Content-Based Music Recommendation

ENGR-E511 Machine Learning for Signal Processing

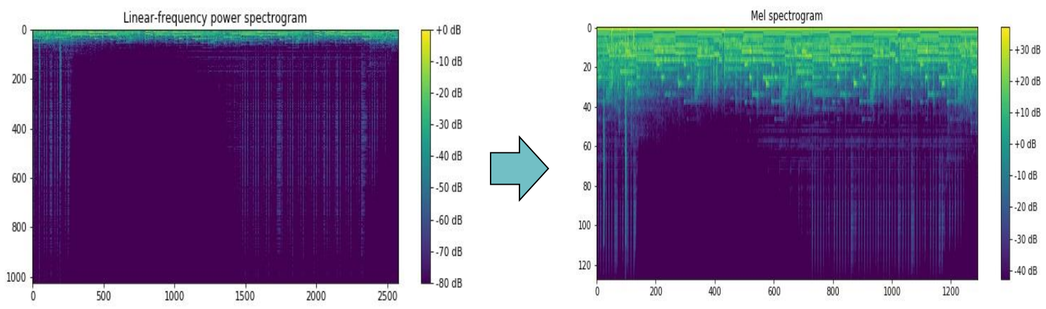
Akshay Naik, Vaishnavi Srinivasan, Rishab Nagaraj

1.Introduction

- The automated generation of music playlists has become a significant problem in the last few years due to the rise of multiple streaming platforms.
- Current state-of-the-art recommendation system depends on the collaborative filtering model.
- This system has some problems like cold start problem, popularity bias.
- Our proposed model will try to tackle these problems by recommending songs purely on the signal data.
- We have used both CNN as well as RNN based approach in our proposed model.

2. Data Collection and Preprocessing

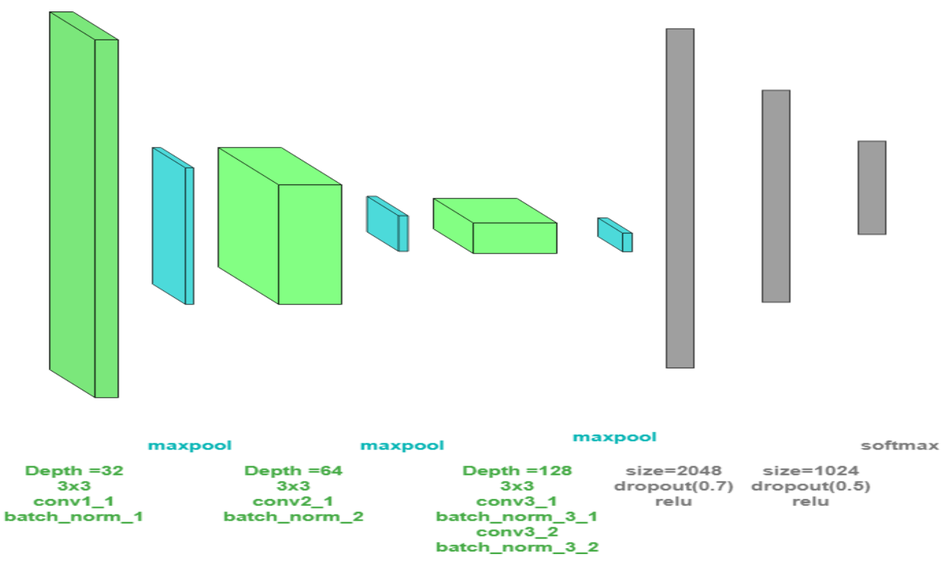
- Collected music data provided by Free Music Archive (FMA) containing 917 GB of creative commons-licensed audios for Music Information Retrieval.[1]
- Used small and medium dataset provided by FMA.
- Small dataset has approximately 4000 tracks with four balanced genres while medium dataset has approximately 13500 tracks with eight unbalanced genres of length 30s.
- The songs are converted to a visual spectrum of frequencies that vary with time. The frequencies are extracted in mel-scale.
- The spectrograms are Short Time Fourier Transforms of the sound signal. They have 2048 long samples window (i.e.) 10ms song period.
- Songs are extracted with a fixed sampling rate of 44100 Hz and are cropped at 29.95 seconds to have consistency between all the music signals.



3. Neural Net Model

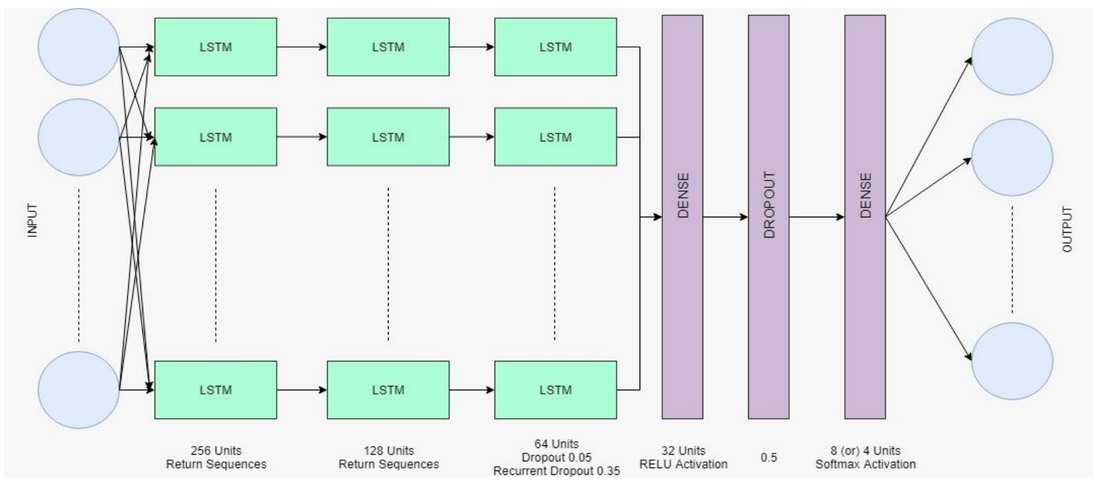
- We trained Genre classification models to transform the data from time-frequency space to genre space wherein each song is represented as a probability distribution over the genre list we considered. Similar songs can be found in this space to generate playlists for the listener.
- We implemented both the neural net architectures using Keras Framework.

3.1. 7-Layer Convolutional Neural Network Model



- Implemented from scratch a 7 layer Convolutional Neural Network.
- The figure mentioned represents the architecture we came up with after fair bit of experimentation.
- The network contains 4 convolution layers and 3 dense layers with batch-normalization and dropout for generalization purpose.
- Used (3*3) convolution filters with ReLU as activation function for hidden layers and softmax for output layer.

3.2. LSTM Neural Network Model

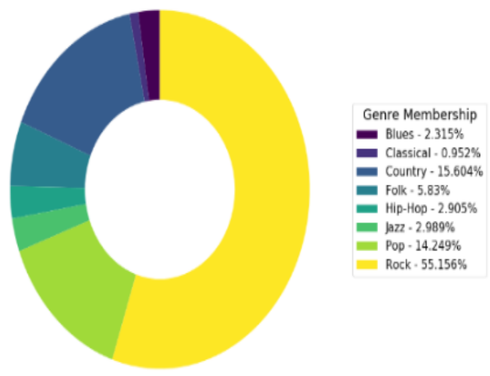


- The mel-spectrogram has a sequence wherein every timestamp strongly relies on immediate predecessor and long-term structure of a whole song.
- We have used LSTM layers to locate this long-term structure of the song and its dependencies across short period of time.
- The model contains 3 LSTM layers with 256, 128, 64 units and 2 dense layers with 32 and 4 or 8 units based on the number of gernes.
- To regularize the model, dropouts are added to the LSTM and dense layer.

Once we convert songs to genre space, we use cosine similarity to find similar songs in this space and create playlists of five closest songs.

4. Sample Results

Genre Membership for 'All the Way to Holloway' which is a Rock song



Playlist	Genre
Cheap Cherry Wine by Junk Boys	Rock
Safe on Your Mountaintop by Fat Spirit	Rock
Black Case by Dlina Volny	Rock
Godspeed (Lyndon Scarfe Remix) by Maxwell Powers	Pop
The Words Don't Work by Orange Peels	Pop

5. Qualitative Results

For the purpose of evaluating our models, we worked under the assumption that a generated playlist for any song would have songs belonging to the same genre, so we assigned each playlist a generated score ranging from 0 to 1 based on the number of songs in the playlist that matched the test/current song's genre.

$$Evaluation\ Score = \frac{\sum_{i=1}^{n} x_i}{n}$$

where n is the number of songs in the playlist (we have considered 5) and $x_i = 1$ if the genre matches and $x_i = 0$ if the genre doesn't match

Neural Network	Dataset	Epochs	Evaluation Score
CNN	Small	100	0.69
CNN	Medium	100	0.57
RNN	Small	50	0.59
RNN	Medium	60	0.48

6. Conclusion

- This project demonstrates how we can apply deep learning techniques to music recommendation task to solve the cold start and popularity bias problem.
- Furthermore, this system can be integrated with collaborative filtering result to create an enhanced hybrid recommendation system.
- In future, we can try training a neural network using the result of collaborative filtering to find latent feature representation like genre space for playlist generation using signal data.

7. References

1] Free Music Archive (FMA): A Dataset for Music Analysis
2] LSTM - Music Genre Classification, Ruoho Ruotsi
3] Recommending Music on Spotify with Deep Learning (CNN), Sander Dieleman
4] Poster template from postersession.com