Music Genre Classification

Madhura Dole
Department of Computer Science
The University of Texas at Dallas
Email: mxd1624430@utdallas.edu

Teja Mukka
Department of Computer Science
The University of Texas at Dallas
Email: txm162230@utdallas.edu

Abstract

Analyzing music audio files based on genres and other qualitative tags is an active field of research in machine learning. When paired with respective classification algorithms, most notably support vector machines (SVMs) and k-nearest-neighbor classifiers (KNNs), certain features, including Mel-Frequency Cepstral Coefficients (MFCCs), Chroma attributes and other spectral properties, have been shown to be effective features for classifying music by genre. In this study, we apply these methods and features across two different combinations of GTZAN dataset. By experimenting different classification techniques on these combinations, we analyze the correlation amongst datasets (if exists), and how this affects the percentage of correctly classified music files.

1. Introduction

Classification is one of the basic important tasks of Machine Learning. Some of the common examples involve E-mail classification, news article classification, music genre classification etc. This study attempts to present performance evaluation of various classification methods for Music Genre classification with supervised learning approach. With this being a category of multi-class classification, the correlation amongst different classes plays an important role in the task of classification. Music can be classified into numerous categories based on its genre, but effective classification into its intended category is the main purpose. Although, no classifier can correctly classify all music samples based on its different features, we try to present the effect of training these classifiers on different combination of music datasets.

2. Dataset

For this project, the dataset used is GTZAN, which is a collection of 1000 tracks of 10 different genres which has approximately 100 tracks per genre. It consists of different genres like jazz, classic, reggae, blues, pop, metal, hip-hop etc. Reason for choosing this dataset is that, this is most standard dataset used for music classification. One of the crucial step in the whole process is to extract the right features that can help us in distinguishing all these genres with the best prediction rate. Different features extraction methods were used to extract the features based on different timbre features like MFCC Mel Frequency Cepstral Coefficients, rhythmic features as well as RMS features.

3. Related & Proposed Work

Existing research and work in the field of Music Genre classification has been primarily focused on training and testing of various machine learning classifiers on music files of various genres. Our work is mostly inspired by the project "'Music Genre Classification', 2011, Stanford, Michael Haggblade, Yang Hong, Kenny Kao". With this as a base, we propose here the classification of music files using Machine Learning classifiers such as Support Vector Machines(SVM), Logistic Regression, Decision Trees, AdaBoost, KNN, Perceptron etc. We primarily intend to analyze and compare the effect of training these classifiers on different combinations of genres and the evaluation of their performance. The result of classification not only depends of features used for training a classifier but also on correlation amongst data of different classes.

Another point we tried to convey in the work we have done is to show how the correlation between the genres

affect the classification accuracy and how the best accuracies are available in genre classification of four specific genres.
We have compared the performance of various algorithms using various feature extraction methods, and compared the genre classification for different number of genres selected randomly.

4. Feature Extraction

Feature extraction is the most crucial task in classification. The accuracy of correctly classifying a data mainly depends on the data that we are using (covering all the possible cases) and features extracted from that data to classify each instance.

1. Mel Frequency Cepstral Coefficients (MFCC)

In sound processing, the mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency.
Here we generate MFCCs from a time series and use it in the form of an array to create feature vector for training on different classifiers.

2. Room-Mean-Square (RMS) Energy

Here we compute root-mean-square(RMS) energy for each frame from audio samples. Computing the energy from audio samples is faster than computing from spectrogram as it does not require an STFT calculation.

3. Chroma Vector

Chroma-based features, which are also referred to pitch class profiles, are a powerful tool for analyzing music whose pitches can be meaningfully categorized (often into twelve categories) and whose tuning approximates to the equal-tempered scale. One main property of chroma features is that they capture harmonic and melodic characteristics of music, while being robust to changes in timbre and instrumentation.

5. Classification

After generating feature vectors, the next step is to train different classifiers on these feature vectors to generate different models. Here we divided the dataset into training, validation and testing in a ratio of approximately 60%, 15% and 25% respectively. From a finite set of music genres $\mathcal{G}$ we must select one class $\hat{g}$ which best represents the genre of the music associated to the signal S. From a statistical perspective, the goal is to find the most likely $\hat{g} \in \mathcal{G}$, given the feature vector $\bar{X}$, that is

$$\hat{g} = \arg\max_{g \in G} P(g|\bar{X})$$

where P($\bar{X}$|g) is the probability in which the feature vector X occurs in class g, P(g) is the a priori probability of the music genre g (which can be estimated from frequencies in the database) and P($\bar{X}$) is the probability of occurrence of the feature vector $\bar{X}$. The last probability is in general unknown, but if the classifier computes the likelihoods of the entire set of genres, then $\sum_{g \in \mathcal{G}} P(g|\bar{X}) = 1$ and we can obtain the desired probabilities for each g ∈ $\mathcal{G}$ by

$$P(g|\bar{X}) = \frac{P(\bar{X}|g).P(g)}{\sum_{g \in \mathcal{G}} P(\bar{X}|g).P(g)}$$

Below experimented with classification methods for this project:

1. Support Vector Machines(SVM)
2. Logistic Regression
3. Decision Trees
4. AdaBoost
5. KNN
6. Perceptron

Support vector machines (SVM) have proven to be effective in a variety of classification problems. The idea behind SVMs is to project data onto a higher dimensional space to separate classes with a LDF, which maximizes the margin between competing classes. The software package SVM light was used for training and classification. The inputs into the SVMs are the feature vectors created using feature extraction methods mentioned above (one feature vector per song). The output of this classifier is a score, with a positive value indicating one class and a negative value indicating the other class. However, the song genre problem is multi-category problem where each song is assigned to the most likely genre.
KNN algorithm utilizes audio feature information of a music's k-nearest neighbor to infer its genres. We experimented with multiple values for k.

We used AdaBoost algorithm to fit a sequence of weak learners on different combinations of dataset.

6. Evaluation Measures

Three measures of performance are often used: precision, recall, and accuracy. Precision and recall are useful measures for comparing individual genres because they are independent to the number of examples that may exist in each genre. We have plotted the confusion matrices for all the classifiers and calculated the precision, recall and other metrics from the confusion matrix. The formulas for these two performance measures are

$$precision = \frac{tp}{tp + fp}$$
$$recall = \frac{tp}{tp + fn}$$
$$accuracy = \frac{tp}{N_t} bh$$

where tp, fp, fn, and N_t are the true positives, false positives, false negatives, and total number of test queries, respectively.

7. Results

Music Data Classification

Classifier	Classical (%)	Jazz (%)	Metal (%)	Pop (%)	Total Accuracy (%)
Perceptron	80	90	70	96.66	84.1
SVM	86.66	86.66	80	93.33	86.66
Logistic Regression	83.33	86.66	76.66	96.66	85.8
KNN	60	56.66	76.66	80	68.33
Decision Tree	61	57	74	78	67.003
AdaBoost	79.6	89.4	68	95.6	83.25

Music Data Classification
(Datasets: Classical, Country, Disco, Jazz, Reggae)

Classifier	Total Accuracy (%)
Perceptron	69.99
SVM	69.33
Logistic Regression	69.33
KNN	49.33
Decision Tree	23.15
AdaBoost	34.29

Logistic Regression Confusion Matrix

Classifier	Classical	Jazz	Metal	Pop	Recall
Classical	25	1	3	1	83.33
Jazz	1	26	3	0	86.7
Metal	0	3	23	4	76.6
Pop	0	1	0	29	86.7
Precision	96.1	83.8	79.31	85.24	

SVM Confusion Matrix

Classifier	Classical	Jazz	Metal	Pop	Recall
Classical	26	1	2	1	86.7
Jazz	0	27	3	0	90
Metal	0	3	25	2	83.33
Pop	0	1	1	28	93.33
Precision	100	84.3	80.61	90.32	

Perceptron Confusion Matrix

Classifier	Classical	Jazz	Metal	Pop	Recall
Classical	24	0	5	1	80
Jazz	0	27	3	0	90
Metal	1	6	21	2	70
Pop	0	1	0	29	97
Precision	96	79.41	72.4	90.62	

KNN Confusion Matrix

Classifier	Classical	Jazz	Metal	Pop	Recall
Classical	18	3	5	4	60
Jazz	1	17	5	7	56.66
Metal	1	3	23	3	76.7
Pop	2	0	4	24	80
Precision	81.81	73.91	62.1	64.8	

AdaBoost Confusion Matrix

Classifier	Classical	Jazz	Metal	Pop	Recall
Classical	26	0	3	1	86.9
Jazz	0	27	2	1	90
Metal	1	4	22	3	73.33
Pop	0	1	0	29	96.66
Precision	96.29	84.375	81.48	85.29	

With MFCC features, we get very good accuracy results whereas, RMSE does not perform that well, as we get limited number of feature vectors for each file, and it becomes very difficult to fit a better model to perform better than this. With Chroma Vectors, we get accuracies in the range of 40-50% only. The best chosen feature vector (MFCC) results are shown above.

8. Future Work

Currently it takes good amount of time to preprocess all the genre files and run it across various classifiers. The optimal way of doing it would be extract the features using various methods and store them in a database storage, so that it can be accessed easily with any classifier that we want to extend in the future. Another better approach of extracting features is to use a mix of couple of methodologies and combine the features and use them to predict the genre classification possibly with better accuracies and precision.
Since genre classification between fairly different genres is quite successful, it makes sense to attempt finer classifications. The exact same techniques used in this project could be easily extended to classify music based on any other labelling, such as artist. In addition, including additional metadata text features such as album, song title, or lyrics could allow us to extend this to music mood classification as well.

9. Conclusion

We dismissed Neural Network and KMeans classifiers as being comparatively ineffective genre classifiers with a spectral feature set. We concluded that to increase the performance of the Neural networks in this scenario, further pre-processing like normalization of data etc. would be needed. Qualitatively, spectral features are likely not independent. On the other hand, SVMs and KNNs do not rest on independence assumptions, explaining why they performed relatively well. These results corroborate past studies in machine learning genre classifiers.
The prediction of Classical genre has the highest precision rate, which clearly concludes that the it overlaps very minimal with the other genres. Most learning algorithms faced difficulty differentiating between Metal and Jazz, except k-Means which had the most difficulty differentiating between Classical and Jazz. This corroborates the idea that qualitatively these genres are the most correlated.

10.Summary

In general, the classification task is very subjective and everyone has their view on how the features are to be extracted and processed for analysis. Here in the project, we got the best results using the MFCC features. We have run the classification of the four specific genres which are very commonly referenced in various papers and we were able to match their accuracies. The selected genres are classical, jazz, metal and pop. In the project, we compared the performances of various pre-processing techniques and concluded that the MFCC feature extraction is the best amongst them. In addition to these, we have chosen different combinations of genres and explored how the accuracies varies with various combinations. We reported the accuracies on the combinations of genes that are most commonly used in the literature and reported the accuracies, confusion matrices, precision, recall etc. for the given algorithm and the feature extraction.

11.References

1. F. Pachet and D. Cazaly. “A Taxonomy of Musical Genres,” in Proc. Of Content-Based Multimedia Conf. on Information Access Conf, April 2000.
2. Omar Diab, Anthony Mainero, Reid Watson. “Musical Genre Tag Classification with Curated and Crowdsourced Datasets,” Stanford University, Computer Science, 2012.
3. Michael Haggblade, Yang Hong, Kenny Kao. “’Music Genre Classification’, Stanford University, 2011”
4. K. West and S. Cox. “Features and Classifiers for the Automatic Classification of Musical Audio Signals,” in ISMIR 2004 Fifth Conf. on Music Inf. Retr. Proc., 2004.
5. A. Meng and J. Shawe-Taylor. “An Investigation of Feature Models for Music Genre Classification Using the Support Vector Classifier,” in ISMIR 2005 Sixth Conf. on Music Inf. Retr. Proc., 2005, pp. 604-609.
6. Rachel Mellon, Dan Spaeth, Eric Theis. “Genre Classification Using Graph Representations of Music,” Stanford University, 2014.
7. L.R. Rabiner. “A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition,” in Proc. Of IEEE, Iss. 2, vol. 77, pp. 257-286, Feb. 198