Emotion Recognition of Music based on Machine Learning Scenarios

Zenan Li *

Beijing No.4 High School International Campus, Beijing, China

* Corresponding author email: lfm2000@cau.edu.cn

Abstract. As the rapid development of computer music, the technique for recognizing the emotion of music also have many progresses. After the brief introduction of the history of computer music, this paper mainly discusses about the current existing machine learning models for the emotion recognition in music. The complexity of emotion has been emphasized in this paper for several reasons. In addition, by comparison different models, this paper summarizes common features, metrics and steps used in music emotion analyzation. Moreover, this study finds out the limitations and disadvantages of different classifications and feature extracting method for different models, pointing out the living problems, e.g., the difficulty of emotion recognition for experimental music. To sum up, this paper summarizes and analyzes the primary studying in the field of music emotion recognition, offering a guideline for implementations of different machine learning approaches in the field. These results shed light on paving a path for further exploration of emotion recognition in computer music.

Keywords: Emotion Recognition; Computer Music; Machine Learning Scenarios.

1. Introduction

It is a commonly accepted fact that the first computer-generated sound was heard in 1957, in Bell labs, America. The man who made this sound and wrote the very first music program was Max Mathews, who firstly developed a program called "MUSIC" which could only make single notes, and the second program "MUSIC II", which appeared one year after "MUSIC", was able to create four-part polyphony. The following "MUSIC III" introduced an important concept in the whole computer music field, which was a "unit generator". The "MUSIC" program was still developing by many other computer musicians in order to fit their own preferences. In 1981, MIDI devices had been wildly used, which laid the foundation of modern computer music [1]. As the inter discipline of computer science and music, computer music has various possibility and blank space waiting for exploring and researching. This paper will mainly discuss about latest research on the emotion recognition of music based on machine learning scenarios. Emotional analyzation among music has a significant meaning for deeper understanding of the essence of music and music theory, which converts those abstract ideas music into some measurable indexes. In this case, it allows musicians to create music in a more efficient way by following rules summarized during the process of emotional analyzation.

Emotional analyzation is a process that needs to collect and analyze countless data and graphs. In order to stimulate the efficiency, researchers develop machine learning techniques to generate new information and convert them into a form that can be master by human. The most widely used one is the paradigm of rules from a set of training examples [2]. The system is able to classify and generalize the given examples by establishing and following several indicators.

Using human intelligence to analyze emotion is a milestone during the development of machine learning, someone even argues that it is more important than mathematical and verbal learnings [3]. Emotion, as an abstract thing, is generally hard to capture and explain. However, preceding researchers have found some way to deal with it.

During the previous research works, several models had been adopted to recognize emotion. Music IR is a system that tend to use either categorical descriptions or parametric models of emotion for classification or recognition. It is supported by many psychological researches. The system categorizes al emotion into four groups and establishes a rectangular plane coordinate system based

on Valence- Arousal (V-A) model. The researchers present the emotion as a single dimensional vector in the plane depends on different indexes (e.g., lyrics, tension, texture) [4]. Another model is related to visualization, recognizing the emotion among music by following the machine learning models of training, analysis and composition. This model is firstly used on detecting the emotion appears in photos. When it comes to use for music visualization, researchers primarily cut music in to segments which have homogenous emotion, and then select photos which present the same kind of emotion as music pieces, that make the visual and auditory consistent. This model also uses many mathematical formulae to determine the relevance of the music and the picture, which include corresponding the brightness of photo and the spectral centroid of music, the color of the photo and the timbre of the music, etc. [5]. Those above-mentioned models have their unique advantages of analyzing emotion and create significant results, which are the key points that will be discussed in the following paper.

Expressing personal or collective emotion is always an essential purpose of creating music, by using the most leading-edge technology, the skills of composing can definitely be improved if people get to know the origin theory of creating different emotions. Models used for emotion recognition provided a different prospective for researchers to consider, enjoy and appreciate music. The primary aim of the study is to compare different methods which are used to analyze emotions in music and give the prediction of the differences between the results. This paper will firstly introduce some common features for emotion recognition, the basic process of machine-based emotion recognition and the way to judge whether the result given by the machine model is correct or not. Afterwards, the applications and comparison between different models will be demonstrated accompanied with clarification of the limitation in present computer emotion analyzation in music and forecasting the future of computer music. Eventually, a brief summary will be presented.

2. Features for Emotion Recognition

To extract concrete feelings from those abstract, consistent and contrapuntal melodies, using indexes which can evaluate and relate music with emotion is a must. Common features in music such as rhyme, tone, harmony and speed take an important role during the recognition of emotion. There are mainly two kinds of feature for emotion recognition: Standard Audio (SA) Features and Melodic Audio (MA) Features. Standard Audio Features applies the most general way of analyzation, which computes and analyze a short-time audio waveform that extracts from the entire audio file. The work will calculate essential numerical value of the graph, e.g., sloop, centricity, climax. Melodic Audio features mainly computes melody-related indexes, e.g., the highest and the lowest pitch that occurred, the duration and range of notes. In the previous studying, one also applied SA+MA method, which exactly was the combination of these two features [6].

There is also a useful tool called MIR tool box that be used to extract acoustic features, including total 34 feature values. These features come from the mathematical analyzation from the wave graph of the music and give us a comprehensive understanding of music. For example, in the 34 features in MIR system, Root Mean Square (RMS) is used to calculate the power of the sound systems by knowing its amplitude. Tempo can be used to detect the repetition, which further determines the speed of the music. Zero-crossing Rate shows the noise degree, frequency and the split [7].

Moreover, the importance of lyrics in the performance of musical emotion cannot be ignored. Selecting lyric features is an even more hard task. Previous researchers developed a tool called the Affective Norms for English Words (ANEW), which can be used to label different words with corresponding emotion, and there also developed Affective Norms for Chinese Words (ANCW), which help to make sure the recognition system is still effective under different language environment. However, the most used approach for lyrics' emotion recognition Bag of Words (BOW), which can compute the frequency and order of appeared words [8].

These features are easily to understand and capture in the music, which help to transfer the audio to some quantified values, in order to provide us a clearer way to detect the emotion expresses in the music. Furthermore, they are the essential base of further developed system and models for emotion

recognition, determining how the future study goes. Table 1 has shown a collection of some common features.

Table 1. Common features used in the emotion classification

Type	Features
Dynamics	RMS Energy
Timbre	MFCCs, spectral shape, spectral contrast
Harmony	Roughness, harmonic change, key clarity, majorness
Register	Chromagram, chroma centroid and deviation
Rhyme	Rhythm strength, regularity, tempo, beat histograms
Articulation	Event density, attack slope, attack time

3. Machine Learning Models and Applications

3.1 Machine Learning Models for Emotional Recognition

The most original way of presenting emotion is through physiological signals that are released by human race. Physiological signal is always a reliable index, although people often unaware of them. In order to recognize specific emotion, previous researchers have generally classified various emotion by four dimensions [9], as shown in Fig. 1. Fig. 2 presents an entire emotional recognition process. In the figure, PPG and GSR indicate two different physiological signals. PPG means photoplethysmography, which can detect the changes of blood volume, heart rate and oxygen saturation. GSR is the galvanic skin response, that can measure skin's electrical conductivity. The system was designed to recognize three kinds of emotion: amusement, sadness and neutral (no emotional feelings) by collecting data from sensors and extracting appointed features.

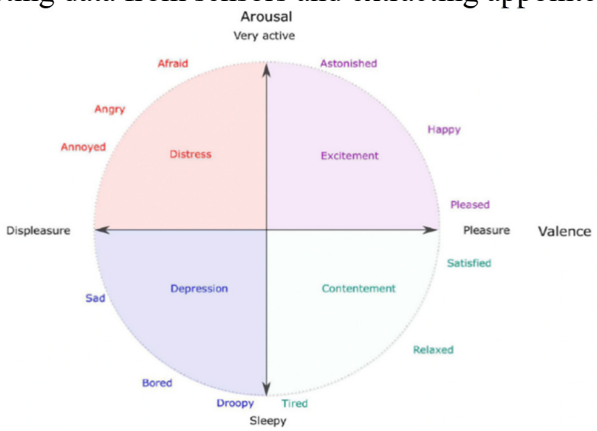


Fig 1. Russels circumplex model of affect.

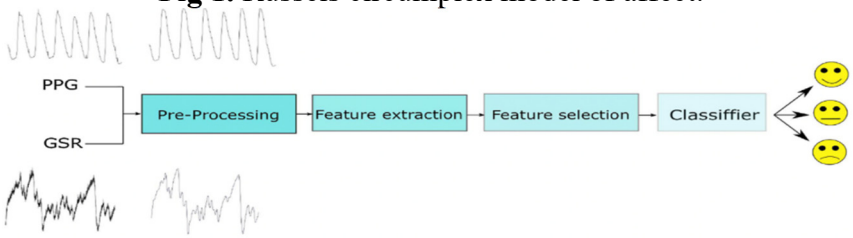


Fig 2. Methodology for emotional recognition

Apart from traditional ways of emotional recognition, machine learning models for emotional analyzation of audio files have its unique pattern and steps because of the great differences between physiological signals and sonic signals. One innovative model for music emotional recognition is CNN-based model. It applies Russell's valence-arousal space as its emotion model, using the spectrogram of a song as its input, which is produced through the performing Short-Time Fourier transform [10]. The model is able to recognize the emotion of music pieces by classifying those

features abstracted by CNN in the spectrograms. Another commonly used methods for music emotional recognition are based on electroencephalograph (EEG). It records brain activity when the experimenter is listening to specific music, extracting EEG features and applying machine algorithms in order to classifying EEG dynamics. The EEG models includes more than 30 features and uses two classifiers, which are multilayer perceptron (MLP) and SVM for classification. It employs many neuroscience research works as its theoretical foundation, and has impressive accuracy among various emotional recognition models. The EEG system successfully associates music emotion and EEG dynamics, and is a brilliant example for disciplinary research [11].

As the developing of human intelligence, researchers even innovate model that can analyze emotion of non-musical sound. This field often receives less attention, although it has similar or even more importance as musical sound for human's daily life. A useful system for non-musical emotional recognition is the International Affective Digitized Sound (IADS) system, which provides 167 difference sounds and their emotional ratings. By analyzing the data provides by the IADS system, researchers can establish a computational model for general sounds. Fig. 3 depicts the whole process of using IADS as the base of the audio emotional model [12].

Sound, as one of the most essential parts of human's daily life, undertaking important roles of representing affections. By developing models of recognizing emotion in music, researchers are able to produce more affective sounds for music, movies, and various multi-media. Moreover, music applications can develop a more advanced classification system so that users can receive better personal daily music recommendation.

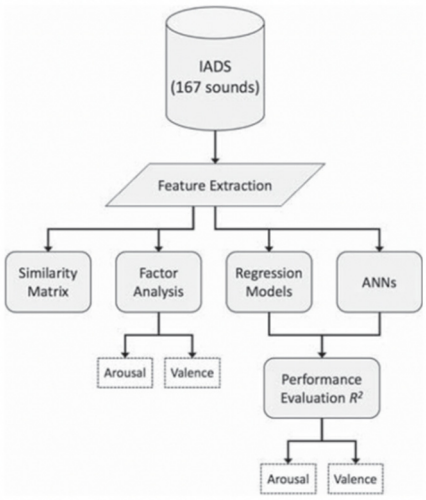


Fig 3. Audio emotion recognition in IADS.

3.2 Metrics of Machine Learning Models

Although computational models seem to own super high accuracy, reliability, and being very persuasive, in order to ensure the accuracy of the model's output, some metrics still need to be set. Many music emotional recognition models start with forming a database, which includes enough music samples that have already been annotated and labelled emotional features. These samples often come from streaming media or previous-established affective systems. The database aims to provide references for the experiment results and determine the precision rate of the experiment result.

For EEG-based learnings, three entropy metrics has been used: Quadratic Sample Entropy (QSE), Amplitude-Aware Permutation Entropy (AAPE), and Permutation Min-Entropy (RPE) [13]. Quadratic Sample Entropy is a development of Sample Entropy (SE), which can represent the irregularity among the signals and find same pattern. It also shows the probability that two sequences that match with m points also match m+1 points. Amplitude-Aware Permutation Entropy is an improvement of permutation entropy (PE), and it can convert original data into a sequence, thereby calculating the repetitiveness of the original signals and measuring the predictability. Permutation Min-Entropy is a very flexible tool in emotional recognition, which can effectively characterize

different signals and estimate signal's suitability for EEG system. Fig. 4 illustrates the metrics steps in EEG system.

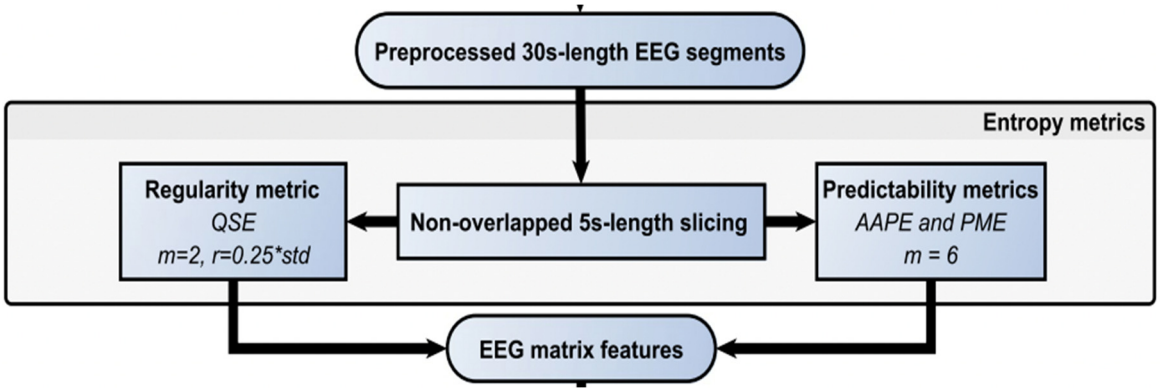


Fig 4. Metrics in EEG model.

Another method to evaluate the experiment result is comparing the output data by using a different computational algorithm. In order to prove the performance of the above-mentioned CNN-based model, researchers compared their result with SVM algorithm, which is a very traditional and commonly used machine learning model, and they also used SGD algorithm to train the SVM. By comparing the data extract through different algorithms, they evaluated their accuracy.

3.3 Comparison of Models

The result of the emotional recognition will easily be affected by each step that involves in the model. In this part, the paper will mainly discuss and compare how different classification methods and feature extraction methods affect the final emotional prediction. Almost all emotional recognition models are based on the V-A model and basically classify different emotions into four dimensional. However, there are still other features need to be identified for further classification. Previous researchers have tested experimenters under different classifier environment. The first experiment was held when the speakers know to the classifier, which means they have already presented to the training set before. Therefore, this experiment was in a speaker dependent framework. The entire successful rate of the first experiment was 83.17% [14]. The second experiment is speaker-independent, which repeated 5 times, each time measure a pair of speakers. By analyzing the data they collected, researchers found out that the MLP performance (55% correct classification) had higher accuracy than Random Forest classifier (49% correct classification). Moreover, Random Forest classifier reached higher recognition rate than MLP, which were 89.4% for high arousal and 82.52% for low arousal. For EEG system, five features have often been extracted in order to recognize emotion: power, entropy, fractal dimension, statistical features and wavelet energy [15]. After comparing two different nature selection methods, Relief algorithm and principle component analysis (PCA), the latter one received better result for its improved performance and higher efficiency.

4. Limitations & Prospects

Although the emotional recognition model and the machine learning techniques are already well developed, there are still many limitations exist in current emotional analyzation system. In the reality, individuals present their emotion in different form and manner, so do what musicians present in their music. Modern music takes different composing methods, often with asymmetry and frequent modulation and transposition, which makes it even harder for machines to identify the concrete emotion that composers want to present based on just a few simple features. Moreover, modern musicians tend to sampling non-music sound, sometimes noise, into their art work. Thus, traditional features setting for music emotional features, such as harmony, pitch and rhyme, might not work for those experimental and progressive music.

Emotion is always a complex thing to understand and analyze, even for human ourselves. The recognition can never reach one hundred percent even under the ideal circumstances, however, the tasks in the real life is far more intricate and implicit than in the laboratory. However, as composing based on computational algorithm become new trend, those music pieces create by machine seem to be easier to recognize by machine themselves since they follow a standard model to compose, and the foundation rule for computer music are based on algorithm as well. The technique of music emotional recognition still has a long way to go, and it will indeed affect the future music and computer industry in a large extent.

5. Conclusion

In summary, this paper investigates emotion recognition based on various machine learning scenarios. Specifically, EEG, CCN and IADS systems are demonstrate. According to the analysis, existing models can already recognize and predict emotion in music and even non-musical sound, but still have noteworthy limitations and disadvantages. In the future, more sophisticate system in order to deal with complex reality tasks still need to be developed. Overall, these results offer a guideline for current existing works for music emotional recognition models and simply compare the differences between different classifications and feature extracting methods for emotional analyzation.

References

[1] Computer Music (so Far). Short History of Computer Music, Retrieved from: http:// artsites. ucsc. edu/ EMS/ Music/equipment/computers/history/history.html.

[2] Peter Clark. Machine Learning: Techniques and Recent Developments. 1990, Retrieved from: http:// www.cs.utexas.edu/users/pclark/papers.

[3] Picard R W, Vyzas E, Healey J. Toward machine emotional intelligence: analysis of affective physiological state. IEEE Trans on Pattern Analysis & Machine Intelligence, 2001, 23(10):1175 - 1191.

[4] Yang X, Dong Y, Li J. Review of data features-based music emotion recognition methods. Multimedia Systems, 2017.

[5] Chen C, Weng M, Jeng S, et al. Emotion-Based Music Visualization Using Photos. Advances in Multimedia Modeling, 14th International Multimedia Modeling Conference, MMM 2008, Kyoto, Japan, January 9-11, 2008, Proceedings. DBLP, 2008.

[6] Rocha B, Panda R, Rui P P. Music Emotion Recognition: The Importance of Melodic Features. 6th International Workshop on Music and Machine Learning – MML'2013 – in conjunction with the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases – ECML/PKDD 2013. 2013.

[7] Kim Y E, Schmidt E M, Migneco R, et al. Music emotion recognition: A state of the art review. Music Emotion Recognition a State-of-the-Art Review, 2010.

[8] Nalini N J, Palanivel S. Music emotion recognition: The combined evidence of MFCC and residual phase. Egyptian Informatics Journal, 2016, 17(1):1-10.

[9] Hoz E, Santos J, Helena C O, et al. A machine learning model for emotion recognition from physiological signals. Biomedical Signal Processing and Control, 2020, 55:101646.

[10] Liu Tong, et al. Audio-Based Deep Music Emotion Recognition. AIP Publishing, AIP Publishing LLC AIP Publishing, 23 May 2018.

[11] Lin Y. P. et al., EEG-Based Emotion Recognition in Music Listening. in IEEE Transactions on Biomedical Engineering, vol. 57, no. 7, pp. 1798-1806, July 2010.

[12] Cunningham Stuart, et al. Supervised Machine Learning for Audio Emotion Recognition - Personal and Ubiquitous Computing. SpringerLink, Springer London, 22 Apr. 2020.

[13] Beatriz García-Martínez, et al. Recognition of Emotional States from EEG Signals with Nonlinear Regularity- and Predictability-Based Entropy Metrics - Cognitive Computation. SpringerLink, Springer US, 16 Nov. 2020.

[14] Theodoros Iliou, Christos-Nikolaos Anagnostopoulos. Comparison of Different Classifiers for Emotion Recognition. IEEE Xplore, 5298878, 2019.

[15] Nawaz, Rab, et al. "Comparison of Different Feature Extraction Methods for EEG-Based Emotion Recognition." Biocybernetics and Biomedical Engineering, Elsevier, 11 May 2020.