



Redes Neurais Profundas: Implementação no Contexto de Reconhecimento de Imagens

Pedro R. A S. Bassi*, Romis Attux.

Resumo

Este trabalho pertence à temática de interfaces cérebro-computador. Nele, utilizamos redes neurais profundas do tipo Triplet para analisar sinais de eletroencefalografia (EEG), os quais haviam sido convertidos em imagens através de transformadas de Fourier. Os sinais foram gerados em um estudo de SSVEP, onde pessoas observam um estímulo piscando e têm sua atividade cerebral monitorada. O objetivo das redes neurais é, a partir das imagens geradas (espectrogramas dos sinais), dizer que frequência estimulava visualmente o observador. Foram obtidos resultados promissores, apesar de mais estudos serem necessários para analisar a perspectiva de aproveitamento adicional das relações não-lineares presentes nos dados.

Palavras-chave: Deep Learning, Brain Computer Interface, Redes Triplet.

Introdução

Neste estudo, propusemos uma nova forma de classificar sinais cerebrais obtidos por eletroencefalografia no âmbito de interfaces cérebro-computador baseadas em potenciais visualmente evocados em estado estacionário (SSVEP, do inglês *steady state visually evoked potentials*).

Nossa abordagem se baseou em transformar os sinais obtidos em espectrogramas, por meio de transformada de Fourier de curto termo (STFT, do inglês *short-term Fourier transform*) para em seguida utilizar uma rede neural artificial estado-da-arte em classificação de imagens para resolver o problema de classificação.

Para isso, foram escolhidas as redes do tipo Triplet (propostas em [1]). Elas se baseiam em uma arquitetura, geralmente profunda e convolucional, a qual realiza um mapeamento não linear das imagens de entrada em um espaço Euclidiano de pequena dimensão, no qual imagens de mesma classe são mapeadas de forma a ficarem próximas e de classes diferentes a ficarem distantes. A rede minimiza uma função chamada de *triplet loss* para aprender a fazer este mapeamento.

Resultados e Discussão

Neste trabalho, utilizamos dados de EEG pertencentes a uma base construída anteriormente pelo grupo de pesquisa, com dez sujeitos. Nós buscamos sempre utilizar nove sujeitos para treinar a rede e um para testá-la. Desta forma, simulamos o caso realista de se usar a rede em pessoas que nunca tenham tido contato com ela. Classificamos os sinais gerados por estímulos visuais de 12 Hz e 15 Hz, os quais devem ser diferenciados pela rede. Verificou-se também que, para obter melhores resultados, dever-se-ia utilizar apenas o sinal do eletrodo Oz (localizado sobre o córtex visual) para treinar a rede.

Para a criação das imagens que seriam classificadas utilizamos STFT, gerando espectrogramas. Em seguida foi aplicado um janelamento sobre elas, seguido por uma remoção de frequências que não eram interessantes para o estudo. Foram criadas imagens que apenas levavam em conta as frequências em torno de 12 Hz e 15 Hz, e imagens que levavam em conta também frequências em torno das primeiras harmônicas destes valores. A figura 1 ilustra uma imagem gerada, considerando também as primeiras harmônicas, e a estrutura da rede Triplet utilizada.

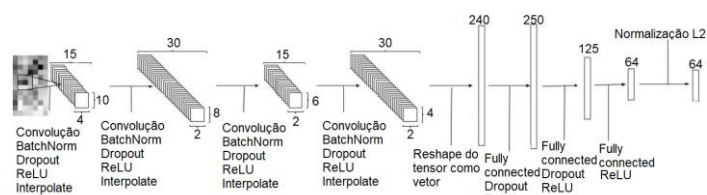


Figura 1. Exemplo de espectrograma (à esquerda) gerado e arquitetura da rede.

Os resultados, em termos de taxa de acerto da classificação (acurácia) no conjunto de testes, estão consolidados na figura 2, onde também são comparados com os resultados de uma Máquina de Vetor-Suporte (SVM, Support-Vector Machine) linear com a mesma entrada (espectrogramas).

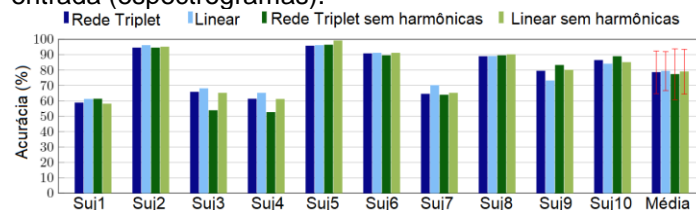


Figura 2. Resultados a rede profunda Triplet e da SVM.

Conclusões

Buscou-se explorar o potencial das redes Triplet, que, até onde pudemos verificar, não haviam sido empregadas no contexto de BCIs baseadas em SSVEP. Procuramos obter um projeto de interface que opere sem a necessidade de treinamento "indivíduo a indivíduo". Pelos resultados obtidos, foi possível constatar que a rede, na implementação atual, é capaz de obter um desempenho que pode ser considerado bom (mais de 70% de acurácia para a maior parte dos indivíduos). Por outro lado, a proximidade de desempenho com respeito a uma estrutura linear requer novas investigações no sentido de avaliar se mudanças na arquitetura e/ou no treinamento podem levar a uma exploração mais ampla da informação subjacente aos dados.

Agradecimentos

Este trabalho foi financiado pelo PIBIC/SAE/UNICAMP.

¹ F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 815-823, 2015.