

***Deep learning*: a Inteligência Artificial que domina a vida do século XXI**

Dora Kaufman¹

Latanya Sweeney, ex-chefe de tecnologia da Comissão Federal de Comércio dos EUA e atualmente professora da Universidade de Harvard, foi informada por uma colega que o Google AdSense associava seu nome a anúncios sugerindo sua prisão. Intrigada, ela digitou o nome de outro de seus colegas, Adam Tanner, e o anúncio da mesma empresa surgiu sem a sugestão de prisão. Testando nomes racialmente associados, Sweeney encontrou discriminação estatisticamente significativa, sendo que um nome estereotipado como de negro era 25% mais propenso a receber um anúncio de registro de detenção – claramente um viés do sistema de busca ao reproduzir os preconceitos raciais da sociedade.²

A participação de pesquisadores chineses nos principais congressos mundiais de Inteligência Artificial (IA) cresceu de 10% em 2012 para 23% em 2017, enquanto de americanos no mesmo período caiu de 41% para 34% (AGRAWAL; GANS; GOLDFARB, 2018). A tendência é a China assumir a liderança na pesquisa e na aplicação comercial de IA, consequência de vários fatores, dentre eles a ausência de proteção à privacidade dos dados de seus cidadãos, significando uma vantagem comparativa vis-a-vis de regiões, como a Europa, com regulamentação e cultura rigorosas.

Os dois eventos, aparentemente não relacionados, têm em comum o processo denominado de *deep learning* (aprendizagem profunda), capaz de transformar grandes volumes de dados em informação útil. Tarefas tradicionalmente desempenhadas pelos seres humanos (reconhecimento visual, tomada de decisão, reconhecimento de voz, tradução) e outras que superam a capacidade humana (manipular e processar grandes bases de dados, *big data*), estão sendo executadas por máquinas inteligentes. Aprendizagem profunda é sobre previsão, e permeia grande parte das atividades do século XXI. Quando digitamos uma consulta ao Google, é ele que seleciona a resposta personalizada e os anúncios apropriados ao perfil do usuário, bem como traduz um

¹ Ver Editorial, p. 5.

² L. Sweeney, "Discrimination in online ad delivery", *Communications of the ACM* 56, no. 5 (2013), p. 44-54, <<https://dataprivacylab.org/projects/onlineads/>>.

texto de outro idioma, assim como filtra os e-mails não solicitados (*spam*). A Amazon e o Netflix recomendam livros e filmes pelo mesmo processo, do mesmo modo o Facebook usa aprendizado de máquina para decidir quais atualizações mostrar e o Twitter faz o mesmo para os tweets. Quando acessamos um computador, em qualquer de seus formatos, provavelmente estamos acessando concomitantemente um processo de *deep learning*. “O aprendizado de máquina faz inferências a partir de dados. E quanto mais dados eles têm, melhor elas ficam. Agora não precisamos programar computadores, eles se programam” (DOMINGOS, 2015, p. xi). Cada interação acessa dois níveis: “O primeiro é conseguir o que você quer. O segundo nível, e no longo prazo o mais importante, é ensinar o computador sobre você. Quanto mais você ensina, melhor ele pode servir – ou manipular você” (ibid., p. 264).

Os volumes de dados gerados atualmente inviabilizam o uso da tradicional programação computacional (com regras definidas a priori). A vantagem dos sistemas de aprendizado é que eles próprios estabelecem os algoritmos, i.e., adaptam-se automaticamente aos requisitos da tarefa. “Para muitas aplicações não fomos capazes de criar algoritmos apropriados apesar das décadas de pesquisa que começaram nos anos 1950. [...] O aprendizado de máquinas agora é a força motriz da Inteligência Artificial”:

Especialmente nos últimos vinte anos ou mais, as pessoas começaram cada vez mais a se perguntar o que poderiam fazer com todos esses dados. Com esta pergunta, toda a direção da computação é revertida. Antes, os dados eram o que os programas processavam e cuspiam – os dados eram passivos. Com esta pergunta, os dados começaram a conduzir a operação; não são mais os programadores, mas os dados em si que definem o que fazer a seguir. (ALPAYDIN, 2016, p. x-xiii)

A grande quantidade de dados não é o único fator restritivo. No reconhecimento de imagem facial, por exemplo, os seres humanos têm certa facilidade, mas não conseguem explicá-lo (conhecimento tácito), o que não permite programar o computador. Ao analisar diferentes imagens de rosto de uma pessoa, um programa de aprendizado captura o padrão específico para essa pessoa e, em seguida, verifica esse padrão em uma dada imagem (ibid.). Uma das aplicações que mais tem surpreendido é a tradução interlingual automática.

Existem muitas expressões sendo utilizadas por diferentes comunidades – reconhecimento de padrões, modelagem estatística, mineração de dados, descoberta de conhecimento, análise preditiva, ciência de dados, sistemas adaptativos, sistemas de auto-organização, e outros – e alguns até o denominam simplesmente Inteligência Artificial (DOMINGOS, 2015). Independente do nome e da funcionalidade, o foco é no desenvolvimento de algoritmos e técnicas para solucionar determinados problemas, executar uma tarefa específica. A *strong* IA ainda é ficção.

A título de introdução

Cunhado em 1955, por John McCarthy, o termo “Inteligência Artificial” deu início a um campo de conhecimento associado à linguagem e à inteligência, ao raciocínio, à aprendizagem e à resolução de problemas (RUSSELL; NORVIG, 2009³). A IA propicia a simbiose entre o humano e a máquina ao acoplar sistemas inteligentes artificiais ao corpo humano (prótese cerebral, braço biônico, células artificiais, joelho inteligente e similares), e a interação entre o homem e a máquina como duas “espécies” distintas conectadas (homem-aplicativos, homem-algoritmos de IA). Tema de pesquisa em diversas áreas – Computação, Linguística, Filosofia, Matemática, Neurociência, entre outras –, a diversidade de subcampos e atividades, pesquisas e experimentações, dificulta descrever o estado da arte atual da IA. Os estágios de desenvolvimento bem como as expectativas variam entre os campos e suas aplicações, que incluem os veículos autônomos, reconhecimento de voz, games, robótica, tradução de linguagem natural, diagnósticos médicos, assim por diante. Atualmente, os sistemas inteligentes estão em todas as áreas de conhecimento.

Existem inúmeras definições de Inteligência Artificial, reflexo das especificidades intrínsecas a cada campo. Russell e Norvig (2009, p. 1-5) listam oito delas agrupadas em duas dimensões – as relativas a processos mentais e raciocínio e as relativas a comportamento – contudo, duas definições generalistas servem ao nosso propósito. Conforme a primeira de John McCarthy, Inteligência Artificial “é a ciência e a engenharia de fazer máquinas inteligentes, especialmente programas de computador

³ Publicado originalmente em 1994 e seguido de várias edições, adotado nas universidades americanas como o livro de referência sobre IA.

inteligentes”.⁴ A segunda, de Russell e Norvig, define IA como o estudo de “agentes inteligentes” capazes de “perceber seu meio ambiente e de realizar ações” com a expectativa de “selecionar uma ação, que maximize seu desempenho” (2009, p. viii, 37).

Em 1959, Arthur Lee Samuel, pioneiro norte-americano no campo de jogos de computador e Inteligência Artificial, cunhou o termo “machine learning” (ML) (enquanto funcionário da IBM), inaugurando um subcampo da IA cuja finalidade é prover os computadores da capacidade de aprender sem serem programados. Evoluindo a partir do estudo do reconhecimento de padrões e da teoria de aprendizagem computacional na IA, o *machine learning* explora o estudo e a construção de algoritmos que, seguindo instruções, fazem previsões ou tomam decisões baseadas em dados – modelos elaborados a partir de entradas de amostras. Originada na estatística, em que migrar de observações particulares a descrições gerais é chamado de “inferência”, a aprendizagem é chamada de “estimativa”, e a classificação é chamada de “análise discriminante” (ALPAYDIN, 2016). O aprendizado de máquina é empregado em uma variedade de tarefas de computação, nas quais programar os algoritmos é difícil ou inviável. Esses modelos analíticos permitem que pesquisadores, cientistas de dados, engenheiros e analistas produzam decisões e resultados confiáveis e replicáveis, e revelem ideias ocultas em relacionamentos históricos e tendências de dados.

Usamos o aprendizado de máquina quando acreditamos que existe uma relação entre observações de interesse, mas não se sabe exatamente como. Porque não sabemos sua forma exata, não podemos simplesmente seguir em frente e anotar o programa de computador. Portanto, nossa abordagem é coletar dados de observações de exemplo e analisá-lo para descobrir o relacionamento. (ALPAYDIN, 2016, p. 29)

Na década de 1980, inspirados no cérebro humano, cientistas da computação criaram um subcampo da ML propondo um processo de aprendizado com base nas redes neurais, com resultados mais concretos nesta década.⁵ O pioneiro foi Geoffrey Hinton, com a ideia de “neural networks” em artigo publicado na revista Nature de

⁴ “Q. What is artificial intelligence? A. It is the science and engineering of making intelligent machines, especially intelligent computer programs” Disponível em: <<http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html>>. Acesso em 29 de jan. 2018.

⁵ “O início da década de 1980 trouxe esperança de que os engenheiros pudessem programar cuidadosamente sistemas especialistas para replicar domínios habilidosos como diagnósticos médicos, mas estes eram caros de se desenvolver, complicados e incapazes de lidar com a miríade de exceções e possibilidades, levando ao que ficou conhecido como o ‘Inverno da IA’” (AGRAWAL; GANS; GOLDFARB, 2018, p. 32).

1986.⁶ O avanço não ocorreu acidentalmente. “Ganhamos capacidade de construir hardware paralelo conectando milhares de processadores, e as redes neurais artificiais despertaram interesse como uma possível teoria para distribuir cálculos em um grande número de unidades de processamento, todas em paralelo” (ibid., p. 28). O processo de aprendizagem profunda começou a florescer na década de 1990, com foco em problemas solucionáveis de natureza prática, relacionados a uma tarefa concreta. “As redes profundas ainda funcionam em domínios relativamente restritos, mas estamos vendo resultados mais impressionantes todos os dias à medida que as redes aumentam e são treinadas com mais dados” (ALPAYDIN, 2016, p. 109). O treinamento consiste em mostrar exemplos e ajustar gradualmente os parâmetros da rede até obter os resultados requeridos, denominado “aprendizagem supervisionada”: são fornecidos os resultados desejados (*output*) e, por “tentativa e erro” chega-se ao resultado – meta.

Uma vez que temos dados – e hoje em dia temos dados “grandes” – uma computação suficiente disponível – e agora temos centros de dados com milhares de processadores – apenas esperamos e deixamos o algoritmo de aprendizagem descobrir tudo o que é necessário por si só. [...] Descobrir essas representações abstratas é útil não só para a previsão, mas também porque a abstração permite uma melhor descrição e compreensão do problema. (ALPAYDIN, 2016, p. 108)

A abordagem é chamada de “retropropagação” (*back propagation*), aprendendo por exemplos. Muitos problemas foram transformados de problemas algorítmicos (Quais são as características de um gato?) em problemas de previsão (Essa imagem é similar a uma imagem que já vi antes?) (AGRAWAL; GANS; GOLDFARB, 2018).

A rede geralmente tem entre 10 e 30 camadas empilhadas de neurônios artificiais. Num reconhecimento de imagem, por exemplo, a primeira camada procura bordas ou cantos; as camadas intermediárias interpretam as características básicas para procurar formas ou componentes gerais; e as últimas camadas envolvem interpretações completas. Na identificação de fotos nas redes sociais, a máquina percebe padrões e “aprende” a identificar rostos, tal como alguém que olha o álbum de fotos de uma família desconhecida e, depois de uma série de fotos, reconhece o

⁶ In: Rummerhart, David E.; Hinton, Geoffrey E.; Williams, Ronald J. Learning representations by back-propagating errors. *Nature*, Vol. 323, p. 533-536. October, 1986. Disponível em: <https://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop_old.pdf>. Acesso em: 29 de jan. 2018.

fotografado. O reconhecimento de voz, que junto com a visão computacional está entre as aplicações mais bem-sucedidas, já permite a comunicação entre humanos e máquinas, mesmo que ainda precária (Siri, Alexa, Google Now). Na cognição houve, igualmente, importantes avanços.

É importante notar que as máquinas inteligentes não reproduzem o funcionamento do cérebro, cuja complexidade ainda é relativamente pouco entendida, inviabilizando qualquer tentativa nessa direção. É mais correto dizer que a construção dessas máquinas é inspirada no cérebro humano. O cérebro é composto de neurônios, que por sua vez são formados por detritos que se conectam por meio de sinapses: cada vez que os detritos dos neurônios se encontram provocam uma sinapse (conexão). Essa configuração é denominada “redes neurais” em que, por analogia, o equivalente aos neurônios no computador são as unidades, ou seja, cada unidade do computador equivale a um neurônio no cérebro humano. Se temos 100 “sinapses” num computador, significa que temos 100 informações chegando e se conectando. As novas unidades, localizadas numa nova camada, recebem as informações, processam e “cospem” o *output* para as unidades de uma nova camada.

No processo de visão, por exemplo, a retina, um sensor de luz, representa a primeira camada. A retina é impactada por feixes de luz, que são as primeiras informações originadas no exterior. O mesmo se passa no ouvido com relação ao som, no olfato do nariz com relação ao cheiro, e no tato da pele com relação a sensibilidade. São informações elétricas e químicas, posteriormente enviadas para o cérebro. O aparelho perceptivo da visão é o único dos sentidos em que a primeira camada contém neurônios (logo, já é “cérebro”). Não por coincidência é o mais sofisticado, correspondendo a 1/3 do cérebro, ou seja, esta parcela do cérebro é dedicada à visão (a segunda atividade predominante no cérebro são os movimentos). A luz inicialmente encontra o sensor da retina, que é a primeira camada, em seguida segue para uma nova camada, neste caso localizada na parte de trás do cérebro chamada de V1, continua se deslocando entre várias camadas, até retomar para a parte frontal do cérebro (*vision path way*). O cérebro tem dez áreas, e cada área cerca de 140 milhões de neurônios. O computador criado pela Microsoft há cerca de dois anos, considerado o mais avançado atualmente na tarefa de reconhecimento de imagem – Image res.Net –, tem 152

camadas, ou seja, as unidades vão se conectando e transmitindo informação a outras unidades ao longo de 152 camadas.

Fazendo um paralelo entre a visão humana e a câmara fotográfica, a nossa retina corresponde ao sensor de imagem da câmara. Em ambos, o que desencadeia o processo é a incidência de luz. O *input* da luz se transforma num número. Como isso é possível? A luz é composta de fótons, então importa calcular quantos fótons “caíram” na retina por unidade de tempo. Simplificando, o que permite diferenciar um objeto de outro é o número de fótons que sensibiliza a retina. Se todos os *inputs* viram números, temos um conjunto de números na primeira camada. Os processos no cérebro e os nas máquinas são semelhantes. Cada unidade que corresponde ao neurônio humano tem a decisão sobre o que será enviado à outra camada (ou não enviado). O que sai de uma camada não é necessariamente igual ao que entrou da camada anterior, significando certo grau de autonomia em relação ao operador.

Cada unidade recebe informações (*inputs*) de muitas unidades da camada anterior. No estado de evolução atual da IA, o operador humano arbitra o número de camadas. No futuro, existe forte indicação neste sentido, as máquinas vão construir outras máquinas inteligentes (sem arbitragem humana). O que define uma máquina inteligente são dois componentes: o valor de cada conexão e a arquitetura, traduzido no número de camadas.

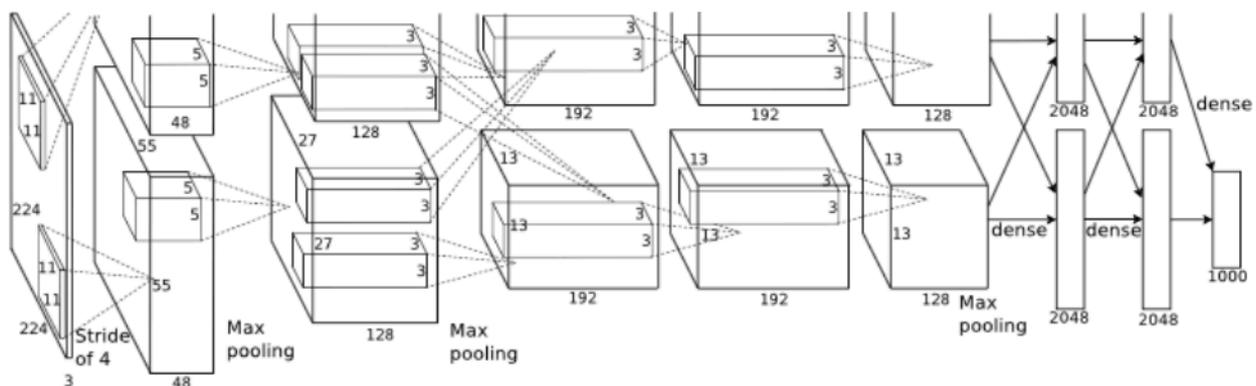


Figura 1. Arquitetura típica – com delimitação de responsabilidades – CPUs.

Fonte: ImageNet classification with deep convolutional neural networks, p. 5.

Disponível em: <<https://www.nvidia.cn/content/tesla/pdf/machine-learning/imagenet-classification-with-deep-convolutional-nn.pdf>>. Acesso em: 16 abr. 2018.

A figura 1 ilustra uma arquitetura típica, com a delimitação de responsabilidades entre duas CPUs (*Central Processing Unit*).⁷

Aprendizagem profunda no século XXI

O avanço do processo de aprendizagem profunda a partir dos anos 2006 a 2010 com a obtenção de resultados explícitos, deve-se fundamentalmente a três fatores: (a) crescente disponibilidade de grande quantidade de dados (*big data*), (b) maior capacidade computacional e (c) evolução dos algoritmos. Vejamos dois desses componentes.⁸

Big data

Em fevereiro de 2008, o *Centers for Disease Control and Prevention* (CDC) identificaram um crescimento de casos de gripe no leste dos EUA; na ocasião, o Google declarou ter detectado um aumento nas consultas sobre os sintomas da gripe duas semanas antes do lançamento do relatório. A partir dessa experiência, sua unidade filantrópica criou um sistema de alerta, o “Google Flu Trends”.⁹ Anteriormente ao aparecimento do vírus H1N1, pesquisadores do Google publicaram um artigo na revista *Nature*, ignorado pelas autoridades, sobre a capacidade de previsão da propagação da gripe de inverno, com base nos dados gerados em sua plataforma. A metodologia basicamente estabelecia correlações entre a frequência de certas consultas e a disseminação da gripe ao longo do tempo e espaço, identificando regiões específicas em tempo real. Esse evento influenciou uma mudança na mentalidade sobre o uso de dados (MAYER-SCHÖNBERGER; CUKIER, 2013). “*Big data* refere-se a coisas que se pode fazer em grande escala para extrair novos insights ou criar novas formas de valor, mudando os mercados, as organizações, a relação entre cidadãos e governos e muito mais” (ibid., p. 6).

Mayer-Schönberger e Cukier indicam como desdobramentos a capacidade de analisar grandes quantidades de dados sobre um tópico específico, e não mais se ater a amostras; a disposição em adotar a desordem do mundo real dos dados, deixando de

⁷ Esse conteúdo foi transmitido diretamente à autora pelo pesquisador em Computer Science Davi Geiger da NYU.

⁸ A capacidade computacional adentra em um campo de conhecimento particular, do domínio das ciências exatas/ tecnologias.

⁹ Artigo do NYT. Disponível em: <<https://www.nytimes.com/2008/11/12/technology/internet/12flu.html>>. Acesso em: 31 de mai. 2018.

privilegiar a exatidão; e o respeito crescente por correlações em vez de causalidades. Trata-se de sacrificar a exatidão para ter acesso à tendência geral. Agrawal, Gans e Goldfarb (2018) destacam três funções desempenhadas pelos dados: (a) primeiro temos os dados de entrada (*input*), que alimentam os algoritmos e são utilizados no processo de previsão; (b) segundo, os dados de treinamento (*training data*), usados para aperfeiçoar os algoritmos; e (c) terceiro, temos os dados de feedback com a função de melhorar o desempenho dos algoritmos com base na experiência dos usuários (*ibid.*, p. 43).

Qualquer interação com tecnologias digitais deixa “rastros”, alguns voluntários como as publicações nas redes sociais – Facebook, Twitter e Instagram –, e outros involuntários, como as informações armazenadas nos bancos de dados digitais na compra com cartão de crédito, na movimentação bancária *online*, no acesso aos programas de fidelidade, no vale-transporte, nas comunicações por telefonia móvel, e inúmeras outras ações presentes em nossa rotina. Esses podem ser usados pelas plataformas originais ou “reusados” por terceiros, ou combinados pela fusão de conjunto de dados, com as mais variadas finalidades, e são responsáveis por inúmeros benefícios da sociedade do século XXI (e, igualmente, por inúmeras ameaças). “O entusiasmo pela ‘internet das coisas’ – incorporando chips, sensores e módulos de comunicação aos objetos do cotidiano – é, em parte, relacionado à rede, mas também sobre digitalizar a informação de tudo que nos rodeia” (MAYER-SCHÖNBERGER; CUKIER, 2013, p. 96). Por meio deles é possível revelar uma infinidade de questões relacionadas à população, desde quais grupos são mais suscetíveis a determinadas doenças até qual é o perfil do cidadão propenso a honrar um empréstimo bancário, até segmentar os consumidores em perfis. Pentland (2015) defende que o *big data* oferece a chance de ver a sociedade em toda a sua complexidade; para ele, uma vez desenvolvida uma visualização mais precisa dos padrões de vida humana, podemos esperar compreender a sociedade de forma mais adequada à nossa rede complexa e interligada de seres humanos e tecnologia.

A importância dos dados cresce gradativamente.¹⁰ Em 2012, a operadora espanhola Telefónica criou uma empresa separada – Telefónica Digital Insights – com o

¹⁰ Ver sobre “Capitalismo de Dados” no artigo de Dora Kaufman desta edição.

propósito de comercializar dados anônimos e agregados de localização de assinantes para varejistas e outros. Uma divisão do cartão de crédito Mastercard – Mastercard Advisors – agrega e analisa cerca de 65 bilhões de transações de 1,5 bilhões de titulares de cartão em 210 países procurando identificar tendências de negócios e consumo para, em seguida, comercializar a informação à terceiros (MAYER-SCHÖNBERGER; CUKIER, 2013).

Algoritmos

“Algoritmos estão em toda parte. Dominam o mercado de ações, compõem música, dirigem carros, escrevem artigos de notícias e autênticas provas matemáticas – e seus poderes de autoria criativa estão apenas começando a tomar forma” (FINN, 2017, p. 15). Como pondera Domingos (2015), atualmente se todos os algoritmos parassem de funcionar, seria o fim do mundo. Algoritmo é um conjunto de instruções matemáticas, uma sequência de tarefas para alcançar um resultado esperado em um tempo limitado. Os algoritmos antecedem os computadores – o termo remonta ao século IX ligado ao matemático al-Khwārizmi, cujo livro ensinava técnicas matemáticas a serem equacionadas manualmente. “Algorismus” era originalmente o processo de calcular numerais hindu-arábicos (FINN, 2017). Ed Finn define um algoritmo como “qualquer conjunto de instruções matemáticas para manipular dados ou raciocínios através de um problema” (ibid., p. 17). Para Ethem Alpaydin, “um algoritmo é uma sequência de instruções que são realizadas para transformar a entrada (*input*) na saída (*output*)” (2016, p. 14). Brian Christian e Tom Griffiths (2016) extrapolam o conceito para além do âmbito da Matemática: “Quando você cozinha pão a partir de uma receita, você está seguindo um algoritmo, o mesmo quando você tricota uma peça com base num determinado padrão. [...] Algoritmo faz parte da tecnologia humana desde a Idade da Pedra” (2016, p. 4). A ideia de associar algoritmo à receita de culinária, contudo, é contestada por Domingos (2015) para quem a receita não especifica exatamente a ordem e as etapas – quanto de açúcar, por exemplo, está contido em uma colherada. “Se quiséssemos programar um robô de cozinha para fazer um bolo, teríamos que dizer como reconhecer o açúcar do vídeo, como pegar uma colher e assim por diante. Portanto, uma receita culinária está muito longe de um algoritmo” (ibid., p.

3). O algoritmo requer instruções precisas e não ambíguas, o suficiente para serem executadas por um computador. “Algoritmos são um padrão exato. Costuma-se dizer que você realmente não entende algo até que possa expressá-lo como um algoritmo” (ibid., p. 4).

Os algoritmos têm sua própria complexidade, que Domingos (2015) agrupa em (a) complexidade do espaço, traduzida no número de bits de informação que precisa armazenar na memória do computador, (b) complexidade do tempo, traduzida no tempo necessário para “rodar”, i.e., quantas etapas à percorrer até produzir resultados, e (c) complexidade humana, traduzida nos limites do cérebro humano, inclusive para identificar os erros e corrigi-los.

Como mencionado anteriormente, a aprendizagem profunda é sobre “previsão” com base em correlações, e é sobre reduzir significativamente o custo de previsão. O objetivo não é identificar causalidades entre distintos fenômenos ou simples eventos, mas descobrir padrões e correlações que geram *insights*. “Antes do *big data*, nossa análise geralmente se limitava a testar um pequeno número de hipóteses que definíamos bem antes de coletar os dados. Quanto deixamos os dados falarem, podemos fazer conexões que nunca imaginamos que existissem” (MAYER-SCHÖNBERGER; CUKIER, 2013, p. 14).

Os algoritmos de aprendizado não funcionam da mesma forma, e suas diferenças impactam os resultados e, por vezes, o próprio modelo de negócio. Domingos (2015) compara os modelos de recomendação da Amazon e do Netflix: “Se cada um tivesse guiado você através de uma livraria física, tentando determinar o que é ‘certo para você’, a Amazon estaria propensa a levá-lo até às prateleiras que você frequentou anteriormente; a Netflix levaria você a seções estranhas da loja” (ibid., p. xvi). O algoritmo da Netflix tem uma compreensão mais profunda das preferências do usuário, e explica-se porque seu modelo depende de expandir a demanda para filmes e vídeos da “cauda longa”, que custam menos do que os *blockbusters*.¹¹ Pelo modelo de negócio da Amazon, a concentração nas mesmas preferências é positiva porque facilita a logística. Além disso, como clientes, as pesquisas indicam que estamos mais dispostos

¹¹ O valor da assinatura seria deficitário se os usuários escolhessem apenas filmes e vídeos “blockbusters”.

a ter uma chance em um item estranho num modelo de assinatura do que se tivermos que pagar por esse item individualmente (ibid.).

A sociedade dos dados e dos algoritmos

Os algoritmos de IA já estão presentes no nosso cotidiano. Parte do sucesso da Netflix, por exemplo, está em seu sistema de personalização, em que algoritmos analisam as preferências do usuário e de grupos de usuários com preferências semelhantes e, com base nelas, sugere filmes e séries. Acessamos sistemas inteligentes para programar o itinerário com o Waze, pesquisar no Google e receber do Spotify recomendações de músicas. A Siri da Apple, o Google Now e a Cortana da Microsoft, são assistentes pessoais digitais inteligentes que nos ajudam a localizar informações úteis com acesso por meio de voz. Existe uma multiplicidade de algoritmos de IA permeando as interações nas redes sociais, dentre eles os algoritmos do Feed de Notícias do Facebook.

Na sua operação diária, a Amazon captura grandes volumes de dados incluindo não apenas os livros que os usuários compram, mas quais os livros que eles só olham e por quanto tempo. Inicialmente, as recomendações derivaram de processar semelhanças entre clientes baseadas em amostras com resultados limitados; a partir de 1998, seu sistema de recomendação passou a buscar associações entre os próprios produtos¹² utilizando todos os dados com resultados mais precisos. “Quando a Amazon realizou um teste comparando a venda produzida por editores humanos com a venda produzida por conteúdo gerado por computador, os resultados nem chegaram perto. O material derivado e dados gerou muito mais vendas” (MAYER-SCHÖNBERGER; CUKIER, 2013, p. 51).

O crescimento exponencial dos dados inviabiliza a programação tradicional, remetendo inevitavelmente às técnicas de aprendizado de máquinas. A Amazon não pode codificar os gostos do conjunto de seus clientes em um programa de computador, assim como o Facebook desconhece como escrever um programa para identificar as melhores atualizações no Feed de Notícias. A Netflix pode ter cem mil títulos de DVD em estoque, mas se os clientes não souberem como encontrar suas

¹² Na ocasião, a Amazon registrou uma patente sobre “filtragem colaborativa item-a-item”, como a técnica é conhecida (MAYER-SCHÖNBERGER; CUKIER, 2013, p. 51).

preferências, o padrão será escolher os hits. A “cauda longa” só decola com os algoritmos de aprendizado (recomendação assertiva). Os algoritmos de aprendizado são os casamenteiros: eles encontram produtores e consumidores um para o outro com o melhor dos dois mundos: a diversidade de opções e o baixo custo da grande escala, com o toque da personalização associado aos pequenos. Um dos efeitos é a concentração de mercado: “Quem tem mais clientes acumula a maior parte dos dados, aprende os melhores modelos, conquista os novos clientes e assim por diante, em um círculo virtuoso” (ibid., p. 12).

Cada um de nós é, simultaneamente, um gerador e um consumidor de dados. “Queremos ter produtos e serviços especializados. Queremos que nossas necessidades sejam compreendidas e nossos interesses sejam previstos” (ALPAYDIN, 2016, p. 16). O paradoxo é que, ao mesmo tempo, queremos preservar a privacidade de nossos dados. O desafio colocado é encontrar um equilíbrio entre a abertura de dados, inclusive como enfrentamento da concentração de mercado, e a transparência sobre o uso dos dados.

Referências

AGRAWAL, Ajay; GANS, Joshua; GOLDFARB, Avi. *Prediction machines: the simple economics of Artificial Intelligence*. Boston, MA: Harvard Business Review Press, 2018.

ALPAYDIN, Ethem. *Machine learning*. Cambridge, MA: MIT Press, 2016.

CHRISTIAN, Brian; GRIFFITHS, Tom. *Algorithms to live by: the computer science of human decisions*. New York, NY: Henry Holt, 2017.

DOMINGOS, Pedro. *The master algorithm: how the quest for the ultimate learning machine will remake our world*. New York, NY: Basic Books, 2018.

FINN, Ed. *What algorithms want: imagination in the age of computing*. Cambridge, MA: MIT Press, 2017.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. *Big data: a revolution that will transform how we live, work, and think*. New York, NY: Houghton Mifflin Harcourt, 2013.

RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: a modern approach*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2009.

PENTLAND, Alex. *Social physics: how social networks can make us smarter*. New York, NY: Penguin, 2015.