Acceleration and Braking Recognition

Felipe Bombardelli

May 7, 2019

1 Introduction

The work aims to recognize the sound of train, when they braking or accelerating, from a WAV file using SVM and Convolution Neural Network (CNN) like classifier. There is no work about train recognizing on the IEEEXplorer, but there some works about environmental sound recognition [1][2].

2 Program Structure

The code of solution was divided on 3 modules for a better organization and reuse of code. The first module called datasetmod is responsible by handle and normalize the dataset. The second, way_classic is responsible by training and evaluation classic classifiers like KNN, Neural Network and SVM. And by last, way_cnn is responsible by training and evaluation classifier using CNN.

3 Dataset and Data Normalize

The dataset has 2996 samples with one channel and rate 22050 Hz divided by 9 classes, as shown in the table 3. How the samples of dataset have different duration, so it needs to divide the samples in little window with same size.

Id	Label	Samples	Average Duration
0	negative/checked	828	9.83
1	accelerating/1_New	493	3.57
2	accelerating/2_CKD_Long	169	3.57
3	accelerating/3_CKD_Short	74	3.56
4	accelerating/4_Old	410	3.38
5	braking/1_New	381	4.58
6	braking/2_CKD_Long	143	4.51
7	braking/3_CKD_Short	62	4.35
8	braking/4_Old	436	3.95

Table 1: Dataset provided

According the work [2] there are some representation used to learning features like short-time Fourier transform (STFT) with linear and Mel scales, constant-Q transform (CQT) and continuous Wavelet transform (CWT). And Mel-STFT, STFT and CQT were consistently good performers [2].

In this work the descriptor used was the Mel-Scaled Spectrogram through the function librosa.feature.melspectrogram() provided by the Librosa library.

3.1 Methodology

The work evaluates 2 CNN architecture, SVM, KNN and Neural Network. For evaluating, the dataset was divided in 70% to train and 30% to test. The first CNN architecture is called in this work by CNN-flat and it similar a Neural Network. The second CNN architecture is called CNN-based-piczak and based on the work [2], which is described in the figure 1. Finally the CNN models were trained using 10 epochs.

It performs some tests to choice the window size and frequency resolution needed to mel-scale. The tests consist in train a CNN-flat varying the resolution frequency and window size, and analyzing their accuracy tested on the 30% of the dataset. Firstly, the window size was fixed in 175(4s), because the 5 classes in the dataset has less than 4s of duration and the frequency resolution was varied from 10 until 40. So the table 3.1 shows, that frequency resolution greater 20 don't improve more the accuracy.

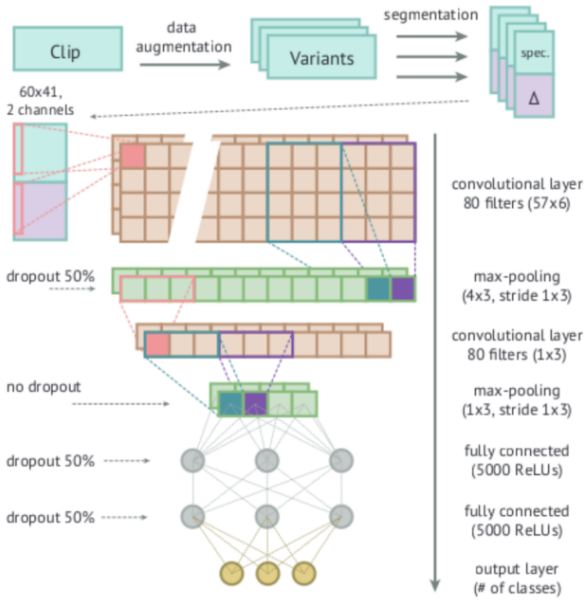

Figure 1: CNN Architecture used in the work [3]

Frequency Resolution	Window Size	Accuracy
10	175 (4s)	0.8254
15	175 (4s)	0.8756
20	175 (4s)	0.8876
25	175 (4s)	0.8869
30	175 (4s)	0.8939
35	175 (4s)	0.8869
40	175 (4s)	0.8939

Table 2: Result of the test varying the frequency resolution

After defined the frequency resolution in 25, also it was realized same tests to define a good window size. So the window size was variable from 50(1.1s) until 200(4.6s). When the sound have duration next to the window size, so the sound is padding or cutting, else the sound is divided in many windows. In the table shows 3.1, that the frequency resolution greater than 150 don't improve more the accuracy. Therefore, the window size adopted was 150.

Frequency Resolution	Window Size	Accuracy
25	50 (1.1s)	0.83
25	75 (1.7s)	0.83
25	100 (2.3s)	0.87
25	125 (2.9s)	0.88
25	150 (3.5s)	0.89
25	175 (4.0s)	0.88
25	200 (4.6s)	0.89

Table 3: Result of the test varying the window size

4 Results and Discussion

The evaluating of the classifiers used 25 frequency resolution and window size of 150 as parameter. So the table show the accuracy obtained for each classifier. The result CNN-flat and Neural Network had the best accuracy.

Label	Accuracy
CNN-flat	0.8933
CNN-based-piczak	0.8251
SVM	0.8718
KNN	0.8379
Neural Network	0.8939

Table 4: Accuracies of the classifiers using frequency resolution=25 and window size=150

Although the result of CNN-flat was 0.89, when the classifier was aplied in the real-world scenario, it don't have apparently the same accuracy, how showed in the figure 2. Because in the same braking or accelerating have multiples classes of the trains.

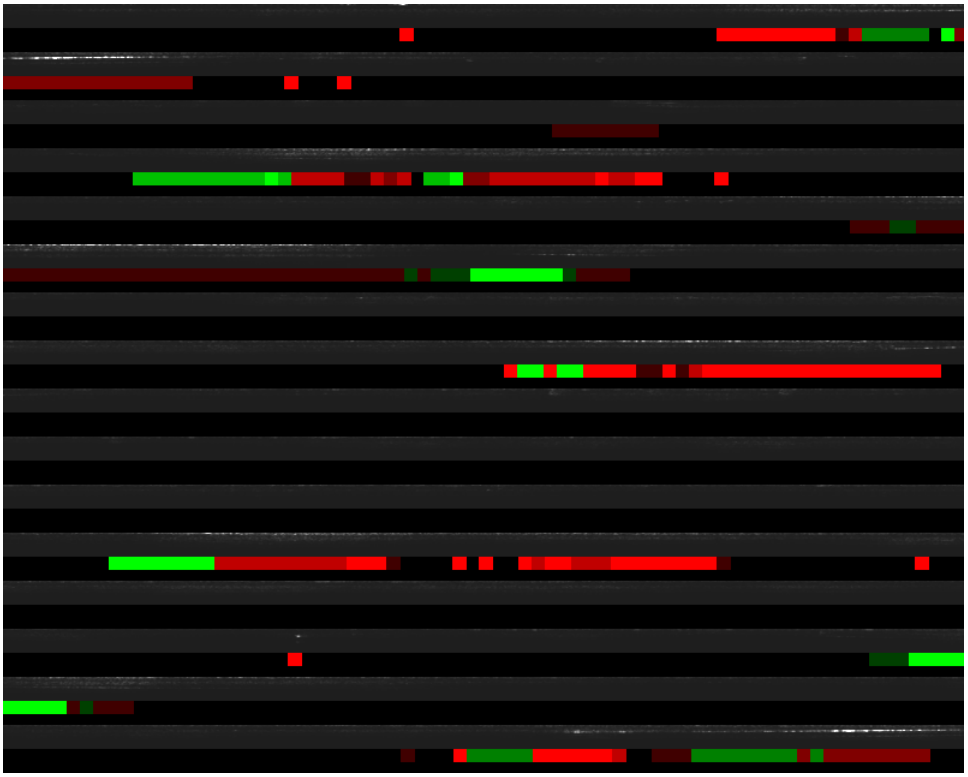

Figure 2: Result using CNN-flat classifier on the file test_files/tram-2018-11-30-15-30-17.wav. Green means a train accelerating and Red means a train braking. The different tons of the greens and red are different classes of the trains.

5 Conclusion

The accuracies on the tests on 30

References

[1] S. Chachada and C. . J. Kuo. Environmental sound recognition: A survey. In *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 1–9, Oct 2013.

[2] Muhammad Huzaifah. Comparison of time-frequency representations for environmental sound classification using convolutional neural networks. *CoRR*, abs/1706.07156, 2017.

[3] K. J. Piczak. Environmental sound classification with convolutional neural networks. In *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6, Sep. 2015.