

RECONHECIMENTO AUTOMÁTICO DE LOCUTOR UTILIZANDO REDES NEURAIS ARTIFICIAIS

Autor

Rodrigo de Toledo Caropreso

Resumo— *Reconhecimento de padrões é um problema bastante estudado e abordado de diversas formas. Particularmente, o reconhecimento de fala mostra-se como uma área onde há ainda muito a ser explorado.* As redes neurais são uma técnica que auxiliam na tarefa de reconhecimento de padrões, dada suas características e vantagens frente à natureza não-estruturada dos padrões. Principalmente nos padrões de fala, onde uma estrutura exata é difícil de ser definida face a grande variabilidade existente, provocada por entonação, timbre, sotaque, etc. Visando diminuir esta variabilidade, algumas restrições são impostas às aplicações, procurando torná-las menos complexas.

Palavras-chave— Redes Neurais, locutor, reconhecimento, padrões, perceptron, cepstrum.

1 - Introdução

O reconhecimento da voz humana e de um interlocutor tem sido objeto de estudo cada vez mais frequente na atualidade [CORSI, 81],[JACOBS,88].

O Reconhecimento Automático de Locutores (RAL) tem despertado, durante as últimas décadas um profundo interesse por parte da comunidade científica e, por conta disso, tem sido fruto de amplas pesquisas.

Como resultado destas pesquisas foi possível elaborar procedimentos e algoritmos capazes de processar e efetuar o RAL. Entretanto, apesar dos avanços obtidos pelas pesquisas, ainda não foi possível obter uma máquina capaz de compreender toda a fala de um ser humano.

Ainda assim, o avanço da tecnologia, dos computadores e das telecomunicações tem servido como fator de ampliação das pesquisas na área do RAL e suas possibilidades: aplicações de automação bancária, controle de acesso à informações, sistemas de segurança e outras [HATON, 81].

De uma forma mais simples, podemos definir o RAL como sendo a capacidade de reconhecer uma determinada pessoa através da sua voz, sendo uma variação do Reconhecimento de Voz, outro assunto amplamente estudado por diversos pesquisadores [PICONE, 93], [FURUI, 89].

Entretanto, no estudo do Reconhecimento de Voz as variações referentes aos locutores não são consideradas, enquanto que no RAL toda e qualquer variação intra-locutor é amplamente levada em consideração para as análises.

Alguns estudos abrangentes sobre o RAL podem

ser vistos em [ATAL, 76], [CORSI, 81] e [GISH, 94].

Em paralelo à estas pesquisas, os estudos na área de Inteligência Artificial e Redes Neurais Artificiais (RNA) têm se intensificado e novas técnicas vêm sendo introduzidas pela comunidade científica ao longo dos últimos anos. a exemplo de [SAMBUR, 75] e [ATAL, 76].

Diversos trabalhos sobre as Redes Neurais Artificiais [CASAGRANDE, 97],[ELMAN, 90], [JORDAN,86], [MAGNI, 98], [TMOSZCZUCK,98] mostram a viabilidade do uso de classificadores neurais. Os Classificadores do tipo Multi-Layer Perceptron (MLP) têm sido amplamente utilizados em problemas de identificação do locutor, devido à sua simplicidade e eficiência.

Por este motivo a rede MLP será usada como modelo de classificador padrão no estudo deste trabalho e seu desempenho servirá como balizador para comparação com as demais redes.

A proposta deste trabalho é avaliar a o desempenho de redes neurais artificiais no processo de reconhecimento automático de locutor.

2 – Descrição do problema

O problema de reconhecimento de fala é de difícil tratabilidade. A maior dificuldade é a sua natureza interdisciplinar. Além dessa, variabilidades acústicas, do transdutor, intra-locutor, entre locutores estão relacionadas com o problema. Mas, em contra partida, várias áreas podem ser beneficiadas com o uso desta técnica, tornando um campo desafiador para ser pesquisado.

O problema a ser abordado no trabalho é:

- qual é a viabilidade, em termos de desempenho ao se utilizar paradigmas de redes neurais artificiais no processo de reconhecimento e identificação de um locutor?

2.2 – Objetivos Gerais

- Apresentar as redes neurais utilizadas no trabalho;
- pretende-se utilizar técnicas de pré-processamento e extração de características do sinal da fala através do modelo cepstrum [41]. apresentando uma descrição sucinta de técnicas de pré-ênfase até os métodos de extração de coeficientes mel-cepstrais e energia.
- avaliar se este modelo pode ser usado de forma consistente para identificar corretamente o locutor, com uma taxa aceitável de acerto.

2.3 – Objetivos Específicos

- descrever o algoritmo de RNA usado;
- descrever o pré-processamento utilizado;
- selecionar as amostras de treinamento e operação da rede;
- criar os modelos computacionais e realizar as simulações (treinamento/operação);
- apresentar os resultados das simulações e a análise crítica da proposta.

2.4 – Metodologia adotada

A base de dados de voz utilizada neste trabalho será a Speaker Recognition v1.0, desenvolvida pelo *Center of Spoken Language Understanding (CSLU) do Oregon Graduate Institute (OGI)*, e consiste na gravação de amostras de voz de 90 participantes ao telefone.

Cada participante gravou sua voz em doze sessões ao longo de um período de vários meses. Por se tratar de uma base americana, as frases foram construídas no idioma inglês.

O desenvolvimento do problema irá consistir na seleção das amostras, treinamento e operação da rede neural escolhida, com a respectiva análise dos resultados em termos de desempenho no reconhecimento.

Referências

[ATAL, 76] - ATAL, B.S. Automatic recognition of speakers from their voices. *Proceedings of the IEEE*, v.64, n.4, p.460-75, Apr. 1976.

[CASAGRANDE, 97] - CASAGRANDE, R. Redes neurais artificiais com retardos temporais aplicadas ao reconhecimento automático do locutor. São Paulo, 1997, 80p. Dissertação de Mestrado, EPUSP.

[CORSI, 81] - CORSI, P. Speaker recognition: a survey. In: *Proceedings of the NATO Advanced Study Institute, Bonas, 1981. Automatic Speech Analysis and Recognition*. Dordrecht, D. Reidel Publishing company, 1982. pp.277-308.

[ELMAN, 90] - ELMAN, J. Finding Structure in Time. *Cognitive Science* 14, pp. 179-211, 1990.

[FURUI, 89] - S. Furui. *Digital Speech Processing, Synthesis and Recognition*. Mareei Decker Inc., NY. 1989.

[GISH, 94] - GISH, H.; SCHMIDT, M. Text-independent speaker identification. *IEEE Signal Processing Magazine*, v.1, n.4, p. 18-32, Oct. 1994.

[HATON, 81] - HATON, J.P. Automatic Speech Analysis and Recognition. *Proceedings of the NATO, France, 1981*.

[JACOBS, 88] - JACOBS, RA. Increased rates of convergence through learning rate adaptation. *Neural Networks*, vol.I, pp.295-307, 1988.

[JORDAN.86] - JORDAN, M Serial Order: A Parallel Distributed Approach. *Institute for Cognitive Science Report 8604, University of Califórnia. San Diego, 1986*.

[MAGNI, 98] - MAGNI, A.B. Reconhecimento automático do locutor com coeficientes Mel-Cepstrais e redes neurais artificiais. São Paulo, 138p., Dissertação de Mestrado, EPUSP, 1998.

[PICONE, 93] - PICONE, J.W. Signal modeling techniques in speech recognition. *Proceedings of the IEEE*, v.81, n.9, p.1215-47, Sept. 1993.

[SAMBUR. 75] - SAMBUR, M.R. Selection of acoustic features for speaker identification. *IEEE Transactions on Acoustics, Speech and Signal Processing*, v.23, n.2, p. 176-82, Apr. 1975.

[TMOSZCZUK. 98] - TMOSZCZUK, A.P. Reconhecimento automático do locutor com redes neurais artificiais do tipo radial basis function (RBF) e Minimal Temporal Information (MTI). São Paulo, Dissertação de Mestrado, 135p. EPUSP, 1998.