Music Feature Extraction And Recommendation Using CNN Algorithm

Aniket Jha[1,*], *Sagar* Gupta[**], *Priyanshu* Dubey[***], *Aditi* Chhabria[****]
[1]Dept. of Computer Engineering Ramrao Adik Institute of Technology, Nerul, India.

Abstract. In this age of technological advancements, it has become considerably easier for an individual to access a variety of music from a significant number of sources. Today, there are a multitude of songs of varying diversity available to users. Therefore, it becomes difficult for users to manually discover new music that may suit their liking. Thus arises the need for a system that will help the music streaming applications to recommend new music to their users that will befit their music taste, based on some predetermined criteria. With the ever-expanding user and song database, the system must also be dynamic and its recommendations must be up-to-date and accurate. Therefore, there is a strong demand for a well-qualified music recommendation system. The proposed system focuses of technical features of audio. The main purpose of this systems is to classify songs in different genre using Deep Learning Algorithm. There are two main approaches for implementing these system, viz, Feature Extraction and Content Based Filtering.

Keywords: Machine Learning, Logistic Regression, KNN, CNN, Decision Tree, Librosa Library.

1 Introduction

Music Recommendation and Classification System focuses on technical features of audio. This extracted features are used to classify songs in different genre using Deep Learning Algorithm. Further we apply Cosine Similarity on lyrics of songs of same genre to find similarity between them. The top 5 songs with most similarity in lyrics compared to current song played by user is recommend. In this way, we are trying to depict the mood of user based on current song that he is playing and then recommending songs according to his mood.

The objective of this project can be summarized as follows:

1. Use Content-Based Filtering.
2. Categorize songs into different Genre and Emotion.
3. Use lyrics as a parameter to find similarity between songs.
4. Identify the mood of user based on current song that he is playing and recommend songs accordingly.

The growth in the amount of information is drastic in the 21st century .In today's time, it is not rare for situations where users have to interact with large sets of items, to occur. In such cases, the Recommendation system comes into picture.

*e-mail: aditi.chhabria@rait.ac.in
**e-mail: ani.jha.rt18@rait.ac.in
***e-mail: sag.gup.rt18@rait.ac.in
****e-mail: pri.dub.rt18@rait.ac.in

In such scenarios, the system recommends the user items that the system thinks the user may be interested in based on contents of data (Content Based Filtering). Therefore, the user gets to see items that he/she may like without having to search for the same.

In this work, we will use technical features of song such chroma_vector, frequency, amplitude, energy, loudness, valence, liveness, mfcc, etc. This parameters will be used for classification of songs into different genre and emotion using CNN algorithm. The songs of same genre will be further filtered based on lyrics using Cosine Similarity. So basically songs recommended will have same genre and they will have most similar lyrics.

People of all age love listening to music may be their music taste differ. With so many songs around people look for best recommendations. This Application will recommend songs based on current song user is playing. So, if user is listening to devotional song then recommended songs will also be devotional in nature.

2 Literature Survey

Jiosaavn, Spotify, Gaana are examples of application which recommends songs to user. They recommend songs to user on basis of their song preferences, songs they have played frequently, by comparing likings with other users. Most often popular or trending songs are recommended [1].

Following are some recommendation parameters used by existing systems:

Collaborative filtering – In Collaborative Filtering

recommendation are done by comparing liking and disliking between users of particular system [2].
For e.g. If user1 and user2 listens to similar songs then now a new song played by user1 will also be recommended to user2.

Popular and Trending songs - Here recommended songs are either most popular songs or songs which are listen more often these days. Songs are sorted by how frequently they are played in recent time and accordingly most played or most liked songs are recommended. This method doesn't take users preference into account.

Most frequent songs played by user - Here songs are recommended from user's previous listen history. These may be songs which has listened most recently or most frequently.

Genre based classification - In this method of recommendation songs belong to some genre i.e. they may be type classical, pop, jazz, etc. For e.g. If user is currently listening a classical song then recommended songs will also be classical. This is one of the most common music recommendation method.

Outcome of Literature Survey

1. Songs played by user some other day may not depict his current mood.
2. Songs of same genre may have totally different lyrics.
3. Popular and Trending songs may not be of same genre which user is looking for.
4. Filtering based on technical features is a better measure as compared to Collaborative filtering for music recommendations [3].

3 Proposed Methodology

We have used Content Based Recommendation technique i.e. we are recommending using technical features of song. We have extracted technical features of song using librosa library. Further songs are classified into different genre and emotion category using CNN algorithm. Also, we have used cosine similarity to find similarity in song's lyrics.

Content-Based Filtering

Content-based Filtering is an algorithm that uses features to make recommendations. If two objects are similar in feature then if user chooses one the other one can be recommended [4]. Refer Fig 1. Below.

	Feature 1	Feature 2	Feature 3	Feature 4
Product 1	1		1	2
Product 2		1	4	
Product 3	3			1
User Data	2		1	1

Fig 1. Content-Based Filtering

Feature Extraction using LIBROSA

Zero-Crossing Rate – It implies the rate at which signal moves from positive side to negative side or vice-versa. It gives the rate of amplitude variation of song.

Spectral Centroid - It gives the center of mass for a sound for the song. For e.g. Let say a classical song has same frequency throughout its length then its center of mass lies in center. If a rock song has more frequency at its end then its center of mass lies at end.

Spectral Rolloff - It helps to visualize shape of song represented as signal. It gives frequency in relation to total spectral energy, e.g., 85%, lies.

Mel-Frequency Ceptral Coefficients – The MFCC of a song is a small set of features that describe the shape of spectral envelope. It gives the human voice characteristics in song i.e. pitch at which song is sung.

Chrome Frequencies – It represents the 12 distinct chroma of the octave. This is a very powerful representation of audio [5, 6, 7].

Convolutional Neural Network :

CNN is a deep learning algorithm which is used for classification of data indifferent category. It applies various filter matrix on given input features [1,9]. Fig 2. Represents various layers of CNN.

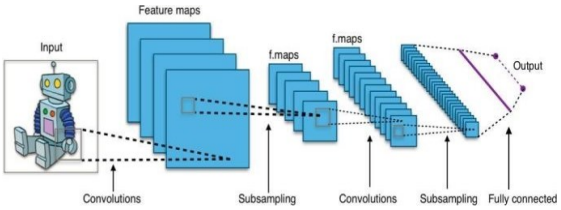

Fig 2. CNN Model (Wikipedia)

CNN for Genre Classification

1. In our CNN model, we have used the Adam optimizer for reducing losses. We have trained our model on 20 epochs.
2. The output layer is using softmax function. All other layers are using RELU as activation function. Sparse categorical crossentropy function is used for evaluating.
3. Adam optimizer gave us the best results as compared to other optimizers.
4. After a certain number of epochs the accuracy of the model achieves threshold i.e. it doesn't increase. We achieved accuracy of 98% for train set and 87% for the test.

Here is a summary of the final architecture:
Layer-1: Input Features (55 Columns)
Layer-2 : 256 Neurons
Layer-3 : 128 Neurons
Layer-4 : 64 Neurons
Output layer: 10 neurons representing 10 different genres
.

Cosine Similarity for lyrics

Cosine Similarity is used to find similarity between two objects represented as vectors. The similarity can

be found by cosine angle between two vectors. The greater the angle the more dissimilar the objects. While if the angle is less, then the objects are similar in nature. For e.g. If angle is 0 (degree) then the objects are similar while If the angle is 90 (degree) then the objects are dissimilar [10, 11, 12]. Refer Fig 3.

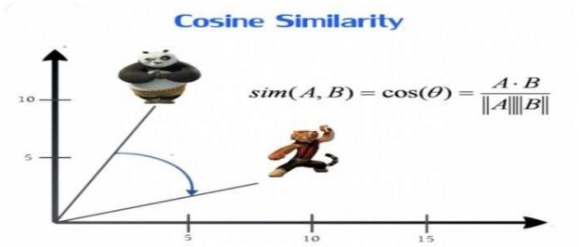

Fig 3. Cosine Similarity

Songs of same genre may have completely different lyrics. For e.g. Songs "Lag ja Gale" and "Lakshya" belongs to the same genre but completely different lyrics and meaning. To remove this inefficiency in recommendation, we found the similarity between songs based on their lyrics. Cosine Similarity will find the similarities in lyrics of two songs by comparing the frequencies of presence of certain words in song lyrics. In "Teri Mitti" song nation-related words are used more frequently. Hence it will be more similar in lyrics with other patriotic songs like "ae watan", "sandhese aate hain", etc. Hence recommended song will have similarity in words and meaning.

Song Genre Classification

We have used Jio Saavn Api to fetch song details such as song url, song name, image, artist name, etc. If the song is played for the first time then its features and lyrics are extracted and stored in a csv. Based on its features we are able to classify the song in a particular genre. Lyrics of the song is used to find top 5 similar songs of same genre. Finally we are recommending songs to the user [13]. Refer Fig. 4.

Fig 4. System flow diagram for Song Recommendation

Song Emotion Classification

We have used python library openCV and haarcascade_frontalface_default XML document to detect face. Then we have used CNN algorithm to classify facial expression into emotion. We also classified songs into different emotion category based on technical features such as energy, beats, chroma_vector, mfcc, tempo, valence, etc. Now we recommend songs with emotion similar to that of user. [14]. Refer Fig. 5.

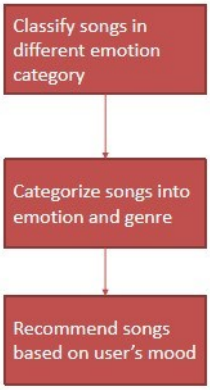

Fig 5. System flow for Song Emotion Classification

4 Results and Simulation

Sr no	Model	Train Accuracy	Test Accuracy
1	CNN	98%	87%
2	KNN	86.04%	82.73%
3	Random Forest	99.81%	82.21%
4	Logistic Regression	73.40%	72.53%

Fig 6. Comparison of different ML Models for Song Genre Classification

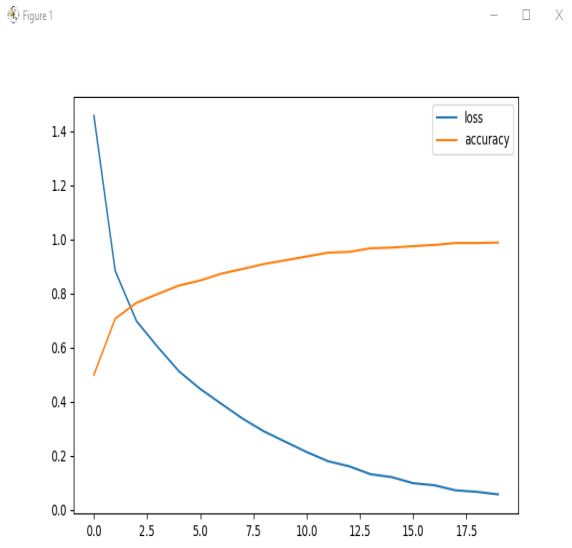

Fig 7. Loss and Accuracy graph for Genre Classification using CNN Model

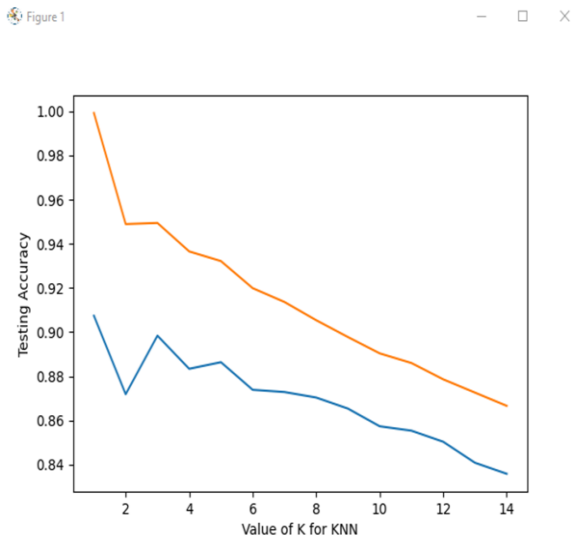

Fig 8. Accuracy Graph for KNN Model

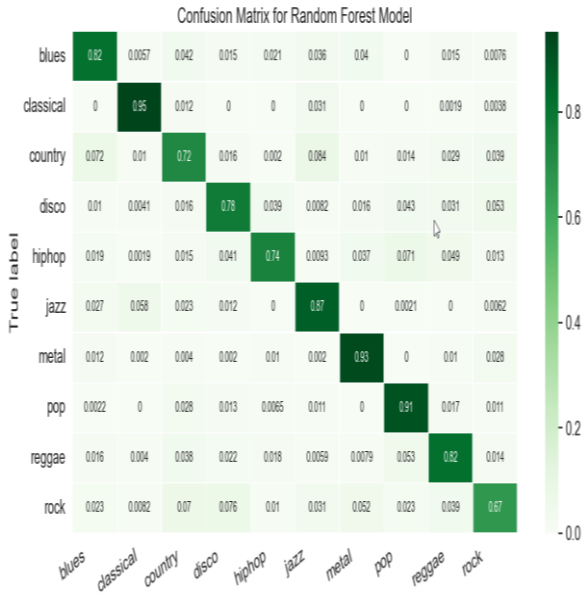

Fig 9. Confusion Matrix for Random Forest

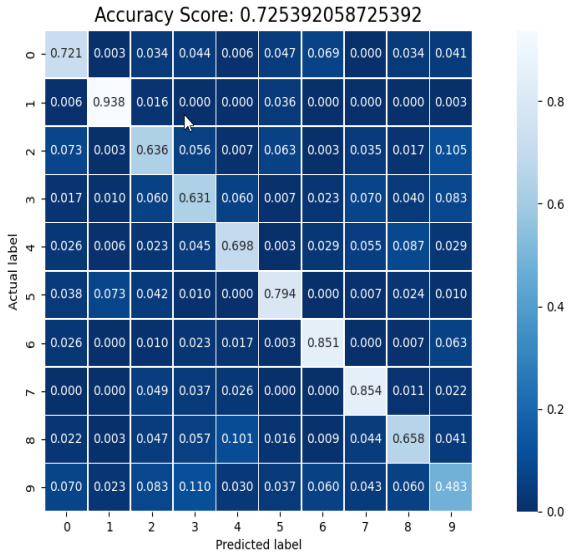

Fig 10. Confusion Matrix for Logistic Regression

Song Emotion Classification :

```
Epoch 7/20
1/1 [==============================] - 0s 999us/step - loss: 0.8017 - accuracy: 0.7000
Epoch 8/20
1/1 [==============================] - 0s 824us/step - loss: 0.7443 - accuracy: 0.7167
Epoch 9/20
1/1 [==============================] - 0s 1ms/step - loss: 0.6911 - accuracy: 0.7333
Epoch 10/20
1/1 [==============================] - 0s 1000us/step - loss: 0.6408 - accuracy: 0.7667
Epoch 11/20
1/1 [==============================] - 0s 664us/step - loss: 0.5921 - accuracy: 0.8167
Epoch 12/20
1/1 [==============================] - 0s 1ms/step - loss: 0.5446 - accuracy: 0.8333

TRAIN ACCURACY : 88.33%
TEST ACCURACY : 83%
```

Fig 11. Result for Song Emotion Classification

5 Conclusion

In this work, we have used feature based recommendation scheme which provides a way to classify song in different genre. Lyrics of the song is our second parameter to find similarity between songs. We have trained our model on GTZAN dataset with 10000 songs and achieved an accuracy 98% in training phase and 87% in test phase. We have classified songs in 10 genre these are reggae, pop, classical, blue, country, disco, hiphop, jazz, metal, rock. Further we have classified songs into different emotion using CNN algorithm and achieved an accuracy of 88% in training phase and 83% in test phase.

6 References

1. T-RECSYS: A Novel Music Recommendation System Using Deep Learning. 2019 IEEE International Conference.
2. Badrul Sarwar, George Karypis, Joseph Konstan, and John Ried. Item-Based Collaborative Filtering Recommendation Algorithms (2004).
3. A Survey of Music Recommendation Systems and Future Perspectives, Conference: The 9th International Symposium on Computer Music Modeling and Retrieval (CMMR) At: London, UK (June, 2012).
4. Application of Content-Based Approach in Research Paper Recommendation System for a Digital Library. (October, 2014) International Journal of Advanced Computer Science and Applications 5(10).
5. librosa: Audio and Music Signal Analysis in Python. Conference: Python in Science Conference (January, 2015)
6. https://librosa.org/doc/latest/index.html.
7. Music Feature Extraction. Sanket Doshi (Dec, 2018)

8. Saad Albawi; Tareq Abed Mohammed; Saad Al-Zawi. Understanding of a convolutional neural network. 2017 IEEE International Conference.

9. Keiron Teilo (O'Shea Aberystwyth University). An Introduction to Convolutional Neural Networks. (November, 2015)

10. Alfirna Rizqi Lahitani; Adhistya Erna Permanasari; Noor Akhmad Setiawan. Cosine similarity to determine similarity measure. 2016 4th IEEE International Conference.

11. Dani Gunawan (University of Sumatera Utara), Mohammad Andri Budiman (University of Sumatera Utara). The Implementation of Cosine Similarity to Calculate Text Relevance. March 2018 Journal of Physics Conference Series 978(1):012120

12. https://towardsdatascience.com/understanding-cosine-similarity-and-its-application-fd42f585296a

13. C. de la Bandera, A.M. Barbancho, L.J. Tard´on, Simone Sammartino, and Isabel Barbancho. Humming Method for Content-Based Music Information Retrieval. ISMIR 2011, (Ismir):49–54, 2011.

14. Cristóbal Veas. Predicting the Music Mood of a Song with Deep Learning. Published in Towards Data Science (Aug 15, 2020).