Music Genre Classification

I. Problem Description

Music genre classification is widely discussed in the MIR (Music Information Retrieval) Society and has been of interest for a variety of reasons, including management of large music collections. As these collections are becoming more digital, genre classification is essential for creating recommendation systems, that can benefit collective music libraries, social networks, etc. Only limited agreement is achieved among human annotators when classifying music by genre and an automatic genre classifier aims to overcome those limitations by defining a standard for extracting features directly from the audio signal for classification. It is important that we extract the most robust set of features and establish a reliable ground truth for classifying into genres.

II. Related Work

[1] describes a neural network approach that is trained for the classification tasks to determine the genre of audio files. Feature vectors are extracted using Fourier Transforms and fed into a neural network classifier. The algorithm - a linear vector quantization network - in this paper can only distinguish between four genres. [2] also uses neural networks for classification, specifically feed-forward neural networks. Their approach is limited to three genres as classification accuracy would drop significantly below the mentioned 72% if that number was increased. [3] applies Gaussian classifiers and Gaussian mixture models. They present a hierarchy of musical genres and an elaborate section on feature extraction. Yet their classification results in only 61% accuracy over ten genres. [4] describes an approach using high-level melodic features for their classification. Various algorithms are compared including support vector machines, random forests, k-nearest neighbour networks and Bayesian networks. Recognition rates of over 90% are reported. This approach though requires the existence of a melody in an audio file, which is not the case for all genres.

III. Implementation

The idea of this project is to implement and compare various algorithms for genre classification. We intend to implement a Gaussian classifier, a Bayesian classifier, a k-nearest neighbor classifier and a classifier using support vector machines. The dataset[1] we will use offers readily extracted feature vectors from a selection of songs of six different genres. Whereas it is not clear that this feature set is ideal for training classifiers, we will build our initial system using these feature vectors to train and evaluate our approach. This basic system will then be extended to accept general feature vectors as input that we intend to extract from audio files directly. We will therefore implement a feature extraction algorithm that operates directly on audio files (similar to [4]). Feature extraction and classification using statistical learning methods lends itself especially well to the feature-rich domain of audio signals. As shown in [3] or [4] there is a variety of different approaches to extracting features and training classifiers. Whereas the mentioned papers focus on the feature extraction part, we will focus on learning, implementing, and comparing different classification algorithms with the expectation that we will be able to use them for related research.

[1] Dataset courtesy of http://modelai.gettysburg.edu/2012/music/

IV. Evaluation

We will evaluate the accuracy of the classification algorithms across two additional datasets; GTZAN[2] which consists of ten genre classes, each genre class containing 100 audio recordings 30 seconds long and the ISMIR 2011 dataset[3] that contains 1458 full audio recordings distributed over six genre classes. We will calculate the classification accuracies as the mean accuracies obtained by 10-fold stratified cross-validation on the full datasets. We will also calculate the corresponding standard deviation. The resulting classification accuracies will be compared to those in [4] (both the GTZAN and ISMIR datasets are used in this paper). This will help us understand and highlight the strengths and weaknesses of each as applied to the genre classification problem and the results will be useful in evaluating different feature extraction sets. We expect a mixture of low-level (such as frequency and beat) and high-level features (such as melody) to perform best in terms of accuracy. Specifically we expect our approach to outperform [4] as it will not depend on the occurence of a melody in an audio file.

References

[1] M. Talupur, S. Nath, and H. Yan, "Classification of music genre," 2001. [Online]. Available: http://www.cs.cmu.edu/~yh/files/GCfA.pdf

[2] A. Galkin, P. Rege, and R. Pocratsky, "24787 artifical intelligence and machine learning for engineering design - classficiation of music using neural networks," 2011. [Online]. Available: agalkin.com/media/music/proj_music_report.pdf

[3] G. Tzanetakis and P. R. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[4] J. Salamon, B. Rocha, and E. Gómez, "Musical genre classification using melody features extracted from polyphonic music signals," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, 25/03/2012 2012. [Online]. Available: files/publications/SalamonRochaGomezICASSP2012.pdf

[2] marsyas.info/download/data_sets

[3] http://labrosa.ee.columbia.edu/millionsong/ismir2011