

PROCESSAMENTO DE SINAIS DE VOZ NA IDENTIFICAÇÃO DE PADRÕES COMPORTAMENTAIS POR REDES NEURAI ARTIFICIAIS

Gustavo Ravanhani Matuck¹ (UNIVAP, Bolsista PIBIC/CNPq)
Dr. José Demisio Simões da Silva² (LAC/INPE)

RESUMO

Este trabalho é uma continuação da pesquisa iniciada em 2004 (Matuck, 2004) para utilizar sistemas classificadores baseados em Redes Neurais Artificiais, no reconhecimento de padrões de voz, com o objetivo de identificar alterações emocionais em falantes. Foram realizadas várias gravações de vozes em situações consideradas de estresse e normais, diretamente de filmes, noticiários, novelas, entre outros.

No trabalho inicial os dados foram diretamente fornecidos para a rede neural durante o treinamento, com o objetivo de fazer a rede descobrir as características embutidas nos dados, a serem utilizados como possíveis padrões para distinguir situações de estresse de situações normais. Os testes de generalização realizados mostraram a adequação do uso de redes neurais para esta finalidade, entretanto, o desempenho da rede neural não foi muito promissor, restringindo-se a uma taxa de reconhecimento da ordem de 52%. Diversas tentativas de arquitetura de redes neurais foram testadas, mas sem implicar em melhorias significativas que pudessem indicar a rede neural como um bom reconhecedor para o problema. Possivelmente o baixo desempenho da rede neural esteja associado à presença de fatores externos (ruído, chiados, etc.) de alta frequência nos dados, oriundos do processo de aquisição.

Visando o aumento do desempenho da rede neural, neste trabalho utilizou-se um pré-processamento dos dados através da aplicação da transformada de Fourier com o intuito de realizar uma suavização dos dados, através da seleção de frequências não associadas com o ruído. Assim, durante a realização da pesquisa foi necessário uma familiarização e implementação dos códigos da transformada de Fourier para permitir um entendimento completo da transformada. Feita a transformação, os dados foram utilizados para treinar a rede e posteriormente foram realizados os testes de generalização. Diversas arquiteturas de rede foram testadas, resultando numa capacidade reconhecimento da rede em torno de 72% de acerto. Deve-se enfatizar também um aumento significativo do desempenho da rede, com cerca de 95% mais rápida em relação ao desempenho conseguido no trabalho anterior (Matuck, 2004).

A rede neural utilizada foi um Perceptron de Múltiplas Camadas com aprendizagem por retro-propagação do erro (error backpropagation). A justificativa para uso desta rede está no fato dela ser a mais referenciada na literatura, para tarefas de reconhecimento de padrões, e por o trabalho exigir uma rede com aprendizagem supervisionada.

Os resultados apesar de melhores, ainda não indicam um ótimo. Muita pesquisa por métodos de pré-processamento pode ser realizada que possivelmente poderá significar em um aumento no desempenho da rede. Neste sentido, é sugerido o uso de outras técnicas de transformação, como as Wavelets, que podem extrair ruídos dos dados sem alterar as características principais dos dados.

¹ Aluno do Curso de Ciência da Computação, UNIVAP. E-mail: gustavomatuck@yahoo.com.br

² Pesquisador na área de Inteligência Artificial e Computação Aplicada. E-mail: demisio@lac.inpe.br