Trabalho apresentado no CNMAC, Gramado - RS, 2016.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Aplicação de Redes Neurais para Reconhecimento de Comandos de Voz

Paulo Hitoshi Sassaki[1]
Tecnologia em Sistemas Biomédicos, FATEC, Bauru, SP.
Leonardo Vinicius Cavacchioli[2]
Engenharia de Automação de Sistemas Elétricos, INATEL, São Paulo, SP
Mauricio Eiji Nakai[3]
Departamento de Engenharia Elétrica e de Computação, USP, São Carlos, SP

1 Introdução

O controle de equipamentos automatizados ou robôs pode auxiliar pessoas portadoras de deficiência física. Pensando em pessoas com dificuldades de locomoção, decidiu-se analisar a aplicação das RNA (Redes Neurais Artificiais) para comandos de voz que possam ser utilizadas em cadeiras de rodas através de um sistema microcontrolado. Para isso optou-se por um sistema mais simples do que os utilizados atualmente no reconhecimento de voz. Utilizou-se para o treinamento o algoritmo Levenberg-Marquardt (LM), devido às suas propriedades de convergência.

Com o aumento da capacidade de processamento e a diminuição do custo computacional o reconhecimento de voz vem sendo largamente pesquisado e utilizado nas últimas décadas. As principais linhas de pesquisa são: Modelos Ocultos de Markov [1], Dynamic Time Warping [2] e Redes Neurais Artificiais [3].

2 Metodologia, Resultados e Discussão

Para a realização do ensaio utilizou-se o *Neural Network Toolbox* –MATLAB. Utilizou-se um microfone de eletreto como transdutor, acoplado à entrada de áudio de um computador. A base de dados é formada por cinco conjuntos de palavras contendo 100 amostras cada. As palavras utilizadas como comandos de voz foram “pare”, “ré”, “esquerda”, “direita” e “devagar”.

Utilizou-se o RMS (*Real Mean Square*) para o processamento dos dados pela sua simplicidade computacional. Para o cálculo do RMS as amostras foram divididas em três janelas de tempo com igual duração. Com o MATLAB, calculou-se o RMS de cada janela

[1]hitoshi_paulo@hotmail.com
[2]leonardoc@cpfl.com.br
[3]nakaimauricio@usp.br

© 2017 SBMAC

obtendo assim três valores RMS por palavra, para depois serem concatenadas e inseridas como entradas na RNA.

Para a análise dos resultados utilizou-se a Matriz de Confusão. Não foram realizados tratamentos para redução ou eliminação dos ruídos, a aquisição das amostras se deu em um ambiente controlado, mas sem tratamento acústico.

A estrutura da rede é de múltiplas camadas composta por três camadas, sendo uma camada de entrada com três neurônios, uma camada oculta com dez neurônios e uma camada de saída com cinco neurônios. A função de ativação utilizada é a do tipo sigmoide por não ter parâmetros fixos, mais adequado aos problemas não lineares.

A utilização da RNA dos comandos de voz para cadeira de rodas mostrou desempenho significativo: esquerda (100%), direita (82%), ré (57%), devagar (93%) e pare (99%). A percentagem de acertos obtidos pela palavra "ré" deu-se devido ao seu tamanho reduzido de amostras, portanto menor variação entres as partes segmentadas.

3 Conclusões

A relação entre a simplicidade no processamento computacional e os resultados obtidos foi muito boa, por não necessitar de filtros de entrada e considerando que o cálculo RMS não é computacionalmente custoso. Devido ao tipo de treinamento, os pesos dos neurônios que constituem a RNA podem ser obtidos previamente, sem a necessidade da implementação do algoritmo de treinamento no sistema embarcado. Indicando que este sistema é funcional e pode ser utilizado em um sistema embarcado de uma cadeira de rodas.

A RNA apresentou desempenho significativo na classificação dos comandos de voz, mesmo sem a utilização de algoritmos mais complexos tais como *Dynamic Time Warping* e Modelos Ocultos de Markov.

Para futuras investigações e pesquisas, tem-se como objetivo a implementação de redes diferentes tais como redes neurais com funções de base radial assim como algoritmos de aprendizado mais complexos tais como *deep learning* ou até mesmo outros sistemas inteligentes tais como algoritmos genéticos.

Referências

[1] H. Farsi e R. Saleh, Implementation and optimization of a recognition system based on hidden Markov model using genetic algorithm, *Iranian Conference on Intelligent Systems*, 2014.

[2] X. Zhang, J. Sun, Z. Luo e M. Li, Confidence index dynamic time warping for language-independent embedded speech recognition, *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013.

[3] V. Mitra, G. Sivaraman, H. Nam, C. Espy-Wilson e E. Saltzman, Articulatory features from deep learning neural networks and their role in speech recognition, *IEEE International Conference on Acoustic, Speech and Signal Processing*, 2014.

© 2017 SBMAC