

Proposta de Trabalho de Graduação

Avaliação da Técnica Type-2 Fuzzy GMM para Verificação de Locutor Independente de Texto em Ambientes Ruidosos

Aluno: Sérgio Renan Ferreira Vieira (srfv@cin.ufpe.br)

Orientador: Tsang Ing Ren (tir@cin.ufpe.br)

Recife, Outubro de 2014

1. Contexto e Motivações

À medida que sistemas computacionais modernos avançam em caminho da ubiquidade, aumenta a necessidade de proteção a dados pessoais intransferíveis para autenticação de usuários. Neste sentido os meios tradicionais como senhas e cartões de acesso apresentam-se inconvenientes, uma vez que podem ser perdidos ou roubados. Sendo assim, sistemas biométricos se tornam uma interessante alternativa para autenticação de usuário já que faz uso de medidas biológicas com características únicas a cada ser humano e, portanto, mais difíceis de serem reproduzidos por impostores - a despeito dos meios tradicionais. Neste cenário a biometria de voz ganha destaque em aplicações de dispositivos móveis podendo-se apontar dois motivos principais. Primeiro, usuários consideram natural fornecer uma autenticação por voz quando já estão falando. Segundo, qualquer dispositivo que disponha de um microfone é capaz de capturar o sinal de voz sem necessidade de transdutores especiais [1].

O problema de autenticação biométrica de voz pode ser endereçado em duas abordagens: **(i) identificação** e; **(ii) verificação** de locutor. No primeiro caso o objetivo é identificar quem está falando dentro de um conjunto de locutores cadastrados. Já na segunda abordagem, o objetivo é verificar se um locutor é ou não quem afirma ser [2]. Em aplicações comerciais observa-se a preferência pela abordagem de verificação de locutor [3]. A tarefa de verificação pode ser classificada com respeito ao discurso do locutor da seguinte maneira: **(i) texto-dependente** quando o locutor deve falar um texto estabelecido ou; **(ii) texto-independente** se o usuário estiver livre para discursar o que quiser. A verificação de locutor texto-independente é preferível por não impor, normalmente, condições incômodas ao usuário [1]. Diante deste cenário pesquisas foram desenvolvidas para diferentes abordagens de aplicações em *Verificação de Locutor Texto-Independente (VLTi)*.

Nos últimos anos o framework *Gaussian Mixture Model (GMM)* se tornou a abordagem dominante [4]. Este framework se apresenta interessante pelo poder de criar um modelo estatístico paramétrico (o GMM) para qualquer densidade de probabilidades de observações acústicas de um locutor. Isto é, ele é capaz de modelar um locutor determinando-se parâmetros GMM. Os parâmetros de um GMM, por sua vez, podem ser obtidos via um algoritmo a partir de locuções de treinamento. De modo a discriminar um locutor de outros, é necessário também criar um GMM, conhecido como *Universal Background Model (UBM)*, que represente todos os outros locutores, que não o locutor em questão. Em [4], Reynolds *et al.* propôs uma abordagem chamada **GMM-UBM** onde o GMM de um locutor é determinado a partir de uma adaptação dos parâmetros do UBM de modo a maximizar o poder discriminante entre o GMM e o UBM.

No entanto, em problemas práticos podemos ter locuções insuficientes ou ruidosas que comprometam o GMM por produzir parâmetros com valores incertos. Zeng *et al.* propôs em [5] o uso do framework *Type-2 Fuzzy* em conjunto com o GMM (**T2 FGMM**) de modo a estimar as

incertezas dos parâmetros de um GMM. Não obstante, Pinheiro *et al.*, em [6], incorporou a abordagem T2 FGMM para o problema VLTl com uma nova formulação para estimação de incertezas apresentada em [5]. A partir disto, Pinheiro *et al.* apresentou avanços na área de VLTl com resultados superiores à abordagem GMM-UBM.

No entanto, na proposta de [6] o modelo de locutor foi treinado e testado em ambientes ruidosos. Desse modo, ainda não conhecemos seu comportamento quando treinado em ambientes livres de ruído e testados sob ruído. Em outras palavras, não sabemos se o T2 FGMM é capaz de discriminar ruído de fala quando treinado sem informação de ruído. Neste sentido, a proposta de [6] deve ser amadurecida e avaliada em um protocolo experimental mais crítico de modo a encontrar avanços ainda maiores em relação ao estado da arte.

2. Objetivos

Diante do cenário exposto, identificamos os seguintes objetivos:

1. Avaliar robustez do T2 FGMM proposto por Pinheiro *et al* [6] sob condições críticas. Para tal, o sistema será experimentado sob treinamento com locuções livres de ruído e receberá locuções de teste sob presença de ruído. Dessa maneira, podemos identificar o poder discriminatório da estimação da incerteza dos parâmetros proposta;
2. Reformular a estimação de incertezas dos parâmetros de acordo com o observado no 1º objetivo.

3. Cronograma

O desenvolvimento deste trabalho está dividido em atividades a serem realizadas, segundo o cronograma presente na tabela abaixo:

	Outubro	Novembro	Dezembro	Janeiro	Fevereiro
Revisão Lit.	X X X X				
Implem. T2FGGM	X X	X X			
Testes		X X			
Reform. T2FGMM		X X	X X X X		
Avaliação				X X X X	
Esc. Relat.			X X X X	X X	X X
Apres. Oral					X X

4. Avaliadores

O avaliador solicitado é o Professor **George Darmiton da Cunha Cavalcanti** (gdcc@cin.ufpe.br) .

Em caso de indisponibilidade, o possível avaliador é o Professor **Carlos Alexandre Barros de Mello** (cabm@cin.ufpe.br).

5. Referências

- [1] F. Bimbot, J. –F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska-Delacretaz, D. A. Reynolds, “A tutorial on text-independent speaker verification”, *EURASIP J. Appl. Signal Process.* 4, p. 430 – 451, 2004.
- [2] Lawrence R. Rabiner, Ronald W. Schafer, “Introduction to Digital Speech Processing”, *Foundations and Trends in Signal Processing*, p. 1-194, January 2007.
- [3] R.H. Woo, A. Park, and T.J. Hazen, “The MIT mobile device speaker verification corpus: Data collection and preliminary experiments,” *The IEEE Odyssey Speaker and Language Recognition Workshop*, p. 1–6, 2006.
- [4] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, “Speaker verification using adapted gaussian mixture models”. *Digital Signal Process.*, v. 10, n. 1, pp. 19–41, 2000.
- [5] Jia Zeng, Lei Xie, Zhi-Qiang Liu, “Type-2 fuzzy Gaussian mixture models”, *Pattern Recognition*, v. 41, n. 12, p. 3636-3643, 2008.
- [6] Hector N. B. Pinheiro, Tsang Ing Ren, George D. C. Cavalcanti, Ing Jyh Tsang and Jan Sijbers. “Type-2 fuzzy GMMs for robust text-independent speaker verification in noisy environments”. *International Conference on Pattern Recognition (ICPR)*, p. 4531-4536, 2014.

6. Assinaturas

Sérgio Renan Ferreira Vleira - **Orientando**

Tsang Ing Ren - **Orientador**

Recife, Outubro de 2014.