

INTERPRETAÇÃO DE SINAIS DA LÍNGUA BRASILEIRA DE SINAIS COM MOVIMENTO E EXPRESSÕES FACIAIS UTILIZANDO REDES NEURAIS ARTIFICIAIS

MATHEUS L. RIBEIRO¹, LUCIANA B. R. DOS SANTOS², LUCAS V. POVOA³

¹ Graduando em Tecnologia em Análise e Desenvolvimento de Sistemas, Bolsista PIBIFSP, IFSP, Câmpus Caraguatatuba, matheusladislau.adm@gmail.com.

² Doutora em Computação Aplicada, Orientadora, Professora, IFSP, Câmpus Caraguatatuba, lurebello@ifsp.edu.br.

³ Mestre em Ciência da Computação, Colaborador, Professor, IFSP, Câmpus Caraguatatuba, venezian@ifsp.edu.br.

Área de conhecimento (Tabela CNPq): Computabilidade e Modelos de Computação – 1.03.01.01-1.

RESUMO: A Língua Brasileira de Sinais (LIBRAS) é a segunda língua oficial do Brasil e principal forma de comunicação e expressão utilizada pela comunidade surda/muda no país. Embora seja uma ferramenta para a inclusão social, há uma falta de intérpretes para a língua, diminuindo seu uso. Visto este contexto, o uso de agentes de Tecnologia Assistiva (TA) para a interpretação da LIBRAS permitiria um crescente uso na língua e, principalmente, maior inclusão da comunidade surda/muda do país. Para caracterização de sinais em LIBRAS há cinco parâmetros fonológicos que causam distinção entre os sinais: Configuração da Mão, Orientação, Expressão, Movimento e Ponto de Articulação. O projeto objetiva o estudo e desenvolvimento de modelos de redes neurais que utilizam como dados de entrada os parâmetros da LIBRAS de Movimento e Expressão Facial, gerando um algoritmo base para a interpretação de sinais em LIBRAS para Português brasileiro. O projeto possui como base trabalhos anteriores, desenvolvidos no âmbito do Instituto Federal de Ciência, Educação e Tecnologia de São Paulo, Câmpus Caraguatatuba, desde o ano de 2015.

PALAVRAS-CHAVE: Língua Brasileira de Sinais; Tecnologia Assistiva; Movimento; Expressão Facial;

1 INTRODUÇÃO

Uma pesquisa demográfica (IBGE, 2010) apresenta que o Brasil contém 9.722.163 deficientes auditivos, sendo 347.481 totalmente surdos. Para a comunidade surda brasileira é fundamental o domínio da Língua Brasileira de Sinais (LIBRAS), sendo a segunda língua oficial do Brasil, conforme Lei no 10.436, de 24 de abril de 2002 que dispõe sobre a língua de sinais brasileira (BRASIL, 2002).

Apesar disso, a LIBRAS é pouco conhecida entre os cidadãos que não estudam ou trabalham com a língua de sinais (GESSER, 2009), dificultando a inclusão de pessoas com deficiência auditiva. Partindo deste contexto, é imprescindível o uso de ferramentas de tecnologias assistivas que permitam maior inclusão social.

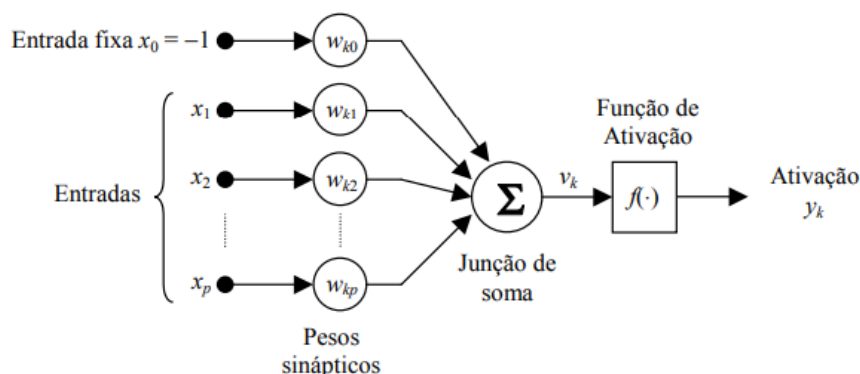
O projeto possui como objetivo geral o estudo e desenvolvimento de redes neurais artificiais que utilizem como dados de entradas os parâmetros da LIBRAS de movimentação e expressão facial para gerar interpretação do sinal para Português brasileiro, tendo como base iniciações científicas e trabalhos anteriores referente ao projeto DeepLIBRAS (DEEPLIBRAS, 2017).

O presente artigo propõe trazer à comunidade parte do projeto desenvolvido em Iniciação Científica do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Câmpus Caraguatatuba.

2 MATERIAL E MÉTODOS

Redes Neurais Artificiais são técnicas de aprendizado de máquina, modeladas de forma semelhante ao cérebro humano e capazes de adquirir conhecimento por meio de entradas e troca de informações por neurônios artificiais. O neurônio é a principal unidade de processamento de dados de uma rede neural (HAYKIN, 1999). O chamado treinamento de máquina é o conjunto de entradas utilizado pela RNA para identificação de padrões. Essa representação pode ser visualizada na Figura 1.

FIGURA 1. Representação da estrutura de neurônio artificial.



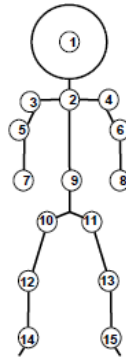
Fonte: LYODA (2000).

Para programação, foi utilizada a linguagem de programação Python (PYTHON, 2018) e o *framework* Keras (KERAS, 2015), para desenvolvimento das RNA. Python é uma linguagem de programação de alto nível, visa a produtividade e legibilidade do código, sendo considerada de fácil aprendizado. Keras é uma API (*Application Programming Interface*) de alto nível, voltada ao desenvolvimento de redes neurais. Foi desenvolvida na linguagem Python, com foco na rápida implementação.

Para treinamento das redes neurais é utilizada uma base de dados gerada em trabalhos anteriores, relativos ao projeto DeepLIBRAS. A base de dados possui diversos vídeos de execução de sinais da LIBRAS. A coleta destas imagens foi realizada por meio do dispositivo *Kinect*. O dispositivo foi desenvolvido pela empresa Microsoft para facilitar a interação entre jogador e console e garantir jogabilidade sem o uso de controles manuais. Conta com um sensor de profundidade, uma câmera de cor, e quatro microfones que provêem captura de movimento 3D de corpo completo, reconhecimento facial e capacidade de reconhecimento de voz (ZHANG, 2012).

Para utilização do parâmetro de movimento da LIBRAS cada imagem do corpo humano sofreu uma segmentação em quinze pontos principais, conforme apresentado na Figura 2. Destes quinze pontos, nove foram considerados os mais importantes para distinção do sinal em LIBRAS, sendo eles: cabeça, entre os ombros, ombros, cotovelos, mãos e tronco, representados pelos números de um a nove.

FIGURA 2. Representação de fragmentação do corpo humano em quinze pontos.



Fonte: LIMA (2018).

Foram valorados os eixos x, y e z de cada um dos pontos do corpo humano, para que fosse realizado o cálculo da distância euclidiana entre os nove pontos da imagem pela sequência até o final da execução do sinal, o resultado foi utilizado como entrada na RNA.

O cálculo da distância euclidiana é definido pela seguinte equação:

$$D(x, y, z) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

em que,

D – distância euclidiana;

X – valor do eixo x do ponto;

Y – valoração do eixo y do ponto;

Z – valoração do eixo z do ponto;

Para treinamento do segundo parâmetro, expressão, foram utilizadas como entradas imagens de expressões faciais retiradas de bases open-source, tais como AKDEF, KDEF e CK+ (KDEF, 1998). A Figura 3 apresenta uma amostra de imagens do conjunto AKDEF. Foi realizada a categorização das imagens com base nas seis emoções principais, alegria, tristeza, medo, repulsa, surpresa e raiva, acrescido de neutro, sem expressão.

FIGURA 3. Exemplo de amostra de imagens da base de dados AKDEF.



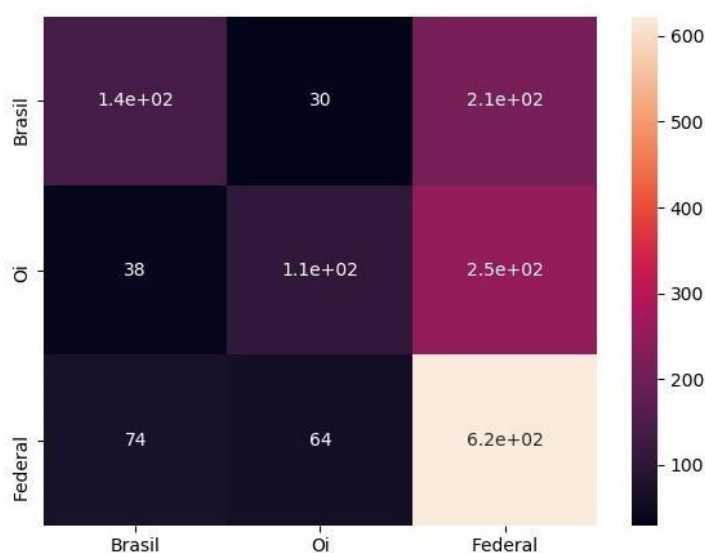
Fonte: KDEF (1998).

3 RESULTADOS E DISCUSSÃO

Desenvolvido de maneira modular, foram utilizados diferentes tipos de redes neurais para interpretação do sinal em LIBRAS, as quais estão atuando de maneira independente. Para o parâmetro de movimento, foi desenvolvida uma Rede Neural Perceptron Multicamadas e, para o parâmetro de expressão facial, uma rede neural convolucional. Em acurácia, ambas estão atingindo acima de 90% de precisão, nos melhores casos, informação esta indicada pelo *framework* Keras.

Um dos métodos para validação do modelo é a utilização da matriz de confusão, a qual gera previsões, apresentando seus erros e acertos. A Figura 4 apresenta a matriz de confusão com entradas dos sinais "Brasil", "oi" e "federal", utilizando o parâmetro de Movimento da LIBRAS. A diagonal principal da matriz apresenta as classificações corretas, enquanto que os outras classificações são incorretas.

FIGURA 4. Matriz de confusão de teste com o modelo que utiliza o parâmetro Movimento da LIBRAS.



Fonte: Autor (2018).

Em 1500 épocas, pode-se observar o acerto com a predição sinal "Brasil" em 140 dos casos, "oi" em 110 e 620 com "federal".

4 CONSIDERAÇÕES FINAIS

Ambos os modelos apresentaram valor alto em acurácia, nos melhores dos casos, no entanto, houve falhas ao realizar classificação de sinais, como apresentado.

Os modelos atuam de forma independente, com entradas separadas para predição do sinal. Um dos possíveis meios de melhoria de acurácia, e um dos objetivos deste trabalho, é o desenvolvimento de um novo modelo, que seja capaz de receber ambas as entradas de movimento e expressão facial, para gerar como saída a classificação do sinal interpretado.

5 AGRADECIMENTOS

Os autores agradecem aos envolvidos com o projeto DeepLIBRAS, que tiveram participação desde 2015 no desenvolvimento dos algoritmos e na coleta de dados. Agradecimentos ao IFSP Câmpus Caraguatatuba, por conceder infraestrutura para desenvolvimento do projeto e apoio financeiro, por meio da bolsa PIBIFSP.

6 REFERÊNCIAS

BRASIL. Decreto nº 5626. <http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2005/decreto/d5626.htm>. Acesso em: 10 de agosto de 2018.

DEEPLIBRAS. DeepLIBRAS - Tradução da Língua Brasileira de Sinais com Técnicas de Aprendizagem de Máquina. 2017. Disponível em: <<http://deeplibras.github.io/pt/>>. Acesso em: 31 jul. 2018.

GESSER, A. LIBRAS, Que Língua É Essa?. São Paulo: Parábola Editorial, 2009.

IBGE. Censo demográfico: características gerais da população, religião e pessoas com deficiência. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2010.

KERAS. Keras Documentation. Disponível em: <<https://keras.io/>>. Acesso em: 12 de junho de 2018.

KDEF. KDEF & AKDEF. Disponível em: <<http://kdef.se/>> Acesso em: 05 de agosto de 2018.

LIMA, B.C.R. Interpretação de Sinais Com Movimento da Língua Brasileira de Sinais Utilizando Redes Neurais Artificiais, 2018.

LYODA, E.M. Inteligência Computacional no Projeto Automático de Redes Neurais Híbridas e Redes Neurofuzzy Heterogêneas, 2000.

PYTHON. About Python™ | Python.org. Disponível em: <<https://www.python.org/about/>>. Acesso em: 12 de junho de 2018.

ZHANG Z., “Microsoft Kinect Sensor and Its Effect”, 2012.