AN ENSEMBLE METHOD FOR MIREX AUDIO TAG CLASSIFICATION

Hung-Yi Lo, Ju-Chiang Wang, Hsin-Min Wang
Institute of Information Science, Academia Sinica, Taipei, Taiwan
{hungyi, asriver, whm}@iis.sinica.edu.tw

ABSTRACT

This paper describes our submitted method for MIREX 2009 audio tag classification task. We extract features to capture different musical information aspects, including dynamics, rhythm, timbre, pitch, tonal information. An ensemble classifier for each tag is exploited to predict on the testing music clips.

1. INTRODUCTION

This MIREX 2009 audio tag classification task aims to compare various algorithms' abilities to associate tags with 10-second audio clips of songs. The tags are metadata which describe different types of musical information such as genre or instrument and can be used for music recommendation. The audio clips and tags in MIREX audio tag classification task come from the MajorMiner[1] website. This paper presents our submitted method.

2. FEATURE EXTRACTION

The audio features of each 10-sec music clip are extracted with and without segmentation. We apply an unsupervised method, "Novelty Curve", which was proposed from [1] for segmentation. The features we used can be categorized as five fields: dynamics, rhythm, timbre, pitch and tonality. In **dynamics** field, we extract rms; the **rhythm** field consists of (1) the peak and centroid of the fluctuation summary, (2) tempo, and (3) attack slop and time of the onset; the **timbre** field contains (1) zero-crossing rate, (2) spectral centroid, spread, skewness and kurtosis, (3) brightness, (4) rolloff with 95% threshold, (5) rolloff with 85% threshold, (6) spectral entropy and flatness, (7) roughness, (8) irregularity, (9) inharmonicity, (10) MFCCs, delta-MFCCs, and delta-delta-MFCCs, (11) low energy rate, and (12) spectral flux; in **pitch** field, we extract (1) pitches, (2) chromagram and its centroid and highest peak; and finally the **tonality** field includes (1) key clarity, (2) key mode, and (3) harmonic change. For training, we calculate the statistics including mean and standard variation of the features if the feature is frame-decomposed. The feature extractor is implemented based on MIRToolbox 1.1[2], which is a free and robust tool for MIR tasks. The detail description of the audio features we used can be comprehended from the user's manual of MIRToolbox [2].

3. CLASSIFICATION METHOD

We treat each tag prediction as a binary classification problem independently. Our predictive model is a classifier ensemble which consists of two classifiers: L2-loss linear support vector machine (SVM) and AdaBoost with decision stump as its base learner. The advantages of these two classifiers are the training efficiency and less parameter needed to be turned. For linear SVM, we only need to determine the cost parameter *C*. For AdaBoost, we have to determine the number of the base learners. These parameters are selected for each tag using cross-validation on training data. We select parameters according to the area under ROC curve since it is suitable for imbalanced binary classification problem.

Literature indicates that combining divergent but high performance models into an ensemble usually leads to a better generalization performance. We predict on the testing set using each classifier, and then transform their scores into their related ranks. The final score of an instance comes from the average of the related ranks from two classifiers' output. For the tag ranking task, we normalize the average rank into [0, 1] as its tag affinity score. For the tag classification task, the threshold is verified by using each tag's class distribution probability which is obtained from training data.

4. REFERENCES

[1] J. Foote and M, Cooper: "*Media Segmentation using Self-Similarity Decomposition*", Proceedings of SPIE Storage and Retrieval for Multimedia Databases, Vol. 5021, pp. 167-75, 2003.

[2] O. Lartillot, P. Toiviainen and T. Eerola: *MIRtoolbox 1.1 User's Manual*, Finnish Center of Excellence in Interdisciplinary Music Research, University of Jyväskylä, Finland, 2008.

[1] http://majorminer.org/

[2] https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox