este processo de quatro passos — vaporizar, extrair, perfilar, classificar — acontece em três segundos ou menos. Desta forma, a nova lâmina pode efetivamente dizer aos cirurgiões onde parar de cortar. E num ensaio que envolveu pacientes cirúrgicos reais, o software de IA da lâmina identificou corretamente o tipo de tecido 91 vezes em 91, tal como se verificou pela histologia pós-operatória.[44]

Alguns equipamentos médicos inteligentes podem até ir além da medição e entrar no reino do tratamento automático. Veja-se por exemplo o pâncreas artificial de circuito fechado, um sistema de IA

que mimetiza a função hormonal de um pâncreas real ao fornecer automaticamente aos diabéticos a quantidade correta de insulina em resposta a alterações dos valores do açúcar no sangue. Um pâncreas artificial envolve três passos: medição, dosagem e administração. O passo da medição usa um monitor contínuo de glicose, ou CGM, que proporciona uma medição em tempo real do açúcar no sangue ao longo de 24 horas diárias. O passo seguinte é um algoritmo para calcular dosagens: uma regra preditiva inteligente pega em todos os dados em tempo real da glicose no sangue do CGM e determina a dose apropriada de insulina. O último passo é uma bomba que fornece insulina à medida que é necessária.

Empresas de equipamentos médicos como a Medtronic, a Insulet e a Tandem estão a alcançar progressos rápidos nesta área — e, como bom presságio, os reguladores estão de facto a manter-se a par delas. Em setembro de 2016, por exemplo, após um surpreendentemente rápido período de revisão de três meses, a agência federal americana para a saúde e alimentação (FDA) aprovou o então mais recente modelo da Medtronic, o primeiro pâncreas artificial concebido para equilibrar tanto os altos como os baixos níveis de açúcar no sangue.[45]

4.2 IA para Imagiologia Médica

A imagiologia de diagnóstico oferece um exemplo ainda mais imediato de onde a IA pode fazer a diferença. Muitas formas comuns de análise de imagem médica, desde olhar para um raio-X a examinar células cancerígenas ao microscópio, envolvem um

problema clássico de reconhecimento de padrão: os *inputs* são as caraterísticas extraídas da imagem e o *output* é o diagnóstico. Como vimos no Capítulo 2, os computadores são absolutamente brilhantes a aprenderem a prever *outputs* a partir de *inputs*, especialmente para imagens — e com mais dados e novos algoritmos de reconhecimento de padrões estão a ficar cada vez melhores.

Para alguns diagnósticos guiados por imagens, em breve nem sequer precisaremos de visitar o consultório do médico. Considere, por exemplo, o problema de diagnosticar uma lesão na pele. Aqui os riscos são elevados: o melanoma causa mais de 10 mil mortes por ano só nos Estados Unidos. A taxa de sobrevivência a cinco anos do melanoma é superior a 99 por cento se for detetado cedo, mas o valor cai para 14 por cento se for detetado muito tarde. As pessoas, por várias razões — tempo, dinheiro, uma aversão geral aos médicos —, muitas vezes resistem a ir ao dermatologista até ser demasiado tarde.

Mas em 2017 um artigo de investigação na *Nature*, de uma equipa multidisciplinar de cientistas liderada por Sebastian Thrun, da Universidade de Stanford, descreveu um sistema de IA que em breve poderá permitir o acesso a um pedaço vital de diagnóstico referente à saúde da pele, de forma gratuita, para qualquer pessoa com um smartphone. A equipa de Stanford sabia imenso acerca de algoritmos para reconhecimento de imagens devido ao seu trabalho anterior com carros-robots, e isto deu-lhes uma ideia simples: e se em vez de treinar estes algoritmos para distinguirem um sinal de stop de um aviso de caça grossa, como fazem para um carro-robot,

os treinassem para distinguir diferentes tipos de cancro da pele através de uma simples fotografia?

A sua tentativa de dermatologia assistida por computador não foi a primeira, mas a equipa de Stanford fez três escolhas cruciais que separaram o seu algoritmo de um largo conjunto de outras abordagens bem menos bem-sucedidas. A primeira foi a escala. As tentativas anteriores de fazer este tipo de coisa tinham usado conjuntos pequenos de dados, com menos de mil imagens de lesões da pele. Os investigadores de Stanford compilaram 19 bases de dados contendo 129.450 imagens, cada uma classificada de acordo com uma taxonomia de 2032 lesões da pele diferentes. Mais dados significa uma amplitude maior de experiência e consequentemente melhor reconhecimento de padrões, como um dermatologista veterano que há décadas olha para lesões da pele e que já viu de tudo.

A segunda escolha foi a abordagem à visão computacional, que envolveu as redes neurais profundas que encontrámos no Capítulo 2. Estas redes conseguem extrair caraterísticas visuais subtis e elas conseguem combinar essas caraterísticas em conceitos visuais de alto nível — como círculos, arestas, riscas, textura ou gradações de variegação — que podem ser usados para distinguir 2000 tipos diferentes de lesões da pele. Elas conseguem fazer isto, aliás, sem que um programador alguma vez lhes diga o que procurar.

A última escolha que a equipa de Stanford fez foi usar imagens de câmaras normais, em vez de imagens médicas altamente padronizadas que apenas podem ser obtidas através de uma

biópsia, ou usando equipamento que apenas um dermatologista teria. Estas imagens exibiam grande disparidade em termos de iluminação, oscilação de cor, ampliação e ângulo — fontes de variação irrelevantes que facilmente podem enganar um algoritmo inferior, levando-o a alucinar diferenças que não existem. Mas o que faltava a estas imagens em qualidade eles mais do que compensavam com a quantidade — teria sido muito difícil recolher mais de 129 mil imagens padronizadas numa clínica de dermatologia.

O resultado de todo este trabalho foi um sistema de IA de ponta a ponta que, a partir de uma fotografia comum, pode fazer duas inferências cruciais acerca de uma lesão na pele. Ele consegue distinguir os dois tipos mais comuns de cancro da pele um do outro, e também consegue distinguir um sinal benigno da forma mais letal de cancro de pele: o melanoma maligno. E consegue fazê-lo, inclusivamente, com uma precisão comparável à de um painel de 21 dermatologistas certificados. Em algumas medições, o desempenho do algoritmo de Stanford até se comporta melhor.*

Em breve, técnicas semelhantes de análise de imagens chegarão a todas as áreas da medicina, à medida que novas especialidades como a radiologia computacional e a patologia computacional atinjam a maturidade. Um laboratório de investigação no Instituto Federal de Tecnologia de Zurique, por exemplo, desenvolveu um algoritmo de IA para classificar a severidade da doença inflamatória intestinal a partir de uma IRM abdominal.[46] Outro laboratório no Memorial Sloan Kettering Cancer Center construiu um sistema para classificar o carcinoma das células renais a partir de lâminas de

microscópio digitais.[47] E o Moorfields Eye Hospital em Londres aliou-se recentemente com a DeepMind da Google para analisar mais de um milhão de imagens de exames oculares. O resultado foi uma rede neural capaz de detetar automaticamente sinais de doença ocular, como retinopatia diabética e degeneração macular.[48]

As empresas de hardware também responderam ao grande aumento de procura de imagiologia médica potenciada pela IA. O fabricante de *chips* Nvidia, por exemplo, é principalmente conhecido pelas suas placas gráficas para computador (GPU) de alta qualidade para jogadores e cineastas. Mas o seu hardware também é bastante cobiçado por investigadores em IA que trabalham com imagens e vídeo. Apercebendo-se disto, a Nvidia começou recentemente a construir supercomputadores equipados com GPU que integram software desenhado explicitamente para análise de imagens médicas. O Massachusetts General Hospital foi um dos seus primeiros clientes e agora a Nvidia quer treinar 100 mil novos programadores de software para utilizarem o sistema de imagiologia baseada em IA.[49]

4.3 Medicina Remota

A frase «medicina remota» invoca imagens de pessoas a viver em lugares distantes com acesso limitado a cuidados de saúde, como uma nave espacial ou uma plataforma petrolífera no mar do Norte.[50] Para muitas pessoas, contudo, a medicina permanece remota não apenas por razões de isolamento físico. Pense nas centenas de milhões de pessoas que vivem no mundo em desenvolvimento, ou nas dezenas de milhares de americanos

carenciados que ficam perdidos entre as brechas dos sistemas de seguros público e privado. Pense, até, na pessoa comum de classe média que tem um emprego e uma vida familiar atarefada e que simplesmente não gosta de ir ao médico.

A medicina remota baseada em IA promete uma melhoria significativa nos cuidados de saúde a cada um destes grupos. Imagine uma versão do algoritmo de deteção de cancro da equipa de Stanford para uma vasta gama de problemas de diagnóstico. Pense em conetar um estetoscópio barato ao seu telemóvel, de modo a que uma rede neural possa ouvir o seu batimento cardíaco. Ou em olhar para uma câmara para permitir que um algoritmo na nuvem examine os seus olhos, em busca de sintomas de doença visual. Agora pense em juntar esses algoritmos em algo como uma Dra. Alexa: uma assistente digital treinada em vastas coletâneas de conhecimentos médicos que foi programada para colocar questões acerca dos seus sintomas e a responder adequadamente. (A equipa do Watson, da IBM, já desenvolveu algo bastante parecido a isto com o objetivo de treinar estudantes de medicina.[51])

Uma nova geração de sensores utilizáveis poderá estimular ainda mais a eficácia da medicina remota baseada em IA. Se julga que o seu *Fitbit* é fixe, espere até que alguém do seu escritório tenha uma tatuagem eletrónica biométrica: um pequeno adesivo com a mesma espessura e elasticidade da pele humana, que pode enviar dados de saúde para o seu telemóvel sem precisar de fios. Esta «eletrónica epidérmica» é como um sistema de monitorização de Fórmula 1 mas para a saúde. Ela pode medir a sua tensão arterial, o esforço muscular, o nível de hidratação, a frequência respiratória, até

mesmo a atividade elétrica do seu coração ou do cérebro — e pode de imediato sinalizar qualquer anomalia. Um tal sistema conseguirá ser usado pelos médicos para monitorizar alguém que acabou de ter alta do hospital, ou por pessoas comuns que queiram controlar a saúde enquanto vivem as suas vidas diárias.

Estas tecnologias não substituirão diagnósticos laboratoriais sofisticados, e certamente não substituirão o cuidado presencial por parte de um médico altamente qualificado. Mas, para uma gama de doenças significativa, elas podem recomendar tratamentos simples e encaminhar o paciente para um médico se, e quando, realmente precisar, tudo a um custo muito baixo. Este tipo de cuidado de saúde de primeira linha de defesa, e baseado em IA, pode aumentar significativamente o alcance dos médicos e ajudá-los a tratar problemas muito antes de se transformarem em algo fatal e dispendioso — um casamento perfeito entre inteligência humana e automática.

As implicações podem ser particularmente dramáticas para a medicina no mundo em desenvolvimento, ao baixar o custo da tecnologia de monitorização, tornando-a bastante mais móvel.

5. O que Acontece a Seguir?

Esperamos que concorde que tudo isto é muito empolgante. Ainda assim, além de algumas destas tecnologias estarem em estádios iniciais, há muitas outras barreiras culturais que têm de

ser resolvidas antes de esperarmos vê-las com uma utilização generalizada.

5.1 Incentivos

Para ilustrar estas barreiras vamos voltar ao exemplo do sistema de alerta precoce de doença hepática baseado em IA. Irão os hospitais sequer aderir? Segundo o Dr. Mark Sendak, a pergunta que todos os hospitais irão colocar a si próprios é: «O que significa para a minha rentabilidade podermos prever melhor a doença hepática?» Não precisamos de ser propriamente cínicos para observar, como faz Sendak, que «os grandes sistemas de saúde fazem dinheiro com as doenças crónicas progressivas».

A questão dos incentivos não é específica dos Estados Unidos. Todos os países, mesmo aqueles onde o governo paga os cuidados de saúde, enfrentam o problema assustador de assegurar que todos no sistema estão tanto motivados como capacitados para uma perspetiva de longo prazo. Sendak reitera este ponto: «Parte de melhorar a ciência de dados nos cuidados de saúde tem a ver com alinhar incentivos, de modo a que todos se preocupem com o que acontece aos seus pacientes quando não estão no hospital.» Desse modo, os médicos irão pressionar os seus chefes para que lhes deem as ferramentas necessárias para tomar decisões tendo em mente o interesse a longo prazo do doente. Neste momento isso não está a acontecer; se somos incentivados para apenas cuidar do doente quando ele está à nossa frente, argumenta Sendak, «então não nos importamos com a maneira como os dados são

armazenados e não nos preocupamos em analisar registos históricos para encontrar padrões».

O sistema legal proporciona outro conjunto de incentivos — ou, melhor, esmorecimentos. Imagine que está na situação da Dra. Katherine Heller, enquanto pondera a sensatez de comercializar, ou até mesmo oferecer, uma aplicação baseada em IA que pode prever a evolução da doença hepática. Essa aplicação provavelmente poderia ajudar muitas pessoas, mas também poderá constituir enormes riscos legais para os seus criadores. Não está claro se os criadores da aplicação, os cientistas de dados ou os médicos que usem a aplicação — ou os três — estariam sujeitos a um julgamento de zilhões de dólares por causa do inevitável primeiro caso de doença hepática não detetado, independentemente de quantas vidas forem salvas pela aplicação e de se o aconselhamento médico tinha todas as ressalvas necessárias. Isto porque os advogados e os decisores políticos não mexeram uma palha para resolver uma questão básica: quem, em última análise, é responsável pelo aconselhamento médico facultado por um algoritmo? Como é que poderemos responder a esta questão de uma maneira que simultaneamente fomente a inovação e proteja os pacientes?

5.2 Partilha de Dados

Isto traz-nos a outra questão importante: irão as equipas de ciência de dados ter acesso a todos os elementos de que precisam para melhorar os sistemas de IA existentes e construir outros novos? Se trabalharmos para um único hospital, possivelmente teremos acesso a milhares de registos de pacientes. Mas não seria

muito melhor ter milhões de registos de vários hospitais? Afinal de contas, uma das grandes razões para as empresas tecnológicas como a *Google* e o *Facebook* terem tão boa IA é a enorme escala dos seus conjuntos de dados. Certamente que há milhões de histórias clínicas de doença renal espalhadas pelas bases de dados médicas por todo o mundo. Em princípio poderiam ser reunidas, e equipas de cientistas de dados poderiam ser contratadas para as analisar utilizando ferramentas de IA de última geração, de uma maneira que continuasse a assegurar a privacidade do paciente. Fazer isto em toda a medicina criaria centenas de milhares de empregos, a par de imensas fontes de valor financeiro e social.

No entanto, há poucas hipóteses de tal vir a acontecer brevemente. Primeiro: os prestadores americanos de cuidados de saúde carecem de um padrão comum para os seus registos eletrónicos, tornando efetivamente impossível partilhar dados e alcançar a escala necessária para que as nossas melhores tecnologias de IA façam a sua magia. Mesmo em países com sistemas de saúde centralizados, como no Reino Unido, a interoperacionalidade das bases de dados — por exemplo, entre as que são usadas por médicos de família e hospitais — continua a ser um problema enorme.

Segundo: mesmo que existisse um padrão de dados comum, a maioria dos hospitais está muito hesitante em formar uma parceria com cientistas de dados, mesmo sob condições que garantam a privacidade dos pacientes. Na verdade, achamos que são completamente paranoicos, e outros investigadores com quem falámos disseram o mesmo. Os hospitais americanos tendem a ver

esses dados — os *seus* dados — como um segredo empresarial muito bem guardado. Ninguém nos diz realmente porquê, mas sempre suspeitámos de que se deve a uma razão bastante cobarde: os hospitais não querem que os seus modelos bizantinos de fixação de preços sejam alvo de engenharia reversa por parte da concorrência, pelo que a posição empresarial padronizada é simplesmente trancar os discos rígidos. Qualquer que seja a razão, todos aqueles registos de saúde eletrónicos são usados para gerar faturas muito detalhadas, mas quase nunca para ajudar as pessoas a, desde logo, necessitarem de menos serviços hospitalares dispendiosos.

Nós achamos que isto é incompreensível e não estamos sozinhos. Consegue imaginar o que aconteceria se permitíssemos que os hospitais tratassem a doação de órgãos da mesma maneira? E se os hospitais pudessem açambarcar os nossos rins após a nossa morte, tal como podem açambarcar os dados acerca dos nossos rins? Não deveria existir um formulário que pudéssemos assinar para contrariá-los, doando os nossos dados para salvar a vida de outra pessoa? Como diz Sendak, «o imperativo moral é que estas são pessoas que nos pagam por cuidados de saúde. Nós recolhemos os dados delas, nós cobramos-lhes e nada fazemos de útil com os dados. Se nós somos as únicas pessoas com estas informações, e elas estão a pagar-nos para cuidarmos delas, como é que não estamos a usá-los?»

Isto traz-nos ao tema dos próprios dados médicos, que tipicamente estão cheios de erros e de entradas em falta — por outras palavras, exatamente o que seria de esperar se

perguntássemos a um grupo de médicos atormentados para digitarem os dados manualmente, o mais depressa possível entre consultas, após lhes ter sido explicado que a maior parte dos dados nunca será utilizada para fazer algo útil. De modo que quando aparece uma equipa de investigação solitária e obtém permissão para utilizar uma ínfima fração dos dados para construir uma ferramenta de IA, eles primeiro têm de limpar e selecionar os dados. Isso requer competência, paciência e trabalho de equipa com os médicos — e neste momento, essas colaborações são tão específicas que é impraticável replicá-las. Imagine uma equipa de investigadores a gastar seis meses a limpar um conjunto de dados com dezenas de milhões de pontos de dados, tudo para colocar uma simples questão — por exemplo, como prever a progressão da doença renal — e publicar dois artigos académicos. Não existe uma maneira clara para que outros beneficiem desse conjunto de dados; nem há qualquer sistema para possibilitar interações semelhantes em larga escala. Imagine que tinha de escrever o seu próprio software de mapeamento via GPS sempre que quisesse apanhar uma boleia com um *Uber* ou um *Lyft*. O mais provável é que optasse por um táxi.

E os próprios hospitais, mais uma vez com algumas exceções, não parecem estar a apressar-se para contratar as suas próprias equipas internas de ciência de dados. O resultado é uma triste má afetação de talento. Os melhores cientistas de dados da nossa geração já podiam estar há anos a trabalhar em cuidados de saúde. Muitos teriam adorado fazer isso e oferecer as maravilhas que criam. Ao invés, têm andado a criar melhores maneiras de fazer-nos clicar em anúncios — porque é aí que estão os dados.

5.3 Privacidade

A próxima questão é acerca da privacidade da sua informação de saúde. Este é um tema muito importante que não podemos tratar na sua totalidade, mas um facto importante a destacar é que estamos a falar de dados de cuidados de saúde que os hospitais já recolhem, aos quais os funcionários do hospital já acedem para enviar as faturas. Construir um sistema de IA implica contratar alguém para analisar estes dados preexistentes no local, ou então dar acesso remoto a um cientista de dados externo num servidor seguro, após se ter removido toda a informação identificativa.

Esse facto tranquiliza muitas pessoas, mas certamente não resolve todas as preocupações possíveis acerca de privacidade ou segurança. Por exemplo: talvez o preocupe que algum cientista da dados malicioso *possa* de facto identificar pessoas com base nos registos médicos delas, mesmo a partir de um conjunto de dados supostamente desidentificado. Na verdade, de todos os problemas que aqui elencámos, este é o único que realmente pode ser resolvido com nova tecnologia — especificamente, por algo chamado «privacidade diferencial». Os estatísticos e os investigadores de aprendizagem automática pensam *bastante* acerca da privacidade dos dados e inventaram uma variedade de truques de análise — técnicas matemáticas com nomes como «subamostragem», «dispersão criptográfica» e «adição de ruídos» — que mantêm os registos de cada pessoa individual completamente seguros. Com estes novos algoritmos de privacidade diferencial, uma equipa de ciência de dados de um hospital poderia armazenar dados de cuidados de saúde de uma

maneira que lhes permitisse construir regras preditivas precisas e, no entanto, continuar a evitar que alguém de má índole possa descobrir algum detalhe específico acerca de qualquer dos pacientes que viole a privacidade. Escusado será dizer que a maioria dos hospitais está longe de implementar este tipo de algoritmos, mas os algoritmos propriamente ditos já existem. Até os pode encontrar em qualquer telemóvel novo que funcione com iOS ou Android — onde, por exemplo, são usados para analisar quais as sugestões de autocorreção que rejeita nas mensagens de texto, mantendo ao mesmo tempo as próprias mensagens encriptadas e seguras.

Depois há a questão da pirataria. A pirataria já assola os hospitais: se se lembra dos grandes ataques de *ransomware*[*] de 2017 (como o *WannaCry*), talvez também se recorde de que os hospitais foram atingidos de forma desproporcionada. Estes hospitais provavelmente não estavam a fazer nada com os seus dados relacionado com IA, mas esse tipo de atividade dificilmente teria constituído um risco de segurança mais elevado do que o já existente. Obviamente, os hospitais deveriam resolver as suas falhas de segurança referentes à informação — provavelmente, como muitos peritos sugerem, através de mover a informação para algum tipo de infraestrutura com base na nuvem, gerida por uma empresa que esteja sempre a pensar na segurança. Mas isto nada tem a ver com se os dados que já estão reunidos em servidores de hospitais devem ser usados para melhorar os cuidados de saúde.

6. Pós-Escrito

Como agora já entende, no que toca à adoção generalizada da IA, o sistema de cuidados de saúde enfrenta muito poucas barreiras de tecnologia, mas enormes impedimentos de cultura, legislação e incentivos. Algumas destas barreiras são específicas dos Estados Unidos, mas muitas outras afetam todos os sistemas de cuidados de saúde do mundo rico.

O resultado é que a próxima revolução da ciência de dados nos cuidados de saúde não precisará apenas de uma pessoa como Florence Nightingale, mas de milhares. Serão necessários investigadores como Katherine Heller, Zoltan Takats, Sebastian Thrun e Mark Sendak — pessoas que continuam a trabalhar em projetos interessantes, continuam a convencer os seus colegas dos cuidados de saúde de que esta coisa da IA realmente funciona e continuam a gerar provas válidas. Serão precisos médicos, enfermeiros, engenheiros de software, advogados, gestores de bases de dados, peritos em segurança, investidores de risco, seguradores, administradores de hospitais, decisores políticos e pacientes, os quais devem unir-se para que esta coisa funcione.

Que a força de vontade de Florence — essa «coisa mais resoluta e férrea» — continue viva em todos eles.

---

[1] Gabrielle Glaser, «Unfortunately, Doctors Are Pretty Good at Suicide», *Journal of Medicine*, 15 de agosto de 2015, https://www.ncnp.org/journalof-medicine/1601-unfortunately-doctors-are-pretty-good-at-suicide.html.

[2] Mark Bostridge, *Florence Nightingale*: *The Making of an Icon* (Nova Iorque: Farrar, Straus & Giroux, 2008), 56-60.

[3] *Ibid*., 35.

[4] *Ibid*., 31-35.

[5] *Ibid*., 68-70.

[6] *Ibid*., 47-50.

[7] *Ibid*., 105.

[8] *Ibid*., 157.

[9] Introdução de *The Collected Works of Florence Nightingale*, vol. 14, *The Crimean War*, ed. Lynn McDonald (Waterloo, Ont.: Wilfrid Laurier University Press, 2010), 9.

[10] Bostridge, Florence Nightingale, 248.

[11] *Ibid*., 219-20.

[12] *Ibid*., 203.

[13] *Ibid*., 220, 225-29.

[14] *Ibid*., 229.

[15] Ibid.

[16] Carta, 7 de agosto de 1855, em *Collected Works of Florence Nightingale*, vol. 14, McDonald, ed., 204.

[17] Bostridge, Florence Nightingale, 237.

[18] Lynn McDonald, «Florence Nightingale and Her Crimean War Statistics: Lessons for Hospital Safety, Public Administration and Nursing», transcrição da apresentação no Gresham College, 30 de outubro de 2014, https://www.gresham.ac.uk/lectures-and-events/florence-nightingaleand-her-crimean-war-statistics-lessons-for-hospital-safety-.

[19] Bostridge, Florence Nightingale, 248.

[20] *Ibid*., 226.

[21] *Ibid*., 229.

[22] *The Times*, 8 de fevereiro de 1855, citado por E. T. Cook, *The Life of Florence Nightingale*, 2 vols. (Londres: Macmillan, 1913), 1:236-37, disponível em https://archive.org/details/lifeofflorencenio1cookuoft. (Indisponível)

[23] Bostridge, *Florence Nightingale*, 260-62.

[24] *Ibid*., 321.

[25] E. H. Sieveking, «Training Institutions for Nurses», *Englishwoman's Magazine* 7 (1852): 294, de Anne Summers, «The Mysterious Demise of Sarah Gamp: The Domiciliary Nurse and Her Detractors», *Victorian Studies* 32, n.º 3 (1989): 365.

[26] Edwin W. Kopf, «Florence Nightingale as Statistician», *Journal of the American Statistical Association* 15, n.º 116 (1916): 390.

[27] John Hall, carta para o Dr. Andrew Smith, 6 de abril de 1856. A carta foi vendida em leilão em 2007 e está transcrita em https://www.bonhams.com/auctions/15231/lot/26/.

[28] Florence Nightingale, «Notes on the Health of the British Army», em *Collected Works of Florence Nightingale*, vol. 14, McDonald, ed., 864.

[29] Kopf, «Florence Nightingale as Statistician», 390.

[30] Ibid.

[31] Bostridge, Florence Nightingale, 345.

[32] *Ibid*., 335-39.

[33] Florence Nightingale, «Notes on the Health of the British Army», em *Collected Works of Florence Nightingale*, vol. 14, McDonald, ed., 854-55.

[34] Kopf;, «Florence Nightingale as Statistician», 394.

[35] Florence Nightingale, «Notes on Hospitals», em *Collected Works of Florence Nightingale*, vol. 16, *Florence Nightingale and Hospital Reform*, ed. Lynn McDonald (Waterloo, Ont.: Wilfrid Laurier University Press, 2012), 215.

[36] Kopf, «Florence Nightingale as Statistician», 397.

[37] Jocelyn Keith, «Florence Nightingale: Statistician and Consultant Epidemiologist», *International Nursing Review* 35, n.º 5 (1988): 147-50.

[38] Jan Beyersmann e Christine Schrade, «Florence Nightingale, William Farr and Competing Risks», *Journal of the Royal Statistical Society*, *Series A (Statistics in Society)* 180, n.º 1 (janeiro de 2017): 285-93.

[39] A maioria das medições da função renal do Joe teriam sido da sua creatinina sérica, em vez de medir diretamente a TFG. Todas estas medições foram aqui convertidas para TFG com o propósito de visualização. Joseph Futoma *et al*., «Scalable Joint Modeling of Longitudinal and Point Process Data for Disease Trajectory Prediction and Improving Management of Chronic Kidney Disease», em *Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence*, ed. Alexander Ihler e Dominik Janzig (Corvallis, Ore.: AUAI Press, 2016), 222-31.

[40] Não se acedeu aos dados de qualquer paciente para criar este gráfico. Estes dados foram simulados para se assemelharem aos dados do verdadeiro paciente. Para encontrar um gráfico com os dados originais, veja Futoma *et al*., «Scalable Joint Modeling», 223.

[41] Kevin C. Oeffinger *et al*., «Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update from the American Cancer Society», *Journal of the American Medical Association* 314, n.º 15 (2015): 1599-1614.

[42] Carta para William Farr, 14 de setembro de 1859, em Lynn McDonald, ed., *Collected Works of Florence Nightingale*, vol. 5, *Florence Nightingale on Society and Politics, Philosophy, Science*, Education and Literature (Waterloo, Ont.: Wilfrid Laurier University Press, 2003), 76.

[43] Futoma *et al*., «Scalable Joint Modeling».

[44] J. Balog *et al*., «Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry», *Science Translational Medicine* 5, n.º 194 (2013): 194rag93; «"Intelligent Knife" Tells Surgeon If Tissue Is Cancerous», Imperial College London, comunicado de imprensa, 17 de julho de 2013, http://www3.imperial.ac.uk/newsandeventspggrp/imperialcollege/newssummary/news17-7-2013-17-17-32.

[45] «MiniMed 670G System Launches in the United States», *Meaningful Information*, blogue da Medtronic, 7 de junho de 2017, https://www.medtronicdiabetes.com/blog/fda-approves-minimed-670g-systemworlds-first-hybrid-closed-loop-system/.

[*] A ressalva é que tanto os médicos como o algoritmo estavam a fazer julgamentos baseados apenas em imagens, o que é um bocado artificial; espera-se que, com mais informação clínica acerca do paciente, ambos tenham um desempenho melhor.

[46] P. J. Schüffler *et al*., «Semi-automatic Crohn's Disease Severity Estimation on MR Imaging», *em Abdominal Imaging: Computational and Clinical Applications*, ed. H. Yoshida, J. Näppi e S. Saini (Heidelberg e Nova lorque: Springer, Cham, 2014), 128-39.

[47] Thomas J. Fuchs e Joachim M. Buhmann, «Computational Pathology: Challenges and Promises for Tissue Analysis», *Computerized Medical Imaging and Graphics*, vol. 35, n.º 7-8 (2011): 515-30.

[48] Varun Gulshan *et al*., «Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs», *The Journal of the American Medical Association* 316, n.º 22 (2016): 2402-10. A parceria de investigação com o Moorfields Eye Hospital está descrita num comunicado de imprensa em http://www.moorfields.nhs.uk/news/moorfields-announces-research-partnership.

[49] Jim McHugh, «Man, Machine and Medicine: Mass General Researchers Using AI», blogue da Nvidia, 7 de dezembro de 2016, https://blogs.nvidia.com/blog/2016/12/07/mass-general-researchers-ai/. Veja também Lee Bell, «Nvidia to Train 100,000 Developers in "Deep Learning" AI to Bolster Healthcare Research», Forbes.com, 11 de maio de 2017, https://www.forbes.com/sites/leebelltech/2017/05/11 /nvidia-to-train-100000-developers-in-deep-learning-ai-to-bolster-health-care-research/.

[50] Veja, por exemplo, Tom Simonite, «The Recipe for the Perfect Robot Surgeon», *MIT Technology Review*, 14 de outubro de 2016, https://www.technologyreview.com/s/602595/the-recipe-for-the-perfect-robot-surgeon/.

[51] David Szondy, «IBM's Watson Adapted to Teach Medical Students and Aid Diagnosis», *New Atlas*, 21 de outubro de 2013, http://newatlas.com/ibm-supercomputer-watsonpath/29415/.

[*] *Ransomware* — tipo de software malicioso que restringe o acesso ao sistema infetado e que cobra um resgate para restabelecer o acesso. [N. T.]

CAPÍTULO 7

O «Yankee Clipper»

Basebol, grandes volumes de dados
e a importância dos pressupostos.

Há uma estirpe peculiar de evangelistas da IA que acredita que as máquinas inteligentes em breve irão tornar as pessoas irrelevantes para o processo de descoberta. Segundo esta narrativa, faltará pouco para não precisarmos de teorias ou pressupostos para aprender acerca do mundo. Tudo do que precisamos são os algoritmos certos de aprendizagem profunda, libertados no conjunto de dados certo, e ficaremos com tanto conhecimento que até nos sairá pelo nariz.

As pessoas fazem este tipo de previsão há já algum tempo. Em 2008, por exemplo, o editor-chefe da *Wire* escreveu: «A ciência pode avançar mesmo sem modelos coerentes, teorias unificadas, ou até sem qualquer explicação mecanicista [...]. Podemos atirar os números para os maiores aglomerados de computação que o mundo alguma vez viu e deixar que os algoritmos estatísticos encontrem padrões onde a ciência não é capaz.»[1]

Nós conseguimos perceber este entusiasmo. A IA é uma coisa poderosa — e ninguém sabe ao certo se um dia as máquinas serão suficientemente inteligentes para criar medicamentos novos, dizer-nos como a mente funciona ou inventar a teoria quântica da gravidade, tudo apenas a partir de dados brutos.

Mas... e atualmente? Não estamos sequer perto. Para ilustrar porquê vamos considerar uma questão científica simples e muito específica: causarão os medicamentos para a osteoporose cancro do esófago? Este é exatamente o tipo de questão a que as pessoas que trabalham em IA para os cuidados de saúde adorariam poder responder de forma automática, usando algoritmos sofisticados libertados em enormes bases de dados de informação de saúde. Na verdade, esta acaba por ser uma pergunta *perfeita* para analisarmos, porque muitas pessoas inteligentes discordam acerca da resposta. Por exemplo: a Dra. Jane Green, uma epidemiologista do cancro na Universidade de Oxford, compilou provas de que os medicamentos para a osteoporose *de facto* causam cancro. O Dr. Chris Cardwell, um investigador em saúde pública na Queen's University, em Belfast, contrapôs que não causam. Não seria fantástico se pudéssemos usar a IA para ajudar a resolver esta disputa?

Primeiro, algum contexto. A muitas pessoas com osteoporose são receitados medicamentos chamados «bifosfonatos». Estes medicamentos podem abrandar ou prevenir a perda de massa óssea, mas também acarretam o risco de perturbar o trato digestivo, resultando em náusea e diarreia. Alguns médicos receiam

que os bifosfonatos também possam aumentar o risco de se desenvolver cancro esofágico, gástrico ou colorretal.

O que dizem as evidências? Vamos começar com a defesa do «não». O Dr. Chris Cardwell e os seus colaboradores de investigação em Belfast estudaram uma enorme base de dados médicos anonimizada, com informação sobre cerca de quatro milhões de pacientes no Reino Unido. A conceção do seu estudo foi simples. Primeiro, eles procuraram um grupo de pacientes na base de dados que tivesse usado bifosfonatos. Depois correram um algoritmo de correspondência sofisticado para encontrar pacientes de «controlo» que fossem semelhantes aos do primeiro grupo, mas que não tivessem usado bifosfonatos. Por fim seguiram ambos os grupos através da base ao longo do tempo. Em última análise eles não encontraram diferenças: os utilizadores e os não utilizadores de bifosfonatos tinham taxas semelhantes de cancro do esófago. Eles publicaram as suas conclusões na *JAMA*, *Journal of the American Medical Association*, uma das revistas médicas mais prestigiadas do mundo, em agosto de 2010.[2]

Agora vejamos a defesa do «sim». A Dra. Jane Green e a sua equipa de investigação em Oxford também estudaram uma enorme base de dados de pacientes no Reino Unido, e a conceção do seu estudo, embora diferente, também foi simples. Primeiro procuraram «casos», ou pacientes, que desenvolveram cancro esofágico. Em seguida usaram um algoritmo de correspondência sofisticado para encontrar pacientes de «controlo» que fossem semelhantes aos casos, mas não tivessem desenvolvido cancro. Por fim compararam os casos com os controlos e concluíram que os

utilizadores frequentes de bifosfonatos tinham o dobro do risco de cancro do esófago do que os não utilizadores. Eles publicaram os seus resultados na britânica *BMJ*, outra das mais prestigiadas revistas médicas do mundo, em setembro de 2010 — apenas um mês após o artigo de Cardwell surgir na *JAMA*.[3]

Assim, para resumir: um estudo diz que *não* há risco extra e outro diz que há o *dobro* do risco.* Pelo menos um deles deve estar errado.

Não é invulgar que dois resultados científicos publicados olhem para conjuntos diferentes de dados e encontrem respostas dissemelhantes para a mesma questão, especialmente acerca de algo tão complicado como a saúde humana. É assim que a ciência frequentemente funciona. Ao início alguns indícios apontam num sentido e outros apontam noutro. Apenas com o passar do tempo as provas se acumulam convincentemente num só sentido.

Ainda assim, há algo bastante significativo acerca dos estudos de Green e de Cardwell, que foram publicados com um mês de intervalo em lados opostos do Atlântico, e que chegaram a conclusões opostas acerca de os bifosfonatos aumentarem o risco de cancro. Na verdade, deixámos de fora uma parte muito importante da história. Sem se aperceberem disso, ambas as equipas estavam a efetuar as suas análises na *mesma base de dados* — nomeadamente, a Base de Dados de Investigação em Medicina Geral do Reino Unido, que está publicamente disponível a qualquer investigador em saúde. As equipas obtiveram respostas diferentes, apesar de terem os mesmos casos de cancro, os

mesmos utilizadores de bifosfonatos, a mesma fonte de sujeitos controlo... o mesmo tudo.

Bem, não exatamente tudo. Os estudos obtiveram respostas diferentes porque estabeleceram pressupostos diferentes. Por exemplo: Cardwell e a sua equipa selecionaram pacientes de controlo com base na exposição aos bifosfonatos (uma conceção de «coorte retrospetiva»), enquanto Green e a sua equipa selecionaram controlos com base no desfecho de cancro (uma conceção de «caso-controlo»). Essa é a maior diferença de pressupostos entre os estudos, mas está longe de ser exclusiva — e não há uma única máquina no planeta que consiga dizer qual dos pressupostos está correto. Isto porque ainda não se inventou um algoritmo que consiga propor, testar e justificar os seus próprios pressupostos. Os algoritmos simplesmente fazem exatamente aquilo que lhes dizem.

Agora já sabe por que estamos tão céticos em relação aos evangelistas da IA neste assunto. Se uma máquina nem sequer é capaz de nos dizer qual dos estudos dos bifosfonatos está certo depois de ver as conclusões deles, então como é que poderia chegar sozinha à resposta certa sem ajuda humana?

A lição é simples. Pode parecer que atualmente dependemos de máquinas inteligentes para tudo mas, na realidade, elas dependem muito mais de nós.

1.1 Um Estudo de Pressupostos

De que maneiras é que a IA depende de pressupostos feitos pelas pessoas? Qual é sequer o aspeto destes pressupostos? Por que são tão importantes e como é que as coisas correm mal quando eles são violados? Estas são as questões que iremos tratar neste capítulo.

Na nossa perspetiva, a existência de IA inteligente não torna de alguma maneira os pressupostos menos importantes. Torna-os mais importantes porque as consequências de um único mau pressuposto podem ser amplificadas milhões de vezes, ou mais, à medida que uma qualquer máquina continua a repetir uma má decisão vezes sem conta. Dito de outra maneira: a IA permite que o fruto de uma árvore venenosa cresça exponencialmente. Quando isto acontece, geralmente é porque as pessoas fizeram más escolhas na preparação do solo.

Existem três formas principais para isto acontecer:

1. Fúria para concluir.

2. Ferrugem no modelo.

3. Viés entra, viés sai.

Para ilustrar estes temas vamos pedir ajuda a um ícone americano de meados do século, Joe DiMaggio.

Nascido em 1914 numa família de imigrantes italianos na Califórnia, Joltin' Joe DiMaggio veio a tornar-se num dos melhores jogadores de basebol de sempre e um homem cuja fama transcendeu a sua modalidade. As pessoas comuns viam-no como um herói popular, e escritores e artistas — de Hemingway a Madonna, Rodgers e Hammerstein a Simon & Garfunkel — mencionaram-no nos seus trabalhos mais duradouros. Um locutor no Yankee Stadium apelidou-o de «Yankee Clipper» por causa de um novo avião da Pan American; ambos eram velozes e glamorosos.

Enquanto dois cromos das probabilidades, iremos sempre recordar DiMaggio principalmente pelo verão de 1941, quando ele teve uma batida em 56 jogos consecutivos. Este recorde de tacadas permanece, no momento em que escrevemos, a mais longa de sempre — na verdade, paira sobre o segundo lugar de séries de batidas de 45 jogos, de «Wee» Willie Keeler, em 1897. A maioria dos fãs de basebol considera que o recorde de DiMaggio é imbatível; Stephen Jay Gould, o eminente biólogo e fã de basebol, certa vez chamou-lhe «a coisa mais extraordinária que alguma vez aconteceu no desporto americano». Como Gould referiu, não só DiMaggio derrotou 56 lançadores da Liga Principal consecutivamente, «como derrotou o mais duro tirano de todos... a Senhora Sorte».[4]

Exatamente quão improvável foi o recorde histórico de DiMaggio de batidas em 56 jogos consecutivos? A resposta é certamente interessante para os fãs de basebol, que adoram comparar proezas desportivas através de épocas e modalidades diferentes — como se a série de batidas de DiMaggio fosse mais impressionante do que

os 1281 golos de Pelé ao longo da sua carreira, ou do que as 23 medalhas olímpicas de Michael Phelps.

Contudo, na verdade estamos interessados nesta questão por uma razão muito diferente. A série de 56 jogos com batidas de DiMaggio pode ensinar-nos uma lição acerca da importância dos pressupostos — especificamente acerca dos perigos de usar maus pressupostos para extrapolar para muito longe dos dados. Esta lição é fundamental para a IA, porque as boas práticas na ciência de dados são essenciais para construir máquinas que consigam aprender e tomar decisões por conta própria. A série de batidas de DiMaggio é o ato de abertura numa parábola acerca de como o lado humano deste processo pode correr mal.

2. Joe DiMaggio e a Fúria para Concluir

2.1 Primeiro Ato: A Série

Para calcular a probabilidade da série de Joe DiMaggio de 56 jogos com batidas, iremos começar com uma metáfora. Suponha que um jogo de basebol é como atirar uma moeda ao ar: cara significa que DiMaggio teve uma batida nesse jogo e coroa que não teve. Esta metáfora torna possível analisar matematicamente uma série de batidas. Vamos começar com uma simples: quais são as hipóteses de obter caras duas vezes de seguida? Para uma moeda verdadeira, todos concordarão que a resposta é $^1/_2$ x $^1/_2$ = $^1/_4$, dado que de cada vez a probabilidade de a cara ficar virada para cima é de $^1/_2$, e dado que o primeiro lançamento não afeta o

segundo. Para a nossa moeda hipotética, Joe DiMaggio é só ligeiramente diferente: aqui a probabilidade de caras é mais de cerca de 80 por cento, uma vez que ele teve uma batida em cerca de 80 por cento dos jogos nas temporadas de 1940-42.* Portanto, a probabilidade de uma série de dois jogos com batidas é de 0,8 x 0,8 = 0,64.

É fácil alargar esta lógica a séries mais longas, usando algo chamado de «regra de composição». Suponha que um certo evento ocorre com probabilidade *P* num qualquer ensaio. Então a probabilidade de ele ocorrer todas as vezes em *N* ensaios independentes é igual a P^N, ou *P* multiplicada por si mesma *N* vezes. Assim, para calcular a probabilidade de Joe DiMaggio ter uma série de 56 jogos com batidas multiplicamos 0,8 por si mesmo 56 vezes de seguida. O resultado é um número bastante pequeno:

P (DiMaggio série de 56 jogos) = 0,8 x 0,8... x 0,8 = 1 em 250 000

Perante isto, uma reação natural é pensar: caramba, o Joe DiMaggio teve imensa sorte, certo? Isso é uma verdade indiscutível. Se olhar para a série dele jogo a jogo, irá indubitavelmente encontrar alguns ressaltos felizes ou pancadas fracas que quase não contaram.

Mas nós maravilhamo-nos com a habilidade de DiMaggio, não com a sua sorte. Para ver porquê vamos efetuar os mesmos cálculos usando as estatísticas de um jogador diferente: Pete Rose, que teve a sua própria famosa série de batidas em 1978. Nessa época da sua carreira, Rose estava a bater de forma segura em cerca de 76 por

cento dos jogos. Isto era apenas quatro por cento mais baixo do que a probabilidade de batidas por jogo de 80 por cento de DiMaggio. No entanto, ao longo de 56 jogos, a regra de composição amplifica esta modesta diferença de um jogo num enorme golfo de probabilidades:

P (Rose série de 56 jogos) = 0,76 x 0,76 ... x 0,76 = 1 em 5 milhões

É 20 vezes mais pequeno do que 1 em 250 mil de DiMaggio. E o próprio Rose era um jogador fantástico. Então, o que aconteceria com um jogador mediano da Liga Principal com uma média de batidas de 0,25, que bate em segurança em cerca de 68 por cento dos jogos?

P (série de 56 jogos) = 0,68 x 0,68... x 0,68 = 1 em 2 mil milhões

Isto quase de certeza nunca acontecerá.

Assim, é certamente verdade que DiMaggio precisou de alguns ressaltos felizes durante a sua série. Mas ele também precisou de à partida ser bastante competente, para que as probabilidades que tinha de ultrapassar fossem *apenas* de 250 mil para 1.

2.2 Intervalo: Modelo versus Realidade

Podemos aprender duas lições acerca de ciência de dados e IA a partir da nossa análise à série de batidas de Joe DiMaggio.

A primeira lição é que a probabilidade se acentua surpreendentemente depressa, como os juros num cartão de crédito. A longo prazo, margens pequenas transformam-se em margens grandes. Pense em como uma pequena diferença de probabilidade num jogo entre DiMaggio (80%) e Rose (76%) se acentuou tão drasticamente, transformando-se numa diferença 20 vezes maior ao longo de uma série de 56 jogos. Na verdade, isto é uma boa metáfora para como as máquinas habitualmente vencem quando jogam contra humanos, quer o jogo seja xadrez, *Go*, ou recomendar filmes: elas descobrem muitas vantagens pequenas que juntas se potenciam para formarem uma vantagem enorme.

A segunda lição é acerca da importância de modelar pressupostos. Como em breve irá descobrir, se aprender a primeira lição mas não esta segunda, o resultado pode ser problemático.

A maioria dos cálculos na ciência de dados requer pressupostos de um tipo ou de outro. Nós implicitamente estabelecemos dois pressupostos ao analisar a série de batidas de DiMaggio. O primeiro foi *probabilidade constante*: a hipótese de DiMaggio obter uma batida é a mesma em todos os jogos (80%). O segundo pressuposto foi *independência*: se DiMaggio obtiver uma batida num jogo, isso nada nos diz em relação ao jogo seguinte. Isto é o mesmo que dizer que se atirarmos uma moeda ao ar duas vezes, o primeiro arremesso não afeta o segundo. Sem estes pressupostos, a metáfora da moeda não funciona e os nossos cálculos também não.

Então, serão estes pressupostos literalmente verdadeiros? Nem por isso! Considere o pressuposto de probabilidade constante.

Alguns jogos, DiMaggio jogou em casa, num cavernoso Yankee Stadium; outros decorreram fora, em estádios mais pequenos. Nalguns dias houve bolas rápidas; noutros houve bolas curvas. Nalguns dias ele enfrentou lançadores do Corredor da Fama; noutros enfrentou substitutos assalariados, acabados de chegar das ligas secundárias. Longe de ter a mesma probabilidade em todos os jogos, Joe DiMaggio tinha uma probabilidade diferente de obter uma batida em cada tacada.

E em relação à independência? Esse pressuposto é mais discutível mas provavelmente também é falso. Um estudo recente apresentado na MIT Sloan Sports Analytics Conference, de 2016, analisou uma enorme quantidade de dados históricos de basebol e descobriu provas claras de um efeito «mão quente» entre batedores de basebol.[5] Por outras palavras: os jogadores que obtêm uma batida uma vez têm uma maior probabilidade estatística de obter outra na vez seguinte. Esta descoberta contraria o nosso pressuposto de independência.

Assim, se os nossos pressupostos estão errados, poderá perguntar: por que se deram sequer ao trabalho de fazer os cálculos? É uma excelente pergunta, com uma resposta complicada.

Qualquer cientista ou engenheiro lhe dirá que os modelos fazem o mundo girar. A Boeing usa modelos de túnel de vento para ajudar a construir aviões. Os biólogos usam moscas-da-fruta como modelo para os ajudar a compreender a genética humana. A Toyota usa manequins de teste de colisão como modelos para o que acontece às pessoas em colisões frontais. Todas estas situações envolvem

alguns pressupostos acerca de quais as caraterísticas do modelo que têm de ser exatas e quais as que podem ser aproximadas. Em muitas questões não poderíamos fazer qualquer progresso sem modelos. Como destacou um engenheiro-chefe no projeto Viking para Marte, o seu trabalho não era conceber uma sonda que pudesse aterrar em Marte, era criar uma sonda que pudesse aterrar no modelo de Marte criado pelos geólogos da NASA.[6]

Os cientistas de dados também usam modelos, como o que nos ajudou a refletir sobre a série de batidas de Joe DiMaggio. Os nossos modelos baseiam-se em probabilidades. Eles são usados para extrair conhecimentos a partir dos dados e para construir sistemas de IA bem-sucedidos, como muitos dos que conheceu ao longo deste livro.

Os cientistas de dados têm uma expressão favorita: todos os modelos estão errados, mas alguns são úteis.[7] Por outras palavras: nenhum modelo pode descrever o mundo real de forma perfeita, mas por vezes o desfasamento é importante e noutras não é. O corolário é que determinar se um modelo é ou não útil implica conhecer o modelo e como esse modelo será usado. Um manequim de vitrina é um modelo perfeitamente adequado de uma pessoa se tudo o que quiser fazer for exibir roupas, mas é um modelo péssimo para ensinar anatomia vascular a estudantes de medicina.

Então, vamos revisitar a nossa afirmação anterior de que Joe DiMaggio tinha uma em 250 mil oportunidades de obter uma batida em 56 jogos de seguida. Isto não é uma afirmação acerca de DiMaggio, o homem, mas acerca de DiMaggio, o modelo. Este

modelo tem pressupostos, como probabilidade constante e independência, que intencionalmente trocam realismo por simplicidade.

Não seria difícil corrigir os problemas mais graves do nosso modelo. Podíamos calcular probabilidades separadas para jogos em casa e fora, ou simplesmente ajustar os números com base nos lançadores que DiMaggio enfrentou.[8] As máquinas fazem este tipo de coisa com facilidade. Claro que, até mesmo para fazer a pergunta, à partida tem de compreender as limitações do modelo e, se tudo o que precisa é de uma aproximação grosseira para um debate no sofá com os seus amigos, estes passos extra provavelmente não valem a pena. Não precisa de aceitar que o modelo está certo, apenas que é suficientemente útil para o propósito em vista — que embora enriquecer o modelo fosse agradável, fazê-lo não traria muito conhecimento suplementar para uma discussão de pouca relevância acerca de séries de batidas no basebol.

O ponto mais importante aqui é que construir modelos é um trabalho adequado apenas a pessoas. Uma máquina pode fazer predições com base nos pressupostos com que foi programada, mas só as pessoas podem verificar esses pressupostos. Uma máquina pode ajustar um modelo, mas só as pessoas podem usá-lo para colocar as questões certas. Uma máquina pode processar milhões de pontos de dados por segundo, mas só as pessoas à partida podem decidir quais os pontos de dados que é apropriado utilizar. A boa ciência de dados requer que as pessoas e as máquinas trabalhem juntas, porque a diferença entre o modelo e a realidade

nem sempre é um assunto tão casual como é um debate sobre basebol.

Para ilustrar este ponto vamos agora olhar para o segundo ato da história do Joe DiMaggio. Aqui verá como um jornal importante levou longe demais o modelo de DiMaggio de séries de vitórias, e pelo meio acabou por assustar desnecessariamente milhões de pessoas. Esta história comporta uma lição muito importante para compreender o papel dos pressupostos na IA.

2.3 Segundo Ato: Quão Eficaz É o Vosso Método?

Os egípcios do Baixo Egito usavam uma mistura de mel e carbonato de sódio. Os mesopotâmios preferiam folhas de acácia e tecido. Os antigos persas usavam excrementos de elefante e couves: os europeus renascentistas, raiz de lírio e glândulas sericígenas de bicho-da-seda.

Nas sociedades modernas, a nossa vida está mais facilitada. A maioria das pessoas escolhe preservativos ou a pílula, ou opta pela esterilização voluntária e indolor.

O controlo de natalidade é pelo menos tão antigo quanto a civilização; a grande diferença em relação aos tempos antigos é que os nossos métodos efetivamente funcionam bem. Desde a década de 1960, quando a contraceção eficaz se tornou amplamente disponível, as taxas de natalidade em todo o mundo industrializado caíram a pique. Hoje, nos países ricos, alguma experiência com

contraceção é praticamente universal entre adultos sexualmente ativos.[9]

Nós reconhecemos que, para muitas pessoas, a escolha de quando usar contracetivos, e qual o método a usar, não pode ser reduzida a uma única variável.[10] Mas uma pergunta importante para toda a gente é qual a probabilidade de engravidar se se usar um determinado método. Foi com esta questão em mente que, em 2014, o *The New York Times* publicou um artigo intitulado «Quão Provável É Que o Controlo de Natalidade o Deixe Ficar Mal?».[11] Os autores do artigo partiram de uma premissa simples: quantas mais vezes usar um método de controlo da natalidade, mais oportunidades existem para ele falhar. Para colocar alguns números por trás desta ideia, os autores do artigo analisaram dados publicados acerca da eficácia a um ano de 15 métodos contracetivos populares. Eles usaram estes dados — em conjunto com os seus próprios cálculos, que descreveremos à frente — para criarem um gráfico interativo engenhoso, que supostamente mostrava a taxa de insucesso de cada método a longo prazo, até dez anos.

Nós usámos os mesmos dados publicados para replicar os cálculos do *The New York Times*,[12] usando a mesma metodologia utilizada pelos autores do artigo, para um subconjunto de nove destes métodos. Os nossos cálculos, que pode ver na Figura 16, concordam com os deles. Cada painel mostra um método contracetivo diferente. O eixo vertical exibe a estimativa do *The New York Times* para a probabilidade de engravidar pelo menos uma vez se se usar esse método durante um longo período.

Se está surpreendido com os números nesta figura não está sozinho: o artigo do *The New York Times* chocou muitas pessoas. Por exemplo: o artigo declarava que a taxa de insucesso a um ano para os utilizadores típicos da pílula era nove por cento*, mas que a taxa de insucesso a dez anos eram uns alarmantes 61 por cento. Os números para o preservativo eram ainda piores: a sua taxa de insucesso a dez anos era estabelecida em 86 por cento. Para muitas pessoas estes números pareciam incrivelmente elevados e implicavam um muito maior risco a longo prazo de gravidezes indesejadas do que o risco para o qual estavam preparadas. Talvez como resultado, o artigo depressa se tornou viral nas redes sociais — e embora não tenha desencadeado uma corrida em massa para entrar para um convento, causou no entanto muita ansiedade entre os leitores habituais do *The New York Times*, muitos dos quais tinham presumivelmente acreditado que os respetivos métodos contracetivos eram mais fiáveis. Até mesmo ginecologistas, que deviam conhecer a investigação tão bem como qualquer outra pessoa, foram diretamente online para partilharem o link e expressarem as suas inquietações.*

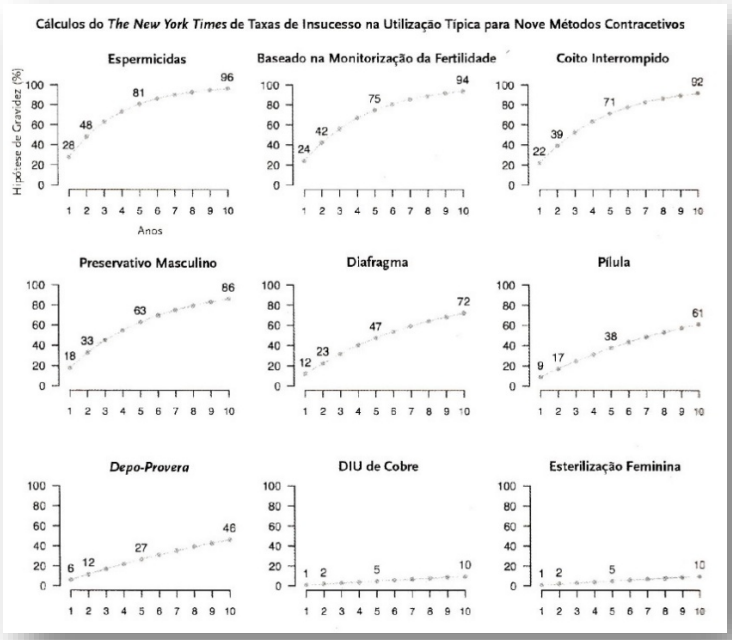


Figura 16.

2.4 Uma História Construída Sobre Maus Pressupostos

Mas havia um problema grave com o artigo do *The New York Times*: as suas putativas taxas de insucesso a longo prazo não tinham qualquer base factual. Elas quase de certeza são demasiado altas. Acontece que ninguém no mundo sabe de facto qual a taxa de insucesso a dez anos para qualquer destes métodos contracetivos.[13] Por razões práticas, a questão simplesmente nunca tinha sido estudada. Apesar desta ausência de provas, contudo, há fortes razões para acreditar que, devido aos maus pressupostos, o artigo do *The New York Times* sobrestimou drasticamente a possibilidade de engravidar durante o uso a longo prazo da maioria dos métodos contracetivos.

Eis como o *The New York* Times calculou a suposta taxa de insucesso a longo prazo de cada método. Primeiro pegaram na taxa

de insucesso a um ano da «utilização típica» obtida em estudos publicados (por exemplo, nove por cento para a pílula). Estes números referentes a um ano foram inicialmente calculados através de dados de ensaios clínicos ou de sondagens representativas a nível nacional. Eram as melhores estimativas disponíveis. Até aqui, tudo bem.

Em seguida, os autores usaram a regra de composição para calcular a probabilidade de uma «série vitoriosa» de ausência de gravidez para vários anos consecutivos.

Efetivamente, os jornalistas do *The New York Times* estavam a tratar uma série de vários anos de uso de contracetivo sem gravidez exatamente como se fosse uma série de batidas de Joe DiMaggio, usando os mesmos dois pressupostos que utilizámos atrás: independência e probabilidade constante ao longo dos anos.

Vejamos um exemplo. Entre os utilizadores típicos da pílula, a probabilidade de evitar eficazmente uma gravidez no primeiro ano é de 91 por cento. Com base neste valor, o *The New York Times* usou a regra de composição para calcular as probabilidades seguintes:

$$P(\text{ausência de gravidez durante um ano}) = 0{,}91$$

$$P(\text{ausência de gravidez durante dois anos}) = (0{,}91)^2 = 0{,}82$$

$$P(\text{ausência de gravidez durante três anos}) = (0{,}91)^3 = 0{,}75$$

E assim por diante. Ao chegar aos dez anos, a probabilidade de uma «série vitoriosa» começa a parecer mais pequena: cerca de 39 por cento. Isto implica uma probabilidade de 61 por cento de pelo menos uma gravidez num período de dez anos de utilização típica da pílula.

2.5 Uma Analogia

Estes cálculos, todavia, têm uma falha enorme. Para ver qual é vamos raciocinar por analogia. Suponha que conduzimos um estudo recrutando 100 pessoas e dando uma moeda a cada uma. Estas moedas foram alteradas de modo a que 90 tenham cara em ambos os lados e 10 coroa em ambas as faces. Agora pedimos aos nossos participantes para começarem a atirar as moedas ao ar. Digamos que obter coroa é como engravidar. A questão é: quantos dos nossos 100 participantes no estudo terão uma série de «ausência de gravidez» durante dez anos, obtendo com êxito cara dez vezes de seguida?

Claramente, a resposta é 90 por cento: 90 em 100 participantes no estudo têm moedas com duas caras. Eles evitarão sempre obter coroa. Mas vamos ver como poderíamos obter a resposta errada usando em alternativa a regra de composição. Suponha que procedíamos da seguinte maneira:

1. Pegar em dados do primeiro ano do estudo, em que 90 pessoas obtiveram cara e 10 pessoas obtiveram coroa.

2. Calcular a probabilidade média de evitar com êxito coroa durante o primeiro ano, a qual será de 90 por cento.

3. Usar a regra de composição para calcular a probabilidade de uma série vitoriosa de dez anos com base na estimativa de um ano: $0{,}9^{10}$, ou cerca de 35 por cento.

4. Concluir que apenas 35 em cada 100 participantes do estudo irão evitar com êxito uma gravidez durante dez anos consecutivos.

Foi mais ou menos isto o que o *The New York Times* fez na sua análise de taxas de insucesso de contracetivos — e está terrivelmente errada. Será correto dizer que a probabilidade média de obter cara entre os participantes do estudo é de 0,9? Absolutamente. Mas significará isso que a probabilidade média de obter cara 10 vezes de seguida é $0{,}9^{10}$, ou 35 por cento? Absolutamente não. No nosso estudo, 10 pessoas irão sempre obter coroa e as outras 90 obterão cara para sempre. Nesta população, a probabilidade média de uma série de vitórias de 10 lançamentos — ou de uma série de qualquer dimensão — é efetivamente 90 por cento e não 35 por cento. Nós não podemos usar a regra de composição como uma aproximação grosseira. A regra simplesmente não funciona *de todo* para médias populacionais.

Eis uma segunda analogia, que é muito mais próxima da nossa questão acerca da eficácia dos contracetivos: quais são as probabilidades de se evitar causar um acidente de carro durante os

próximos dez anos? Em cada ano há dois milhões de condutores que causam acidentes nos Estados Unidos, o que corresponde a um por cento dos cerca de 200 milhões de condutores no país. Assim, a probabilidade de o americano «típico» passar um só ano sem causar um acidente de carro é cerca de 99 por cento. Para calcular a probabilidade de o fazer durante dez anos, talvez houvesse a tentação de usar a regra de composição, multiplicando 0,99 por si mesmo 10 vezes:

$$P(\text{série de 10 anos sem acidentes}) = 0{,}99 \times 0{,}99 \times \ldots \times 0{,}99 = 0{,}904$$

Mas isto está errado. Para perceber porquê, vamos retroceder o relógio até ao final do primeiro ano. Após o primeiro ano, a população americana clivou-se em dois grupos: dois milhões de pessoas que causaram um acidente de carro e 198 milhões que não causaram. Agora coloque a si próprio duas questões simples: o que irá acontecer às taxas de seguro automóvel de cada grupo e porquê?

A resposta é clara. O grupo 1, com dois milhões de pessoas que causaram acidentes, verá as suas taxas aumentarem. O grupo 2, com 198 milhões de pessoas que não causaram acidentes, verá as suas taxas manterem-se ou descerem. Porquê? As companhias de seguros não estão a fazer isto para punir ou premiar as pessoas. Elas estão a fazê-lo para fixar apropriadamente o preço do risco *futuro* de um acidente — e acidentes em anos sucessivos não são independentes. Acidentes passados predizem acidentes futuros; algumas pessoas têm maior probabilidade de obterem cara e outras têm maior probabilidade de obterem coroa.

Então, o que acontecerá no segundo ano? É quase certo que mais de um por cento das pessoas no grupo 1 irão causar um acidente no segundo ano. Os condutores neste grupo são estatisticamente menos cuidadosos, pelo menos em média. Do mesmo modo, *menos* de um por cento das pessoas no segundo grupo irá causar um acidente. Os condutores neste grupo são estatisticamente mais cuidadosos — de novo, pelo menos em média. Em ciência dos dados chamamos a isto uma variável de confundibilidade: algo que tem um efeito importante no desfecho de interesse mas que não é medido diretamente.

O problema da variável de confundibilidade explica o que estava tão errado com o nosso cálculo anterior, quando pegámos na probabilidade média de 99 por cento de ausência de acidente e a potenciámos a dez anos. A questão é: de quem era a probabilidade que estávamos a multiplicar? E a resposta será: de ninguém! A probabilidade anual de um por cento é uma propriedade de uma população — ou, no melhor dos casos, de um certo *Homo mediocritus* imaginário que tem um risco de um por cento de ter um acidente de carro, 2,1 crianças, meia licenciatura, um testículo e um ovário. Mas toda a pessoa real tem um risco que ou é superior ou inferior à média de um por cento. Se tivermos um acidente no primeiro ano, o nosso risco parece mais alto; se não tivermos parece mais baixo. O cálculo da regra de combinação está errado *literalmente para toda a gente*.

2.6 De Volta à Pílula

Vamos regressar agora à taxa de insucesso a dez anos da pílula. Usando uma probabilidade de êxito de 91 por cento a um ano, em conjunto com a regra de combinação, o *The New York Times* chegou a uma probabilidade de 39 por cento para uma «série vitoriosa» de dez anos sem gravidezes. Mas, como vimos, não podemos simplesmente potenciar uma probabilidade de uma população média, porque ao fazê-lo não estamos a ter em conta as variáveis de confundibilidade. E existe aqui uma variável de confundibilidade muito importante: algumas pessoas não usam um método da maneira que deviam, pelo que é mais provável que lhes saia coroa e engravidem no início do estudo. Outras pessoas são utilizadores consistentes, pelo que é mais provável que lhes saia cara ano após ano, evitando a gravidez até ao final do estudo.

Na verdade, num estudo de contraceção não existe realmente um utilizador típico, apenas um *grupo* típico.[14] A investigação da contraceção não é uma espécie de jogo de basebol voyeurístico, onde os cientistas vasculham os quartos de dormir da nação em busca do jogador médio da Liga Principal, que bate 0,25. É sobre esperar e contar: acompanha-se um grupo típico de pessoas que está a usar um método — alguns deles de forma errática, outros de forma consistente — e conta-se quantos engravidam ao longo do tempo.

Em qualquer situação deste tipo, se usarmos a regra de combinação para refletir sobre o que acontecerá ao grupo com base

na sua média a um ano obteremos a resposta errada. Para continuar com as nossas analogias anteriores:

- No primeiro lançamento do nosso estudo hipotético de lançamento da moeda, 10 por cento das pessoas obterão coroa. Esse valor inclui tanto moedas de duas caras como de duas coroas. Assim, se não obtiver coroa aprendemos algo acerca da sua moeda: ela tem duas caras. A sua hipótese de obter coroa no lançamento seguinte é zero por cento.

- Num determinado ano, cerca de um por cento dos americanos irá causar um acidente de carro. Esse valor inclui tanto bons condutores como maus. De modo que, se não causar um acidente nesse ano, aprendemos algo sobre os seus hábitos de condução. A sua hipótese de causar um acidente no ano seguinte é inferior a um por cento.

- No primeiro ano, cerca de nove por cento de utilizadores típicos da pílula, irão engravidar. Esse valor inclui tanto utilizadores consistentes como erráticos. Assim, se não engravidar neste ano, aprendemos algo acerca do seu hábito de uso da pílula. A sua hipótese de engravidar no ano seguinte é provavelmente inferior a nove por cento. Talvez seja oito por cento, talvez seja dois por cento — ninguém sabe porque ninguém realizou o estudo. Mas sabemos que os utilizadores mais erráticos, que contribuíram mais para a taxa de insucesso de nove por cento no primeiro ano, agora estão de fora.

A Figura 17 transmite esta ideia. Ela compara as taxas de gravidezes cumulativas entre utilizadores típicos da pílula, sob dois conjuntos de pressupostos. A linha curva tracejada cinzenta assume o que o *The New York Times* assumiu: que as mulheres que permanecem no estudo em anos mais avançados continuam a engravidar à mesma irrealista taxa elevada de nove por cento ao ano. Isto prevê uma taxa de insucesso cumulativa de 61 por cento durante dez anos, com insucessos a ocorrerem com igual frequência tanto no último como no primeiro ano.

Entretanto, a curva a negro assume que, nos últimos anos, as mulheres que permanecem no estudo têm menos de nove por cento de probabilidades de engravidarem, dado que os utilizadores menos aderentes já abandonaram o estudo. Este efeito intensifica-se com o passar do tempo, pelo que no final do estudo apenas permanecem os utilizadores mais cuidadosos. Esta curva prevê uma taxa cumulativa de gravidez de algo mais como 25 por cento a dez anos, com a grande maioria de insucessos contracetivos a acontecerem no início da janela de dez anos e a utilizadores erráticos.

Devemos salientar que a única coisa que os investigadores realmente sabem a partir dos dados é que nove por cento dos utilizadores da pílula numa coorte de «utilização típica» irão engravidar no primeiro ano. Do segundo ano em diante ambas as curvas são extrapolações, baseadas apenas em pressupostos de modelação.

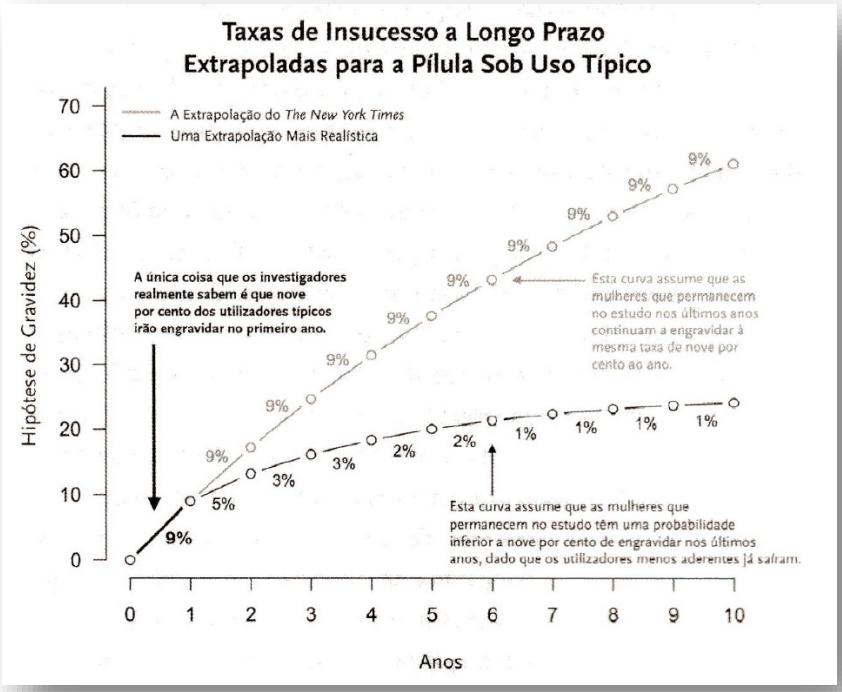


Figura 17.

Mas, ainda que todos os modelos estejam errados, alguns estão mais errados do que outros.

2.7 Epílogo: A «Mania Mais Funesta e Estéril»

Nós vemos o artigo do *The New York Times* sobre taxas de insucesso de contracetivos como um exemplo daquilo a que o guru da análise de dados, Edward Tufte, certa vez chamou «viés de fúria para concluir». Ele foi buscar o nome a um aforismo de Flaubert: «A fúria de querer concluir é uma das manias mais funestas e estéreis da humanidade.»[15]

Tufte referia-se à tendência humana para ver padrões na aleatoriedade, mas o fenómeno de fúria para concluir certamente não pára por aí. Por vezes um conjunto de dados é inerentemente

incapaz de responder a uma questão. Quando isso acontece, deve mesmo ir à procura de dados que consigam responder à questão. Por exemplo: os dados sobre a taxa de insucesso da pílula no primeiro ano não o podem informar acerca da taxa de insucesso a dez anos; para saber o que acontece ao fim de dez anos tem de esperar dez anos. Mas se está mesmo impaciente para saber a resposta *neste momento*, é lamentavelmente tentador torturar os dados de que dispõe para obter uma confissão, usando pressupostos dúbios: Essa confissão pode acabar por provocar danos relevantes. Uma coisa é usar pressupostos idealizados para analisar uma série vitoriosa de Joe DiMaggio; quase nada dependerá do resultado. Outra inteiramente diferente é usar os mesmos pressupostos para analisar a eficácia de contracetivos — um domínio onde esses pressupostos estão *bastante* errados e onde notícias falsas podem prejudicar milhões de pessoas.

Isto pode ter sido um erro de poucos dados, mas a lição para o mundo dos grandes conjuntos de dados da IA é clara. Imagine agora que aqueles pressupostos maus não eram apenas usados para escrever um artigo pontual de jornal. Em vez disso eram cozinhados num sistema de IA que toma decisões automáticas sem um humano no circuito. É exatamente assim que se chega a situações como as seguintes.

- Em abril de 2011 havia 17 cópias disponíveis na *Amazon* de *The Making of a Fly*, um livro clássico de biologia do desenvolvimento. O mais barato de 15 exemplares usados custava 35,54 dólares, enquanto o mais barato de dois exemplares novos custava mais de 23 milhões de dólares.

Descobriu-se que dois algoritmos, geridos por dois vendedores de livros diferentes, tinham entrado numa guerra de licitações inversa, sob pressupostos maus acerca do comportamento de outros vendedores.[16]

- Um revendedor de roupa online chamado *Solid Gold Bomb* criou um algoritmo que automaticamente executava novos desenhos para t-shirts de impressão por encomenda, baseados em inserir frases aleatórias em chavões populares como «Keep Calm and Carry On». Devido a uma fraca supervisão, a empresa acabou por publicitar acidentalmente t-shirts estampadas com frases terrivelmente misóginas, incluindo algumas sobre abuso sexual. Foi uma experiência traumática para muitos que encontraram as criações online e a empresa fechou portas devido às reações negativas.[17]

- A 6 de maio de 2010, as ações nos Estados Unidos experienciaram uma «queda-relâmpago», na qual o mercado perdeu mil milhões de dólares de valor numa questão de minutos — tudo por causa de algoritmos que correram mal. De acordo com o Departamento de Justiça dos EUA, um operador bolsista imprudente, sediado em Londres, tinha submetido 200 milhões de dólares em transações fantasma que eram modificadas 19 mil vezes num período muito curto, antes de finalmente serem retiradas. Isto criou uma sensação falsa de pessimismo no mercado acerca de algumas ações. Em resposta, os algoritmos de negociação de alta frequência de todas as

> outras pessoas — cujos pressupostos não abrangiam a possibilidade de tais fintas — ficaram completamente descontrolados e emitiram milhões de ordens de venda reais. Antes de as pessoas perceberem o que estava a acontecer, o Dow Jones Industrial Average tinha perdido nove por cento do seu valor em menos de meia hora.[18] (Felizmente, o índice recuperou quase de imediato.)

Estes algoritmos não estavam cientes das consequências das suas decisões ou do contexto de negócio que em primeiro lugar levou à sua criação. Eles simplesmente estavam a fazer aquilo que lhes foi dito para fazerem, por pessoas que fizeram pressupostos errados.

Contudo, mesmo ao reconhecermos o perigo dos maus pressupostos, também é importante sermos comedidos no nosso ceticismo, para não sermos encurralados num canto inútil onde permanecemos para sempre, relutantes em fazer quaisquer pressupostos. Nem todos os pressupostos são maus e nem todos os maus pressupostos criam problemas. A inteligência artificial depende de impelir as fronteiras da ciência de dados tão longe quanto possível, e por vezes isso significa depender de pressupostos e aproximações para extrapolar para lá do domínio original de um conjunto de dados. Por exemplo:

- Os epidemiologistas usam a IA para explorar bases de dados enormes de registos médicos, para responderem a questões importantes de saúde.

- Os psicólogos estão a estudar as publicações no Instagram para detetarem alterações no estado de espírito de alguém, que possam prever uma depressão emergente ou ansiedade.

- Os observadores do mercado usam as conversas nas redes sociais como indicador avançado da atividade económica.

- A *Zillow* usa dados disponíveis publicamente, em conjunto com relatórios gerados pelos utilizadores, para prever o valor de mercado de basicamente todas as casas dos Estados Unidos.

Estas ideias, e milhares de outras, podem funcionar e de facto funcionam. Mas para o fazerem têm de contornar um facto básico: a maioria dos conjuntos de dados na Era da Internet são coligidos sob condições nada científicas para o propósito de outra pessoa, e só acidentalmente são úteis para qualquer outro propósito. Para ultrapassar isso — ou, igualmente importante, para saber quando não se pode ultrapassar — é melhor que tenha considerado os seus pressupostos honestamente.

Agora está em marcha uma grande democratização da IA. Os esforços contínuos para recolher e organizar dados, para que estes poderosos algoritmos possam realizar o seu trabalho eficazmente sem cometerem erros flagrantes, irão criar reservatórios gigantescos de valor social e económico. Em breve quase todas as empresas, grandes e pequenas, irão contar com este tipo de dados

para desenvolverem a sua atividade. Nesta nova Era é essencial acalmarmos a impaciência para concluir, lembrando-nos de que cada pressuposto não verificado é um substituto — uma aproximação para ser utilizada, para o melhor ou para o pior, apenas até que estejam disponíveis mais dados.

3. Ferrugem do Modelo

Agora já viu como pressupostos maus, inseridos no próprio ADN de um modelo, podem resultar em erros pavorosos. No entanto, os modelos nem sempre nascem deteriorados. Eles ficam assim devido a demasiada ferrugem.

Uma aplicação de IA particularmente famosa e malfadada ilustra este fenómeno na perfeição. Este sistema ficou online em 2008, na esperança de resolver um problema importante de saúde pública, com dinheiro e vidas em risco. Com o passar do tempo, contudo, as suas predições afastaram-se cada vez mais da realidade. Em 2012, o modelo estava a errar *severamente*. Contudo, mesmo quando o seu desempenho se deteriorou, o modelo continuou a receber muita publicidade — afinal de contas, usava «grande volume de dados», um chavão deveras sedutor e, todavia, tão protegido por um campo de forças de tédio matemático que frequentemente desencoraja o escrutínio.

Este sistema chamava-se *Google Flu Trends* (GFT). Esta é a história de como correu mal e por que foi finalmente encerrado em 2015.

3.1 Usar Grandes Volumes de Dados para Prever Surtos de Gripe

A gripe mata anualmente centenas de milhares de pessoas em todo o mundo e é fonte de sofrimento para dezenas de milhões de outras. E isso é apenas a gripe sazonal. Os especialistas em doenças infeciosas perdem mais horas de sono a pensar na gripe pandémica, como o surto de «gripe de Hong Kong» de 1968, ou o surto de «gripe espanhola» de 1918, a qual matou 50 milhões de pessoas — três vezes mais do que a Primeira Guerra Mundial.

Para informar acerca dos seus esforços de prevenção e tratamento da gripe, os Centros para Controlo e Prevenção de Doenças dos Estados Unidos (ou CDC) há muito que usam algo chamado ILINet: a Rede de Vigilância de Doenças do Tipo da Influenza. A ILINet é uma rede de abrangência nacional com mais de 2700 prestadores de cuidados de saúde que enviam dados e espécimes laboratoriais diretamente para os CDC sempre que veem pacientes com sintomas parecidos com os da gripe. Os CDC usam esta informação para produzir um índice semanal de atividade gripal para cada estado.

Infelizmente, há um grande problema com a ILINet: pode demorar uma semana ou mais para que todos os espécimes sejam processados e os dados brutos possam ser analisados. Então, embora as agências de saúde pública em todo o país dependam da ILINet para tomarem todo o tipo de decisões importantes relacionadas com a prevenção da gripe, testes e distribuição de medicamentos — e embora seja a melhor ferramenta que eles têm

para alerta situacional —, geralmente está desatualizado em duas semanas. Isso é tempo suficiente para um surto de gripe infetar uma enorme quantidade de pessoas.

Os cientistas de dados na Google, contudo, acreditaram que podiam resolver este problema usando uma abordagem engenhosa de IA. A ideia deles era tanto simples quanto brilhante: a frequência de certas questões de pesquisa na Internet deveria estar fortemente correlacionada com a atividade gripal. Por exemplo, a Figura 18 mostra quão frequentemente os americanos estavam a perguntar ao *Google* «Quanto tempo dura a gripe?» entre 2008 e 2012.

As buscas para este termo atingem um máximo todos os invernos, por volta da mesma altura em que a época da gripe também atinge o pico. O *Google* consegue analisar estas questões de pesquisa muito mais depressa do que os CDC examinam amostras laboratoriais provenientes de 2700 clínicas. Deveria por isso ser possível rastrear a atividade gripal com um desfasamento de notificação muito mais curto, usando um sistema de IA que agregasse questões de pesquisa e produzisse uma previsão.

O mapeamento entre as pesquisas no *Google* e a atividade gripal, contudo, é imperfeito e ruidoso. Nem todos usam os mesmos termos de pesquisa; nem todas as pesquisas por um termo específico significam que alguém tem gripe. Por isso, não quererá construir um sistema de deteção da gripe de forma ingénua, contando um caso de gripe por cada pesquisa no *Google* que inclua

a palavra «gripe». Tem de ser mais esperto do que isso, construindo uma regra preditiva que use dados históricos.

Os cientistas de dados da *Google* fizeram exatamente isso. Os *inputs* do seu modelo eram as frequências de 50 milhões de diferentes termos de pesquisa possíveis. O *output* era uma previsão para os valores publicados da ILINet fornecidos semanalmente pelos CDC — o padrão de excelência para quantificar a atividade gripal nos Estados Unidos. A equipa da *Google* descreveu o seu método num artigo na *Nature*[19] e começou a publicar as suas previsões de atividade gripal do seu modelo num site dedicado a «Tendências da Gripe», para grande alarde da comunidade de saúde pública.

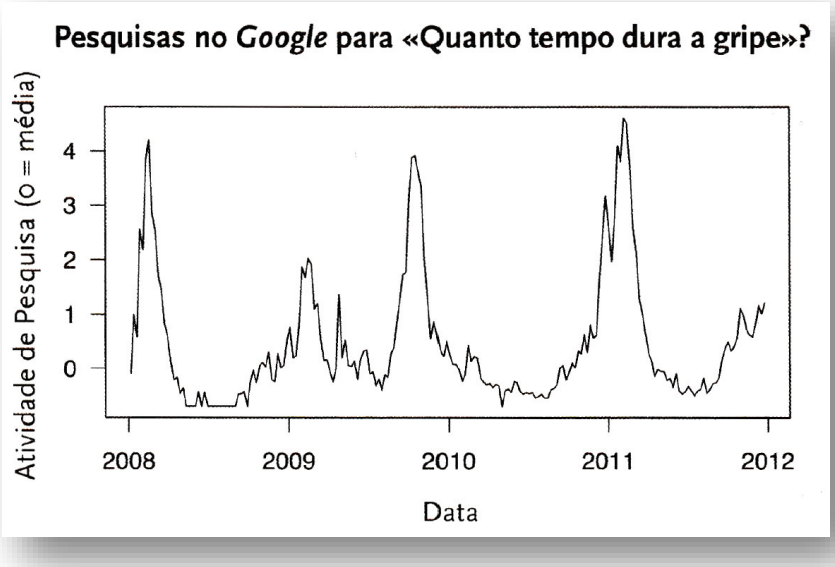


Figura 18.

Infelizmente, o «Tendências da Gripe» teve um início atribulado. Em 2009 falhou completamente um pico enorme de gripe fora de estação provocado pela pandemia do H1N1 («gripe A»). Em

resposta, os engenheiros da *Google* fizeram alguns ajustes ao algoritmo, e nos dois anos seguintes, do outono de 2009 ao verão de 2011, o «Tendências da Gripe» teve um bom desempenho: acompanhou os números dos CDC muito de perto, mas sem o desfasamento de duas semanas.[20]

No início do outono de 2011, contudo, tudo correu mal. Na época de 2011-12, o modelo sobrestimou a atividade gripal em cerca de 50 por cento, fazendo soar o alarme entre os profissionais de saúde pública que confiavam nele. Depois ficou ainda pior: na época de 2012-13, a previsão do «Tendências da Gripe» ultrapassou o pico de inverno em quase 150 por cento.[21] No seu conjunto, de agosto de 2011 a setembro de 2013, as estimativas do *Google* foram demasiado elevadas em 100 de 108 semanas.[22] Se os agentes de saúde pública tivessem em conta estas estimativas, teriam acabado por alocar recursos para lidar com dezenas de milhares de casos de gripe que simplesmente não existiram.

3.2 Como os Modelos Envelhecem

O *Google Flu Trends* é um ótimo exemplo de um princípio geral: na IA, os modelos não permanecem novos de fábrica durante muito tempo.

Na verdade, gostamos de pensar nos modelos a envelhecerem como uma frigideira de ferro fundido. Se cuidarmos bem de uma frigideira de ferro, efetivamente torna-se melhor com o tempo: desenvolve-se uma pátina bonita e a comida não aderirá tão facilmente. O mesmo é verdade para um modelo na IA. Se fizer a

manutenção do modelo e o *temperar* regularmente com novos dados — esse é o processo de tentativa e erro de ajuste do modelo de que falámos no Capítulo 2 —, fará melhores previsões com o passar do tempo. Mas se negligenciar o modelo — se deixar que uma crosta de pressupostos velhos e enegrecidos se desenvolva demasiado —, então a pátina facilmente pode transformar-se em ferrugem. Com um pouco mais de negligência o modelo acabará por apodrecer e transformar-se em nada.

O «Tendências da Gripe» sofria de um caso sério de ferrugem do modelo, a roçar o podre. Para perceber porquê, falámos com a Dra. Rosalind Eggo, uma investigadora de doenças infeciosas na Escola de Higiene e Medicina Tropical de Londres. Ela desde logo louvou a *Google* por ter colocado um recurso tão rico ao serviço do bem público, mas também achou que a maneira como o «Tendências da Gripe» falhou a pandemia de H1N1 em 2009 devia ter lançado um alerta vermelho. «Embora a *Google* tenha sido muito opaca em relação aos detalhes do algoritmo», explicou Eggo, «algumas pessoas conjeturaram que os termos de pesquisa que foram escolhidos não eram de todo termos referentes à gripe, mas a termos referentes ao *inverno*, como pesquisas por basquetebol na escola secundária. Em resultado, apenas estavam a captar uma sazonalidade geral em padrões de pesquisa, em vez de algo específico da gripe». Eggo citou um artigo de 2014 da *Science*, do Dr. David Lazer e colegas, que analisou o desempenho do *Google Flu Trends* ao longo do tempo e concluiu que o algoritmo original era «parte detetor de gripe, parte detetor de inverno».[23] Isto devia ter deixado as pessoas um pouco mais desconfiadas acerca das escolhas de conceção incluídas no algoritmo.

Outro grande problema foi que o *Google* efetivamente incentivou os próprios utilizadores a violarem os pressupostos de modelação do «Tendências da Gripe». O *Google* está constantemente a ajustar de milhares de maneiras, grandes e pequenas, a forma como os seus algoritmos de pesquisa funcionam — e em resposta as pessoas ajustam os seus próprios padrões de pesquisa. Um exemplo é a função de preenchimento automático do *Google*, que sugere termos de pesquisa à medida que digitamos. Isto poupa tempo às pessoas, mas também altera o seu comportamento. Eggo salienta que quando elas colocam dedos ranhosos no teclado para obterem conselhos acerca dos seus sintomas gripais, o «preenchimento automático irá influenciar as suas pesquisas finais». Por exemplo, talvez alguém que anteriormente pesquisou por «gripe melhor remédio» agora pesquise por «gripe melhor tratamento», porque «tratamento» era a primeira sugestão do preenchimento automático. Mas o *Google Flu Trends* dependia do pressuposto de uma relação estável entre termos de pesquisa e atividade gripal. Se esse pressuposto falha — se algo *diferente da gripe* causa grandes alterações nos termos que as pessoas pesquisam —, então o modelo de previsão também falha. Provavelmente foi isso que aconteceu depois de 2009: estes milhares de pequenas alterações aos algoritmos impulsionadas pelos negócios, de acordo com Eggo, «não estavam a ser rastreados pelo *Google Flu Trends*, e o efeito delas na qualidade do ajustamento não estava a ser monitorizado».

Há dois aspetos tristes nesta história. Primeiro, em princípio, não teria sido difícil, a uma empresa com os recursos da *Google*, concretizar um compromisso permanente com a «prevenção da ferrugem», deixando o seu modelo adaptar-se a novos

comportamentos de pesquisa. Os modelos «dinâmicos», do tipo dos que conseguem rastrear um conjunto de relações estruturais em mudança entre *inputs* e *outputs*, são um componente-padrão da caixa de ferramentas da ciência de dados. Para nós é uma espécie de mistério a razão para os cientistas de dados da *Google* não terem conduzido as coisas nesta direção — e, tanto quanto sabemos, nenhum explicou o motivo em público.

Outro facto triste é que agora muitos investigadores foram dissuadidos de perseguir uma ideia com um potencial tão fantástico. Perguntámos a Eggo se a comunidade de saúde pública teria aprendido a lição no rescaldo do «Tendências da Gripe», e ela respondeu:

> Acho que eles aprenderam a lição um pouco em demasia. Ficaram assustados com este fracasso. Se o Google simplesmente tivesse permitido que o seu modelo se adaptasse com o algoritmo de pesquisa, provavelmente teria funcionado bem. E poderia ter-nos oferecido informação muito mais detalhada ao nível das cidades do que a que os sistemas de vigilância formais alguma vez poderiam fornecer. Isso é bastante promissor. Mas imagino que o Google não queria mais má publicidade. E, para os investigadores, é um caso de gato escaldado de água fria tem medo.

Por vezes, tudo aquilo de que precisa é prevenir a ferrugem. Esperamos que a *Google* esteja disposta a tentar outra vez.

4. Viés Entra, Viés Sai

Outro assunto importante na IA surge quando treinamos os nossos modelos usando um conjunto de dados com um viés inerente. Eis uma analogia. Na eleição presidencial de 2016, todos

os agregadores de sondagens previram a vitória de Hillary Clinton. Mas os seus modelos de previsão, por mais inteligentes que fossem, estavam inerentemente limitados pela qualidade dos dados de *input*. O problema era um viés pequeno, mas persistente, nas sondagens subjacentes, o qual subestimou o apoio a Donald Trump.

Muitos algoritmos na IA padecem de um problema semelhante: viés entra, viés sai. Há uma parábola clássica acerca de um modelo de rede neural, que em tempos o exército dos Estados Unidos produziu, para detetar tanques que estavam parcialmente escondidos na orla de uma floresta.[24] Os cientistas do Exército treinaram o seu modelo usando um conjunto de dados de fotografias catalogado, algumas com e outras sem tanques. A rede neural revelou ter uma precisão surpreendentemente elevada. Até se saiu bem quando o Exército reteve parte dos dados de treino originais e os usou exclusivamente para testar o desempenho do modelo. (Esta prática de validação de resultados, usando dados teoricamente «fora da amostra», é comum na IA.)

Mas quando o Exército tentou usar o método para detetar tanques no mundo real, não funcionou — mais valia que estivessem a atirar uma moeda ao ar. Os peritos estavam confusos. O que poderia explicar a dramática queda do desempenho? Então alguém descobriu que os dados de treino tinham um viés escondido. Todas as fotografias com tanques tinham sido tiradas num dia com sol brilhante, enquanto as fotografias sem tanques haviam sido obtidas num dia nublado. O que o modelo tinha efetivamente aprendido a fazer foi a distinguir uma floresta com e sem sombras provenientes

das árvores — uma competência que era completamente inútil para identificar tanques.

Muitas aplicações imprudentes ou incorretas da IA fracassam por uma razão semelhante: viés escondidos nos dados de treino. Quantos mais dados tiver, pior este problema pode tornar-se. Conjuntos de dados maiores não eliminam forçosamente o viés. Por vezes, centram-se simplesmente de forma assíntota nos vieses que sempre estiveram ali.

Veja, por exemplo, a maneira como a inteligência artificial tem sido usada no sistema de justiça criminal. Os juízes responsáveis pelas sentenças sempre tentaram aferir o perigo que um condenado representa para a sociedade. Tradicionalmente, têm feito isto de maneira não científica, recorrendo à sua própria sabedoria, intuição e experiência para efetuar julgamentos acerca do caráter e do cadastro do réu. Agora, contudo, essas avaliações começam a ser fundamentadas em dados — e atualmente alguns juízes estão até a confiar em algoritmos de aprendizagem automática, treinados com dados históricos do sistema de justiça, que preveem a probabilidade de alguém reincidir.

Um algoritmo popular de previsão de reincidência chama-se COMPAS (Correctional Offender Management Profiling for Alternative Sanctions). O COMPAS, como todos os sistemas deste tipo, está explicitamente impedido de «saber» coisas como a etnia ou o género do réu como um dos seus *inputs*. Mas isso não é suficiente para evitar que o viés entre de modo sorrateiro: toda a premissa da aprendizagem automática estipula que é possível

aprender por aproximação acerca de atributos não observados. Então, para testar a neutralidade do algoritmo, jornalistas do *ProPublica* investigaram os resultados da previsão de reincidência de 10 mil pessoas presas no condado de Broward, na Florida, sentenciadas por juízes usando o COMPAS.[25] Eles verificaram quem tinha sido preso de novo nos dois anos seguintes, e depararam-se com uma evidente disparidade étnica. Entre as pessoas que não cometeram mais crimes, os réus negros tinham uma taxa de falsos positivos superior: a probabilidade de serem erradamente classificados como de alto risco era superior à dos réus brancos. Pelo contrário, entre as pessoas que cometeram crimes adicionais, os brancos tinham uma taxa de falsos negativos mais elevada: a probabilidade de serem classificados como de baixo risco era superior à dos réus negros.

Como é que isto pôde acontecer, sendo o algoritmo subjacente verdadeiramente «neutral em relação à etnia»? Uma explicação muito plausível para a discrepância é os *próprios dados*. Lembre-se: os algoritmos de IA são concebidos para encontrar e recriar os padrões dos conjuntos de dados com que treinaram. Se esses padrões forem inerentemente discriminatórios, então um algoritmo aprenderá a discriminar. Imagine que aceita os argumentos, como alguns académicos avançaram, de que é mais provável a polícia deter uma pessoa negra pelo mesmo crime; de que é mais admissível os procuradores públicos levarem adiante casos que envolvam suspeitos negros; de que é mais plausível que o júri condene suspeitos negros e de que é mais provável que os brancos tenham melhores advogados. Se qualquer dessas alegações for verdadeira, então é claro que os dados irão refletir taxas de

reincidência mais elevadas para os negros do que para os brancos, por razões que nada têm a ver com a propensão de alguém para reincidir. Se a pele escura sugere que é mais provável alguém ser detido, acusado e condenado por um dado crime, então qualquer algoritmo de previsão de reincidência a realizar o seu trabalho irá procurar afincadamente representantes de pele escura. E dado o longo e triste historial de desigualdade racial nos Estados Unidos, esses representantes são numerosos e o algoritmo poderá perguntar, por exemplo, se o réu tem algum familiar na prisão.

Infelizmente, não podemos determinar se este tipo de racismo por aproximação foi responsável pelo padrão de viés nas previsões do algoritmo do COMPAS no condado de Broward. Isto porque o algoritmo é secreto.[26] A empresa que vende o software não informa os réus ou os juízes acerca do seu funcionamento interno — portanto, se o algoritmo o classificar como sendo de alto risco, e em consequência o juiz lhe der uma sentença mais longa, *não poderá sequer perguntar porquê*. Isto moralmente é obsceno. Culturalmente não aceitamos secretismo nos algoritmos usados para classificar equipas de futebol americano universitário. Então por que devemos aceitá-lo numa situação em que está em risco a liberdade de alguém?

Muitas pessoas, quando ouvem falar de algo tão chocante como um algoritmo secreto a atribuir sentenças de prisão de uma forma etnicamente enviesada, chegam a uma conclusão simples: que a inteligência artificial não deveria desempenhar de todo qualquer papel no sistema de justiça criminal.

Apesar de estarmos tão chocados e revoltados quanto qualquer um, achamos que essa é a conclusão errada. Sim, todos devemos lutar contra o viés algorítmico quando ele surge. Para fazer isso precisamos de vigilância permanente por parte de peritos: pessoas que conhecem a lei mas que também conhecem a IA e têm poder para agir perante uma ameaça à justiça. Mas mesmo ao reconhecermos as armadilhas de usar a IA para ajudar as pessoas a tomarem decisões importantes, e ao reiterarmos o apelo para que a transparência e a justiça passem a ser valores definidores desta nova Era, não nos esqueçamos de que também há aqui um potencial incrível. Afinal de contas são as pessoas, e não os algoritmos, as principais responsáveis pela embrulhada no condado de Broward:

- Aqueles, no sistema de justiça, que trataram um algoritmo para sentenciar da mesma maneira que tratariam um micro-ondas, simplesmente introduzindo alguns números e virando as costas.

- Os legisladores e tribunais superiores, que permitiram que estas decisões fossem tomadas usando algoritmos patenteados cujo funcionamento interno não pode ser interpretado, contestado ou até analisado.

- Acima de tudo a polícia, procuradores, juízes e júris, cujas ações coletivamente codificaram um viés racial *humano* nos conjuntos de dados que treinam estes algoritmos.

É este último grupo — que inclui cada um de nós — que mais o deve preocupar. Se ouvir a história do COMPAS e chegar à conclusão de que a IA deveria ser mantida a quilómetros de distância de decisões importantes, colocamos-lhe uma questão simples: será o estado das coisas aqui verdadeiramente seu amigo? As decisões importantes no sistema de justiça criminal sempre foram tomadas usando algoritmos enviesados treinados com dados imperfeitos. Acontece apenas que esses algoritmos simplesmente viviam dentro das mentes das pessoas. Não pode sujeitar os vieses destes algoritmos de «*wetware*[*] humano» a um escrutínio numérico direto, como pode fazer com uma regra preditiva na IA. Mas tudo o que precisa de fazer é olhar para o rol de criminosos no corredor da morte em Huntsville, no Texas, ou analisar as taxas de encarceramento na América estratificadas por etnia — 0,45 por cento para brancos, 2,31 por cento para negros — e poderá ver os danos que esses vieses humanos forjaram.[27]

E não é apenas no sistema de justiça criminal. Gostaria, por exemplo, de deixar o seu futuro nas mãos destes decisores?

- O gestor de recursos humanos, que é mais provável que entreviste pessoas com nomes estereotipicamente brancos do que com nomes carateristicamente negros.

- O chefe que dá avaliações de desempenho mais elevadas a empregados atraentes.

- O diretor de admissões à universidade que exige padrões mais elevados aos asiáticos do que aos brancos.

- O executivo que paga 80 cêntimos a uma mulher e um dólar a um homem para fazerem o mesmo trabalho.

- As pessoas bem-intencionadas, diversas e basicamente decentes, no comité de recrutamento, que têm de olhar para uma pilha enorme de currículos para uma única vaga e que por isso são indevidamente influenciadas por uma formatação e escrita elegantes, com verbos ativos.

Os algoritmos de tomada de decisão enviesados e mal informados não são menos perniciosos apenas porque funcionam com pequenas células cinzentas em vez de chips de silicone. Não seria o mundo um lugar melhor se as pessoas que sofrem nas mãos de decisores preconceituosos tivessem de facto acesso a uma segunda opinião da IA — um algoritmo cujo raciocínio e vieses estivessem a descoberto, e por isso pudessem ser corrigidos?

5. Pós-Escrito

Imagine que podia viajar no tempo até à década de 1990, quando descarregou o seu primeiro navegador de Internet, ou para a década de 2000, quando comprou o primeiro smartphone e criou contas no *Facebook* e no *Twitter* (atualmente *X*). À luz do que aprendeu desde então, qual o conselho que daria a si próprio — acerca de que informação partilhar, que fotos publicar e que hábitos

cultivar? Ou, se tivesse a atenção de chefes de empresas e reguladores governamentais, o que quereria que eles soubessem? Que histórias contaria acerca de como essas tecnologias mudaram a sua vida para melhor? Que patologias lhes pediria que evitassem?

Em breve a inteligência artificial irá desempenhar um papel em decisões que são muito mais importantes do que os filmes que as pessoas veem na *Netflix*, as músicas que ouvem no *Spotify* ou as notícias que o *Facebook* lhes recomenda. Ela irá informar quais os tratamentos médicos que as pessoas recebem, quais os trabalhos em que competem, quais as universidades que frequentam, quais os empréstimos para que se qualificam — e, sim, quais as sentenças de prisão que recebem quando cometem um crime. Ao refletir sobre estes assuntos complexos não podemos depender dos conselhos de um viajante no tempo. Somos apenas nós, e *temos de fazer isto bem*. Há muito a ganhar mas também muito a perder, e o equilíbrio que alcançarmos entre custos e benefícios será enormemente afetado por se as pessoas no comando percebem como as tecnologias de IA verdadeiramente funcionam. Se nos desenrascarmos — ou, pior ainda, se deixarmos que as empresas tecnológicas do mundo avancem depressa e partam coisas enquanto o resto das pessoas perde o seu tempo a preocupar-se em vão com pesadelos de ficção científica — iremos matar a credibilidade destes sistemas de IA antes mesmo de terem uma hipótese de amadurecer, e iremos privar a humanidade de tantas promessas...

Mas agora imagine um mundo onde somos efetivamente inteligentes acerca dos nossos esforços — um mundo onde

colocamos os peritos certos e as proteções legais adequadas nos lugares certos, e onde estamos eternamente vigilantes face aos vieses e pressupostos dos nossos algoritmos. Nesse mundo, os nossos protocolos de tomada de decisão podem tornar-se radicalmente melhores do que os assolados pelo enviesamento que temos agora — aqueles que dão uma vantagem imerecida aos que são mais atraentes, ou que têm atitudes mais dinâmicas, ou o pai mais rico, ou a pele mais branca. A nossa visão coletiva e a tecnologia atingiram um ponto em que podemos com êxito ensinar as máquinas a conduzir um carro, prever doenças renais e manter uma conversa. Certamente que podemos ensinar essas máquinas a serem justas. E elas poderão até ensinar-nos.

Todos concordam que alguns assuntos são demasiado importantes para serem decididos por um algoritmo inimputável a operar sozinho. Alguns de nós simplesmente iriam um passo mais além e diriam o mesmo acerca das pessoas. Quando se trata das decisões importantes na vida, podemos e devemos combinar a inteligência artificial com o conhecimento e os valores humanos. Só é preciso que as pessoas e as máquinas trabalhem juntas.

---

[1] Chris Anderson, «The End of Theory: The Data Deluge Makes the Scientific Method Obsolete», *Wired*, 23 de junho de 2008, https://www.wired.com/2008/06/pb-theory/.

[2] C. R. Cardwell *et al.*, «Exposure to Oral Bisphosphonates and Risk of Esophageal Cancer», *The Journal of the American Medical Association* 304, n.º 6 (11 de agosto de 2010): 657-63.

[3] J. Green *et al.*, «Oral Bisphosphonates and Risk of Cancer of Oesophagus, Stomach, and Colorectum: Case-Control Analysis Within a UK Primary Care Cohort», BMJ 2010;341:C4444.

[*] Um ponto importante: duas vezes um número pequeno continua a ser um número pequeno. O risco basal de cancro esofágico para pessoas com idades entre 60 e 79 anos é cerca de 1 em 1000 ao longo de cinco anos. Green *et al.* estimaram que este risco aumenta para cerca de 2 em 1000 com o uso de bifosfonatos durante cinco anos.

[4] Stephen Jay Gould, «The Streak of Streaks», *New York Review of Books*, 18 de agosto de 1988, http://www.nybooks.com/articles/1988/08/18/thestreak-of-streaks/.

[*] Estamos a agregar dados ao longo de três temporadas para obter uma amostra maior e para evitar escolher seletivamente as estatísticas de DiMaggio nos jogos da sua série de batidas, o que iria inflacionar artificialmente a sua probabilidade real de batidas por jogo.

[5] Brett Green e Jeffrey Zweibel, «The Hot-Hand Fallacy: Cognitive Mistakes or Equilibrium Adjustments? Evidence from Major League Baseball», artigo apresentado na MIT Sloan Sports Analytics Conference, março de 2016, http://www.sloansportsconference.com/wpcontent/uploads/2016/02/1422-Baseball.pdf. (Página não encontrada)

[6] Agradecemos a Peter Norvig da Google por esta história. Ouvimo-lo contá-la durante uma visita à Universidade do Texas em Austin em 2011; também é narrada em «On Chomsky and the Two Cultures of Statistical Learning», http://norvig.com/chomsky.html.

[7] Na verdade a expressão é atribuída ao estatístico George Box.

[8] Veja, por exemplo, Edward Beltrami e Jay Mendelsohn, «More Thoughts on DiMaggio's 56-Game Hitting Streak», *Baseball Research Journal* 39, n.º 1 (verão de 2010), disponível em https://sabr.org/research/morethoughis-dimaggio-s-56-game-hitting-streak.

[9] Kimberly Daniels, William D. Mosher, e Jo Jones, «Contraceptive Methods Women Have Ever Used: United States, 1982-2010», *National Health Statistics Reports* n.º 62, 14 de fevereiro de 2013, http://www.cdc.gov/nchs/data/nhsr/nhsro62.pdf.

[10] Veja, por exemplo, Guttmacher Institute Fact Sheet, «Contraceptive Use in the United States», setembro de 2016, https://www.guttmacher.org/fact-sheet/contraceptive-use-united-states.

[11] Gregor Aisch e Bill Marsh, «How Likely Is It That Birth Control Could Let You Down?», *The New York Times*, secção de resenha de domingo, 13 de setembro de 2014.

[12] James Trussell, «Contraceptive Failure in the United States», *Contraception* 83, n.º 5 (maio de 2011): 397-404.

[*] «Utilização típica» não significa «utilização correta». Se usar a pílula exatamente como indicado, a taxa de insucesso é muito menor, menos de um por cento ao ano.

[*] Por exemplo, @hricciot: «Chocante — até para uma ginecologista como eu! #LARCisBest.» «LARC» significa «contracetivo reversível de longa duração», como um DIU.

[13] A única exceção é a esterilização feminina, para a qual existem de facto dados a longo prazo.

[14] Este padrão foi amplamente adotado na literatura desde que foi proposto por Trussell e Kost na década de 1980, em «Contraceptive Failure in the United States: A Critical Review of the Literature», *Studies in Family Planning* 18, n.º 5 (1987): 237-83.

[15] Gustave Flaubert, *Correspondance* (Paris: Louis Conard, 1929), 5:111. Na citação original em francês lê-se: «La rage de vouloir conclure est une des manies les plus funestes et les plus stériles qui appartiennent à l'humanité.»

[16] Joshua Klein, «When Big Data Goes Bad», *Fortune*, 5 de novembro de 2013, http://fortune.com/2013/11/05/when-big-data-goes-bad/.

[17] Catherine Talbi, «"Keep Calm and Rape" T- Shirt Maker Shutters After Harsh Backlash», *The Huffington Post*, 25 de junho de 2013, https://www.huffingtonpost.com/2013/06/25/keep-calm-and-rape-shirtn3492411.html.

[18] Silla Brush, Tom Schoenberg e Suzi Ring, «How a Mystery Trader with an Algorithm May Have Caused the Flash Crash», *Bloomberg News*, 21 de abril de 2015, https://www.bloomberg.com/news/articles/2015-04-22/mystery-trader-armed-with-algorithms-rewrites-flash-crash-story.

[19] J. Ginsberg *et al.*, «Detecting Influenza Epidemics Using Search Engine Query Data», *Nature* 457 (19 de fevereiro de 2009): 1012-14.

[20] D. Lazer *et al.*, «The Parable of Google Flu: Traps in Big Data Analysis», *Science* 343 (14 de março, 2014): 1203-5.

[21] D. R. Olson *et al.*, «Reassessing Google Flu Trends Data for Detection of Seasonal and Pandemic Influenza: A Comparative Epidemiological Study at Three Geographic Scales», *PLOS Computational Biology* 9, n.º 10 (2013), https://doi.org/10.1371/journal.pcbi.1003256.

[22] Lazer *et al.*, «The Parable of Google Flu».

[23] Ibid.

[24] Embora esta história possa ter uma origem anterior, encontrámo-la primeiro em Hubert L. Dreyfus e Stuart E. Dreyfus, «What Artificial Experts Can and Cannot Do», *AI & Society* 6, n.º 1 (1992): 18-26.

[25] Julia Angwin e Jeff Larson, «Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say», *ProPublica*, 30 de dezembro de 2016, https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say.

[26] Julia Angwin, Jeff Larson, Surya Mattu e Lauren Kirchner, «Machine Bias», *ProPublica*, 23 de maio de 2016, https://www.propublica.org/arti-cle/machine-bias-risk-assessments-in-criminal-sentencing.

[*] O cérebro humano ou um ser humano considerado não ter particularmente qualquer lógica ou capacidades computacionais humanas. [*N. T.*]

[27] Leah Sakala, «Breaking Down Mass Incarceration in the 2010 Census», relatório da Prison Policy Initiative, 28 de maio de 2014, https://www.prisonpolicy.org/reports/rates.html.

Agradecimentos

Queremos agradecer em conjunto às duas pessoas mais responsáveis por impulsionarem este livro desde as suas fases iniciais: Lisa Gallagher, da DeFiore & Company, e Tim Bartlett, da St. Martin's Press. Esta é a primeira coisa que algum de nós alguma vez escreveu que não se destina a uma audiência académica, e começámos com quase nenhuma noção do que realmente implicaria escrever e editar um livro «a sério». Estamos bastante gratos a Lisa por ter visto e alimentado o potencial daqueles primeiros rascunhos rabiscados, que agora parecem tão terrivelmente toscos. Estamos igualmente agradecidos a Tim, por apostar em dois cientistas de dados suficientemente imprudentes para porem à prova as suas aptidões como escritores, e por nos dar tantos conselhos certeiros pelo caminho. Estamos também gratos a Doug Young, da Transworld, pelos seus valiosos comentários editoriais.

Também estamos reconhecidos às muitas outras pessoas da DeFiore, St. Martin's Press e Macmillan, que têm sido tão prestáveis, incluindo Robert Allen, Alan Bradshaw, Jeff Capshew, Laura Clark, Jennifer Enderlin, Tracey Guest, Leah Johanson, Linda Kaplan, Alice Pfeifer, Gabrielle Piraino, Jason Prince, Sally Richardson, Brisa Robinson, Mary Beth Roche, Robert Van Kolken, Laura Wilson e George Witte. Um agradecimento especial a India Cooper, cujo magnífico trabalho de edição colocou nitidamente em evidência a diferença entre um escritor profissional e dois amadores como nós. Obrigado também a Larry Finlay, Bill Scott-Kerr, e ao resto da equipa da Transworld, pelo seu apoio.

Agradecemos a Ellen Zippi pela sua preciosa ajuda com a pesquisa para este livro. Estamos também gratos a muitos dos nossos colegas por partilharem histórias e conhecimentos, muito especialmente a Steven Levitt por nos ter apresentado à Lisa Gallagher, e a David Madigan por nos chamar a atenção para os dois estudos sobre o uso de bifosfonatos descrito no Capítulo 7. Agradecemos a Rosalind Eggo, Katherine Heller e Mark Sendak pelo seu tempo e esforço ao aceitarem ser entrevistados. Agradecemos também àqueles membros da família que incansavelmente leram rascunhos e deram as suas opiniões: Catherine Aiken, Patricia e Josh Lowry, Anne e George Scott, e Brian Woods.

Agradecimentos Pessoais

Estou reconhecido ao meu coautor, James Scott. Acima de tudo, agradeço à minha família pelo seu amor e apoio: à minha mulher, Anne Gron, e aos nossos filhos, Emma, Michael e Sarah.

— Nick Polson

Obrigado a Nick Polson. Devo a Nick tanta coisa na minha carreira que não me é possível listar tudo aqui; este livro é apenas o último de uma longa série de projetos e ideias que ele tão generosamente partilhou comigo. Suponho que durante as próximas décadas olharei para trás e verei o Nick como a influência singular mais importante na minha vida profissional, e o melhor amigo que alguma vez tive neste campo. Também quero agradecer aos três professores mais importantes que já tive: Bill Jeffreys, Jim Berger e John Trimble. Sem Bill e Jim nunca me teria tornado um estatístico. Sem a gentileza e a generosidade de John nunca teria sabido como «abreviar/aperfeiçoar /abrilhantar» o meu caminho para uma prosa melhor. Agradeço também aos meus pais, que deram tanto — sobretudo o seu exemplo. Por último, estou grato à minha mulher, Abigail Aiken, por praticamente tudo. Amo-te, e não poderia ter ajudado a escrever este livro sem o teu apoio.

— James Scott