Generate Xi'an Drum Music Based on Compressed Coding

Wen Ru[1], Haochen Wang[2], Jiabao Li[3], Yuxuan Lin[3, 4, 5], Jiang Wu[5]
1. School of Humanities and Social Science, Xi'an Jiaotong University, Xi'an, 710043 China
2. Computer Science, Vrije Universiteit Amsterdam & Universiteit van Amsterdam, Amsterdam, Netherlands
3. School of Qianxuesen, Xi'an Jiaotong University, Xi'an, 710043 China
4. School of Artificial Intelligence, Xi'an Jiaotong University, Xi'an, 710043 China
5. School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, 710043 China
Mail: haochen.wang@student.uva.nl, linyuxuan@stu.xjtu.edu.cn

Abstract: Xi'an Drum Music is a traditional form of Chinese music and its notes are recorded by Chinese Characters. Because Xi'an Drum Music is composed and translated by the elder musicians, Xi'an Drum Music becomes difficult to be protected. In this article, we use sparse coding and compressed coding to transfer the Chinese Character recording to genre and lyrics of Xi'an Drum Music. Based on our dataset of Xi'an Drum Music, we set up a method to generate Xi'an Drum Music similar to Huffman coding, a model named Xi'an Drum Music Generation via Variational AutoEncoder (DMGVAE) and the accuracy of Xi'an Drum Music generation increases to 0.73. Compressed coding on Xi'an Drum Music shows a novel method to generate Xi'an Drum Music by compressed format, making potential application for generating the sparse traditional Chinese Music such as Xi'an Drum Music.

Key Words: Xi'an Drum Music, Compressed Coding, Autoencoder, Deep Learning

1 Introduction

Xi'an Drum Music is an important part of Chinese traditional music and reflects the history of Chinese traditional music and ancient culture (CAO and LI 2017). The generation of the Xi'an Drum Music is by the musicians of the Xi'an Drum Music and the notes of Xi'an Drum Music are recorded by Chinese character, making it difficult to record the genre and lyrics of Xi'an Drum music and can only be composed by fork artists of Xi'an Drum Music (Cheng 2005). As time goes by, Xi'an Drum Music is getting lost because there are less and less musicians of Xi'an Drum Music to transfer the original Chinese character notes of Xi'an Drum Music to its genre and lyrics. To understand Xi'an Drum Music much more and generate the genre of Xi'an Drum Music based on Chinese character notes, Xi'an Drum Music also needs to be protected by the means of the digital method and deep learning.

In recent years, deep learning has enabled music generation better and better performance with different methods of deep network (Briot, Hadjeres, and Pachet 2017). The innovative application of original models such as LSTM, VAE can generate music with its own music style, melody, and lyrics. However, the notes of Xi'an Drum Music are recorded by Chinese characters and the generation of Xi'an Drum Music is by folk artists of Xi'an Drum Music. These two processes make Xi'an Drum Music hard to be digitalized, stored, and fertilized in music generation. New method remains to be developed to protect Xi'an Drum Music, which is an important role of the traditional Chinese music. There remains to be explored whether different coding methods will make sense to the dataset of Xi'an Drum Music in the deep learning model for music generation to protect Xi'an Drum Music.

Consequently, there remains to be some problems on music generation based on the coding of Xi'an Drum Music. First, no efficient methods are discovered to code Xi'an Drum Music and generate new music based on music genre, especially based on the original music genre. Thus, an effective coding method for music genre transfer and generation such as Xi'an Drum Music genre is deadly needed. Second, it is difficult to model after music genre with the limitation of crowd of music dataset. Since the dataset of Xi'an drum music is not enough for us to generate the music, it is also a challenge to generate music by expanding the Xi'an drum music dataset. In this article, we use different way of coding format to generate Xi'an Drum Music by the means of deep learning model and compare with the Huffman coding of Xi'an Drum Music.

2 Related Works

Music generation is a hot topic in artificial intelligence application (Briot, Hadjeres, and Pachet 2017). Based on the pre-trained deep learning model for music generation, the users can compose their own music based on a certain music style, which also makes it easy for modern musicians to compose and develop new music. A well-trained music generation model can generate the music with a customized genre or to imitate the music genre of the famous musicians and shows the potential of recreating the new music with ancient music style and genre. It is also possible for Xi'an Drum Music to be protected and generated by deep learning models, which is one of the best ways to model after the Xi'an Drum musicians to generate the Xi'an Drum Music based on music generation model.

With the approach to the deep learning and cloud computing, there are more and more related deep learning models to generate music. Besides, it is easier and easier to utilize the methods of neural network and deep learning methods to create new music and protect ancient music. There are many classical music generation models to generate music such as LSTM, autoencoder, VAE, and stacked autoencoder (Ghosal and Kolekar 2018). Music transferring is related to music genre recognition, which is well researched. Recently, Deepanway, et al design a music genre recognition model by deep learning and transfer learning, which can work well in music genre recognition.

As for music genre transfer, Chien-Yu Lu, et al designed a model which can transfer the multi-modal music (Lu et al. 2019). Gino Brunner, et al designed a symbolic music genre transfer by combining VAEs with GAN and the model is able to transfer the music genre (Brunner et al. 2018).

However, coding for the notes and genre is the first and important step of Xi'an Drum Music generation models because the notes of Xi'an Drum Music are recorded by Chinese characters, which is seldom studied. The total number of Chinese characters recording notes of Xi'an Drum Music is ten and there are many duplicates for the pitch and duration of each note. Therefore, coding methods for the music wave or the music genre also matter in the process of music genre generation (Briot, Hadjeres, and Pachet 2017), especially for Xi'an Drum Music. There are not thousands of Xi'an Drum Music recorded, so the dataset for Xi'an Drum Music is small regarding different songs of Xi'an Drum Music. Pitches and durations of Xi'an Drum Music are confined to a small range. Using regular coding methods such as music wave or MIDI for Xi'an Drum Music may lead to the wrong normal distribution for durations, because the unbalanced distribution of pitches.

In this article, we will analyze, and discuss the influence of coding methods, such as as 0-1 standardization, Z-Score standardization, compressed coding and Huffman coding methods of Xi'an Drum Music dataset towards the results of the Xi'an Drum Music genre generation.

3 Methodology

3.1 Workflow

The dataset of Xi'an Drum Music is collected by us and we transfer the original sources of Xi'an Drum Music into MIDI files. We then get the mapping of the pitches and durations from MIDI files. The Xi'an Drum Music dataset then encoding and decoding by normalized transferring, Z-score normalized transferring and Huffman coding. Variant autoencoder (VAE) model is developed to generate the Xi'an Drum Music based on three types of coding methods and compared with the original mapping of Xi'an Drum Music dataset. We use the accuracy based on VAE model as the metric of music generation performance.

3.2 Dataset

We use Xi'an Drum Music data as our raw data, which is a set of drum music staffs. The drum music staffs consist of Chinese character staffs (named Suzipu in Chinese) and standard staffs in drum music. Suzipu staffs are the traditional Chinese musical characters to record the pitches of the music and there are totally 10 Chinese characters. Generated by the drum music by hearts and brains of Xi'an Drum Music musicians, Suzipu staffs will be translated into standard staffs. The standard music staffs are record by experts of Chinese traditional music researchers. The original dataset of Xi'an Drum Music is then combined of Suzipu staffs together with standard music staffs, and we transferred them into the standard MIDI files to get the input dataset of Xi'an drum music dataset.

Then the pitches, durations and velocities are extracted from MIDI files to make up the elements of the dataset. Before loading into the model, we try various standardization methods to re-code the MIDI dataset such as 0-1 standardization, Z-Score standardization, compressed coding and Huffman coding.

Since the pitches and durations of Xi'an Drum Music dataset reflect the genre of Xi'an Drum Music, the value of pitches and durations should be considered as continuous number. To make the values of pitches and durations in the same scale, 0-1 standardization is adopted to preprocess our dataset as follows:

$$f_{0-1}(pitch_i) = \frac{pitch_i - pitch_{min}}{pitch_{max} - pitch_{min}}$$

$$f_{0-1}(duration_i) = \frac{duration_i - duration_{min}}{duration_{max} - duration_{min}}$$

where $pitch_i$, $duration_i$ is the pitch and duration of the ith notes in Xi'an Drum Music dataset, $pitch_{min}$ and $duration_{min}$ denote the minimum value of pitches and durations, $pitch_{max}$ and $duration_{max}$ denote the maximum value of pitches and durations in Xi'an Drum Music dataset. By 0-1 standardization, the value of pitches and durations are in range from 0 to 1 and be flattened. Besides, we use Z-Score standardization to normalize the pitches and durations of Xi'an Drum Music Dataset. The Z-Score standardization is shown as follows:

$$f_{z-score}(pitch_i) = \frac{pitch_i - \mu_p}{\sigma_p}$$

$$f_{z-score}(duration_i) = \frac{duration_i - \mu_d}{\sigma_d}$$

where $pitch_i$, $duration_i$ is the pitch and duration of the ith notes in Xi'an Drum Music dataset, μ_p, μ_d is the average value of pitches and durations, σ_p, σ_d is the standard deviation of pitches and durations in Xi'an Drum Music dataset.

To compress the repetitive sequences of Xi'an Drum Music dataset, we use the compressed coding to recode the dataset. The pitches and durations of Xi'an Drum Music dataset are highly repetitive in a continuous sequence and the most optimal coding method is run-less encoding (Hauck 1986). The compressed coding algorithm based on run-less encoding is shown as follows:

$$X = \{x_1, x_1, \dots, x_2, x_2, \dots, x_n, x_n, \dots\}$$

Here, X denotes the set of our observation from Xi'an Drum Music dataset. By calculate the repetitive number of each value, the compressed coding will be following:

$$f_{compress}(X) = \{N(x_1), x_1, \dots, N(x_2), x_2, \dots, N(x_n), x_n, \dots\}$$

Where $N(x_i)$ denotes the number of repetitive x_i in the continuous sequence. Eventually, we try the Huffman coding for our Xi'an Drum Music dataset (Liu and Žalik 2005). Suppose we have the same repetitive and continuous sequence from Xi'an Drum Music dataset as $X = \{x_1, x_1, \dots, x_2, x_2, \dots, x_n, x_n, \dots\}$, then the preprocess for Xi'an Drum Music by Huffman coding algorithm is as follows:

$$f_{Huffman}(X) = \{g(x_1), g(x_1), \dots, g(x_2), g(x_2), \dots, g(x_n), g(x_n), \dots\}$$

where $g(x_i)$ denotes the Huffman coding of the pitches or durations for our Xi'an Drum Music dataset.

By generating the dataset, we can get a larger dataset of Xi'an Drum Music to train our model well. The coding methods show the same information regards on Xi'an Drum Music staffs, thus they are identical to music each other, even if their coding methods are different. Simultaneously, we joint the different methods of the coding methods together to

generate a mixed-style dataset of Xi'an Drum Music. The assumption is that even it is a little different to use dataset mixed to enlarge the data from Xi'an Drum Music, it makes a little different to the style (genre) for Xi'an Drum Music.

3.3 Model and architecture

Our model is based on the autoencoder. Autoencoder is one of the best models for music generation and is effective for our Xi'an Drum Music generation. The coding methods of Xi'an Drum Music play the main role for Xi'an Drum Music generation and the same model of music generation can be applied to validate the coding methods. We design the encoder layers based on embedding layers, dropout layers and LSTM layers to fit the different input formats from different coding methods. Basically, the encoder layers need a universal input for prepared Xi'an drum music pitch sequences, which will be regarded as different strings to be conveyed into the input layer of our model. Then the features of autoencoder based on LSTM are included in our model as the main architecture of our model. Ultimately, the sequences are obtained by the decoder layer of our model.

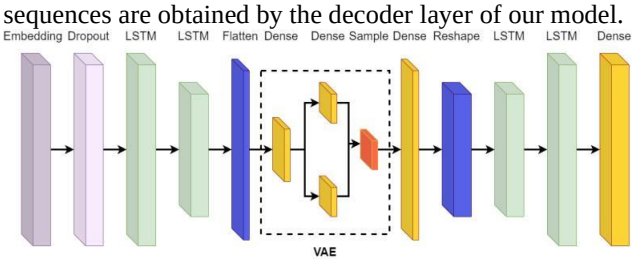

Fig. 1: The encoder architecture of Xi'an Drum music generation model

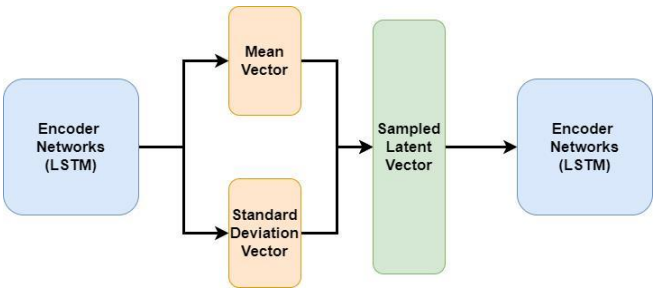

Fig. 2: The decoder model of Xi'an Drum music generation model

3.4 Metrics

To validate the consequences of our datasets, models, and results, we calculate the SSD of the prediction and validation dataset. Xi'an Drum Music genre is nothing different with the duration time but with the styles of the duration, so that the outline of the duration style should also be minimized and optimized when validated the predictions for Xi'an Drum Muaic genre in our music genre generation model. Verification of Xi'an Drum Music can also be verified by the experts because the music style of Xi'an Drum Music is different from original. When Xi'an Drum Music generation is good enough, we can have less value of the loss function and this can be the best method to simplify our verification metrics, although it may be a little tough for our model training for the least loss value.

4 Experiments and Results

4.1 Dataset Generation

To get the dataset file, we transfer the image files of Suzi staffs and standard staffs into MIDI files manually. Then the pitches, durations and velocities can be extracted from MIDI files by using python pretty MIDI module. Therefore, we get the raw data for our original Xi'an Drum Music dataset by directly extracting from MIDI files which are processed manually.

To check whether the coding styles will influence the results of our models based on our dataset, we not only use the raw original Xi'an Drum Music datasets but also attempt to code the dataset by different coding methods. The 0-1 normalization is calculated by the minimum and maximum of the pitch series, duration series and velocity series. The Z-Score normalization is depended by the mean value of the dataset series and the standard deviation. In contrast to the two methods above, the compressed coding method doesn't base on the calculate of the statistic features of the dataset series but to be determined by the features of the dataset itself. The same value of the pitches and durations will be conflated, resulting in a new data structure similar to dictionary for Xi'an Drum Music raw data.

The statistical results of Xi'an Drum Music can be listed as following figures:

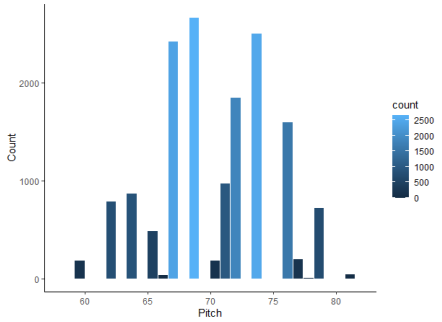

Fig. 3: The distribution of pitch in Xi'an Drum Music dataset

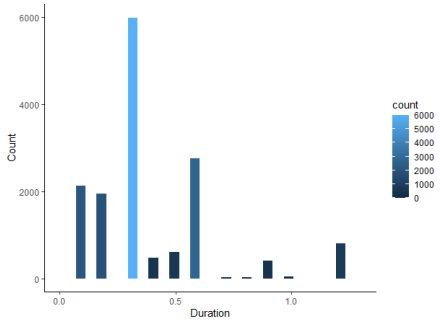

Fig. 4: The distribution of duration in Xi'an Drum Music dataset

The dataset of MIDI is a totally of 39 songs of Xi'an Drum music, which is also not enough for the dataset. The reason for constructing different datasets is that the data is not enough for the deep learning model to train as the dataset. Apart from the conclusions above, the distribution of Xi'an Drum Music dataset shows that the pitches and duration of Xi'an Drum Music are unbalances in series and the re-coding for Xi'an Drum Music is necessary.

4.2 Music Generation

Xi'an Drum Music generation model is based on autoencoder and the dataset is based on different format of Xi'an Drum Music. Drum music of MIDI format is transferred into the mixed dataset by extracting the pitch and the duration, because the pitch and the duration are two of the most important elements for Xi'an Drum Music. The music sequence is divided into sub-sequences when using different pitches.

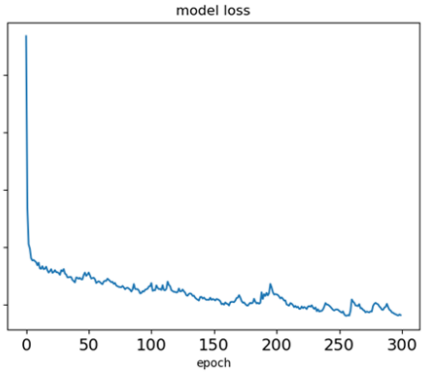

Fig. 5: The trend of model loss in DMGVAE

The results of Xi'an Drum Music generation model show that the training process is successful with the shake of the low level. And the accuracy is gradually increasing with shake at a large level, finally stabilizing at 0.73.

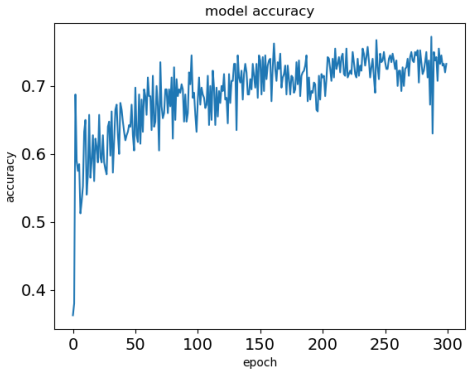

Fig. 6: The model accuray training in DMGVAE

5 Discussion and Conclusions

From the different coding methods of Xi'an Drum Music, we find that it is possible and effective with the MIDI coding to the mixed pitch-duration coding. Based on the autoencoder model for Xi'an Drum Music generation, the architecture we set is suitable for the secrete dataset of Xi'an Drum Music dataset with the representation of 0.73 accuracy for Xi'an Drum Music.

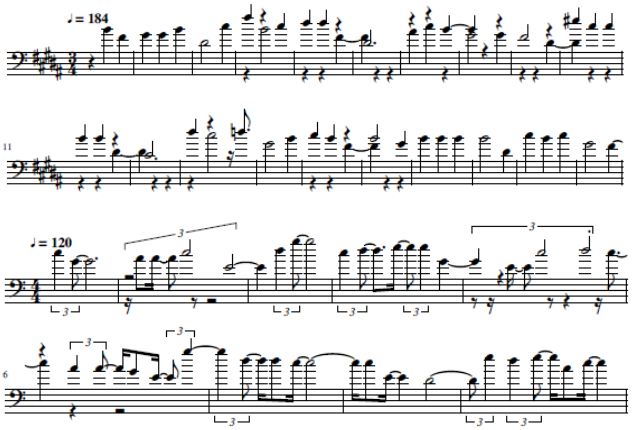

Fig. 7: The original Xi'an Drum Music sequence and generation results in DMGVAE

Compared with original coding method of Xi'an Drum Music, the coding of 0-1 normalized transferring, Z-score normalized transferring and Huffman coding performs better than the original unbalanced dataset, which leads to the accuracy of 0.73 via our VAE model. This result shows that the coding method for Xi'an Drum Music can decrease the degree of unbalanced data and improves the accuracy of Xi'an Drum Music, proposing to be used in music generation as well in future.

References

[1] Briot, Jean-Pierre, Gaëtan Hadjeres, and François-David Pachet. 2017. "Deep Learning Techniques for Music Generation--a Survey." *arXiv preprint*, arXiv:1709.01620.

[2] Brunner, Gino, Yuyi Wang, Roger Wattenhofer, and Sumu Zhao. 2018. "Symbolic Music Genre Transfer with Cyclegan." *In 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, 786–93.

[3] CAO, Gengxian, and Donghe LI. 2017. "Research on the Value of Xi'an Drum Inheritance under the Complex Information System." *Journal of Xi'an Jiaotong University (Social Sciences)* (3): 17.

[4] Cheng, Yu. 2005. "Xi'an Old Music in New China:" Living Fossil" or" Flowing River"?"

[5] Ghosal, Deepanway, and Maheshkumar H Kolekar. 2018. "Music Genre Recognition Using Deep Neural Networks and Transfer Learning." *In Interspeech*, 2087–91.

[6] Hauck, Edward L. 1986. "Data Compression Using Run Length Encoding and Statistical Encoding."

[7] Liu, Yong Kui, and Borut Žalik. 2005. "An Efficient Chain Code with Huffman Coding." *Pattern Recognition*, 38(4): 553–57.

[8] Lu, Chien-Yu et al. 2019. "Play as You like: Timbre-Enhanced Multi-Modal Music Style Transfer." *In Proceedings of the AAAI Conference on Artificial Intelligence*, 1061–68.