

Text-Independent Speaker Recognition Using Gaussian Mixture Models

Eduardo Martins Barros de Albuquerque Tenório

Centro de Informática
Universidade Federal de Pernambuco
Trabalho de Graduação em Engenharia da Computação
embat@cin.ufpe.br

Recife, 25 de Junho de 2015

Conteúdo

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor
- 3 Extração de Características
- 4 Modelos de Mistura Gaussianas
- 5 Experimentos
- 6 Conclusão
- 7 Referências

Conteúdo

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor
- 3 Extração de Características
- 4 Modelos de Mistura Gaussianas
- 5 Experimentos
- 6 Conclusão
- 7 Referências

Status quo...

Reconhecimento de usuários ainda é um processo bastante “manual”

Status quo...

Reconhecimento de usuários ainda é um processo bastante “manual”

Verificação O usuário deve fornecer uma senha correta para:

Status quo...

Reconhecimento de usuários ainda é um processo bastante “manual”

Verificação O usuário deve fornecer uma senha correta para:

- Efetuar login em um computador

Status quo...

Reconhecimento de usuários ainda é um processo bastante “manual”

Verificação O usuário deve fornecer uma senha correta para:

- Efetuar login em um computador
- Sacar dinheiro em um ATM

Status quo...

Reconhecimento de usuários ainda é um processo bastante “manual”

Verificação O usuário deve fornecer uma senha correta para:

- Efetuar login em um computador
- Sacar dinheiro em um ATM
- Ouvir seu saldo bancário pelo telefone

Status quo...

Reconhecimento de usuários ainda é um processo bastante “manual”

Verificação O usuário deve fornecer uma senha correta para:

- Efetuar login em um computador
- Sacar dinheiro em um ATM
- Ouvir seu saldo bancário pelo telefone

Identificação Um perito deve:

Status quo...

Reconhecimento de usuários ainda é um processo bastante “manual”

Verificação O usuário deve fornecer uma senha correta para:

- Efetuar login em um computador
- Sacar dinheiro em um ATM
- Ouvir seu saldo bancário pelo telefone

Identificação Um perito deve:

- Ouvir diversas ligações de telefone

Status quo...

Reconhecimento de usuários ainda é um processo bastante “manual”

Verificação O usuário deve fornecer uma senha correta para:

- Efetuar login em um computador
- Sacar dinheiro em um ATM
- Ouvir seu saldo bancário pelo telefone

Identificação Um perito deve:

- Ouvir diversas ligações de telefone
- Detectar características (sotaque, vícios de linguagem, gírias)

Status quo...

Reconhecimento de usuários ainda é um processo bastante “manual”

Verificação O usuário deve fornecer uma senha correta para:

- Efetuar login em um computador
- Sacar dinheiro em um ATM
- Ouvir seu saldo bancário pelo telefone

Identificação Um perito deve:

- Ouvir diversas ligações de telefone
- Detectar características (sotaque, vícios de linguagem, gírias)
- Conhecer detalhes do usuário e do assunto tratado

... e suas falhas

E justamente por ser “manual”, apresenta resultados pouco satisfatórios

... e suas falhas

E justamente por ser “manual”, apresenta resultados pouco satisfatórios

Verificação A senha pode ser facilmente:

... e suas falhas

E justamente por ser “manual”, apresenta resultados pouco satisfatórios

Verificação A senha pode ser facilmente:

- Esquecida

... e suas falhas

E justamente por ser “manual”, apresenta resultados pouco satisfatórios

Verificação A senha pode ser facilmente:

- Esquecida
- Anotada em um papel e perdida

... e suas falhas

E justamente por ser “manual”, apresenta resultados pouco satisfatórios

Verificação A senha pode ser facilmente:

- Esquecida
- Anotada em um papel e perdida
- Quebrada

... e suas falhas

E justamente por ser “manual”, apresenta resultados pouco satisfatórios

Verificação A senha pode ser facilmente:

- Esquecida
- Anotada em um papel e perdida
- Quebrada

Identificação Pessoas apresentam diversos defeitos:

... e suas falhas

E justamente por ser “manual”, apresenta resultados pouco satisfatórios

Verificação A senha pode ser facilmente:

- Esquecida
- Anotada em um papel e perdida
- Quebrada

Identificação Pessoas apresentam diversos defeitos:

- Análises enviesadas

... e suas falhas

E justamente por ser “manual”, apresenta resultados pouco satisfatórios

Verificação A senha pode ser facilmente:

- Esquecida
- Anotada em um papel e perdida
- Quebrada

Identificação Pessoas apresentam diversos defeitos:

- Análises enviesadas
- Precisam de anos de treinamento e prática para se especializarem

... e suas falhas

E justamente por ser “manual”, apresenta resultados pouco satisfatórios

Verificação A senha pode ser facilmente:

- Esquecida
- Anotada em um papel e perdida
- Quebrada

Identificação Pessoas apresentam diversos defeitos:

- Análises enviesadas
- Precisam de anos de treinamento e prática para se especializarem
- Demoram e se cansam com facilidade

Solução

Utilizar alguma biometria

Solução

Utilizar alguma biometria

Preferivelmente pouco invasiva

Solução

Utilizar alguma biometria

Preferivelmente pouco invasiva

A voz é uma ótima escolha:

Solução

Utilizar alguma biometria

Preferivelmente pouco invasiva

A voz é uma ótima escolha:

- Bastante natural para humanos

Solução

Utilizar alguma biometria

Preferivelmente pouco invasiva

A voz é uma ótima escolha:

- Bastante natural para humanos
- Peculiaridades referentes ao locutor

Solução

Utilizar alguma biometria

Preferivelmente pouco invasiva

A voz é uma ótima escolha:

- Bastante natural para humanos
- Peculiaridades referentes ao locutor
- (quase) Todo mundo fala

Solução

Utilizar alguma biometria

Preferivelmente pouco invasiva

A voz é uma ótima escolha:

- Bastante natural para humanos
- Peculiaridades referentes ao locutor
- (quase) Todo mundo fala

Soluções atuais apresentam progresso contínuo:

Solução

Utilizar alguma biometria

Preferivelmente pouco invasiva

A voz é uma ótima escolha:

- Bastante natural para humanos
- Peculiaridades referentes ao locutor
- (quase) Todo mundo fala

Soluções atuais apresentam progresso contínuo:

- Apple Siri

Solução

Utilizar alguma biometria

Preferivelmente pouco invasiva

A voz é uma ótima escolha:

- Bastante natural para humanos
- Peculiaridades referentes ao locutor
- (quase) Todo mundo fala

Soluções atuais apresentam progresso contínuo:

- Apple Siri
- Google Now

Solução

Utilizar alguma biometria

Preferivelmente pouco invasiva

A voz é uma ótima escolha:

- Bastante natural para humanos
- Peculiaridades referentes ao locutor
- (quase) Todo mundo fala

Soluções atuais apresentam progresso contínuo:

- Apple Siri
- Google Now
- Samsung S Voice

Reconhecimento de ...

Fala **O que** está sendo dito

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem
- Estado emocional do locutor

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem
- Estado emocional do locutor
- Sotaque ou dificuldade de articulação

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem
- Estado emocional do locutor
- Sotaque ou dificuldade de articulação

Locutor **Quem** está falando

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem
- Estado emocional do locutor
- Sotaque ou dificuldade de articulação

Locutor **Quem** está falando

- Identificar uma pessoa na multidão

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem
- Estado emocional do locutor
- Sotaque ou dificuldade de articulação

Locutor **Quem** está falando

- Identificar uma pessoa na multidão
- Autenticar um usuário

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem
- Estado emocional do locutor
- Sotaque ou dificuldade de articulação

Locutor **Quem** está falando

- Identificar uma pessoa na multidão
- Autenticar um usuário

É comum o foco em apenas um tipo

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem
- Estado emocional do locutor
- Sotaque ou dificuldade de articulação

Locutor **Quem** está falando

- Identificar uma pessoa na multidão
- Autenticar um usuário

É comum o foco em apenas um tipo

Fala Aplicações de busca por voz

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem
- Estado emocional do locutor
- Sotaque ou dificuldade de articulação

Locutor **Quem** está falando

- Identificar uma pessoa na multidão
- Autenticar um usuário

É comum o foco em apenas um tipo

Fala Aplicações de busca por voz

Locutor Prevenção de fraude via telefone

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem
- Estado emocional do locutor
- Sotaque ou dificuldade de articulação

Locutor **Quem** está falando

- Identificar uma pessoa na multidão
- Autenticar um usuário

É comum o foco em apenas um tipo

Fala Aplicações de busca por voz

Locutor Prevenção de fraude via telefone

Ambos Transcrição de conversas gravadas

Reconhecimento de Locutor

Identificação Determina a identidade de um locutor dentro de um conjunto não unitário

Reconhecimento de Locutor

Identificação Determina a identidade de um locutor dentro de um conjunto não unitário

- 1 para N

Reconhecimento de Locutor

Identificação Determina a identidade de um locutor dentro de um conjunto não unitário

- 1 para N
- Problema de **conjunto fechado**

Reconhecimento de Locutor

Identificação Determina a identidade de um locutor dentro de um conjunto não unitário

- 1 para N
- Problema de **conjunto fechado**

Verificação Determina se o locutor é quem diz ser

Reconhecimento de Locutor

Identificação Determina a identidade de um locutor dentro de um conjunto não unitário

- 1 para N
- Problema de **conjunto fechado**

Verificação Determina se o locutor é quem diz ser

- 1 para 1

Reconhecimento de Locutor

Identificação Determina a identidade de um locutor dentro de um conjunto não unitário

- 1 para N
- Problema de **conjunto fechado**

Verificação Determina se o locutor é quem diz ser

- 1 para 1
- Problema de **conjunto aberto**

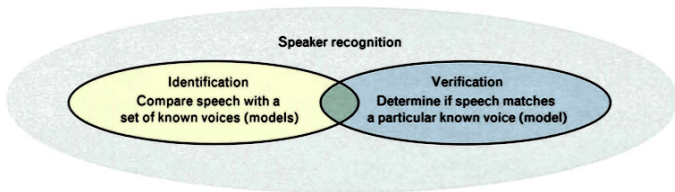
Reconhecimento de Locutor

Identificação Determina a identidade de um locutor dentro de um conjunto não unitário

- 1 para N
- Problema de **conjunto fechado**

Verificação Determina se o locutor é quem diz ser

- 1 para 1
- Problema de **conjunto aberto**



Dependência de texto

Dependente Teste \in Treinamento

Dependência de texto

Dependente Teste \in Treinamento

- Diversos graus de dependência

Dependência de texto

Dependente Teste \in Treinamento

- Diversos graus de dependência
- Teste \notin Treinamento \implies Retreinamento

Dependência de texto

Dependente Teste \in Treinamento

- Diversos graus de dependência
- Teste \notin Treinamento \implies Retreinamento

Independente Teste \neq Treinamento

Dependência de texto

Dependente Teste \in Treinamento

- Diversos graus de dependência
- Teste \notin Treinamento \implies Retreinamento

Independente Teste \neq Treinamento

- Características não textuais

Dependência de texto

Dependente Teste \in Treinamento

- Diversos graus de dependência
- Teste \notin Treinamento \implies Retreinamento

Independente Teste \neq Treinamento

- Características não textuais
- Presentes em diferentes sotaques e até *gibberish*

Modelos de Mistura Gaussiana

GMM Combinação de Gaussianas

Modelos de Mistura Gaussiana

GMM Combinação de Gaussianas

UBM GMM gerado por diversas locuções de fundo

Modelos de Mistura Gaussiana

GMM Combinação de Gaussianas

UBM GMM gerado por diversas locuções de fundo

AGMM GMM adaptado a partir de um UBM

Modelos de Mistura Gaussiana

GMM Combinação de Gaussianas

UBM GMM gerado por diversas locuções de fundo

AGMM GMM adaptado a partir de um UBM

FGMM GMM utilizando Fractional Covariance Matrix (FCM)

Objetivos

Implementar sistemas de reconhecimento de locutor e analisar

Objetivos

Implementar sistemas de reconhecimento de locutor e analisar

- Taxas de sucesso para **identificação**

Objetivos

Implementar sistemas de reconhecimento de locutor e analisar

- Taxas de sucesso para **identificação**
- Comparar identificação utilizando GMM e FGMM

Objetivos

Implementar sistemas de reconhecimento de locutor e analisar

- Taxas de sucesso para **identificação**
- Comparar identificação utilizando GMM e FGMM
- Falsa detecção e falsa rejeição para **verificação**

Objetivos

Implementar sistemas de reconhecimento de locutor e analisar

- Taxas de sucesso para **identificação**
- Comparar identificação utilizando GMM e FGMM
- Falsa detecção e falsa rejeição para **verificação**
- Comparar verificação utilizando GMM e AGMM

Conteúdo

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor**
- 3 Extração de Características
- 4 Modelos de Mistura Gaussianas
- 5 Experimentos
- 6 Conclusão
- 7 Referências

Sistemas de Reconhecimento de Locutor

Conteúdo

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor
- 3 Extração de Características**
- 4 Modelos de Mistura Gaussianas
- 5 Experimentos
- 6 Conclusão
- 7 Referências

Extração de Características

Conteúdo

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor
- 3 Extração de Características
- 4 Modelos de Mistura Gaussianas**
- 5 Experimentos
- 6 Conclusão
- 7 Referências

Modelos de Misturas Gaussianas

Conteúdo

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor
- 3 Extração de Características
- 4 Modelos de Mistura Gaussianas
- 5 Experimentos**
- 6 Conclusão
- 7 Referências

Experimentos

Conteúdo

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor
- 3 Extração de Características
- 4 Modelos de Mistura Gaussianas
- 5 Experimentos
- 6 Conclusão**
- 7 Referências

Conclusão

Conteúdo

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor
- 3 Extração de Características
- 4 Modelos de Mistura Gaussianas
- 5 Experimentos
- 6 Conclusão
- 7 Referências**

Referências

Obrigado