Music Generation of Traditional Chinese Pentatonic Scale

Contributor: Fanzhi Jiang, Liumei ZHANG, Kexin Wang, Xi Deng, Wanyan Yang

Recent studies demonstrate that algorithmic music attracted global attention not only because of its amusement but also its considerable potential in the industry. Thus, the yield increased academic numbers spinning around on topics of algorithm music generation. The balance between mathematical logic and aesthetic value is important in music generation.

1. Introduction

During the past few decades, the field of computer music has precisely addressed challenges surrounding the analysis of musical concepts [1][2]. Indeed, only by first understanding this type of information can we provide more advanced analytical and compositional tools, as well as methods to advance music theory [2]. Currently, literature on music computing and intelligent creativity [1][3][4] focuses specifically on algorithmic music. We have observed a notable rise in literature inspired by the field of machine learning because of its attempt to explain the compositional textures and formation methods within music on a mathematical level [5][6]. Machine learning methods are well accepted as an additional motivation for generating music content. Instead of the previous methods, such as grammar-based [1], rule-based [7], and metaheuristic strategy-based [8] music generation systems, machine learning-based generation methods can learn musical paradigms from an arbitrary corpus. Thus, the same system can be used for various musical genres.

Driven by the requirement for widespread music content, more massive music datasets have emerged in the genres of classical [9], rock [10], and pop music [11], for instance. However, a publicly available corpus of traditional folk music seems to pay little attention to the niche corner. Historically, research investigating factors associated with music composition from large-scale music datasets has focused on deep learning architectures, stemming from its ability to automatically learn musical styles from a corpus and generate new content [5].

Although music possesses its special characteristics that distinguish it from text, it is still classified as sequential data because of its temporal sequential relationship. Hence, recurrent neural networks (RNN) and its variants are adopted by most music-generating neural network models that are currently available [12][13][14][15][16][17]. Music generation sequence models were often characterized by the representation and prediction of a number of events. Then, those models can use the conditions formed by previous events to generate the current event. MelodyRNN [18] and SampleRNN [19] are representatives of this approach, with the shortcoming that the generated music lacks segmental integrity and a musical recurrent structure. Neural networks have studied this musical repetitive structure, called translation invariance [20]. Convolutional neural network (CNN) has been influential in the music domain, stemming from its excellence in the image domain. This regional learning capability is sought to migrate to the translational invariance of the musical context. Some representative work has emerged [21][22][23] to use deep CNN for music generation, although there have been few attempts. However, it seems to be more imitative than creative in music, stemming from its over-learning of the local structure of music. Therefore, inspired by whether it is possible to combine the advantages of both structures, they used compound architectures in music generation research [12][24][25][26].

Compound architecture combines at least two architectures of the same type or of different types [5] and can be divided into two main categories. Some cases are homogeneous composite architectures that combine various instances of the same architecture, such as the stacked autoencoder. Most cases are heterogeneous compound architectures that combine various types of architectures, such as a RNN Encoder-Decoder that combines the RNN and autoencoder. From an architectural point of view, we can conduct compositing using different methodologies.

- *Composition*—Combination of two architectures of the same type or of different types. For instance, the bidirectional LSTM [15] combines two RNNs to analyze music semantic contexts from temporal forward and inverse; and RNN-RBM architectures combine RNN architectures and RBM architectures [14].

- *Refinement*—Refinement and specialization of a model by additional constraints. The sparse autoencoder architecture is an example of a specialized solution to the note sparse coding problem on top of the autoencoder architecture [27] and the variational autoencoder (VAE) [28].

- *Nesting*—Nesting one model into another structure to form a new model. Examples include stacked autoencoder architectures [29] and RNN encoder-decoder architectures, where two RNN models are nested in the encoder and decoder parts of an autoencoder, so we can also call them autoencoders (RNN, RNN) [16].

- *Instantiation*—The architectural pattern is instantiated into a given architecture. For a case in point, the Anticipation-RNN architecture instantiates a conditional reflection architectural pattern onto an RNN and the output of another RNN as a conditional reflection input, which we can call conditional reflection (RNN, RNN) [17]. The C-RBM architecture is a convolutional architectural pattern instantiated onto an RBM architecture, which we can note as convolutional (RBM) [30].

2. Deep Learning-Based Music Generation

- *RNN-based music generation*. This work [12] is an RNN architecture with a hierarchy of cyclic layers that generates not only melodies but also drums and chords. The model [13] well demonstrates the ability of RNNs to generate multiple sequences at the same time. However, it requires prior knowledge of the scale and some contours of the melody to be generated. The results demonstrate that the text-based Long short-term Memory (LSTM) performs better in generating chords and drums. MelodyRNN [18] is probably one of the best-known examples of neural networks generating music in the symbolic domain. It includes three RNN-based variants of the model, two variants aimed at musical structure learning, lookback RNN and Attention RNN. Sony CSL [31] proposed DeepBach, which could specifically compose a polyphonic four-part choral repertoire in the style of J. S. Bach. It is also an RNN-based model that allows the execution of user-defined constraints, such as rhythms, notes, parts, chords, and allegro. However, this direction remains challenging for the following reasons. From the outside, the overall musical structure seems to have no hierarchical features, and parts do not have a unified rhythmic pattern. Musical features are considered extremely simplified in terms of musical grammar, ignoring key musical features such as note timing, tempo, scale, and interval. Regarding the connotation of music, the music style is uncontrollable, the aesthetic measurement is invalid, and there is a significant gap between the sense of hearing and the music created by the musicians.

- *CNN-based music generation*. Some CNN architectures have been identified as an alternative to RNN architectures [21] [22]. The paper [21] is proposed as a representative work for CNN-based generative model building, which enables speech recognition, speech synthesis, and music generation tasks. WaveNet architecture presents a number of causal convolutional layers, somewhat similar to recurrent layers. However, it has two limitations: its inefficient computation reduces the use of real time, and it was created to be mainly oriented to acoustic data. MidiNet [22] architecture for symbolic data is inspired by WaveNet. It includes an adjustment mechanism that incorporates historical information (melody and chords) from previous measurements. The authors discuss two ways to control creativity and constrain the condition. One method is to insert adjustment data only in the middle convolutional layers of the generator architecture. The other method is to reduce the value of two control parameters of feature-matching regularization, thus reducing the distribution of actual and generated data.

- *Compound architecture-based music generation*. Bretan et al. [32] implemented the encoding of musical input by developing a deep autoencoder and reconstructed the input by selecting from a library. Subsequently, they established a deep-structured semantic model DSSM combined with LSTM to perform unitary prediction of monophonic melody. However, because of the limitations of unitary prediction, the quality of the generated content is sometimes poor. Bickerman et al. [24] proposed a music-coding scheme for learning jazz using deep belief networks. The model can generate flexible chords of different tones. It demonstrates that if the jazz corpus is large enough to generate chords, there is reason to believe that more complex jazz corpora can be performed. While some interesting pieces of jazz melodies have been created, the phrases generated by the model are not sufficient to represent all the features of the jazz corpus. Lyu et al. [11] combined the capability of LSTM in long-term data training and the advantages of the Restricted Boltzmann Machine (RBM) in high-dimensional data modeling. The results demonstrate that the model has a good generalization effect in the generation of chord music, but some high-quality music clips are rare. Chu et al. [12] proposed a hierarchical neural network model for generating pop music based on note elements. The lower layer

handles melody generation, and the upper layer produces chords and drums. Two practical applications of this model are related to neural dance and neural storytelling at the cognitive level. However, the shortcoming of this model also lies in the note-based generation mode, which does not include music theory research, thus limiting its musical creativity and stylistic integrity. Lattner et al. [25] learn the local structure of music by designing a C-RBM architecture that utilizes convolutions only in the temporal dimension in order to model time invariance, instead of pitch invariance, breaking the concept of pitch. Its core idea is to grammatically reduce the structure of music generation before music generation, such as music mode, rhythm pattern, etc. The disadvantage is that the musical structure is plagiarized. Huang et al. [26] proposed a transformer-based model for music generation. The core of the algorithm is to reduce the intermediate memory requirement to the length of the linear sequence. Finally, it's possible to generate a combination of tiny segment steps for a few minutes and use it in JBS Choir. Despite the experimental comparison of Maestro's two classic public music datasets, the qualitative evaluation was relatively crude.

3. Traditional Chinese Music Computing

To our knowledge, there are few available MIDI-based datasets of Chinese folk music. Luo et al. [33] proposed an algorithm to generate genre-specific Chinese folk songs based on auto-encoder. However, the results could only produce simpler fragments and did not perform qualitive analysis of the music from a musical genre perspective. Li et al. [34] presented a combined approach based on Conditional Random Field (CRF) and RBM for classifying traditional Chinese folk songs. Such an approach is noteworthy for being the first in-depth qualitative analysis of the classification results from a music-theoretical perspective. Zheng et al. [35] reconstructed the speed-update formula and proposed a Chinese folk music creation model based on the spatial particle swarm algorithm. Huang [36] collected data from two musical elements based on Chinese melody, and analyzed the application value of Chinese melody imagery in creating traditional Chinese folk music. Zhang et al. [37][38] conducted music data textualization and cluster analysis on Chinese traditional pentatonic groups. In summary, our motivation is to produce Chinese pentatonic music with hierarchical structure and local pentatonic music with multiple musical features and uniform rhythms, as shown in **Figure 1** [39].

Figure 1. The five main scales and four partial scales in traditional Chinese pentatonic music.

References

1. Cope, D. The Algorithmic Composer; AR Editions, Inc.: Middleton, WI, USA, 2000; Volume 16.

2. Nierhaus, G. Algorithmic Composition: Paradigms of Automated Music Generation; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009.

3. Herremans, D.; Chuan, C.H.; Chew, E. A functional taxonomy of music generation systems. ACM Comput. Surv. 2017, 50, 1–30.

4. Fernández, J.D.; Vico, F. AI methods in algorithmic composition: A comprehensive survey. J. Artif. Intell. Res. 2013, 48, 513–582.

5. Briot, J.P.; Pachet, F. Deep learning for music generation: Challenges and directions. Neural Comput. Appl. 2020, 32, 981–993.

6. Liu, C.H.; Ting, C.K. Computational intelligence in music composition: A survey. IEEE Trans. Emerg. Top. Comput. 2016, 1, 2–15.

7. Fiebrink, R.; Caramiaux, B. The machine learning algorithm as creative musical tool. arXiv 2016, arXiv:1611.00379.

8. Acampora, G.; Cadenas, J.M.; De Prisco, R.; Loia, V.; Munoz, E.; Zaccagnino, R. A hybrid computational intelligence approach for automatic music composition. In Proceedings of the 2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011), Taipei, Taiwan, 27–30 June 2011; pp. 202–209.

9. Kong, Q.; Li, B.; Chen, J.; Wang, Y. Giantmidi-piano: A large-scale midi dataset for classical piano music. arXiv 2020, arXiv:2010.07061.

10. Aiolli, F. A Preliminary Study on a Recommender System for the Million Songs Dataset Challenge. In Proceedings of the ECAI Workshop on Preference Learning: Problems and Application in AI, State College, PA, USA, 25–31 July 2013; pp. 73–83.

11. Lyu, Q.; Wu, Z.; Zhu, J.; Meng, H. Modelling high-dimensional sequences with lstm-rtrbm: Application to polyphonic music generation. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.

12. Chu, H.; Urtasun, R.; Fidler, S. Song from PI: A musically plausible network for pop music generation. arXiv 2016, arXiv:1611.03477.

13. Choi, K.; Fazekas, G.; Sandler, M. Text-based LSTM networks for automatic music composition. arXiv 2016, arXiv:1604.05358.

14. Johnson, D.D. Generating polyphonic music using tied parallel networks. In Proceedings of the International Conference on Evolutionary and Biologically Inspired Music and Art, Amsterdam, The Netherlands, 19–21 April 2017; pp. 128–143.

15. Lim, H.; Rhyu, S.; Lee, K. Chord generation from symbolic melody using BLSTM networks. arXiv 2017, arXiv:1712.01011.

16. Sun, F. DeepHear—Composing and Harmonizing Music with Neural Networks. 2017. Available online: https://fephsun.github.io/2015/09/01/neural-music.html (accessed on 1 September 2015).

17. Hadjeres, G.; Nielsen, F. Interactive music generation with positional constraints using anticipation-RNNs. arXiv 2017, arXiv:1709.06404.

18. Waite, E.; Eck, D.; Roberts, A.; Abolafia, D. Project Magenta: Generating Long-Term Structure in Songs and Stories. Available online: https://magenta.tensorflow.org/2016/07/15/lookback-rnn-attention-rnn (accessed on 15 July 2016).

19. Mehri, S.; Kumar, K.; Gulrajani, I.; Kumar, R.; Jain, S.; Sotelo, J.; Courville, A.; Bengio, Y. SampleRNN: An unconditional end-to-end neural audio generation model. arXiv 2016, arXiv:1612.07837.

20. Myburgh, J.C.; Mouton, C.; Davel, M.H. Tracking translation invariance in CNNs. In Proceedings of the Southern African Conference for Artificial Intelligence Research, Muldersdrift, South Africa, 22–26 February 2021; pp. 282–295.

21. Oord, A.v.d.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; Kavukcuoglu, K. Wavenet: A generative model for raw audio. arXiv 2016, arXiv:1609.03499.

22. Yang, L.C.; Chou, S.Y.; Yang, Y.H. MidiNet: A convolutional generative adversarial network for symbolic-domain music generation. arXiv 2017, arXiv:1703.10847.

23. Dong, H.W.; Hsiao, W.Y.; Yang, L.C.; Yang, Y.H. Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.

24. Bickerman, G.; Bosley, S.; Swire, P.; Keller, R.M. Learning to Create Jazz Melodies Using Deep Belief Nets. In Proceedings of the First International Conference on Computational Creativity, Lisbon, Portugal, 7–9 January 2010.

25. Lattner, S.; Grachten, M.; Widmer, G. Imposing higher-level structure in polyphonic music generation using convolutional restricted boltzmann machines and constraints. J. Creat. Music. Syst. 2018, 2, 1–31.

26. Huang, C.Z.A.; Vaswani, A.; Uszkoreit, J.; Shazeer, N.; Hawthorne, C.; Dai, A.; Hoffman, M.; Eck, D. Music transformer: Generating music with long-term structure. arXiv 2018, arXiv:1809.04281.

27. Roberts, A.; Engel, J.; Raffel, C.; Hawthorne, C.; Eck, D. A hierarchical latent vector model for learning long-term structure in music. In Proceedings of the International Conference on Machine Learning, PMLR, Stockholm, Sweden, 10–15 July 2018; pp. 4364–4373.

28. Roberts, A.; Engel, J.; Raffel, C.; Simon, I.; Hawthorne, C. MusicVAE: Creating a Palette for Musical Scores with Machine Learning. 2018. Available online: https://magenta.tensorflow.org/music-vae (accessed on 15 March 2010).

29. Chung, Y.A.; Wu, C.C.; Shen, C.H.; Lee, H.Y.; Lee, L.S. Audio word2vec: Unsupervised learning of audio segment representations using sequence-to-sequence autoencoder. arXiv 2016, arXiv:1603.00982.

30. Boulanger-Lewandowski, N.; Bengio, Y.; Vincent, P. Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription. arXiv 2012, arXiv:1206.6392.

31. Hadjeres, G.; Pachet, F.; Nielsen, F. Deepbach: A steerable model for bach chorales generation. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 1362–1371.

32. Bretan, M.; Weinberg, G.; Heck, L. A unit selection methodology for music generation using deep neural networks. arXiv 2016, arXiv:1612.03789.

33. Luo, J.; Yang, X.; Ji, S.; Li, J. MG-VAE: Deep Chinese folk songs generation with specific regional styles. In Proceedings of the 7th Conference on Sound and Music Technology (CSMT); Springer: Berlin/Heidelberg, Germany, 2020; pp. 93–106.

34. Li, J.; Luo, J.; Ding, J.; Zhao, X.; Yang, X. Regional classification of Chinese folk songs based on CRF model. Multimed. Tools Appl. 2019, 78, 11563–11584.

35. Zheng, X.; Wang, L.; Li, D.; Shen, L.; Gao, Y.; Guo, W.; Wang, Y. Algorithm composition of Chinese folk music based on swarm intelligence. Int. J. Comput. Sci. Math. 2017, 8, 437–446.

36. Kuo-Huang, H. Folk songs of the Han Chinese: Characteristics and classifications. Asian Music 1989, 20, 107–128.

37. Liumei, Z.; Fanzhi, J.; Jiao, L.; Gang, M.; Tianshi, L. K-means clustering analysis of Chinese traditional folk music based on midi music textualization. In Proceedings of the 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 9–11 April 2021; pp. 1062–1066.

38. Zhang, L.M.; Jiang, F.Z. Visualizing Symbolic Music via Textualization: An Empirical Study on Chinese Traditional Folk Music. In Proceedings of the International Conference on Mobile Multimedia Communications, Guiyang, China, 23–25 July 2021; Springer: Cham, Switzerland; pp. 647–662.

39. Xiaofeng, C. The Law of Five Degrees and pentatonic scale. Today's Sci. Court. 2006, 5.