Music Generation Based on Feature Engineering and LSTM

Shiwei Mao, Yujia Chen, Guocheng Shao

Tsinghua-Berkeley Shenzhen Institute

Scan it and try our music!

ABSTRACT

Music generation has been an attractive research topic in recent years. Many music generation methods have been developed, including MusicVAE, MuseGAN, Music Transformer, etc. Most previous methods have thrown all the music in the training dataset as input into the network. However, music has different styles. Hence, dividing the different styles of music first and learning them separately may help the model better generate musics. Besides, since music can be considered as a time sery, long short-term memory (LSTM) may be able to implement music generation as well. Therefore, we propose a new music generation method based on feature engineering and LSTM.

INTRODUCTION

The importance of music generation.

- Music is essential in expressing and conveying emotions.
- Music generation can help enlarge the music data.

The problems we want to solve.

- Teach the deep learning network to create music by itself.
- Generate different types of music according to the training data.
- Improve the performance of the network and find a more suitable method to generate music.

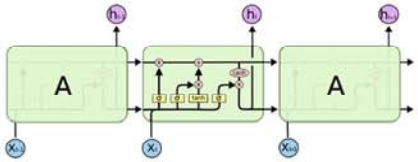
The works we have done.

- Realize a music generation method based on LSTM.
- Cluster the training data set according to feature extraction.
- Compare the generated music with other methods.
- Conduct a survey to better evaluate the music and experiments show that people generally accept LSTM-generated music.

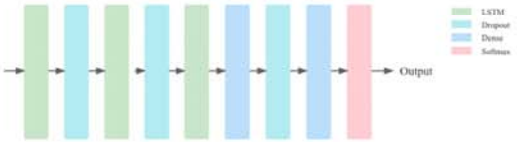
METHOD

LSTM

- A special type of RNN
- Key: the states of cells



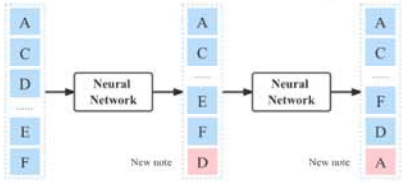
Network Structure



We use LSTM as the main part of the network and cross-entropy as the loss function:

$$L = \frac{1}{N}\sum_{i=1}^{N}\left[-\sum_{j=1}^{M} y_{ij}\ln\left(p_{ij}\right)\right]$$

The cross-entropy loss will compute the probabilities of generating notes. We will then choose the note with the highest probability as the new note.



RESULTS

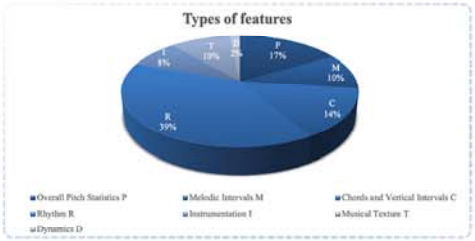
Data set:

Lakh MIDI data set is a collection of 176,581 unique MIDI files.

Data analysis and pre-processing:

Step one: Feature Extraction

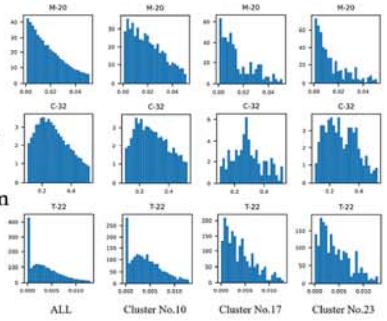
244 features ↓ 7 types



Step two: Feature Selection

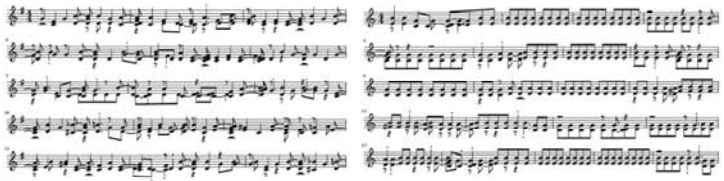
Step three: Music Clustering

We then cluster music by performing clustering on the 29 selected features. According to the dendrogram of hierarchical clustering, the number is set to 30.



Music generation:

We train the network based on the clustered training data and we use different training sets to find the best model.



No.10 Epoch 50 loss 0.32 More variable 😊

No.10 Epoch 89 loss 0.10 Too monotonous 😭

One interesting finding is that the loss value for some clustered training data should not be as small as possible.

Survey:

- Among all the candidates, the highest score is LSTM No.17 Epoch90, a piece of music with variable rhythms and a light and lively style.
- Most pieces by LSTM achieved scores higher than MusicVAE.
- The music generated by LSTM can be accepted by most people.

CONCLUSION

- We extract features from MIDI files and cluster the reduced features by using K-means algorithm.
- We generate several music by LSTM on the clustered data.
- We conduct a survey to better evaluate our musics and the result shows that our method is comparable to the recent works.

REFERENCE

[1] C. Hernandez-Olivan and J. R. Beltran, "Music composition with deep learning: A review," arXiv preprintarXiv:2108.12290, 2021.

[2] A. Roberts, J. Engel, C. Raffel, C. Hawthorne, and D. Eck, "A hierarchical latent vector model for learning long-term structure in music," in International conference on machine learning. PMLR, 2018, pp. 4364 – 4373