Sistemas de Verificação de Locutores Independentes de Texto

Agosto, 2014

Sérgio R. F. Vieira e Eduardo M. B. de A. Tenório
{srfv,embat}@cin.ufpe.br

Centro de Informática, Universidade Federal de Pernambuco, Brasil
Adaptação da apresentação de Hector N. B. Pinheiro

Roteiro

- Reconhecimento de Locutores
  - Definição
  - Taxonomia
- Sistemas de Verificação de Locutores Independentes de Texto
  - Definição
  - Arquitetura Geral
  - Pré-processamento
  - Extração de características
  - Estimação dos modelos
  - GMM-UBM
- Experimentos
  - MIT Mobile Device Speaker Verification Corpus
  - Resultados GMM-UBM

Roteiro

- Reconhecimento de Locutores
  - Definição
  - Taxonomia
- Sistemas de Verificação de Locutores Independentes de Texto
  - Definição
  - Arquitetura Geral
  - Pré-processamento
  - Extração de características
  - Estimação dos modelos
  - GMM-UBM
- Experimentos
  - MIT Mobile Device Speaker Verification Corpus
  - Resultados GMM-UBM

Reconhecimento de Locutores

- Definição
  - Reconhecer o locutor a partir de informações extraídas de locuções.
  - A partir de uma determinada locução, inferir afirmações sobre a identidade da pessoa que a produziu.

Roteiro

- Reconhecimento de Locutores
  - Definição
  - Taxonomia
- Sistemas de Verificação de Locutores Independentes de Texto
  - Definição
  - Arquitetura Geral
  - Pré-processamento
  - Extração de características
  - Estimação dos modelos
  - GMM-UBM
- Experimentos
  - MIT Mobile Device Speaker Verification Corpus
  - Resultados GMM-UBM

Reconhecimento de Locutores

- Taxonomia
  - Com respeito à tarefa
  - Identificação (1:n)
    Dada uma locução, identificar, dentre um conjunto fechado de possíveis locutores, qual deles produziu a locução.
  - Problema de conjunto fechado.
  - Mais fácil, uma vez que temos a certeza de que a locução foi produzida por um dos locutores registrados no sistema.

  - Verificação (1:1)
  - Dada uma locução (X) e um determinado locutor (S), decidir se X foi produzida por S.
  - Problema de conjunto aberto.
  - Mais difícil, tendo em vista que é necessário considerar um conjunto potencialmente infinito de impostores.

Reconhecimento de Locutores

- Taxonomia
  - Com respeito às restrições impostas às locuções
  - Dependente de texto
  - A locução utilizada no reconhecimento deve possuir uma determinada palavra ou frase.
  - Alternativa amplamente utilizada para assegurar segurança na autenticação de pessoas (torna mais difícil o ataque de impostores).
  - Nesse caso, as locuções de treino e teste possuem o mesmo conteúdo fonético. Por essa razão, o sistema não precisa aprender as diversas características fonéticas do indivíduo.

  - Independente de texto
  - O reconhecimento deve ocorrer a partir de qualquer locução, independente do seu conteúdo.
  - O sistema precisa generalizar bem as diversas variações fonéticas de um determinado indivíduo.
  - Amplamente utilizado para aplicações forenses.

Reconhecimento de Locutores

- Veremos agora um pouco sobre os Sistemas de Verificação de Locutores Independentes de Texto.

Roteiro

- Reconhecimento de Locutores
  - Definição
  - Taxonomia
- Sistemas de Verificação de Locutores Independentes de Texto
  - Definição
  - Arquitetura Geral
  - Pré-processamento
  - Extração de características
  - Estimação dos modelos
  - GMM-UBM
- Experimentos
  - MIT Mobile Device Speaker Verification Corpus
  - Resultados GMM-UBM

Sistemas de Verificação de Locutores Independentes de Texto

- Na literatura, observamos que desde o princípio, a definição do problema seguiu uma abordagem bayesiana.

  D. A. Reynolds, “A Gaussian mixture modeling approach to text independent speaker identification,” Ph.D. thesis, Georgia Institute of Technology, 1992.

- As mesmas definições são seguidas até hoje. Além disso, nenhuma outra abordagem conseguiu o mesmo sucesso.

- Essa mesma abordagem estocástica também pode ser observado em outros problemas que envolvem a análise da voz. Tais como reconhecimento de fala, por exemplo.

Sistemas de Verificação de Locutores Independentes de Texto

- Dada uma locução Y e um locutor S, o sistema opera sobre duas hipóteses:

$$H_0 = Y\ foi\ produzida\ por\ S.$$
$$H_1 = Y\ não\ foi\ produzida\ por\ S.$$

Teste da razão das verossimilhanças:

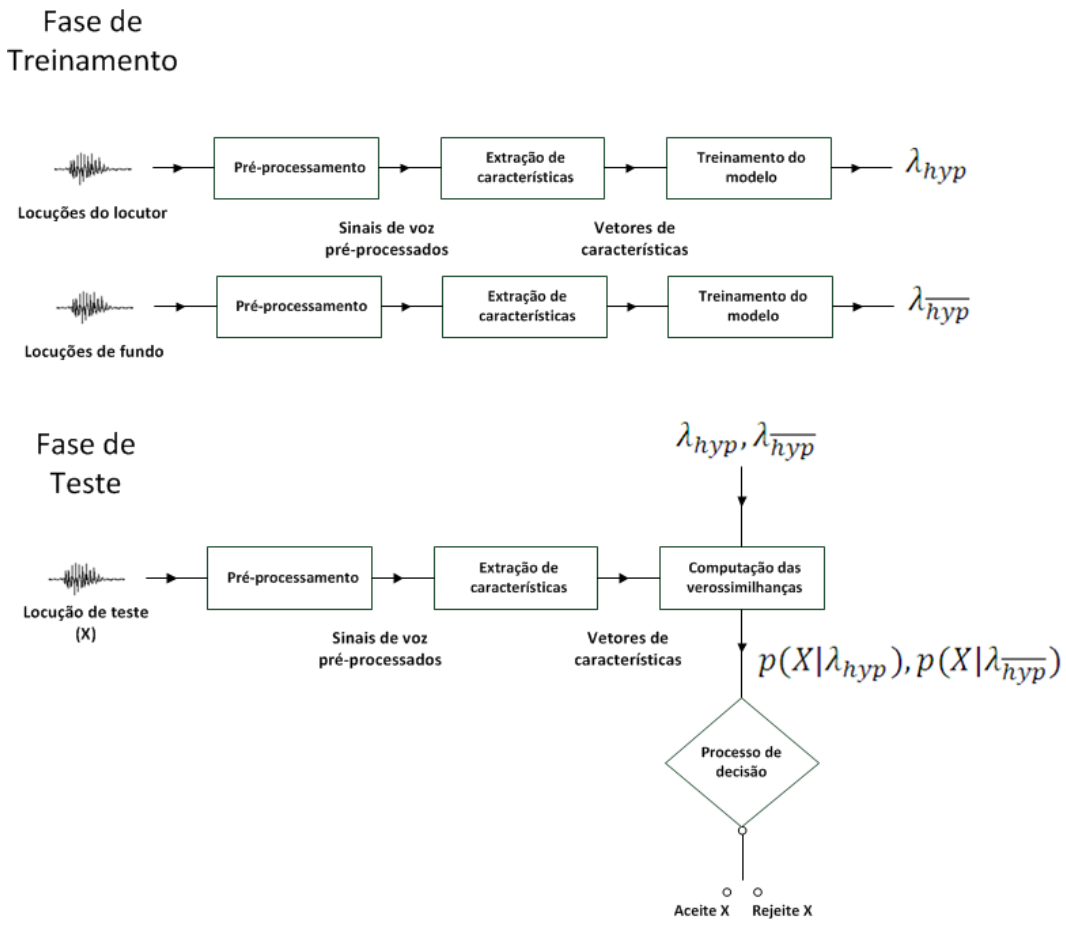
$$\frac{p(Y|H_0)}{p(Y|H_1)} = \begin{cases} \geq \theta, aceite\ H_0, \\ < \theta, rejeite\ H_0. \end{cases}$$

Roteiro

- Reconhecimento de Locutores
  - Definição
  - Taxonomia
- Sistemas de Verificação de Locutores Independentes de Texto
  - Definição
  - Arquitetura Geral
  - Pré-processamento
  - Extração de características
  - Estimação dos modelos
  - GMM-UBM
- Experimentos
  - MIT Mobile Device Speaker Verification Corpus
  - Resultados GMM-UBM

Sistemas de Verificação de Locutores Independentes de Texto

- Arquitetura geral:



Roteiro

- Reconhecimento de Locutores
  - Definição
  - Taxonomia
- Sistemas de Verificação de Locutores Independentes de Texto
  - Definição
  - Arquitetura Geral
  - Pré-processamento
  - Extração de características
  - Estimação dos modelos
  - GMM-UBM
- Experimentos
  - MIT Mobile Device Speaker Verification Corpus
  - Resultados GMM-UBM

Sistemas de Verificação de Locutores Independentes de Texto

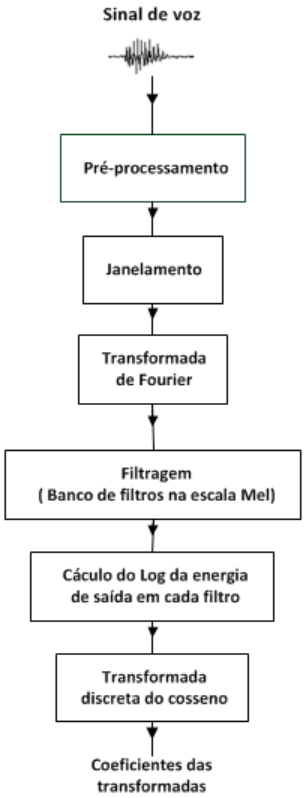
- Pré-processamento
  - Técnicas que possuem o objetivo de processar o sinal de voz a fim de enfatizar as informações úteis presentes no mesmo.
  - Uma tarefa que poderia ser executada nesse módulo é a limpeza do sinal, a fim de retirar a influência de possíveis ruídos (proveniente do canal de comunicação ou do ambiente).
  - Porém, a Literatura evoluiu para técnicas de extração de características que se propõem a serem robustas nesse sentido.
  - Ainda hoje, podemos observar que nesse módulo ainda são executados os chamados VADs (Voice Activity Detectors). Técnicas são executadas para descartar as partes do sinal que não possuem voz (silêncio) ou que são bastante ruidosas.

Roteiro

- Reconhecimento de Locutores
  - Definição
  - Taxonomia
- Sistemas de Verificação de Locutores Independentes de Texto
  - Definição
  - Arquitetura Geral
  - Pré-processamento
  - Extração de características
  - Estimação dos modelos
  - GMM-UBM
- Experimentos
  - MIT Mobile Device Speaker Verification Corpus
  - Resultados GMM-UBM

Sistemas de Verificação de Locutores Independentes de Texto

- Extração de características
  - Hoje em dia o conjunto de características mais utilizados para aplicações de voz são os chamados MFCCs (Mel Frequency Cepstral Coeficients) e seus coeficientes delta de primeira e segunda ordem.



Coeficientes delta:

Valor sugerido:

K=2, hk = 2K+1

$$\frac{\delta c_m(t)}{\delta t} \approx \Delta c_m = \frac{\sum_{k=-K}^{K} k h_k c_m(t+k)}{\sum_{k=-K}^{K} h_k k^2}$$

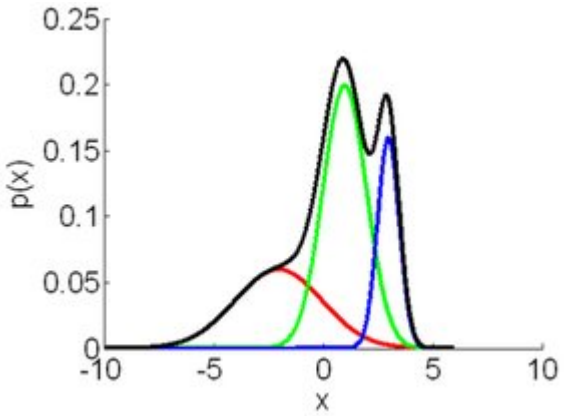
Roteiro

- Reconhecimento de Locutores
  - Definição
  - Taxonomia
- Sistemas de Verificação de Locutores Independentes de Texto
  - Definição
  - Arquitetura Geral
  - Pré-processamento
  - Extração de características
  - Estimação dos modelos
  - GMM-UBM
- Experimentos
  - MIT Mobile Device Speaker Verification Corpus
  - Resultados GMM-UBM

Sistemas de Verificação de Locutores Independentes de Texto

- Treinamento dos modelos
  - Objetivo:
  - Estimação do modelo de um determinado locutor, referente à hipótese nula, e de um modelo dos impostores, referente à hipótese alternativa.
  - Para sistemas independentes de texto, o modelo mais bem sucedido até então é o GMM (Gaussian Mixture Model).

$$p(x|\theta) = \sum_{i=1}^{M} \omega_i \, N(x; \mu_i, \Sigma_i)$$



Roteiro

- Reconhecimento de Locutores
  - Definição
  - Taxonomia
- Sistemas de Verificação de Locutores Independentes de Texto
  - Definição
  - Arquitetura Geral
  - Pré-processamento
  - Extração de características
  - Estimação dos modelos
  - GMM-UBM
- Experimentos
  - MIT Mobile Device Speaker Verification Corpus
  - Resultados GMM-UBM

Sistemas de Verificação de Locutores Independentes de Texto

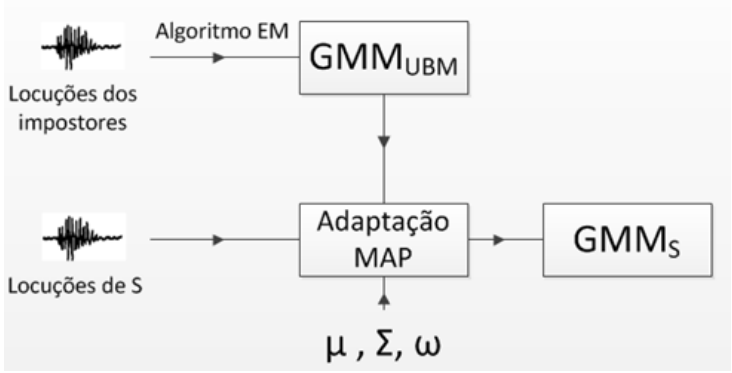
- GMM-UBM
  - A técnica padrão para geração dos modelos consiste no chamado sistema GMM-UBM, proposto também por Reynolds, em 2000.

D. A. Reynolds, T. F. Quatieri, and R. Dunn, “Speaker verification using adapted Gaussian mixture models,” *Digital Signal Process.*, 2000.

  - Como mencionado anteriormente, o sistema necessita de um modelo específico para modelagem da hipótese alternativa.
  - Esse modelo consiste no UBM (Universal Background Model).
  - O UBM é um GMM com matrizes de covariâncias diagonais estimado utilizando locuções de diversos (muitos!) locutores diferentes.
  - Especificamente, ele modela a distribuição esperada dos vetores de características para vozes.
  - Essa estimação é realizada através do algoritmo EM (Expectation Maximization), que é o algoritmo tradicional para maximização de verossimilhanças de distribuições paramétricas.

Sistemas de Verificação de Locutores Independentes de Texto

- GMM-UBM
  - Apesar de o modelo de um determinado locutor poder também ser estimado via EM, os autores mostraram um método de estimação desses modelos a partir do UBM.
  - Utilizando locuções do locutor específico, uma nova adaptação é realizada no UBM a fim de produzir o modelo do locutor.
  - A adaptação realizada consiste na chamada adaptação MAP (Maximum a Posteriori), que maximiza a probabilidade a posteriori dos vetores, diferente do EM, que maximiza as verossimilhanças do modelo com respeito aos vetores.



Sistemas de Verificação de Locutores Independentes de Texto

- GMM-UBM
  - Quando a adaptação MAP opera sobre GMMs, as misturas do modelo que possuem maior probabilidade a posteriori são mais adaptadas que as outras. Isso deriva diretamente das equações de adaptação.
  - Dado o UBM e um conjunto de vetores extraídos das locuções de um determinado locutor, $X = \{\boldsymbol{x_1}, \ldots, \boldsymbol{x_T}\}$, primeiramente calcula-se a probabilidade a posteriori referente a cada uma das i misturas do UBM.
  - Depois, exatamente como no algoritmo EM, as estatísticas de interesse são calculadas:

$$\Pr(i \mid \boldsymbol{x}_t) = \frac{w_i \, p_i(\boldsymbol{x}_t)}{\sum_{j=1}^{M} w_j \, p_j(\boldsymbol{x}_t)}$$

$$n_i = \sum_{t=1}^{T} \Pr(i \mid \boldsymbol{x}_t) \qquad E_i(\boldsymbol{x}) = \frac{1}{n_i} \sum_{t=1}^{T} \Pr(i \mid \boldsymbol{x}_t) \boldsymbol{x}_t \qquad E_i(\boldsymbol{x}^2) = \frac{1}{n_i} \sum_{t=1}^{T} \Pr(i \mid \boldsymbol{x}_t) \boldsymbol{x}_t^2$$

Sistemas de Verificação de Locutores Independentes de Texto

- GMM-UBM
  - Finalmente, as adaptações das misturas são realizadas utilizando as estatísticas:

$$\hat{w}_i = \left[\alpha_i^w n_i / T + (1 - \alpha_i^w) w_i\right] \gamma$$

$$\hat{\boldsymbol{\mu}}_i = \alpha_i^m E_i(\boldsymbol{x}) + (1 - \alpha_i^m) \boldsymbol{\mu}_i$$

$$\hat{\boldsymbol{\sigma}}_i^2 = \alpha_i^v E_i(\boldsymbol{x}^2) + (1 - \alpha_i^v)(\boldsymbol{\sigma}_i^2 + \boldsymbol{\mu}_i^2) - \hat{\boldsymbol{\mu}}_i^2$$

  - Essas equações de adaptação são as equações produzidas quando realizamos a adaptação MAP, com exceção dos chamados fatores de relevância para cada um dos parâmetros de cada mistura.

$$\alpha_i^\rho = \frac{n_i}{n_i + r^\rho}, \qquad \rho \in \{w, m, v\}$$

Sistemas de Verificação de Locutores Independentes de Texto

- GMM-UBM
  - Por fim, adaptações sucessivas são executadas até que a probabilidade a posteriori das misturas estabilize.
  - Um importante resultado apresentado pelos autores mostra que ao invés de adaptarmos os três parâmetros (média, peso e matriz de covariância) do UBM, melhores resultados são alcançados quando apenas as médias são adaptadas.
  - Intuitivamente, podemos observar que o modelo de um determinado locutor é bem parecido com o UBM. Suas misturas possuem os mesmos pesos e as mesmas matrizes de covariância. Além disso, boa parte das médias também é bem similar, uma vez que apenas uma pequena parte das misturas são adaptadas pela adaptação MAP.
  - Isso cria uma relação intrínseca entre o modelo do locutor e o UBM, melhorando o desempenho do teste de razão de verossimilhanças.

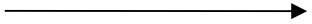
Roteiro

- Reconhecimento de Locutores
  - Definição
  - Taxonomia
- Sistemas de Verificação de Locutores Independentes de Texto
  - Definição
  - Arquitetura Geral
  - Pré-processamento
  - Extração de características
  - Estimação dos modelos
  - GMM-UBM
- Experimentos
  - MIT Mobile Device Speaker Verification Corpus
  - Resultados GMM-UBM

Experimentos

- MIT Mobile Device Speaker Verification Corpus
  - 48 locutores registrados (26 homens, 22 mulheres)
  - 3 sessões de áudio
    - 2 sessões com os locutores registrados.
    - 1 sessão com 40 locutores impostores.
    - Para cada locutor, 54 locuções em cada sessão.
    - As locuções foram gravadas em 3 ambientes distintos.
    - Escritório silencioso, com pouco ruído,
    - Hall de entrada de um prédio, com nível médio de ruído de fundo,
    - Cruzamento entre ruas movimentadas, com nível alto de ruído de fundo.
    - 18 locuções por ambiente.

Sessão	Treinamento	Teste
Enroll session 1	X	
Enroll session 2		X
Imposter session		X



Para cada locutor, temos um total de 54 testes de autenticação correta (18 em cada ambiente) e 2160 testes de ataques de impostores (720 em cada ambiente).

Roteiro

- Reconhecimento de Locutores
  - Definição
  - Taxonomia
- Sistemas de Verificação de Locutores Independentes de Texto
  - Definição
  - Arquitetura Geral
  - Pré-processamento
  - Extração de características
  - Estimação dos modelos
  - GMM-UBM
- Experimentos
  - MIT Mobile Device Speaker Verification Corpus
  - Resultados GMM-UBM

Experimentos

- Resultados do GMM-UBM
  - Pré-processamento:
  - Extração de características:
  - 19 coeficientes MFCC + Coeficientes delta de primeira e segunda ordem.
  - Geração do UBM:
  - Execução do algoritmo EM utilizando os vetores extraídos de todas as locuções de treino de todos os locutores.
  - Número de distribuições do UBM igual a 512, sendo uma fusão de 2 GMMs: um de homens com 256 misturas e outro de mulheres com 256 misturas.

Experimentos

- Resultados do GMM-UBM
  - Equal Error Rate de 40,9%, aparentemente devido à falta do VAD durante o pré-processamento.