

PROBLEMAS DA 'INTELIGÊNCIA' ARTIFICIAL

Valdemar W. Setzer

Prof. Titular Sênior, Depto de Ciência da Computação
Instituto de Matemática e Estatística da USP

1. Introdução

Devo declarar de início que vou utilizar, a contragosto, as denominações usuais, 'inteligência' artificial (IA) e 'aprendizado' de máquina (AM, *machine learning*). Isso se deve ao fato de que elas estão erradas e são enganosas. Não sabemos o que é inteligência, portanto não se pode classificar algum tipo dela como 'artificial'. Não sabemos como o ser humano aprende (se soubéssemos, o curso de medicina não levaria 6 anos), portanto, não se pode dizer que uma máquina aprende. No primeiro caso, uma denominação correta poderia ser 'simulação de comportamentos humanos'. No segundo caso, poderia ser 'programas baseados em cálculo de parâmetros'. Faço essas ressalvas porque essas denominações usuais produzem uma dupla mistificação: elevam as máquinas, os computadores, ao nível humano, e degradam a imagem que se deveria fazer do ser humano. Por detrás dessas denominações há uma concepção de mundo, a materialista ou fisicalista, que considera que só há substâncias e processos físicos no universo e no ser humano. Essa concepção leva à ideia de que o ser humano é uma máquina e age como uma, o que para mim é uma tragédia pior do que se considerar que o ser humano é um animal. Pode-se ter dó dos animais, pois eles sentem dor como nós; mas seria uma aberração psicológica ter dó de uma máquina, por exemplo, ter dó de desligar o computador. Existem sociedades protetoras dos animais, mas não há sociedades protetoras das máquinas. O que se está fazendo e se vai fazer com o ser humano, devido a essa mentalidade, será uma aplicação de uma das 'Leis de Setzer' (ver a referência [1]): "A miséria causada pelo ser humano ultrapassa o pior pessimismo."

Tenho muitas evidências para a minha hipótese de trabalho de que o ser humano não é um ser puramente físico, o que abordo em vários artigos disponíveis em meu *site*. Aliás, aplico a mesma hipótese para animais e plantas.

Quando se usam as denominações IA e AM, dever-se-ia estar consciente de que, se elas são usadas, degradam a imagem que se deveria fazer do ser humano e da natureza.

Para mais detalhes sobre IA e outros aspectos não abordados aqui, veja-se meu artigo a respeito [2].

2. Evolução da IA

A expressão *artificial intelligence* foi inventada pelo grande pioneiro da ciência da computação John McCarthy em 1955, por ocasião de um simpósio que houve no Dartmouth College, nos EUA [3, 4]. Ela suscitou muitas esperanças de que os computadores poderiam equiparar-se ao ser humano em todas as suas funções e ações, especialmente as intelectuais. Durante muito tempo, a IA não cumpriu as visões dos que tinham essa concepção, para decepção dos fisicalistas futuristas. Alguns sucessos foram totalmente mal interpretados, como o do xadrez eletrônico.

Considerava-se que jogar bem xadrez era uma mostra de inteligência. Quando a máquina Deep Blue da IBM venceu em 1997 o segundo torneio da empresa (a primeira máquina, que perdeu por 4 a 2, chamava-se, pouco modestamente, Deep Thought) contra o então campeão mundial Garry Kasparov [5], os visionários da IA cantaram vitória, inclusive afirmando que a máquina

tinha ultrapassado o ser humano. Essa última afirmação precisa se colocada em seu devido contexto: desde sua introdução em 1946 (com a máquina ENIAC) e o primeiro comercial em 1951 (UNIVAC) os computadores já haviam ultrapassado o ser humano em uma área muito restrita: fazer contas aritméticas repetitivas – com muito mais rapidez e confiabilidade. O mesmo passou-se com o xadrez eletrônico. Mas o que significou, realmente, o sucesso da Deep Blue? Ocorre que, em primeiro lugar, o desfecho do torneio foi de 2 vitórias da DB, uma do Kasparov e 3 empates. Em segundo lugar, o xadrez é um jogo matemático, isto é, o tabuleiro, as peças e as suas regras podem ser descritos matematicamente. Em terceiro lugar, a DB fazia 200 de milhões de previsões (cálculos) de lances por segundo. Em quarto lugar, Kasparov fazia mentalmente apenas algumas dezenas de lances.

Assim sendo, como foi possível que um ser humano pudesse empatar três partidas e vencer uma contra uma supermáquina matemática, jogando um jogo matemático, quando a máquina tinha capacidade milhões de vezes maior? Essa pergunta não foi formulada na época e constou de meu artigo sobre o xadrez eletrônico [6].

Minha conjectura é que isso foi possível porque Kasparov, ao contrário da DB, *não calculava!* Ele provavelmente olhava a posição do tabuleiro e tinha rapidamente a *intuição* do lance correto; aí ele certamente testava sua ideia, simulando mentalmente lances adiante, para ver se a intuição tinha sido correta. Ora, a intuição é uma ideia vinda do nada, portanto, algo inexplicável. De fato, os grandes mestres em xadrez não conseguem explicar como chegam mentalmente ao lance correto, de modo que nunca foi possível colocar as suas estratégias numa máquina. Qualquer concepção de que o cérebro trabalha – inconscientemente – e gera os pensamentos é mera especulação, pois não se sabe como o cérebro funciona. Mais recentemente, o programa AlphaZero usou técnicas de AM para, partindo do zero, apenas com as regras do jogo, jogar contra a própria máquina e calcular parâmetros que o fizeram ganhar de todas as outras máquinas de jogar xadrez.

3. ‘Aprendizado’ de máquina

O maior sucesso recente da IA é o ‘aprendizado’ de máquina (AM) [7]. Trata-se de uma técnica de dar a um programa uma enorme quantidade de conjuntos de dados de entrada e de saída; o programa então, em uma fase denominada de ‘treinamento’, calcula parâmetros para transformar um conjunto de dados de entrada em um conjunto de dados de saída correspondente, também fornecido ao programa. Posteriormente, depois de muitos casos desses, dá-se um novo conjunto de dados de entrada, obtendo-se então dados de saída que devem se aproximar o máximo possível do que seria esperado. Trata-se, portanto, de um problema de otimização matemática: diminuir o erro dos dados de saída em relação ao que seria de esperar. Uma das técnicas é usar uma ‘rede neural artificial’ [8] – novamente uma denominação errada: não se sabe como um neurônio funciona, e muito menos a rede dos neurônios no cérebro (calcula-se que haja 86 bilhões de neurônios e mais de um trilhão de conexões entre eles, as sinapses). As redes neurais artificiais contêm várias camadas de nós abstratos interligados. A primeira camada de nós recebe os dados de entrada, cada nó recebendo um dado. Na segunda camada de nós, cada nó é ligado a alguns nós da primeira camada; um cálculo combinando os dados de cada nó da primeira camada produz um resultado em cada nó da segunda camada, segundo uma fórmula que usa parâmetros de transformação do dado que sai de cada nó da camada anterior. Uma terceira camada tem cada um de seus nós ligado a alguns da segunda camada, recebendo os resultados desses últimos e combinando esses dados por meio de um cálculo. E assim por diante, até a última camada de nós, a de saída. Note-se que a topologia da rede, isto é, as interligações entre os nós, também podem ser consideradas como parâmetros da rede.

Um exemplo que se tornou clássico foi dar como entrada a uma rede os dados de muitas fotos digitalizadas de cães e de lobos. Para cada uma, especificava-se se a saída devia indicar um cão

ou um lobo. Depois de muito ‘treinamento’, a rede era capaz de prever com grande chance de acerto se uma nova foto era de um cão ou de um lobo. Voltaremos a esse exemplo mais adiante.

O AM está sendo usado em inúmeras aplicações, por exemplo, no sistema financeiro, para indicar se uma pessoa desejando tomar um empréstimo tem chance de honrá-lo ou de se tornar inadimplente. Ou em entrevistas para contratação de funcionários, dando-se o perfil de um candidato e o sistema concluindo se ele tem boa chance de ser um bom funcionário.

4. Problemas do ‘aprendizado’ de máquina

Há 3 problemas fundamentais no AM.

4.1 Uso de programas fechados

Muitas aplicações de AM utilizam programas fechados, isto é, o funcionário não sabe como o programa funciona, isto é, quantas camadas a rede tem, quantos nós existem em cada camada e como as saídas de cada nó são calculadas a partir dos nós da camada anterior a ele ligados. Isto é, o usuário de tal rede não sabe que critérios são usados para transformar os dados de entrada nos de saída.

4.2 Parâmetros desconhecidos

Em geral, usuários empregam um sistema de AM para transformar certos dados de entrada nos de saída, mas não examinam os parâmetros calculados pelos sistemas, isto é, não sabem como o sistema chega aos dados de saída. Em outras palavras, não se sabe que algoritmo o programa usa para chegar aos resultados.

O exemplo da distinção entre cães e lobos tornou-se famoso, pois os pesquisadores tiveram a curiosidade de examinar o que os parâmetros significavam. Tiveram a surpresa de verificar que as conformações das cabeças dos animais não estavam sendo usadas, como esperado. Em lugar disso, os parâmetros faziam o programa usar as imagens do ambiente: cães em geral estavam em ambientes domésticos, ruas etc., e as fotos dos lobos mostravam savanas, neve, florestas etc.

Somente em casos de pesquisas os parâmetros são examinados depois do ‘treinamento’.

4.3 Combinação dos dois fatores

Muitas aplicações de AM são usadas sem que o usuário saiba como funcionam os algoritmos de cálculos dos parâmetros e, além disso, não conhece esses parâmetros. Isto é, o programa é uma caixa preta.

5. Impacto social

Se um sistema de IA é utilizado para controlar máquinas, ele pode ser muito útil. Por exemplo, a quase totalidade dos acidentes de trânsito é devida a falhas humanas. Se um sistema de automação total de veículos diminuir os acidentes, seu emprego é recomendável. A respeito disso, é interessante fazer uma observação: aqui no Brasil, motoristas muitas vezes não respeitam as regras de trânsito, por exemplo, avançando em faixas em semáforos quando há congestionamentos. Um carro autônomo seguirá as regras, não vai ocupar essas faixas e com isso terá que esperar um tempo extenso até que o trânsito não esteja mais congestionado, o que provocará o desespero dos motoristas atrás dele. Ou todos os motoristas passam a respeitar as regras de trânsito, ou o carro autônomo terá que ser colocado em modo manual para também se poder desrespeitá-las.

O grande problema de IA ocorre quando um sistema é usado para tomadas de decisões sociais ou humanas, sem que seres humanos participem da última fase das decisões. É muito importante reconhecer que computadores não tomam decisões, eles fazem escolhas lógicas.

Decisões humanas envolvem sentimentos, e máquinas não têm sentimentos e em minha conjectura jamais terão [2]. Em particular, computadores são máquinas puramente sintáticas, máquinas de seguir regras matemáticas cegamente.

Se uma máquina é usada sozinha para substituir decisões humanas, ela trata as pessoas, às quais as escolhas dela se aplicam, como se fossem máquinas, como seriam os casos citados na seção 3, e mais aplicações desse tipo no sistema judiciário, diagnóstico e procedimentos médicos etc.

Como o ser humano tem sentimentos, toda lei social é estabelecida levando-se-os em conta. Um contraexemplo seria considerar-se que há excesso de população no mundo. Assim sendo, um raciocínio puramente lógico seria o seguinte: vamos passar leis permitindo pessoas matarem-se umas as outras, até que a população diminua. É por uma questão de sentimentos que uma tal lei parece revoltante – fora o fato de o legislador correr o risco de ser ele próprio morto...

Foi dito que os computadores são máquinas sintáticas. Isso significa que elas não têm semântica (compreensão) e muito menos pragmática, ambas características humanas. Por exemplo, em um julgamento no sistema judiciário poderiam ser levados em conta atenuantes, como por exemplo, uma mãe faminta roubar um pão para seus filhos. Um computador aplicará a lei dentro das regras que lhe foram dadas, e talvez essa atenuante ou outras, inclusive não previstas, não sejam parte do sistema.

Um exemplo é o sistema COMPAS, da Northpointe Inc., usado nos EUA para proferir sentenças judiciais, e que emprega um algoritmo ‘proprietário’, portanto secreto; o sistema calcula a probabilidade de uma pessoa vir a cometer novamente um crime. É conhecido o caso de um motorista guiando um carro roubado que não parou em uma perseguição policial, o que pela jurisprudência devia resultar em 2 ou 3 anos de cadeia. No entanto, examinando os antecedentes da pessoa, opinião de outros sobre ela etc., o sistema atribuiu-lhe 6 anos de prisão. Um recurso contra essa escolha da máquina não foi aceito [9].

Um outro exemplo é a geração de timbre de voz de uma pessoa, bem como movimentação dos lábios e expressões faciais que acompanham uma certa fala. Com isso é possível gerar vídeos extremamente convincentes de que alguém disse certa coisa que jamais pronunciou. Aqui não se trata de *fake news*, notícias falsas, mas de discurso falso (*fake voice*), o que está sendo denominado de *deep fake* (falsidade profunda). Tornaram-se conhecidos o vídeo falso de um discurso do ex-presidente Obama [10] e da presidente do Congresso americano Nancy Pelosi, em que ela aparece embriagada [11].

6. ‘Inteligência’ artificial vs. inteligência humana

Uma das diferenças fundamentais entre a ‘Inteligência’ artificial e a inteligência humana é que a primeira é sempre especializada. Por exemplo, os programas de jogar xadrez só jogam xadrez, não fazem mais nada. O sistema de AM relatado acima, para distinguir cães de lobos, não distingue mais nada. Tenho a impressão de que um sistema de AM que tivesse várias aplicações distintas iria resultar na maior confusão. Qualquer ser humano, ao contrário, tem uma inteligência extremamente multifacetada, por exemplo, distinguindo uma quantidade muito grande de animais, sabendo fazer múltiplas tarefas diferentes com as mãos etc.

Uma outra diferença fundamental é que os computadores são máquinas matemáticas, usando uma matemática muito particular, a discreta e finita. De fato, qualquer dado que esteja dentro de um computador tem que ser expresso numericamente, o que vale também para os programas, que também são dados. É impossível colocar em um computador o conceito de contínuo (por exemplo, dos números reais, ou os pontos de uma reta) e de infinito, apesar de serem usados na matemática geral. Não se pode afirmar que a inteligência humana seja discreta e finita. Aliás, a nossa memória é aparentemente ilimitada, pois jamais alguém teve a

vivência de querer memorizar algo (p.ex. um número de telefone, o nome de uma pessoa etc.) e não haver mais 'lugar' para isso. Isso é uma das indicações de que o ser humano tem elementos que transcendem a matéria, pois se nossa memória fosse simplesmente física, teria necessariamente que ser limitada, finita.

No meu artigo sobre IA [2], caracterizei dois tipos de inteligência: a 'incorporada' e a 'criativa'. Todas as máquinas têm uma inteligência incorporada, devida ao seu projeto, sua construção e funcionalidade, todos devidos a ações humanas. Mas também seres humanos a têm, na infinita sabedoria revelada pelo seu corpo físico. Animais e plantas idem. Mas até a Terra tem uma inteligência incorporada; por exemplo, se ela estivesse um pouco mais próxima ou afastada do Sol, não estaríamos aqui; se houvesse muito mais oxigênio na atmosfera, 'queimaríamos'. Se o eixo de rotação da Terra não fosse inclinado, não haveria estações do ano, e provavelmente também não estaríamos aqui. E também o mundo físico tem essa inteligência: um pequeno desvio nas constantes físicas, e até mesmo os átomos não existiriam.

Mas só o ser humano tem o que denominei de 'inteligência criativa'. É devido a ela que aparecem novas ideias e novos impulsos. É devido a ela que o ser humano está mudando as condições da Terra – infelizmente, em vários aspectos mais para pior do que para melhor.

7. Conclusão: os grandes perigos

E por falar em mudar para pior, o grande sucesso recente da IA, especialmente o AM, representa quatro enormes perigos para a humanidade.

Em primeiro lugar, o SH passar a confiar cada vez mais nas máquinas, substituindo decisões humanas pelas escolhas lógicas, frias, das máquinas.

Em segundo lugar, os dados de entrada que são usados podem ser viciados, levando a conclusões também viciadas. Por exemplo, uso de dados racistas levarão a escolhas racistas. Um outro exemplo é o de uso de dados apenas de homens, e depois os sistemas serem aplicados também a mulheres.

Em terceiro lugar, a tendência é de armazenar em computadores cada vez mais dados pessoais, usados posteriormente em escolhas. Por exemplo, a empresa Google traça um perfil de cada um de seus usuários, examinando para isso os *e-mails* que ele enviou pelo processador gmail, quais *sites* visitou usando o navegador Chrome etc. Mesmo se uma pessoa não usa nenhum produto da Google, se ela enviar um e-mail para alguém com um endereço do gmail, a Google já começa a traçar o perfil da segunda. Com esses perfis, o sistema de busca do Google elenca os resultados da busca segundo o perfil, sugerindo *sites* comerciais. Afinal, a Google desenvolveu um sistema genial: ela cobra pelo número de acessos que são feitos aos *sites* dos seus anunciantes.

Em quarto lugar, existe um grande perigo de dados pessoais ficarem armazenados eternamente. Por exemplo, alguma pessoa cometeu uma bobagem quando adolescente. Esse dado poderá ser usado por um sistema de IA muitos anos depois, na seleção de candidatos a algum emprego, sem que a pessoa saiba disso.

Em quinto lugar, notícias falsas e falas falsas com som subvertem totalmente a confiança que se tinha, especialmente na mensagem de uma pessoa devido ao seu timbre de voz. Já estamos naquilo que denomino 'era da mentira'. Isso vai abalar um dos pontos fundamentais do relacionamento humano, a confiança.

O que fazer? É absolutamente essencial que se saiba como cada programa e sistema funciona, isto é, quais são os critérios de escolha embutidos neles. Além disso, quaisquer dados pessoais devem poder ser examinados pelos cidadãos, e deve haver mecanismos judiciais rápidos (rapidez no sistema judiciário é, quem sabe, a maior utopia no Brasil...) para que eles sejam apagados quando inconvenientes, ou corrigidos. Finalmente, não se pode permitir que

empresas desenvolvam sistemas que substituam autonomamente decisões humanas; o funcionamento (algoritmos) dos programas que fornecem dados de apoio a decisões humanas devem ser conhecidos pelos seus usuários.

8. Referências

(Propositamente, para simplificar, não será seguido o padrão usual de referências bibliográficas)

- [1] Leis de Setzer. Acesso em 11/8/19: <https://www.ime.usp.br/~vwsetzer/jokes/leis.html>
- [2] Meu artigo sobre IA. Idem: <https://www.ime.usp.br/~vwsetzer/IAtrad.html>
- [3] Artigo bastante completo sobre histórico e detalhes da IA. Acesso em 10/8/19: <https://www.oezratty.net/wordpress/wp-content/themes/Ezratty5/forcedownload.php?file=/Files/Publications/Usages%20intelligence%20artificielle%202018%20Olivier%20Ezratty%20Compressed.pdf>
- [4] Proposta do simpósio no Dartmouth College. Idem: <https://web.archive.org/web/20070826230310/http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>
- [5] O torneio Kasparov vs. Deep Blue. Idem: <https://rafaelleitao.com/o-homem-e-a-mquina-o-match-kasparov-x-deep-blue/>
- [6] Meu artigo sobre xadrez eletrônico. Idem: <https://www.ime.usp.br/~vwsetzer/xadrez.html>
- [7] 'Aprendizado' de máquina. Idem: https://pt.wikipedia.org/wiki/Aprendizado_de_m%C3%A1quina
- [8] Redes 'neurais' artificiais, Idem: https://pt.wikipedia.org/wiki/Rede_neural_artificial
- [9] Artigo do New York Times sobre o caso do uso do sistema judiciário Compas e de outros sistemas. Idem: <https://www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html?module=inline>
- [10] Vídeo falso de Barack Obama. Idem: <https://www.youtube.com/watch?v=AmUC4m6w1wo>
- [11] Vídeo falso de Nancy Pelosi bêbada. Idem: <https://www.irishtimes.com/culture/tv-radio-web/the-fake-facebook-video-that-turned-nancy-pelosi-into-an-alcoholic-old-hag-1.3905862>