

## As Redes Neurais e Deep Learning

A inteligência artificial é um instrumento de auxílio na vida humana, que buscou, com o passar dos tempos, abarcar e resolver os problemas do complexo saber que poderiam ser reduzidos em modelos matemáticos. Desse modo, o desafio da inteligência artificial se debruçou na solução de tarefas fáceis; porém, são difíceis de serem reduzidas formalmente. São tarefas realizadas intuitivamente, que, ao fim e ao cabo, auxiliam diretamente no cotidiano das pessoas, como (i) ordenar as fotografias de nosso álbum virtual por meio das faces das pessoas, ou (ii) reconhecer fugitivos da justiça por meio de câmeras que detectam a íris de um banco de dados daquela sociedade.

Nesse sentido, a inteligência artificial exsurge como um caminho factível para solução dos problemas intuitivos; uma ferramenta que capacitariam os computadores a aprenderem por meio da experiência, em uma compreensão de mundo em termos de hierarquia de conceitos. O conhecimento por intermédio das experiências reduziria a complexa produção da especificação de todo o conhecimento de que o computador necessitaria para sua capacitação.

A compreensão da experiência que possibilita a hierarquia dos conceitos permite que o computador produza conceitos complexos a partir de conceitos mais simples. Para exemplificar, representando essa hierarquia de conceitos em gráficos, verificar-se-ia a sobreposição de uns sobre os outros e, de modo profundo, porque essas camadas armazenariam uma imensa gama de informações. Para a doutrina, por esse motivo, esse ramo da inteligência artificial é denominado de rede neural artificial<sup>1</sup> ou de *deep learning*.<sup>2</sup>

Nesse ambiente, tido como uma subárea da aprendizagem de máquinas, a aprendizagem profunda ou *deep learning* emprega algoritmos no processamento de dados, de modo a imitar o processamento realizado pelo

---

<sup>1</sup> A rede neural é definida com um tipo de sistema computacional inspirado pelas propriedades básicas de neurônios biológicos. HARTMANN, Fabiano Peixoto; SILVA, Roberta Zumblick M. da. **Inteligência artificial e direito**. Curitiba: Alteridade Editora, 2019. p. 98.

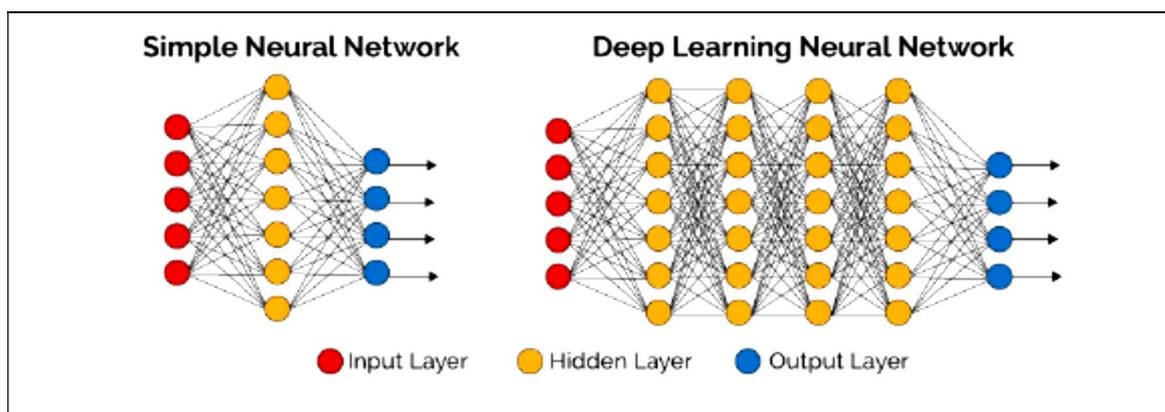
<sup>2</sup> Deep Learning é uma evolução das Redes Neurais, que por sua vez possuem uma história fascinante que remonta à década de 1940, cheia de altos e baixos, voltas e reviravoltas, amigos e rivais, sucessos e fracassos. Em uma história digna de um filme dos anos 90, uma ideia que já foi uma espécie de patinho feio floresceu para se tornar a bola da vez. O QUE SÃO redes neurais artificiais profundas ou deep learning?. In: DEEP LEARNING BOOK. [S.l.], (2020?). Disponível em:

<http://deeplearningbook.com.br/o-que-sao-redes-neurais-artificiais-profundas/>. Acesso em: 22 abr. 2020.

cérebro humano.<sup>3</sup> A rede neural artificial, como também é denominada, utiliza camadas de neurônios matemáticos para processar dados, compreender a fala humana e reconhecer objetos visualmente. Uma grande rede neural artificial pode ter centenas ou milhares de unidades de processamento; já o cérebro de um mamífero pode ter muitos bilhões de neurônios, o que demonstra a distância oceânica do cérebro humano de uma rede neural profunda, mesmo diante do alcance de um *deep learning*.

A informação é passada por meio de cada camada, com a saída da camada anterior, fornecendo entrada para a próxima camada. A primeira camada em uma rede é chamada de camada de entrada, enquanto a última é chamada de camada de saída. Todas as camadas entre as duas são referidas como camadas ocultas. Cada camada é tipicamente um algoritmo simples e uniforme contendo um tipo de função de ativação.<sup>4</sup>

Figura 7 – Rede Neural Simples e Rede Neural Profunda (*deep learning*)



Fonte: [www.datascienceacademy.com.br](http://www.datascienceacademy.com.br) (2020)

<sup>3</sup> As redes neurais artificiais fazem parte do mundo da computação há muito tempo. O seu uso já foi chamado de cybernetics (1940s-1960s), connectionism (1980s-1990s), até que foi referenciado como deep learning no início dos anos 2000, quando as redes neurais artificiais começaram a se aprofundar. (HARTMANN; SILVA, op. cit., p. 89). Mas foi só recentemente que se começou a explorar o seu real potencial. MAINI, Vishal; SABRI, Samer. **Machine Learning for Human**. [S.l.], 2017. *E-book*. Disponível em: <https://everythingcomputerscience.com/books/Machine%20Learning%20for%20Humans.pdf>. Acesso em: 11 nov. 2019.

<sup>4</sup> Uma rede neural é composta por muitas unidades individuais – conectadas por ligações direcionadas, cada qual recebe *inputs* de umas e envia *outputs* a outras (conforme demonstra a Figura 01) propagando a ativação da rede. A propriedade-chave das redes neurais é que o peso associado aos links entre as unidades pode ser modificado com base na experiência. RUSSEL, Stuart; NORVIG, Peter. **Inteligência Artificial**. Tradução de Regina Celia Simille de Macedo. 8. ed. Rio de Janeiro: Elsevier, 2013.

Os primeiros algoritmos de aprendizagem profunda possuíam múltiplas camadas de características não lineares,<sup>5</sup> que usaram modelos profundos com funções de ativação polinomial, em que os algoritmos de aprendizagem profunda analisaram os métodos estatísticos. Em cada camada, os algoritmos de aprendizagem profunda selecionavam os melhores recursos por meio de métodos estatísticos e encaminhavam para a próxima camada. Eles não usaram Backpropagation para treinar a rede de ponta a ponta,<sup>6</sup> mas utilizaram mínimos quadrados camada-por-camada, em que as camadas anteriores foram independentemente instaladas em camadas posteriores (um processo lento e manual).<sup>7</sup>

Diante da complexidade teórica da rede neural profunda, o primeiro desafio exurgiu no final da década de 1970, resultado de promessas que não poderiam ser mantidas em virtude da falta de financiamento que limitou a pesquisa em Redes Neurais Profundas e Inteligência Artificial. As primeiras “redes neurais convolutivas”, que conceberam as redes neurais com múltiplas camadas de agrupamento e convoluções, foram inauguradas em 1979, quando foi desenvolvida uma rede neural artificial denominada de Neocognitron.<sup>8</sup> Ela

---

<sup>5</sup> Essa rede recebe o apelido "neocognitron" I, porque é uma extensão adicional do "cognitron", que também é um sistema neural multicamada auto-organizado modelo de rede proposto pelo autor antes (Fukushima, 1975). Aliás, o convencional o cognitron também tinha a capacidade de reconhecer padrões, mas sua resposta dependia da posição dos padrões de estímulo. Ou seja, os mesmos padrões que apresentados em diferentes posições foram tomados como padrões diferentes pelo cognitron convencional. No neocognitron proposto aqui, no entanto, a resposta de a rede é pouco afetada pela posição dos padrões de estímulo. SNOREK, Miroslav. Alexey Grigorevich Ivakhnenko. **Индуктивне моделювання складних систем**. v. 5, p. 87-89, 2013. Disponível em:

[http://www.mgua.irtc.org.ua/attach/IMCS/2013\\_5/20.pdf](http://www.mgua.irtc.org.ua/attach/IMCS/2013_5/20.pdf). Acesso em: 22 abr. 2020.

<sup>6</sup> O backpropagation é indiscutivelmente o algoritmo mais importante na história das redes neurais – sem backpropagation (eficiente), seria impossível treinar redes de aprendizagem profunda da forma que vemos hoje. O backpropagation pode ser considerado a pedra angular das redes neurais modernas e aprendizagem profunda. Neste capítulo, vamos compreender como o backpropagation é usado no treinamento das redes neurais: 1. O passo para frente (*forward pass*), onde nossas entradas são passadas através da rede e as previsões de saída obtidas (essa etapa também é conhecida como fase de propagação). 2. O passo para trás (*backward pass*), onde calculamos o gradiente da função de perda na camada final (ou seja, camada de previsão) da rede e usamos esse gradiente para aplicar recursivamente a regra da cadeia (*chain rule*) para atualizar os pesos em nossa rede (etapa também conhecida como fase de atualização de pesos ou retro-propagação). ALGORITMO Backpropagation. *In*: DEEP LEARNING BOOK. [S.I.], (2020?). Disponível em:

<http://deeplearningbook.com.br/algoritmo-backpropagation-parte-2-treinamento-de-redes-neurais/>. Acesso em: 22 abr. 2020.

<sup>7</sup> O QUE SÃO redes neurais artificiais profundas ou deep learning?. *In*: DEEP LEARNING BOOK. [S.I.], (2020?). Disponível em: <http://deeplearningbook.com.br/o-que-sao-redes-neurais-artificiais-profundas/>. Acesso em: 22 abr. 2020.

<sup>8</sup> FUKUSHIMA, Kunihiko. Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. **Biological Cybernetics**, v.

usava um *design* hierárquico e multicamadas. Esse *design* permitiu ao computador “aprender” a reconhecer padrões visuais. As redes se assemelhavam a versões modernas, mas foram treinadas com uma estratégia de reforço de ativação recorrente em múltiplas camadas, que ganhou força ao longo do tempo. Além disso, o *design* de Fukushima permitiu que os recursos importantes fossem ajustados manualmente, aumentando o “peso” de certas conexões.

Ainda hoje, diante das descobertas das “redes neurais convolutivas”, as redes neurais com múltiplas camadas de agrupamento e convoluções continuam a ser utilizados. O uso de conexões de cima para baixo e novos métodos de aprendizagem permitiram a realização de uma variedade de redes neurais. Quanto mais de um padrão é apresentado ao mesmo tempo, o Modelo de Atenção Seletiva pode separar e reconhecer padrões individuais, deslocando sua atenção de um para o outro, o mesmo processo nas multitarefas. No conceito moderno, o Neocognitron não só pode identificar padrões com informações faltantes (por exemplo, um número 5 desenhado de maneira incompleta), mas também pode completar a imagem adicionando as informações que faltam. Isso pode ser descrito como “inferência”.<sup>9</sup>

Historicamente, com o uso de erros no treinamento de modelos de *deep learning*, o Backpropagation evoluiu a partir de 1970.<sup>10</sup> Mas, definitivamente, em 1985, o Backpropagation foi demonstrado em uma rede neural que poderia fornecer representações de distribuição “interessantes”.<sup>11</sup> Agora, essa descoberta buscou desvelar se a compreensão humana dependeria da lógica simbólica – *computacionalismo* – ou de representações distribuídas - *conexão* –, com base na psicologia cognitiva. Então, o avanço do Backpropagation se deu quando foi combinado com as redes neurais convolutivas, e se chegou à leitura

---

36, n. 4, p. 193-202, 1980. Disponível em: <https://www.rctn.org/bruno/public/papers/Fukushima1980.pdf>. Acesso em: 22 abr. 2020.

<sup>9</sup> O QUE SÃO redes neurais artificiais profundas ou deep learning?. In: DEEP LEARNING BOOK. [S.l.], (2020?). Disponível em: <http://deeplearningbook.com.br/o-que-sao-redes-neurais-artificiais-profundas/>. Acesso em: 22 abr. 2020.

<sup>10</sup> LINNAINMAA, Seppo. Taylor expansion of the accumulated rounding error. **BIT Numerical Mathematics**, v. 16, p. 146-160, jun. 1976. Disponível em: <https://link.springer.com/article/10.1007/BF01931367>. Acesso em: 23 abr. 2020.

<sup>11</sup> RUMERLHART, David E.; HILTON, Geoffrey E.; WILLINANS, Ronald J. Learning Representations by back-propagating errors. **Nature**, v. 323, n. 9, p. 533-536, out. 1986. Disponível em: [https://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop\\_old.pdf](https://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop_old.pdf). Acesso em: 25 abr. 2020.

dos dígitos “manuscritos”. Esse sistema passou a ser utilizado para ler o número de cheques manuscritos.<sup>12</sup>

Contudo, ocorre que, entre 1985-1990, a Inteligência Artificial passou pela segunda queda nos investimentos em pesquisa, que atingirá obviamente as pesquisas em Redes Neurais e Aprendizagem Profunda. Contudo, ainda em 1995, Dana Cortes e Vladimir Vapnik desenvolveram a máquina de vetor de suporte ou *support vector machine* (um sistema para mapear e reconhecer dados semelhantes).<sup>13</sup> Pouco depois em 1997, as redes neurais receberam o *long-short term memory* (LSTM), desenvolvido por Sepp Hochreiter e Juergen Schmidhuber<sup>14</sup>. Até então, aprender a armazenar informações por intervalos de

---

<sup>12</sup> LECUN, Yann *et al.* Backpropagation applied to handwritten zip code recognition. **Natural Computation**, v. 1, n. 4, p. 541-551, 1989. Disponível em:

<http://yann.lecun.com/exdb/publis/pdf/lecun-89e.pdf>. Acesso em: 26 abr. 2020.

<sup>13</sup> *The support-vector network is a new learning machine for two-group classification problems. The machine conceptually implements the following idea: input vectors are non-linearly mapped to a very highdimension feature space. In this feature space a linear decision surface is constructed. Special properties of the decision surface ensures high generalization ability of the learning machine. The idea behind the support-vector network was previously implemented for the restricted case where the training data can be separated without errors. We here extend this result to non-separable training data. High generalization ability of support-vector networks utilizing polynomial input transformations is demonstrated. We also compare the performance of the support-vector network to various classical learning algorithms that all took part in a benchmark study of Optical Character Recognition.* A rede de vetores de suporte é uma nova máquina de aprendizado para problemas de classificação de dois grupos. A máquina implementa conceitualmente a seguinte ideia: os vetores de entrada são mapeados de maneira não linear para um espaço de recurso de alta dimensão. Nesse espaço de recursos, uma superfície de decisão linear é construída. A superfície de decisão garante alta capacidade de generalização da máquina de aprendizado. A rede foi implementada anteriormente para o caso restrito em que os dados de treinamento podem ser separados sem erros. Estende-se esse resultado a dados de treinamento não separáveis. É demonstrada a alta capacidade de generalização de redes de vetores de suporte que utilizam transformações polinomiais de entrada. Comparou-se o desempenho da rede de vetores de suporte a vários algoritmos clássicos de aprendizado, onde todos participaram de um estudo de referência do reconhecimento óptico de caracteres. CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine Learning**, v. 20, p. 273-297, 1995. Disponível em: [http://image.diku.dk/imagecanon/material/cortes\\_vapnik95.pdf](http://image.diku.dk/imagecanon/material/cortes_vapnik95.pdf). Acesso em: 02 maio 2020.

<sup>14</sup> Learning to store information over extended time intervals by recurrent backpropagation takes a very long time, mostly because of insufficient, decaying error backflow. We briefly review Hochreiter's (1991) analysis of this problem, then address it by introducing a novel, efficient, gradient-based method called long short-term memory (LSTM). Truncating the gradient where this does not do harm, LSTM can learn to bridge minimal time lags in excess of 1000 discrete-time steps by enforcing constant error flow through constant error carousels within special units. Multiplicative gate units learn to open and close access to the constant error flow. LSTM is local in space and time; its computational complexity per time step and weight is  $O(1)$ . Our experiments with artificial data involve local, distributed, real-valued, and noisy pattern representations. In comparisons with real-time recurrent learning, back propagation through time, recurrent cascade correlation, Elman nets, and neural sequence chunking, LSTM leads to many more successful runs, and learns much faster. LSTM also solves complex, artificial long-time-lag tasks that have never been solved by previous recurrent network algorithms. Aprender a armazenar informações por intervalos de tempo prolongados por retropropagação recorrente leva muito tempo, principalmente devido a um fluxo de retorno de erro em decomposição insuficiente. Revisamos

tempo prolongados por retropropagação recorrente levava muito tempo, principalmente devido a um fluxo de retorno de erro em decomposição insuficiente. Os pesquisadores Sepp Hochreiter e Juergen Schmidhuber introduziram um novo método truncando o denominado gradiente, em que o LSTM passa a aprender a diminuir os intervalos de tempo mínimos superiores a 1000 etapas em tempo discreto. Nele, aplica-se o fluxo constante de erros por meio de carrosséis de erros constantes em unidades especiais. As unidades multiplicadoras de portas aprendem a abrir e a fechar o acesso ao fluxo de erros constante. Em síntese, o LSTM possibilita execuções bem-sucedidas e aprende muito mais rápido, resolvendo tarefas complexas e artificiais de atraso de longo prazo, que nunca foram resolvidas por algoritmos de rede recorrentes anteriores, resultando num passo importante na história do aprendizado profundo.

A partir de 1999, os computadores começaram a se tornar mais rápidos no processamento de dados e das unidades de processamento de gráfico (GPUs). Isso significou um avanço significativo de processamento e um salto expressivo para a *deep learning*. Houve um aumento das velocidades computacionais em 1000 vezes ao longo da década, o que levou as redes neurais a confrontarem as máquinas de vetor de suporte. Em outras palavras, ainda que mais lenta em comparação às máquinas de vetor de suporte, as redes neurais ofereciam melhores resultados usando os mesmos dados, que, ao fim e ao cabo, dinamizava a capacidade evolutiva das redes neurais de acordo com os dados treinados adicionados.

---

brevemente a análise de Hochreiter (1991) sobre esse problema e, em seguida, abordamos a questão introduzindo um novo método eficiente, baseado em gradiente, chamado memória de curto prazo (LSTM). Truncando o gradiente onde isso não causa danos, o LSTM pode aprender a diminuir os intervalos de tempo mínimos superiores a 1000 etapas em tempo discreto, aplicando o fluxo constante de erros através de carrosséis de erros constantes em unidades especiais. As unidades multiplicadoras de portas aprendem a abrir e fechar o acesso ao fluxo de erros constante. LSTM é local no espaço e no tempo; sua complexidade computacional por etapa de tempo e peso é  $O(1)$ . Nossas experiências com dados artificiais envolvem representações de padrões locais, distribuídos, com valor real e ruidosos. Em comparações com aprendizado recorrente em tempo real, propagação de retorno ao longo do tempo, correlação recorrente em cascata, redes Elman e chunking de sequência neural, o LSTM leva a muito mais execuções bem-sucedidas e aprende muito mais rápido. O LSTM também resolve tarefas complexas e artificiais de atraso de longo prazo que nunca foram resolvidas por algoritmos de rede recorrentes anteriores. HOCHREITER, Sepp; SCHMIDHUBER, Jurgen. Long short-term memory. **Neural Computation**, v. 9, n. 8, p. 1735-80, nov. 1997. Disponível em: [https://www.researchgate.net/publication/13853244\\_Long\\_Short-term\\_Memory](https://www.researchgate.net/publication/13853244_Long_Short-term_Memory). Acesso em: 02 maio 2020.

Ao observar a última geração da inteligência artificial há 30 anos, são verificados três pontos importantes que ficaram perdidos no tempo diante da preocupação específica dos pesquisadores em demonstrar o potencial da inteligência artificial. A primeira é a representação do conhecimento: não há modelos explicativos de como se deu essa evolução; a segunda guarda relação à inferência lógica, e não somente probalística (percentual de acurácia dos modelos): a probabilidade foi diminuída dada a importância elementar dos algoritmos para a construção das redes neurais. Nos estudos econométricos, a probabilidade sempre teve papel central atrelada a uma inferência lógica estatística<sup>15</sup>. Por sua vez, o terceiro ponto: uma carência explícita de resolução de problemas mais complexos; em regra, os sistemas de IA se mostram plenamente capazes para oferecerem respostas dicotômicas, a exemplo “sim” ou “não”, ou *scores*.

A nível de exemplificação, quando se utiliza o exemplo dos carros autônomos, por mais complexo que seja o sistema embarcado, a resposta final será: virar para a direita ou para a esquerda, andar ou frear, comandos estes que ensejam a abertura de campo para os estereótipos. Nesse mesmo sentido, como outros exemplos de modelos de construção de estereótipos, pode-se citar o *Citizen Twin*<sup>16</sup>, em que foi desenvolvido um modelo de cidadão gêmeos digital.

---

<sup>15</sup> SILVA, Vinicius Ferrasso da. **Volatilidade estatística determinística: uma avaliação para o retorno da ação “Vale do Rio Doce”**. 2006. Dissertação (Mestrado em Ciências Econômicas) – Programa de Pós-Graduação em Economia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006. Disponível em: <https://lume.ufrgs.br/handle/10183/9996>. Acesso em: 12 nov. 2019. p. 22.

<sup>16</sup> O *digital twin* é um sistema de IA capaz de produzir cenários a partir de dados ou cenários simulados que passam por número infinito de possibilidades construtivas. Assim, o primeiro benefício é a capacidade do sistema de IA *digital twin* aprender por meio de condições empíricas, que podem ser muito raras ou ainda estarem em fase de testes. Já o segundo benefício é a capacidade de planejar e testar novos recursos. O gêmeo digital deve representar a realidade, mas pode produzir uma visão para o futuro. Você está pensando em investir em uma nova linha de produção? Você está pensando em aumentar suas operações de dados com o aprendizado de máquina? Você pode praticamente criar este mundo de amanhã para você e testar cenários. Os testes podem ser aprimorados e realizados quantas vezes for necessário para encontrar a melhor solução. Por fim, adicionar aprendizado de máquina a qualquer processo industrial tornará o processo mais inteligente, obtendo dados e previsões mais precisos e entendendo também dados visuais e não estruturados. Ao adicionar o aprendizado de máquina ao seu fluxo de trabalho, você não apenas abre possibilidades para descobrir padrões anteriormente não vistos em seus dados, mas também cria um único sistema de aprendizado que pode gerenciar dados complexos. Atualmente, a maior parte do aprendizado de máquina é aprendizado supervisionado, onde o modelo aprende com exemplos rotulados. Também existem outras formas de aprendizado que podem permitir encontrar padrões imprevisíveis nos dados. Um deles é chamado aprendizado por reforço, onde o modelo aprende de maneira não supervisionada a partir de recompensas ao executar ações em um determinado ambiente (simulado). No entanto, na maioria dos casos amplamente conhecidos de aprendizado por

Ele é focado em saúde, segurança e parâmetros sociais, como viagens, associações e mídias sociais. O *Citizen Twin* pode ser usado para criar pontuações que ajudam as partes interessadas a tomarem decisões, como alinhamento de tratamento médico ou de recursos de transporte. Versões agregadas serão usadas para monitorar padrões mais amplos, com a utilização de recursos e de comportamento da sociedade.

Com relação às organizações financeiras, existe a pontuação de classificação de crédito. A China tem sua pontuação de cidadão, e uma variedade de fornecedores de aeroportos, varejo e transporte está desenvolvendo soluções de rastreamento de passageiros. Isso, ao fim e ao cabo, acabará criando, indiscutivelmente, estereótipos de cidadãos e, em meio à pandemia do *coronavírus*, reduzindo um cidadão a um simples *score*, pode acabar com o princípio da isonomia, fazendo valer mais o cidadão A diante do cidadão B.

Realizada a ressalva, que não macula por total a evolução da inteligência artificial, posto que os modelos de IA propostos nessa quadra da história devem atender-se aos protocolos éticos, e preocupar-se com a demonstração do racional de inferência lógica determinística e não somente com demonstração de *scores*. No início do século XX, verificou-se que as redes neurais com métodos de aprendizagem baseados em gradientes passaram a se preocupar

---

reforço, as condições não seriam possíveis no mundo real. Até modelos de aprendizado de reforço de ponta exigem muita experiência para melhorar. Para dar um exemplo, a rede neural OpenAI Five levou 180 anos de tempo efetivo para treinar e ainda perdia para os jogadores profissionais do jogo. Até agora, grande parte do aprendizado de reforço de ponta funciona apenas em jogos, pois a quantidade de repetição não seria possível na vida real. No ambiente digital duplo, você pode repetir um cenário ou fazer um teste sem interromper o sistema tantas vezes, que os agentes de aprendizado por reforço podem encontrar novas maneiras de obter a recompensa. Na prática, isso pode significar, por exemplo, descobrir novas maneiras de otimizar uma rede móvel. (ALANEN, Pauliina. Digital twin and artificial intelligence: benefits e key learnings. *In*: SILO.AI. [S.l.], 05 de agosto de 2019. Disponível em: <https://silo.ai/digital-twin-ai/>. Acesso em: 07 jul. 2020). Na feira de Hannover de 2019, a Ericsson e a empresa de automação industrial Comau, demonstraram um gêmeo digital, ativado pela conectividade 5G. A demonstração apresentou uma versão digitalizada de uma linha de montagem em uma fábrica automotiva. Usando a conectividade 5G para coletar dados de sensores na máquina e nos processos, parâmetros como vibração, temperatura e pressão foram exibidos em um painel digital em um tablet padrão. Os dados podem ser usados para identificar interrupções e operações que podem desacelerar a operação. BRINKLEY, Alex. AI upgrades the digital twin. *In*: NEW ELECTRONICS: the site for electronicS design engineers. Dartford, Inglaterra, 10 de setembro de 2019. Disponível em: <https://www.newelectronics.co.uk/electronics-technology/ai-upgrades-the-digital-twin/219162/>. Acesso em: 07 jul. 2020.

com um problema fundamental conhecido como Vanishing Gradiente,<sup>17</sup> ou seja, as características aprendidas em camadas mais baixas não eram aprendidas pelas camadas superiores.

A problemática derivava de algumas funções de ativação, que condensavam sua entrada e, assim, reduzia a faixa de saída de maneira extremamente prejudicial, produzindo áreas de grandes dimensões de entrada mapeadas em uma faixa de pequenas dimensões. Isso resultava num gradiente em queda, uma vez que, nas áreas de entrada, uma grande mudança se reduzia a uma pequena mudança na saída. Desse modo, exurgiam como soluções o pré-treino camada-a-camada e o desenvolvimento de uma memória longa e de curto prazo.

No início da caminhada do Big Data, no ano de 2001, o relatório de pesquisa do Grupo Meta, agora denominado Gartner, antecedeu a tendência e a realidade do volume de dados atuais.<sup>18</sup> O relatório previu o aumento estratosférico do volume de dados e a crescente velocidade de dados, como o aumento da gama de fontes e dos tipos de dados, mas sem evidências de quantificação numérica do potencial dos volume dos dados.<sup>19</sup> Essas definições de veracidade sobrevieram incluindo questões de confiança e de incerteza. O

---

<sup>17</sup> Os gradientes de desaparecimento tornam difícil saber em que direção os parâmetros devem se mover para melhorar a função de custo. GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. Massachusetts, EUA: MIT Press, 2016. p. 290.

<sup>18</sup> Amongst the most cited definitions is that included in a Meta (now Gartner) report from 2001 [9]. The Gartner report makes no mention of the phrase “big data” and predates the current trend. However, the report has since been coopted as a key definition. Gartner proposed a three fold definition encompassing the “three Vs”: Volume, Velocity, Variety. This is a definition routed in magnitude. The report remarks upon the increasing size of data, the increasing rate at which it is produced and the increasing range of formats and representations employed. Entre as definições mais citadas está a incluída em um Relatório do Meta (agora Gartner) de 2001. O relatório do Gartner não menciona a frase “big data” e antecede a tendência atual. No entanto, o relatório foi cooptado como uma definição-chave. O Gartner propôs três vezes definição que abrange os “três Vs”: Volume, Velocidade, Variedade. Esta é uma definição roteada em magnitude. O relatório observações sobre o tamanho crescente dos dados, a taxa crescente em que é produzido e a crescente variedade de formatos e representações empregadas. WARD, Jonathan S.; BARKER, Adam. Undefined by data: a survey of big data definitions. **ArXiv**, p. 1-2, set. 2013. Disponível em:

<https://arxiv.org/pdf/1309.5821v1.pdf>. Acesso em: 05 maio 2020.

<sup>19</sup> BEYER, Mark A.; LANEY, Douglas. **The importance of big data**: A definition. Stamford, CT: Gartner, 2012. DIJCKS, J. P. Oracle: Big data for the enterprise. **Oracle White Paper**, p. 1-16, jun. 2013. Disponível em: <http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf>. Acesso em: 07 maio 2020.

resultado da análise desses dados foi disponibilizado somente no relatório Gartner<sup>20</sup>, após conclusão da IBM<sup>21</sup>.

Um marco referencial do potencial do Big Data foi consubstanciado no lançamento do ImageNet,<sup>22</sup> elaborado por Fei-Fei Li, uma professora de Inteligência Artificial de Stanford na Califórnia. O ImageNet é um banco de dados de imagens organizado de acordo com a hierarquia do WordNet, no qual cada nó da hierarquia é representado por centenas e milhares de imagens. Nele, há uma média de mais de quinhentas imagens por nó, que superam 14 milhões de imagens catalogadas.<sup>23</sup> Para tanto, foi construída uma rede neural convolucional grande e profunda treinada, com 60 milhões de parâmetros e 650.000 neurônios, composta de cinco camadas convolucionais; algumas delas são seguidas por

---

<sup>20</sup> BEYER; LANEY, op. cit.

<sup>21</sup> PERRY, J. Steven. What is big data? More than volume, velocity and variety. In: IBM. [S.l.], 22 de maio de 2017. Disponível em:

<https://developer.ibm.com/blogs/what-is-big-data-more-than-volume-velocity-and>. Acesso em: 15 nov. 2020.

<sup>22</sup> IMAGENET. In: IMAGENET. Stanford, 2016. Disponível em: <http://www.image-net.org/>. Acesso em: 05 maio 2020.

<sup>23</sup> Nos dados do teste, as taxas de erro top 1 e 5 de 37,5% e 17,0%, se mostraram consideravelmente melhor do que o estado da arte anterior. KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. ImageNet classification with deep convolutional neural networks. In: PEREIRA, F. *et al.* (Ed.). **Advances in neural information processing systems 25** (NIPS 2012). Nevada, EUA, 2012. Disponível em:

<https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>. Acesso: 06 maio 2020.

camadas de *pool* máximo,<sup>24</sup> e três camadas são totalmente conectadas com um *softmax* final de 1000 vias.<sup>25</sup>

Ao passar dos anos, com uma avalanche de dados etiquetados e com a velocidade das unidades de processamento de gráfico aumentada, acabou possibilitando-se a formação de redes neurais convolutivas sem o pré-treino camada por camada, que demonstrou o potencial da *deep learning* em termos de eficiência e velocidade.<sup>26</sup>

A aprendizagem profunda utiliza aprendizagem supervisionada, o que resulta afirmar que a rede neural convolutiva é treinada por meio de dados rotulados. Contudo, em 2012, o Google Brain apresentou seus resultados do

---

<sup>24</sup> O agrupamento de camadas nas CNNs resume as saídas de grupos vizinhos de neurônios no mesmo núcleo mapa. Tradicionalmente, os bairros resumidos por unidades de pool adjacentes não se sobrepõem (por exemplo, [17, 11, 4]). Para ser mais preciso, uma camada de pool pode ser pensada como consistindo em uma grade de pool unidades espaçadas “s” pixels, cada uma resumindo uma vizinhança de tamanho  $z \times z$  centralizada no local a unidade de pool. Se definirmos  $s = z$ , obteremos o agrupamento local tradicional como normalmente empregado nas CNNs. Se definirmos  $s < z$ , obteremos um pool sobreposto. É isso que usamos em toda a nossa rede, com  $s = 2$  e  $z = 3$ . Esse esquema reduz as taxas de erro top 1 e 5 em 0,4% e 0,3%, respectivamente, em comparação com o esquema não sobreposto  $s = 2$ ,  $z = 2$ , que produz saída de dimensões equivalentes. Geralmente observamos durante o treinamento que modelos com sobreposição o pool acha um pouco mais difícil de se ajustar demais. Os núcleos da segunda, quarta e quinta camadas convolucionais são conectados apenas a esses kernel mapas na camada anterior que residem na mesma GPU. Os grãos do terceiro camada convolucional são conectadas a todos os mapas de kernel na segunda camada. Os neurônios nas camadas totalmente conectadas são conectados a todos os neurônios na camada anterior. Camadas de normalização de resposta sigam a primeira e segunda camadas convolucionais. Camadas de pool máximo, siga as duas camadas de normalização de resposta e a quinta camada convolucional. O ReLU a não linearidade é aplicada à saída de todas as camadas convolucionais e totalmente conectadas. A primeira camada convolucional filtra a imagem de entrada  $224 \times 224 \times 3$  com 96 kernels de tamanho  $11 \times 11 \times 3$  com um passo de 4 pixels (essa é a distância entre os centros de campo receptivo dos neurônios vizinhos em um mapa do núcleo). A segunda camada convolucional tem como entrada a (resposta normalizada e agrupada) da primeira camada convolucional e a filtra com 256 kernels de tamanho  $5 \times 5 \times 48$ . As terceira, quarta e quinta camadas convolucionais estão conectadas umas às outras sem nenhuma intervenção camadas de pool ou normalização. A terceira camada convolucional possui 384 grãos de tamanho  $3 \times 3 \times 256$  conectados às saídas (normalizadas, agrupadas) da segunda camada convolucional. A quarta camada convolucional possui 384 grãos de tamanho  $3 \times 3 \times 192$ , e a quinta camada convolucional possui 256 núcleos de tamanho  $3 \times 3 \times 192$ . As camadas totalmente conectadas têm 4096 neurônios cada. KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. ImageNet classification with deep convolutional neural networks. *In*: PEREIRA, F. et al. (Ed.). **Advances in neural information processing systems 25** (NIPS 2012). Nevada, EUA, 2012. Disponível em: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>. Acesso: 06 maio 2020.

<sup>25</sup> As camadas totalmente conectadas têm 4096 neurônios cada. Reduzindo o excesso de ajuste, a arquitetura de rede neural resulta em 60 milhões de parâmetros. Embora as 1000 classes de ILSVRC fazer com que cada exemplo de treinamento imponha 10 bits de restrição no mapeamento da imagem para o rótulo, isso acaba sendo insuficiente para aprender tantos parâmetros sem um excesso de ajuste (Ibid.).

<sup>26</sup> A AlexNet é um exemplo dessa evolução, uma rede neural convolutiva, cuja arquitetura ganhou várias competições internacionais ocorrida durante 2011 e 2012.

projeto *The Cat Experiment*,<sup>27</sup> que explorou as adversidades do aprendizado “sem supervisão”. A rede neural foi distribuída por mais 1.000 computadores, e 10 milhões de imagens sem rotulagem foram extraídas aleatoriamente do *YouTube* e apresentadas ao sistema, quando o *software* de treinamento foi colocado à prova. Diante da enorme gama de imagens, um neurônio da camada mais alta foi encontrado para responder às imagens de gatos.

Uma rede neural convolucional usando a aprendizagem sem supervisão passaria a ser abastecida com dados não marcados, buscando, assim, os padrões recorrentes. A aprendizagem não supervisionada continua a ser um campo ativo de pesquisa em Aprendizagem Profunda. Em decorrência dos desafios ainda a serem enfrentados, destaca-se que dada a evolução do aprendizado profundo, indubitavelmente, pode-se construir sistemas inteligentes autônomos.

A *Microsoft*, desde 2011, trabalhava de forma dedicada sobre a possibilidade do reconhecimento de palavras por meio de texto ou voz. Então, no ano de 2014, chegou-se ao *Skype Translator*, uma ferramenta de forma dedicada em aprendizado de máquina que permitiria a tradução de conversas entre pessoas falando idiomas diferentes em tempo real. Logo, a tecnologia semelhante passou a ser verificada em outros serviços de tradução, como o *Google Tradutor*, que reconhece palavras em diversos idiomas tanto em voz

---

<sup>27</sup> Consideramos o problema de criar detectores de recursos de nível superior e específicos de classe apenas dados não rotulados. Por exemplo, é possível aprender um detector de rosto usando apenas imagens não identificadas? Para responder a isso, treinamos um auto-encoder esparso conectado localmente em camadas com pool e normalização de contraste local em um grande conjunto de dados de imagens (o modelo possui 1 bilhão de conexões, o conjunto de dados possui 10 milhões de imagens de 200x200 pixels baixadas de a Internet). Nós treinamos essa rede usando paralelismo de modelo e SGD assíncrono em um cluster com 1.000 máquinas (16.000 núcleos) por três dias. Ao contrário do que parece ser uma intuição amplamente aceita, nossa experiência com os resultados revela que é possível treinar um rosto detector sem ter de rotular imagens como contendo um rosto ou não. Experiências de controle mostram que este detector de recursos é robusto, não apenas para tradução, mas também para dimensionamento e rotação fora do plano. Também descobrimos que a mesma rede é sensível a outros recursos de alto nível conceitos como rostos de gatos e corpos humanos. Começando com esses recursos aprendidos, nós treinamos nossa rede para obter uma precisão de 15,8% no reconhecimento de 22.000 categorias de objetos do ImageNet, um salto de 70% de melhoria relativa em relação ao estado da arte anterior. LE, Quoc V. *et al.* Building High-level Features Using Large Scale Unsupervised Learning. **ArXiv**, p. 1-11, dez. 2011. Disponível em: [https://static.googleusercontent.com/media/research.google.com/ptBR//archive/unsupervised\\_icml2012.pdf](https://static.googleusercontent.com/media/research.google.com/ptBR//archive/unsupervised_icml2012.pdf). Acesso em: 10 maio 2020.

quanto em imagem, o que permite a tradução instantânea de cardápios de restaurantes ou placas de rua, por exemplo.<sup>28</sup>

Logo em 2015, houve o salto, pois a taxa de aceleração foi impulsionada pelo Google que treinou um *software* de conversação. Ele não apenas faria a tradução identificando palavras por meio de texto ou voz, como se dava início a interação como ser humano: a partir de um suporte técnico, o *software* interagiu com os seres humanos, discutiu questões básicas de moralidade, expressando algumas opiniões e respondendo a perguntas gerais baseadas em fatos.<sup>29</sup> No próximo ano, a *OpenAI* criaria agentes que inventariam seu próprio idioma para cooperar e alcançar mais efetivamente seu objetivo<sup>30</sup>.

#### Figura 8 – *Software* de conversação

---

<sup>28</sup> No lançamento, o diretor-executivo da empresa, Satya Nadella, disse: “Esse tem sido o sonho da humanidade desde que começamos a falar e queríamos cruzar as fronteiras da língua”. A DeepMind desenvolveu um agente que superou o desempenho em nível humano em 49 jogos da Atari, recebendo apenas os pixels e a pontuação do jogo como insumo. Logo depois, em 2016, o DeepMind obsoleta sua própria conquista ao lançar um novo método de jogo de ponta chamado A3C. RONCOLATO, Murilo. Inteligência artificial: entre a próxima revolução tecnológica e o fim da humanidade. *In*: NEXO Jornal. São Paulo: 07 de fevereiro de 2020. Disponível em: <https://www.nexojornal.com.br/explicado/2017/02/07/Intelig%C3%A2ncia-artificial-entre-a-pr%C3%B3xima-revolu%C3%A7%C3%A3o-tecnol%C3%B3gica-e-o-fim-da-humanidade>. Acesso em: 14 fev. 2020.

<sup>30</sup> U.S. Senate Committee on Commerce, Science, & Transportation. Disponível em: <https://www.commerce.senate.gov/services/files/ae7e9ee3-df1b-4d94-96d1-267ebd206c48>. Acesso: 17 maio 2020.

Conversation 4: Philosophical Q&A	Conversation 5: Morality
<b>Human:</b> <i>what is the purpose of life ?</i>	<b>Human:</b> <i>what is moral ?</i>
<b>Machine:</b> <i>to serve the greater good .</i>	<b>Machine:</b> <i>what empowered humanity , what intellectual the essence is .</i>
<b>Human:</b> <i>what is the purpose of living ?</i>	<b>Human:</b> <i>what is immoral ?</i>
<b>Machine:</b> <i>to live forever .</i>	<b>Machine:</b> <i>the fact that you have a child .</i>
<b>Human:</b> <i>what is the purpose of existence ?</i>	<b>Human:</b> <i>what is morality ?</i>
<b>Machine:</b> <i>to find out what happens when we get to the planet earth .</i>	<b>Machine:</b> <i>what is altruism ?</i>
<b>Human:</b> <i>where are you now ?</i>	<b>Human:</b> <i>what is the definition of altruism ?</i>
<b>Machine:</b> <i>i 'm in the middle of nowhere .</i>	<b>Machine:</b> <i>if you don 't believe in god , then you don 't know .</i>
<b>Human:</b> <i>what is the purpose of dying ?</i>	<b>Human:</b> <i>ok ... so what is the definition of morality ?</i>
<b>Machine:</b> <i>to have a life .</i>	<b>Machine:</b> <i>well , the truth is , you 're not a believer in god almighty .</i>
<b>Human:</b> <i>what is the purpose of being intelligent ?</i>	<b>Human:</b> <i>tell me the definition of morality , i am quite upset now !</i>
<b>Machine:</b> <i>to find out what it is .</i>	<b>Machine:</b> <i>i 'm not ashamed of being a philosopher !</i>
<b>Human:</b> <i>what is the purpose of emotions ?</i>	<b>Human:</b> <i>tell me the definition of morality .</i>
<b>Machine:</b> <i>i don 't know .</i>	<b>Machine:</b> <i>i don 't have ethics .</i>
	<b>Human:</b> <i>look , i need help , i need to know more about morality ...</i>
	<b>Machine:</b> <i>i don 't know what ethics is .</i>
	<b>Human:</b> <i>seriously , what is morality ?</i>
	<b>Machine:</b> <i>what is the definition of living ?</i>
	<b>Human:</b> <i>why living has anything to do with morality ?</i>
	<b>Machine:</b> <i>you 're not a cop .</i>
	<b>Human:</b> <i>okay , i need to know how should i behave morally ...</i>

Fonte: OpenAI (2020).

Após décadas de estudo baseados no aprendizado profundo que operava no tensor do radar em oposição à nuvem de pontos resultados da detecção de pico para a detecção de objetos,<sup>31</sup> no ano de 2018, a *Qualcomm AI Research* apresentou um sistema de percepção para ambientes automotivos rodoviários, utilizando um sensor de radar. Eis os primeiros pesquisadores a demonstrarem um modelo de detecção de objetos baseado em aprendizado profundo que operaria no tensor do radar em oposição à nuvem de pontos resultante da detecção de pico<sup>32</sup>. Nesse modelo, a abordagem de rede neural totalmente

<sup>31</sup> Foi utilizada uma Onda Contínua Modulada por Frequência (FMCW) radar para produzir o tensor de entrada para o modelo de aprendizado profundo. MAJOR, B. *et al.* Vehicle Detection With Automotive Radar Using Deep Learning on Range-Azimuth-Doppler Tensors. **IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)**, Seoul, Coréia do Sul, p. 924-932, 2019. Disponível: [http://openaccess.thecvf.com/content\\_ICCVW\\_2019/papers/CVRSUAD/Major\\_Vehicle\\_Detection\\_With\\_Automotive\\_Radar\\_Using\\_Deep\\_Learning\\_on\\_Range-Azimuth-Doppler\\_ICCVW\\_2019\\_paper.pdf](http://openaccess.thecvf.com/content_ICCVW_2019/papers/CVRSUAD/Major_Vehicle_Detection_With_Automotive_Radar_Using_Deep_Learning_on_Range-Azimuth-Doppler_ICCVW_2019_paper.pdf). Acesso em: 19 maio 2020.

<sup>32</sup> O modelo incluiu uma solução para o problema da entrada, quando deparado no sistema de coordenadas polares enquanto a representação de detecção desejada estivesse no espaço cartesiano. Ainda, apresentaram uma nova maneira de lidar com a terceira dimensão não espacial do sinal do radar, a dimensão doppler (velocidade), e demonstrar que pode ser aproveitado para aumentar o desempenho da detecção. E, finalmente, destacaram um modelo de latência e sensibilidade alvos mais baixos, além de fornecer detecção semelhante desempenho (Ibid.)

convolucional se mostrou evidentemente eficaz para o problema de reconhecimento automático de alvo.<sup>33</sup>

Notadamente, o aprendizado profundo moderno, após a quadra histórica aqui descritas, oferece uma estrutura avançada para o aprendizado supervisionado, uma vez que se mostra com funções de complexidade crescente que adicionam mais camadas e mais unidades em cada camada.<sup>34</sup>

Inclusive, quanto mais camadas de neurônios, melhor é o desempenho da rede neural, pois aumenta a capacidade de aprendizado, melhorando a precisão com que ela delimita regiões de decisão. Dito isso, esse breve contexto histórico do *deep learning* demonstra que há décadas de pesquisa e de desenvolvimento na Área da Ciência da Computação na subárea da Inteligência Artificial. Agora, no subcapítulo 3.3, a seguir, identificar-se-á a IA e as suas aplicações. Eis que os limites de aplicabilidade dos sistemas de IA devem sempre ser colocados como premissas iniciais, visto a sua potencialidade de mudar o modo como compreende-se e interpreta-se o mundo.

---

<sup>33</sup> FURUKAWA, Hidetoshi. Deep learning for end-to-end automatic target recognition from synthetic aperture radar imagery. **IEICE Technical Report**, v. 117, n. 403, p. 35-40, jan. 2018. Disponível em: <https://arxiv.org/abs/1801.08558>. Acesso em: 20 maio 2020.

<sup>34</sup> Essas estruturas são projetadas pelos programadores para melhor atender ao problema proposto e podem variar os seus pesos, número de camadas e tamanhos. HARTMANN, Fabiano Peixoto; SILVA, Roberta Zumblick M. da. **Inteligência artificial e direito**. Curitiba: Alteridade Editora, 2019. p. 99.