

Desenvolvimento de um jogo da memória usando classificação de imagens com base em técnicas de Inteligência Artificial

Sofia Souza Luiz

¹Departamento de Informática
Universidade Federal de Viçosa (UFV) – Viçosa, MG – Brasil

sofia.luiz@ufv.br

Abstract. *This meta-paper was produced as a final project for the Artificial Intelligence course, aimed at applying techniques such as Convolutional Neural Networks (CNN) and Multilayer Perceptron (MLP), learned throughout the discipline, in a memory game related to image classification of people based on gender. Therefore, it was possible to explore, analyze and compare the performance of these models in identifying visual patterns, understanding how the network architecture affects the final result of the classification.*

Resumo. *Este meta-artigo foi produzido como projeto final da disciplina de Inteligência Artificial, visando a aplicação de técnicas como Redes Neurais Convolucionais (CNN) e Perceptron Multicamadas (MLP) em um jogo da memória relacionado à classificação de imagens de pessoas com base no gênero. Dessa forma, foi possível explorar, analisar e comparar o desempenho desses modelos na identificação de padrões visuais, compreendendo como a arquitetura da rede afeta o resultado final da classificação.*

1. Objetivo

Diferentemente de um jogo da memória convencional, em que o objetivo do jogador é encontrar pares de imagens iguais, a versão apresentada nesse projeto concentra-se em imagens pertencentes a uma mesma classe, especificamente 'Homem' ou 'Mulher'. Em outras palavras, foi desenvolvido um jogo da memória automatizado no qual o computador precisa identificar pares de pessoas do mesmo gênero, o que só é possível por meio da classificação das imagens pela CNN e MLP. Nesse sentido, foi necessário treinar esses modelos utilizando o dataset CelebA (um conjunto de dados de cerca de 200 mil imagens rotuladas para 40 atributos diferentes, inclusive gênero, que é o foco do trabalho), além de implementar etapas de validação e teste para garantir a confiabilidade da classificação.

2. Trabalhos relacionados

A utilização de redes neurais como solução para problemas de classificação é um tópico amplamente explorado em diversos estudos ao redor do mundo. Pesquisas sobre a importância de Redes Neurais Convolucionais nessa tarefa são particularmente comuns, como a apresentada na International Conference on Computational Intelligence and Data Science de 2018 (ICCIDIS 2018) e intitulada "An Analysis Of Convolutional Neural Networks for Image Classification"[N. Sharma 2018]. Nela, foram comparadas diversos tipos de CNNs para extrair e analisar características de uma imagem, avaliando a

performance de cada uma. Além disso, pode-se destacar a relevância de diversos estudos relativos ao uso de Perceptrons Multilayers na tarefa de classificação como o MLP-Mixer [Tolstikhin et al. 2021], um tipo de arquitetura MLP que se baseia apenas na multiplicação de matrizes e escalas não lineares, sem o uso de convoluções, e o Multi-Scale MLP-Mixer [H. Zhang 2022], que combina vários MLP-Mixers de diferentes escalas.

Tais pesquisas representam uma forte base teórica para o desenvolvimento de uma análise comparativa entre o uso de CNNs e MLPs em problemas de classificação, inseridos em um contexto lúdico de jogo da memória que visa tornar essa comparação interessante e demonstrar a aplicabilidade dessas redes.

3. Metodologia

Classificar imagens a partir do treinamento de redes neurais é uma tarefa vinculada a muitos fatores, como o volume e complexidade dos dados e a arquitetura da rede. Com isso em mente, foram propostos dois modelos para realizar a classificação binária de imagens por gênero, visando encontrar o de melhor desempenho. O primeiro é uma Rede Neural Convolutacional (CNN), um dos mais usados e conhecidos para essa tarefa, obtendo alta acurácia na maioria dos casos, e o segundo é o Perceptron Multicamadas (MLP), que também consegue produzir bons algoritmos classificadores quando bem projetado.

O código está disponível no link do GitHub: <https://github.com/sousofi/Trabalho-Final-de-IA>

3.1. Redes Neurais Convolucionais

A Rede Neural Convolutacional (CNN) é um poderoso algoritmo de deep learning capaz de processar e reconhecer padrões visuais em imagens (e até vídeos), sendo, portanto, uma excelente solução para problemas de classificação. A arquitetura dessa rede é composta por camadas especializadas, como a de convolução, de pooling e de ativação, que, juntas, conseguem extrair características dos dados da imagem. Em suma, a camada de convolução aplica filtros (kernels reconhecedores de padrões) com pesos sobre a entrada, conseguindo captar features e padrões específicos, enquanto a de pooling reduz a dimensão espacial dela, de modo a diminuir o número de parâmetros. Já a camada de ativação introduz não-linearidades ao modelo por meio de funções, o que aumenta a capacidade da rede de reconhecer relações complexas entre os dados.

Tendo isso como base, é importante salientar que a arquitetura da CNN utilizada nesse projeto foi otimizada para tentar obter o melhor desempenho no processamento de imagens de homens e mulheres. Primeiramente, os dados de entrada foram balanceados e divididos igualmente entre as classes, além de serem normalizados para facilitar o aprendizado da rede. Ademais, a arquitetura inclui camadas de convolução (que aplicou função de ativação ReLU) e pooling, seguidas de 4 blocos residuais, que aplicam mais camadas de convolução e pooling, dobrando o número de filtros a cada iteração. E, por se tratar de um problema de classificação binário, a classificação final foi realizada utilizando a função de ativação sigmoide, que transforma a saída da rede em uma probabilidade, no caso, interpretada como a de pertencer ou não a uma classe. Toda essa estrutura permitiu ao modelo conseguir um excelente desempenho na missão que lhe foi dada. A seguir, são mostrados alguns exemplos de previsões feitas.



Figura 1. Exemplos do resultado da classificação com a CNN

Com base na análise das imagens, é possível perceber que a última foi classificada incorretamente pela CNN, sendo atribuída a uma classe diferente da classe desejada. Esse erro, entretanto, é previsível, porque, apesar do bom desempenho geral, o modelo não possui 100% de acurácia, portanto, podendo confundir características ou padrões de uma classe com os de outra, resultando em classificações equivocadas.

3.2. Perceptron Multicamadas

O Perceptron Multicamadas é um tipo de rede neural artificial composta por uma ou mais camadas ocultas, além de uma camada de entrada e outra de saída. Cada camada oculta é formada por neurônios interconectados que possuem pesos ajustáveis e utilizam funções de ativação não lineares para capturar padrões complexos nos dados, característica que torna o MLP um modelo poderoso para tarefas de classificação. O algoritmo utilizado para treinar a rede é o de backpropagation, que ajusta os pesos da rede com base na minimização do erro calculado a partir dos gradientes da função de perda com relação aos parâmetros. Portanto, o Perceptron Multicamadas proporciona flexibilidade e boa capacidade de aprendizado, sendo uma ferramenta valiosa em diversos tipos de aplicação.

Após entender como funciona um MLP, é importante descrever como ele foi adaptada para classificar imagens de homens e mulheres de forma eficiente. Para implementar

o modelo, optou-se por utilizar a biblioteca scikit-learn, que fornece uma implementação do MLP, permitindo ajustar os parâmetros conforme as especificidades da tarefa e a complexidade dos dados, por exemplo. Inicialmente, logo após balancear as imagens para garantir que cada classe estivesse igualmente representada, os dados foram normalizados para acelerar o aprendizado e melhorar o desempenho da rede. Em seguida, foram definidas três camadas ocultas com 512, 256 e 128 neurônios, respectivamente, que utilizaram a função de ativação ReLU para introduzir a não linearidade no modelo e permitir a captação de padrões complexos. Além disso, o modelo foi treinado utilizando otimizador Adam, que ajusta adaptativamente a taxa de aprendizado durante o treinamento. Com essa configuração, o modelo foi capaz de fazer boas previsões, demonstrando que é adequado para a tarefa proposta. Entretanto, assim como a CNN, as previsões do MLP também apresentam inconsistências na classificação da imagem, fator que será melhor abordado no próximo tópico.

4. Comparação de resultados

Em redes neurais, é possível utilizar diversos tipos de modelos em diferentes tarefas, como regressão, previsão, reconhecimento de padrões e classificação. Entretanto, a escolha do modelo ideal para uma tarefa específica depende de treiná-lo adequadamente, a partir do ajuste de hiperparâmetros, a fim de alcançar uma configuração que consiga generalizar bem e, assim, maximizar o sucesso da aplicação. Nesse sentido, para a tarefa de classificação binária proposta, foi analisado o desempenho de dois desses modelos: uma Rede Neural Convolucional (CNN) e um Perceptron Multicamadas (MLP).

Em primeira instância, é importante destacar que a tarefa desse projeto não inclui dados extremamente complexos, sendo que as classes estão balanceadas e divididas igualmente entre si. Além disso, todos os dados foram normalizados antes do treinamento, uma vez que esse processo melhora a estabilidade e convergência do modelo.

Durante o treinamento da CNN foram usadas 40000 imagens para treino e 22000 para validação e alguns hiperparâmetros foram ajustados. Um ponto relevante foi a escolha da função de ativação da camada convolucional: inicialmente, foi usada a sigmóide, apropriada para classificação binária. No entanto, os resultados não foram tão satisfatórios quanto os obtidos com a ReLU. Isso, provavelmente, se deve, em parte, ao fato de que a ReLU, é computacionalmente mais eficiente e ativa apenas uma parte das unidades de forma esparsa, tornando a rede menos propensa ao overfitting. Por outro lado, a sigmoide, ao fornecer uma interpretação de probabilidade de pertencimento a uma classe, demonstrou excelente performance na camada de saída. A variação da taxa de aprendizado também foi analisada, e a de melhor desempenho foi $1e-4$, permitindo atualizações mais suaves nos pesos de modo a estabilizar a convergência. Além disso, apenas 10 épocas foram suficientes para garantir um bom modelo. Testes com um número maior de épocas mostraram que o modelo rapidamente convergia para um bom local mas, posteriormente, apresentava poucas melhorias adicionais na precisão, o que, algumas vezes, resultou em overfitting.

Considerando todos esses fatores, a acurácia final do modelo para dados de treino e validação foram, respectivamente, 99.6% e 96.5%. Abaixo, segue o gráfico de variação da acurácia durante o treinamento.

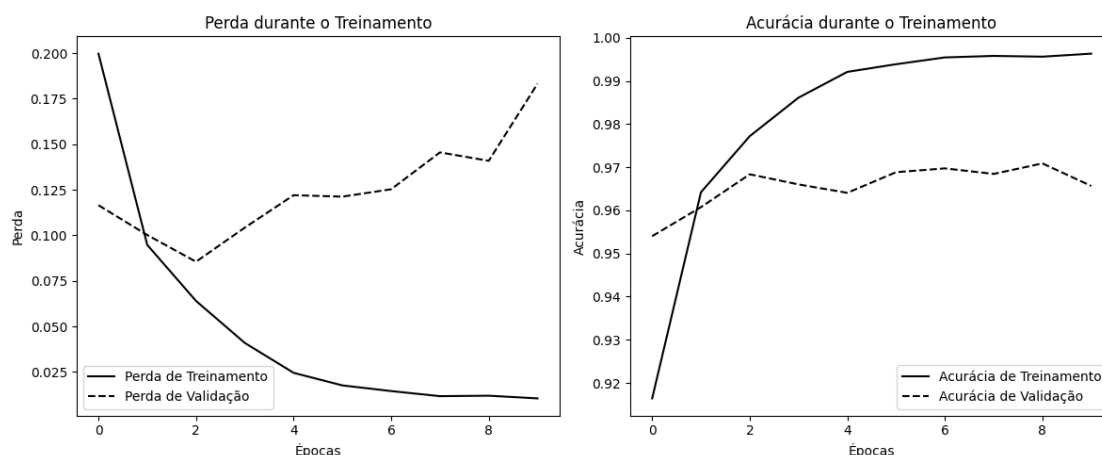


Figura 2. Representação gráfica da acurácia e perda durante o treinamento

Já para o dataset de testes, o modelo obteve 96.6% de acurácia e 0.174 de perda. Esses valores, comparados aos de treino, indicam que o modelo conseguiu generalizar bem para novos dados, sendo capaz de realizar boas previsões. Outros resultados referentes às métricas estão representados abaixo.

	precision	recall	f1-score	support
Female	0.98	0.97	0.98	11541
Male	0.96	0.98	0.97	8459
accuracy			0.98	20000
macro avg	0.97	0.98	0.97	20000
weighted avg	0.98	0.98	0.98	20000

Figura 3. Métricas para classificação da CNN

Para MLP, considerando a complexidade não tão alta dos dados, optou-se por incluir três camadas ocultas. Nesse contexto, foram testados diversos valores de neurônios, e a configuração que apresentou melhor desempenho foi a redução progressiva (512, 256, 128). Isso sugere que a camada de entrada conseguiu captar os padrões mais complexos, enquanto as finais se concentraram em abstrações mais simples e relevantes para saída. Ademais, após definido o número de neurônios nas camadas ocultas, foram avaliadas diferentes quantidades de iterações. Quanto a isso, resultados mostraram que valores muito menores ou muito maiores que 300 comprometeram o desempenho do modelo, o que indica que essa configuração é a ideal para que o modelo consiga aprender características relevantes sem correr o risco de overfitting.

Em comparação com a acurácia apresentada pela CNN, o modelo MLP obteve um valor inferior, alcançando 93% para os dados de teste. No entanto, a acurácia desse modelo ainda é considerada boa, indicando que ele também consegue generalizar bem e fazer previsões confiáveis apesar das limitações. Nesse contexto, para uma visão mais

ampla da diferença de resultado entre os modelos, segue o reporte de classificação desse MLP.

	precision	recall	f1-score	support
Male	0.94	0.92	0.93	7000
Female	0.92	0.94	0.93	7000
accuracy			0.93	14000
macro avg	0.93	0.93	0.93	14000
weighted avg	0.93	0.93	0.93	14000

Figura 4. Métricas para classificação do MLP

Após análise da performance dos dois modelos, ficou evidente que a Rede Neural Convolutacional obteve valores superiores nas métricas, alcançando melhor acurácia e precisão para os casos de teste. Tal constatação indica que a CNN consegue generalizar melhor e identificar características das imagens com maior precisão, o que resulta em previsões mais confiáveis. A exemplo de comparação, têm-se a mesma imagem classificada pelos dois modelos.



Figura 5. CNN

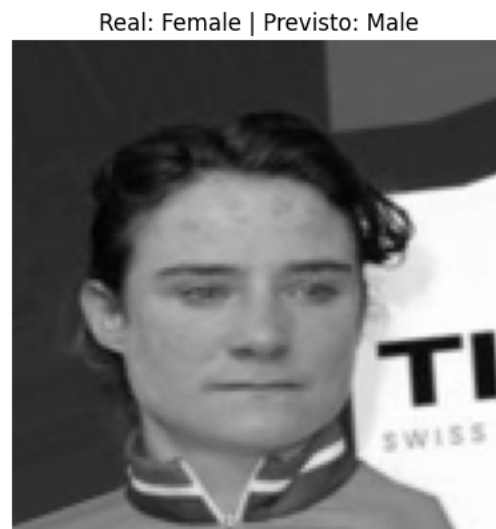


Figura 6. MLP

É possível observar que a CNN conseguiu fazer a previsão correta, enquanto a MLP se equivocou no resultado apresentado, situação que ocorreu com maior frequência nesse último modelo. Essa diferença de desempenho reflete-se nas métricas, que indicaram uma superioridade da CNN para essa tarefa específica de classificação.

5. Modelos aplicados no jogo da memória

A importância das redes neurais em diversas aplicações foi amplamente discutida ao longo deste documento, demonstrando o potencial desse método de resolver problemas complexos. Um contexto particularmente curioso para a aplicação de técnicas de aprendizado

de máquina é o dos jogos. Nesse sentido, visando tornar o tema mais interessante e envolvente, o problema da classificação de imagens foi integrado em um jogo da memória, demonstrando como as redes neurais conseguem aprimorar experiências lúdicas.

Na versão implementada, inicialmente, as imagens das cartas são analisadas por um dos modelos, a fim de determinar a classe a qual cada uma pertence. O computador desempenha o papel do jogador, virando duas cartas simultaneamente e, caso elas pertençam a uma mesma classe, as cartas permanecem viradas; caso contrário, elas são viradas para baixo novamente e novas cartas são escolhidas de modo aleatório. Além disso, se o modelo fizer uma previsão incorreta com base nos rótulos pré-definidos das imagens, resultando em uma combinação equivocada, o jogo é perdido; por outro lado, se todos os pares tiverem sido identificados corretamente, o computador vence a partida. Tal abordagem conseguiu demonstrar, de forma prática e divertida, a eficácia de redes neurais na identificação de padrões, embora ainda apresentem limitações e suscetibilidade a erros.

6. Conclusão

O experimento desenvolvido ao longo desse documento teve como objetivo fazer uma análise comparativa do desempenho de dois modelos de redes neurais na classificação de imagens em diferentes classes de gênero. Para se chegar a esses modelos finais, foram necessários muitos testes para selecionar os melhores hiperparâmetros e otimizar o retorno final. Nesse sentido, os resultados obtidos evidenciaram uma clara superioridade da Rede Neural Convolucional (CNN) em relação ao Perceptron Multicamadas (MLP) nessa tarefa específica, como demonstrado pelas métricas de avaliação. Tal superioridade está relacionada à capacidade da CNN de aprender representações espaciais complexas, se aproveitando da hierarquia de características presente numa imagem para fazer previsões.

Entretanto, é importante salientar que, embora a CNN tenha se mostrado mais eficaz neste contexto particular, isso não significa que esse padrão se mantenha em todas as situações. Sabe-se que a eficácia de um modelo está intimamente atrelada à complexidade e natureza dos dados, objetivo final da tarefa e arquitetura implementada pelo modelo. Portanto, para que seja possível definir qual a melhor abordagem de rede neural para um problema, é de suma importância analisar o contexto envolvido e compreender as especificidades da tarefa em questão.

Referências

- H. Zhang, Z. Dong, B. L. S. H. (2022). Multi-scale mlp-mixer for image classification. *Knowledge-based Systems*, 258.
- N. Sharma, V. Jain, A. M. (2018). *An Analysis Of Convolutional Neural Network for Image Classification*. International Conference on Computational Intelligence and Data Science (ICCIDS).
- Tolstikhin, I. O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., Lucic, M., and Dosovitskiy, A. (2021). Mlp-mixer: An all-mlp architecture for vision. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 24261–24272. Curran Associates, Inc.