

Monografia apresentada para cumprimento da disciplina de estudos especiais

André Furlan - ensismoebius@gmail.com

Universidade Estadual Paulista Júlio de Mesquita Filho

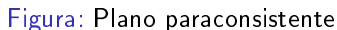
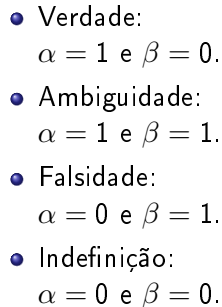
2019

Estrutura da apresentação

- Revisão de conceitos utilizados.
- Trabalhos correlatos.
- Contextualização.

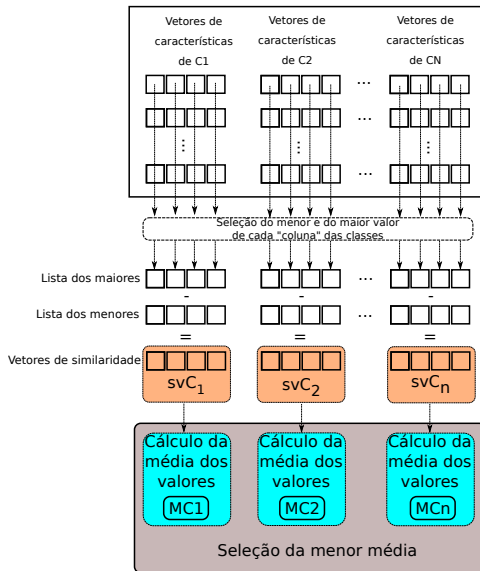
Revisão de conceitos utilizados

Os vetores de características proporcionam uma boa separação interclasses?



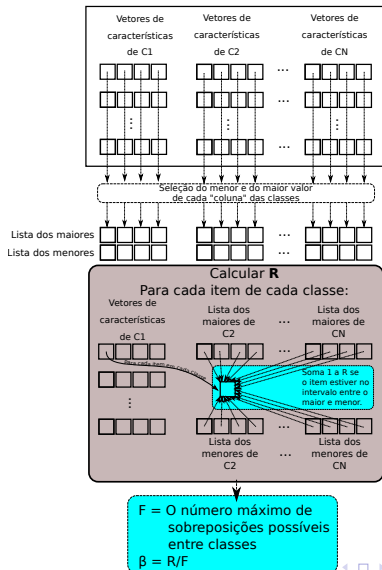
Engenharia paraconsistente de características

Cálculo de α



Engenharia paraconsistente de características

Cálculo de β



Engenharia paraconsistente de características

Graus de certeza e contradição

- Grau de certeza $\rightarrow G_1 = \alpha - \beta$.
- Grau de contradição $\rightarrow G_2 = \alpha + \beta - 1$.

Onde: $-1 \leq G_1 \leq 1$ e $-1 \leq G_2 \leq 1$.

Seja $P = (G_1, G_2)$

- $(-1, 0) \rightarrow$ Falsidade;
- $(1, 0) \rightarrow$ Verdade;
- $(0, -1) \rightarrow$ Indefinição;
- $(0, 1) \rightarrow$ Ambiguidade.

Engenharia paraconsistente de características

Distancias no plano paraconsistente

As distâncias(D) do ponto $P = (G_1, G_2)$ dos limites supracitados. Tal cálculo pode ser feito da seguinte forma:

$$D_{-1,0} = \sqrt{(G_1 + 1)^2 + (G_2)^2} \quad ,$$

$$D_{1,0} = \sqrt{(G_1 - 1)^2 + (G_2)^2} \quad ,$$

$$D_{0,-1} = \sqrt{(G_1)^2 + (G_2 + 1)^2} \quad ,$$

$$D_{0,1} = \sqrt{(G_1)^2 + (G_2 - 1)^2} \quad ,$$

Filtros digitais *wavelet*

Propriedades

- Tamanho de janelas variável.
- Análise multiresolução.
- Análise detalhada em altas e baixas frequências.

Filtros digitais *wavelet*

Restrição de escopo

- Apenas transformadas diretas.
- Não haverá reconstrução do sinal.
- Construção de vetores de características.
- Domínio discreto.

Filtros digitais *wavelet*

Wavelets de Daubechies

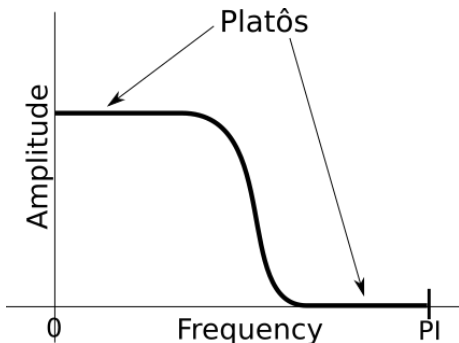


Figura: platôs maximamente planos em um filtro digital

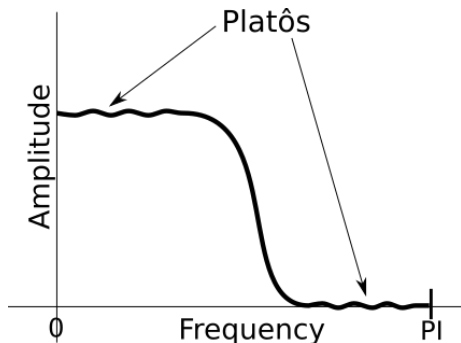


Figura: platôs não maximamente planos de um filtro digital

Filtros digitais *wavelet*

Algoritmo de Malat

- *Wavelet* Haar: $h[\cdot] = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$.
- Par ortogonal: $g[\cdot] = [\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}]$.
- sinal: $s[\cdot] = [1, 2, 3, 4]$.

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} \frac{3}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \\ \frac{7}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{pmatrix} \Rightarrow \left[\frac{3}{\sqrt{2}}, \frac{7}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right] .$$

Amostragem, quantização e o formato do arquivo Wave

Intervalo
em bytes

Tamanho
do trecho

0		
4	Id do trecho	4
8	Tamanho do trecho	4
12	Formato	4
16	Id do subtrecho fmt	4
20	Tamanho do subtrecho fmt	4
22	Formato do audio	2
24	Quantidade de canais	2
28	Taxa de amostragem	4
32	Taxa de bytes	4
34	Alinhamento de bloco	2
36	Bits por amostra	2
40	Id do subtrecho data	4
44	Tamanho do sub trecho data	4
	Dados do sinal	

■ Descritor do trecho **RIFF**

Contêm os descritores iniciais do arquivo indicando o formato "WAVE" o qual requer mais dois sub-trechos: "fmt" e "data"

■ Descritor do trecho **fmt**

Contêm as informações principais do sinal, aqui, por exemplo é possível saber se o sinal é mono ou estéreo, se é de 9 ou 16 bits, etc.

■ Descritor do trecho **data**

Contêm os dados em si e seu respectivo tamanho

- Formato *wave*.
- *Pulse-code modulation* (PCM).
- Taxa de amostragem de 44100hz.
- Resolução de 16bits

Caracterização dos processos de produção da voz humana

Áreas de estudo

- Fisiológica ou fonética articulatória.
- Acústica ou fonética acústica.
- Perceptual.

Foco apenas na acústica

Caracterização dos processos de produção da voz humana

Vozeada versus não-vozeada

- Vozeada: Pregas vocais.
- Não vozeada: Sem pregas vocais.

Caracterização dos processos de produção da voz humana

Frequência fundamental da voz

- Conhecida como F_0 .
- Componente periódico resultante da vibração das pregas vocais.

Caracterização dos processos de produção da voz humana

Formantes

- $F_1 \rightarrow$ amplificação sonora na cavidade oral posterior, posição da língua no plano vertical.
- $F_2 \rightarrow$ cavidade oral anterior, posição da língua no plano horizontal.
- $F_3 \rightarrow$ cavidades à frente e atrás do ápice da língua.
- $F_4 \rightarrow$ formato da laringe e da faringe.

Trabalhos correlatos

- [RFLC19]: Energia e outras várias características do espectro do sinal, SVM.
- [YXWW19] *Modelo oculto de Markov* (HMM); *Wavelets*, SVM.
- [SSAL17] *Wavelets*, coeficientes cepstrais (SCCs), Modelos de mistura Gaussiana (GMM).
- [AV19] *"Zero time windowing"*(ZTW), análise cepstral do espectro GMM.
- [YLYL19] Coeficientes cepstrais, GMM.
- [Han18] Predição linear, coeficientes cepstrais, GMM.
- [RBABA19] "Texturas de voz", padrões binários locais (LBP) e seus respectivos histogramas, SVM.

- [TDE17] *Transformada de constante Q (CQT)*, Processamento cepstral, fusão de dois classificadores GMM, sendo que um deles usa *coeficientes cepstrais de frequência MEL(MFCC)* e o outro usa características CFCC-IF [PP15]; STC [NKL⁺16]; SJTU [KMM⁺16].
- [SPM18] Partes não vozeadas da fala, três GMMs.
- [KP18] amplitude instantânea vinda de flutuações de energia, GMM.
- [WIAE18] Diferenças entre bandas de frequências específicas, *predição linear em domínio de frequência(FDLP)*, GMMs.
- [SSWA18] *Modulation spectral centroid frequency, long term spectral average*, GMM.
- [KP17] Envelopamento das amplitudes e das frequências instantâneas em cada banda estreita filtrada, GMM.
- [DGK⁺16] *Gammatone frequency cepstral coefficients(MGFCC)*, GMM.

- [AKY⁺18] *Hashing* sensível a locus(LSH), MFCC e LSH, tabela *hash*.

Contextualização

- Características mais disjuntas possíveis para de “locutor autêntico” e “ataque de *voice spoofing*”.
- *Wavelet*: Boa resolução em relação às dimensões de tempo e frequência.
- Análise paraconsistente de acordo com o trabalho [Gui19].



A. Awais, S. Kun, Y. Yu, S. Hayat, A. Ahmed, and T. Tu.

Speaker recognition using mel frequency cepstral coefficient and locality sensitive hashing.

In 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), pages 271–276, May 2018.



KNRK Raju Alluri and Anil Kumar Vuppala.

Replay spoofing countermeasures using high spectro-temporal resolution features.

International Journal of Speech Technology, 22(1):271–281, 2019.



K. Arun Das, Kuruvachan K. George, C. Santhosh Kumar, S. Veni, and Ashish Panda.

Modified Gammatone Frequency Cepstral Coefficients to Improve Spoofing Detection.

pages 50–55, 345 E 47TH ST, NEW YORK, NY 10017 USA, 2016.

LNM Inst Informat Technol; IEEE Commun Soc; IEEE Syst Man & Cybernet Soc, IEEE.

International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, INDIA, SEP 21-24, 2016.



R. C. Guido.

Paraconsistent feature engineering [lecture notes].

IEEE Signal Processing Magazine, 36(1):154–158, Jan 2019.



Cemal Hanilçi.

Linear prediction residual features for automatic speaker verification anti-spoofing.

Multimedia Tools and Applications, 77(13):16099–16111, Jul 2018.



Pavel Korshunov, Sébastien Marcel, Hannah Muckenhirn, André R Gonçalves, AG Souza Mello, RP Velloso Violato, Flávio O Simoes, M Uliani Neto, Marcus de Assis Angeloni, José Augusto Stuchi, et al. Overview of btas 2016 speaker anti-spoofing competition. pages 1–6. IEEE, 2016.



Madhu R. Kamble and Hemant A. Patil.

Novel Energy Separation Based Frequency Modulation Features For Spoofed Speech Classification.

pages 326–331, 2017.

9th International Conference on Advances in Pattern Recognition (ICAPR), Indian Stat Inst Bangalore, Bangalore, INDIA, DEC 27-30, 2017.



Madhu R. Kamble and Hemant A. Patil.

Novel Variable Length Energy Separation Algorithm using Instantaneous Amplitude Features For Replay Detection.

Interspeech, pages 646–650. Int Speech Commun Assoc, 2018.

19th Annual Conference of the

International-Speech-Communication-Association (INTERSPEECH 2018), Hyderabad, INDIA, AUG 02-SEP 06, 2018.



S. Novoselov, A. Kozlov, G. Lavrentyeva, K. Simonchik, and V. Shchemelinin.

Stc anti-spoofing systems for the asvspoof 2015 challenge.
pages 5475–5479, March 2016.



Tanvina Patel and Hemant Patil.

Combining evidences from mel cepstral, cochlear filter cepstral and instantaneous frequency features for detection of natural vs. spoofed speech.
09 2015.



Raoudha Rahmeni, Anis Ben Aicha, and Yassine Ben Ayed.

On the contribution of the voice texture for speech spoofing detection.
pages 501–505, 345 E 47TH ST, NEW YORK, NY 10017 USA, 2019.
IEEE.



Yanzhen Ren, Zhong Fang, Dengkai Liu, and Changwen Chen.
Replay attack detection based on distortion by loudspeaker for voice authentication.
Multimedia Tools and Applications, 78(7):8383–8396, Apr 2019.



M. S. Saranya, R. Padmanabhan, and Hema A. Murthy.
Replay Attack Detection in Speaker Verification Using non-voiced segments and Decision Level Feature Switching.
International Conference on Signal Processing and Communications SPCOM, pages 332–336. IEEE, 2018.
12th International Conference on Signal Processing and Communications (SPCOM), Indian Inst Sci, Bangalore, INDIA, JUL 16-19, 2018.



K. Sriskandaraja, V. Sethu, E. Ambikairajah, and H. Li.
Front-end for antispoofing countermeasures in speaker verification:
Scattering spectral decomposition.
IEEE Journal of Selected Topics in Signal Processing, 11(4):632–643,
June 2017.



Gajan Suthokumar, Vidhyasaharan Sethu, Chamith Wijenayake, and
Eliathamby Ambikairajah.
Modulation dynamic features for the detection of replay attacks.
pages 691–695, 2018.



Massimiliano Todisco, Héctor Delgado, and Nicholas Evans.
Constant q cepstral coefficients: A spoofing countermeasure for
automatic speaker verification.
Computer Speech & Language, 45:516 – 535, 2017.



Buddhi Wickramasinghe, Saad Irtza, Eliathamby Ambikairajah, and Julien Epps.

Frequency Domain Linear Prediction Features for Replay Spoofing Attack Detection.

Interspeech, pages 661–665, C/O EMMANUELLE FOXONET, 4 RUE DES FAUVETTES, LIEU DIT LOUS TOURILS, BAIXAS, F-66390, FRANCE, 2018. Int Speech Commun Assoc, ISCA-INT SPEECH COMMUNICATION ASSOC.

19th Annual Conference of the International-Speech-Communication-Association (INTERSPEECH 2018), Hyderabad, INDIA, AUG 02-SEP 06, 2018.



Y. Ye, L. Lao, D. Yan, and L. Lin.

Detection of replay attack based on normalized constant q cepstral feature.

pages 407–411, April 2019.



Diqun Yan, Li Xiang, Zhifeng Wang, and Rangding Wang.

Detection of hmm synthesized speech by wavelet logarithmic spectrum.

Automatic Control and Computer Sciences, 53(1):72–79, Jan 2019.