

Aprendizagem com pequenas amostras na era do Big Data: uma revisão de literatura sobre perspectivas e desafios

João Carlos P. Alves^{1†}, Eric B. Ferreira², Iago A. Carvalho³

¹*Programa de Pós-graduação em Estatística Aplicada e Biometria (PPGEAB), Universidade Federal de Alfenas (UNIFAL-MG).*

²*Departamento de Estatística, Universidade Federal de Alfenas (UNIFAL-MG).*

³*Departamento de Ciência da Computação, Universidade Federal de Alfenas (UNIFAL-MG).*

Resumo: *O avanço do Big Data trouxe um grande volume de dados que possibilitou o uso de técnicas de aprendizagem de máquina para a tomada de decisões em diferentes áreas. Contudo, a efetividade desses modelos depende da disponibilidade de grandes quantidades de dados, levantando o desafio de como lidar com a aprendizagem com pequenas amostras. Aprender a partir de pequenas amostras é importante em muitas aplicações quando grandes volumes não estão disponíveis, como no caso de pesquisas de opinião, devido aos desafios da coleta de informações, principalmente devido à falta de engajamento e participação em questionários. Esta abordagem pode facilitar ou minimizar o uso de questionários de opinião. No entanto, existem desafios que precisam ser superados para obter um desempenho preciso em tarefas de aprendizagem com pequenas amostras, como a seleção de amostras relevantes, a escolha de métodos de treinamento adequados, entre outros. Diante disso, discutiremos as perspectivas e desafios da aprendizagem com pequenas amostras na era do Big Data. Será realizada uma revisão das técnicas de Few-shot learning e suas aplicações como uma alternativa para lidar com pequenas amostras. A revisão dessas técnicas e sua aplicação em conjuntos de dados limitados podem fornecer insights valiosos para a melhoria de modelos de aprendizagem de máquina em diferentes áreas de aplicação.*

Palavras-chave: *Aprendizagem de máquina; Few-shot learning; Big Data; pequenas amostras; Questionários de opinião.*

Few-shot learning in the era of Big Data: a literature review on perspectives and challenges

Abstract: *The advancement of Big Data has brought a large volume of data that has enabled the use of machine learning techniques for decision-making in different fields. However, the effectiveness of these models depends on the availability of large amounts of data, raising the challenge of dealing with learning from few samples. Learning from few samples is important in many applications when large volumes are not available, such as opinion surveys, due to the challenges of data collection, primarily due to lack of engagement and participation in questionnaires. This approach can facilitate or minimize the use of opinion surveys. However, there are challenges that need to be overcome to achieve accurate performance in few-shot learning tasks, such as the selection of relevant samples, and the choice of appropriate training methods, among others. In light of this, we will discuss the perspectives and challenges of learning from few samples in the era of Big Data. A review of Few-shot learning techniques and their applications will be conducted as an alternative to deal with few samples. Reviewing these techniques and their application on limited datasets can provide valuable insights for improving machine learning models in different application domains.*

Keywords: *Machine Learning; Few-shot learning; Big Data; Few samples; Opinion surveys.*

Introdução

Nos últimos anos, o avanço do *Big Data* tem proporcionado um crescimento exponencial no volume de dados disponíveis, impulsionando o uso de técnicas de aprendizagem de máquina para a tomada de decisões em uma ampla variedade de áreas. Esses avanços têm demonstrado o potencial da análise de dados em larga escala para obter *insights* valiosos e melhorar a precisão dos modelos de aprendizado. No entanto, é importante ressaltar que a efetividade desses modelos depende da disponibilidade de grandes quantidades de dados para treinamento.

† Autor correspondente: joao.carlos@sou.unifal-mg.edu.br.

Essa dependência de dados em grande escala pode se tornar um desafio significativo em situações em que a quantidade de dados disponíveis é limitada. Em muitas aplicações do mundo real, é comum encontrar cenários em que a coleta de um grande número de amostras é impraticável, custosa ou até mesmo inviável. Essas limitações são especialmente relevantes em pesquisas de opinião, estudo de doenças raras, desenvolvimento de novos dispositivos, por exemplo, os dados disponíveis são escassos.

Nesses contextos, a aprendizagem com pequenas amostras desempenha um papel crucial. Trata-se de uma abordagem que visa aprender com um número limitado de exemplos, permitindo a tomada de decisões mesmo quando grandes volumes de dados não estão prontamente disponíveis. A ideia fundamental por trás da aprendizagem com pequenas amostras é aproveitar ao máximo as informações limitadas disponíveis para generalizar para novos exemplos e realizar inferências precisas.

Uma técnica popular para lidar com a aprendizagem com pequenas amostras é o *Few-shot learning* (FSL) (WANG et al., 2020), que se concentra em treinar modelos de aprendizado capazes de generalizar para novas tarefas ou classes com base em um número limitado de exemplos de treinamento. Essa abordagem tem sido explorada como uma alternativa promissora para superar os desafios da escassez de dados.

O conceito de FSL foi primeiramente registrado nos anos 2000 (LU et al., 2020). Não obstante, essa abordagem começou a ganhar destaque e projeção mais significativa com o advento das Redes Neurais Convolucionais (CNN) a partir de 2015. Esse desenvolvimento pode ser atribuído, em parte, ao aprimoramento e à popularização de técnicas de treinamento de modelos, à disponibilidade de grandes conjuntos de dados, e ao aumento substancial no poder computacional, possibilitando a implementação de modelos mais complexos e robustos. Ademais, o interesse crescente nesta abordagem é justificado pela sua capacidade de desenvolver modelos de aprendizagem de máquina que podem realizar inferências precisas, mesmo quando o volume de dados de treinamento é restrito, conferindo assim um potencial considerável para uma variedade de aplicações práticas e cenários onde a coleta de dados é desafiadora.

O FSL se destaca por sua capacidade de aprender com eficiência a partir de um número reduzido de amostras de treinamento, permitindo a adaptação rápida a novos contextos e a tomada de decisões mesmo diante da escassez de dados. Ao invés de depender exclusivamente de enormes conjuntos de dados rotulados, o FSL concentra-se em extrair informações úteis e relevantes dos poucos exemplos disponíveis.

Uma característica chave do FSL é a sua habilidade de aproveitar o conhecimento prévio adquirido em tarefas ou domínios relacionados. Por meio de técnicas como a transferência de conhecimento, os modelos de aprendizado são capazes de generalizar a partir de experiências anteriores e aplicar esse conhecimento em novos contextos com pequenas amostras disponíveis. Essa capacidade de transferir aprendizados anteriores permite que o modelo obtenha um desempenho promissor, mesmo quando os dados de treinamento são limitados.

Além disso, a escolha de métodos de treinamento apropriados também é um desafio importante na aprendizagem com pequenas amostras. Diferentes algoritmos e estratégias podem ser empregados para lidar com a escassez de dados, como técnicas de transferência de conhecimento, modelagem de distribuições latentes ou uso de redes neurais generativas. Cada abordagem tem suas vantagens e desvantagens, e a escolha adequada depende do contexto específico do problema e dos recursos disponíveis.

Um desafio crucial é a capacidade de generalização para novas amostras com base em um conjunto reduzido de exemplos. A capacidade de extrapolar informações a partir de um número limitado de amostras é essencial para garantir a aplicabilidade dos modelos de aprendizado em cenários do mundo real. Estratégias como regularização, aumento de dados e técnicas de adaptação de domínio podem ajudar a melhorar a capacidade de generalização dos modelos de aprendizado com pequenas amostras.

Diante dessas questões, este artigo tem como objetivo discutir em detalhes as perspectivas e desafios da aprendizagem com pequenas amostras na era do *Big Data*. Para isso, será realizada uma revisão abrangente das técnicas de FSL e suas aplicações como uma alternativa para lidar com a escassez de dados. Ao serem examinadas as abordagens existentes e explorados exemplos de aplicação do FSL, espera-se fornecer uma visão abrangente das estratégias mais eficazes e identificar lacunas de pesquisa que precisam ser abordadas. A compreensão e a superação dos desafios da aprendizagem com pequenas amostras são fundamentais para impulsionar o avanço da inteligência artificial em um contexto em que o acesso a grandes conjuntos de dados pode ser limitado.

Espera-se que esta revisão das técnicas de FSL e sua aplicação em conjuntos de dados limitados forneça *insights* valiosos para a melhoria de modelos de aprendizagem de máquina em diferentes áreas de aplicação. Através dessa abordagem, podemos explorar maneiras inovadoras de lidar com a escassez de dados e avançar na busca por soluções eficazes para problemas de aprendizagem com pequenas amostras na era do *Big Data*.

Este artigo está organizado da seguinte forma: na seção 2, apresenta-se a metodologia do trabalho, incluindo a descrição do método usado para selecionar os estudos incluídos na revisão. Na sessão 3, é apresentada uma revisão da literatura sobre a contextualização de aprendizagem com pequenas amostras, incluindo conceitos fundamentais e principais abordagens. Na seção 4, discutiremos em detalhes os desafios enfrentados na aprendizagem com pequenas amostras, abordando questões relacionadas à seleção de amostras, métodos de treinamento e capacidade de generalização. Na seção 5, apresentaremos exemplos de aplicações práticas do FSL em diferentes domínios. Por fim, na seção 6, concluiremos o artigo, resumindo as principais contribuições e apontando direções para pesquisas futuras.

Metodologia

Para a elaboração desta revisão de literatura, foi empregada uma estratégia de busca sistemática visando identificar trabalhos relevantes ao tema de FSL na era do *Big Data*. Os critérios de inclusão para os estudos foram: (1) trabalhos publicados recentemente, no período da última década (exceto trabalhos históricos de introdução a conceitos importantes); (2) artigos escritos em português ou inglês; e (3) estudos que fossem experimentais e teóricos, focalizando especificamente modelos de FSL em Aprendizado Baseados em Similaridade;

Com base nesses critérios, foi realizada uma busca detalhada em bases de dados acadêmicas mais conhecidas, incluindo [Google Acadêmico](#), [IEEE Xplore](#) e [ACM Digital Library](#), visando a obtenção de uma amostra representativa e atualizada de trabalhos na área. Foram excluídos da revisão trabalhos que não atendiam aos critérios de inclusão previamente definidos, assim como aqueles que não eram diretamente pertinentes ao tema de interesse, garantindo assim a relevância e a especificidade dos estudos incluídos.

Os artigos selecionados foram, então, analisados de forma crítica e sistemática, com enfoque nas abordagens metodológicas empregadas, nos resultados alcançados e nas conclusões apresentadas pelos autores. Desta forma, esta revisão busca proporcionar um panorama abrangente e conciso dos avanços recentes, das tendências emergentes e dos desafios persistentes no campo da aprendizagem de pequenas amostras na contemporaneidade do *Big Data*.

Fundamentação Teórica

Nessa sessão é apresentado conceitos-chave que formam a base estrutural do estudo, tais como Inteligência Artificial, um ramo da ciência da computação que visa criar sistemas capazes de realizar tarefas que, normalmente, necessitariam da inteligência humana; Aprendizado de Máquina (*Machine Learning*), uma subárea da Inteligência Artificial que foca no desenvolvimento de algoritmos que permitem que os sistemas aprendam e façam previsões ou tomadas de

decisão; *Big Data*, que se refere a conjuntos de dados extremamente grandes e complexos, os quais são analisados computacionalmente para revelar padrões, tendências e associações; e *Few-shot Learning*, uma técnica de aprendizado de máquina que procura construir modelos precisos com o emprego de um número muito reduzido de exemplos.

Inteligência Artificial

De acordo com Russell e Norvig (2013), a Inteligência Artificial (IA) abrange a disciplina científica dedicada ao projeto de máquinas capazes de realizar tarefas que exigem inteligência semelhante à humana. O principal propósito da IA está em criar máquinas e computadores capazes de executar atividades atualmente consideradas intrínsecas ao domínio humano, como compreensão da linguagem falada, dedução lógica, assimilação e adaptação.

Os estudos e aplicações da IA se iniciaram na década de 40, sendo um dos primeiros trabalhos realizados em 1943 por Warren McCulloch e Walter Pitts na Universidade de Illinois. Eles propuseram um modelo de neurônios artificiais, onde cada neurônio se caracteriza por “ligado” ou “desligado”, desse modo, o estado de um neurônio era analisado como “equivalente em termos concretos a uma proposição que definia seu estímulo adequado”. Sendo essa a marco inicial da chamada *neural networks* ou redes neurais (MCCULLOCH; PITTS, 1943).

Graças aos recentes avanços, a IA se tornou imensamente relevante para inúmeras áreas do conhecimento além da computação. Desde o controle de carros autônomos com processamento de imagens até diagnósticos de doenças, a IA está cada vez mais presente. Sendo esta uma poderosa ferramenta para a solução de problemas no qual se possui a necessidade cumprir tarefas específicas ao processar abundância de dados ou até mesmo reconhecer padrões em base de dados reduzidas.

Aprendizado de Máquina e Big Data

O Aprendizado de Máquina (AM) ou *Machine Learning (ML)* pode ser vista como uma subárea da IA, centralizando seu foco em desenvolver modelos capazes de adquirir conhecimento de forma autônoma, aprimorando-se continuamente mediante experiências acumuladas. O AM é fundamentado na implantação de algoritmos aptos a processar dados e, a partir destes, fazer inferências e tomar decisões, utilizando como referência experiências pregressas e soluções previamente consolidadas. O funcionamento do ML é sustentado por aprendizado indutivo, uma modalidade de inferência lógica que possibilita a generalização de conclusões partindo de um conjunto específico de dados. Este aprendizado é categorizado, primariamente, como supervisionado, não supervisionado e semisupervisionado.

No aprendizado supervisionado, o algoritmo é treinado com um conjunto onde a classe de cada exemplo é conhecida, desta forma obtendo um estimador que possa definir a nova classe de novos exemplos. Diferente do aprendizado supervisionado, o não-supervisionado a classe de cada exemplo não é conhecida. Desta forma é comum o trabalho de agrupamento, analisando os exemplos fornecidos e determinar se alguns deles podem ser agrupados de alguma maneira (devido suas características), formando agrupamento ou *clusters*. Por fim, o aprendizado semisupervisionado seria o meio-termo entre aprendizado supervisionado e não-supervisionado. No qual o algoritmo de aprendizado trabalha com um conjunto de treinamento formado por exemplos classificados e um conjunto de dados não classificados (SANCHES, 2003).

Apesar do poder analítico inerente ao ML, é imperativo reconhecer a inexistência de um algoritmo universal, apto a solucionar quaisquer problemas de forma eficaz. Esta constatação evidencia a necessidade de compreender as limitações intrínsecas aos algoritmos de ML, bem como empregar métricas de otimização como a Precisão, Acurácia, Erro Quadrático Médio (EQM), raiz do EQM (REQM), entre outras, para refinar os modelos e mensurar a taxa de erro em situações de classificação ou regressão (MONARD; BARANAUSKAS, 2003).

O campo de ML e o fenômeno do *Big Data* estão intrinsecamente interligados, representando dois pilares fundamentais na transformação digital e na inovação tecnológica contemporânea. O

Big Data refere-se ao processamento e à análise de volumes massivos de dados, que são gerados incessantemente por diferentes fontes, como redes sociais, sensores e transações online. Essa avalanche de dados, caracterizada por sua velocidade, variedade e volume, oferece um terreno fértil para o desenvolvimento e a aplicação de algoritmos de ML. Estes algoritmos, por sua vez, têm a capacidade de identificar padrões, fazer previsões e gerar *insights* valiosos a partir desses dados, permitindo a extração de conhecimento e a tomada de decisões informadas em diversos domínios, como saúde, finanças e ciências ambientais (RAUTENBERG; CARMO, 2019).

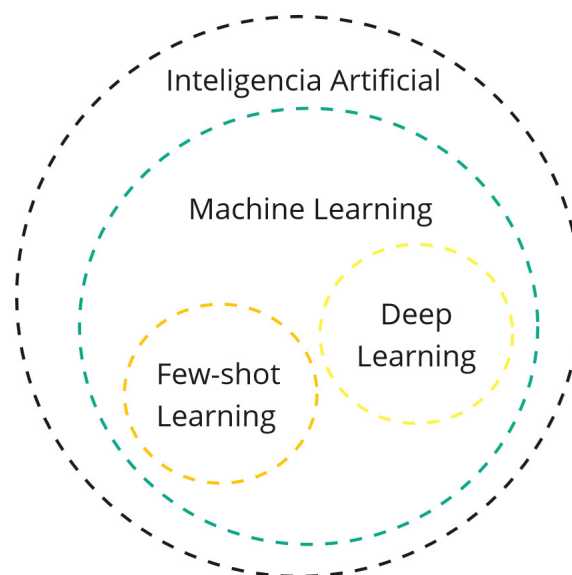
Na conjuntura do *Big Data*, onde volumes imensos de dados são gerados continuamente, o ML apresenta-se como uma ferramenta valiosa para extrair *insights* e conhecimentos. Contudo, paradoxalmente, em diversas situações práticas, como em pesquisas de opinião, estudo de doenças raras, desenvolvimento de novos dispositivos e equipamentos, entre outros, os dados disponíveis são escassos. Nesse cenário, os modelos de ML baseado em aprendizagem com pequenas amostras tornam-se indispensáveis, oferecendo soluções alternativas quando a disponibilidade de dados é limitada.

Few-shot Learning

A aprendizagem com pequenas amostras, também conhecida como *Few-Shot learning* (FSL) é uma área de pesquisa em aprendizado de máquina que se dedica a treinar modelos capazes de generalizar e realizar inferências precisas em tarefas ou classes para as quais há disponibilidade limitada de exemplos de treinamento. Redes de aprendizado profundo convencionais conseguem um bom desempenho na extração de estatísticas complexas, entretanto precisam de uma grande quantidade de dados para aprender. Contudo, os modelos convencionais sofrem com baixa eficiência para pequenas quantidades de dados. Essa abordagem se torna essencial em cenários em que a coleta de grandes volumes de dados é impraticável, custosa ou inviável (ZHENG et al., 2019).

Pode-se representar relação entre FSL com ML em um diagrama de Venn, como mostra a Figura 1. No qual, FSL juntamente com DL são sub-áreas do ML, que por sua vez seria uma subárea da IA.

Figura 1: Diagrama de Venn: Relação entre IA, AM, *Deep Learning* (DL) e FSL

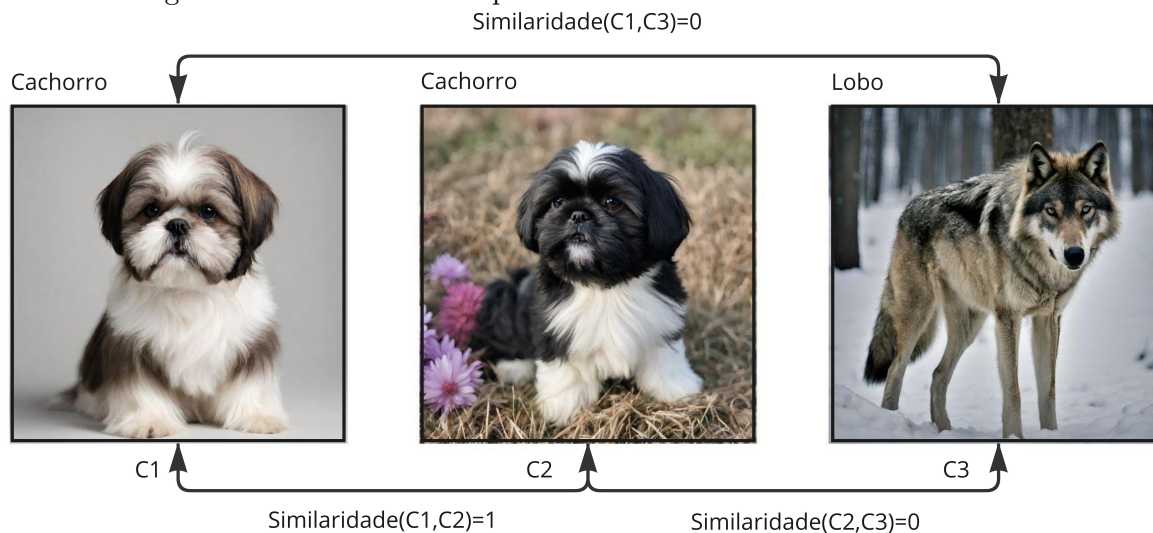


Fonte: Autores.

O objetivo principal na aprendizagem FSL é aprender uma função de similaridade que possa mapear as semelhanças entre as classes nos conjuntos de suporte e de consulta. Normalmente,

funções de similaridade produzem um valor de probabilidade para a semelhança. No exemplo da Figura 2, uma função de similaridade ideal deveria resultar em um valor de 1,0 ao comparar duas imagens de cachorro (C1 e C2). Nos outros dois casos, onde as imagens de cachorros são confrontadas com uma imagem de um lobo, o resultado da similaridade deveria ser 0.0. No entanto, isto é um cenário idealizado. Na prática, os valores poderiam ser 0,95 para C1 e C2 e um valor pequeno superior a 0 para os outros dois casos (como 0,02 e 0,03).

Figura 2: Um cenário ideal para uma medida de similaridade em FSL.

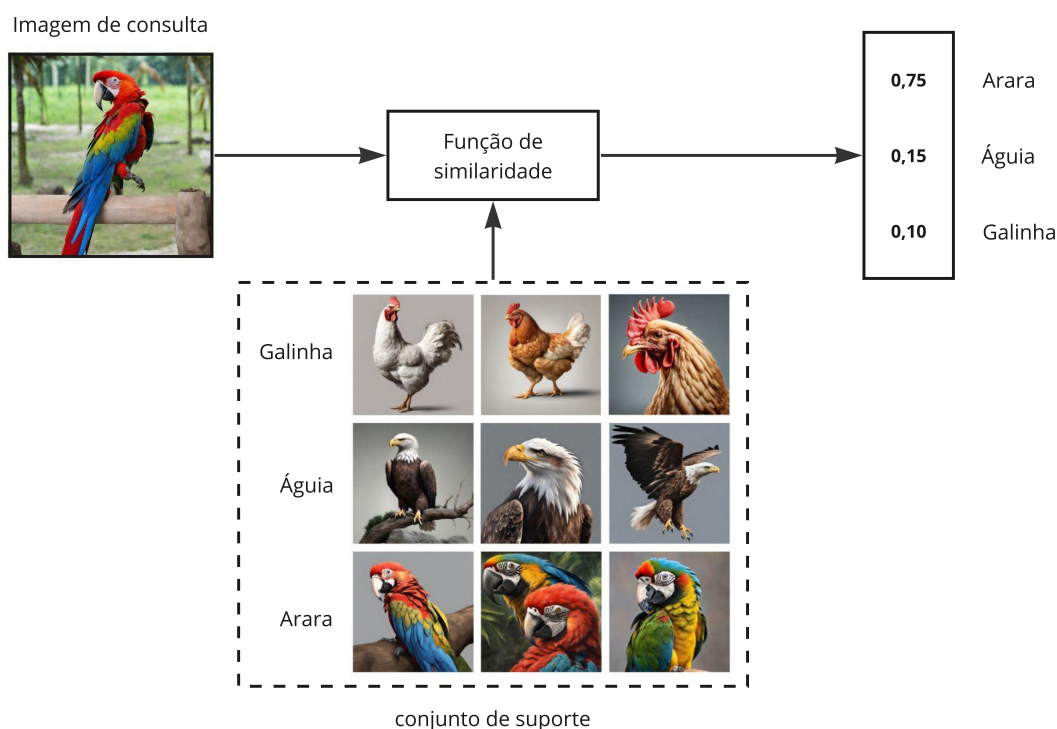


Fonte: Autores.

Normalmente o exemplo dado na Figura 2 é empregado em um conjunto de dados rotulados para treinar os parâmetros de tal função de similaridade. O conjunto de treinamento utilizado para o pré-treino do modelo profundo de maneira supervisionada pode servir a este propósito. Uma vez que os parâmetros da função de similaridade estão treinados, eles podem ser empregados na fase de FSL para determinar a similaridade no conjunto de consulta utilizando as informações do conjunto de suporte. Então, para cada amostra do conjunto de consulta, a classe com a maior similaridade proveniente do conjunto de suporte será inferida como a previsão da etiqueta de classe pelo modelo de FSL. Um exemplo de tal procedimento é ilustrado na Figura 3, no qual uma função de similaridade resulta em um valor de 0,75 ao classificar a imagem da arara.

A aprendizagem FSL, é uma ideia relativamente nova na área de estudo, com suas raízes se estendendo até o começo dos anos 2000, conforme citado por Lu et al. (2020). O desenvolvimento desta abordagem se deu inicialmente em entender como humanos aprendem, visando criar algoritmos capazes de generalizar a partir de um número limitado de exemplos. Como no estudo de Miller, Matsakis e Viola (2000), no qual está entre os primeiros que exploraram este campo de estudo. Não muito tempo depois, o trabalho de Fe-Fei et al. (2003) começaram a desenvolver e formalizar o conceito de aprendizado baseado em um único exemplo, também conhecido como *one-shot learning*. Em outra fase, a de aprendizado profundo, a ascensão do FSL foi, em parte, catalisada por trabalhos como o de Koch et al. (2015). Eles introduziram um dos primeiros estudos que integram Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNN) com aprendizagem FSL, marcando um importante desenvolvimento na interseção destas tecnologias.

Figura 3: Panorama de como um modelo de FSL realiza uma previsão



Fonte: Autores.

Os trabalhos iniciais que mesclam CNN e FSL aproveitaram os benefícios das CNNs na extração de características representativas, elevando o FSL a um novo patamar de desenvolvimento acelerado. Koch et al. (2015) introduziram um modelo de redes siamesas para classificação em tarefas one-shot. As estratégias empregadas focaram no refinamento dos modelos profundos para abordar os desafios do FSL, integrando aprendizagem, métrica e meta-aprendizagem, elementos presentes em abordagens contemporâneas de ponta em FSL.

O progresso em FSL evidenciou significativas transformações nas estruturas de modelo durante a transição para técnicas mais avançadas. A interseção de CNNs, aprendizagem métrica, e meta-aprendizagem tornou-se uma estratégia recorrente. Por mais que seja um campo recente, o FSL é um tema recorrente em estudos de visão computacional, recebendo considerável interesse da comunidade de aprendizagem de máquinas.

Modelos de Aprendizado Baseados em Similaridade

Os modelos baseados em similaridade, métricos ou *learninf-to-measure - L2M* (LU et al., 2020), são caracterizados por adquirirem conhecimento através da avaliação da semelhança em imagens. Um exemplo característico de modelo L2M é o apresentado na Figura 3 constituído por um extrator de características de imagem baseado em CNN, juntamente com uma função de similaridade, empregada para distinguir imagens no espaço de características. Alguns dos atributos distintivos desses modelos métricos são sua simplicidade e eficácia ao comparar características para a generalização de imagens, mesmo quando baseados em um número reduzido de exemplos durante o treinamento.

Quando os dados relativos a um específico desafio são escassos para o adestramento de modelos de aprendizagem, a aquisição de conhecimento por meio de problemas que possuem dados análogos emerge como uma alternativa viável. Isso justifica por que a aprendizagem FSL, é frequentemente interpretada como um dilema de meta-aprendizagem por diversas estratégias. Nesta ótica, a meta-aprendizagem se destaca por capacitar modelos na extração de *insights* a

partir de tarefas com características semelhantes, otimizando a capacidade de generalização do modelo em cenários data-sparse, promovendo assim uma adaptabilidade mais eficaz a novos contextos e desafios.

Meta-aprendizagem

No contexto da aprendizagem com pequenas amostras, existem algumas abordagens e técnicas fundamentais que merecem destaque. Uma delas é a abordagem de “meta-aprendizado” (*meta-learning*), que visa treinar modelos capazes de aprender a aprender. Segundo Finn e Others (2019), tomam a abordagem de que aprendendo uma inicialização de rede que pode se adaptar rapidamente a novas tarefas, esta seria uma forma de meta-aprendizagem ou “aprendendo a aprender”. Para facilitar o aprendizado com poucos dados e corrigir o modelo rapidamente, a meta-aprendizagem é usado para simular várias tarefas durante o treinamento, para projetar o algoritmo correto para adaptação ou aprendizado em um ambiente de recursos compartilhados para classificação baseada em protótipos.

Tarefa K-shot e N-way

A competência de um algoritmo em realizar aprendizado FSL é comumente mensurada pelo seu desempenho em atividades de n-shot, k-way. Neste cenário, um modelo é confrontado com uma amostra de consulta de uma classe inédita e, simultaneamente, recebe um conjunto de suporte, S , formado por n exemplos de k classes distintas e também inéditas. Adicionalmente, existe um parâmetro suplementar e implícito, q , que representa a quantidade de exemplos de consulta para cada classe. Em contraste com métodos de aprendizagem convencionais, que necessitam de um acervo mais amplo de dados, a aprendizagem few-shot opera com uma quantidade limitada de exemplos para treinamento em cada classe, resultando numa variação comum do parâmetro K entre 1 e 10. O parâmetro N pode, contudo, variar consoante o modelo ou a metodologia experimental utilizada (KNAGG, 2022).

Os conjuntos de imagens de suporte e de consulta em uma tarefa são simbolizados por $S = N \times K$ e $Q = N \times q$, respectivamente. A meta do modelo é categorizar as imagens do conjunto de consulta Q , correlacionando-as com as N classes da tarefa, baseando-se em um critério de semelhança. O conjunto original de treinamento é fragmentado em imagens de suporte e consulta. Estas últimas são empregadas para aferir a capacidade do modelo em cada atividade de classificação.

O treinamento e os testes de modelos de FSL são executados por meio de atividades de classificação, utilizando um conjunto destinado, ou conjunto alvo, também segmentado entre imagens de suporte e de consulta. Isso permite a determinação da acurácia média do modelo few-shot. Este método, inserido no contexto de meta-aprendizagem, emprega estratégias distintas de treinamento e teste para otimizar o aprendizado e validação dos modelos.

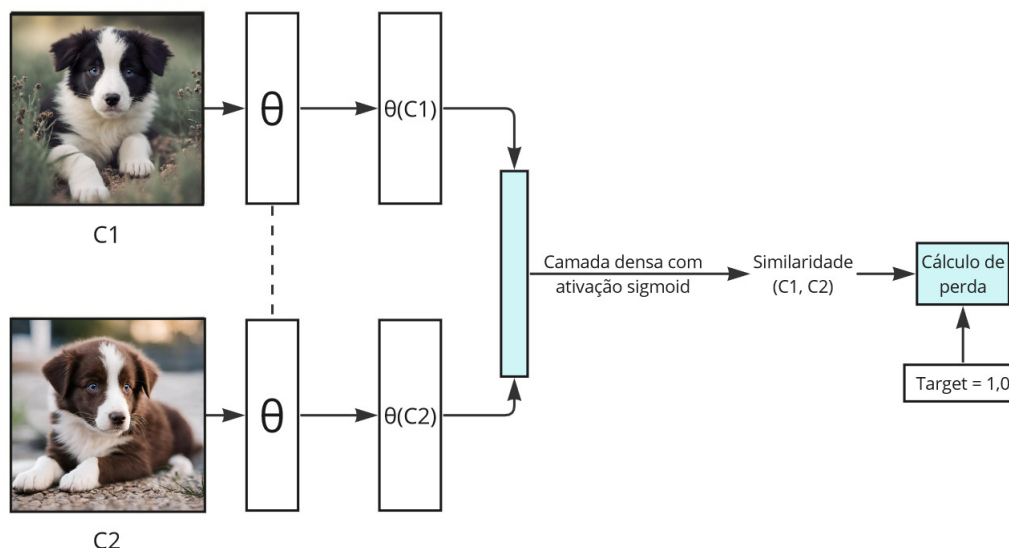
Rede Siamesa

Na literatura de FSL, funções de similaridade não necessitam ser estritamente “funções”. Elas também podem ser, e comumente são, redes neurais; um dos exemplos mais populares é a Rede Siamesa. O nome é derivado do fato de que “gêmeos siameses” são fisicamente conectados. Em contraste com as redes neurais tradicionais, que possuem um ramo de entrada e um de saída, uma Rede Siamesa tem dois ou três ramos de entrada (dependendo do método de treinamento) e um ramo de saída. Existem duas formas de treinar uma Rede Siamesa, sendo a Similaridade de Pares e Redes de Trigêmeos (KUNDU, 2022).

Na primeira, por similaridade de pares, uma Rede Siamesa recebe duas entradas juntamente com seus rótulos correspondentes (utilizando o conjunto de treinamento empregado para o extrator de características pré-treinado). Primeiramente, selecionamos uma amostra de forma aleatória de um conjunto de dados, por exemplo, escolhemos a imagem de um cachorro (Figura

4). Posteriormente, escolhemos novamente uma amostra de forma aleatória do conjunto de dados. Se a segunda amostra pertence à mesma classe da primeira, ou seja, se a segunda imagem também é de um cachorro, então atribuímos um rótulo de “1,0” como verdade básica para a Rede Siamesa. Para todas as outras classes, um rótulo de “0,0” é atribuído como verdade básica.

Figura 4: Visão geral da aprendizagem de similaridade de pares em redes siamesas.



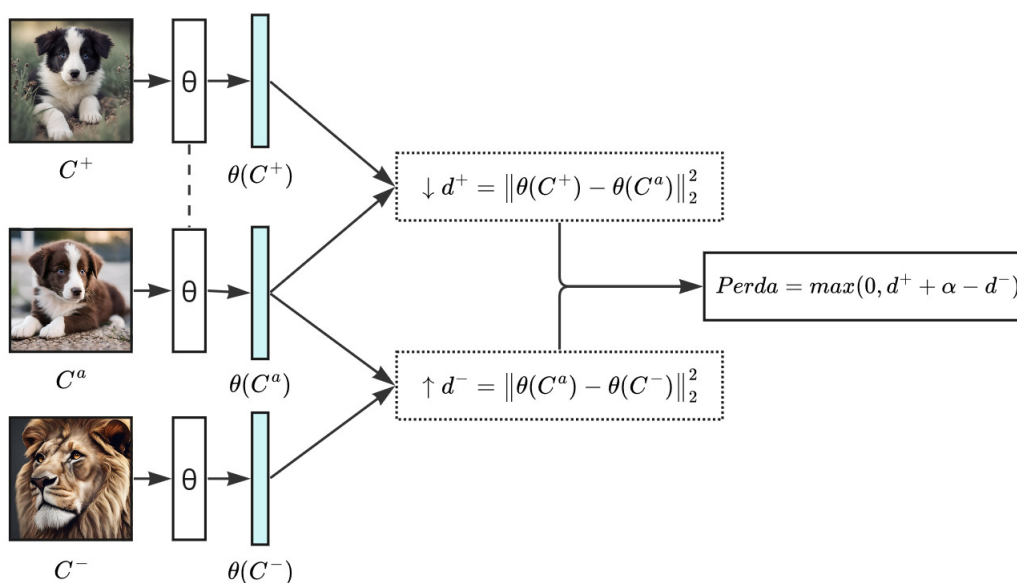
Fonte: Autores.

Portanto, essa rede aprende, fundamentalmente, um critério de correspondência de semelhança por meio de exemplos rotulados, como foi demonstrado com exemplo (Figura 4). Inicialmente, as imagens são processadas individualmente por um extrator de características pré-treinado (normalmente, uma CNN) para obter suas representações correspondentes. Em seguida, as duas representações adquiridas são concatenadas e submetidas a camadas densas e uma função de ativação sigmoide para obter a pontuação de semelhança. Já é de nosso conhecimento se as amostras pertencem ou não à mesma classe, então essa informação é utilizada como a pontuação de semelhança verdadeira para o cálculo da perda e a execução da retropropagação.

Na segunda, Redes de Trigêmeos, fundamenta-se em um critério de “perda tripla”, e pode ser vista como uma expansão da Similaridade de Pares, embora empregue uma estratégia de treinamento distinta. Inicialmente, selecionamos aleatoriamente uma amostra de dados do conjunto de treinamento, denominada amostra “âncora”. Posteriormente, escolhemos outras duas amostras de dados: uma da mesma classe da amostra âncora, chamada de amostra “positiva”, e outra de uma classe diferente, denominada amostra “negativa”.

Uma vez selecionadas, estas três amostras são processadas pela mesma rede neural para obter suas respectivas representações no espaço de incorporação. Posteriormente, calculamos a distância L2 normalizada entre as representações da amostra âncora e da amostra positiva (designemo-la por “ d^+ ”) e a distância L2 normalizada entre as representações da amostra âncora e da amostra negativa (designemo-la por “ d^- ”). Estes parâmetros permitem-nos definir uma função de perda que deve ser minimizada, conforme ilustrado na Figura 5.

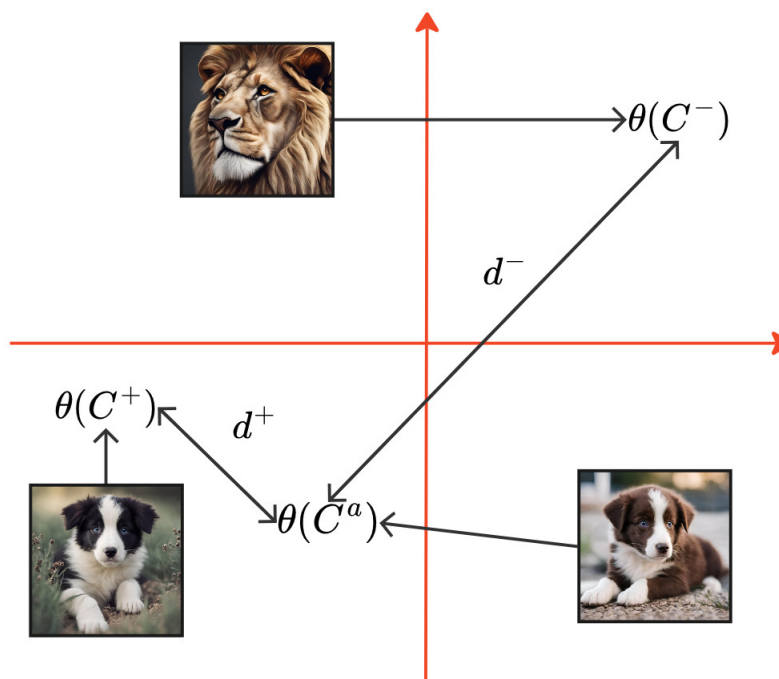
Figura 5: Visão geral da aprendizagem de redes de trigêmeos em redes siamesas.



Fonte: Autores.

Onde se tem “> 0” representa uma margem que impede que os dois termos da função max sejam equivalentes. O objetivo é distanciar ao máximo as representações da âncora e das amostras negativas no espaço de incorporação, enquanto aproxima o máximo possível as representações da âncora e das amostras negativas. Na Figura 6 é possível observar representações de amostras de dados no espaço de incorporação, no qual o ideal é que o d^+ seja mínimo.

Figura 6: Exemplo de como as representações de amostras de dados no espaço de incorporação são alinhadas.

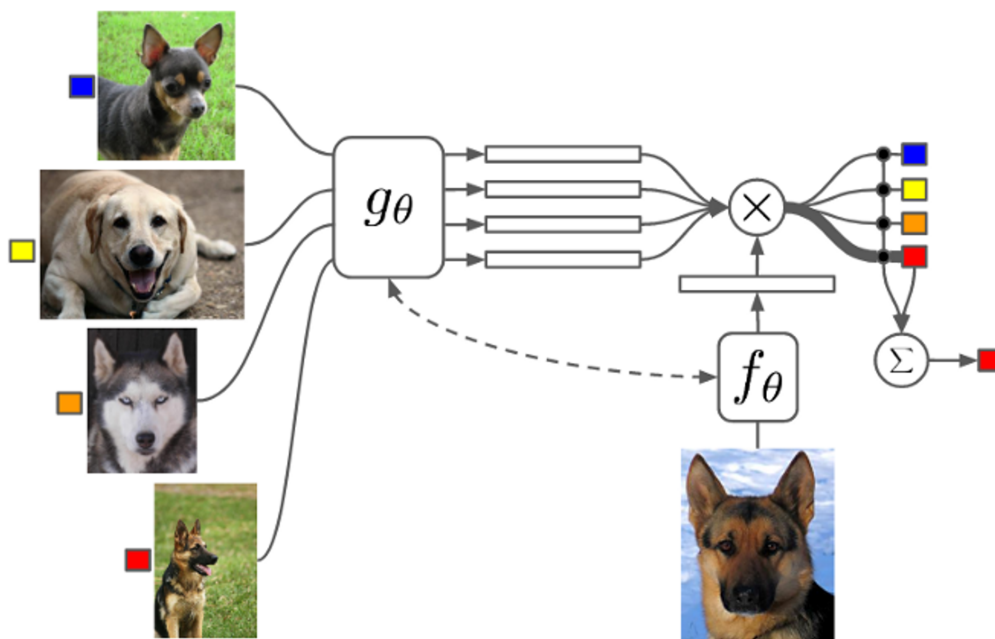


Fonte: Autores.

Redes de correspondência

Redes de correspondência, ou redes Matching (VINYALS et al., 2016), foram concebidas para analisar a concordância entre dois conjuntos distintos de dados. Essas são estruturas voltadas para classificação de múltiplas categorias, que empregam sistemas específicos para a extração de atributos das imagens de suporte e de consulta, avaliando a semelhança através da distância de cosseno entre os *embeddings*, oferecendo assim uma solução refinada e consistente para desafios de classificação. Conforme ilustrado na Figura 7, as imagens de suporte e consulta são processadas por estruturas f_θ e g_θ respectivamente para a obtenção de atributos (podendo ser, em determinados contextos, consideradas como $f_\theta = g_\theta$). A concordância é aferida pela distância do cosseno entre os embeddings ao invés da distância direta, fornecendo um panorama mais acurado e detalhado da semelhança intrínseca entre os conjuntos analisados.

Figura 7: Redes de correspondência com 4-way e 1-shot, com 1 imagem de consulta q .



Fonte: Adaptado de Vinyal et al. (2016).

Inicialmente, as Redes de correspondência incorporam uma amostra de alta dimensão em um espaço de baixa dimensão e, em seguida, realizam uma forma generalizada de classificação baseada em vizinhos mais próximos, conforme descrito pela equação 1.

$$\hat{y} = \sum_{i=1}^k a(\hat{x}, x_i) y_i \quad (1)$$

A previsão do modelo se dá pela soma ponderada dos rótulos y_i do conjunto de suporte, sendo os pesos uma função de similaridade par a par, $a(\hat{x}, x_i)$, entre o exemplo de consulta, \hat{x} , e as amostras do conjunto de suporte x_i . Os rótulos y_i nesta equação são vetores de rótulos codificados em one-hot.

Se escolher a função de similaridade como $1/k$ para as k amostras mais próximas da amostra de consulta e 0 para as outras, recuperamos o algoritmo k-vizinhos mais próximos. O ponto crucial é que as Redes de Correspondência são totalmente diferenciáveis, desde que a função de atenção $a(\hat{x}, x_i)$ também seja.

No trabalho de Vinyals et al. (2016), os autores optam por uma softmax simples sobre similaridades de cosseno no espaço de incorporação como sua função de atenção. A função de

incorporação utilizada é uma CNN, que por sua vez, é diferenciável, tornando a Redes de Correspondência totalmente diferenciável, tornando simples a tarefa de ajustar o modelo inteiro se necessário.

$$a(\hat{x}, x_i) = \frac{e^{c(f(\hat{x}), g(x_i))}}{\sum_{j=1}^k e^{c(f(\hat{x}), g(x_j))}} \quad (2)$$

Na equação 2, c representa a similaridade de cosseno e as funções $f(\hat{x})$ e $g(x_i)$ são as funções de incorporação para as amostras do conjunto de consulta e suporte, respectivamente. Outra interpretação é que o conjunto de suporte atua como uma forma de memória, e ao visualizar novas amostras, a rede gera uma previsão recuperando os rótulos de amostras semelhantes dessa memória.

Na prática, os autores do trabalho de Vinyals et al. (2016) utilizam um *Long Short-Term Memory (LSTM)*, um tipo de especial de CNN que adequado para aprender padrões em sequências de dados para calcular o *Full Context Embeddings (FCE)*, um processo introduzido pelos autores de calcular incorporações (embeddings) levando em consideração o contexto completo das informações disponíveis. Com atenção para modificar a incorporação da amostra de consulta, os autores obtiveram um aumento notável de desempenho, apesar de introduzir mais cálculos e uma ordenação um pouco arbitrária do conjunto de suporte. Sendo esse, um artigo inovador que aprimora a ideia de um algoritmo de vizinhos neural totalmente diferenciável.

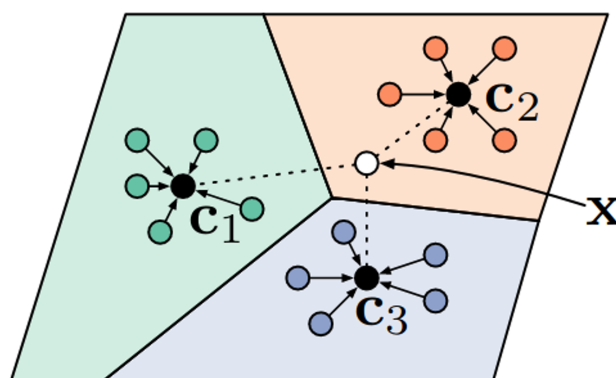
Redes Prototípicas

Nas Redes Prototípicas, Snell, Swersky e Zemel (2017) aplicam um viés indutivo persuasivo na forma de protótipos de classe para alcançar um desempenho de few-shot notável, superando as Redes de Correspondência sem a complicação do FCE. A suposição central é de que existe uma incorporação onde amostras de cada classe se agrupam em torno de uma única representação prototípica, que é simplesmente a média das amostras individuais. Essa concepção simplifica a classificação n-shot quando $n > 1$, já que a classificação é realizada simplesmente atribuindo o rótulo do protótipo de classe mais próximo. A equação 3 de Redes Prototípicas calcula protótipos de classe. S_k é o conjunto de suporte pertencente à classe k e f_ϕ é a função de incorporação.

$$c_k = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} f_\phi(x_i) \quad (3)$$

O modelo, basicamente, adquire uma métrica no espaço de atributos de imagens que simboliza uma semelhança por distância, visando a projeção de uma imagem com base no centroide da classe. As imagens de consulta recebem rótulos ao identificar o protótipo de classe mais próximo, conforme ilustrado na Figura 8.

Figura 8: Redes Prototípicas.



Fonte: SNELL et al., 2017.

Aplicações Práticas

Nessa sessão é apresentado trabalhos e estudos que demonstram o emprego efetivo do FSL em distintas áreas de aplicação. Será dada ênfase em como o FSL está sendo utilizado em tarefas como a classificação de imagens, uma técnica crucial para o reconhecimento e categorização de imagens em diversos contextos; a detecção de objetos, processo essencial para identificar e localizar objetos dentro de imagens; e a geração de linguagem natural, que modela a capacidade de sistemas de computador de gerar texto coerente e contextualmente relevante. Através da exposição desses trabalhos, busca-se mostrar como o FSL está sendo usado em diferentes campos.

Classificação com FSL de imagens

No trabalho de Leao et al. (2022), é apresentado um método que emprega FSL para detectar pneumonia associada à COVID-19 em imagens radiológicas de raio-x. Os dados radiológicos desempenham um papel vital, agindo como substratos cruciais para o diagnóstico e subsequente tratamento de quadros clínicos associados à COVID-19 e pneumonia. A utilização do FSL empregado a bases de imagens de raio-x para detecção de COVID-19 se mostrou especialmente útil em conjuntos de dados com pequenas amostras.

O estudo utiliza a abordagem de FSL acoplada a uma Rede Siamesa visando categorizar imagens radiográficas do tórax associadas à COVID-19. O modelo utilizado foi uma Rede Siamesa, que consiste em duas redes neurais idênticas que compartilham os mesmos pesos. Essas redes recebem como entrada duas imagens de raio-x: uma imagem de teste e uma instância do conjunto de suporte, que é um pequeno conjunto de imagens rotuladas utilizado no momento do teste. A ideia é comparar a imagem de teste com as instâncias do conjunto de suporte e determinar a similaridade entre elas.

Diferentemente de outros trabalhos, este estudo realiza uma avaliação externa dos modelos, comparando os resultados com a avaliação interna. Duas abordagens de geração do conjunto de suporte são investigadas: seleção aleatória de imagens para cada classe e cálculo da representação média das imagens de cada classe. O modelo de Rede Siamesa utilizado é baseado na MobileNet (HOWARD et al., 2017) e a perda é calculada utilizando a Entropia Cruzada Categórica. Quatro repositórios de imagens foram utilizados para compor os conjuntos de dados.

No cenário de avaliação interna, o modelo obteve resultados promissores, alcançando mais de 96% de acurácia, precisão e sensibilidade. No entanto, no cenário de avaliação externa, o modelo não conseguiu generalizar para fontes de dados diferentes. Os resultados obtidos na avaliação interna demonstraram um desempenho promissor, com alta acurácia, precisão e sensibilidade. No entanto, os desafios encontrados na avaliação externa ressaltam a necessidade de aprimorar a representatividade dos conjuntos de dados e a seleção de protótipos mais adequados. Esses *insights* podem direcionar futuras pesquisas e contribuir para o desenvolvimento de métodos mais eficazes na detecção de pneumonia relacionada à COVID-19, auxiliando no diagnóstico e tratamento dessa doença.

Detecção de objetos com FSL via reponderação de atributos

O trabalho conduzido por Kang et al. (2019) apresenta um modelo inovador de detecção de objetos com base em pequenas amostras, o qual se destaca pela sua habilidade de aprender a identificar novos objetos utilizando apenas alguns exemplos anotados. Baseado em FSL utilizando CNNs para a detecção de objetos, este modelo se constrói sobre classes de base plenamente rotuladas e demonstra uma adaptabilidade acelerada a novas classes, por intermédio de um aprendiz de características meta e um módulo de reponderação, ambos integrados em uma arquitetura de detecção unificada.

O modelo delineado neste artigo, articula-se em torno de dois módulos cruciais: um aprendiz de características meta e um módulo de reponderação. O aprendiz de características meta é

meticulosamente treinado empregando classes de base integralmente rotuladas, visando a inferência de características altamente generalizáveis, aptas a operar na identificação de objetos abrangendo uma diversidade de classes. Em paralelo, o módulo de reponderação emprega exemplos de suporte provenientes de classes emergentes para discernir a relevância das características meta na detecção de objetos correlatos.

O procedimento de treinamento do modelo é executado de forma episódica, ancorado em uma estratégia de aprendizado que valoriza a otimização baseada em pequenas amostras. Em cada episódio, um conjunto diversificado de classes de base é escolhido aleatoriamente, juntamente com um compilado de exemplos de suporte para cada classe nova introduzida. Subsequentemente, o aprendiz de características meta é induzido a extrair peculiaridades relevantes desses exemplos, enquanto o módulo de reponderação se aprimora na conversão desses exemplos de suporte em coeficientes de ponderação. Estes últimos são então instrumentalizados para recalibrar as características meta no contexto da detecção de objetos.

Uma vez concluído o treinamento, o modelo demonstra proficiência em identificar novos objetos com um escopo reduzido de exemplos anotados. Os experimentos realizados validam a preeminência deste modelo em comparação com métodos referenciados, evidenciando superioridade na detecção de objetos com escassas amostras em diversos datasets. Adicionalmente, o modelo exibe uma adaptabilidade dinâmica e eficaz a classes inéditas, consolidando sua capacidade de aprender a detectar objetos de forma precisa, mesmo quando confrontado com uma quantidade limitada de exemplos anotados.

O valor intrínseco deste artigo é ressaltado pelo impacto significativo do modelo apresentado. Este é notadamente útil em contextos onde a aquisição de dados rotulados é uma tarefa desafiadora, oferecendo soluções práticas para áreas como visão computacional, robótica e segurança, onde a precisão na detecção de objetos é imperativa. Portanto, o modelo introduzido nesta pesquisa manifesta-se como uma inovação relevante, com potencial para transformações práticas em domínios que demandam rigor e adaptabilidade na identificação de objetos.

Geração de Linguagem Natural com FSL para Diálogo Orientado a Tarefas

O trabalho de Peng e Others (2020), apresenta um novo conjunto de dados de referência chamado *FEW-SHOT WOZ*, um conjunto de dados de referência que simula cenários de aprendizado com pequenas amostras para avaliar a geração de linguagem natural (NLG) em sistemas de diálogo orientados por tarefas. Os autores propõem um modelo chamado SC-GPT, que é pré-treinado em um grande corpus de NLG e refinado com rótulos limitados específicos do domínio. Os resultados experimentais mostram que o SC-GPT supera os métodos existentes no conjunto de dados de referência e alcança um desempenho de ponta no conjunto de dados de diálogo orientado por tarefas em múltiplos domínios (MultiWOZ).

O novo conjunto de dados de referência introduzido no trabalho, contém sete domínios e possui um número menor de instâncias de treinamento por domínio, visando avaliar a capacidade de aprendizado com pequenas amostras. Com a proposta dos autores do modelo pré-treinado (SC-GPT), obtiveram resultados que supera em termos de generalização e controle na geração de respostas de diálogo o SC-LSTM. SC-LSTM é um modelo de NLG baseado em LSTM (Long Short-Term Memory) que utiliza um vetor adicional de ato de diálogo e um mecanismo de leitura para informar a geração de respostas em sistemas de diálogo orientados por tarefas.

O SC-GPT é treinado em três etapas. Primeiro, é realizado um pré-treinamento em larga escala. Em seguida, é realizado um pré-treinamento controlado por ato de diálogo, onde o modelo é treinado para gerar respostas condicionadas a atos de diálogo específicos. Por fim, o modelo é refinado com rótulos limitados específicos do domínio usando o conjunto de dados *FEW-SHOT WOZ*.

Os resultados experimentais mostram que o SC-GPT supera outros métodos de referência em termos de métricas de avaliação automática e humana. O modelo alcança um desempenho de ponta no conjunto de dados MultiWOZ e demonstra uma capacidade de generalização e controle

na geração de respostas de diálogo. Com isso, pode-se afirmar que o conjunto de dados *FEW-SHOT WOZ* e o modelo SC-GPT fornecem uma base sólida para o desenvolvimento de sistemas de diálogo orientados por tarefas com capacidade de aprendizado com pequenas amostras.

O conjunto de dados *FEW-SHOT WOZ* permite simular cenários de aprendizado com pequenas amostras, enquanto o modelo SC-GPT demonstra uma capacidade de generalização e controle na geração de respostas de diálogo. Essas contribuições são importantes para avançar o campo da geração de linguagem natural em sistemas de diálogo e podem ser aplicadas no desenvolvimento de sistemas de diálogo mais flexíveis e eficazes. Isso tem um impacto significativo no desenvolvimento de sistemas de diálogo mais flexíveis, controláveis e capazes de generalizar para novos domínios. O artigo contribui para avançar o campo da geração de linguagem natural em sistemas de diálogo e fornece uma base sólida para pesquisas futuras nessa área.

Desafios e Perspectivas

O FSL emergiu como um campo proeminente na aprendizagem de máquinas, onde se procura construir modelos capazes de reconhecer padrões e tomar decisões acertadas com a exposição a um número pequeno de exemplos. A abordagem tem o potencial de simular a habilidade humana de aprender rapidamente conceitos novos, apontando para inovações significativas na área de IA. No entanto, a implementação eficiente de técnicas de FSL enfrenta um conjunto de desafios e contempla diversas perspectivas futuras.

Um dos principais desafios em FSL é a construção de modelos que possam generalizar bem a partir de um conjunto muito limitado de dados, uma tarefa notoriamente difícil dada a vasta diversidade e complexidade dos dados na realidade. A preservação da precisão e robustez do modelo, na presença de restrições de dados, exige uma reavaliação contínua de métodos e estratégias empregados, orientando a busca por otimizações e melhorias na arquitetura, nos métodos de treinamento e nas técnicas de regularização.

Outro desafio é a capacidade de generalização para novas tarefas ou classes com pequenas amostras. Os modelos de aprendizagem com pequenas amostras devem ser capazes de extrapolar informações a partir de um número limitado de exemplos e aplicar esse conhecimento a situações inéditas. Essa capacidade de generalização é crucial para garantir a aplicabilidade do modelo em cenários do mundo real.

Além disso, o desenvolvimento de métodos de FSL que possam ser aplicados de maneira universal a uma variedade de tarefas e domínios é crucial. A transposição de métodos bem-sucedidos de um domínio para outro ainda é um obstáculo, e explorar abordagens que promovam a adaptabilidade e a flexibilidade é um aspecto preponderante para avançar na aplicação prática de técnicas de FSL.

As perspectivas futuras para o FSL são amplas e promissoras, à medida que a comunidade científica explora novas fronteiras e se aprofunda nas nuances deste paradigma de aprendizagem. A contínua investigação de métodos de meta-aprendizagem, técnicas de transferência de aprendizado e o desenvolvimento de estratégias de aprendizagem não supervisionadas são vistas como etapas essenciais para a realização do potencial pleno do FSL. Ademais, a integração de FSL com outros avanços em aprendizagem de máquinas, como aprendizagem profunda e aprendizagem por reforço, pode resultar em sistemas mais coesos e eficientes, expandindo as capacidades e aplicações do FSL.

A necessidade de expandir a aplicabilidade do FSL a domínios mais amplos e diversificados impulsionará a inovação, e a resolução dos desafios atuais poderá abrir portas para a emergência de tecnologias mais avançadas e sistemas de aprendizagem de máquinas mais versáteis e intuitivos. Portanto, é imprescindível a continuidade do aprofundamento teórico e experimental no campo de FSL, visando superar os obstáculos existentes e explorar todo o espectro de possibilidades que este ramo promissor da IA tem a oferecer.

Considerações Finais

O foco central deste artigo foi a intersecção entre o FSL e o *Big Data*, com especial atenção aos Modelos de Aprendizado Baseados em Similaridade. Uma variedade de trabalhos foi revista, permitindo um entendimento profundo sobre métodos, aplicações e desafios inerentes à aplicação de FSL na era de grandes volumes de dados.

Uma visão integrada e multifacetada dos desenvolvimentos e avanços recentes foi proporcionada através da revisão da literatura sobre FSL. Uma compreensão mais detalhada sobre a operacionalização e as abordagens metodológicas empregadas foi obtida, revelando *insights* valiosos sobre a eficácia e aplicabilidade destes modelos em contextos e domínios diversos.

Foi evidenciado que os Modelos de Aprendizado Baseados em Similaridade são instrumentos vitais no universo do FSL, apresentando soluções inovadoras para os desafios associados ao aprendizado com conjuntos de dados limitados. Ao se basearem na comparação e identificação de similaridades entre instâncias, estratégias de aprendizado mais robustas e adaptáveis são habilitadas, sendo essenciais em cenários onde dados são restritos.

Contudo, também foram revelados diversos desafios e limitações, destacando-se a necessidade de desenvolvimento e aprimoramento contínuo das técnicas de FSL. A qualidade dos dados, a representatividade das amostras e a capacidade de generalização dos modelos são questões que precisam de atenção e pesquisas adicionais, assegurando a validade e eficácia dos modelos em cenários práticos.

Espera-se que, ao consolidar um conjunto de conhecimentos sobre FSL e Modelos de Aprendizado Baseados em Similaridade, este trabalho possa contribuir para o entendimento científico deste importante campo da aprendizagem de máquina e servir como um fundamento sólido para futuras inovações e explorações de novas abordagens e soluções.

Contudo, é imperativo reconhecer as limitações inerentes a este estudo. A seleção e o enfoque temático dos trabalhos analisados correspondem a um segmento específico, abrindo caminho para futuras investigações e análises numa gama mais extensa de modelos e técnicas de aprendizado de poucos exemplos, com o objetivo de tratar problemas distintos que vão além da classificação de imagens, objetos e geração de linguagem natural. Ao analisar campos como pesquisa de opinião, por exemplo, nos quais a disponibilidade de dados pode ser limitada devido à dificuldade de respostas, evidencia-se que a criação de estratégias de FSL pode tornar-se crucial. Assim sendo, em estudos futuros, é vivamente recomendada a exploração de métodos de FSL em contextos diversificados, com carência de dados, possibilitando a obtenção de uma percepção mais diversificada das competências e implementações do FSL.

Conclui-se ressaltando a importância do FSL em um contexto contemporâneo de abundância de dados. Deseja-se que este artigo seja um recurso informativo e inspiracional para aqueles interessados em avançar no campo da aprendizagem de máquina com pequenas amostras. A continuidade da pesquisa e o intercâmbio entre diferentes áreas do conhecimento são vistos como fundamentais para explorar novas fronteiras e construir um conhecimento mais holístico e transformador no campo da aprendizagem de máquina.

Referências

- FE-FEI, L. et al. A bayesian approach to unsupervised one-shot learning of object categories. In: IEEE. *proceedings ninth IEEE international conference on computer vision*. [S.l.], 2003. p. 1134–1141.
- FINN, C.; OTHERS. Online meta-learning. In: PMLR. *International Conference on Machine Learning*. [S.l.], 2019. p. 1920–1930.
- HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDREETTO, M.; ADAM, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

- KANG, B.; LIU, Z.; WANG, X.; YU, F.; FENG, J.; DARRELL, T. Few-shot object detection via feature reweighting. In: ICCV. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. [S.l.], 2019. p. 8420–8429.
- KNAGG, O. *Advances in Few-Shot Learning: A Guided Tour*. 2022. Acessado em 31 de dezembro de 2023. Disponível em: (<https://shorturl.at/cEHX3>).
- KOCH, G.; ZEMEL, R.; SALAKHUTDINOV, R.; OTHERS. Siamese neural networks for one-shot image recognition. In: *ICML Deep Learning Workshop*. [S.l.: s.n.], 2015. v. 2, p. 0.
- KUNDU, R. *Few-Shot Learning*. 2022. Acessado em 31 de dezembro de 2023. Disponível em: (<https://blog.paperspace.com/few-shot-learning/>).
- LEAO, P. P. de S.; SANTOS, E. M. dos; PINTO, R. A.; EVANGELISTA, L. G. C. Detecção de pneumonia causada por covid-19 utilizando few-shot learning. In: SBC. *Anais do XXII Simpósio Brasileiro de Computação Aplicada à Saúde*. [S.l.], 2022. p. 391–400.
- LU, J.; GONG, P.; YE, J.; ZHANG, C. *Learning from very few samples: A survey*. 2020. ArXiv preprint arXiv:2009.02653.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, p. 115–133, 1943.
- MILLER, E. G.; MATSAKIS, N. E.; VIOLA, P. A. Learning from one example through shared densities on transforms. In: IEEE. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000)*. [S.l.], 2000. p. 464–471.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, v. 1, n. 1, p. 32, 2003.
- PENG, B.; OTHERS. *Few-shot natural language generation for task-oriented dialog*. 2020. ArXiv preprint arXiv:2002.12328.
- RAUTENBERG, S.; CARMO, P. R. V. D. Big data e ciência de dados: complementariedade conceitual no processo de tomada de decisão. *Brazilian Journal of Information Science*, v. 13, n. 1, p. 56–67, 2019.
- RUSSELL, S.; NORVIG, P. *Inteligência Artificial*. 3. ed. São Paulo: Elsevier, 2013.
- SANCHES, M. K. *Aprendizado de máquina semi-supervisionado: proposta de um algoritmo para rotular exemplos a partir de poucos exemplos rotulados*. Tese (Doutorado) — Universidade de São Paulo, 2003.
- SNELL, J.; SWERSKY, K.; ZEMEL, R. Prototypical networks for few-shot learning. In: *Advance in Neural Information Processing Systems*. [S.l.]: Curran Associates, Inc., 2017. v. 30. (<https://shorturl.at/afmO>).
- VINYALS, O.; BLUNDELL, C.; LILLICRAP, T.; WIERSTRA, D.; OTHERS. Matching networks for one shot learning. *Advances in neural information processing systems*, v. 29, p. 3630–3638, 2016.
- WANG, Y.; YAO, Q.; KWOK, J. T.; NI, L. M. Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)*, ACM New York, NY, USA, v. 53, n. 3, p. 1–34, 2020.
- ZHENG, Y.; OTHERS. Principal characteristic networks for few-shot learning. *Journal of Visual Communication and Image Representation*, v. 59, p. 563–573, 2019.