What to play next? A RNN-based music recommendation system

Miao Jiang, Ziyi Yang
School of Computing, Informatics and Engineering
Indiana University
Bloomington, IN, 47401
miajiang@iu.edu

Chen Zhao
Department of Intelligent Interaction Technologies
University of Tsukuba
Tsukuba, Ibraki, Japan
iacz489@gmail.com

***Abstract*— In the very recent years, development of music recommendation system has been a more heated problem due to a higher level of digital songs consumption and the advancement of machine learning techniques. Some traditional approaches such as collaborator filtering, has been widely used in recommendation systems, have helped music recommendation system to give music listeners a quick access to the music. However, collaborative filtering or model based algorithm have limitations in giving a better result with the ignorance of combination factor of lyrics and genre.**

In our paper, we will propose an improved algorithm based on deep neural network on measure similarity between different songs. The proposed method will make it possible that it could make recommendations in a large system to make comparisons by "understand" the content of songs. In this paper, we propose an end-end model, which is based on recurrent neural network to predict user's next most possible song by similarity. We will make experiments and evaluations based on Million Song Dataset and demonstrate how it outperformed the traditional methods.

Keywords—Music Information Retrieval; Machine Learning

I. INTRODUCTION

Many music listeners have turned to listen online music. The big data technology has made it possible that music listeners could get access to music as they want.

In this case, online service of music subscription has been increasingly popular in the era of cloud computing. The advancement of cloud techniques eases the users to get access to unlimited number of songs. Some streaming music company such as Spotify, Pandora and YouTube are affording users with access to songs to their paid members. It is important that these companies need to maintain their members to keep subscription to gain more revenue.

Playlist is a special function of these streaming apps. Many users feel difficult to create a list from a long list of music. As a result, users tend to play next song by in a random mode or by recommendation. In this case, a successful personalized music recommendation technique has become key to stay their members from jumping to another service.

Techniques of music recommendation have been developed in the previous decade. Collaborative filtering has been widely used in previous research. In addition to that, Millions of Songs data challenge has attracted many researchers to work in this area. However, previous researches only have shown their results in these two aspects[5][8]:

1. Recommending music by collaborative filter on meta data.
2. Recognizing music by man-made humming or recognizing the music note.
3. Recommendation songs by information retrieval

Though these breakthroughs have already shown good results, they are still not enough to recommend music. First of all, these techniques did not measure similarity between different songs by either lyrics and audio, which limits ability of systems. Secondly, if we solely recommend music based on incomplete information, we may lose chances to recommend cross-culture music. Thirdly, traditional information retrieval will rank the music. However, in a personalized system, ranking may make recommend music based on popularity. As a result, we may find that popular songs or famous singers will occupy a great number of subscription, but less famous songs will gradually lose the chance to get recommended.

Recent years, Recurrent neural network(RNN) model[4] has made advancement in sequence data classification. This deep learning model has advances of natural language processing. Machine translation, is a successful application. Also, in field of computer vision, we also have found a good result in sequence video analysis. From those successful stories, we found that it is possible to apply RNN model could to music recognition and comparison due to these reasons:

1. Music and songs are sequential data.
2. Lyrics itself is made of language.
3. Genre may contain meanings.

In our paper, we will apply RNN model to make comparisons of different songs by similarity, which will help recommendation system to give ranking score solely based on a combination factors of music themselves.

II. RELATED WORK

Before we made this paper. Very few literatures have been done to apply deep learning frame work in music recommendation especially on audio. Oord et al. (2013)[6] have proposed a deep learning model in music recommendation, however, his method focus on semantic gap in lyrics.

To make comparisons with different music, researchers have done on distance measurement. Clausen et al. (2000)[2] proposed a search method that views scores and queries as sets of notes. Notes are defined by note onset time, pitch, and duration. Exact matches are supersets of queries, and approximate matching is done by finding supersets of subsets of the query or by allowing alternative sets.

Typke et al. (2003) [7] also view scores and queries as sets of notes, but instead of finding supersets, they use transportation distances such as the Earth Mover's Distance for comparing sets.

As for RNN model, Long-short term memory (LSTM)[3] have been developed to make it possible to make classifications on sequential data.

III. MODEL AND APPROACH

A. Approach

Our general approach is that we divided song's genre or lyrics into sequences and make them into vectors. And then we compare pairwise for different songs.

To make it more specific, suppose we have two different songs P and Q. For P, we make it as sequence (p_1,p_2,…p_n) and we also make Q represented as sequence (q_1,q_2,…q_n). Then our comparison is between two vectors, which element is a chunk of music.

Then what we need to do next is, we need to make formulations:

$$y=f(P,Q) \tag{1}$$

Here, y means the similarity score between two songs, and f is our similarity measurement function. We could treat this function as a model. And we would use a RNN model to train the predicted model.

B. Model

The proposed model in our paper is based on a Long-short term memory based architecture. The motivation behind using an LSTM-based architecture stems from the fact that audio is inherently sequential in nature, and the similarity between two songs (particularly between their audio signals) must in at least some way be determined by the similarities between their sequences over time. While all recurrent neural networks (RNNs) serve the general purpose of modeling patterns in sequential data, LSTMs are often able to "remember" longer patterns and model them better than vanilla recurrent neural networks.

IV. EXPREIMENT AND RESULTS

A. Dataset:

The Million Song Dataset (MSD) [Bertin-Mahieux et al.,2011[1] is a repository that contains more than a million of audio features, metadata, and unique track IDs. This dataset has a great advantage is that, one huge advantage of this dataset is that, all sources have contained unique links to several other interesting sources of data and are provided by this dataset.

We have found that one of the largest datasets that the MSD links to is the LastFM dataset1. And more fortunately, this dataset already contains lists of similar songs for over 500,000 tracks in the MSD, and provides a similarity score for each pair of songs, which definitely meet out needs, which just provide the ground truth labels for the classification problems.

In this dataset, each pair of tracks has already given a similarity score between 0.0 and 1.0. The next step we need to do is set a threshold for this score at 0.5, if the score is over 0.5, we could consider these two songs as similar songs, otherwise we would consider them as not similar.

It is possible to query addition information by API because the MSD designer has already made it possible to query. API2 is one of the example, Among these are the which provides access to lyrics snippets, and the 7digital API 3,which provides audio samples for a large number of tracks contained in the MSD.

We collected the lyrics for a total of 34412 songs, and audio samples for 4240 songs, resulting in a lyrics dataset consisting of 28,000 pairs, and an audio dataset consisting of 1000 pairs. We use cross validations for both to split the datasets into training and testing sets.

B. The experiment Result:

Audio-based :

For Audio based result, we have trained the model by tuning different hyper-parameters.

Based on our experiment, we found that if we set 1 convolutional layer, we could show the best result of 76.5% of accuracy.

Lyrics-based :

For Lyrics based result, we also have trained the model by tuning different hyper-parameters.

Based on our experiment, we found that if we set 1 convolutional layer, we could also see best result of 86.4 % of accuracy.

ACKNOWLEDGMENT

I appreciate Chen Zhao has conduct part of the experiment on Recurrent Neural network. Also, I appreciate that Ziyi Yang has design a UI interface to test the result.

REFERENCES

[1] Thierry Bertin-Mahieux; Daniel P. W. Ellis; Brian Whitman; Paul Lamere. THE MILLION SONG DATASET, Proceedings of the 12th International Society for Music Information Retrieval Conference, 2011

[2] M. Clausen, R. Engelbrecht, D. Meyer, and J. Schmitz. PROMS: a web-based tool for searching in polyphonic music. In ISMIR Proceedings, 2000.

[3] Sepp Hochreiter and J¨urgen Schmidhuber. Long short-term memory. Neural computation, 9(8): 1735–1780, 1997.

[4] Y LeCun, Y Bengio, G Hinton. Deep Learning Nature 521 (7553), 436-444

[5] H. Lee and J. S. Downie. Survey of music information needs, uses, and seeking behaviours: Preliminary findings. In ISMIR Proceedings, pages 441–446, 2004.

[6] A Van den Oord, S Dieleman, B Schrauwen, Advances in neural information processing systems, 2643-2651, 2013

[7] R. Typke, P. Giannopoulos, R. C. Veltkamp, F. Wiering, and R. van Oostrum. Using transportation distances for measuring melodic similarity. In ISMIR Proceedings, pages 107–114, 2003.

[8] R Typke, F Wiering, RC Veltkamp, Proc. 6th International Conference on Music Information Retrieval, 153-160, 2005