IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 3, March 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903191 |

Music Genre Classification Using Machine Learning

Jegadeesan, Pranav K M, Abbis Kumar V J, Thiruselvam M

Department of Information Technology, Velammal College of Engineering and Technology, Madurai, Tamilnadu, India

ABSTRACT: The Categorizing music files according to their genre is a challenging task in the area of Music Information Retrieval(MIR). This paper describes a new technique that uses Support Vector Machine(SVM)algorithm to classify songs.Support Vector Machine classify audio into their respective classes by learning from training data . The GTZAN genre collection dataset was collected.It consists of 1000 audio files each having 30 seconds duration. There are 10 classes(10 music genres) each containing 100 audio tracks. The features that contribute the most towards this classification task are identified. The experiments are conducted on the Audio set data set and we report an AUC value of 0.894 for an ensemble classifier which combines the two proposed approaches.

KEYWORDS*:* Python, Machine Learning, Support Vector Machine, Light Gradient Boosting Machine.

1. INTRODUCTION

There are numerous studies that are investigated in the field of digital music and how it would be possible to enhance users experience. However automatic genre classification is not an easy task considering music evolving within short periods. In addition to this,the advancement in digital signal processing and data mining techniques has led to intensive study on music signal analysis like content-based music retrieval,music genre classification,duet analysis,Musical transcription,Musical Information Retrieval and Music Instument detection and classification.Music Instrument detection techniques have many potential applications such as detecting and analysing solo passages,audio and video retrieval,music transcription,playlist generation,acoustic environment classification,video scene analysis and annotation etc. Music has also been divided into genre and sub genres not only on the basis on music but also on the lyrics as well.This makes classification harder. To make things more complicate the definition of music genre may have very well changed over time.For instance,rock songs that were made fifty years ago are different form the rock songs we have today. XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, decision tree based algorithms are considered best-in-class right now.
We use LightGBM,short for Light Gradient Boosting Machine,is a free and open source distributed gradient boosting framework by Microsoft. It is based on decision tree algorithms and used for ranking,classification and other machine learning taks.

II. LITERATURE SURVEY

Music genre classification has been a widely studied area of research since the early days of the internet Machine learning techniques have been used for music genre classification for decides now. In 2002, G.Tzanetakis and P.cook used both the mixture of Gaussians model and K-nearest neighbours along with three sets of carefully hand-extracted features representing timbral texture,rhythmic content and pitch content.They achieved 61% accuracy.As a benchmark,human accuracy averages around 70% for this kind of genre classification work .Tzanetakis and cook used MFCCs a close cousin of melspectrograms and essentialy all work has followed in their footsteps in transforming their data in this manner . In the following years ,methods such as support vector machines were also applied to this task such as in 2003 when C.Xu et al used multiple layers of SVMs to achieve over 90% accuracy on a dataset containing only four genres. In the past 5-10 years however convolutional neural networks have shown to be incredibly accurate music genre classifiers, with excellent results reflecting both the complexity provided by having multiple layers and the ability of convolutional layers of effective identity patterns within images(which is essentially what mel-spectrograms and MFCCs) .These results have far excedded human capacity for genre classification with our research finding that

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903191 |

current state of the art models perform with an accuracy of around 91% when using the full 30s track length . Many of the papers which implemented CNNs compared their models to other ML thechniques , including KNN,mixture of Gaussians,and SVMs,and CNNs performed favourably in all cases.In already existing system, it will have some number of music which will be categorized under respective genre where as the end user can not upload any music and find the genre of that particular music. Therefore we decided to focus our efforts on predicting Music genre using SVM(Support Vector Machine).

III. ABOUT THE DATASET

3.1 Support Vector Machine

A machine learning technique which is based on the principle of structure risk minimization is support vector machines. It has numerous applications in the area of pattern recognition. SVM constructs linear model based upon support vectors in order to estimate decision function. If the training data are linearly separable, then SVM finds the optimal hyper plane that separates the data without error.The support vectors are the (transformed) training patternsand are equally close to hyperplane of separation. The support vectors are the training samples that define the optimal hyperplane and are the most difficult patterns to classify.

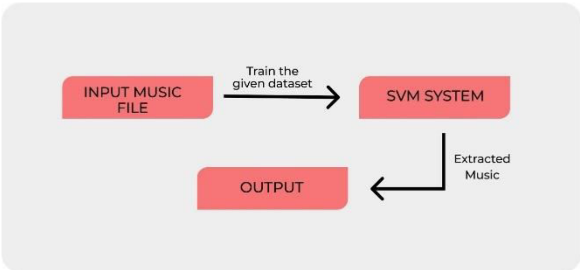

Figure 1: SVM Work flow

3.2 Data Set

GTZAN Genre Collection dataset was used to perform the classification. The dataset has been taken from the popular software framework MARSYAS. Marsyas (Music Analysis, Retrieval and Synthesis for Audio Signals) is an open source software framework for audio processing with specific emphasis on Music Information Retrieval applications. It has been designed and written by George Tzanetakis (gtzan@cs.uvic.ca). Marsyas has been used for a variety of projects in both academia and industry. Dataset consists of 1000 audio tracks each 30seconds long. It contains 10 genres(Blues, Classical, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae and Rock), each represented by 100 tracks. The tracks are all 22050Hz Mono 16-bit audio files in .wav format.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903191 |

Genre	Number of Tracks
Blues	100
Classical	100
Country	100
Disco	100
Hip-Hop	100
Jazz	100
Metal	100
Pop	100
Reggae	100
Rock	100
TOTAL	1000

Data Pre-Processing was done in the following manner:

1. Database of the complete collection was created and stored in a particular file.

2.Feature Vector Extraction is done using the libROSA package in python as shown in figure 1. libROSA is a python package for music and audio analysis which provides the building blocks necessary to create music information retrieval systems.

3.Each audio file is taken and from that, its feature vector is extracted. The extracted feature vector is called MFCC(Mel-Frequency Cepstral Coefficients). The MFCCs as shown in figure 3 encode the timbral properties of the music signal by encoding the rough shape of the log-power spectrum on the Mel frequency scale. A Zero Crossings graph is plotted as shown in figure 4 for each audio track. This graph visualizes the number of times the signal crosses zero level.

4. Fourier Transforms are applied on the music signal. Frequency Spectrum is thus obtained. Mel Scale Filtering is applied on the frequency spectrum to obtain a Mel Frequency Spectrum. A log() function is applied on this Mel Frequency Spectrum which is transformed into Cepstral Coefficients on applying discrete cosine transforms. Finally, the Feature Vector is obtained by finding out the derivatives of the Cepstral Coefficients

3.3 Classifications in music genre

The classification in music genre are done by using many types.In this project we have done the classification using Symbolic based feature and audio based feature. The symbolic based feature and audio based feature consist of Pitch, Rhythm, Timbre, Harmony. Symbolic based feature and audio based feature are divided into two types. The two types are High level features and Mid level feature.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |
|| Volume 9, Issue 3, March 2021 ||
| DOI: 10.15680/IJIRCCE.2021.0903191 |

High level feature

Timbre	Harmony
➢ instrument patches ➢ instrument classes ➢ percussion sets ➢ prevalent instrument ➢ histogram of used instruments ➢ number of used instruments ➢ fraction of notes from unpitched instruments - fraction of notes from pitched instruments	➢ Chord names/degree progression ➢ Used tonal scales ➢ Number of accidental notes ➢ Key changes ➢ Number of distinctric harmonic ➢ Dominant interval histogram ➢ Chord dictionary/histogram

Low level feature

Pitch	Rhythm
➢ pitch/pitch class histogram ➢ sequence of pitch/pitch class ➢ sequence of melodic intervals ➢ absolute/relative pitch contour ➢ melodic intervals histograms - pitch extension/range ➢ number of pitch contour changes ➢ dominant pitch/pitch class prevalence - pitch volume ➢ pitch class variety o pitch bend fraction - number/ proportion of specifc intervals (e.g., major seconds)	➢ Duration histogram ➢ Sequence of inter onset interval ➢ Sequence of duration ratio or IOIR ➢ Sequence of tempo/meter changes ➢ Sequence of duration counter-time attacks ➢ Duration range ➢ Number of meter/tempo changes ➢ Dominant time ratio,tempo

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903191 |

IV. HIGH LEVEL DESIGN

4.1 Methods

We decided to first implement two simpler models as baseline measures of performance, then progress to more complicated models in order to increase accuracy. We implemented variants of k-nearest neighbours, support vector machine, a fully connected neural network, and a convolutional neural network. In addition to mel-spectrogram features, we used principal component analysis to reduce dimensionality for the input to k-NN and SVM.

4.2 K-Nearest Neighbours

After reducing the dimensionality to 15 features using PCA , we applied the k-nearest neighbours algorithm. Predictions are made on a per-case basis by finding the closest training examples to our test or cross-validation example 2 that we wish to classify and predicting the label that appeared with greatest frequency among their ground-truth labels. Through trial and error, we found that best accuracy resulted from setting k = 10 and weighting the label of each neighbour by distance.

4.3 Support Vector Machine

After reducing dimensionality using PCA, we trained an SVM classifier as well. SVMs are optimal margin classifiers that can use kernels to find more complex relations in the input data. This kernel, also sometimes called the Gaussian Kernel, corresponds to an infinite dimensional feature space is related to Euclidean distance. This function as a whole is often minimized using sequential minimization optimization.

4.3 Feed-Forward Neural Network

We used a fully connected neural network as well, with ReLU activation and 6 layers, with cross-entropy loss. As the input to our model was 1D, when using mel-spectrograms, we flattened the data. Our model is fully connected, which means each node is connected to every other node in the next layer. At each layer, we applied a ReLU activation function to the output of each node.

4.4 Convolutional Neural Networks

This was our most advanced model, using 3 convolution layers, each with its own max pool and regularization, feeding into 3 fully connected layers with ReLU activation, softmax output, and cross entropy loss. This approach involves convolution windows that scan over the input data and output the sum of the elements within the window. This then gets fed into a max pool layer that selects the maximum element from another window. Afterwards, the output is fed through a model described in section 4.1. This was implemented with TensorFlow and Keras.

	With data processing			Without data processing		
	Train	CV	Test	Train	CV	Test
Support Vector Machine	.97	.60	.60	.75	.32	.28
K-Nearest Neighbors	1.00	.52	.54	1.00	.21	.21
Feed-forward Neural Network	.96	.55	.54	.64	.26	.25

With the results from various methods and analysing them ,we developed the project using SVM(Support Vector Machine). The efficiency in this algorithm give the accuracy level that satisfies the need of this project.

4.5 Work Flow

1. The dataset is split into two parts, Training data and Test data.

2. Each track from the train dataset is pre processed and a support vector is extracted for the same. A Support

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903191 |

Vector is generated from the extracted feature vectors.

3. The model is trained using the obtained feature vector .

4. Each track from the test dataset is also pre processed and a feature vector is extracted for the same.

5. The trained model operates on the feature vector obtained at the end of step 4 to perform classification on test data.

6. Finally, output is genre of the music track.

V. RESULT AND CONCLUSION

This research work provides the details of an application which performs Music Genre Classification using Machine Learning techniques. The application uses a Convolutional Neural Network model to perform the classification. A Mel Spectrum of each track from the GTZAN dataset is obtained. This is done by using the libROSA package of python. A piece of software is implemented which performs classification of huge database of songs into their respective genres. For the GTZAN dataset, the model we used achieved a training accuracy of about 98% and validation accuracy of 73%. Python was the language used to develop the model. A number of packages such as Keras, numpy, pandas were used to build the model. TensorFlow package is used for deep-learning.

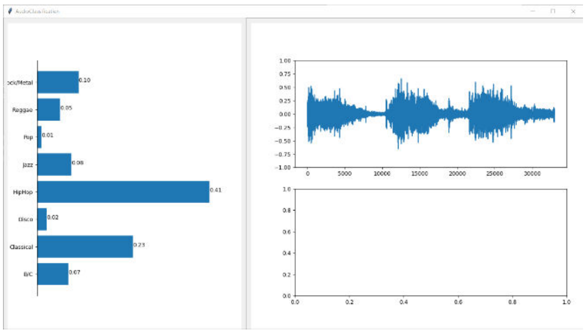

Figure 2: accuracy of prediction HIP-HOP genre

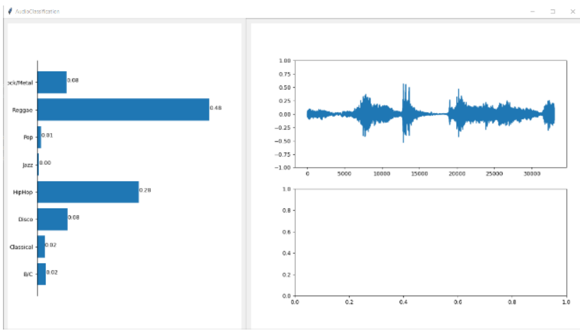

Figure 3: accuracy of prediction Reggae genre

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903191 |

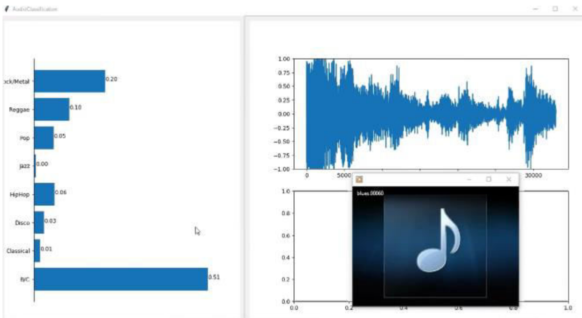

Figure 4: accuracy of prediction Blues and country

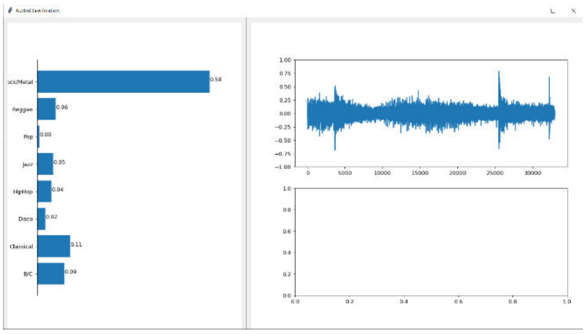

Figure 5: accuracy of prediction Rock and metal

REFERENCES

[1] Vishnupriya S, and K.Meenakshi, "Automatic Music Genre Classification using Convolution Neural Network", IEEE Conference 2018.

[2] G.Tzanetakis and P. Cook. "Musical genre classification of audio signals, Speech and Audio Processing", IEEE Transactions, July 2002.

[3] Chandsheng Xu, Mc Maddage, Xi Shao, Fang Cao, and Qi Tan, "Musical genre classification using support vector machines", IEEE Proceedings of International Conference of Acoustics, Speech, and Signal Processing, Vol. 5, pp. V-429-32, 2003.

[4] Matthew Creme, Charles Burlin, Raphael Lenain, "Music Genre Classification ", Stanford University, December 15, 2016.

[5] T. Feng, "Deep learning for music genre classification", 2014.

[6] Daniel Grzywczak, Grzegorz Gwardys, "Deep image features in music information retrieval", 10th international conference, AMT 2014, Warsaw, Poland, August 11-14, 2014 proceedings, pp 187-199.

[7] Muhammad Asim Ali, Zain Ahmed Siddqui, "Automatic Music Genres Classification using Machine Learning", International Journal of Advanced Computer Science and Applications, Vol 8, No 8, 2017.

[8] Hareesh Bahuleyan, "Music Genre Classification using Machine Learning Techniques", University of Waterloo, ON, Canada, 2018

[9] Sam Clark, Danny Park, Adrien Guerard, "Music Genre Classification using Machine Learning Techniques", 2012.

[10] Eve Zheng, Melody Moh, Teng-Sheng Moh, "Music Genre Classification: A N-gram based Musicological Approach", 7th International Advance Computing Conference, 672-677, 2017.

[11] A comparative study of Internet Architecture of Online Music Streaming services. The impact on the Global Music industry growth Publisher :IEEE

[12] D PW Ellis, classifying music audio with timbral and chroma features. In ISMIR