MIREX AUDIO TAG CLASSIFICATION

Konstantinos Trohidis
Dept. of Journalism & Mass Communication
Aristotle University of Thessaloniki
trohidis2000@yahoo.com

Grigorios Tsoumakas
Dept. of Informatics
Aristotle University of Thessaloniki
greg@csd.auth.gr

George Kalliris
Dept. of Journalism & Mass Communication
Aristotle University of Thessaloniki
gkal@auth.gr

Ioannis Vlahavas
Dept. of Informatics
Aristotle University of Thessaloniki
vlahavas@csd.auth.gr

ABSTRACT

This extended abstract details a submission to the Music Information Retrieval Evaluation eXchange in the audio tag classification task, a new task introduced this year. We model the problem as a multilabel classification task and employ suitable learning algorithms from the Mulan toolkit [1] .

1 INTRODUCTION

Traditional *single-label* classification is concerned with learning from a set of examples that are associated with a single label λ from a set of disjoint labels L, $|L| > 1$. In *multilabel* classification, the examples are associated with a set of labels $Y \subseteq L$.

Multilabel classification methods can be categorized into two different groups [3]: i) *problem transformation* methods, and ii) *algorithm adaptation* methods. The first group contains methods that are algorithm independent. They transform the multilabel classification task into one or more single-label classification, regression or ranking tasks. The second group contains methods that extend specific learning algorithms in order to handle multilabel data directly.

The submitted system uses one algorithm from each category; the RAKEL [4] problem transformation algorithm for outputting binary predictions, and the MLkNN [6] algorithm adaptation algorithm for outputting real valued affinity scores, for each tag and audio clip pair. Algorithms are obtained from the Mulan toolkit. An experimental comparison of such algorithms for a different task (emotion/mood detection in music) can be found in [2].

2 FEATURE EXTRACTION

For the feature extraction process, the Marsyas tool [5] was used. The extracted features fall into two categories: rhythmic and timbre.

[1] http://mlkd.csd.auth.gr/multilabel.html

2.1 Rhythmic Features

The rhythmic features were derived by extracting periodic changes from a beat histogram. An algorithm that identifies peaks using autocorrelation was implemented. We selected the two highest peaks and computed their amplitudes, their BMPs (beats per minute) and the high-to-low ratio of their BPMs. In addition, 3 features were calculated by summing the histogram bins between 40-90, 90-140 and 140-250 BPMs respectively. The whole process led to a total of 8 rhythmic features.

2.2 Timbre Features

Mel Frequency Cepstral Coefficients (MFCCs) are used for speech recognition and music modeling [1]. To derive MFCCs features, the signal was divided into frames and the amplitude spectrum was calculated for each frame. Next, its logarithm was taken and converted to Mel scale. Finally, the discrete cosine transform was implemented. We selected the first 13 MFCCs.

Another set of 3 features that relate to timbre textures were extracted from the Short-Term Fourier Transform (FFT): Spectral centroid, spectral rolloff and spectral flux.

For each of the 16 aforementioned features (13 MFCCs, 3 FFT) we calculated the mean, standard deviation (std), mean standard deviation (mean std) and standard deviation of standard deviation (std std) over all frames. This led to a total of 64 timbre features.

3 REFERENCES

[1] B. Logan. Mel frequency cepstral coefficients for music modeling. In *Proceedings of the 1st International Symposium on Music Information Retrieval (ISMIR 2000)*, Plymouth, Massachusetts, 2000.

[2] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas. Multilabel classification of music into emotions. In *Proc. 9th International Conference on Music Information Retrieval (ISMIR 2008), Philadelphia, PA, USA, 2008*, 2008.

[3] G. Tsoumakas and I. Katakis. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining*, 3(3):1–13, 2007.

[4] G. Tsoumakas and I. Vlahavas. Random k-labelsets: An ensemble method for multilabel classification. In *Proceedings of the 18th European Conference on Machine Learning (ECML 2007)*, pages 406–417, Warsaw, Poland, September 17-21 2007.

[5] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, July 2002.

[6] M-L Zhang and Z-H Zhou. Ml-knn: A lazy learning approach to multi-label learning. *Pattern Recognition*, 40(7):2038–2048, 2007.