Sin City, Classification of Regional Songs in Turkey

Muhammet Emin Özgür
b21427229@cs.hacettepe.edu.tr

Ufuk Umut Şentürk
b21427435@cs.hacettepe.edu.tr

Muazzez Şule Karaşlar
b21427066@cs.hacettepe.edu.tr

Machine Learning Lab, Department of Computer Engineering
Hacettepe University, Ankara, TURKEY

Abstract

In most cultures we can define a generic type of songs that defines the culture. We can even pinpoint the differences between similar cultures but relatively in a different region which might be city, county or state. This is the point where idea came to classify folk songs in Turkey region by region. Turkey has 7 geographic regions which can be seen as cultural regions. Thus, those regions do not have completely different cultures,but differences are visible which one of them is the regional songs. For instance, tulum, kemençe(endemic musical instruments for Karadeniz) can be heard in Karadeniz songs which are distinguishable from other regions songs. However, some regions have same musical infrastructure which are rhythm, instruments, chords or notes patterns and vocal type which is hard to distinguish for a human being. This one is probably the hardest challenge we will face. Hence, feature extraction is the crucial point for this matter and there are certain features can be used for our classification problem and there are some must be tried before used. We will try to implement and analyze different models and different features for this classification problem.

1. Introduction

In machine learning most important step is the feature extraction. Just by improving features will increase dramatically for contemporary machine learning models. In the audio feature extraction, current state-of-art is being provided by MIR(Music Information Retrieval) research field. MIR is a huge field that has been under the spotlight since Tzanetakis and Cook released a paper on Music Genre Classification in 2002[5]. MIR includes audio feature extraction, signal processing, audio analyse etc. Audio feature extraction is what we are interested in MIR research field. We are planning mainly to use features: Chroma-gram, MFCC(Mel-frequency cepstral coefficients), Spectral centroid, Spectral roll-of, Zero-crossing rate, etc. We will also use other features for reference points but these will be the main features. Machine learning model selection for the job is not without importance. State-of-art models are plentiful, however they are mostly specific for a given problem. In order to find the best classification, we will use various models which will be CNN(Convolutional Neural Network), SVM(Support Vector Machine), K-NN(K Nearest Neighbor). These machine learning models will be implemented with various contemporary features discussed in MIR. Subsequently these will be analysed. Analyzing will be last step in our paper. We will gather results after the testing which in a matrix as rows for machine learning models and columns for features. Best model and features will be apparent after the testing is done. Confusion matrix will be constructed by using our best model and features.

2. Related Works

In this paper, classification problem is more specific than any other work. since Tzanetakis and Cook presented music genre classification[5], a lot of work has been done. For Audio processing and analyzing, it is hard to build-up a reliable feature extractor system for audio seems to be a more challenging task than classification problem. For this purpose some researches have been done about features extraction such as; MFCC[8], Chromogram[4], Zero-crossing rate, Spectral-Centroid[11].

The most known study for folk song classification which is based on region has been done for China. On paper in interest, 74 features have been extracted from audio files of the songs, and classified by an audio classifier on SVM which is one of our concern. Research shows usthat SVM without feature selection is a very effective classification method with the combination of 13-dimension MFCC and 10-dimension without feature selection. It points us to

1

post processing and segmentation plays key role in here. Other research is called "Regional Classification of Traditional Japanese Folk Songs" by Akihiro KAWASE, Akifumi TOKOSUMI has been made. That research use music corpora consisting of 202,246 tones from 1,794 song pieces from 45 prefectures in Japan which is big enough.

In another research[12], presented a methodology to automatically extract musical patterns features from audio music using the CNN which is easy to use for feature extractors, it because need minimal prior knowledge to construct. Some other paper uses[9] various machine learning algorithms that are including k-nearest neighbor (kNN), k-means, multi-class SVM, and neural networks to classify the following four genres: classical, jazz, metal, and pop and only Mel Frequency Cepstral Coefficients (MFCC) is used to characterize data. Using CNN is become quite popular in other ML fields with each other. Qiuqiang Kong, et al[10] stated that many manual-selected features such as MFCC have been applied to music processing but they are not effective for music genre classification. In their work, it is presented an algorithm based on spectrogram and CNN and compared with MFCC, the spectrogram contains more details of music components such as pitch, flux, etc.

3. Methodology

3.1. Data Set

We have gathered our data from youtube playlist. There are various regional playlist which we have found with quick search[1]. We obtained these in ogg format and as smallest bitrate as possible we can find in order to save space. JDownloader[2] used to download these playlist with ease. Each download placed in respective folder for label in their classification. In these playlist there might be noise such as given song might not be in that regional as suggested in the name of the playlist. These will be most likely a miniscule noise. Moreover we will add more songs in the data set, hence the noise will be smaller with each growth of the data set.

3.2. Features

As hitherto cited, MFFC, Spectral Centroid, Spectral roll-of, Zero-crossing rate[5] are main focus in this paper. Librosa[3] is used for feature extraction library. It is not the fastest library, it is only used because of convenience. These features will have tremendous dimension space which would need dimension reduction with PCA(Principle Component Analysis) and feature selection techniques.

Above figure shows us the calculation steps of MFCC[6]. MFCC is a convenient feature used in speak recognition,
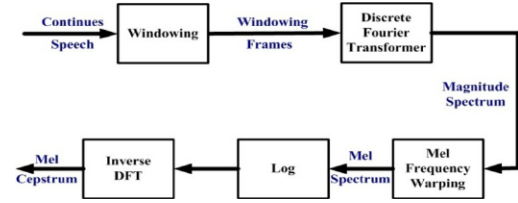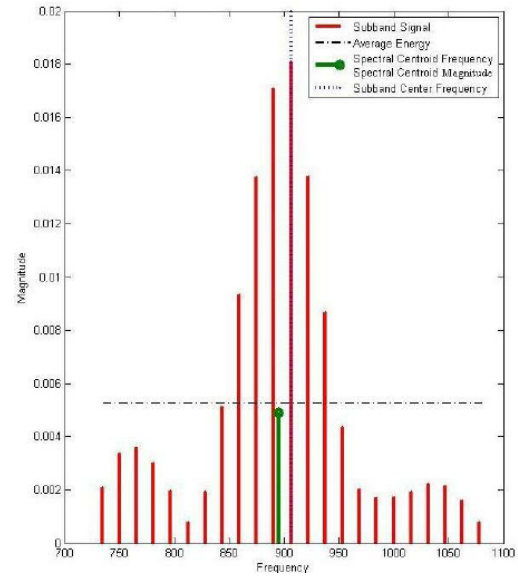

Figure 1. MFCC Calculation Process.[6]

Figure 2. Subband signal, average energy, spectral centroid frequency and spectral centroid[7].

signal processing. Alternatively, spectral centroid feature is proposed and used widely. Below figure shows spectral centroid.[7]

3.3. Classification

We will implement 3 classification model, namely SVM, CNN, K-NN. In the CNN implementation we will use tensorflow. CNN architecure will have 2 convolution layer in which each layer followed by max pooling and drop out. However it is still in progress, hence it is subject to change. SVM and K-NN will be implemented by scikit-learn python module.

4. Feature Works

Implementation of our classification will be our next step. Following our implementation we will test it with our focused features and along-the-way features that we thought might improve accuracy. Feature extraction libraries might change, following that if the new library does not have our features, they might change too—little chance. However we will not change library if we might loose critical feature we

have found out after our tests.

References

[1] Data set of turkish folk songs. `https://we-l-ee.github.io/bbm406-Sin.City/data-set`.

[2] Jdownloader. `http://jdownloader.org/`.

[3] Librosa library. `https://librosa.github.io/`.

[4] D. P. W. Ellis. Classifying music audio with timbral and chroma features.

[5] P. C. G. Tzanetakis. Musical genre classification of audio signals. 10(5):294–295, 2002.

[6] P. . A.-A. A. . A. S. . A. M. Khdour, Thair Muaidi. Arabic audio news retrieval system using dependent speaker mode, mel frequency cepstral coefficient and dynamic time warping techniques. *Research Journal of Applied Sciences, Engineering and Technology*, (7):5082–5097, 2014.

[7] T. . N.-H. . E. A. . E. J. Kua, J.M.K. Thiruvaran. Investigation of spectral centroid magnitude and frequency for speaker recognition/. 2010.

[8] G. S. Md.Sahidullah. Design, analysis and experimental evaluation of block based transformation in mfcc computation for speaker recognition. *Speech Communication*, 54(4):543–565, 2012.

[9] K. K. Michael Haggblade, Yang Hong. Music genre classification.

[10] Y. L. Qiuqiang Kong, Xiaohui Feng. Music genre classification using convolutional neural network.

[11] E. Schubert, J. Wolfe, and A. Tarnopolsky. Spectral centroid and timbre in complex, multiple instrumental textures. 08 2004.

[12] A. B. C. A. H. C. Tom LH. Li. Automatic musical pattern feature extraction using convolutional neural network.