

Redes Neurais Convolucionais aplicada no reconhecimento de LIBRAS

Elaine Figueiredo, Lucas Alves Lira Pauletti, Lucas Lacerda Bernardo, Lucas Waiteman Bastos e Thauan da Trindade Moreno da Silva

¹ Centro Universitário Campo Limpo Paulista (UNIFACCAMP)

² Universidade Anhembi Morumbi (UAM)

ebfigueiredo@anhembi.br; luucasalves.p@gmail.com; lucaaslb@hotmail.com; lucas060994@gmail.com; thauantrindade@outlook.com

Resumo:

A Língua Brasileira de Sinais é um sistema linguístico legítimo e natural, utilizado pela comunidade surda brasileira, de modalidade gestual-visual e com estrutura gramatical independente da Língua portuguesa falada no Brasil (Azeredo, 2006).

Redes Neurais Convolucionais (Convolutional Neural Network - CNN) é uma classe de redes neurais artificiais multicamadas (Perceptron Multicamadas), consiste numa rede com uma ou mais camadas ocultas entre a camada de entrada e a camada de saída. Para análise e classificação de imagens em uma Perceptron Multicamada a entrada pode contar uma dimensão muito alta, por que a arquitetura da rede não leva em consideração a estrutura espacial, ela trata os pixels de entrada no mesmo nível independente da distância entre eles. Uma CNN usa conceito de campo receptivo local com base no funcionamento do córtex visual biológico (Hubel, D. H. et al Wiesel, 1962).

Para treinamento criou-se um script para transformar o dataset de imagens em um arquivo csv. Cada linha no arquivo corresponde a um vetor, seu primeiro valor representa a classe, ou seja, a letra que será treinada, e as demais informações são os pixels da imagem, exceto pela primeira linha que deve conter o cabeçalho. Ao realizar a leitura da imagem para preencher o arquivo csv, as imagens passaram por um redimensionamento, do qual foi alterado o número de canais, passando de 3 (RGB) para 1 em preto e branco, ficando com a dimensão 64x64x1. A Rede foi implementada em python utilizando a biblioteca Keras. A rede que melhor teve resultados foi a CNN com 04 camadas convolucionais.

A função ReLU é definida como $f(x) = \max(0, x)$, onde x é o valor do neurônio. Ao aplicar essa função tem-se um pequeno problema de desativação precoce de neurônios. A ReLU pode ser corrigida com a função **Leaky ReLU**, onde temos $f(x) = \max(\alpha x, x)$, onde x é o valor do neurônio e α uma constante de valor baixo, normalmente entre 0,01 e 0,2. A derivada da função na região negativa ainda é positiva diferentemente da ReLU e assim é evitado desligamentos precipitados de neurônios, sendo possível obter melhores resultados e mais características da imagem analisadas nas demais camadas. Essa alteração trouxe uma melhoria no treinamento da rede em comparação com valores de perda, no entanto decidiu-se manter o uso da **LReLU**.

Estrutura criada, parâmetros treinamento definidos. A rede tem 500 épocas, foi utilizada a função *Early Stopping*, com objetivo observar a partir de uma época definida se o valor de perda se mantém estável para finalizar o treinamento de forma antecipada, evitando treinar excessivamente o modelo antes que comece a aprender alguma regularidade aleatória contida no dataset, problema conhecido como **overfitting**, ou que o modelo seja incapaz de capturar a variabilidade dos dados (JABBAR, KHAN, 2014), chamado de **underfitting**.

O arquivo csv foi importado e dividido em duas partes, treinamento e validação. Para melhorar o processo de treinamento durante as épocas foi utilizado o método de validação cruzada com o parâmetro *validation split*, foi definido, 20% das imagens seriam utilizadas para validar o modelo.

Ao todo foram realizados quatro (4) testes alterando a estrutura da rede, treinados inicialmente com uma camada convolucional, chegando até quatro camadas do modelo final.

No geral as quatro estruturas apresentaram alta acurácia e baixa perda na fase de treinamento, mas ao inserir novas imagens, que não fizeram parte do treinamento, a diferença de acurácia entre os modelos ficou mais evidente. Nos testes em real time o modelo não obteve a mesma acurácia do que nos testes controlados com o dataset criado.

Nos resultados obtidos observou-se pontos de melhoria. Ocorreu uma generalização durante o processo de reconhecimento devido ao dataset estar muito padronizado, sendo necessário ampliar e diversificar ambientes e luminosidade.

É sabido que muitas palavras em LIBRAS possuem movimentos e expressões faciais, dessa forma o modelo aqui proposto deve ser aperfeiçoado para análise de imagens com movimento e reconhecimento de palavras, termos e expressões por meio da aplicação novas técnicas para obtenção dos resultados desejados.

Referências

- AS 10 PRINCIPAIS ARQUITETURAS DE REDES NEURAIS. DEEP LEARNING BOOK. Disponível em: <<http://deeplearningbook.com.br/as-10-principais-arquiteturas-de-redes-neurais>>. Acesso em: 20 mar. 2019.
- AZEREDO, E. Língua Brasileira de Sinais: uma conquista histórica. Brasília: Ática, 2006. GIBSON, A.;
- PATTERSON, J. Deep learning: a practitioner's approach. O'reilly, 2017.
- HUBEL, D. H.; WIESEL, T. N. Receptive fields and functional architecture of monkey striate cortex: vol. 195, no. 1, p. 215–243: The Journal of Physiology, 1968.
- JABBAR, H. K.; KHAN, R. Z. Methods to Avoid Over-Fitting and Under-Fitting in Supervised Machine Learning (Comparative Study). Computer Science, Communication and Instrumentation Devices, 2014.
- KARPATY, A., JOHNSON, J., CLASS CS231n CONVOLUTIONAL NEURAL NETWORKS FOR VISUAL RECOGNITION. UNIVERSIDADE DE STANFORD. Disponível em: <<http://cs231n.github.io/convolutional-networks>>. Acesso em: 05 abr. 2019.
- LECUN, Y.; BOTTOU, L.; BENGIO, Y. Gradient-based learning applied to document recognition: vol. 86, no. 11, p. 2278–2324: Proceedings of the IEEE, 1998.
- LECUN, Y.; CORTES, C.; B. C.; THE MNIST DATABASE OF HANDWRITTEN DIGITS. Disponível em: <<http://yann.lecun.com/exdb/mnist>>. Acesso em: 12 abr. 2019
- LOURENÇO, K. R. C.; BARANI, E. Educação e Surdez: Um Resgate Histórico pela Trajetória Educacional dos Surdos no Brasil e no Mundo. Revista Virtual de Cultura Surda e Diversidade: p.03-24, 2011
- PIZZOLATO, E. B.; ANJO, M. S.; PEDROSO, G. C. Automatic recognition of finger spelling for LIBRAS based on a two-layer architecture: Proceedings of the 2010 ACM Symposium on Applied Computing - SAC '10, 2010.
- QUADROS, R. M. Educação de surdos: Porto Alegre: Artes Médicas, 1997.
- SAMER, C. H.; RISHI, K.; ROWEN. Image Recognition Using Convolutional Neural Networks: p.1-12: Cadence Whitepaper, 2015.
- STUART, J. R.; Peter N. Artificial intelligence: a modern approach, Prentice-Hall, Inc: Upper Saddle River: NJ, 1995.
- SZEGEDY, C.; Inception-resnet and the impact of residual connections on learning: Computing Research Repository, 2016.
- SZEGEDY, C.; LIU, W.; JIA, Y. Going deeper with convolutions: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
- VARGAS, A. C.; CARVALHO, Aline M. P.; VASCONCELOS, C. N. Um estudo sobre Redes Neurais Convolucionais e sua aplicação em detecção de pedestres, p.1-4.