

INTELIGÊNCIA ARTIFICIAL DO APRENDIZADO DE MÁQUINA AO DEEP LEARNING: BREVE REVISÃO HISTÓRICA E CONCEITUAL

Sergio Di Fiore¹, Gustavo Kimura Montanha²Adriane Belluci Bellorio De Castro³

¹ Graduando em Análise e Desenvolvimento de Sistemas, Faculdade de Tecnologia de Botucatu
sergiodifiore@gmail.com,

² Docente da Faculdade de Tecnologia de Botucatu, gustavo.montanha@fatec.sp.gov.br,

³ Docente da Faculdade de Tecnologia de Botucatu, adriane.castro@fatec.sp.gov.br.

RESUMO

O objetivo do trabalho é, a partir de uma revisão de literatura, expor o desenvolvimento histórico e as principais características da Inteligência Artificial e destacar especificamente a tecnologia de *Deep Learning*. Para tanto, recorreu-se à pesquisa bibliográfica com demonstração da evolução de modelos de inteligência, evidenciando a diferença entre essas duas tecnologias.

Palavras-chave: Aprendizado de máquina. *Deep Learning*. Inteligência artificial.

1 INTRODUÇÃO

Por muito tempo, a humanidade sonhou com máquinas que fossem capazes de pensar de forma autônoma. Não surpreendentemente, as primeiras referências dos homens fazendo cópias deles mesmos, sem o auxílio dos deuses, vem da mitologia clássica Romana, Grega e Egípcia (NEWQUIST, 2018). Desde a Grécia antiga, há exemplos como Galatea, Talos e Pandora que hoje são entendidas como modelos de inteligências artificiais (MARTIN, 2004; BISPHAM, 2006; POMEROY; NEWQUIST, 2018). Quando os primeiros computadores começaram a ser concebidos, cem anos antes mesmo do primeiro ser construído, já existia a pergunta se máquinas poderiam algum dia ser inteligentes (GRUNSPAN, 2018). O presente trabalho mergulha em detalhes históricos relacionados a esta área de conhecimento, seus conceitos trazendo as raízes de sua evolução e detalhes do que é hoje conhecido como Aprendizado de Máquina e *Deep Learning*.

2 DESENVOLVIMENTO DO ASSUNTO

Em torno de 1940, alguns cientistas de computação imaginaram que a forma de se obter a resolução de problemas em nível humano seria tomando o modelo de como o cérebro humano trabalha e conectar isso de alguma forma em algum padrão. Em 1951

Marvin Minsky, hoje conhecido como o Pai da IA, que era então um recém graduado de Princeton, desenvolveu a Calculadora Estocástica Neural Analógica por Reforço (SNARC - *Stochastic Neural Analog Reinforcement Calculator*), uma rede de quarenta neurônios artificiais (GRUNSPAN, 2018).

Em 1950, Alan Turing desenvolveu o que chamado de "Teste de Turing": Uma máquina é capaz de imitar o processo de pensamento humano se ao ser interrogada por um indivíduo humano real (conversando, por exemplo), exibe habilidades cognitivas compatíveis com as humanas, e o indivíduo consegue interagir sem ser capaz de dizer que se trata de uma máquina (ALEMI, 2020).

Ainda na década de 1950, Minsky desistiu das redes neurais e começou a trabalhar principalmente na inteligência artificial simbólica que visa espelhar os níveis mais elevados humanos por meio de representação de símbolos e regras. Foi para o MIT e, em 1967, juntou-se a Seymour Papert nessa escola (GRUNSPAN, 2018).

A Inteligência Artificial é a ciência de fazer com que os computadores resolvam tarefas que requerem uma mente humana para serem solucionadas. O termo foi cunhado em 1955 por John McCarthy, Marvin Minsky, Nathaniel Rochester, e Claude Shannon em sua proposta para o “Projeto de Verão DartMouth para a Pesquisa da Inteligência Artificial”, um trabalho de dois meses e dez pessoas mantido pela faculdade DartMouth. Por esse, Minsky foi agraciado com o prêmio A. M. Turing em 1969 e McCarty em 1971 (GRUNSPAN, 2018).

O Departamento de Defesa norte-americano criou a agência ARPA que em 1963 cedeu US\$ 2 milhões para o projeto MAC (Machine-Aided Cognition), o primeiro de muitos projetos na área financiado por essa agência e envolvendo a pessoa de Minsky. Desse trabalho junto com a ARPA também saíram os programas e modelos de Minsky usados por psicólogos para entender o raciocínio (NEWQUIST, 2018).

Em 1965, McCarthy desenvolveu o projeto de programação heurístico (Heuristic Programming Project - HPP) usado por Edward Geigenbaum no Dendral de Joshua Lederberg, Bruce G. Buchanan, Edward Feigenbaum e Carl Djerassi. Foi a primeira incursão pelos sistemas especialistas, buscando uma forma de suporte às pesquisas de exobiologia. Por esse trabalho, Feigenbaum recebeu em 1994 o prêmio A. M. Turing (GRUNSPAN, 2018).

Em 1969, Minsky e Papert publicaram o livro “Perceptrons: An Introduction to Computational Geometry”, provando que as rede de então empregadas, as assim

denominadas Perceptron, poderiam somente executar funções básicas (SOMMERS, 2017).

Por volta de 2001, praticamente todas as empresas que fizeram a história até aqui resumida tinham quebrado. Os pioneiros como Minsky e McCarthy gozavam de suas aposentadorias. Sistemas especialistas e LISP deixaram de ser usados, embora em aplicações bem restritas ainda restavam em soluções tão improváveis como o corretor ortográfico do Microsoft Word, ou o portal de busca Alta Vista que oferecia o *Babel Fish*, um tradutor online. Foi quando a AI viu seu recomeço. Primeiro sugerindo rotas em sistemas GPS, depois a Siri, a Cortana e a Alexa, assistentes da Apple, Microsoft e Amazon respectivamente resgatando essa ciência do limbo através desses nichos. E, aos poucos, literalmente do nada, ressurgiu em todo o vigor (NEWQUIST, 2018).

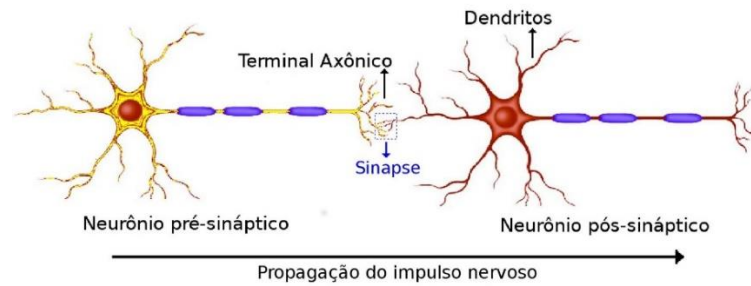
Geoffrey Hinton, hoje conhecido como o pai do *Deep Learning*, desenvolveu o treinamento de mais camadas, uma técnica que, iria além do tradicional aprendizado de máquina, permitindo encontrar pequenos padrões, que anteriormente não conseguiriam ser detectados (HAO, 2018). O seu projeto reconhecia imagens por padrões e não por lógica. Ele conseguiu que as redes neurais reconhecessem padrões pela combinação de um processo de treinamento inicial e então tentativa e erro com o uso de processadores não tradicionais. A novidade era que as suas redes neurais, possuíam diversas camadas, as chamadas camadas ocultas, o que foi chamado de *Deep Learning*. Logo se percebeu que um dos segredos desses sistemas era que funcionavam melhor, quanto mais exemplos de aprendizagem eles tinham (NEWQUIST, 2018).

Em 2011 o Watson da IBM vencia no programa Jeopardy, em 2015 a Google introduziu o TensorFlow, em 2016 o AlphaGo® venceu o mestre sul-coreano Lee Sedol no jogo de Go.

De uma maneira bem simples, entende-se por um sistema de IA como um sistema computacional capaz de aprender padrões os quais será capaz de reconhecer posteriormente (HAO, 2018). É importante ressaltar que, ao contrário do modelo computacional tradicional, no qual o sistema é instruído passo a passo como processar, a IA aprende por si só, o que é chamado de Aprendizado de Máquina, que é o primeiro conceito levantado por este trabalho.

Os sistemas de Aprendizado de Máquina reproduzem modelos que simulam o funcionamento do sistema nervoso humano onde neurônios processam um pacote de informações e o transmite para uma camada sucessiva o que continuam até que, em algum ponto posterior da cadeia, o pensamento é produzido (Figura 1).

Figura 1. Dois neurônios transmitindo um impulso nervoso

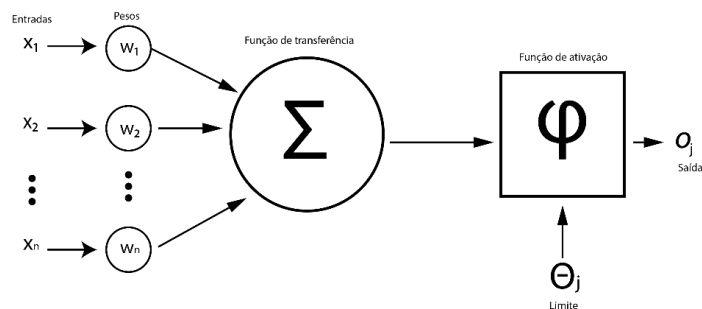


Fonte: <https://www.scielo.br>

Analogicamente, “neurônios artificiais”, ou simplesmente “nós”, serão responsáveis por um certo nível de processamento e, em seguida, transmitem os resultados para um seguinte neurônio artificial, que por sua vez também será responsável por alguma unidade de processamento.

A Figura 2 apresenta um nó dessa rede neural, que trabalha produzindo saídas binárias. Para as entradas (x_1, x_2, \dots, x_n) são atribuídos pesos (w_1, w_2, \dots, w_n) que serão processados pela chamada função de transferência e o seu resultado comparado a um certo limite predeterminado, a função de ativação. Excedido esse limite, a saída terá o valor 1, se não, zero.

Figura 2. Neurônio artificial



Fonte: Autor

Como é binário, pode-se entender que a sua saída é representada pela Equação 1:

Equação 1. Cálculo da saída de um neurônio

$$o_j = \begin{cases} 0 & \text{se } \sum x_n \cdot w_n \leq \Theta_j \\ 1 & \text{se } \sum x_n \cdot w_n > \Theta_j \end{cases}$$

É evidente que um único nó seria um fator decisório muito básico, motivo pelo qual ele toma formas mais complexas, conforme ilustrado na Figura 3, onde camadas

adicionais intermediárias são adicionadas, as ditas “camadas ocultas”. A existência dessas camadas ocultas junto à técnica de aprendizado ligada à essa configuração é o que passa a constituir uma “Rede Neural Profunda” ou então: “*Deep Learning*”, outro conceito elaborado por este presente trabalho.

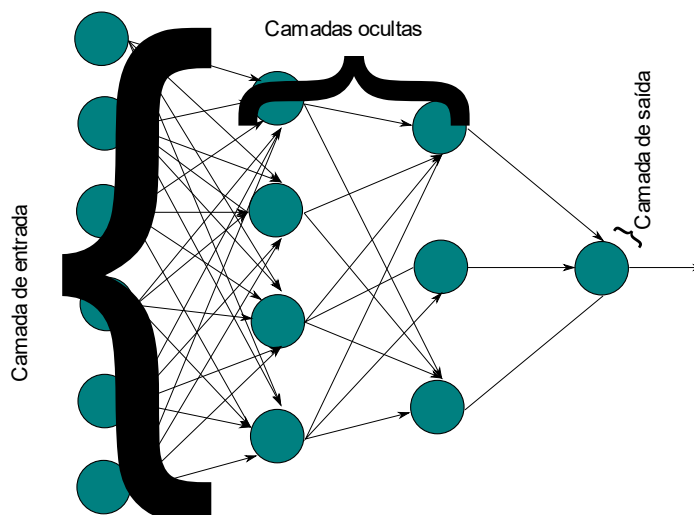
Como é possível inferir, outras diferentes composições estruturais são possíveis, dando origem a diferentes tipos de redes neurais e diferentes classificações. Um exemplo clássico, o neurônio sigmoide, usa como lógica Todas as redes apresentadas até o momento, tem como característica que a informação viaja só em um sentido, isso é da entrada para a saída. Essas redes são denominadas redes *FeedForward* e podem ser em camadas simples ou multicamadas.

Existem casos nos quais as saídas são retroalimentadas fazendo o papel de sinais de entrada para outros nós em sistemas com variação no tempo. Alguns tipos dessas redes são: Hopfield, e Perceptron Multicamada com Retroalimentação. Uma rede neural recursiva (*Recurrent Neural Network*) é capaz de retroalimentar os próprios nós (HAO, 2108).

Resumindo, o *Deep Learning* nada mais é do que o aprendizado de máquinas com “anabolizantes”, isto é, tem a capacidade de encontrar, e amplificar, mesmo os menores padrões. O grande diferencial está na técnica necessária para treiná-las, que foi desenvolvida, como mencionado, por Hinton.

A figura 3 traz um perceptron com duas camadas adicionais às de entrada e de saída, o que, como já dito, caracteriza uma *Deep learning*. A primeira camada pesa as entradas e toma alguma decisão. Neste exemplo, a saída dessas entradas é alimentada numa nova camada (a segunda do exemplo) que com os seus próprios pesos toma alguma outra decisão e encaminha para mais uma camada (no exemplo, a terceira) que repete o processo e encaminha para a camada de saída, para só então determinar o resultado de saída. Dessa forma esse perceptron de muitas camadas pode tomar decisões sofisticadas.

Figura 3. O Perceptron ou o Neurônio Sigmoide com multicamadas configurandou a Rede Neural Profunda



Fonte: Autor

2.1 ANATOMIA DE UM ALGORITMO DE APRENDIZADO

Pode-se entender um algoritmo de aprendizado como sendo composto por três partes:

- 1) Função custo;
- 2) Critério de otimização baseado na função de perda (função de custo, por exemplo); e
- 3) Rotina de otimização aproveitando dados de treinamento para encontrar uma solução para a otimização critério (BURKIV, 2019).

O desejável é que o algoritmo que se está buscando permita definir todos os pesos e tendência de tal forma que o resultado de saída da rede aproxime $y(x)$ para todos os x na entrada. Isto é quantificado pela função custo:

Equação 2. Função Custo

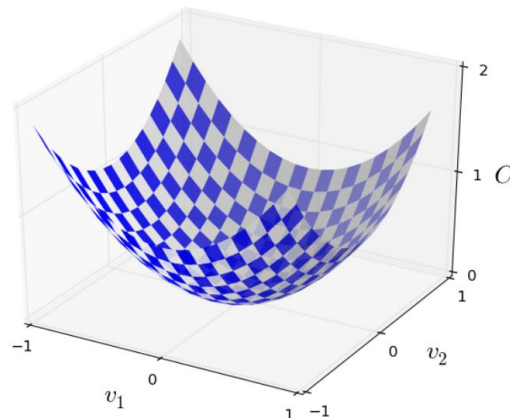
$$(w, b) \equiv \frac{1}{2n} \sum ||y(x) - a||^2$$

onde n é o conjunto de todos os pesos na rede, b todos as tendências, n o número total de entradas de treinamento, a é o vetor de saída da rede quando x é a entrada, e a soma é sobre todos os x de entrada.

2.2 O ALGORITMO DO MÉTODO DO GRADIENTE

O objetivo imediato ao treinar uma rede neural é encontrar valores para pesos e tendência que minimizem a função. Para fazê-lo é útil o exercício mental de se imaginar uma função de somente duas variáveis, v_1 e v_2 que pode ser facilmente representada graficamente como na figura 4:

Figura 4. Função de duas variáveis v_1 e v_2



Fonte: <http://neuralnetworksanddeeplearning.com/chap1.html>

Basicamente o que se procura é o mínimo dessa função, isto é, a equação 3

Equação 3. Vetor gradiente

$$\nabla C \equiv \left(\frac{\partial C}{\partial v_1}, \frac{\partial C}{\partial v_2} \right)^T$$

Onde se chega à equação 4.

Equação 4: Componente tendência com o gradiente de descida estocástico

$$b_l \rightarrow b'_l = b_l - \frac{\eta}{m_j} \sum \frac{\partial C_{x_j}}{\partial b_l}$$

(GOODFELLOW, 2016).

3 CONSIDERAÇÕES FINAIS

O Aprendizado de Máquina nada mais é do que uma entre muitas técnicas de fazer com que um sistema, basicamente estatístico, aprenda por si só a abstrair resultados a partir de um conjunto de dados de treinamento. Já o *Deep Learning*, redes com mais de duas camadas, as camadas ocultas, chega como uma teoria (e prática) que permite enfrentar desafios e modelos muito mais complexos, acompanhando a evolução das tecnologias de hardware com capacidade de processamento suficiente para enfrentá-las. Hoje o *Deep Learning* é uma tecnologia consistente e capaz de resolver eficientemente uma série de novos desafios.

4 REFERÊNCIA

ALEMI, M. **The Amazing Journey of Reason from DNA to Artificial Intelligence**. Springer International Publishing. Kindle Edition, 2020.

BURKIV, A. **The Hundred-Page Machine Learning Book**. Burkiv, Andiy, 2019
 BISPAM, E. HARRISON, T. SPARKES, B. A. **The Edinburgh Companion to Ancient Greece and Rome**. Edinburgh University Press, 2006

GRUNSPAN, S. L. **The Computer Book: From the Abacus to Artificial Intelligence.** Sterling, 2018

HAO, K. **What is machine learning?** MIT Technology Review. Novembro 2018.
Disponível em: <https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/>. Acesso em 18 set. 2021.

GOODFELLOW, *et al.* **Deep Learning.** MIT Press. 2016. Disponível em:
<http://www.deeplearningbook.org>. Acesso em 18 set. 2021.

MARTIN, C. **Ovid Metamorphoses.** W. W. Norton & Company. Kindle Edition, 2004

NEWQUIST, H.P. **The Brain Makers: The History of Artificial Intelligence – Genius, Ego, And Greed In The Quest For Machines That Think.** The Relayer Group. Kindle Edition, 2018.

POMEROY, S. *et al.* (2018). **Ancient Greece - A Political, Social, And Cultural History.** Oxford University Press, 2018.