MUSIC GENRE CLASSIFICATION

Abstract

With the growth of online music databases and easy access to music content, people find it increasing hard to manage the songs that they listen to. Music genre classification is a vital activity that involves categorizing music genres from audio data. In the field of music information retrieval, music genre classification is frequently utilized. The proposed framework deals with three main steps: data pre-processing, feature extraction, and classification. Convolution Neural Network (CNN) is the method used to tackle music genre classification. The proposed system uses feature values of spectrograms generated from slices of songs as the input into a CNN to classify the songs into their music genres. A recommendation system is also implemented after the classification process. The recommendation system aims to recommend songs on each user's preferences and interests. Extensive experiments carried out on the GTZAN dataset show the effectiveness of the proposed system with respect to other methods

Keywords: K-Nearest Neighbor (k-NN); Support Vector Machine (SVM); music; genre; classification; features; Mel Frequency Cepstral Coefficients (MFCC); GTZAN Dataset; Convolution Neural Network (CNN); Recommendation System;

Authors

Anantha Krishna G.K
Department of Computer Science
Visvesvaraya Technological University, Belagavi, India
ananthak.cs18@ sahyadri.edu.in

Suhas A Bhyratae
Assistant Professor
Department of Computer Science
Visvesvaraya Technological University, Belagavi, India
suhasabhyratae.cs@sahyadri.edu.in

Advithiya A Bangera
Department of Computer Science
Visvesvaraya Technological University, Belagavi, India
advithiya.cs18@sahyadri.edu.in

Maruthi
Department of Computer Science
Visvesvaraya Technological University, Belagavi, India
maruthi.me18@sahyadri.edu.in

Deepraj Majalikar
Department of Computer Science
Visvesvaraya Technological University, Belagavi, India
majalikar.cv18@ sahyadri.edu.in

I. INTRODUCTION

In today's world, an individual's music collection generally contains hundreds of songs, while the professional collection normally contains tens of thousands of music files. Music databases are incessantly gaining reputation in a relations to specialized archives and private sound collections. With improvements in an internet services and increase in network bandwidth there is also an increase in number of people accessing the music database. Dealing with extremely large music databases is exhausting and time consuming. Most of the music files are stored according to the song title or the artist's name. This may cause trouble in searching for a song related to a specific genre. The popular music genres are Blues, Classical, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae and Rock. Music has also been divided into Genres and sub genres not only on the basis on music but also on the lyrics as well. This makes music genre classification difficult. Also, the definition of music genre has changed over time. For instance, pop songs that were made fifty years ago are different from the pop songs we have today. Fortunately, the progress in music data and its storage has improved considerably over the past few years. In music genre classification, there are mainly two steps feature extraction, model building. Feature extraction is the technique of extracting distinct characteristics from audio files. Zero crossing rate, Spectral centroid, Spectral roll-off, Spectral bandwidth, Spectral contrast, and Mel-frequency cepstral coefficients are features extracted from the music. The feature extraction module extracts the most relevant data from the raw music data and affects the performance and design of the classifier. There is some amount of work in extracting features of music speech discrimination and speech recognition than extracting features from music signals. All these features are obtained by using the Librosa python library

II. PROBLEM DEFINITION

With the growth of online music databases and easy access to music content, people find it increasingly hard to manage the songs that they listen to. Categorizing music files according to their genre is a challenging task in the area of music information retrieval (MIR). One way to categorize and organize songs is based on the genre. The aim of this project is:

1. To build a machine learning model which classifies music into its respective genre.
2. To compare the accuracies of this machine learning model and the pre-existing models, and draw the necessary conclusions.

III. LITERATURE AND COMPETITION SURVEY

Kris West and Stephen Cox [1] in 2004 prepared a confounded classifier on many sorts of sound elements. They demonstrated capable outcomes on 6-way type characterization errands, with almost 83 % grouping precision on behalf of their greatest framework. As indicated by them the detachment of Reggae and Rock music was a specific issue for the component extraction plan which was assessed by them. They also shared comparative spectral characteristics as well as comparable proportions of harmonic to non-harmonic substance.

Hareesh Bahuleyan [2] in his paper, the task of music genre classification is studied using the Audio-set data. He used two different approaches to solving this problem. The first involves generating a spectrogram of the audio signal and treating it as an image. The second approach consists of extracting time domain and frequency domain features from the audio signals.

Tzanetakis and Cook (2002) [3] addressed this problem with supervised machine learning approaches such as Gaussian Mixture model and k- nearest neighbour classifiers. They introduced 3 sets of features for this task categorized as timbral structure, rhythmic content and pitch content.

Rajeeva Shreedhara Bhat, Rohit B and Mamatha K [4] their research provides the details of an application which performs Music Genre Classification using Machine Learning techniques. The application uses a CNN 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

Alexandros Tsaptsinos,[5] have shown that a HAN (Hierarchical Attention Networks) and other neural-based methods can improve on the genre classification accuracy. They have shown by classifying over a large dataset of nearly half a million song lyrics they have been able to outperform other basic methods. From one standpoint, classification by lyrics will always be inherently flawed by vague genre boundaries with many genres borrowing lyrics and styles from one another. For example one merely need consider cover songs which utilize the same lyrics but produce songs in vastly different genres, or songs which have no lyrical content. In addition, the HAN could be extended to include both a layer at the line and section level, or even at the character level to explore performance.

Rudolf Mayer and Andreas Rauber [6] have employed feature sets derived from the lyrics content, capturing rhyme structures, part-of-speech of the employed words, and style, such as diversification of the words used, sentence complexity, and punctuation. These feature sets were introduced and applied to genre classification. This approach has further been extended to a bigger test collection and a combination of lyrics and audio features, reporting results superior to single feature sets. The combination based on simple feature fusion (early fusion), i.e., concatenating all feature subspaces is however simplistic. Here, they applied late fusion, combining classifier outcomes rather than features. We create a two-dimensional ensemble system, a Cartesian classifier, combining different feature sub spaces from different domains, and different classification algorithms

Athulya K M and Sindhu S [7] they performed classification by using spectrogram feature values from time slices of songs and also used an unknown audio clip to classify into music genre using majority voting system. After the classification task, implement a recommendation system using cosine similarity based on features extracted from the song. The experimental result shows that the proposed system is better performance on the GTZAN dataset than other methods. The future work is to improve the performance of the system and also uses RNN as a model.

Yijie Xu and Wuneng Zhou [8] they have proposed a music genre classification model to implement a small part of the music recommendation system. Based on the spectrogram dataset, this model constructs a convolutional neural network containing SE-Block for effective training.

M. D. Nevetha, A.Nithyasree, A. Parveenbanu and Mrs. JetlinCP [9] developed a classifier on audio files to predict its genre. They worked through this project on GTZAN music genre classification data-set. It explains how to extract important features from audio files. They have implemented a K nearest neighbor.

Rui Yang; Lin Feng; Huibing Wang; Jianing Yao; Sen Luo IEEE Access Year: 2020 [10]. Deep learning-based methods have achieved great success in computing (AI) oriented applications. To advance the event of AI-based IOT systems, effective and efficient algorithms are in urgent need for IOT Edge Computing. Time-series data classification is an on-going problem in applications for mobile devices (e.g., expressive style classification on mobile phones). However, the standard methods require field expertise to extract handcrafted features from the time-series data. Deep learning has been demonstrated to be effective and efficient during this reasonably data. Nevertheless, the prevailing works neglect a number of the sequential relationships found within the timeseries data, which are significant for time-series data classification. Considering the aforementioned limitations, we propose a hybrid architecture, named the parallel recurrent convolutional neural network (PRCNN). The PRCNN is an end-to-end training network that mixes feature extraction and time-series data classification in one stage. The parallel CNN and Bi-RNN blocks specialize in extracting the spatial features and temporal frame orders, respectively, and also the outputs of two blocks are fused into one powerful representation of the time-series data

Daniel Kostrzewa, Piotr Kaminski and Robert Brzeski [11] used convolutional (CNN) and recurrent neural networks (RNN), were used for the genre classification task. In this research, mel-spectrograms with size 128 x 128 were used. They have used only mel-spectrogram and MFCCs as the input data. Precise enlargement of the number of input signals will likely allow for a further increase in classification quality. In this research three types of ensembles were proposed, with two methods of results mixing. The first one is merging outcomes by the usage of a single fully connected layer (FCL). The input of meta-classifier is the output of first-level models, transformed by soft max function and combined together into 3-D matrix. The other one is voting (Vote), where the final classification results from majority voting of base models' predictions. All experiments were performed on Free Music Archive Dataset (FMA) one of the most popular, publicly available databases.

IV. IMPLEMENTATION

1. **Data collection** In order to evaluate Music Rec Net in terms of classification accuracy, the GTZAN data set has been utilized. It contains 1000 music (sampling frequency 22,500 Hz, 16 bits resolution, and 30 s duration). Genres in the GTZAN are blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae and rock. Each genre contains 100 different examples.

2. **Data preprocessing** The split convert function calls both to mel-spectrograms and split songs functions. The split convert function is used to splitting the songs into small chunks. These small chunks are fed into the Librosa python library. The Librosa library is used to visualize the spectrograms. The Librosa library internally uses a Fast Fourier

Transform (FFT) algorithm to analyze the audio signals in time domain and convert spectrograms in the frequency domain. The two mel-spectrograms function is used to convert small chunks of songs into spectrograms and assign labels to small chunks. Before Feature extraction, feature scaling and hot encoding is performed. Feature scaling is the process of normalizing the data ranging between 0 and 1. The one-hot encoding is a method used to convert categorical data into integer data.

The architecture of the proposed system is shown below (see in Fig. 1 & Fig. 2).

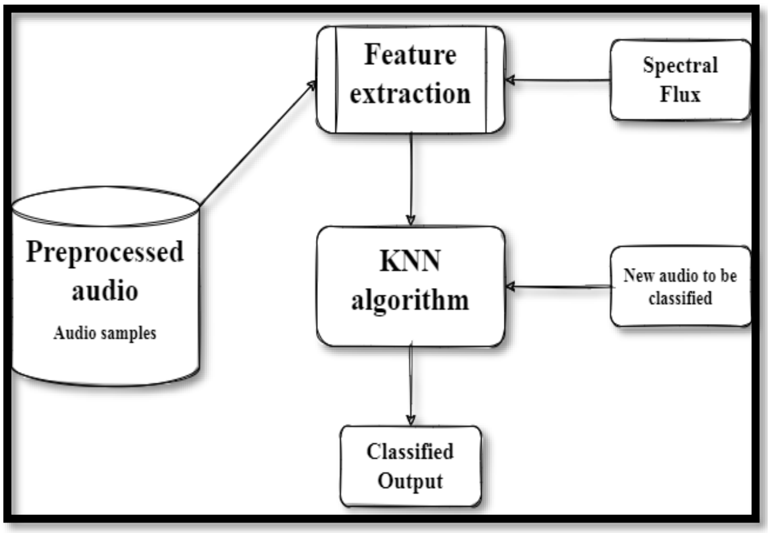

Figure 1: Proposed System

3. **Feature extraction** In feature extraction, extracting meaningful features from slices of spectrograms such as Mel-spectrograms, Spectral centroid, Spectral roll-off, Zero-crossing rate, Spectral bandwidth, and Chromo frequencies. Feature extraction is a valuable activity for analysing and understanding the relations between songs and genres. Extracting the feature values from the slices of songs using Librosa python library. The Librosa python library performs feature extraction using different signal processing techniques. The mel-spectrograms function not to convert spectrograms into images and only using the spectrogram features. These features are stored in a NumPy array (arr spec) with all the numerical values comprising the spectrogram. The arr genres are composed of categorical variables of genres upto 0 to 9. After feature extraction, data augmentation is performed. The function GTZAN generator is created using the sequence method from TensorFlow. It belongs to the data generator classes. The advantage of using data augmentation is to increase the amount of training data. It reduces overfitting and gradually increases the accuracy of the model.

4. **K-Nearest Neighbours (KNN)** algorithm uses ‘feature similarity’ to predict the values of new datapoints which further means that the new data point will be assigned a value based on how closely it matches the points in the training set. Steps to follow:

- The first step of KNN, we must load the training as well as test data.
- We need to choose the value of K, i.e., the nearest data points. K can be any.
- For each point in the test data do the following:

- Calculate the distance between test data and each row of training data, most commonly used method to calculate distance is Euclidean.
- Based on the distance value, sort them in ascending order.
- It will choose the top K rows from the sorted array.
- It will assign a class to the test point based on most frequent class of these rows.

- End

5. **Classification** Compile the model by giving the loss function and optimizer. The optimizer is used as adam optimizer. These will be the parameters for the neural network training process. ReduceLRon Plateau is used to reduce the learning rate when a matric has stopped im- proving. Batch size is mentioned which takes into a number of input samples at each epoch of training. More batch size can increase the accuracy as well as training time. Train the model for 200 epochs. The training is done purely in CPU using TensorFlow. After training, the model is saved using the function 'model.save' function in Keras. For testing purposes, reloading the model. Then pass the test data into it and generate the model accuracy classification report and confusion matrix. Next stage, put the file want to test in the 'unknown' folder. Then reuse the read data function to read the unknown music. This will extract the features from the music and convert it into a format that is readable for the neural network. The unknown data is split into sub-samples. Each sub-sample is then passed into the neural network to predict and give the output label. From the output labels, implement a majority voting system that checks which genre has the most count and then classifies the music belonging to that genre. There are30 slices of unknown audio that will pass into the model and predict the output. All the 30 predicted labels are added to a list. Use a concept of dictionary element in python to find out which label is the most predicted from among the 10 labels.

6. **Recommendation System** after classification of music genre, perform a recommendation system. Users may utilize the recommendation system to propose goods based on their interests. The recommendation of songs using features extracted from time slices of songs. Feature extraction was carried out by means of digital signal processing methods. In the recommendation engine, bring out the top 5 songs that are similar to the input songs. This is done using feature similarity and cosine similarity. The cosine similarity is used to measure the similarity between two items. Using the method find all features when spanned in vector plane are lying close to one another. The test data folder is considered a database for the recommendation. The output will contain songs from different genres.

V. RESULTS AND DISCUSSION

The proposed system architecture is implemented using Jupyter Notebook. Most of the experiments are being carried out on Jupyter Notebook and it is helping to write and run python code through the browser. Compile the model by giving all loss functions and an optimizer to use. We work through this project on GTZAN music genre classification data-set. It explains how to extract important features from audio files. In this classification system, we used K-Nearest Neighbor (K-NN) and Support Vector Machine (SVM) which is developed in Convolutional Kernal with the help of Convolutional Neural Network (CNN) that provide more accuracy compared to the previous system. The testing data-set gives an accuracy of more than 62% (see in Table 1).

Table 1: Statistics of Classifier

k-nearest neighbors (old)	k-nearest neighbors (new)
Correct Predictions: 214	Correct Predictions: 270
Total Predictions: 345	Total Predictions: 328
Accuracy: 0.6202898550	Accuracy: 0.8231707317

VI. CONCLUSION

Music genre classification is a process used for classifying music genres from audio files using k-NN.

In this work, music genre classification is done by using spectrogram feature values from time-slices of songs and also used an unknown audio clip to classify into music genre using majority voting system. The experimental result shows that the proposed system is better performance on the GTZAN dataset than other methods.
The future work is to improve the performance of the system.

REFERENCES

[1] Elbir, A., Çam, H. B., Iyican, M. E., Öztürk, B., & Aydin, N. (2018, october). Music genr -e classification and the recommendation by using machine learning techniques. In 2018 inno - vations in intelligent systems and applications conference (ASYU) (pp. 1-5). IEEE Cheng, y. H., Chang, P. C., and Kuo, C. N. (2020, november). Convolutional neural networks approach for music genre classification. In 2020 international symposium on computer, consumer and control (IS3C) (pp. 399-403). IEEE.

[2] Wu, w., Han, F., Song, G., and Wang, Z. (2018, November). Music genre classification using independent recurrent neural network. In 2018 Chinese automation congress (CAC) (pp. 192-195). IEEE.

[3] Yang,Rui, et al.,2020, "Parallelrecurrent convolutionalneuralnetworks-basedmu-sic genre classification method for mobile devices", vol. 8, pp. 19629-19637, IEEE.

[4] Yu,Yang, et al, 2020, "Deep attention based music genre classification." Neurocom - puting, vol. 372, pp. 84- 91, Elsevier.

[5] Karatana, Ali, and Oktay Yildiz, 2017, "Music genre classification with machine learning techniques" 25th Signal Processing and Communications Applications Conference (SIU), IEEE.

[6] Wu,Wenli, et al, 2018, "Music genre classification using independent recurrent neural network." 2018 Chinese Automation Congress(CAC), IEEE.

[7] Kereliuk, C., Sturm, B.L., Larsen, J.: Deep learning and music adversaries. IEEE Transactions on Multimedia 17(11), 2059-2071 (2015)

[8] Tzanetakis, G., and Cook, P. (2002). Musical genre classification of audio signals: IEEE. IEEE Transactions on Speech and Audio Processing, 10(5), 292–302.

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

[10] Lu, L., Zhang, H. J., and Jiang, H. (2002). Content analysis for audio classification and segmentation. IEEE Transactions on Speech and Audio Processing, 10(7), 504–516.

[11] LonceWyse. 2017. Audio spectrogram representations for processing with convolutional neural networks. arXiv preprint arXiv:1706.09559. Music Genre Classification 15 Department of CSE, SCEM

[12] Vishnupriya S, and K.Meenakshi, "Automatic Music Genre Classification using Convolution

Neural Network", IEEE Conference 2018.
[13] T. Feng, "Deep learning for music genre classification", 2014.
[14] Hareesh Bahuleyan, "Music Genre Classification using Machine Learning Techniques", University of Waterloo, ON, Canada, 2018
[15] Guangxiao Song, Zhijie Wang, Han Fang, and Shenyi Ding. Transfer learning for music genre classification. In International Conference on Intelligence Science, 2017.
[16] George Tzanetakis, Student Member, and Perry Cook. Automatic musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5):293–302, 2002.
[17] Feng et al. Music genre classification with paralleling recurrent convolutional neural network. ArXiv, abs/1712.08370, 2017.
[18] Seonhoon Kim, Daesik Kim, and Bongwon Suh. Music genre classification using multimodal deep learning. In HCI Korea 2016, 2016.
[19] Christine Senac, Thomas Pellegrini, Florian Mouret, and Julien Pinquier. Music feature maps with convolutional neural networks for music genre classification. In the 15th International Workshop, 2017.