Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/48/20241494

Composing music based on machine learning

Jiarui Zhang

Department of Computer Science, Qingdao University, Qingdao 308, China

2021204462@qdu.edu.cn

Abstract. As a matter of fact, the music is an important tool and pathway for human emotional communication, with great significance, especially in recent years. However, it should be noted that traditional music production methods require a lot of time, manpower, and funds, and have high costs. With the rapid development of machine learning technology as well as computation ability, the computer automatic composition based on machine learning technology is receiving more and more attention. With this in mind, this study introduces and discusses the automatic emotion recognition and emotion based music generation based on this advanced technology, as well as describes the process involved. According to the analysis, the advanced applications and implementations results have been demonstrated. At the same time, the current limitations as well as the future prospects will be clarified and discussed. Overall, these results shed light on guiding further exploration of machine learning implementations of music composition.

Keywords: Music emotion recognition, automated composition, machine learning.

1. Introduction

Long before the emergence of computer technology, there was a method of generating music through random algorithms. Mozart used dice to randomly generate music. Use the number of dice to randomly connect different styles of music clips to form a complete piece [1]. In the 1950s, after the birth of the first computer, the computer used algorithms to create the first music sheet The Illiac Suiteni and used abstract models [2]. However, this abstract model has serious drawbacks, as its syntax and specifications are complex and prone to errors.

In recent years, products and technologies that support perceptual interaction capabilities have received increasing attention, which has promoted the development of emotional computing. Emotional computing has entered its era, becoming more popular than ever before. And music is a very important and indispensable part of spreading emotions, and musical emotions are naturally a very important part of emotional calculation. Currently, more and more researchers are focusing on the study of music emotions. Over the past decade, people have done a lot of work and exploration in this field.

The soul of musical works is emotions, and all musical works have emotions and thoughts. So, these emotional features can be utilized to achieve music search and classification. Therefore, the technology of retrieving and generating music through emotional features can be achieved through automatic analysis and recognition of music emotional information. In recent years, the research on extracting physical and lyrical content features of music to automatically recognize emotional information in music has become increasingly important and urgent. MER(Music Emotion Recognition) is a field of research on computational models for identifying song emotions. MER first needs to extract the features of music

DOI: 10.54254/2755-2721/48/20241494

and perform some column analysis, and then establish a mapping relationship between music features and emotional space. Through these operations, it is possible to recognize the emotions expressed by music.

At present, MER has broad application prospects in the fields of music recommendation, music therapy, and music search. In 2007, MEC first appeared as a task in MIREX, reflecting the recognition of the importance of MER. At the same time, more and more music software and music social networking sites are also beginning to apply music emotions. The emotions of music on these websites and software can serve as a tag to recommend music and increase interaction with users. The application of MER is also favored by many music player manufacturers. MER can efficiently recognize the emotions of music to enhance human-computer interaction.

In the second part of this article, basic descriptions of emotion recognition of music and the classification of music emotions are introduced, and the problem of how to recognize music emotions is explained. In the third part of this article, it introduces Machine learning for emotion recognition and how to recognize music emotions based on machine learning, as well as its recent applications. It provides a detailed explanation of the process of MER. In the fourth part of this article, machine learning for music generation of certain emotions and its applications in recent years are introduced. Finally, in the fifth part, the limitations and future prospects of emotion analysis based on machine learning are explained as well as concludes the remark.

2. Basic Descriptions of Emotion recognition of music

Music has the ability to induce or convey emotions, which makes it play a very important role in human life. Music emotion recognition has recently gained a lot of attention from academia and business due to its major contributions to areas including music visualization, automatic music production, psychotherapy, and recommendation systems. Deep learning-based music emotion recognition is steadily growing in importance as a study area due to the rapid growth of artificial intelligence [3]. Since the 1930s, people have conducted a series of studies on music and the emotions it expresses. In the early 21st century, more and more research focused on how to automatically extract music emotions from music feature data. Music Information Retrieval and Evaluation Exchange (MIREX) is an international audio retrieval and evaluation competition. In 2007, the Audio Emotion (AMC) classification was added to MIREX, reflecting the extraordinary significance and importance of music emotion recognition [4].

Many characteristics of music can affect emotions, such as harmony, timbre, interpretation, and lyrics. And the emotions of the music are also variable. One wants to infer the emotional types of music based on these characteristics, but in this process, it often encounters a very tricky problem. The understanding and perception of music emotions often have strong subjectivity, which leads to certain differences in the determination of music emotions, making it difficult to obtain a clear answer.

Determining the emotional content of music audio through computation is essentially an interdisciplinary effort that not only spans signal processing and machine learning, but also requires an understanding of auditory perception, psychology, and music theory. Music emotion recognition (MER) is a field of study that looks at computational models for interpreting the emotions included in songs [5]. MER can be described as a set of processes. Firstly, computers were used to extract music features from a piece of music to be recognized, and the obtained music feature information is mapped to Emotion Space by an effective computational model. Finally, the emotion of the music is recognized. The audio signal, symbolic music notation, lyric texts, and even biological data like the electroencephalogram (EEG) may all be used to extract musical feature. A finite number of discrete categories or an infinite number of points in a continuous multidimensional space can be used to depict the emotions space [3].

3. Machine Learning for Emotion Recognition

The domain definition, the feature extraction, and the model training process make up a typical MER method's three steps [6]. Firstly, select music records and emotion models in different formats as much as possible, and secondly extract music features and ground truth data which is two common kinds of data features. Finally, one establishes a mapping between A and B using machine learning methods to

DOI: 10.54254/2755-2721/48/20241494

complete model training. To ensure a good mapping between music features and emotional space, it is important to correctly select data features. There are two types of music record, audio and symbol files, which can extract high-level and low-level music features, respectively [7]. Ground truth data is divided into two types: label- type and numerical-type, and the selection of Ground truth data directly affects the selection of machine learning methods. After extensive research and summary, emotional models can be divided into classification models and dimension models. Usually, one uses classification models with label- type ground truth data more frequently. In short, the choice of data features was very important.

The purpose of extracting data features is to reduce the information in music, so that symbols describing music can be obtained. And before extracting music features, one needs to first transform the format of the extracted music. Then the music is segmented. These music clips should be as short as possible, so that the distribution of emotions on these clips can be as uniform as possible [8]. However, it should not be too short, as it can lead to music clips that cannot accurately express emotional content.

In the data extraction step, music feature and ground truth data are important components of data features, so the following will describe these two types of data features. The characteristics of music can affect emotional expression, for example, slow paced and soothing music often expresses a comfortable and relaxed mood, while fast-paced music is the opposite. Intensity is a very important feature that can generally be obtained in one frame. Timbre is also a very basic feature that can be represented by DWCHs [9]. Rhythm is a common feature that can describe some repetitive patterns in music. The most noteworthy aspect of rhythm is his strength, regularity, and tempo [10], which are closely related to emotional expression. Finding suitable features is a huge challenge, and performance indicators are not optimistic when the feature dimension is high. Thus, one needs to find a suitable method. The KL method used by Lu et al. can remove some correlated features from the original features and achieve the goal of reducing computational complexity by extracting irrelevant features [11]. There is another technique, such as principal component analysis

The relationship between music and ground truth data has been studied for a long time. Hevner proposed a model to describe music emotions, and Model C summarized 67 adjectives about the emotions expressed by music, which were divided into 8 clusters. This model can explore the relationship between multiple features, including pitch and rhythm, and musical emotions [12]. There are three methods for collecting ground truth data: Choose from a choice of adjectives, indicate the volume of musical characteristics, or name dimensional models explicitly [3].

There are four types of methods for implementing MEC systems using machine learning: Song level categorical MER, Song level dimensional MER, categorical MEVD, and dimensional MEVD. Song level categorical MER commonly used models include: the support vector machines (SVM), k-nearest neighbors (KNN), decision tree (DT), random forests (RF), and naive Bayes (NB). Laurier et al. used a fusion of SVM and KNN methods to obtain more ideal experimental results [13]. The fusion of methods in MEC can obtain more accurate prediction results. The commonly used models in Song level dimensional MER are mainly regression models, such as linear regression and multivariate linear regression (MLR). It is worth mentioning that there is a model that is more accurate in predicting music emotional outcomes than models such as SVR and MLR. It was proposed by A et al. in 2012 and is called the Acoustic Emotion Gaussian Model [14]. The most typical machine learning model in MEVD is generally based on SVM. Sometimes, Gaussian mixture models are also used. In addition, research has found that commonly used regression models can also be used in MEVD [15]. Xianyu et al. proposed a method of combining two SVR models, where one SVR model is used to compare emotional changes in different songs and the other SVR model is used to explore the emotional changes of a song over time [16].

4. Machine Learning for Music Generation of Certain Emotion

Music, as an important artistic form and means of emotional communication, is widely needed by human society. However, traditional music production methods are time-consuming and costly. In recent years,

DOI: 10.54254/2755-2721/48/20241494

with the rapid development of machine learning and artificial intelligence technology, automatic composition based on machine learning has received increasing attention.

In 2007, Radford et al. proposed an LSTM neural network model that can generate a paragraph of text that matches the emotion based on input emotional data in the form of text [17]. When combined with logistic regression, the emotion analysis function will be more accurate. This model can also be used to generate music with specified emotions. Lucas N. Ferreira et al. also proposed a music dataset represented by emotionally annotated symbols to evaluate this method [18]. There are also some other similar jobs. Williams et al. developed a model for generating background music for game scenes [19]. Davis and Mohammad proposed a system for creating music for novel plots [20]. They first extract emotions from game scenes or novel plots, and then generate melodies with these emotions. Radford et al. created a single-layer mLSTM with over 4000 units to process text content. The model updates the state of each byte and predicts the next byte [17]. Radford et al. trained a logistic regression trainer using Amazon's comment content dataset. After examining the performance of features on different datasets, they found that a unit can be associated with emotions. So, it is possible to artificially set the value of emotions and generate comments corresponding to emotions [17]. Based on this work, music generation that is easy to give emotions can be achieved. Music composition can be likened to language modeling, so that one can use mLSTM and logical regression to compose together. Music expresses emotions through certain features such as timbre, melody, tone, etc., so these features can be encoded. Lucas N. Ferreira et al. encoded the pitch, duration, velocity, velocity, and tempo of music to represent its composition and characteristics [18]. In order to use this method to create music, they used two MIDI files to train LSTM and logistic regression, and created a new dataset labeled based on emotions. The source of this dataset is game music in MIDI format. They used a highly flexible model called the Valence routing model, which supports effective annotation of music. And the same set of data supports both classification and regression models. Therefore, the valence-arousal model is a very commonly used model [17]. A typical results are shown in Fig. 1.

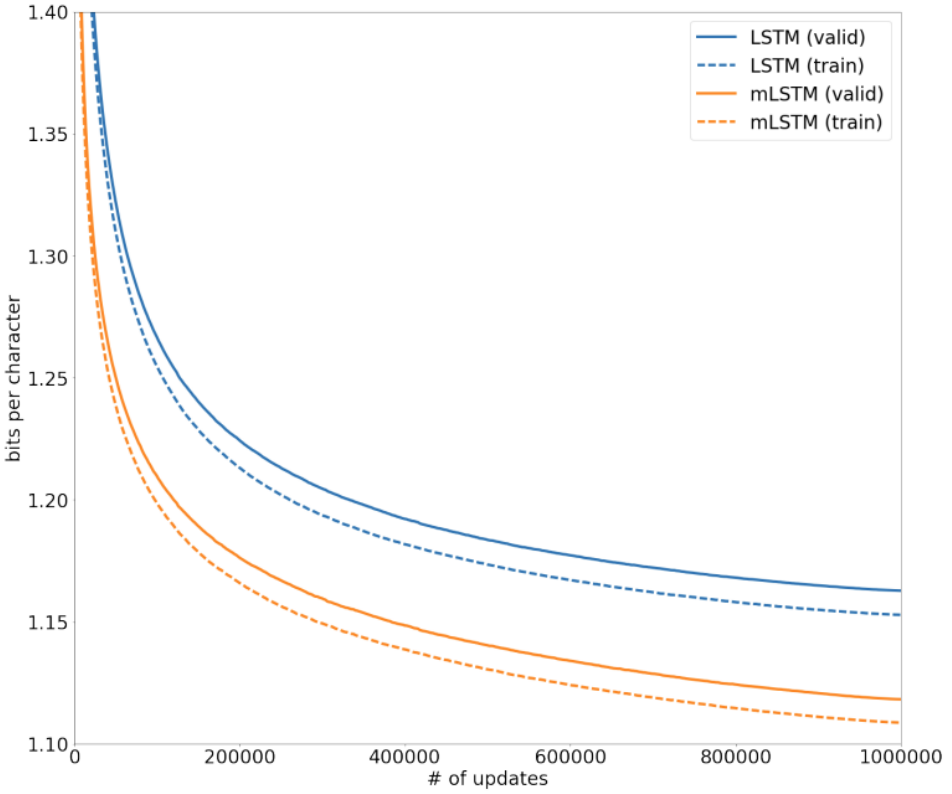

Figure 1. Bits per character as a function of update.

DOI: 10.54254/2755-2721/48/20241494

Then, the next step is annotation and data collection. Firstly, recruit a group of volunteers, inform them of the annotation tasks to be done and teach them how to use annotation software, and explain the purpose of the tasks to them. Then play a few pieces of music and have the subjects describe the music clips they hear. After they complete the calibration and annotation tasks, collect their information. Collect the information obtained from the annotation task just now. Through the collection process, one can provide valuable values for each fragment. Nevertheless, only a portion of the information is needed, so it needs to be preprocessed to form a sequence about time and divided into three clusters. This can eliminate the negative effects of the generated noise and achieve better results. Firstly, the noise cluster needs to be discarded, and then the average of the time series in other clusters needs to be calculated. Negative numbers are considered negative emotions, while positive values represent positive emotions. Through these steps, some words can be generated, including words with positive emotions and words with negative emotions.

5. Conclusion

To sum up, the field of music creation and generation based on machine learning is very popular. It is based on artificial neural networks to generate music. With the progress and development of machine learning, the field of music generation is also facing more and more opportunities and challenges. There are still some limitations and the technology is still not very mature. For example, controlling the generated music (controlling tonal consistency, controlling the maximum number of repeated syllables), structural issues in music, interactivity in music, and innovation in music (issues of plagiarism and imitation) are all challenges faced. In addition, the automation of music creation makes it more convenient to create, and the personality of music is becoming increasingly diverse. At the same time, the personalization of music content can blur the relationship between music creation and music consumption, which is also a challenge. In the future, the technology of utilizing machine learning for music creation will become increasingly mature and face more challenges. One should promptly study solutions to address challenges and adapt to the development of technology.

References

[1] Briot J P 2021,From artificial neuralnetworks to deep learning for music generation: history, concepts and trends, Neural Computing and Applications vol 33(1) pp 39-65.
[2] Music E 1959 Composition with an electronic computer (New York: McGraw-Hill).
[3] Han D, Kong Y, Han J, et al. 2022 ,A survey of music emotion recognition,Frontiers of Computer Science vol 16(6) p 166335.
[4] Kim Y E, Schmidt E M, Migneco R, et al. 2010 ,Music emotion recognition: A state of the art review,Proc. ismir. vol 86 pp 937-952.
[5] Yang X, Dong Y and Li J 2018 ,Review of data features-based music emotion recognition methods,Multimedia systems vol 24 pp 365-389.
[6] Bartoszewski M, Kwasnicka H, Markowska-Kaczmar U, et al. 2008,Extraction of emotional content from music data, 7th Computer Information Systems and Industrial Management Applications pp 293-299.
[7] McKay C 2004 ,Automatic genre classification of MIDI recordings (Berling: Springer).
[8] MacDorman S O and Chin-Chang Ho K F 2007 Journal,Automatic emotion prediction of song excerpts: Index construction, algorithm design, and empirical comparison, New Music Research vol 36(4) p 281-299.
[9] Chin Y H, Lin P C, Tai T C, et al. 2015,Genre based emotion annotation for music in noisy environment,International Conference on Affective Computing and Intelligent Interaction (ACII) pp 863-866.
[10] Liu D, Lu L and Zhang H J 2003,Automatic mood detection from acoustic music data, Journal of Hunan University p 10.
[11] Lu L, Liu D and Zhang H J 2005,Automatic mood detection and tracking of music audio signals, IEEE Transactions on audio, speech, and language processing vol 14(1) pp 5-18.

DOI: 10.54254/2755-2721/48/20241494

[12] Abrol V, Abtahi A, Agrawal P, et al. 2020 ,Automatic mood detection and tracking of music audio signals,IEEE/ACM Transactions on Audio, Speech, and Language Processing vol 28.
[13] Laurier C, Grivolla J and Herrera P 2008,Multimodal music mood classification using audio and lyrics,seventh international conference on machine learning and applications pp 688-693.
[14] Wang J C, Yang Y H, Wang H M, et al. 2012 Proceedings of the 20th ACM international conference on Multimedia pp 89-98.
[15] Yang Y H, Lin Y C, Su Y F, et al. 2008 ,The acoustic emotion Gaussians model for emotion-based music annotation and retrieval,IEEE Transactions on audio, speech, and language processing vol 16(2) pp 448-457.
[16] Xianyu H, Li X, Chen W, et al. 2016,SVR based double-scale regression for dynamic emotion prediction in music, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) pp 549-553.
[17] Radford A, Jozefowicz R and Sutskever I 2017,Learning to generate reviews and discovering sentiment, arXiv preprint arXiv:1704.01444.
[18] Ferreira L N and Whitehead J 2021,Learning to generate music with sentiment,arXiv preprint arXiv:2103.06125, 2021.
[19] Williams D, Kirke A, Eaton J, et al. 2015 ,Dynamic game soundtrack generation in response to a continuously varying emotional trajectory,Audio engineering society conference: 56th international conference: Audio for games p 15.
[20] Davis H and Mohammad S M 2014 ,Generating music from literature,arXiv preprint arXiv:1403.2124.