© June 2021| IJIRT | Volume 8 Issue 1 | ISSN: 2349-6002

Music Genre Classification

Miss. Geetanjali[1], Utkarsh Upadhyay[2], Prateek Dubey[3], Aman Singh[4], Sonu Kumar[5]
[1,2,3,4,5] *Department of CSE, RKGIT/AKTU*

***Abstract* - Classification of music genre has always been an interest in the area of music and musical data. Classification of genre can be very important to explain some interesting problems such as creating song references, exploring related songs, finding groups which will like that specific song. The aim of our project is to find the machine learning algorithm that predicts the genre of songs using k-nearest neighbor (k-NN) and Support Vector Machine (SVM). This paper also gives the difference between k-nearest neighbor (k-NN) and Support Vector Machine (SVM) with the help of principal component analysis (PCA). The Mel Frequency Cepstral Coefficients (MFCC) is used to get the information for the data set. Also, the MFCC features are used for a particular track. From the outcome of the project, we found that without the dimensionality reduction both k-nearest neighbor and Support Vector Machine (SVM) gave more accurate results than dimensionality reduction. Overall, the Support Vector Machine (SVM) is a much more effective classifier for classification of music genres. It had an accuracy of around 75%.**

***Index Terms* - K-nearest neighbor (k-NN), Support Vector Machine (SVM), music genre, Mel Frequency Cepstral Coefficients (MFCC).**

I.INTRODUCTION

In today's time, one's music collection contains various types of songs, and the professional music collection usually has thousands of songs. The music files are differentiated by different artist names and title of the song [1], and this causes difficulty in classifying the songs in different genres. As the internet and networking is increasing very fast, the number of people in the field of music is also increasing. Due to the large range of music database, the warehouses require an exhausting and time-consuming work, particularly when classifying audio genres manually. Nowadays, Music has also been categorized into different genres and subgenres on the basis of music sound as well as its lyrics [2]. This adds more difficulty to classify the music. The definition of genres has also changed now. For example, rock songs which were composed fifty years ago and today are different from the rock songs. There is a lot of work done on the music data in the last few years and there is a lot to do. Aucouturier and Pachet, 2003 [4] found that the genre of music is possibly the best general information for the music content clarification. Music engineering heartened the practice of categories and family-based operators like to organize their sound storage by this clarification, so the classification of genres has improved. Also, the latest progress in category organization here is still an issue to accurately describe a type, or whether it mostly relies on a consumer's understanding and flavor. In order to establish and explore increasing composition groups we implemented an automatic technique that can be used for data mining for valuable data about audio composition direct from the audio file. Such data also include rhythm, tempo, energy distribution, pitch, timbre etc. Most of the classifications depend on spectral statistical features. Content collections relating to further music contents such as pitch and rhythm are suggested, however their execution is very fast and moreover they are closed by tiny info collections pointing at different audio arrangements. The k nearest neighbor is non-linear, but also it can sense linear or non-linear spread information. It tends to do very well with a lot of data points. Support Vector Machine can be used in both of the methods, once we have a partial set of points in many dimensions the Support Vector Machine inclines to be very good because it easily unearth the linear separation that should exist. Support Vector Machine is good with outliers as it will only use the most related points to find a linear separation (support vectors).

II. LITERATURE REVIEW

The prominence of programmed music genre classification has been established relentlessly for as far back as a couple of years. Many papers have expected frameworks that either model songs as a whole or apply SVM to build models for classes of music. Below some of the similar work is mentioned. Kris West and Stephen Cox [10] in 2004 planned a confounded classifier on many sorts of sound elements. They demonstrated efficient outcomes on 6-way type characterization errands, with almost 83% grouping attention on behalf of their greatest framework. As indicated by them the detachment of Reggae and Rock music was a different issue for the component extraction plan which was assessed by them. They also shared comparative phantom characteristics as well as comparable capacity of harmonic to nonharmonic material. Aucouturier & Pachet [11] perform on single songs through Gaussian combination Model (GMM) [12] and apply Monte Carlo conduct to assess the KL divergence [11] among them. Their structure was focused on an audio information restoration structure where the position is calculated in articulations of recovery perfection. Authors did not apply a propelled classifier, as their results are positioned by k-NN. They move some important part sets for a few models that we use in our scanning, in particular the MFCC.

A wide scale of information is invisible inside a music waveform which ranges from audible to perceptual. In a trial by Logan and Salomon they arrange playlists with the closest neighbors of a seed song. As specified by them they represent a technique to scan songs constructed exclusively in light of their sound material. They evaluate their separation measure on a database of more than 8000 songs. Preliminary goal and subjective result demonstrated that their separation calculates numerous parts of emotive comparability. For the twenty songs evaluated by two clients they saw that all things advised 2.5 out of the main 5 songs returned are perceptually equivalent. They additionally notice that their estimate is powerful to basic humiliation of the sound.

III. METHODOLOGY

Before starting, we added necessary toolboxes to the search path of MATLA. These were as follows:

1. Utility Toolbox.
2. Machine Learning Toolbox.
3. Speech and Audio Processing Toolbox.
4. Automatic Speech Recognition Toolbox.

We wrote a script to read in the audio files of the hundred tracks per category and removed the MFCC features used for individual tracks. We additionally decreased the dimension of each track because removed features are based on MFCC's statistics [8] comprising mean, std, min, and max along respectively dimension. Since MFCC has 39 dimensions, the removed file-based features have 39*4=156 dimensions.

To finish we used k-NN and SVM machine learning techniques via compact features as well as with all features set of each track.

Feature Extraction

For every song, we differentiate the comparing feature vector for classification. We used the function mgcFeaExtract.m (which MFCC and its measurements) for feature removal. We additionally put all the dataset into a single variable ―dataset‖ which is less demanding for further griping which includes classifier development and assessment. Since the feature removed is extensive, we just loaded the dataset.mat. As discussed above the removed features are based on MFCC's, so the removed file-based features had 39*4=156 dimensions.

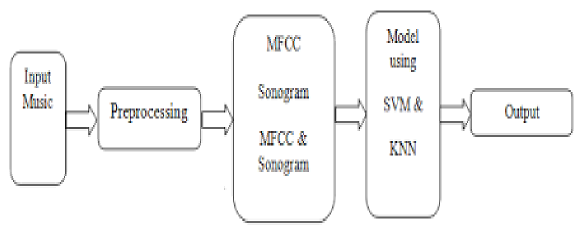

Fig: Processing of the project

Data Visualization

Since we had all the necessary information stored in ―dataset‖, we tried different functions of machine learning toolbox for data visualization and classification. For example, we displayed the size of each class.

IV. ALGORITHMS

A. *K-Nearest Neighbour (k-NN)*

The first machine learning technique we used was the k-closest neighbors (k-NN) [5] as it is very famous for its simplicity of implementation. The k-NN is by design non-linear and it can detect direct or indirect spread information. It also inclines with a huge amount of data. The essential calculation in our k-NN is to measure the distance between two tunes. We handled this by methods of the Kullback-Leibler divergence [10].

B. *Support Vector Machine (SVM)*

The second technique we used is the support vector machine [6] which is a directed organization method that discovers the extreme boundary splitting two classes of information. During this the information is not directly distinct in the feature space; if this is the case then they can be put into an upper dimensional space through the method of Mercer kernel. Actually, the internal results of the information focused in this higher dimensional space are essential, so the projection can be understood if such an inner item can be figured forthrightly.

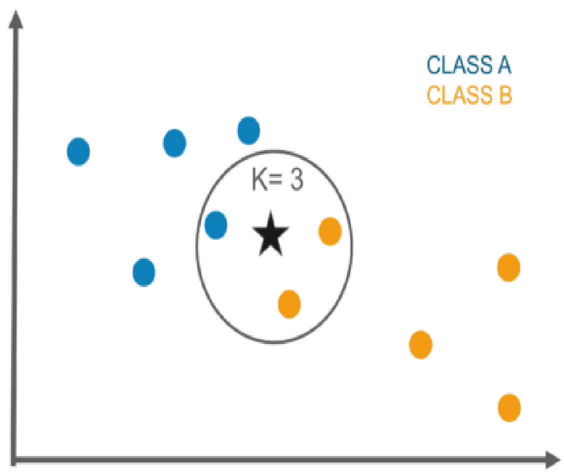

Fig: K-NN Algorithm

V. DATASET

Music Analysis, Retrieval, and Synthesis for Audio Signals (Marsyas) is an open source World Wide Web for sound handling with particular complement on audio data uses. For our examinations we used the GTZAN dataset which has an anthology of thousand sound files. Each of the files is thirty seconds in length. Ten genres are attending in this dataset comprising hundred tracks each. Each track has a 16-bit audio file 22050 Hz Mono in .au format [10]. We have chosen ten genres: blues, classical, rock, jazz, reggae, metal, country, pop, disco and hip-hop. Our total data set consists of more than 1000s of songs files.

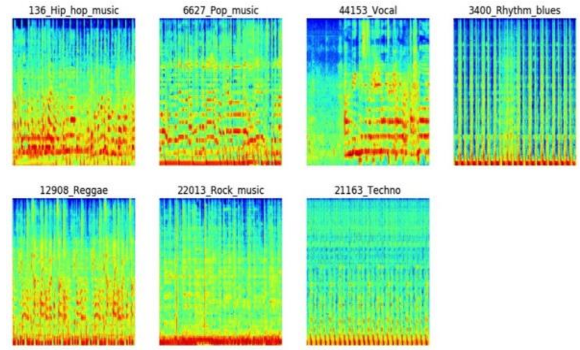

Fig: Sample of few music genres

VI. CONCLUSION

The precision of classification by various genres and different machine learning algorithms is different. The success rate of SVM was 83% but the blues genre was miscalculated as a rock or metal genre. The k-NN did inadequately while comprehending blues with a recognizing percentage of 49%. The SVM also misidentified the classical genre like jazz or hip-hop, but the rock genre was accurately observed with an achievement rate of 94%. The K-NN did also well when specifying classical with a success rate of 90%. Furthermore, the SVM did also well with comprehending entire classifications but on the other hand it also incorrectly specified disco with rock and reggae with hip-hop. The success rate of the country was 70% but with the rock genre, it was just 12%. The Hip-hop genre had a success rate of 74% but had difficulty distinguishing between reggae with the highest inaccuracy of 14%. Jazz was recognized with a precision rate of 90% but had difficulty in comprehending classical genres. Rock has the lowest success rate of 59% having drawbacks with many other genres.

REFERENCES

[1] Chathuranga, Y. M., & Jayaratne, K. L. (2013). Automatic Music Genre Classification of Audio Signals with Machine Learning Approaches. GSTF International Journal of Computing, 3(2).

[2] Serwach, M., & Stasiak, B. (2016). GA-based parameterization and feature selection for automatic music genre recognition. In Proceedings of 2016 17th International Conference Computational Problems of Electrical Engineering, CPEE 2016.

[3] Dijk, L. Van. (2014). Radboud Universiteit Nijmegen Bachelor Thesis Information Science Finding musical genre similarity using machine learning techniques, 1–25.

[4] Aucouturier, J., & Pachet, F. (2003). Representing Musical Genre: A State of the Art. Journal of New Music Research, 32(February 2015), 83–93.

[5] Leif E. Peterson (2009) K-nearest neighbor. Scholarpedia, 4(2):1883.

[6] Mandel, M. I., Poliner, G. E., & Ellis, D. P. W. (2006). Support vector machine active learning for music retrieval. Multimedia Systems, 12(1), 3–13.

[7] Jolliffe, I. T. (2002). Principal Component Analysis, Second Edition. Encyclopedia of Statistics in Behavioral Science, 30(3), 487.

[8] Logan, B. (2000). Mel Frequency Cepstral Coefficients for Music Modeling. International Symposium on Music Information Retrieval, 28, 11p.

[9] Fu, Z., Lu, G., Ting, K. M., & Zhang, D. (2011). A survey of audio based music classification and annotation. IEEE Transactions on Multimedia, 13(2), 303–319.

[10] West, K., & Cox, S. (2004). Features and Classifiers for the Automatic Classification of Musical Audio Signals. Proc. International Society for Music Information Retrieval Conference, 1–6.

[11] Aucouturier, J.-J., & Pachet, F. (2004). Improving timbre similarity: How high's hte sky? Journal of Negative Results in Speech and Audio Sciences, 1(1), 1–13. Retrieved from