

Gabriel Yuri Hosaki

Aluno ADS/FATEC Taquaritinga
gabrielyuri42@gmail.com

Douglas Francisco Ribeiro

Professor/FATEC Taquaritinga
douglas.ribeiro@fatectq.edu.br

RESUMO

Aprendizado profundo atualmente é considerado a principal técnica de aprendizado de máquina na solução de problemas, particularmente em problemas de classificação. Contudo, há ainda muita dificuldade no entendimento e utilização dessas técnicas. Nesse trabalho pretende-se apresentar as Redes Neurais Profundas e as Redes Convolucionais. Como resultado desta pesquisa, busca-se um entendimento mais sistematizado sobre o funcionamento dessas técnicas. Pode-se concluir que a partir das modernas técnicas de aprendizado de máquina existentes, os problemas envolvendo variáveis mais difíceis de serem descritas para a máquina como, por exemplo, o instinto humano, estão ao alcance com as infinitas possibilidades de aprendizado, como se a própria máquina fosse uma criança dando seus primeiros passos.

Palavras-chave: Aprendizado de Máquina. Redes Neurais Profundas. Redes Convolucionais. Aprendizado profundo.

ABSTRACT

Deep learning is currently considered the main machine learning technique in solving problems, particularly in classification problems. However, there is still much difficulty in understanding and using these methods. In this work we intend to present the Deep Neural Networks and the Convolutional Networks. The main focus will be on understanding how these methods work. One may conclude that, based on state-of-the-art machine learning techniques, it's possible to solve problems containing a more complex logic for computers such as the human instinct through infinite learning possibilities. Just like if the machine itself was a child giving its first steps.

Keywords: Machine Learning. Deep Neural Networks. Convolutional Networks. Deep learning.

Correspondência/Contato

Faculdade de Tecnologia de Assis - FATEC

Rua Senhor do Bonfim, 1226
CEP 19802-130
Fone (18) 3322-3941
Rgecontato.fatecassis@fatec.sp.gov.br
<http://fatecassis.edu.br>

Editores responsáveis

Taciana Maria Lemes de Luccas
taciana.luccas@fatec.sp.gov.br

Rafael Oliva
rafael.oliva@fatec.sp.gov.br

1 INTRODUÇÃO

Métodos de aprendizado profundo possibilitaram a evolução de diversas áreas de aprendizado de máquina especialmente por duas razões: disponibilidade de bases de dados com milhares de dados e computadores eficazes no processamento e em tempo consideravelmente reduzido (DENG et al., 2009; RUSSAKOVSKY et al., 2015).

Estudos anteriores pesquisaram a utilização de representações hierárquicas com redes neurais como o neocognitron de Fukushima (1988) e a rede neural para reconhecimento de dígitos de LeCun et al. (1998).

O interesse neste tema surgiu apenas após a difusão do artigo e código-fonte que possibilitava a implementação da rede neural convolucional AlexNet (KRIZHEVSKY, 2012). Os efeitos apontavam para um conclusivo desempenho estatístico na classificação de imagem onde o aprendizado profundo popularmente conhecido como Deep Learning passou a ser consumido entre as áreas de Visão Computacional, Processamento de Imagens, Computação Gráfica e Redes Neurais convolucionais consideradas como base para métodos do estado da arte em diversas aplicações.

De acordo com Deng et. Al. (2009), o projeto ImageNet teve uma importante contribuição nesse processo em encontrar um modelo que superasse o modelo anterior na tarefa de classificar e segmentar imagens, reconhecer objetos entre outros desafios.

Deep Learning oferece um conjunto de métodos para analisar sinais como áudio e fala, conteúdos visuais, incluindo imagens e vídeos, e ainda conteúdo textual. Entretanto, tais métodos incluem diversos modelos, componentes e algoritmos trazendo dificuldade para que um leigo entenda e acompanhe os estudos recentes.

Embora seja considerável o uso de redes neurais profundas entre os cientistas para resolução de problemas, há também uma dificuldade na compreensão desses métodos.

2 OBJETIVO

Nesse trabalho pretendemos revisar os conceitos e conhecimentos que contemplam Deep Learning e as redes neurais convolucionais. O foco principal será no entendimento do funcionamento desses métodos.

3 METODOLOGIA

Para realização deste artigo, foram revisados artigos científicos que contemplassem a temática por meio das bases de dados Scielo e Scholar Google sendo selecionados aqueles que eram pertinentes ao assunto com o objetivo de apresentar os conceitos que contemplam: Machine Learning e Deep Learning. Desta forma, este estudo pretende contribuir para a consolidação do conhecimento dentro deste contexto.

4 FUNDAMENTAÇÃO TEÓRICA

4.1 Conceitos sobre Deep Learning

Deep learning trata-se de um conjunto de técnicas de Machine Learning que utilizam redes neurais artificiais profundas, com muitas camadas intermediárias entre a camada de entrada e a de saída (LECUN et al., 2015).

Essa técnica tem como diferencial tecnológico os melhores resultados obtidos na resolução de problemas, superando o desempenho dos melhores especialistas em áreas como reconhecimento de características semânticas em imagens, jogos de estratégia como xadrez.

O surgimento do Deep Learning iniciou uma nova área de pesquisa de aprendizagem de máquina a partir do ano de 2006 (LIU et al., 2017; BENGIO 2009; LECUN et al., 2015). Atualmente técnicas de pesquisas de aprendizagem profunda têm impactado em uma série de trabalhos de processamento e reconhecimento de sinais e padrões.

É muito perceptível a abertura de novos campos de problemas que podem ser resolvidos com Deep Learning, especialmente os da aprendizagem de máquina e da inteligência artificial (HINTON et al. 2012; YANG et al., 2015).

As principais empresas do mercado tecnológico estão fazendo altos investimentos no estudo de Deep Learning e gerando novas aplicações.

Deep Learning utiliza algoritmos de aprendizado de máquina que trabalham em vários níveis, com diferentes graus de abstração. Esses modelos aprendidos correspondem a diferentes graus de conceitos, onde os conceitos de alto nível são definidos a partir de níveis inferiores, e os mesmos conceitos de nível inferior ajudam definir muitos conceitos de nível superior concentrando a aprendizagem de múltiplos níveis de representação e abstração que ajudam a dar sentido aos dados, como imagens, sons e textos.

Na literatura sobre Deep Learning, destacam-se alguns conceitos como sendo modelos que consistem em múltiplas camadas ou estágios de processamento de informação não linear; e também são métodos para a aprendizagem supervisionada ou não supervisionada da representação de características em camadas consecutivamente mais altas e mais abstratas. O Deep Learning abrange as áreas de pesquisa de redes neurais, inteligência artificial, modelagem gráfica, otimização, reconhecimento de padrões e processamento de sinais.

Dentre as razões mais importantes para a notoriedade atual da aprendizagem profunda destacam-se: o aperfeiçoamento da capacidade de processamento de chips, crescimento do tamanho dos dados usados em treinamentos e os avanços atuais em pesquisas em aprendizagem de máquina e processamento de sinais e esses avanços possibilitaram que tais métodos analisassem de forma eficaz funções complexas e não lineares, aprendessem representações de recursos distribuídos e hierárquicos, além de permitirem o uso eficaz de dados rotulados e não rotulados (CHUNG et al., 2015; LECUN et al., 2010).

Deep Learning refere-se a uma classe ampla de técnicas e arquiteturas de aprendizagem de máquina que usam muitas camadas de processamento de informações não lineares de natureza hierárquica. Dependendo de como as arquiteturas e técnicas são destinadas para uso como síntese/geração ou reconhecimento/classificação, pode-se categorizar amplamente a maior parte do trabalho nessa área em três grupos principais (BENGIO, 2009; PANG et al., 2017):

1. Redes profundas para aprendizado não supervisionado, que se destinam a capturar a correlação de alta ordem dos dados visíveis para análise de padrões quando não há informações sobre os rótulos das classes disponíveis na base. Aprendizado não supervisionado na literatura refere-se à categoria de redes profundas. Quando usado no modo generativo, também busca a caracterização de distribuições estatísticas conjuntas dos dados visíveis e suas classes associadas quando disponíveis, podendo ainda serem tratadas como parte dos dados visíveis.

2. Redes profundas para aprendizagem supervisionada, que se destinam a fornecer diretamente poder discriminativo para fins de classificação de padrões, frequentemente caracterizando as distribuições a posteriori de classes condicionadas aos dados visíveis. Os dados do rótulo-alvo estão sempre disponíveis de forma direta ou indireta para essa aprendizagem supervisionada. Elas também são chamadas de redes profundas discriminativas.

3. Redes profundas híbridas, se destacam pela discriminação que é assistida, muitas vezes de forma significativa, com os resultados de redes profundas generativas ou não supervisionadas. Isso pode ser alcançado através de uma melhor otimização e regularização das redes profundas supervisionadas. A proposta dessa rede também pode ser alcançada quando critérios discriminativos para a aprendizagem supervisionada são usados para estimar os parâmetros em qualquer das redes profundas generativas ou não supervisionadas (GOODFELLOW et al., 2016).

Para entender o funcionamento do Deep Learning, necessitamos de alguns pré-requisitos que abrangem conhecimento básico de Aprendizado de Máquinas e Processamento de Imagens, em particular conhecimentos básicos sobre aprendizado supervisionado, classificação, redes neurais multilayer perceptron, uma rede neural semelhante à perceptron, mas com mais de uma camada de neurônios em alimentação direta (NORVIG; RUSSEL, 2004). Essa rede é composta por camadas de neurônios ligadas entre si por sinapses com pesos. O aprendizado nesse tipo de rede é geralmente feito através do algoritmo de retro-propagação do erro (HAYKIN, 2001). Existem outros algoritmos para este fim, como a Rprop, aprendizado não-supervisionado, fundamentos de processamento de imagens, representação de imagens, filtragem e convolução e conhecimento prévio de álgebra linear, cálculo, probabilidade e otimização (RIEDMILLER; BRAUN, 1993).

De acordo com Goodfellow et al. (2016), os métodos de Deep Learning têm por objetivo descobrir um modelo a partir de um conjunto de dados e método que servirão como guia para o aprendizado do modelo a partir desses exemplos. O processo de aprendizado ao final, terá uma função adequada para receber como entrada os dados em sua forma bruta e fornecerá como saída uma representação adequada para o problema em questão.

Segundo Ponti e Costa (2017) para entender como Deep Learning se diferencia de Machine Learning apresentaremos os exemplos abaixo:

- **classificação de imagens:** recebe como entrada uma imagem no formato RGB e como saída produzirá as probabilidades se essa imagem pertence a um conjunto possível de classe onde uma função será capaz de diferenciar imagens que contém cães, gatos, tartarugas e corujas. Dessa forma pretende-se aprender uma função como $f(x) = y$, em que x representa a imagem, e y a classe mais provável para x .
- **detecção de anomalias em sinais de fala:** recebe como entrada um sinal de áudio e como saída produzirá a representação desse sinal capaz de apresentar se a entrada representa um caso normal ou se existe alguma anomalia na fala de uma pessoa. A função aprendida deverá estar no formato $f(x) = p$, em que p representa a probabilidade de x ser um áudio anômalo (PONTI; COSTA, 2017).

Ponti e Costa (2017) afirmam que em ambos os exemplos as definições são análogas aos de aprendizado de máquina: o primeiro exemplo é um cenário de aprendizado supervisionado, enquanto o segundo exemplo se trata de um cenário não supervisionado.

Quando se trata de Deep Learning a diferença está em como se aprende a função $f(\cdot)$. Métodos superficiais ou shallow em inglês buscam diretamente por uma única função que possa gerar o resultado desejado a partir de um conjunto de parâmetros. Deep Learning por outro lado tem métodos que aprendem $f(\cdot)$ por meio da composição de funções, como por exemplo: $f(x) = f_L(\dots f_2(f_1(x_1)) \dots)$ (onde o índice l se refere a uma camada) onde cada função $f_l(\cdot)$ toma como entrada um vetor de dados x_l , gerando como saída o próximo vetor x_{l+1} (PONTI; COSTA, 2017).

Ainda de acordo com Ponti e Costa (2017), cada função usa um conjunto de parâmetros para transformar dos dados de entrada. Denota-se o conjunto desses parâmetros em W_l , relacionado a cada função f_l , e então podemos escrever por exemplo: $f_L(\dots f_2(f_1(x_1, W_1); W_2) \dots), W_L$, onde x_1 representa os dados de entrada, cada função tem seu próprio conjunto de parâmetros e sua saída será passada para a próxima função. Na equação acima, temos a composição de L funções, ou L camadas.

Um dos principais objetivos de Deep Learning é aprender representações sucessivas dos dados, intermediárias, ou seja, os $x_l, l = 1 \dots L$ acima. Os algoritmos de Deep Learning resolvem o problema de encontrar os parâmetros W diretamente a partir dos dados e definem cada representação como combinações de representações anteriores e mais simples. Assim a profundidade permite aprender uma sequência de funções que transformam vetores mapeando-os de um espaço a outro, até atingir o resultado desejado. Por isso é de fundamental importância a hierarquia das representações: cada função opera sobre uma entrada gerando uma representação que é então passada para a próxima função (PONTI; COSTA, 2017).

Em Deep Learning temos a hipótese de que se tivermos um número suficiente de camadas L espaços com dimensionalidade alta o suficiente como o número de parâmetros W em cada função e dados suficientes para aprender os parâmetros W_l para todo l , então será possível obter o escopo das

relações nos dados originais, encontrando assim a representação mais adequada para a tarefa desejada. Matematicamente, podemos interpretar que essa sequência de transformações separa as múltiplas variedades que nos dados originais estariam todas enoveladas (CHOLLET, 2017).

A capacidade de aprender com as experiências é extremamente importante para a evolução das espécies, assim a implementação de técnicas de aprendizado automático tem se tornado uma das áreas de grande interesse em Inteligência Artificial. Embora o processo seja usado pelas pessoas, seu funcionamento não é conhecido o suficiente para auxiliar a sua implementação em uma máquina. Usando o conceito de agentes, o aprendizado pode ser resumido como a alteração do programa de controle, baseando-se em suas próprias experiências. Por causa da sua capacidade de extrair conhecimento, técnicas de aprendizado de máquina têm sido intensamente usadas para extrair informações, de modo automático, a partir de diversas bases de dados (ARTERO, 2009).

De acordo com Artero (2009), o aprendizado de máquina pode ser classificado de diferentes maneiras, embora o mais comum é classificá-lo em supervisionado e não supervisionado. No primeiro caso, deve ser apresentado um conjunto de pares entradas e saídas, sendo as entradas formadas por alguns atributos relativos aos objetos que se deseja reconhecer, ou seja, um padrão ou tendência e as saídas correspondem às classes desejadas. Por outro lado, no aprendizado não supervisionado, apenas as entradas são fornecidas e, neste caso, o padrão de saída deve ser deduzido pelo próprio sistema de aprendizado. Alguns exemplos de técnicas de aprendizado de máquina nestas duas categorias são:

- Aprendizado supervisionado: redes neurais artificiais, com treinamento supervisionado; algoritmos genéticos e as árvores de decisão. O aprendizado supervisionado nas redes neurais artificiais ocorre com o ajuste dos pesos, a partir das diferenças entre os valores de saída obtidos pela rede e o valor desejado, informado no conjunto de treinamento. No caso dos algoritmos genéticos, o treinamento é supervisionado pela função de aptidão adotada, que vai priorizando algumas características nas novas gerações.
- Aprendizado não supervisionado: redes neurais com treinamento não supervisionado, algoritmos de agrupamento, mineração de dados com regras de associação. A grande vantagem das técnicas desta categoria é que não é necessário conhecer a classe de saída de cada item de dado e, alguns casos, nem a quantidade de classes de saída (ARTERO, 2009).

4.2 Machine Learning vs Deep Learning

De acordo com Santana (2018), Deep Learning é um subgrupo específico de técnicas de Machine Learning, que utilizam redes neurais profundas e dependem de muitos dados para o treinamento. Muitos fatores diferenciam Deep Learning das técnicas clássicas de Machine Learning e alguns deles favorecem a utilização dessas técnicas em áreas como visão computacional e processamento de linguagem natural.

Outra diferença é que as técnicas clássicas chegam no limite máximo que conseguem extrair de informação dos dados. Enquanto com Deep Learning isso não acontece, pois são preparadas para trabalhar com maior quantidade de dados. Esses dois problemas dos algoritmos clássicos de Machine

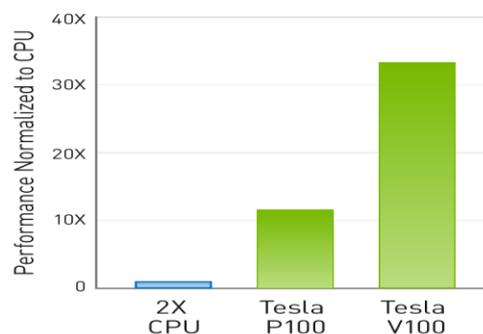
Learning reforçam um dos principais benefícios do uso de Deep Learning, quanto mais dados, melhor (SANTANA, 2018).

Deep Learning tem arquitetura complexa e necessita de muitos dados para seu treinamento. Dessa forma, é fatal a dependência de muito poder computacional para aplicar essas técnicas. Embora alguns métodos clássicos utilizem muito poder computacional como memória ou CPU, as técnicas de Deep Learning estão em uma escala superior. As pesquisas relacionadas a computação paralela e utilização de GPU's com CUDA, contribuíram significativamente com Deep Learning, pelo simples fato de ser algo inviável utilizando apenas CPU (SANTANA, 2018).

A Ilustração 1 apresenta uma comparação do treinamento de uma rede neural profunda utilizando apenas CPU e utilizando duas outras placas da NVidia. Ao rodar na Tesla V100 o treinamento da rede fica 30x mais rápido em comparação com o CPU, se levarmos isso a utilização real, em que uma rede dessas pode treinar por semanas, utilizando apenas CPU seria algo impossível de se obter resultados. Outras arquiteturas de hardware com GPUs trazem uma melhoria de 100X em relação a CPU (NVidia DGX-1). Pois, acompanha a complexidade das redes profundas e dos problemas solucionados. Uma rede complexa pode demorar semanas treinando em um hardware de ponta, o que levaria anos em uma CPU (SANTANA, 2018).

Ilustração 1 - Comparação do treinamento de uma rede neural profunda utilizando apenas CPU e utilizando duas outras placas da NVidia

30x Higher Throughput than CPU Server on Deep Learning Inference



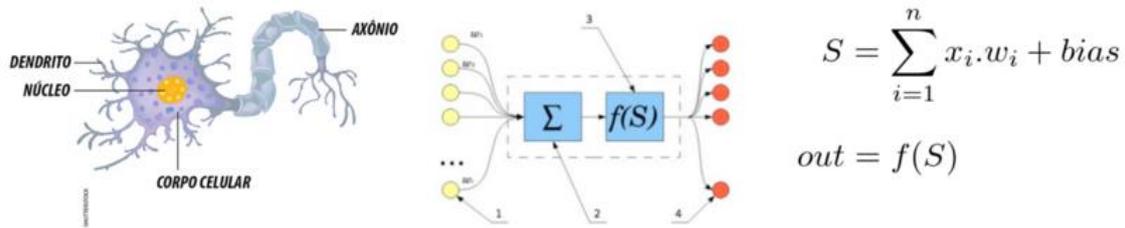
Workload: ResNet-50 | CPU: 2X Xeon E5-2660 v4, 2GHz | GPU: add 1X Tesla P100 or V100 at 150W | V100 measured on pre-production hardware.

Fonte: Santana (2018)

De acordo com Santana (2018), embora existam técnicas clássicas que são de propósito geral, a estrutura de Deep Learning e sua unidade mais básica, o neurônio, consegue ser o mais genérico e flexível possível. Resumidamente, o neurônio é composto por entradas (os dendritos), um núcleo de processamento (o núcleo) e saídas (o axônio). Dessa forma um sinal entra, é processado e sai diferente. Matematicamente, o neurônio pode ser modelado como um somatório das entradas multiplicado por pesos, em que esse valor passa por uma função de ativação. Essa proposta foi feita por McCulloch em 1943 e posteriormente refinada por Rosenblatt em 1950. A formulação matemática de um neurônio é o

mais simples possível, e fazendo um paralelo, temos a exemplificação gráfica da Ilustração 2 (SANTANA, 2018).

Ilustração 2 - Simulação matemática de um neurônio



Fonte: McCulloch; Pitts (1943).

Para Souza (2018), embora apenas um neurônio não seja útil, a flexibilidade de interligar vários neurônios em estruturas mais complexas para resolver problemas é o grande diferencial das arquiteturas de Deep Learning. Embora Machine Learning tenha se tornado parte integrante do processamento de dados. Uma das principais diferenças em relação ao Deep Learning é que exige uma intervenção manual na seleção dos recursos a serem processados, enquanto Deep Learning é intuitivo.

A extração de recursos na aprendizagem de máquina requer o trabalho do cientista de dados para pré-processar os dados e entregar aos algoritmos dados que possam ser explorados em busca de padrões. Já Deep Learning é um paradigma para a realização de Machine Learning. Assim, torna-se um dos temas mais quentes da atualidade devido aos resultados inigualáveis em aplicações como visão computacional, reconhecimento de fala e compreensão de linguagem natural.

4.3 Rede Neural Artificial

Em ciência da computação e campos relacionados, redes neurais artificiais são modelos computacionais inspirados pelo sistema nervoso central de um animal que são capazes de realizar o aprendizado de máquina bem como o reconhecimento de padrões.

Em sua forma mais geral, uma rede neural é um sistema projetado para modelar a maneira como o cérebro realiza uma tarefa particular, sendo normalmente implementada utilizando-se componentes eletrônicos ou é simulada por propagação em um computador digital. Para alcançarem bom desempenho, as redes neurais empregam uma interligação maciça de células computacionais simples, denominadas de “neurônios” ou unidades de processamento (HAYKIN, 2001).

O surgimento das RNAs deu-se com o modelo matemático do neurônio biológico proposto por Warren McCulloch e Walter Pitts em 1943 (McCULLOCH e PITTS, 1943).

O modelo, denominado neurônio MCP (McCulloch-Pitts), é descrito por um conjunto de n entradas, as quais são multiplicadas por um determinado peso e, em seguida, os resultados são somados e comparados a um limiar (NIED, 2007).

Em 1958, Frank Rosenblatt propôs uma topologia de rede denominada de percéptron constituída por neurônios MLP (Percéptrons de Múltiplas Camadas) e arranjada em forma de rede composta de duas camadas (ROSENBLATT, 1958), a qual possibilitou um aumento de trabalhos relacionados a redes neurais até 1969. Neste mesmo ano, a publicação de Minsky e Papert (MINSKY; PAPERT, 1969) mostrou deficiências e limitações do modelo MLP, provocando um desinteresse pelos estudos relacionados às RNAs. Somente a partir de 1982, com a publicação do trabalho de Hopfield (HOPFIELD, 1982), foi novamente despertado o interesse pelos estudos relacionados às redes neurais.

As RNAs são comumente utilizadas na resolução de problemas complexos, onde o comportamento das variáveis não é rigorosamente conhecido. Uma de suas principais características é a capacidade de aprender por meio de exemplos e de generalizar a informação aprendida, gerando um modelo não-linear, tornando sua aplicação na análise espacial bastante eficiente (SPÖRL et al., 2011).

Em termos de topologia, para implementar uma RNA deve-se definir diferentes variáveis, dentre as quais: a) o número de nós na camada de entrada (tal variável corresponde ao número de variáveis que serão usadas para alimentar a rede neural, sendo normalmente as variáveis de maior importância para o problema em estudo), b) o número de camadas escondidas e o número de neurônios a serem colocados nessas camadas e, c) o número de neurônios na camada de saída (SANTOS et al., 2005) Uma das maiores dificuldades encontradas no uso das redes neurais é a escolha da melhor arquitetura, uma vez que esse processo é experimental e demanda um grande tempo de execução. Na prática, o processo deve ser aplicado com o intuito de testar vários métodos de aprendizado e as diferentes configurações que uma rede possa ter para a resolução de um problema em específico (MIRANDA et al., 2009). Em geral, podemos identificar três classes de arquiteturas de rede fundamentalmente diferentes: redes alimentadas adiante com camada única, redes alimentadas diretamente com múltiplas camadas e redes recorrentes.

As RNAs são algoritmos computacionais que apresentam um modelo matemático inspirado na estrutura de organismos inteligentes, os quais possibilitam inserir simplificada o funcionamento do cérebro humano em computadores. Dessa forma, a exemplo do cérebro humano, a RNA é capaz de aprender e tomar decisões baseadas em seu próprio aprendizado. A RNA corresponde a um esquema de processamento capaz de armazenar conhecimento baseado em aprendizagem e disponibilizar este conhecimento para a aplicação à qual se destina (SPÖRL et al., 2011).

De acordo com Haykin (2001), a rede neural se assemelha ao cérebro humano em dois aspectos básicos: a) o conhecimento é adquirido pela rede a partir de seu ambiente, por intermédio do processo de aprendizagem; b) forças de conexão entre neurônios (pesos sinápticos) são utilizadas para armazenar o conhecimento adquirido.

As redes neurais são compostas por uma determinada quantidade de entradas e unidades de processamento, as quais são ligadas através de pesos sinápticos. As entradas são propagadas através da topologia da RNA, sendo transformadas pelos pesos sinápticos e pela função de ativação (AF) dos neurônios (MACHADO E FONSECA JÚNIOR, 2013).

Recebendo entradas de n neurônios (y_i), o neurônio k calcula a sua saída através de:

$$y_k = AF \left(\sum_{i=1}^n (y_i w_{ki}) + b_k \right)$$

Onde y_i é a saída calculada pelo neurônio i , w_{ki} representa o peso sináptico entre o neurônio i e o neurônio k e b_k é o peso entre um valor constante e diferente de zero ao neurônio k , conhecido como bias. Se o neurônio estiver ligado às entradas, o termo y_i é substituído pela entrada correspondente.

É necessário determinar os pesos sinápticos e bias para se utilizar uma RNA. A estimação de tais parâmetros, conhecida como treinamento, se constitui em um processo iterativo onde os parâmetros iniciais são utilizados até a convergência do processo.

Considerando a interação j , o peso w_{ki} é utilizado através de:

$$j = w(j-i) + \Delta w(j) \quad (2)$$

Com $i \Delta w(j)$ sendo o vetor de correção ao parâmetro w_{ki} na iteração j . Devido às características não-lineares inerentes ao mapeamento entre camadas de RNAs feedforward, estas se caracterizam como ferramentas de modelamento bastante apropriadas para o modelamento e controle de sistemas. Ferramentas matemáticas disponíveis para o controle de sistemas não-lineares normalmente utilizam técnicas de linearização, transformando a tarefa de controle não-linear em pequenas tarefas de controle linear. Apesar de muitas vezes eficiente, esta abordagem não retrata a realidade dos sistemas físicos, podendo resultar em soluções subótimas. Devido à capacidade de modelar com fidelidade ações não lineares, as RNAs se constituem uma importante ferramenta para o controle de sistemas (BRAGA, 2000).

De acordo com Braga et al. (2007) um modelo básico de RNA possui diferentes componentes, dentre os quais:

- Conjunto de sinapses: conexões entre os neurônios da RNA. Cada uma delas possui um peso sináptico;
- Integrador: realiza as somas dos sinais de entrada da RNA, ponderados pelos pesos sinápticos;
- Função de ativação: restringe a amplitude do valor de saída de um neurônio;
- Bias: valor aplicado externamente a cada neurônio e tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação.

4.3.1 Rede Neural Convolutiva

Redes neurais convolucionais CNNs são provavelmente o modelo de rede Deep Learning mais conhecido e utilizado atualmente. O que caracteriza esse tipo de rede é ser composta basicamente de camadas convolucionais, que processa as entradas considerando campos receptivos locais. Adicionalmente inclui operações conhecidas como pooling, responsáveis por reduzir a dimensionalidade

espacial das representações. Atualmente as CNNs de maior destaque incluem as redes residuais - ResNet e Inception (PONTI; COSTA, 2017).

A principal aplicação das CNNs é para o processamento de informações visuais, em particular imagens, pois a convolução permite filtrar as imagens considerando sua estrutura bidimensional. Considere o exemplo da base de dados MNIST apresentado na seção anterior. Note que ao vetorizar as imagens estamos desprezando toda a estrutura espacial que permite entender a relação entre os pixels vizinhos em uma determinada imagem. É esse tipo de estrutura que as CNNs tentam capturar por meio da camada convolucional.

Na camada convolucional cada neurônio é um filtro aplicado a uma imagem de entrada e cada filtro é uma matriz de pesos.

Segundo Ponti e Costa (2017), dado uma imagem RGB de tamanho $224 \times 224 \times 3$, o 3 indica os canais de cor RGB, que serve de entrada para uma camada convolucional. Cada neurônio dessa camada irá processar a imagem e produzir uma transformação dessa imagem por meio de uma combinação linear dos pixels vizinhos e tem-se um peso para cada elemento da imagem. Como exemplo uma camada fully connected *FC* obtém-se para cada neurônio, 150528 pesos, um para cada valor de entrada. Porém, foi definido filtros de tamanho $k \times k \times d$, em que k é a dimensão espacial do filtro a ser definida e d a dimensão de profundidade que depende da entrada da camada.

Se definirmos $k = 5$ para a primeira camada convolucional, obtém-se filtros $5 \times 5 \times 3$, pois como a imagem possui 3 canais, RGB, então $d = 3$, e assim cada neurônio terá $5 \times 5 \times 3 = 75$ pesos.

Cada região da imagem processada pelo filtro é chamada de campo receptivo local; um valor de saída é uma combinação dos pixels de entrada nesse campo receptivo local. No entanto todos os campos receptivos são filtrados com os mesmos pesos locais para todo pixel. Isso é o que torna a camada convolucional diferente da camada *FC*. Assim, um valor de saída ainda terá o formato bidimensional. Por exemplo podemos dizer que $f_{l+1}(i, x, y)$ é o pixel resultante da filtragem da imagem vinda da camada anterior l , processada pelo filtro i a partir dos valores da vizinhança centrados na posição x, y . No nosso exemplo com $k = 5$, teremos uma combinação linear de 25 pixels da vizinhança para gerar um único pixel de saída (PONTI; COSTA, 2017).

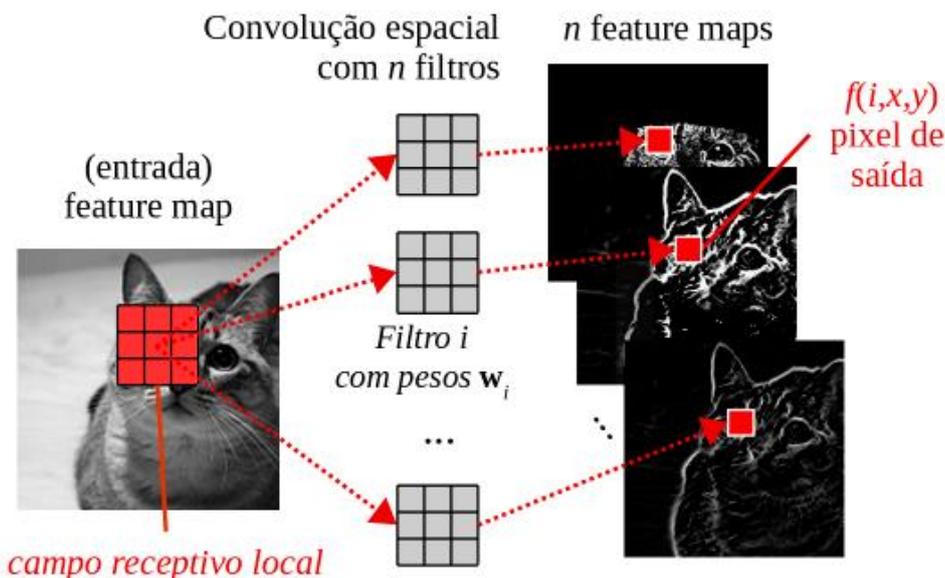
Ainda de acordo com Ponti e Costa (2017), os tamanhos de filtros mais utilizados são $5 \times 5 \times d$, $3 \times 3 \times d$ e $1 \times 1 \times d$. Como trabalhamos com matrizes multidimensionais com profundidade d , utilizamos o termo tensor para denotá-las.

Os autores afirmam que, considerando um problema em se tem como entrada imagens RGB, de tamanho $64 \times 64 \times 3$. Sejam então duas camadas convolucionais, a primeira com 4 filtros de tamanho $k_1 = 5$, e a segunda com 5 filtros de tamanho $k_2 = 3$. Considere ainda que a convolução é feita utilizando extensão da imagem com preenchimento por zero padding de forma que conseguimos realizar a filtragem para todos os pixels da imagem, mantendo seu tamanho. Nesse cenário teríamos a seguinte composição: $\hat{y} = f(x) = f_2(f_1(x; W_1; b_1); W_2; b_2)$, em que W_1 possui dimensão $4 \times 5 \times 5 \times 3$ sendo 4 filtros de tamanho 5×5 , entrada com profundidade 3, e portanto a saída da camada 1, $x_2 =$

$f1(x1)$ terá tamanho: $64 \times 64 \times 4$. Após a convolução utiliza-se uma função de ativação que trunca para zero os pixels negativos.

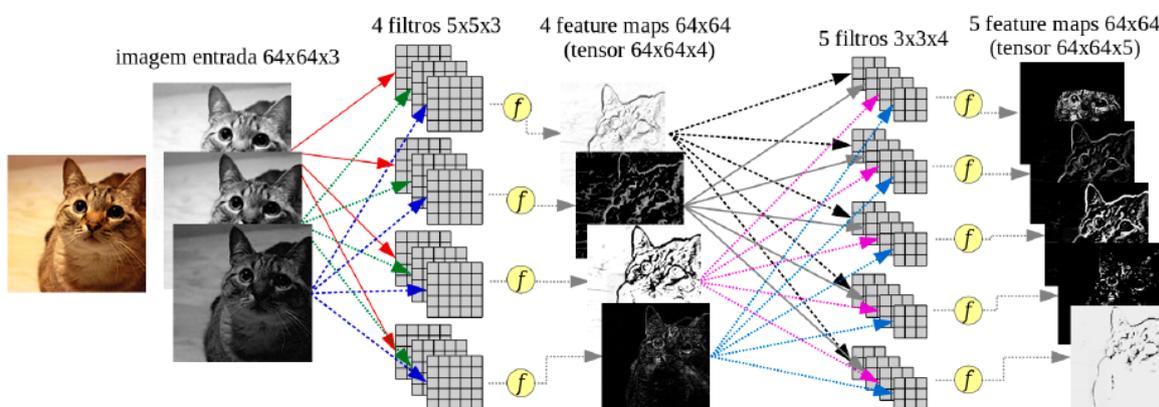
As Ilustrações 3 e 4 ilustram esse processo, bem como o da camada 2, que recebe por entrada o tensor $64 \times 64 \times 4$. Essa segunda camada possui 5 filtros $3 \times 3 \times 4$ porque a profundidade do tensor de entrada tem $d = 4$, e gera como saída $x3 = f2(x2)$, um tensor de tamanho $64 \times 64 \times 5$ (PONTI; COSTA, 2017).

Ilustração 3 - Processamento de informações locais utilizando cada posição (x, y) como centro.



Fonte: Ponti e Costa (2017)

Ilustração 4 - Ilustração de duas camadas convolucionais.



Fonte: Ponti e Costa (2017)

Chamamos de mapa de características a saída de cada neurônio da camada convolucional. Outro aspecto importante para se citar é o passo ou stride. A convolução convencionais é feita com passo/stride 1, ou seja, filtramos todos os pixels e portanto para uma imagem de entrada de tamanho 64×64 , geramos

uma nova imagem de tamanho 64×64 . O uso de strides maiores que 1 é comum quando deseja-se reduzir o tempo de execução, pulando pixels e assim gerando imagens menores como o $stride = 2$, teremos como saída uma imagem de tamanho 32×32 (PONTI; COSTA, 2017).

5 RESULTADOS E DISCUSSÃO

Deep Learning contribuiu com resultados estatisticamente notáveis com a utilização de múltiplas camadas. Mesmo assim, suas limitações no uso de redes neurais profundas devem ser consideradas, visto que essas são necessariamente uma forma de aprender as transformações a serem aplicadas aos dados de entrada dadas por um imenso conjunto de parâmetros atualizados durante a etapa de treinamento.

Chollet (2017), afirma que a primeira limitação para que seja possível realizar o treinamento está relacionado ao mapa entre a entrada e saída da rede que deve ser contínuo e tênue. A segunda limitação diz respeito à abstração e adaptação visto que, redes neurais profundas precisam de quantidades massivas de dados rotulados para aprender conceitos simples; em contrapartida, seres humanos são capazes de aprender um conceito a partir de pequenas quantidades de exemplos.

Deep Learning tem uma valiosa contribuição para as múltiplas áreas em que o aprendizado de máquina é empregado na busca de soluções de problemas. Ainda com suas limitações, é importante continuar pesquisando métodos para obtermos avanços importantes com o uso dessas aplicações (NAZARE et al., 2017).

Há muitos estudos apontando incertezas sobre a capacidade de generalização desses modelos, em particular: as perturbações nos dados de entrada, tais como ruído que podem comprometer de forma significativa os resultados, o uso de imagens invertidas na fase de teste ocasionando erros nos resultados, possibilidade de criar imagens que visualmente parecem conter somente ruído aleatório, mas que fazem com que as redes neurais convolucionais as classifiquem com 99% de certeza, desenvolvimento de perturbações imperceptíveis a humanos que fazem as redes produzirem saídas completamente erradas (NGUYEN et al., 2015).

6 CONSIDERAÇÕES FINAIS

Uma estratégia de explorar ainda mais as técnicas de aprendizagem profunda é abordar a eficiência computacional do processo de treinamento. Mesmo que a aplicação de técnicas de paralelismo para agilizar o processamento, tais como o uso de unidades gráficas de processamento, não seja uma estratégia recente, o crescente tamanho dos conjuntos de treinamento e a grande quantidade de parâmetros envolvidos no treinamento de arquiteturas profundas trouxeram a necessidade do uso desse recurso. Nota-se como forte tendência o uso de bibliotecas especializadas que realizem, com eficiência e menor custo computacional, operações comuns em aprendizagem profunda e que tirem proveito de hardware para paralelismo.

Um desafio ainda pertinente na pesquisa em aprendizagem profunda, assim como em aprendizagem de máquina, é projetar algoritmos que possam treinar sistemas com poucos dados e, principalmente, aproveitar de forma eficiente o uso de dados não supervisionados.

A técnica de pré-treinamento não supervisionado define um caso de uso de dados não supervisionados. Essa mesma técnica permitiu que a pesquisa em aprendizagem profunda crescesse ainda mais a partir de meados de 2006 (HINTON et al., 2012).

Mesmo com esse avanço, os problemas têm se concentrado em trabalhos com dados supervisionados, que apresentaram maior disponibilidade desde esse período.

E ainda, as abordagens supervisionadas alcançam bom desempenho na presença de uma grande quantidade de dados dessa natureza.

7 REFERÊNCIAS

- ARTERO, A. O. Inteligência Artificial: teórica e prática. **Editora Livraria da Física**, 1.ed., 2009.
- BENGIO, Y. Learning Deep Architectures for AI. **Foundations and Trends in Machine Learning**, 2(1):1–127, 2009.
- BRAGA, A. de.; et al. Redes Neurais Artificiais teoria e aplicações. **Rio de Janeiro: LTC – Livros Técnicos e Científicos Editora S.A.**, 2000.
- BRAGA, A. P.; et al. Redes Neurais Artificiais: teorias e aplicações. **Livros Técnicos e Científicos, Rio de Janeiro**. 2007.
- CHOLLET, F. Deep Learning with Python. **Manning**, 2017.
- CHUNG, J., et al. Gated feedback recurrent neural networks. **In Proceedings of the 32nd International Conference on International Conference on Machine Learning**, volume 37 of ICML'15, pages 2067-2075, 2015.
- DENG, J. et al. ImageNet: A Large Scale Hierarchical Image Database. **In CVPR09**, 2009.
- FUKUSHIMA, K. Neocognitron: A hierarchical neural network capable of visual pattern recognition. **Neural networks**, 1(2):119-130, 1988.
- GOODFELLOW, I. et al. Deep Learning. **MIT Press**, 2016. Disponível em <<http://www.deeplearningbook.org>>. Acesso em 08 de out de 2018.
- HAYKIN, S. Redes Neurais. Princípios e prática. 2 ed., **Bookman**. 2001.
- HINTON, G. et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. **IEEE Signal Processing Magazine**, 29(6):82-97, 2012.
- HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. **Proceedings of the National Academy of Science of the USA**. v.79, n.8, p.2554-2558, 1982.
- HOSSEIN, H., RADHA, P. Deep neural networks do not recognize negative images. **arXiv preprint arXiv:1703.06857**, 2017.
- KRIZHEVSKY, A., et al. Imagenet classification with deep convolutional neural networks. **In Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems.**, pages 1106-1114, 2012.
- LECUN, Y.; BENGIO, Y. Convolutional Networks for Images, Speech, and Time-Series. **The Handbook of Brain Theory and Neural Networks**. **Cambridge: MIT Press**, 1995.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, 86(11):2278-2324, 1998.
- LECUN, Y. et al. Convolutional networks and applications in vision. **In Proceedings of 2010 IEEE International Symposium on Circuits and Systems**, pages 253-256, 2010.
- LECUN, Y. et al. Deep learning. **Nature**, 521(7553):436-444. 2015.

- LIU, C. et al. A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure. **IEEE Transactions on Services Computing**, p. (99):1-13, 2017.
- MACHADO, W, C.; FONSECA JÚNIOR, E. S. Redes Neurais Artificiais aplicadas na previsão do VTEC no Brasil. **Boletim de Ciências Geodésicas**, v.19, n.2, p. 227-246, 2013.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, 5(4):115–133, 1943.
- MINSKY, M. L.; PAPERT, S. A. Perceptrons. **MIT Press, Cambridge**, MA. 1969.
- MIRANDA, F. A. et al. Integração e interpolação de dados de anomalias ar livre utilizando a técnica de RNA e krigagem. **Boletim de Ciências Geodésicas**, v.15, n.3, p. 428-443, 2009.
- NAZARE, T. et al. Deep convolutional neural networks and noisy images. In **Iberoamerican Conference on Pattern Recognition (CIARP)**, 2017.
- NGUYEN, A. et al. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In **IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, pages 427–436, 2015.
- NORVIG, P; RUSSEL, S. Inteligência Artificial. 2. ed., **Elsevier**, 2004.
- PANG, Y., et al. Convolution in convolution for network in network. **IEEE Transactions on Neural Networks and Learning Systems**, p. (99):1-11, 2017.
- PONTI, M. A.; COSTA, G. B. P. Como funciona o Deep Learning. in: **Tópicos em Gerenciamento de Dados e Informações, SBC**, 1.ed., 2017.
- RIEDMILLER, M.; BRAUN, H. A direct adaptative method for faster backpropagation learning: the RPROP algorithm. **IEEE International Conference on Neural Networks**, 1993.
- ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, v.65, n.6, p. 386-408, 1958.
- RUSSAKOVSKY, O. et al. ImageNet Large Scale Visual Recognition Challenge. **International Journal of Computer Vision (IJCV)**, 115(3):211-252, 2015.
- SANTANA, M. Deep Learning: do conceito às aplicações. **Medium.com**. Disponível em <<https://medium.com/data-hackers/deep-learning-do-conceito-%C3%A0s-aplica%C3%A7%C3%B5es-e8e91a7c7eaf>>. Acesso em 08 de out de 2018.
- SANTOS, A. M.; et al. Usando Redes Neurais Artificiais e Regressão Logística na predição da Hepatite A. **Revista Brasileira de Epidemiologia**, v.8, n.2, p. 117-126, 2005.
- SOUZA, J. Quais as diferenças entre Machine Learning e Deep Learning? Disponível em: <<https://www.lecom.com.br/blog/machine-learning-e-deep-learning/>>. Acesso em 08 de out de 2018.
- SPÖRL, C. et al. Aplicação de Redes Neurais Artificiais na construção de modelos de fragilidade ambiental. **Revista do Departamento de Geografia**, v. 21, n.1, p. 113-135, 2011.
- YANG, S. et al. From facial parts responses to face detection: A deep learning approach. In **Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV)**, pages 3676-3684, 2015.