
Music Genre Classification

Yuchi Tian

Department of Computer Science
University of Virginia
yt8mn@virginia.edu

Abstract

For the project of CS 6316 Machine Learning, I work on music genre classification. Music genre classification is common and significant task for MIR (Music Information Retrieval) and plays an important roles in MER (Music Emotion Recognition) which has been a prevalent study area in recent years. Music genre classification is more fundamental for MIR than music emotion recognition and has generally accepted genre tags and categorical description. Music genre classification and music emotion recognition have shared challenges and methodology, so studying on the former will shed light on the latter.

In this project, I extract low level music features with the tools “yaafe” and “librosa”, both of which are implemented in python. And I separate these low level music features into two groups, perform feature selection separately and use SVM classifier to perform the initial classification. I extract a mid level music feature “chord sequence” by HMM Viterbi, which was proposed by a paper. I define the distance as $1/\text{Length}(\text{longest common subsequence})$ and use k nearest neighbors method to improve the initial classification performance. The experimental results show that chord sequence can effectively distinguish jazz songs from country songs and improve the genre classification accuracy. Finally, the classification accuracy reaches to 0.76.

1 Introduction

Music information retrieval (MIR) is the interdisciplinary science of retrieving information from music. It involves the combination of musicology, psychology, signal processing and machine learning. Music genre classification and music emotion recognition are two important fields in MIR. In many commercial music websites such as Last.fm or music companies such as AMG, some studies in these two fields have been applied for music collection, categorization and recommendation. A lot of researchers work on music genre classification and music emotion recognition and many machine learning techniques have been applied in these two fields [1].

In recent years, MER has been a prevalent research field. Some machine learning algorithms have been applied to learning the relationship between music features and emotion labels. Besides the machine learning algorithms, the conceptualization of emotion and the associated emotion taxonomy have been a considerable challenge for MER. [2] There is still no consensus on emotion taxonomy. It is not sensible to spend much time on the emotion conceptualization and taxonomy which deviates the track of machine learning course. So I decide to focus on music genre classification in the project.

For music genre classification, there are generally accepted genre taxonomy and tagging [3]. And music genre classification is a fundamental musical information retrieval process, which has been applied to MER. In a study on MER, different models are trained for associated music genre, and music emotion is predicted by a selection of different models [4]. There are many different kinds of music features for music genre classification, which are shown in Table 1 [5]. The majority of

previous studies on music genre classification focus on only low level music features. But mid level features[6] such as chord sequence[7] and instruments[8] are rarely discussed. A chord, in music theory, is any harmonic set of three or more notes which is played simultaneously. A chord is often presented with a letter and major (or minor). In my project, each chord is associated to a value from 0 to 23, which represents the chord (C, C#, D, D#, E, F F#, G, G#, A, A#, B, c, c#, d, d#, e, f f#, g, g#, a, a#, b).

In my project, I use GTZAN dataset[10] for music genre classification experiments. I extract low level music features with two python tools “yaafe”[27] and “librosa”[29]. I separate low level music features into two groups and perform feature selection separately. Then I use SVM classifier and these low level music features to perform the initial classification. I use HMM viterbi[31-33] to predict the chord sequence and after I pre-process the predicted chord sequence I use k nearest neighbors to perform further classification. The experimental results show that chord sequence can effectively distinguish jazz songs from country songs and thus improve the accuracy.

The following sections are organized as follows. In the second section, I introduce the previous studies on music genre classification. I explain why this project is related to machine learning in the third section and describe the proposed method and experimental design in my project in the forth section. In the following three sections, I describe the dataset, parameters selection and discuss the experimental results. And in the last section, I conclude this paper and discuss the future work for this project.

Table 1: Music Features

Music Features	Description
Energy	Dynamic loudness, Audio power, Total loudness, Specific loudness sensation coefficients
Rhythm	Beat histogram, Rhythm pattern, Rhythm histogram, Rhythm strength, Rhythm regularity, Rhythm clarity, Average onset frequency, Tempo, Average tempo
Tempora	Zero-crossings, Temporal centroid, Log attack time
Spectrum	Spectral centroid, Spectral rolloff, Spectral flux, Spectral flatness measures, Spectral crest factors, Spectral contrast, Daubechies wavelets coefficient histogram, Tristimulus, Even-harm, Odd-harm, Mel-frequency cepstral coefficients
Harmony	Salient pitch, chromagram centroid, key clarity, musical mode, harmonic change, Pitch histogram, Sawtooth waveform inspired pitch estimate

2 Previous Solutions

The machine learning algorithms that have been used in previous solutions are shown in Table 2[9].

Table 2: machine learning techniques

Algorithms	Studies
K-Nearest Neighbor (KNN):	[10,11]
Gaussian Mixture Models (GMM):	[10]
Hidden Markov Model (HMM):	[7,12]
Linear Discriminant Analysis (LDA):	[13,14]
Support Vector Machines (SVM):	[13,15,16,17]
Fuzzy Clustering	[15]
Combination of Classifiers	[18]
Sparse Representation-based Classification	[25]

3 Why is this related to machine learning

MRI is an interdisciplinary science which involves musicology, psychology, signal processing and machine learning. Music genre classification is an important part of MRI and many machine learning techniques have been applied to automatic musical similarity detection and music genre classification.

4 Proposed Method and Experimental Design

4.1 Overall solution

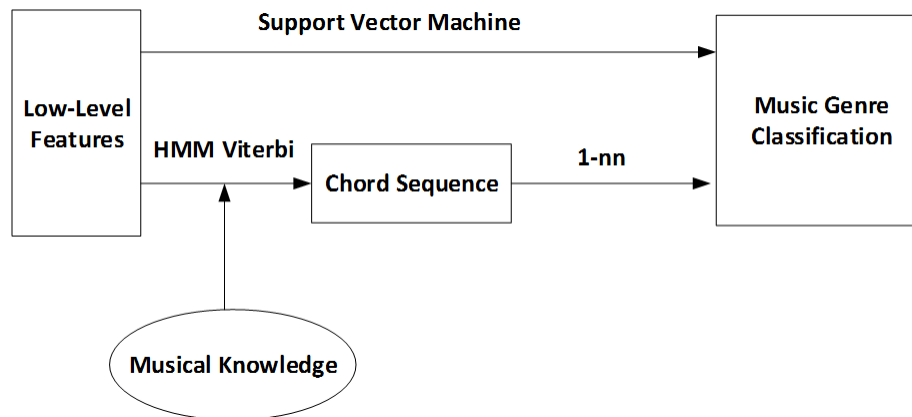


Figure 1: Overall solution

Figure 1 shows the overall proposed solution for music genre classification. Firstly, the low-level features are directly used for training and predicting by a SVM classifier[19,20]. Secondly, the musical knowledge is extracted as templates and the low-level features and the templates are used for predicting chord sequence by the HMM Viterbi algorithm. After preprocessing the predicted chord sequence, I use k nearest neighbors and these processed chord sequence for further classification. In figure 1, it shows 1-nn. Before the presentation, I only did experiments with 1-nn. After the presentation, I also perform 2-nn and 3-nn in the experiments for classification. And for k nearest neighbors algorithm I used, I define the distance between two samples of chord sequence as $1/\text{length of the longest common subsequence}$. I will discuss about the preprocessing of predicted chord sequence in detail later.

4.2 Feature extraction and chord recognition

Figure 2 shows the feature extraction process. I use “yaafe” and “librosa” to extract low-level features and separate these features into two groups. The dimensionality of the feature group 1 is 24 in total. The feature group 2 is composed of all the statistical information of some low-level features. However the dimensionality of the feature group 2 is as many as 3206, even though I only extract seven pieces of statistical information for every feature in group 2. In the next section I will discuss how I reduce the dimensionality by feature selection.

Figure 3 shows the process of chord recognition[31-33].

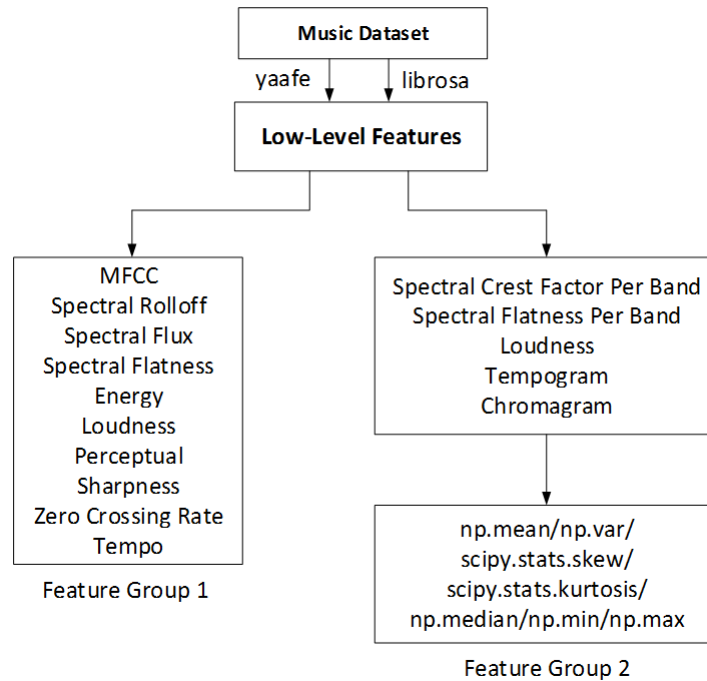


Figure 2: Feature extraction

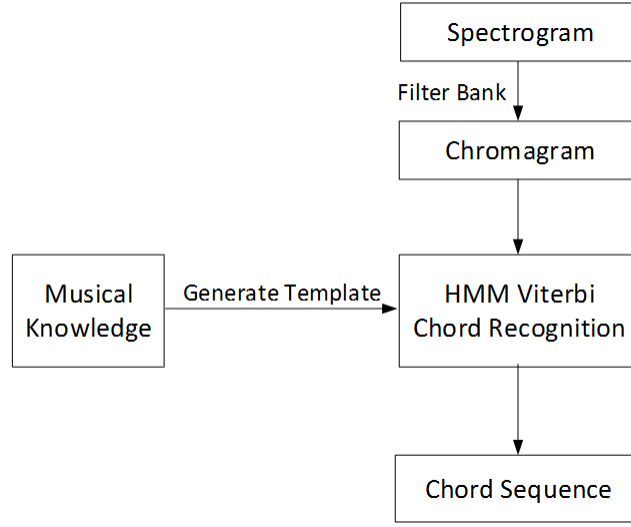


Figure 3: Chord recognition

4.3 Feature selection and chord sequence preprocessing

Two feature selections using SVM classification are performed separately for feature group 1 and feature group 2. After the feature selection for feature group 1, the features are reduced from $X[:,0:25]$ to $X[:,15] + X[:,22:25]$. After the feature selection for feature group 2, the features are reduced from $X[:,0:3206]$ to $X[:,266:434] + X[:,3122:3206]$.

The predicted chord sequence for each song is discrete and has 2538 values for 30 seconds. The chord sequence for a period of time is stable, so the values for a period of time do not change. I take the period of time (tunit < 2538) as a parameter to preprocess the predicted chord sequence. For the period of time, I set the chord to be the value in the majority of time. Figure 4 shows the feature selection for two feature groups and chord sequence preprocessing.

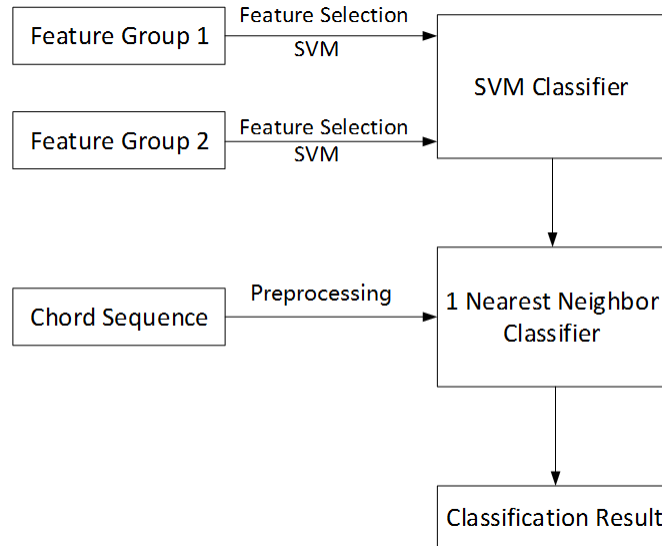


Figure 4: Feature selection and chord sequence preprocessing

5 Dataset

I use GTZAN as my dataset. This dataset was used for the well known paper in genre classification “Musical genre classification of audio signals ” by G. Tzanetakis and P. Cook in IEEE Transactions on Audio and Speech Processing 2002. The files were collected in 2000-2001 from a variety of sources including personal CDs, radio, microphone recordings, in order to represent a variety of recording conditions. The dataset consists of 1000 audio tracks each 30 seconds long. It contains 10 genres, each represented by 100 tracks. The tracks are all 22050Hz Mono 16-bit audio files in .au format. Figure 5 shows the distribution of music dataset

For the classification of this dataset, the training dataset is composed of 50 music files in each category and the testing dataset is composed of the other 50 music files in each category.

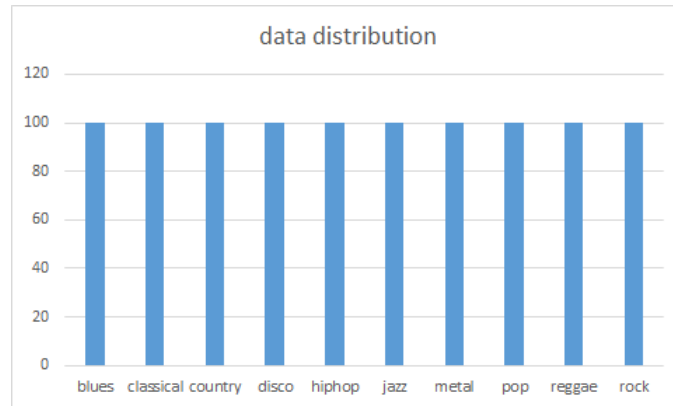


Figure 5: distribution of music dataset

There are 10 genres in total in this dataset. But I only choose 6 of them including classical, jazz, country, pop, rock and metal, which are more popular.

6 Parameter Selection

For support vector machine, I conduct grid search for kernel, C, gamma, and degree.

kernel: linear, poly, rbf

C: 2^i , $i=\text{range}(-10,10)$

gamma: 2^j , $j=\text{range}(-8,5)$

degree: range(2,8)

For k nearest neighbors, I choose k from 1,2,3.

For chord sequence preprocessing, I choose period of time tunit from 80(1s) to 360(4s) with step 20.

7 Experimental Results

The following results are 5 experimental results with optimal parameters acquired by grid search for each experiment . The first three experiments do not include chord sequence feature. Of the three experiments, the third experiment which includes both the feature group 1 and feature group 2 has the best accuracy. However from the confusion matrix, we can see that the jazz songs are confused with country songs to a great extent. The forth experiments and the fifth experiments have included the chord sequences feature and the accuracy has been improved and reached 0.76.

Feature Group 1

kernel = poly

C = 2

$\gamma = 2$
 $\text{degree} = 2$
 classification accuracy: 0.67
 The confusion matrix is as follows (figure 6).

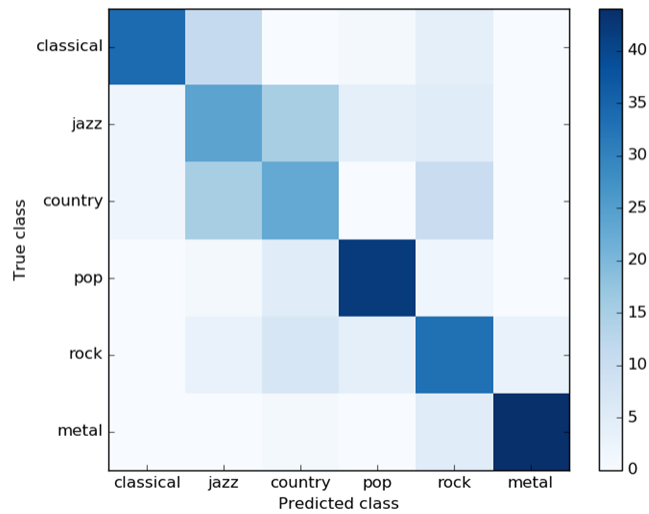
$$\begin{bmatrix}
 34 & 11 & 0 & 1 & 4 & 0 \\
 2 & 24 & 15 & 4 & 5 & 0 \\
 2 & 15 & 23 & 0 & 10 & 0 \\
 0 & 1 & 5 & 42 & 2 & 0 \\
 0 & 3 & 7 & 4 & 33 & 3 \\
 0 & 0 & 1 & 0 & 5 & 44
 \end{bmatrix}$$


Figure 6: Feature Group 1

Feature Group 2
 $\text{kernel} = \text{rbf}$
 $C = 4$
 $\gamma = 0.125$
 classification accuracy: 0.71
 The confusion matrix is as follows (figure 7).

$$\begin{bmatrix}
 40 & 7 & 1 & 0 & 2 & 0 \\
 0 & 29 & 16 & 1 & 4 & 0 \\
 2 & 4 & 29 & 0 & 15 & 0 \\
 0 & 1 & 6 & 7 & 30 & 3 \\
 0 & 0 & 0 & 0 & 8 & 42
 \end{bmatrix}$$

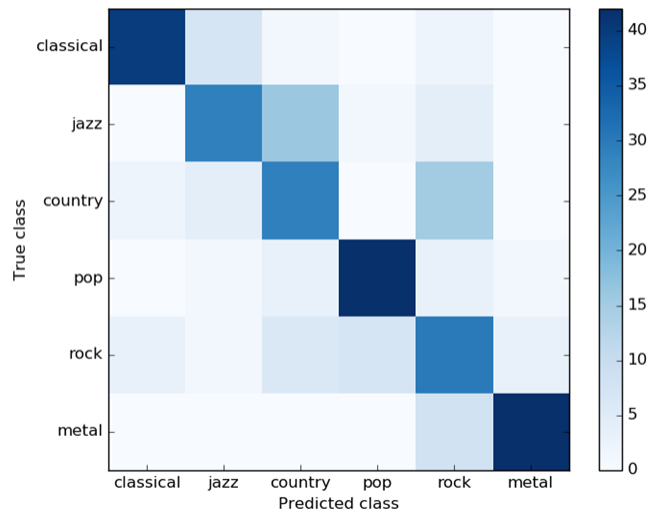


Figure 7: Feature Group 2

Combined features of feature group 1 and feature group 2

kernel = rbf

C = 1

gamma = 0.125

classification accuracy: 0.74

The confusion matrix is as follows (figure 8).

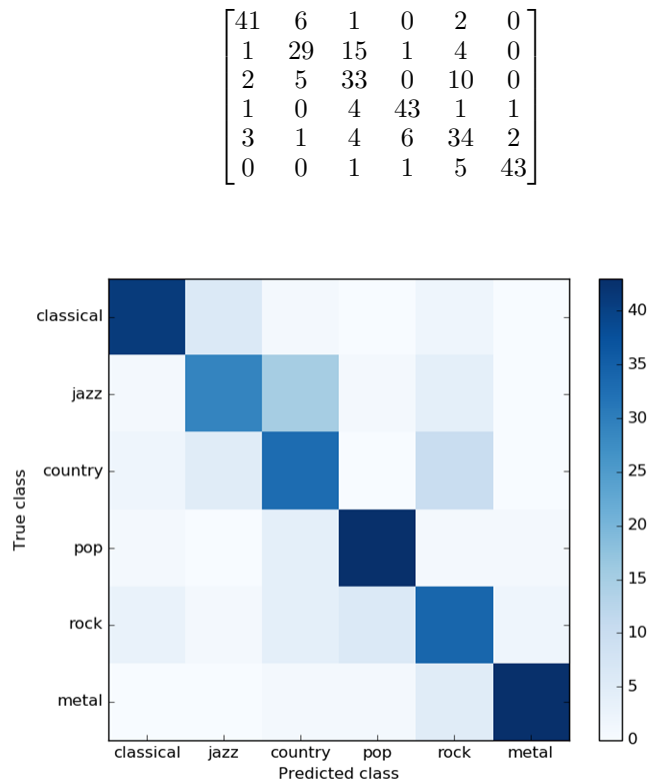


Figure 8: Combined features of feature group 1 and feature group 2

Combined features with Chord(1-nn)
kernel = rbf
C = 1
gamma = 0.125
k = 1
tunit = 175
classification accuracy: 0.76
The confusion matrix is as follows(figure 9).

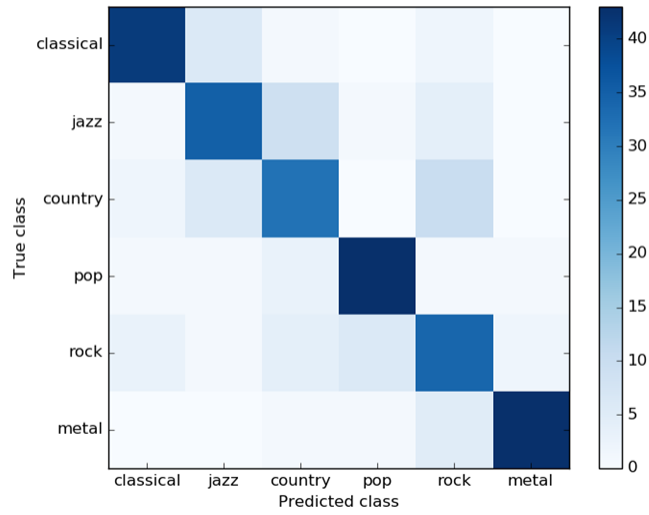
$$\begin{bmatrix} 41 & 6 & 1 & 0 & 2 & 0 \\ 1 & 35 & 9 & 1 & 4 & 0 \\ 2 & 6 & 32 & 0 & 10 & 0 \\ 1 & 1 & 3 & 43 & 1 & 1 \\ 3 & 1 & 4 & 6 & 34 & 2 \\ 0 & 0 & 1 & 1 & 5 & 43 \end{bmatrix}$$


Figure 9: Combined features with Chord(1-nn)

Combined features with Chord(3-nn)
kernel = rbf
C = 1
gamma = 0.125
k = 3
tunit = 270
classification accuracy: 0.76
The confusion matrix is as follows(figure 10).

$$\begin{bmatrix} 41 & 6 & 1 & 0 & 2 & 0 \\ 1 & 36 & 8 & 1 & 4 & 0 \\ 2 & 7 & 31 & 0 & 10 & 0 \\ 1 & 2 & 2 & 43 & 1 & 1 \\ 3 & 2 & 3 & 6 & 34 & 2 \\ 0 & 1 & 0 & 1 & 5 & 43 \end{bmatrix}$$

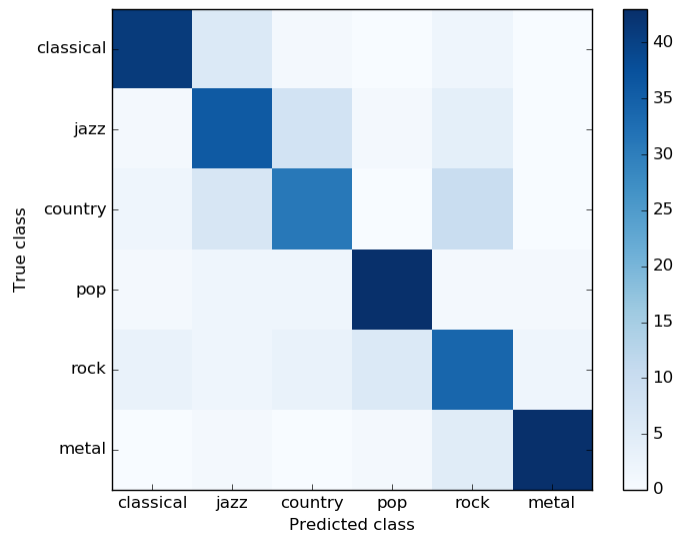


Figure 10: Combined features with Chord(3-nn)

8 Conclusion and Future Work

8.1 Conclusion

By analyzing the experimental results, we can see that the feature group 2 has better classification performance than feature group 1 and combining feature group 1 and feature group 2 will have the best accuracy. However, by observing the confusion matrix, we may find that jazz songs are confused with country songs. And by comparing the confusion matrix of the third experiment and the forth or fifth experiment, we can conclude that chord sequence is able to classify jazz songs from country songs better and it improve the classification accuracy. The accuracy finally reaches 0.76.

8.2 Future Work

In future study, we may improve the classification performance by improving the chord recognition. And we could use other mid-level feature such as instrument. We may also replace 1/length of longest common subsequence with other property to define distance for k nearest neighbors.

9 Why are you the right person for implementing this plan

I like music. Specifically, I like playing piano and guitar and I am crazy about some chord progression. I often analyze music mathematically. I am taking the machine learning class and interested in machine learning classification algorithms. So I am the right person for this project.

References

- [1] Youngmoo E. Kim, Erik M. Schmidt & et al. Music emotion recognition: a state of the art review. 11th International Society for Music Information Retrieval Conference (ISMIR 2010)
- [2] Yang, Yi-Hsuan, & Homer H. Chen. Machine recognition of music emotion: A review. ACM Transactions on Intelligent Systems and Technology (TIST) 3.3 (2012): 40.
- [3] Aucouturier, Jean-Julien, & Francois Pachet. Representing musical genre: A state of the art. Journal of New Music Research 32.1 (2003): 83-93.

- [4] LIN, Y.-C., YANG, Y.-H., & CHEN, H.-H. 2009. Exploiting genre for music emotion classification. In Proceedings of the IEEE International Conference on Multimedia and Expo. 618621.
- [5] Yang, Y.-H & Chen, H.H. (2011) Music Emotion Recognition. CRC Taylor & Francis Books, Feb. 2011
- [6] J. P. Bello and J. Pickens, A robust mid-level representation for harmonic content in music signals, in Proc. ISMIR, pp. 304311, 2005.
- [7] CHENG, H.-T., YANG, Y.-H., LIN, Y.-C., LIAO, I.-B., & CHEN, H.-H. 2008. Automatic chord recognition for music classification and retrieval. In Proceedings of the IEEE International Conference on Multimedia and Expo. 15051508.
- [8] BENETOS, E., KOTTI, M., & KOTROPOULOS, C. (2007) Large scale musical instrument identification. In Proceedings of the International Conference on Music Information Retrieval.
- [9] Scaringella, Nicolas, Giorgio Zoia, & Daniel Mlynek. Automatic genre classification of music content: a survey. Signal Processing Magazine, IEEE 23.2 (2006): 133-141.
- [10] Tzanetakis, G. & Cook, P. 2002. Musical genre classification of audio signals. IEEE Trans. Speech Audio Process. 10, 5, 293302. <http://marsyas.sness.net/>.
- [11] Elias Pampalk, Arthur Flexer, & Gerhard Widmer. (2005) IMPROVEMENTS OF AUDIO-BASED MUSIC SIMILARITY AND GENRE CLASSIFICATION.
- [12] Shao, Xi, Changsheng Xu, & Mohan S. Kankanalli. Unsupervised classification of music genre using hidden markov model. Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on. Vol. 3. IEEE, 2004.
- [13] Li, Tao, Mitsunori Ogihara, & Qi Li. A comparative study on content-based music genre classification. Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2003.
- [14] K. West, S. Cox, Finding an optimal segmentation for audio genre classification, in Proceedings of the 6th Int. Symposium on Music Information Retrieval, London, UK, 2005.
- [15] Poria, Soujanya, et al. Music genre classification: A semi-supervised approach. Pattern Recognition. Springer Berlin Heidelberg, 2013. 254-263.
- [16] N. Scaringella, G. Zoia, On the modeling of time information for automatic genre recognition systems in audio signals, in Proceedings of the 6th Int. Symposium on Music Information Retrieval, London, UK, 2005.
- [17] N. Scaringella, D. Mlynek, A mixture of support vector machines for audio classification. Music Information Retrieval Evaluation exchange (MIREX) website.
- [18] Silla Jr, Carlos N., Celso A. Kaestner, & Alessandro L. Koerich. Automatic music genre classification using ensemble of classifiers. Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on. IEEE, 2007.
- [19] CHANG, C.-C. AND LIN, C.-J. 2001. LIBSVM: A library for support vector machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [20] CORTES, C. AND VAPNIK, V. 1995. Support vector networks. Machine Learn. 20, 3, 273297.
- [21] PEETERS, G. 2008. A generic training and classification system for MIREX08 classification tasks: Audio music mood, audio genre, audio artist and audio tag. In Proceedings of the International Conference on Music Information Retrieval.
- [22] SCARINGELLA, N., ZOIA, G., & MLYNEK, D. 2006. Automatic genre classification of music content: A survey. IEEE Signal Process. Mag. 23, 2, 133141.
- [23] Richert, Willi. Building Machine Learning Systems with Python. Packt Publishing Ltd, 2013.
- [24] Kaichun K. Chang, Jyh-Shing Roger Jang, & Costas S. Iliopoulos. (2010) Music Genre classification via compressive sampling. 11th International Society for Music Information Retrieval Conference.
- [25] Yannis Panagakis, Constantine Kotropoulos, & Gonzalo R. Arce (2009) Music Genre Classification via Sparse Representations of Auditory Temporal Modulations.
- [26] Chen, Ling, Phillip Wright, & Wolfgang Nejdl. Improving music genre classification using collaborative tagging data. proceedings of the second ACM international conference on web search and data mining. ACM, 2009.
- [27] <http://yaafe.sourceforge.net/>

- [28] http://www.ifs.tuwien.ac.at/~schindler/lectures/MIR_Feature_Extraction.html
- [29] <https://github.com/bmcflee/librosa>
- [30] A. Sheh & D. P. Ellis, Chord segmentation and recognition using EM-trained hidden Markov models, in Proc. ISMIR, pp. 185-191, 2003.
- [31] H. Papadopoulos & G. Peeters, Large-scale study of chord estimation algorithms based on chroma representation, in Proc. CBMI, pp. 53-60, 2007.
- [32] K. Lee & M. Slaney, A unified system for chord transcription and key extraction from audio using hidden Markov models, in Proc. ISMIR, pp. 245-250, 2007.