



APRENDIZAGEM PROFUNDA E IA VERDE: CAMINHOS PARA UM FUTURO MAIS SUSTENTÁVEL

¹Vivian Rique Gil Ferraro (IC- discente de IC sem bolsa); ²Daniel da Silva Costa (Coorientador); ¹Pedro Nuno de Souza Moura (Orientador).

1 - Departamento de Informática Aplicada; Centro de Ciências Exatas e Tecnologia; Universidade Federal do Estado do Rio de Janeiro

2 - Programa de Pós-Graduação em Informática (PPGI); Centro de Ciências Exatas e Tecnologia; Universidade Federal do Estado do Rio de Janeiro

Palavras-chave: Aprendizagem profunda; IA verde; Sustentabilidade; Consumo de energia; Desenvolvimento sustentável.

Resumo: Este trabalho aborda o consumo elevado de energia, de recursos computacionais, financeiros e os impactos ambientais produzidos durante treinamentos e inferências de modelos de aprendizagem profunda. Foram realizados experimentos com métricas utilizadas para comparar modelos de aprendizagem profunda. Para isso, utilizamos a base de dados MNIST e dois modelos bastante referenciados na literatura: LeNet-5 e AlexNet. Durante as fases de treinamento e de inferência, foram medidos o tempo de processamento tomado, a quantidade de parâmetros, o número de operações de ponto flutuante (*Flops*) e o consumo energético do hardware.

Introdução: A Inteligência Artificial (IA) é um campo interdisciplinar que abrange uma ampla gama de tópicos, incluindo Aprendizado de Máquina, Processamento de Linguagem Natural, Visão Computacional, robótica, dentre outros. Dentre as técnicas de Aprendizado de Máquina, a Aprendizagem Profunda se destaca por utilizar as chamadas redes neurais artificiais para aprender a partir de dados. Essas redes são inspiradas em estruturas cerebrais e são capazes de aprender padrões complexos em dados, sendo utilizadas para uma variedade de tarefas. No entanto, apesar dos avanços proporcionados pelo modo tradicional de pesquisa em Aprendizagem Profunda, os modelos, pela necessidade de grande capacidade computacional, consomem muita energia, gerando impactos negativos no meio ambiente. Para minimizar esses impactos, um campo emergente é a IA verde, “pesquisa de IA que produz novos resultados sem aumentar o custo computacional e, idealmente, reduzi-lo, enquanto a IA vermelha resulta em custos computacionais (e, portanto, de carbono) rapidamente crescentes, a IA verde tem o efeito oposto.” (SCHWARTZ et al., 2020). Essa abordagem leva em consideração não apenas o desempenho, que é medido pela métrica de avaliação de interesse, mas também a eficiência do método e o custo computacional utilizado, promovendo uma redução desses custos. O objetivo é encontrar formas de fazer os modelos usarem menos energia e recursos sem perder muito desempenho.

Objetivo: Este estudo tem como objetivo realizar um levantamento e validação das principais métricas adotadas na literatura para comparação de modelos de Aprendizagem Profunda em termos de eficiência, gasto energético e custo computacional. Para isso, foi realizado um conjunto de experimentos com os modelos selecionados e as métricas obtidas foram aplicadas para comparar o desempenho de cada modelo. Além disso, foi discutida a importância da adoção de práticas pela comunidade acadêmica que possam levar a mudanças no paradigma de como as pesquisas envolvendo Inteligência Artificial vêm sendo conduzidas, com o intuito de avaliar como essas medidas podem contribuir para a construção de um futuro mais sustentável para todos.

Metodologia: Para adquirir base teórica e maior conhecimento sobre o assunto, foi realizada a leitura de livro (KROHN et al., 2020) e de diversos artigos. Em seguida, foi feito um levantamento das principais métricas utilizadas na literatura para comparar modelos de aprendizagem profunda. Posteriormente foram realizados experimentos utilizando a base de dados MNIST, que consiste em um conjunto de imagens de dígitos escritos à mão normalizadas em termos de tamanho e centralizadas em uma imagem de tamanho fixo, contendo um conjunto de treinamento com 60.000 exemplos e um conjunto de teste com 10.000 exemplos. Para o treinamento, foram utilizados dois modelos: LeNet-5 (LECUN et al, 1998), uma rede neural convolucional proposta por Yann LeCun em 1998 conhecida por sua arquitetura simples e direta, e AlexNet (KRIZHEVSKY et al., 2017), uma arquitetura de rede neural convolucional projetada por Alex Krizhevsky em colaboração com Ilya Sutskever e Geoffrey Hinton, que venceu o *ImageNet Large Scale Visual Recognition Challenge* em 2012 e é considerada um marco no progresso dos estudos em aprendizado profundo. Os modelos foram treinados para classificar as imagens do conjunto MNIST e, durante as etapas de treinamento e inferência, foram mensurados o tempo de treino, tempo de inferência, número de parâmetros, números de *Flops* (operações de ponto flutuante) e consumo energético do hardware.

Resultados: As redes foram treinadas com o objetivo de atingir as métricas estabelecidas na metodologia, o tamanho do batch foi estabelecido em 128 e os treinos foram feitos em 20 épocas. A Figura 1 mostra a disposição das camadas da rede neural baseada no modelo LeNet-5. Neste modelo, a função de ativação ReLU (unidade linear retificada) foi utilizada nas camadas convolucionais e na camada densa para introduzir não-linearidade na rede. A camada de saída utiliza a função de ativação Softmax. O modelo foi compilado com a função de perda *categorical_crossentropy*, o otimizador *adam* e a métrica de acurácia. Foram realizados 10 treinos e os modelos gerados apresentaram valor médio de acurácia em 0,990556, com desvio padrão de 0,001211; o tempo médio dos treinos foi de 33,8262 segundos, com desvio padrão de 3,6201 segundos; e consumo energético médio de 3956288 μ J (microjoules), com desvio padrão de 169868,8 μ J.

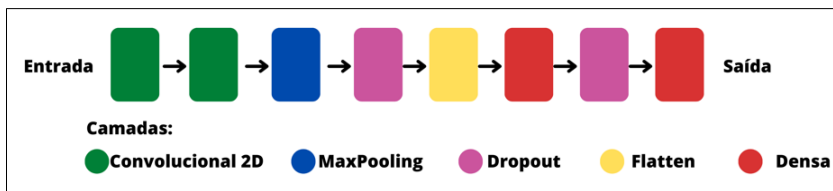


Figura 1 – Esquematisação da arquitetura utilizada no modelo LeNet-5.

Os modelos possuem 609.354 parâmetros; e o número de operações foi de 2,93 GFLOPS em cada treino. O modelo com maior acurácia (sétimo modelo treinado) foi escolhido para a inferência, que apresentou os seguintes resultados: acurácia: 0,9929; Precisão: 0,992841; *recall*: 0,992858; *F1-Score*: 0,992845; o tempo da inferência foi de 854 milissegundos. Para obter o número de FLOPS, utilizamos a biblioteca *keras_flops* e, para o consumo energético utilizamos a biblioteca *pyJoules*.

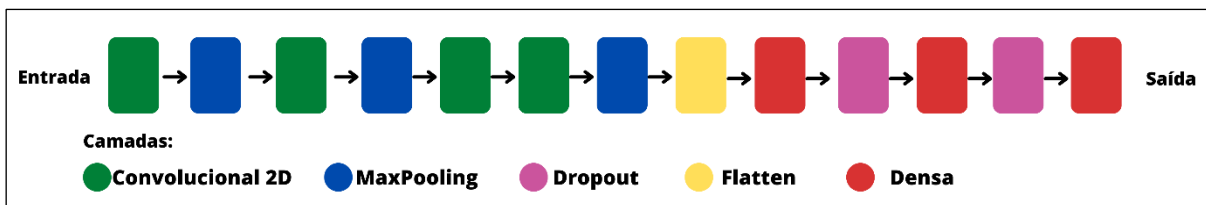


Figura 2 – Esquematisação da arquitetura utilizada no modelo AlexNet.

No modelo AlexNet, além da função de ativação ReLU, a utilização da função tanh (tangente hiperbólica) nas camadas densas auxilia no desempenho da tarefa de classificação. A camada de saída utiliza a função de ativação Softmax. A disposição das camadas dessa rede está descrita na figura 2. O modelo foi compilado com a função de perda *categorical_crossentropy*, o otimizador *adam* e a métrica de acurácia. Após uma sequência de 10 treinos, os modelos apresentaram valor médio das acurácias em 0,988175, com desvio padrão de 0,003202; tempo médio dos treinos em 76,6458 segundos, com desvio padrão de 10,3636 segundos; e consumo energético médio de 6889739,4 μ J, com desvio padrão de 451840,72 μ J. Os modelos possuem 1.938.922 parâmetros e o número de operações foi de 5,67 GFLOPS por treino. O modelo com maior acurácia (segundo modelo treinado) foi escolhido para a inferência, que apresentou os seguintes resultados: acurácia: 0,985; precisão: 0,984997; *recall*: 0,984776; *F1-Score*: Biblioteca; o tempo da inferência foi de 1 segundo e 737982 milissegundos.

Conclusões: Este estudo teve como objetivo avaliar e comparar modelos de aprendizagem profunda em termos de eficiência, gasto energético e custo computacional. Para isso, foram realizados experimentos com os modelos LeNet-5 e AlexNet, utilizando a base de dados MNIST. Os resultados mostraram que ambos os modelos apresentaram desempenho semelhante em termos de acurácia, precisão, *recall* e *F1-Score*, no entanto o modelo AlexNet apresentou um consumo energético consideravelmente maior, cerca de 74,1% superior ao modelo LeNet-5. Esses resultados serviram para validar o uso das métricas energéticas levantadas, pois corresponde ao que se esperava, isto é, que o modelo AlexNet obtivesse valores maiores das métricas adotadas em relação ao modelo LeNet-5. Isso destaca a importância de considerar não apenas a precisão do modelo, mas também outros fatores ao comparar modelos de aprendizagem profunda. Além disso, este estudo enfatiza a necessidade de mudanças no paradigma de como as pesquisas envolvendo Inteligência Artificial vêm sendo conduzidas, visando à construção de um futuro mais sustentável para todos.

Referências:

ANTHONY, Lasse F. Wolff; KANDING, Benjamin; SELVAN, Raghavendra. Carbontracker: tracking and predicting the carbon footprint of training deep learning models. **arXiv preprint**, arXiv:2007.03051, 2020.

DESLAVOV, Radosvet; MARTÍNEZ-PLUMED, Fernando; HERNÁNDEZ-ORALLO, José. Trends in AI inference energy consumption: beyond the performance-vs-parameter laws of deep learning. **Sustainable Computing: Informatics and Systems**, [S.L.], v. 38, p. 100857, abr. 2023.

DOUWES, Constance; ESLING, Philippe; BRIOT, Jean-Pierre. Energy consumption of deep generative audio models. **arXiv preprint**, arXiv:2107.02621, 2021.

INRIA - Institut National de Recherche en Informatique et en Automatique. **pyJoules 0.2.0 documentation**. Disponível em: <<https://pyjoules.readthedocs.io/en/latest/>>. Acesso em: 22 ago. 2023.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey. ImageNet Classification with Deep Convolutional Neural Networks. **Communications of the ACM**, [S.L.], v. 60, n. 6, p. 84–90, 2017.

KROHN, Jon; BEYLEVELD, Grant; BASSENS, Aglaé. **Deep Learning Illustrated: a visual, interactive guide to artificial intelligence**. Boston: Pearson Addison-Wesley, 2020. 416 p.

LACOSTE, Alexandre; LUCCIONI, Alexandra; SCHMIDT, Victor; DANDRES, Thomas. Quantifying the carbon emissions of machine learning. **arXiv preprint**, arXiv:1910.09700, 2019.

LECUN, Yann et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278-2324, 1998.

LENHERR, Nicola; PAWLITZEK, René; MICHEL, Bruno. New universal sustainability metrics to assess edge intelligence. **Sustainable Computing: Informatics and Systems**, [S.L.], v. 31, p. 100580, set. 2021.

SCHWARTZ, Roy; DODGE, Jesse; SMITH, Noah A.; ETZIONI, Oren. Green AI. **Communications of the ACM**, [S.L.], v. 63, n. 12, p. 54-63, 2020.