A BASIC SYSTEM FOR MUSIC GENRE CLASSIFICATION

Enric Guaus
Music Technology Group
Universitat Pompeu Fabra
enric.guaus@iua.upf.edu

Perfecto Herrera
Music Technology Group
Universitat Pompeu Fabra
perfecto.herrera@iua.upf.edu

ABSTRACT

This paper describes the algorithm submitted to the Audio Genre Classification task organized for the MIREX 2007 contest. The algorithm has been designed after many tests using different sets of descriptors, classifiers databases. The purpose of all these tests is to create a plain classifier capable to of dealing with different environments and serving as a baseline for further improvements

1 INTRODUCTION

According to the literature, in the last few years automatic genre classification has recruited many efforts from the MIR community. The most common schema for classifiers (computation of a set of descriptors and training a supervised machine learning algorithm) is also followed in our implementation. In order to decide the parameters for the final implementation, different experiments have been carried out using different descriptors (covering different musical facets), classifiers and labelled databases. After that, we select the schema that is most suitable for all the possible environments. Our implementation is built in C++ and the system does not take advantage of any pretrained model

2 DESCRIPTION

The algorithm has been developed as a set of C++ classes. The final implementation uses three well known C++ libraries: libsndfile (for i/o of audio files), FFTW (for FFT computations) and libSVM (for Support Vector Machine train and test processes). Three bash scripts have been developed to provide compatibility with MIREX specifications.

2.1 Features

As mentioned in Section 1 a set of tests with different descriptors have been performed. Results show how timbre related features provide better results in different environments, followed by rhythmic descriptors. Other descriptors related to musical facets (melody, tonality, tempo, etc.) seem to have less discriminative power in the different databases. Although the accuracies obtained by the rhythm features are about $5..10\%$ lower than those obtained with timbre features, the combination of both descriptor sets increases about 1 to 4% points the overall performance.

2.1.1 Timbre descriptors

The timbre descriptors used in this experiment are not new. As shown in literature, they have proven to be quite robust in automatic classification[1]. In our experiments, we use a set of timbre descriptors comprising: 12 $MFCC$, 12 $\Delta MFCC$, 12 $\Delta^2 MFCC$, Spectral Centroid, Spectral Flatness, Spectral Flux and Zero Crossing Rate. The frame size we use is $92.9ms$ and 50% overlap. The Mean, Variance, Skewness and Kurtosis for all these descriptors are computed for each audio excerpt.

2.1.2 Rhythm descriptors

The rhythmic description used in the experiments is based on the *Rhythm Transformation* proposed in [2]. Although many successful approaches on rhythmic description can be found in literature, this algorithm has proved to be a good and compact representation of rhythm, even for signals such as speech, or for some excerpts of classical music where rhythm is not present at all. This method is based on the periodogram computation of the derivative of the energy for each sub-band of the input signal. This data is compacted in a similar way that MFCC does with the spectrum. The frame size we use is $92.9ms$ and 50% overlap, $1/3rd$ filterbanks and a $3s$ window size to compute rhythm. The Mean, Variance, Skewness and Kurtosis for all these descriptors are computed for each audio excerpt.

2.2 Databases

In order to build a classifier that is capable of dealing with different music collections, our algorithm has been tested in two databases: (1) a particular database defined by some musicologists that uses 8 different musical genres (classical, dance, hiphop, jazz, pop, rhythm&blues, rock, speech), and 50 full songs per genre without artist redundancy. This database is focused on the most common music broadcasted by radiostations. (2) The database

DB	Descriptors	IB1	SVM1	SVM2	AdaBoost	RandomForest
1	Timbre	63.342%	80.299%	81.296%	71.820%	75.062%
1	Rhythm	53.117%	59.850%	62.594%	58.354%	56.608%
1	Timbre + Rhythm	69.576%	82.294%	83.791%	77.057%	74.564%
2	Timbre	80.578%	90.030%	90.030%	83.484%	37.361%
2	Rhythm	45.619%	52.467%	60.020%	57.905%	57.301%
2	Timbre + Rhythm	84.390%	91.239%	90.533%	83.685%	80.765%

Table 1. Results of genre classification for 2 databases and 2 sets of descriptors using 4 classification techniques. Accuracies are obtained using 10-fold cross validation

proposed by Tzanetakis [3] that uses 10 musical genres (blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, rock), 100 audio fragments per genre without artist redundancy. Each audio excerpt is 30 seconds long, 1 channel, WAV at $sr = 22050Hz$

2.3 Classifiers

The most representative classification algorithms we have tested are compared in Table 1, using different descriptors and databases. The best results are obtained using Timbre and Rhythm features and a Support Vector Machine with exp=2 classifier. This is the approach we have implemented for the MIREX.

3 RESULTS

The results we obtain after the MIREX evaluation are the following: Average. Hierarchical Classification Accuracy: 71.87% (best submission: 76.56%). Average Raw Classification Accuracy: 62.89% (best best submission: 68.29%). Run Time for descriptors: $22740s$ (best submission: 6879). Classification: 194 folds (best submission: 51). The confusion matrix is shown in Figure 1.The most relevant confusions in our implementation are (in descending order): (1) Baroque, Classical and Romantic, (2) Blues and Jazz, (3) Rock'n'Roll and Country and (4) Dance and Rap-HipHop. All these confusions are musically coherent with the selected taxonomy, which is not the same taxonomy used in our previous experiments. Results are quite close to the best submission using this basic approach (less than 5% below). The Rap-HipHop category is our best classified genre (as in the other approaches) while Rock'n'Roll is our worst classified genre (as in other 3 approaches, but others show worse results for Dance, Romantic, Classical).

4 CONCLUSIONS

In this paper, we have reported our approach submitted to the MIREX 2007 contest. Similar types of confusion between genres have been found for most of the participants. Our plain classifier is not so far from the best submitted classifier. SVM seems to be one of the best techniques for this task and the feature computation seem to be crucial in the overall accuracy. Further research should deal with

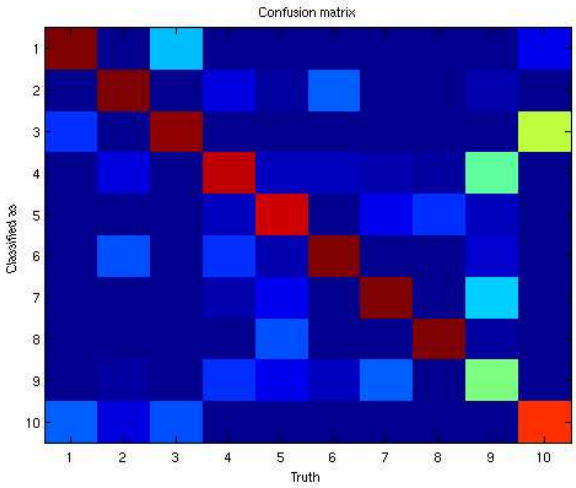

Figure 1. Table 2: Confusion matrix of the classification results: 1:Baroque, 2:Blues, 3:Classical, 4:Country, 5:EDance, 6:Jazz, 7:Metal, 8:Rap-HipHop, 9:Rock'n'Roll, 10:Romantic

new descriptors that are able to represent the particularities of each musical genre.

5 ACKNOWLEDGEMENTS

This work was partially funded by the SALERO (IST FP6-027122, http://www.salero.eu) and CANTATA (ITEA 05010) projects. We also thank George Tzanetakis for kindly providing the database used in [3].

6 REFERENCES

[1] Logan B, "Mel Frequency Cepstral Coefficients for Music Modeling", *International Symposium on Music Information Retrieval*, 2000.

[2] Guaus E, "New approaches for rhythmic description of audio signals", *Doctoral Pre-Thesis Work*, UPF, 2004.

[3] Tzanetakis G, Cook P, "Musical Genre Classication of Audio Signals", *IEEE Transactions on Speech and Audio Processing*, 10(5), 2002.