Experimento para determinar melhores janelas de curto e médio prazos

Sumário executivo

De acordo com o experimento realizado, as configurações das janelas tem pouca influência na acurácia de um classificador, foi detectado que janelas de tamanho menor tem um custo computacional maior. Portanto, recomenda-se a utilização de Short windows de duração maior (cerca de de 100ms).

Experiment plan and results	O que queremos medir: > *"Influência das janelas de curto prazo (sw ou ShortWindow, médio prazo ou MidWindow, sobreposições das respectivas janelas (shortWindow step ou Mid Windows step) na acurácia de em um classificador de áudio."* Como será realizado o experimento: 1. Dispomos de uma base de dados com 81 áudios onde ocorrem uma violência física contra uma mulher e 107 áudios onde não ocorre violência. Os áudios estão tratados (normalizados, com frequência máxima em 16kHz, taxa de amostragem de 16kHz, resolução de 16 bits / amostra, resultando em uma taxa de bit ("bit rate") de 256kbps, mono canal) 2. Criamos uma tabela com os ensaios a serem realizados, variando os parâmetros de interesse. Para cada combinação, será criado um classificador SVM e calculada a acurácia do classificador, utilizando cross validação (n=100) 3. Avaliaremos qual a influência dos parâmetros de entrada na acurácia dos classificador implementados.
Experiment owner	@ Tiago Beltrão Lacerda
Reviewers	☐ @ anapaula.furtado ☐ @ Péricles Miranda ☐ @ GILMÁRIO PERIRA DOS SANTOS
Approver	☐ @ anapaula.furtado
Optimizely link	
Jira ticket(s)	https://herproject.atlassian.net/browse/HER-1
Status	CONCLUÍDO
On this page	

Variáveis de entrada

O que vamos variar	Nivel baixo	Nível alto
Short Window (sw)	0,02s ou 20ms	0,1s ou 100ms
Short window Step (ss)	0% de sobreposição, ou seja: ss = **0,02 ou 0,1**	50% de sobreposição, ou seja: ss = 0,5 * sw = **0,01 ou 0,05**

Mid Window (mw)	1s	10s
Mid Windows Step (ms)	0% de sobreposição, ou seja: ms = 1s ou 10s	50% de sobreposição, ou seja: ms = 0,5s ou 5s

Tipo	Efeito
Frames longos..	Melhor representação em frequência, mais componentes porém perde-se qualidade no domínio do tempo, no sentido em que o intervalo não pode ser considerado estacionário de forma que podem ocorrer mais de um evento no tempo.
Frames curtos…	Pior representação na frequência (menos amostras). Melhor representação no tempo, a janela de tempo pode ser considerada constante com um delta T suficientemente pequeno.

📋 Planejamento do experimento

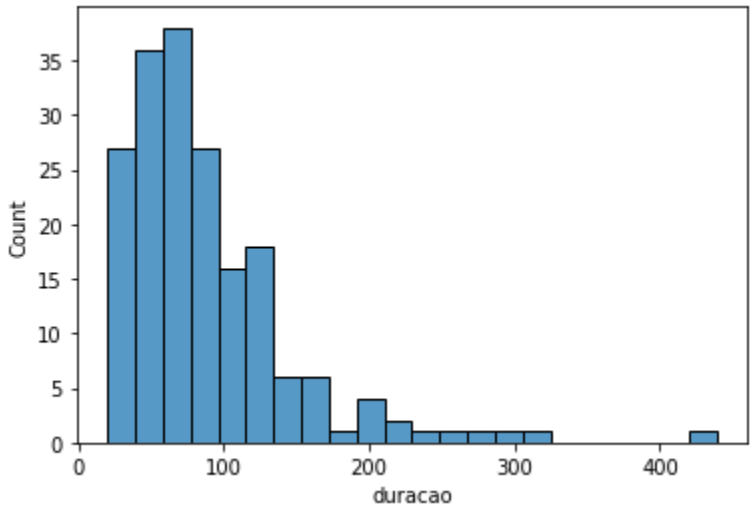
Visão geral

De acordo com Barros Neto (2001), no livro “Como fazer experimentos: pesquisa e desenvolvimento na indústria”, no planejamento de qualquer experimento a primeira coisa que devemos fazer é decidir quais são os fatores e as respostas de interesse.

Este é um experimento fatorial 2^k onde deseja-se conhecer o efeito dos tamanhos das janelas de curto e médio prazos e seus respectivos passos na acurácia final de um classificador. Com esta informação, poderemos determinar uma configuração ótima de janelas para o nosso contexto.

Dados:

Como dito anteriormente, dispomos de 86 áudios onde ocorre uma violência física contra a mulher e outros 107 onde não ocorre violência física, os áudios tem tamanhos variados, conforme histograma abaixo:



Recomendação de configuração de janelas de outros autores:

Autor	Contexto	Setup
Batista-Duran (2016)	Violência	janelas de 400ms com 95% de sobreposição (step de 20ms)
Giannakopoulos (2015)	Geral	20 a 100ms para janelas de curto prazo e 1 a 10s para janelas de médio prazo. Sobreposição de 50%

Setup utilizado

- Notebook AVELL A40, AMD® Ryzen 5 3500u, 16Gb de RAM utilizando Linux 5.11.0-7612-generic Pop!_OS 20.10_64bits
- Ambiente Python 3.8 e biblioteca para tratamento de sinais sonoros PyAudioAnalysis 0.3.8

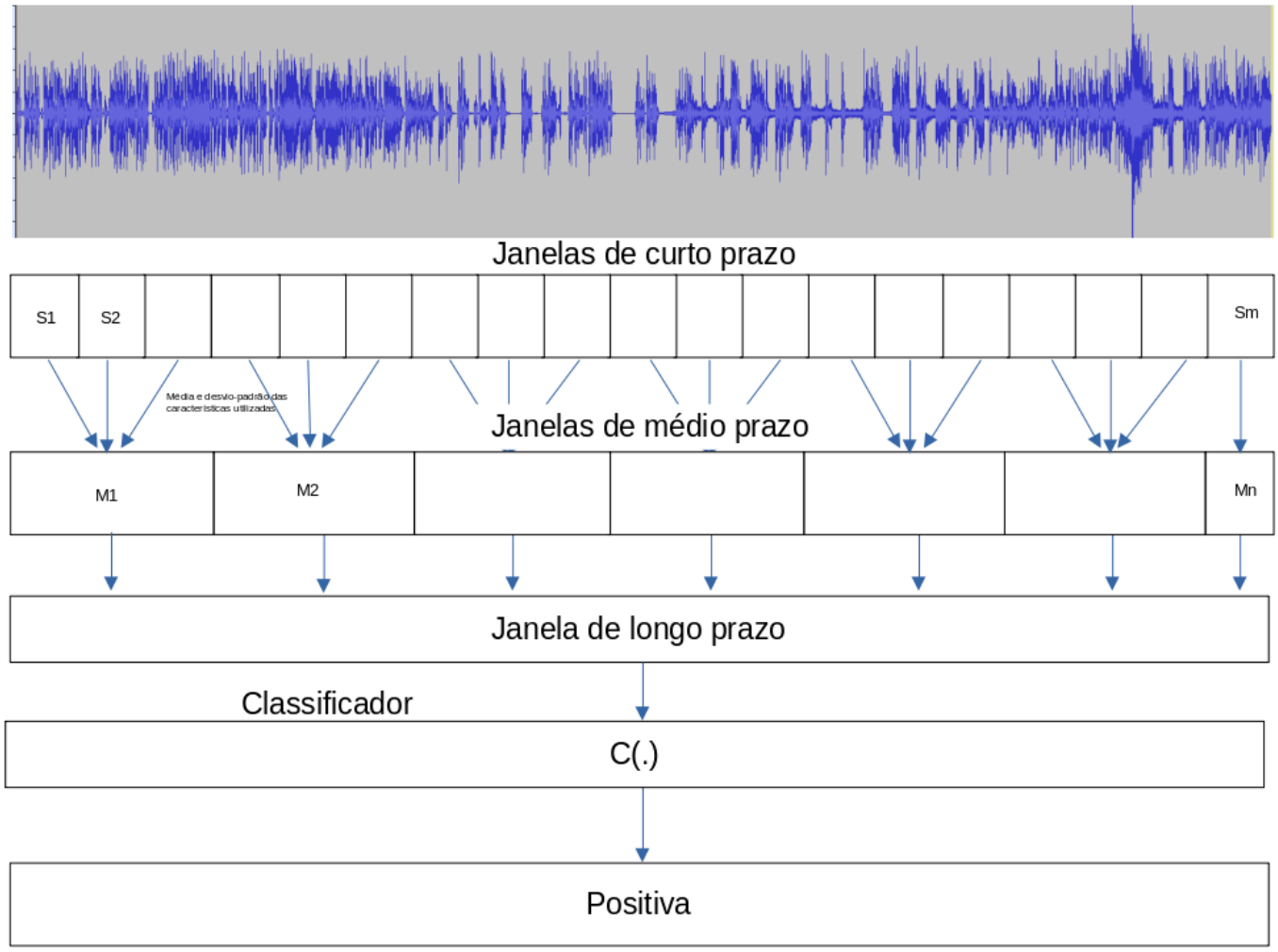
Código fonte utilizado:

```
from pyAudioAnalysis import MidTermFeatures as aF
import os
import numpy as np
from sklearn.svm import SVC
import plotly.graph_objs as go
import plotly
import pandas as pd
from pyAudioAnalysis.audioTrainTest import extract_features_and_train
import datetime

#                sw,    ss, mw, ms
ensaios = (
('Ensaio 0',  [0.02, 0.02 , 1 , 1  ]),
('Ensaio 1',  [0.02, 0.02 , 1 , 0.5]),
('Ensaio 2',  [0.02, 0.02 , 10, 10 ]),
('Ensaio 3',  [0.02, 0.02 , 10, 5  ]),
('Ensaio 4',  [0.02, 0.01, 1 , 1  ]),
('Ensaio 5',  [0.02, 0.01, 1 , 0.5]),
('Ensaio 6',  [0.02, 0.01, 10, 10 ]),
('Ensaio 7',  [0.02, 0.01, 10, 5  ]),
('Ensaio 8',  [0.10, 0.10 , 1 , 1  ]),
('Ensaio 9',  [0.10, 0.10 , 1 , 0.5]),
('Ensaio 10', [0.10, 0.10 , 10, 10 ]),
('Ensaio 11', [0.10, 0.10 , 10, 5  ]),
('Ensaio 12', [0.10, 0.050, 1 , 1  ]),
('Ensaio 13', [0.10, 0.050, 1 , 0.5 ]),
('Ensaio 14', [0.10, 0.050, 10, 10 ]),
('Ensaio 15', [0.10, 0.050, 10, 5  ]),
)

df_resposta = pd.DataFrame(columns =
['Ensaio','sw','ss','mw','ms','acuracia'])

dirs = ["hear/negativos", "hear/positivos"]
class_names = [os.path.basename(d) for d in dirs]

sw, ss, mw, ms = ensaios[1][1]
sw, ss, mw, ms
extract_features_and_train(dirs, mw, ms, sw, ss, "svm_rbf", "Ensaio 1")
```

Exemplo de resposta:

```
                negativos                       positivos               OVERALL
        C       PRE     REC     f1      PRE     REC     f1      ACC     f1
        0.001   57.9    100.0   73.3    50.0    0.0     0.0     57.9    36.7
        0.010   57.9    100.0   73.3    50.0    0.0     0.0     57.9    36.7
        0.500   74.0    92.0    82.0    83.5    55.5    66.7    76.6    74.3
        1.000   79.9    89.0    84.2    82.0    69.1    75.0    80.6    79.6    best f1         best Acc
        5.000   80.2    84.5    82.3    77.0    71.4    74.1    78.9    78.2
        10.000  79.2    81.6    80.4    73.6    70.5    72.0    76.9    76.2
        20.000  80.8    83.1    81.9    75.8    72.9    74.3    78.8    78.1
Confusion Matrix:
        neg     pos
neg     51.53   6.37
pos     13.00   29.11
Selected params: 1.00000
[ 16:14:20 ] Fim

real    33m11,806s
user    33m7,512s
sys     0m2,148s
```

Utilizou-se a função `extract_features_and_train` do pyAudioAnalisys. Ela, por sua vez, utiliza a `multiple_directory_feature_extraction`, que utiliza a `directory_feature_extraction`, que retorna um único vetor com um valor médio e o desvio-padrão de cada *feature* calculada para cada arquivo de áudio, conforme Figura abaixo:



```
def directory_feature_extraction(folder_path, mid_window, mid_step,
                                 short_window, short_step,
                                 compute_beat=True):
    """
    This function extracts the mid-term features of the WAVE files of a
    particular folder.

    The resulting feature vector is extracted by long-term averaging the
    mid-term features.
    Therefore ONE FEATURE VECTOR is extracted for each WAV file.

    ARGUMENTS:
        - folder_path:        the path of the WAVE directory
        - mid_window, mid_step:    mid-term window and step (in seconds)
        - short_window, short_step:    short-term window and step (in
seconds)
    """
```

Features utilizadas

Nesta avaliação foram utilizadas todas as *features* fornecidas pelo pyAudioAnalysis. Mais informações sobre elas em https://github.com/tyiannak/pyAudioAnalysis/wiki e literatura relevante.

```
['zcr_mean', 'energy_mean', 'energy_entropy_mean', 'spectral_centroid_mean', 'spectral_spread_mean',
'spectral_entropy_mean', 'spectral_flux_mean', 'spectral_rolloff_mean', 'mfcc_1_mean',
'mfcc_2_mean', 'mfcc_3_mean', 'mfcc_4_mean', 'mfcc_5_mean', 'mfcc_6_mean', 'mfcc_7_mean',
'mfcc_8_mean', 'mfcc_9_mean', 'mfcc_10_mean', 'mfcc_11_mean', 'mfcc_12_mean', 'mfcc_13_mean',
'chroma_1_mean', 'chroma_2_mean', 'chroma_3_mean', 'chroma_4_mean', 'chroma_5_mean',
'chroma_6_mean', 'chroma_7_mean', 'chroma_8_mean', 'chroma_9_mean', 'chroma_10_mean',
'chroma_11_mean', 'chroma_12_mean', 'chroma_std_mean', 'delta zcr_mean', 'delta energy_mean', 'delta
energy_entropy_mean', 'delta spectral_centroid_mean', 'delta spectral_spread_mean', 'delta
spectral_entropy_mean', 'delta spectral_flux_mean', 'delta spectral_rolloff_mean', 'delta
mfcc_1_mean', 'delta mfcc_2_mean', 'delta mfcc_3_mean', 'delta mfcc_4_mean', 'delta mfcc_5_mean',
'delta mfcc_6_mean', 'delta mfcc_7_mean', 'delta mfcc_8_mean', 'delta mfcc_9_mean', 'delta
mfcc_10_mean', 'delta mfcc_11_mean', 'delta mfcc_12_mean', 'delta mfcc_13_mean', 'delta
chroma_1_mean', 'delta chroma_2_mean', 'delta chroma_3_mean', 'delta chroma_4_mean', 'delta
chroma_5_mean', 'delta chroma_6_mean', 'delta chroma_7_mean', 'delta chroma_8_mean', 'delta
chroma_9_mean', 'delta chroma_10_mean', 'delta chroma_11_mean', 'delta chroma_12_mean', 'delta
chroma_std_mean', 'zcr_std', 'energy_std', 'energy_entropy_std', 'spectral_centroid_std',
'spectral_spread_std', 'spectral_entropy_std', 'spectral_flux_std', 'spectral_rolloff_std',
'mfcc_1_std', 'mfcc_2_std', 'mfcc_3_std', 'mfcc_4_std', 'mfcc_5_std', 'mfcc_6_std', 'mfcc_7_std',
'mfcc_8_std', 'mfcc_9_std', 'mfcc_10_std', 'mfcc_11_std', 'mfcc_12_std', 'mfcc_13_std',
'chroma_1_std', 'chroma_2_std', 'chroma_3_std', 'chroma_4_std', 'chroma_5_std', 'chroma_6_std',
'chroma_7_std', 'chroma_8_std', 'chroma_9_std', 'chroma_10_std', 'chroma_11_std', 'chroma_12_std',
'chroma_std_std', 'delta zcr_std', 'delta energy_std', 'delta energy_entropy_std', 'delta
spectral_centroid_std', 'delta spectral_spread_std', 'delta spectral_entropy_std', 'delta
spectral_flux_std', 'delta spectral_rolloff_std', 'delta mfcc_1_std', 'delta mfcc_2_std', 'delta
mfcc_3_std', 'delta mfcc_4_std', 'delta mfcc_5_std', 'delta mfcc_6_std', 'delta mfcc_7_std', 'delta
mfcc_8_std', 'delta mfcc_9_std', 'delta mfcc_10_std', 'delta mfcc_11_std', 'delta mfcc_12_std',
'delta mfcc_13_std', 'delta chroma_1_std', 'delta chroma_2_std', 'delta chroma_3_std', 'delta
chroma_4_std', 'delta chroma_5_std', 'delta chroma_6_std', 'delta chroma_7_std', 'delta
chroma_8_std', 'delta chroma_9_std', 'delta chroma_10_std', 'delta chroma_11_std', 'delta
chroma_12_std', 'delta chroma_std_std']
```

Resultados

Ensaio	Resumo	ShortWindow sw (s)	ShortWindow Step ss (s)	MidWindow mw (s)	MidWindow Step ms (s)	Acurácia do Classificador	Duração
0		0,02	0,02(0%)	1	1 (0%)	82,0%	51m58s
1	ms	0,02	0,02(0%)	1	0,5 (50%)	80,5%	54m25s
2	mw	0,02	0,02(0%)	10	10 (0%)	81,3%	49m38s
3	mw/ms	0,02	0,02(0%)	10	5 (50%)	81,8%	50m43s
4	ss	0,02	0,01 (50%)	1	1 (0%)	82,6%	101m29s
5	ss/ms	0,02	0,01 (50%)	1	0,5 (50%)	81,7%	104m45s
6	ss/mw	0,02	0,01 (50%)	10	10 (0%)	82,7%	99m56s
7	ss/mw/ms	0,02	0,01 (50%)	10	5 (50%)	81,7%	99m47s
8	sw	0,1	0,1(0%)	1	1 (0%)	77,9%	20m33s
9	sw/ms	0,1	0,1(0%)	1	0,5 (50%)	81,2%	22m38s
10	sw/mw	0,1	0,1(0%)	10	10 (0%)	78,6%	17m45s
11	sw/mw/ms	0,1	0,1(0%)	10	5 (50%)	79,6%	18m01s
12	sw/ss	0,1	0,05 (50%)	1	1 (0%)	80,6%	33m11s
13	sw/ss/ms	0,1	0,05 (50%)	1	0,5 (50%)	79,3%	35m46s
14	sw/ss/mw	0,1	0,05 (50%)	10	10 (0%)	79,6%	30m58s
15	sw/ss/mw/ms	0,1	0,05 (50%)	10	5 (50%)	79,7%	31m38s

Cálculos dos efeitos principais

- Mid window step (ms): **0,00%**
- Mid window (mw): **0,00%**
- Short window step (ss): **0,01%**
- Short window (sw): **-0,02%**

Cálculos dos efeitos de interação de dois fatores

- Mid window step (ms) E Mid window (mw): **0,12%**
- Mid window step (ms) E Short window step (ss): **-0,80%**
- Mid window step (ms) E Short window (sw): **0,75%**
- Mid window (mw) E Short window step (ss): **-0,16%**
- Mid window (mw) e Short window (sw): **-0,28%**
- Short window step (ss) e Short window (sw): **-0,15%**

Cálculo dos efeitos de interação dos quatro fatores

- Os quatro fatores: **0,73%**

Cálculo do efeito de interação de três fatores

- Mid window step (ms), Mid window (mw) e Short window step (ss): **0,20%**
- Mid window step (ms), Mid window (mw) e Short window (sw): **-0,48%**
- Mid window step (ms), Short window step (ss) e Short window (sw): **-0,58%**
- Mid window (mw), Short window step (ss) e Short window (sw): **0,10%**

Métricas

- PRIMARY METRIC Acurácia do classificador SVM utilizando os dados extraídos conforme as configurações de janelas informadas
- SECONDARY METRIC Tempo de execução do ensaio, que compreende a extração das features dos áudios e o treinamento do classificador

Results

Experiment start	9 de abr de 2021
Experiment end	14 de abr de 2021

Link to results in Optimizely	
Conclusion	**ALCANÇADA** . As configurações das janelas tem pouca influência (< 1%) no resultado final do classificador, medida por meio da acurácia de um classificador SVM. No entanto, utilizar janelas de curto prazo muito pequenas aumentam bastante o custo computacional de processamento de áudio. Sendo assim, recomenda-se utilizar janelas de curta duração maiores. Numericamente: utilizando as janelas curtas de 0,02s (20ms) fizeram os ensaios durarem, em média 80,10 min (desvio-padrão de 26,76 min) enquanto que com as janelas curtas com 0,1s (100ms) fizeram os ensaios durarem. em média, 26,31 min (desvio-padrão de 7,32 min), o que se configura em um achado relevante.

✨ Conclusões

Highlights

- Primary goal
  - As diferentes configuraçõe de short window, short window step, mid window e mid window step tem pouca influência na acurácia de um classificador SVM.

Takeaways

- Short windows de duração maior (100ms) são preferíveis pois tem custo computacional menor e não interferem significativamente na acurácia do classificador que foi desenvolvido.