einstein
Publicação Oficial do Instituto Israelita de Ensino e Pesquisa Albert Einstein

ISSN: 1679-4508 | e-ISSN: 2317-6385

EDITORIAL

Desembalando a caixa preta

Unpacking the black box

Thiago Gonçalves dos Santos Martins[1], Paulo Schor[2]

[1] Universidade de Coimbra, Coimbra, Portugal.
[2] Universidade Federal de São Paulo, São Paulo, SP, Brasil.

DOI: 10.31744/einstein_journal/2021ED6037

Na atualidade, existem grandes bancos de dados (*Big Data*) de prontuários eletrônicos e imagens digitais, que permitem o reconhecimento de padrões em grandes volumes de informações em curto período de tempo, contribuindo para a criação de uma Medicina personalizada com o auxílio da inteligência artificial.[1,2]

Os algoritmos utilizados nos programas de diagnóstico médico são formados por uma rede neural profunda, que consiste em diversas camadas que se assemelham aos processos biológicos do córtex humano, no qual cada neurônio responde a um estímulo específico para uma região dentro de uma imagem, semelhante à forma como o neurônio cerebral responderia aos estímulos visuais, que ativariam um região específica do espaço visual. A rede neural aprende sozinha, conferindo valores a esses filtros, embora os parâmetros específicos, como número de filtros, tamanho do filtro, arquitetura de rede, ainda precisem ser definidos antes da fase de treinamento.

A fase na qual o dispositivo treina para melhorar suas previsões é a de aprendizado, que pode ser dividida em dois tipos: supervisionado e não supervisionado. Na aprendizagem supervisionada, os dados são atribuídos aos dados de treinamento, conforme são inseridos no computador, enquanto no aprendizado não supervisionado, o dispositivo cria seu próprio algoritmo de entrada. A fase de treinamento é seguida pela de validação dos dados, que devem idealmente ser diferentes dos usados na fase anterior. Nessa fase, a máquina consegue classificar os dados e começar a realizar previsões. Para essa fase de treinamento, um algoritmo precisa de milhares de dados para ter uma acurácia aceitável.[3,4]

Contudo, muitos pesquisadores ainda não conseguem explicar como os algoritmos chegaram a certas conclusões, e isso diminui um pouco a confiança do mundo científico no uso da inteligência artificial na Medicina, já que tendemos a refutar o que não conseguimos explicar. Esse termo é conhecido como "*black box*", na qual não temos acesso às informações do desenho interno e da implementação do algoritmo. Esse termo se opõe a "*white box*", na qual o componente é completamente exposto ao usuário. Entre esses algoritmos, temos a "*gray box*", que é quando temos acesso a alguns dados.[5]

Os algoritmos mais desenvolvidos são formados por uma estrutura não linear e com várias camadas em sua arquitetura, tornando possíveis previsões de alta complexidade. Contudo, isso dificulta a explicação de como chegamos a um resultado, o que é mais fácil em algoritmos mais simples e lineares.[6]

O acesso à informação dos algoritmos é difícil, já que esta não fica armazenada em um local específico, assim como a memória humana, que forma diversas sinapses durante o aprendizado. O aprendizado das máquinas acaba formando esses caminhos de aprendizado, fornecendo pesos diferentes aos filtros durante a fase de treinamento.

Como citar este artigo:
Martins TG, Schor P. Desembalando a caixa preta. einstein (São Paulo). 2021;19:eED6037.

Autor correspondente:
Thiago Gonçalves dos Santos Martins
Rua Botucatu, 821 – Vila Clementino
CEP: 04023-062 –São Paulo, SP, Brasil
Tel.: (11) 5521-2571
E-mail: thiagogsmartins@yahoo.com.br

Desvendar a *black box* poderia criar a possibilidade de enxergar a Medicina de outra forma. Alguns animais, como o beija-flor, possuem um número e tipo de cones diferentes dos seres humanos, desenvolvendo a capacidade de identificar cores imperceptíveis ao nosso olho, o que permite que eles enxerguem o mundo de outra forma.[7] Em analogia, podemos citar o funcionamento dos algoritmos, cuja lógica de processamento não e a mesma do ser humano. Sob essa ótica, mesmo atestando os resultados alcançados pelos algoritmos, por não entendermos o caminho percorrido, temos relutância em confiar nos resultados. Quando conseguimos explicar o funcionamento dos algoritmos, podemos justificar suas decisões e minimizar os erros, tornando a vulnerabilidade do mesmo mais perceptível. O aprendizado com a máquina pode ser exemplificado com o desenvolvimento de programas como o AlphaGo, que permitiu o ensino de novas técnicas de jogar *Go* . Esse jogo de tabuleiro conseguiu ser desvendado pelo aprendizado de máquina, tendo desempenho superior ao humano.[8] Dessa forma, aprendendo com as maquinas podemos ampliar nossos conhecimentos em outras áreas da ciência.

Já foram realizados alguns estudos procurando explicar o funcionamento de algoritmos. Caruana et al.[9] descreveram um algoritmo para diagnóstico de pneumonia e estudaram sua conduta com casos de pacientes reais. Lembrando que os algoritmos desenvolvidos para o diagnostico e acompanhamento de doenças devem ser preferencialmente validados com dados de populações diferentes que foram utilizados na etapa de treinamento e com a participação de pesquisadores independentes do seu desenvolvimento.[10]

Dessa forma, o conhecimento de como funciona o cérebro humano pode nos ajudar a entender melhor a *black box*, assim como o melhor entendimento do aprendizado da máquina pode aperfeiçoar nossas formas de entender o mundo, podendo enxergar a Medicina com “olhos de beija-flor”, e começar a compreender os mistérios da Medicina que, até então, não foram solucionados, além de possibilitar o desenvolvimento de melhores algoritmos que precisem de menos exemplos para seu aprendizado.

INFORMAÇÃO DOS AUTORES

Martins TG: http://orcid.org/0000-0002-3878-8564
Schor P: http://orcid.org/0000-0002-3999-4706

REFERÊNCIAS

1. Martins TG, Costa AL. A new way to communicate science in the era of Big Data and citizen science. einstein (São Paulo). 2017;15(4):523.
2. Martins TG, Costa AL, Martins TG. Big Data use in medical research. einstein (São Paulo). 2018;16(3):eED4087.
3. Martins TG, Francisco Kuba MC, Martins TG. Teaching ophthalmology for machines. Open Ophthalmol J. 2018;12:127-9.
4. Lu W, Tong Y, Yu Y, Xing Y, Chen C, Shen Y. Applications of artificial intelligence in ophthalmology: general overview. J Ophthalmol. 2018;2018:5278196. Review.
5. Suman RR, Mall R, Sukumaran S, Satpathy M. Extracting state models for Black-Box software components. J Object Technol. 2010;9(3):79-103.
6. Tan S, Sim KC, Gales M. Improving the interpretability of deep neural networks with stimulated learning. IEEE Workshop Autom, Speech Recognit. 2015;617-23.
7. Herrera G, Zagal JC, Diaz M, Fernández MJ, Vielma A, Cure M, et al. Spectral sensitivities of photoreceptors and their role in colour discrimination in the green-backed firecrown hummingbird (Sephanoides sephaniodes). J Comp Physiol A Neuroethol Sens Neural Behav Physiol. 2008;194(9):785-94.
8. Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, et al. Mastering the game of Go without human knowledge. Nature. 2017; 550(7676):354-9.
9. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. In: KDD ‘15: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2015. Pages 1721-1730 [cited 2020 Dec 3]. Sydney, NSW, Australia; 10-13 Aug. Available from: https://dl.acm.org/doi/10.1145/2783258.2788613
10. Faes L, Liu X, Wagner SK, Fu DJ, Balaskas K, Sim DA, et al. A Clinician’s Guide to Artificial Intelligence: How to Critically Appraise Machine Learning Studies. Transl Vis Sci Technol. 2020;9(2):7. Erratum in: Transl Vis Sci Technol. 2020; 9(9):33.