Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/68/20241432

Analyzing principles and applications of machine learning in music: Emotion music generation, and style modeling

Jadon T. Lam

Pleasant Grove High School, Elk Grove CA 95624, United States

The corresponding author's e-mail address: katelynlam123@berkeley.edu

Abstract. The combination of machine learning with music composition and production is proving viable for innovative applications, enabling the creation of novel musical experiences that were once the exclusive domain of human composers. This paper explores the transformative role of machine learning in music, particularly focusing on emotion music generation and style modeling. Through the development and application of models including DNNs, GANs, and Autoencoders, this study delves into how machine learning is being harnessed to not only generate music that embodies specific emotional contexts but also to transfer distinct musical styles onto new compositions. This research discusses the principles of these models, their operational mechanisms, and evaluates their effectiveness through various metrics such as accuracy, precision, and creative authenticity. The outcomes illustrate that these technologies not only enhance the creative possibilities in music but also democratize music production, making it more accessible to non-experts. The implications of these advancements suggest a significant shift in the music industry, where machine learning could become a central component of creative processes. These results pave a path to the understanding of the potential and limitations of machine learning in music and forecasts future trends in this evolving landscape.

Keywords: Machine learning, emotion music generation, music style transfer.

1. Introduction

Music and technology have a rich and evolving history that extends over several decades. This history is marked by the continuous advancement of computational methods used in crafting and analyzing music. One of the first milestones in this journey was the creation of the first computer music in 1951. As the mid-20th century approached, the field of computer music began to take shape, transitioning from basic sound synthesis to the complex algorithms that drive today's music generation and classification systems. The development of digital computers in the 1950s and 1960s provided the initial platform for computer music, with pioneers like Max Mathews and John Chowning [1]. In the following years, the field expanded to include various approaches such as granular synthesis, physical modeling, and computer-assisted composition. Each of these techniques contributed to the rich and varied landscape of computer music, opening up new possibilities for music creation and research [2]. The 1970s and 1980s saw further developments with the introduction of digital synthesizers and samplers, which allowed for greater flexibility and creativity in music production. The advent of MIDI revolutionized the way electronic musical instruments could communicate with each other and with computers, leading to new forms of composition and performance. In the recent decades, the rise of the

DOI: 10.54254/2755-2721/68/20241432

internet and advancements in artificial intelligence and machine learning have further transformed the field of computer music. Online platforms and software have democratized music production, allowing anyone with a computer to create and share music. Meanwhile, AI-driven systems are now capable of composing music autonomously, analyzing and generating music in ways that were once the exclusive domain of human musicians.

Contemporarily, the developments in machine learning had a profound impact on the field of music composition, revolutionizing the way music is created. Machine learning algorithms have the capability to analyze vast datasets and make complex decisions, and have been increasingly applied to mimic human-like composition techniques [3]. Deep learning, a specialized branch of machine learning, has been instrumental in this development. RNNs and GANs have been harnessed to create music that encapsulates the nuances of musical structure and style [4-8]. MuseNet, developed by OpenAI, is one of the well-known products with a variety of instruments and styles, from country to classical to pop. It is trained from a vast dataset of MIDI files on patterns of harmony, rhythm, and style without being explicitly programmed with musical knowledge [5]. The Magenta Project, an open-source initiative by Google, applies machine learning to artistic expression, including music. It focuses on creating tools that assist in blending music loops, generating new sounds, and even composing entirely new pieces using machine learning [6]. These advancements demonstrate the potential of machine learning to revolutionize4 music composition, offering new possibilities for creativity and innovation in the field.

The motivation behind the study stems from the escalating interest in the application of machine learning for music generation and classification, as well as the potential benefits it offers for a wide range of applications within the music industry. By delving into the principles of machine learning within the contexts of emotion music generation, style modeling, and music genre classification, this study will provide significantly to the understanding of how these technologies can be effectively utilized to augment the creative process and enhance the accessibility of music for a broader audience.

2. Descriptions of machine learning in music

In the realm of music, machine learning models have been applied in various capacities, revolutionizing the way music is composed, analyzed, and experienced. One of the key applications is in music recommendation systems, where models like collaborative filtering and content-based filtering are employed to personalize music suggestions [9, 10]. Another significant application is in automatic music generation, where models such as LSTM networks and Variational Autoencoders (VAEs) are used to create novel musical compositions by learning from a vast corpus of existing music [3].

Machine learning models have also been applied in music classification tasks, where models like CNNs and SVMs are utilized to categorize music into genres, moods, or other attributes based on audio features extracted from the music [11-18]. Furthermore, models such as Deep Belief Networks (DBNs) and Recurrent Neural Networks (RNNs) have been employed in music transcription, where the goal is to convert audio signals into musical notation [19-24]. Evaluations of these models typically involve metrics, and measures like BLEU score or ROUGE score for generative tasks. Additionally, subjective evaluations are often conducted to assess the perceptual quality and creativity of the generated music.

3. Emotion-based music generation

The field of music generation has seen significant advancements with the integration of machine learning techniques. One fascinating area of application is emotion-based music generation, where the goal is to create music that can convey specific emotions to the listener. This application is particularly interesting because it combines the creative aspect of music composition with the analytical power of machine learning. The principle behind emotion-based music generation involves training machine learning models on a dataset of music pieces that are annotated with emotional labels. These labels could represent various emotions such as happiness, sadness, anger, or calmness. The models learn to recognize the patterns and features in the music that are associated with these emotions and use this knowledge to generate new pieces of music that evoke the desired emotions. Music generation is grouped into two types: symbol domain generations (MIDIs or piano sheets [8] and audio domain

DOI: 10.54254/2755-2721/68/20241432

generation sound waves [23]). There are various models for describing emotion and they can be mainly divided into four categories: discrete, dimensional, music-specific models, and miscellaneous [24].

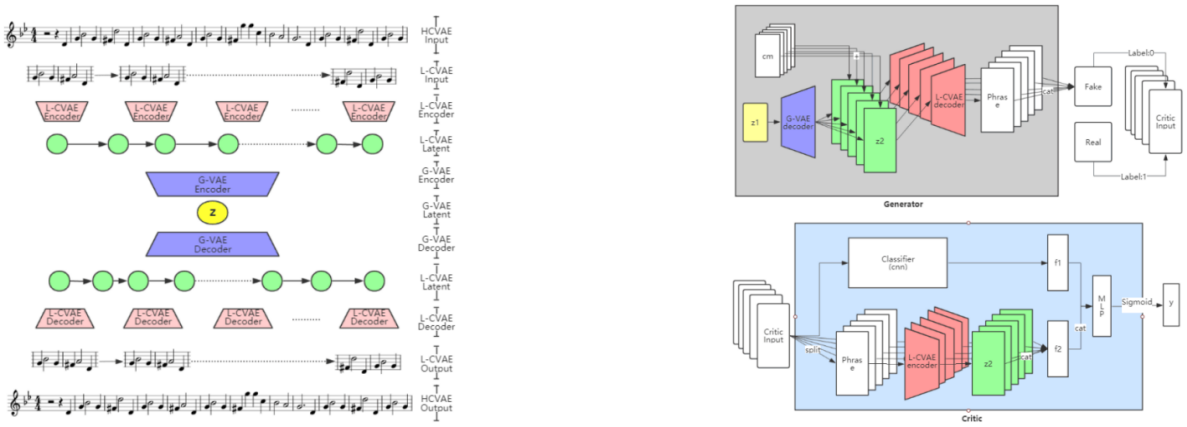

Figure 1. Proposed system flow chart [25].

Table 1. Scoring statistic for the generated music

Emotion: Calm, relax, and at ease		Emotion A	Emotion B	Emotion C	Emotion D
Generated Music 1	Average	2.525	1.525	2.275	4.100
	Std	1.240	0.784	1.062	0.900
	Variation	1.538	0.615	1.128	0.810
	Mode	3	1	3	4
	Median	3	1	2	4
Generated Music 2	Average	2.400	1.700	2.500	4.225
	Std	1.317	0.966	1.177	0.768
	Variation	1.733	0.933	1.385	0.589
	Mode	1	1	3	4
	Median	2	1	2	4

Conditional Variational Autoencoders (CVAEs): CVAEs are a type of generative model that can generate music conditioned on a specific emotion. They learn a latent representation of the input music and can generate new samples by sampling from this latent space while conditioning on the desired emotion. This study examines the CVAE-GAN architecture as its primary framework, as depicted in Fig. 1 [25]. The architecture integrates an encoder and decoder in a sequential manner using the Seq2Seq model, while the other components, including additional decoders, discriminators, and classifiers, follow the conventional CGAN approach. Each element of the model is constructed using multilayer

DOI: 10.54254/2755-2721/68/20241432

GRU networks. Initially, when music data is input into the model, it is first represented as a One-hot Vector. This data is then embedded, reducing its dimensionality which not only conserves computational resources but also minimizes the inefficiencies associated with sparse representations typical of One-hot Vectors. According to the ADAM optimization algorithm, which is designed for the first-order gradient-based optimization of stochastic objectives with adaptive estimates of lower-order moments, the embedding process effectively reduces the dimensionality of the One-hot Vector data from 99 to 24 dimensions. In this reduced format, 8 dimensions represent the pitch and 16 dimensions cover the pitch length. In the coding of the model, the input data structure is defined by the shape (number of songs, maximum number of notes per song, number of pitches). For instance, a shape of (4, 6, 8) indicates there are four input instances, each with 6 notes described by 8-dimensional pitch vectors. Once encoded, the data undergoes a transformation through a tiling function, which adapts the emotional attributes associated with the music- referred to as 'Attribute' in the experimental code. The attribute is expanded to match each note's length, and the Concat function merges the emotion attributes with the musical data. Table 1 shows the scoring statistics for the generated music pieces 1 and 2.

Emotion-based music generation using machine learning models like CVAEs, GANs, and RNNs presents a novel approach to music composition. By leveraging the ability of these models to learn from emotional annotations in music, composers and researchers can create music that not only sounds pleasing but also effectively communicates specific emotions.

4. Music style transfer

Music style transfer is an advanced application of machine learning that involves transforming a piece of music to adopt the style of another, while preserving its original content. This technique has gained significant interest due to its potential to innovate in the field of music production and personalization. Music style transfer allows artists and enthusiasts to experiment with different musical genres and styles, thus broadening the creative horizons of musical expression. The principle behind music style transfer involves the decomposition of music into content and style components. The model then learns to apply the stylistic elements of one piece to the content of another. This process requires understanding of both musical structure and the characteristics that define different musical styles.

Deep Neural Networks are pivotal in the style transfer due to their ability to analyze and extract complex features from music tracks. These networks delve into the audio files, identifying and isolating distinct elements that represent the 'content' and 'style' of the pieces. The content typically pertains to the melody and basic structure, while the style encompasses aspects like rhythm, instrument timbre, and articulation. The separation of these components is crucial for effective style transfer, allowing for the precise application of one music piece's style onto the content of another, maintaining the original's identity while adopting a new auditory aesthetic [15, 17].

In music style transfer, GANs have been innovatively adapted to handle two primary functions — generation and discrimination. The generator creates music that aims to emulate a particular style, learning from various examples of that style. Meanwhile, the discriminator evaluates whether the generated music convincingly resembles the target style or if it can be distinguished from authentic samples of that style. This dynamic training process helps refine the generator's output, enhancing its ability to produce musically coherent and stylistically accurate pieces [25, 26].

Autoencoders, especially Variational Autoencoders (VAEs), are extensively used to encode music into a compact latent space where each point captures meaningful musical attributes. When employed in style transfer, VAEs can manipulate these encoded representations to modify a piece's style while retaining its original content. This involves adjusting aspects of the latent space that correspond to stylistic features, thereby allowing for the synthesis of music that merges the original composition's content with the stylistic elements of another music piece. This model's ability to handle continuous changes in the music space makes it particularly useful for creating seamless transitions and variations in style [16, 24].

The application of these models in music style transfer has shown promising results. Studies have demonstrated that DNNs effectively preserve the original musical content while applying new stylistic

DOI: 10.54254/2755-2721/68/20241432

features. GANs have shown proficiency in generating music that closely resembles target styles, making them highly valuable for tasks requiring high fidelity in style reproduction. Autoencoders, particularly VAEs, have provided a balanced approach by offering good style transfer capabilities along with easier control over the degree of style application. Subjective evaluations, involving human listeners, have generally reported high satisfaction with the transformed music, noting that the transferred styles are perceptible while the original music's integrity is maintained. Objective assessments, such as style accuracy measurements and content preservation metrics, have further validated the effectiveness of these models in achieving convincing music style transfers. Music style transfer using machine learning represents a significant advancement in the intersection of artificial intelligence and creative arts. By leveraging models like DNNs, GANs, and Autoencoders, this technology not only enhances the creative capabilities of musicians and producers but also opens new possibilities for personalized music experiences. As this field evolves, further improvements in model performance and user interfaces are expected to make music style transfer more accessible and widely used in various musical applications.

5. Limitations and prospects

Despite the advancements in machine learning for music composition, plenty of severe limitations are necessary to be addressed and overcome. One of the main challenges is the computational complexity of the models, which needs a large amounts of training and inference ability and capacity. As a matter of fact, it is also difficult to achieve high-quality results for less-represented musical styles or instruments. Future prospects for machine learning in music composition include the development of more efficient and scalable models, as well as the exploration of new applications including music emotion recognition, interactive music systems, and the integration of machine learning with music therapy. Additionally, the continued exploration of interdisciplinary approaches, combining insights from music theory, cognitive science, and machine learning, holds promise for advancing the understanding of music composition and perception.

6. Conclusion

In summary, machine learning has significantly impacted the field of music, offering innovative approaches to music generation, style modeling, and genre classification. The advancements in machine learning models have enabled the creation of music that mimics human composition techniques, leading to new possibilities for creativity and innovation. While there are still challenges to be addressed, the future of machine learning in music composition holds exciting prospects for further exploration and development. These analysis and evaluations of this research extend to various applications in the music industry, enhancing the creative process and accessibility of music.

References

[1] Sze Roads C 1996 The Computer Music Tutorial MIT Press.
[2] Miranda E R 2001 Composing Music with Computers Focal Press.
[3] Briot J P, Hadjeres G and Pachet F 2017 Deep learning techniques for music generation - A survey arXiv preprint arXiv:1709.01620.
[4] Huang A and Wu R 2016 Deep learning for music. arXiv preprint arXiv:1606.04930.
[5] OpenAI 2019 MuseNet Retrieved from: https://openai.com/blog/musenet/
[6] Flynn S 2020 The Magenta Project. ReHack
[7] Choi K, Fazekas G, Sandler M and Cho K 2016 A Tutorial on Deep Learning for Music Information Retrieval. arXiv preprint arXiv:1709.04396.
[8] Dong, H. W., Hsiao, W. Y., Yang, L. C., & Yang, Y. H. (2018, April). Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 32, No. 1).
[9] Ferreira, L. N., Mou, L., Whitehead, J., & Lelis, L. H. (2022, October). Controlling perceived emotion in symbolic music generation with monte carlo tree search. In Proceedings of the

DOI: 10.54254/2755-2721/68/20241432

AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (Vol. 18, No. 1, pp. 163-170).

[10] Hawthorne, C., Elsen, E., Song, J., Roberts, A., Simon, I., Raffel, C., ... & Eck, D. (2018). ĒOnsets and frames: Dual-objective piano transcription, ē in. In Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018, Paris, France.

[11] Miranda, E. R. (2001, September). Evolving cellular automata music: From sound synthesis to composition. In Proceedings of 2001 Workshop on Artificial Life Models for Musical Applications.

[12] Orio, N. (2006). Music retrieval: A tutorial and review. Foundations and Trends® in Information Retrieval, 1(1), 1-90.

[13] Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. IEEE Transactions on speech and audio processing, 10(5), 293-302.Transactions on speech and audio processing, 10(5), 293-302.

[14] Aggarwal, S., Selvakanmani, S., Pant, B., Kaur, K., Verma, A., & Binegde, G. N. (2022). Audio segmentation techniques and applications based on deep learning. Scientific Programming, 2022.

[15] Briot, J. P., Hadjeres, G., & Pachet, F. D. (2020). Deep learning techniques for music generation (Vol. 1). Heidelberg: Springer.

[16] Li, S., Zheng, Z., Dai, W., Zou, J., & Xiong, H. (2020, May). Rev-ae: A learned frame set for image reconstruction. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1823-1827). IEEE.

[17] Zheng, K., Meng, R., Zheng, C., Li, X., Sang, J., Cai, J., & Wang, J. (2021). EmotionBox: a music-element-driven emotional music generation system using Recurrent Neural Network. arXiv preprint arXiv:2112.08561.

[18] Yu, X., Ma, N., Zheng, L., Wang, L., & Wang, K. (2023). Developments and applications of artificial intelligence in music education. Technologies, 11(2), 42.

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

[20] Huang, C. F., & Huang, C. Y. (2020, October). Emotion-based AI music generation system with CVAE-GAN. In 2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE) (pp. 220-222). IEEE.

[21] Subramani, K., Rao, P., & D'Hooge, A. (2020, May). Vapar synth-a variational parametric model for audio synthesis. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 796-800). IEEE.

[22] Vuoskoski, J. K., & Eerola, T. (2017). The pleasure evoked by sad music is mediated by feelings of being moved. Frontiers in psychology, 8, 245046.

[23] Huang, Y. S., & Yang, Y. H. (2020, October). Pop music transformer: Beat-based modeling and generation of expressive pop piano compositions. In Proceedings of the 28th ACM international conference on multimedia (pp. 1180-1188).

[24] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2015). Advances in Neural Information Processing Systems, Vol. by F. Pereira, Curran Associates, Red Hook, NY, 2672.

[25] Liang X, Wu J and Yin Y 2019 MIDI-sandwich: multi-model multi-task hierarchical conditional VAE-GAN networks for symbolic single-track music generation. arXiv preprint arXiv:1907.01607.

[26] Engel, J., Resnick, C., Roberts, A., Dieleman, S., Norouzi, M., Eck, D., & Simonyan, K. (2017, July). Neural audio synthesis of musical notes with wavenet autoencoders. In International Conference on Machine Learning (pp. 1068-1077). PMLR.