Classification of Music into Different Genres Using Machine Learning

ROHIT RANA
Department of Electrical Engineering
IIT Ropar, India
Entry Number: 2020EEM1016
2020eem1016@iitrpr.ac.in

Dr SATYAM AGARWAL
Department of Electrical Engineering
IIT Ropar, India
satyam@iitrpr.ac.in

ABSTRACT-Nowadays a huge number of different types of music are released each day. Managing or classifying this huge number of music manually is very difficult. Machine learning models give the various classification techniques for classification of music into different genres. So, in this paper we use Recurrent neural network (RNN) with Long Short-Term Memory (LSTM) model for music genre classification. By this model we can classify into 10 different genres such as Blues, Classical, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae, Rock. First, we read all the audio files using librosa and extract features such as MFCC(Mel frequency cepstral coefficient), spectral centroid, chroma to distinguish unique genre. Then we used LSTM RNN to train our model and then used Adam optimizer to achieve optimum parameter.

KEYWORDS
LSTM, Music Genres Classification, RNN, Librosa, MFCC

I. INTRODUCTION

At Present, Machine learning has been extensive application to many several areas, for examples healthcare, marketing, security and information retrieval. In this paper, we implement an artificial neural network to music genre classification. We target on to classify music in different-different genres, such as Blues, Classical, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae, Rock.

The main objective of this project is to classify various musical sounds. Today apps like Shazam works on the principle of sound classification. People have a very diverse taste of the music that they hear. So automatic music classification is the need of the hour. In future research, these classification techniques can be used in the specialized search for a user among vast libraries of music from its sound rather than its name.

The classes chosen in this project to classify sounds samples into the following:

i) Blues ii) Classical iii) Country iv) Disco v) Hip hop vi) Jazz vii) Metal viii) Pop ix) Reggae x) Rock

In this paper we use Recurrent neural network (RNN) with Long Short-Term Memory (LSTM) model for music genre classification. Here data used are standard audio sample from public GTZAN. In this dataset 100 samples of each type that is 100*10 = 1,000 sample data. First, we use librosa library to read the all the audio files, Input consists of audio samples from GTZAN. Row of the data consist of following features MFCC (for short term power spectrum(amplitude) of a sound), Spectral Centroid (conveys impression of the brightness of a sound), Chroma (representing 12 distinct semitones of the musical octave which gives musical information about the audio).

After the feature extraction we are going to Splitting the dataset into train test and validation data then build the model for these models. Training the model using an LSTM Recurrent neural Network (RNN), Validating and testing the performance of the model on validation and testing data
respectively. From this approach we show that this music genres classification improves the accuracy.

The remaining section of this paper arranged as follows. In section II we introduce Literature review of similar works. In section III Data source, in section IV Exploratory Data analysis. In section V Model Architecture is discussed. In section VI we discuss experimental results. In section VII we conclude the paper and references are attached in section VIII.

II. LITERATURE REVIEW OF SIMILAR WORK

There is a lot of work that has already been done on Music genres classification. But few of them actually advanced to working on models and accuracy. Here few research works in this field below:

Chun Pui Tang, Ka Long Chui, Ying Kin Yu, Zhiliang Zeng, Kin Hong Wong ," Music Genre classification using a hierarchical Long Short Term Memory (LSTM) model" In this paper direct Long Short Term Memory(LSTM) used for music genres classification and also used hierarchical divide-and-conquer approach to achieve 10 genres classification(Blues, Classical, Country,Disco, Hip-Hop, Jazz, Metal, Pop, Reggae, Rock.). In this paper achieved an accuracy of 50-60% for a direct approach of using LSTM. And in hierarchial divide-and conquer approach achieved from 48.7% to 57.45%(average = 52.975%) accuracy.

Lin Feng, Shenlan Liu, Jianing Yao,"Music Genre Classification with Paralleling CRNN" In this paper proposed a hybrid architecture PRCNN for improving the performance of music genre classification. Here model paralleling CNN and Bi-RNN blocks for extract the features. achieved 88% accuracy without RNN and 92% accuracy with RNN.

Nilesh M. Patil, Dr. Milind U. Nemade,"Music Genre Classification Using MFCC, K-NN and SVM Classifier"

In this paper proposed music genre classification using MFCC features, Chroma features, spectral centroid, spectral roll-off, ZCR are used for feature extraction and train the model using three classifiers k-NN, linear and polynomial kernel SVMs, achieved accuracy 78% by using polynomial kernel SVM.

Many of the papers which implemented CNN's compared their models to other ML techniques including a mixture of Gaussians, SVMs, and CNN's performed favorably in all cases. In one of the papers, their future work was mentioned as more accuracy can be achieved by using the RNN model. Therefore, we decided to focus our efforts on implementing a high-accuracy RNN, with other models used as a baseline.

III. DATA SOURCE

All the musical data is obtained from the public GTZAN dataset from Kaggle. This dataset provides us with 1000 30-second audio clips, all labeled as one out of 10 possible genres and presented as .au files. 100 samples of each type that is 100*10 = 1,000 sample data.

Link: https://www.kaggle.com/andradaolteanu/gtzan-dataset-music-genre-classification

A. *Data Set*

The data used are a standard audio sample from public GTZAN. 100 samples of each type i.e., 100*10 = 1,000 sample data.

B. *Input*

Input consists of audio samples from GTZAN. The row of the data consists of the following features:

- MFCC - for short term power spectrum(amplitude) of a sound
- Spectral centroid - conveys the impression of the brightness of a sound.
- Chroma - representing 12 distinct semitones of the musical octave which gives musical information about the audio.
- Spectral contrast - represents the relative spectral distribution of musical clips.

C. *Output*

The model predicts in one of the above genres.

IV. EXPLORATORY DATA ANALYSIS

Label	ENMFP	manual	last.fm	# tags
Blues	63	96	96	1549
Classical	63	80	20	352
Country	54	93	90	1486
Disco	52	80	79	4191
Hip hop	64	94	93	5370
Jazz	65	80	76	914
Metal	65	82	81	4798
Pop	59	96	96	6379
Reggae	54	82	78	3300
Rock	67	100	100	5475
All	60.6	88.3	80.9	33814

Tabel shows the percentage of GTZAN: identified with Echo Nest Musical Fingerprint (ENMF); Identified with manual search, tagged in the last FM. Number of last FM having count greater than 0 (July 3, 2012).

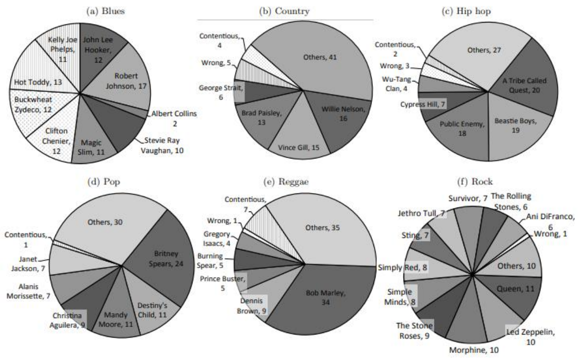


Figure Number of excerpts by a specific artist in 6 categories of GTZAN. Mislabeled excerpts are labeled.

All the above Data analyses for the GTZAN dataset were findings of research [4].

V. ML MODEL ARCHITECTURE

A sequential model consisting of two LSTM layers was used. The model had 0.35 dropout in the recurrent layer and 0.05 dropout between the LSTM units. The number of LSTM units in the first layer was 128 and in the second layer were 32. After it, a dense layer was connected with nodes equal to the number of classes and the activation of SoftMax. The architecture was very simple with few layers. The purpose behind it was to prevent the model from overfitting. The addition of a large number of extra layers may lead to over-fitting.

Figure 1 shows the overview of music genres classification. Here we use GTZAN music dataset to train our model. Then we use Librosa library for extracting the audio features that is Mel frequency cepstral coefficients (MFCC) from the input data, Then we train our model by using Long short-term memory (LSTM) with Recurrent neural network (RNN).

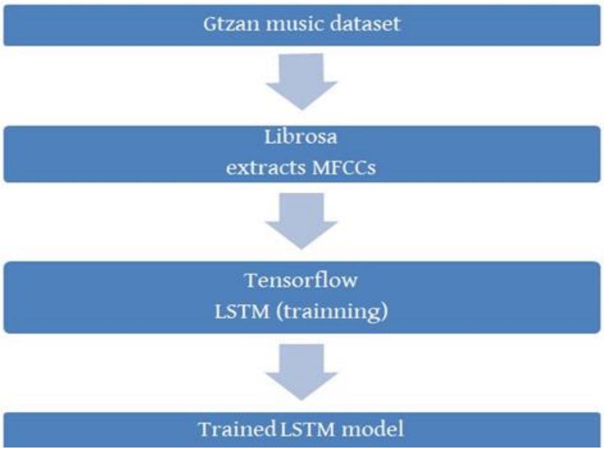


Figure 1: Overview of music genre classification process

1. Mel frequency cepstral coefficients (MFCC)

MFCC (Mel frequency cepstral coefficients) are used for recognising the speech audio signal, music genres classification and similarity measurement of audio signals.

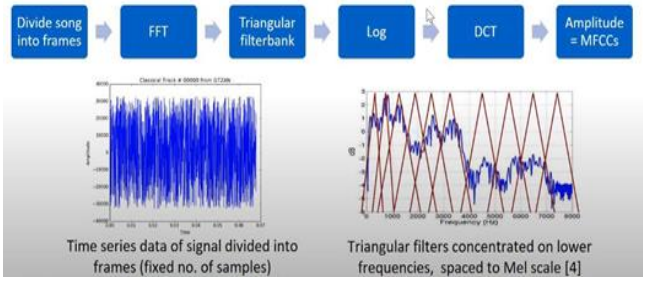


Figure 2: Pipeline of MFCC

Here we used the Librosa library to extract the features that is MFCCs from the audio tracks. Figure 3 shows the Librosa library MFCC feature extraction and figure 4 shows the sample outputs of librosa library.

```
y, sr = librosa.load(file)
mfcc = librosa.feature.mfcc(y=y, sr=sr, hop_length=hop_length, n_mfcc=13)
```

Figure 3: Librosa MFCC feature extraction

```
array([[ -5.229e+02,  -4.944e+02, ...,  -5.229e+02,  -5.229e+02],
       [  7.105e-15,   3.787e+01, ...,  -7.105e-15,  -7.105e-15],
       ...,
       [  1.066e-14,  -7.500e+00, ...,   1.421e-14,   1.421e-14],
       [  3.109e-14,  -5.058e+00, ...,   2.931e-14,   2.931e-14]])
```

Figure 4: Sample outputs of Librosa

2. The Long Short Term Memory Network (LSTM)

The LSTM network used in this paper is a part of RNN.

RNN is different than Neural Network. It stores and uses previous data to predict the new output. LSTM is new better part of RNN which overcomes the long-term dependencies problems. RNN can stores and uses previous data to predict the new output but when the gap between previous state and current state is large then it fails. Figure 5 shows a typical LSTM model

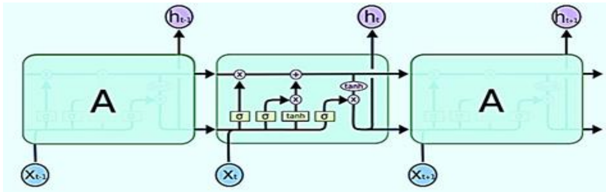


Figure 5: A typical LSTM model

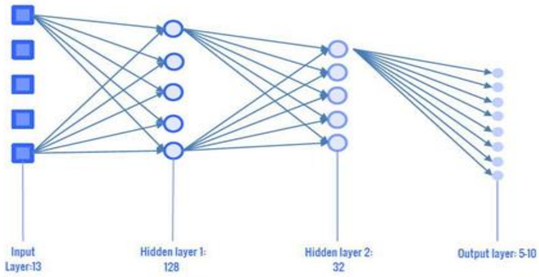


Figure 6: The LSTM network used in our music genre classification problem

VI. Results

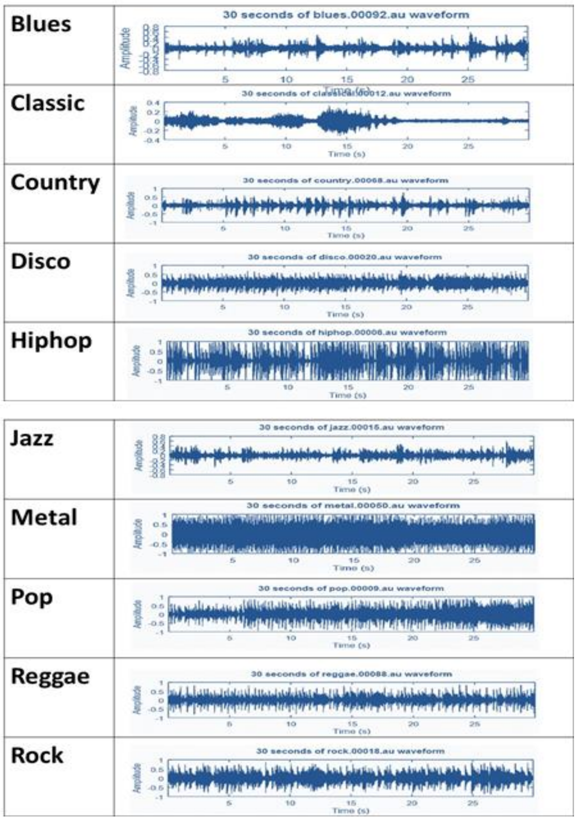
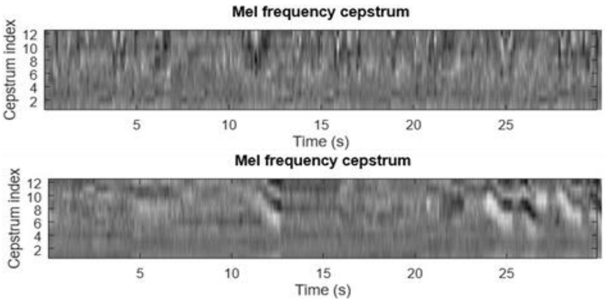
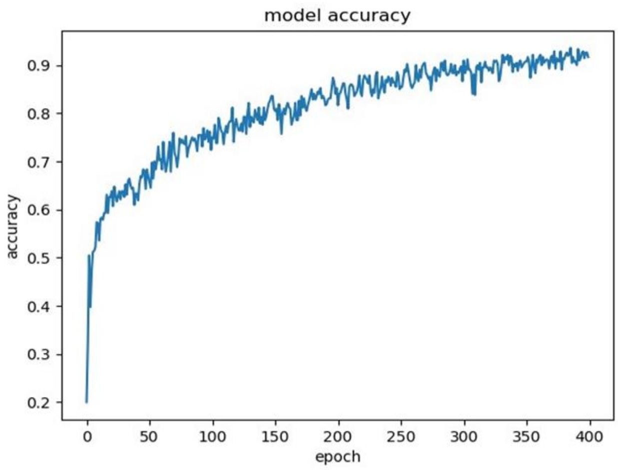


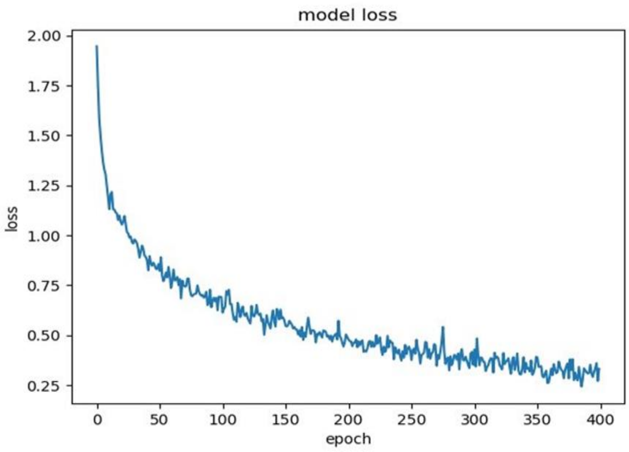
Figure 7: Sample Waveforms of different music genres

We used the GTZAN dataset and it contains Blues, Classic, Country, Disco, Hip-hop, Jazz, Metal, Pop, Reggae and Rock music genres. We split the dataset into train test and validation data. the waveforms shown in figure 7 we have found some similar genres such as Blue, Jazz, Country, Rock, Pop and Reggae.



MFCCs extracted from the audio files were used for the training of the architectures. The model was trained for 400 epochs. The training model accuracy can be observed in the figure below:





As we see the loss decrease with an increase in the number of epochs; similarly, the accuracy increases with the number of epochs.

VII. CONCLUSION

It was observed that through proper feature engineering i.e., MFCCs. The accuracy of the simple sequential model increased up to 97%. It could also be concluded that simple Dense layer architecture would have failed to achieve high accuracy with that simple architecture. LSTM. LSTM is well-suited to classify, process, and predict time series. It also performed excellently on the music genres classification. In future research different Data Augmentation techniques could be applied to increase the data to reduce the problem of over-fitting. Complex architectures could be applied to further increase the Accuracy.

VIII. REFERENCES

[1] http://cs229.stanford.edu/proj2018/report/21.pdf
[2] https://arxiv.org/abs/1802.09697
[3] Mingwen Dong. Convolutional neural network achieves human-level accuracy in music genre classification. CoRR, abs/1802.09697, 2018.
[4] Hareesh Bahuleyan. Music genre classification using machine learning techniques. CoRR, abs/1804.01149, 2018.
[5] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing,10(5):293–302, July 2002.
[6] Brian McFee, Matt McVicar, Stefan Balke, Carl Thomé, Vincent Lostanlen, Colin Raffel, Dana Lee, Oriol Nieto, Eric Battenberg, Dan Ellis, Ryuichi Yamamoto, Josh Moore , WZY, Rachel Bittner, Keunwoo Choi, PiusFriesch, Fabian-Robert Stöter, Matt Vollrath, Siddhartha Kumar, nehz, Simon Waloschek, Seth, Rimvydas Naktinis, Douglas Repetto, Curtis "Fjord" Hawthorne, CJ Carr, João Felipe Santos, JackieWu, Erik, and Adrian Holovaty. librosa/librosa: 0.6.2, August 2018.
[7] Chun Pui Tang, Ka Long Chui, Ying Kin Yu, Zhiliang Zeng, Kin Hong Wong, "Music Genre classification using a hierarchical Long Short Term Memory (LSTM) model", International Workshop on Pattern Recognition IWPR 2018 ,University of Jinan, Jinan, China, May 26-28, 2018.
[8] Lin Feng, Shenlan Liu, Jianing Yao, "Music Genre Classification with Paralleling CRNN", Dec 22, 2017.
[9] Nilesh M. Patil, Dr. Milind U. Nemade,"Music Genre Classification Using MFCC, K-NN and SVM Classifier", IJCERT, February, 2017.