Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/68/20241415

Analysis of the realization for emotion recognition based on machine learning

Fuhong Shen[1, *]

[1]The Affiliated High School to Hangzhou Normal University, Hangzhou, China

[*]The corresponding author's e-mail address: 1714030113@stu.hrbust.edu.cn

Abstract. As a matter of fact, music emotion recognition based on machine learning has been a trend in recent years. This study describes the analysis of the realization for emotion recognition based on machine learning. In this paper, the study expresses some methods of analyzing the emotion of the music composition. The paper represents some classifier molds that enable us to determine the emotion of the music. It aims to provide readers with a deeper comprehension of the music and composers' feelings and to know more about the background of the music. The study may let more people to improve their own ability to appreciate music, if this process has let great amounts of individuals known. It also allows people to have more awareness and feelings with more voices. In this paper, the research will also show some about the evolution and original of the synthesizer. Moreover, the study will go to illustrate this through scientific, real examples and personal experiences and pictures.

Keywords: Emotion, music, machine learning.

1. Introduction

Contemporarily, a number of technology advancements have both benefited from and prompted the expansion of Computer Music as an art and academic subject. Among these, a piece of software called a music programming system is one of the most durable. Many discoveries and innovations were made possible by these packages as composers and researchers shaped and developed them; some of these innovations if the music business later took it as a crucial component of the widely used digital audio seen in production and performance [1]. Computer music has a long history of more than 50 years. Robert invented the world's first analog sound synthesizer called the Miniature Moog, which has A variety of timbre. Since then, the synthesizer has entered the public's vision and entered a revolution of electronic music. The early electronic music laboratory mainly consists of many sine wave signal generators, white noise signal generators, rectangular wave oscillators, filters, ring modulators, variable speed recorders, dynamic suppressors, and paper tape punchers for recording control information. These were the early synthesizers, and they could synthesize all kinds of sounds, even ones that didn't exist [2].

After producing an Analog voice synthesizer, scholars also made a digital synthesis software. The initial software for direct digital synthesis, MUSIC 1, created by Max Mathews in 1957, and its offspring, cleared the way for contemporary methods for programming music. Their development was marked by more flexibility and generality, ranging from basic synthesis programs to completely functional programming languages. Along the way, certain key innovations were suggested. For example, Mathews presented the table-lookup oscillator in MUSIC II, this is now the component of digital synthesizers that

DOI: 10.54254/2755-2721/68/20241415

is used the most. In MUSIC I, for computer instruments, the ideas of a unit generator, synthesis module, and compiler were introduced. These concepts are still crucial to contemporary systems [1].

Furthermore, one can modify real-world sound samples to produce new sounds. The 1960s saw the invention of analog voltage control circuit technology, which produces distinct vibration cycles and overlaid polyphony, or musical waveforms, when many sets of circuits are utilized simultaneously. It was not used at the time, though, because the technology was not developed sufficiently, the computer was pricey, and each synthesizer's interface varied. The digital synthesizer emerged in the early 1970s. Its benefit is that it can create sounds that are not found anywhere else, but because the sounds are created based on the musician's emotions, they are hard to replicate and utilize， and the sound is different every time. To meet this need, people have invented sampling playback and direct sampling synthesizers. After the late 1980s, people began to experiment with computer arrangements [3]. Besides, 1900 saw the invention of the direct current arc oscillator, and in 1906 Lee De Forest received a patent for the vacuum-tube triode amplifier valves he same year that the dynamic phone was first demonstrated. The industry was well established by the conclusion of the war, and several engineers succeeded in look into the potential of employing the new technology to build electronic musical instruments. Progress was made slowly but steadily. The majority of these designs were primarily driven by the desire to expand the traditional orchestral range, with the underlying goal being that composers would be convinced to supply appropriate repertoire. Thus, the primary goal of the emerging gadgets was to fulfill conventional notions of musical writing. Certain pieces, like the Neo-Bechstein Piano [2].

This study will search some information from the people who study the realization for emotion recognition based on machine learning and they found that it is common practice to extract musical elements from audio signals, textual lyrics, symbolic music scores, and even biological features such as electroencephalograms (EEGs). An unlimited number of points in a continuous multidimensional space or a finite number of discrete categories can be used to depict emotion space [4]. Numerous earlier studies have classified emotions into various emotion classes in an effort to understand the connection between emotion and music. They have then trained a classifier using the usual pattern recognition approach. While the method acknowledges the ambiguity of adjectives and defines emotional categories in terms of arousal (excitement or calm) and valence (positive or negative), basic emotions such as happiness, anger, sadness, and fear are adopted as emotional categories [5].

The objective of this study is to research the emotion of the music composition. This study will examine the composer's emotions when they writing. In this case, it taking advantages of analysis the emotion of the music composition, because one can better know the express of the song writers. Thus, one can have a great understanding of a song. In addition, in the recent years the academic and industry world has been paying a lot of attention to music emotion detection because of its many applications in areas such as music visualization, automated music composition, psychotherapy, recommendation systems, and more. Deep learning-based music emotion recognition is steadily making a breakthrough, especially given the speed at which artificial intelligence is developing [4]. This study will strengthen the understanding of the music. The rest part of the paper is organized as follows. First, this study will describe the classification of musical mood and the basic expression of emotions. Secondly, this study will show some models of the realization for emotion recognition based on machine learning. Third, this study will do some conclusion about the research. Eventually, this research will point out some limitations for present inventions and provide some suggestions.

2. Basic descriptions for emotions of music

Personally, it is thought that there are plenty of emotions in a song. Scholars have a question about so what is the emotions. Previous studies have already found that the characteristics of feeling. Clarifying what is meant by "emotion" more precisely is one of the first steps. This term, which is frequently used in ordinary speech, is a hypothetical construct, meaning it defines an underlying phenomena conceptually and operationally and serves as the subject of theory and investigation. A componential approach to emotion has been taken by the majority of contemporary emotion theorists, who contend that coordinated changes in multiple components make up an emotion episode. The trinity of emotional

DOI: 10.54254/2755-2721/68/20241415

response: motor expression, subjective feeling, and physiological arousal— have been recognized as the three main reaction components of emotion across the ages. Behavior preparation (action readiness, action inclinations) and associated added to the list of components are cognitive processes that recognize and generate distinct emotional reaction patterns [6].

According to the Wilhelm Wundt's research he proposed using introspection to evaluate three aspects of emotions: tension–relaxation, rest–activation, and pleasantness–unpleasantness. This three-dimensional paradigm has significantly influenced affect and emotion psychology. Since it has been challenging to consistently determine the third dimension using factor analyses, sensation has been frequently charactery in terms of the two-dimensional space that activation and valence take up. These two-dimensional models are appealing because they may be used to show how different emotions are comparable in terms of their spatial vicinity. Tension-relaxation is a crucial component in musical analysis [6].

One often has the question about what is the classification of the moods and how can one knows the emotions of the music correctly. Normally, one knows there are happy, angry, depress, excited even have some feeling like falling in love etc. It is found that to defined the emotions of the music not only depend on listening one also needs to use other ways to accurate positioning the emotion of music. For instance, Comparing Marsyas8 with the best-performing lyrical feature set revealed that Marsyas8 was the best-performing audio system in the MIREX 2007 AMC task evaluation to see if this was true for music mood categorization. Marsyas employs sixty-three spectrum features, such as the spectrum Centroid mean and variance, roll off, Flux, Cepstral Coefficients of Mel-Frequency, etc. For classification, it uses a linear kernel and LIBSVM as well. All of the dataset's audio tracks were converted to 44.1KHz stereo before being imported into Marsyas.wav files. Once the spectral characteristics were obtained, the data was processed using SVM classifiers [7].

3. Models

The goal of the developing transdisciplinary study topic of affective computing is enables the recognition, inference, prediction, and interpretation of human emotions by intelligent machines. It includes social science, cognitive science, computer science, artificial intelligence, neuroscience, and neuropsychology. To recognize emotional cues is the goal of affective computing. HCI and evoking sentimental reactions. What affective computing is the collection of methods for recognizing impact from data at various granularities and modalities. Research in affective computing primarily consists of the subjects of sentiment analysis and emotion detection. The previous. Does coarse-grained affect recognition, which is typically a task with three classes—positive, negative, and neutral—or binary positive vs. negative attitudes the former entails detailed analysis (often classification of multiple classes of enormous information get into a bigger collection), whereas the latter are normally good at analyzing fine-grained.[8]. It is completely insufficient and difficult to analyze the emotion of a song only through people's feelings, because everyone is different, and everyone's feelings are different, so one can use machines to analyze it more scientifically and reliably. Hence, it is found an experiment [9]. about the analysis of the emotions in the music composition. This study is for the listener, but it is assumed that it can reflect the emotions which composers have if one has enough people and combine some of the composer's back stories from that time. There are also have other methods to achieve the goal. A database called the DEAP database for physiological signal-based emotion analysis founded on the emotion paradigm of valence, arousal, and dominance, is employed in this study. The DEAP dataset has thirty-two people. For every subject, one-minute music videos were played in order to activate the visual and auditory cortex. Each participant saw 40 music videos, and seven distinct recording modalities were used. This study used EEG; further details are available in. To ensure that the 40 video clips had a sufficiently wide valence/arousal scope, their valence/arousal time length was predetermined. Each participant was asked to assign a score of 1 to 9 for dominance, likeability, valence, and arousal to each music video. Hence, if the grade was higher than 4.5 the arousal will be strong, vice versa [9]. As a result, one can use the database to directly analysis the emotions of the people, because there are so

DOI: 10.54254/2755-2721/68/20241415

many people, one can avoid some accidents for the study [9]. The classification of human emotions looks like the results shown in Fig. 1.

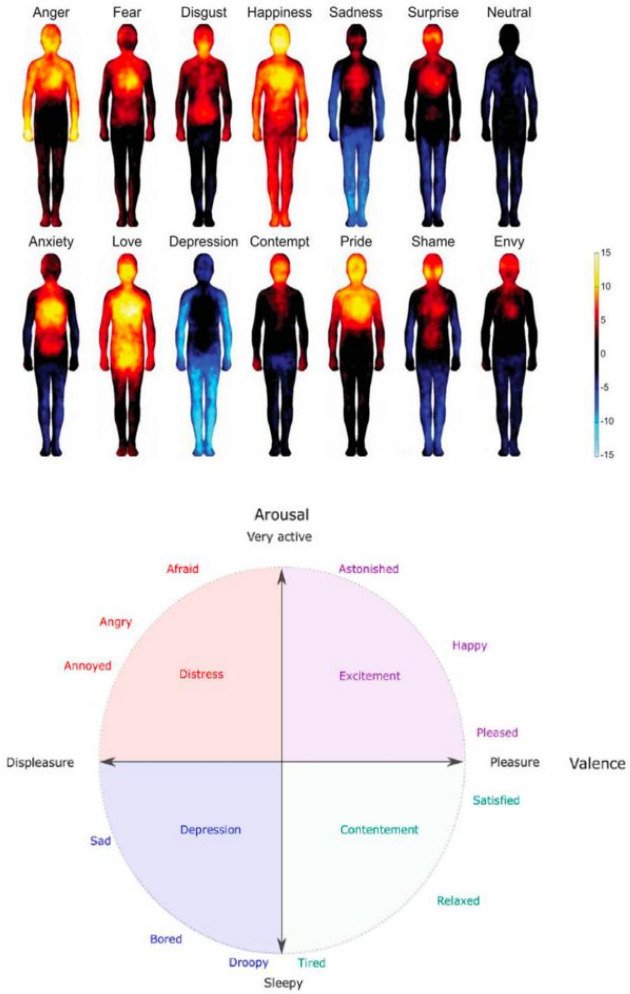

Figure 1. The emotion visualization matrix [9].

Table 1. Summary of cross validated value for certain channels.

Cross-validated Accuracy (%)	F3- F4	F7-F8	Channels FC1-FC2	FC5-FC6	FP1-FP2
Arousal-SVM	90.8	87.9	87.9	85.1	88.1
Valence-SVM	90.6	94.9	84.9	85.5	88.1
Arousal-KNN	76.4	79	79	77.6	71.5
Valence-KNN	79	80.4	80.4	75.5	73.6
Arousal-ANN	82.1	83.2	83.2	77.7	80.4
Valence-ANN	84.7	83.1	83.1	74.4	78.9
Arousal-SVM	85.7	79.7	80.3	83.1	82.3
Valence-SVM	84.8	81.2	82.2	82.8	83.2
Arousal-KNN	68.5	70.2	65.3	66.6	64.5
Valence-KNN	69	68	68.2	69.9	63.5
Arousal-ANN	74.3	76.2	73.6	70.2	73.4
Valence-ANN	71.2	79.2	69.4	68.5	74.6

DOI: 10.54254/2755-2721/68/20241415

Table 2. Summary of cross validated value for different tune.

Cross-validated Accuracy (%)	Gamma	Beta	Frequency Bands Alpha	Theta
Arousal-SVM	89.8	91.3	90.4	89.4
Valence-SVM	89.7	91.1	90.9	89.1
Arousal-KNN	75	75.1	72.6	73.8
Valence-KNN	77	78.1	78	76.1
Arousal-ANN	81.3	86.5	83.5	80.7
Valence-ANN	84.8	88.3	97.2	82.6
Arousal-SVM	81.4	89	88.68	86.6
Valence-SVM	82.43	89.34	89.72	86.1
Arousal-KNN	76	73	73.7	75.6
Valence-KNN	74.3	73.2	73.8	72.4
Arousal-ANN	83.2	78.4	82.7	81.9
Valence-ANN	79.1	75.2	86.1	84.3

Some of the results are shown in Table 1 and Table 2. The SVM, ANN, and KNN are trained using all of characteristics that were taken out of every channel in each frequency band. Table 2 displays the cross-verified accuracy of each classifier that was drilled using the important portions of each band. The optimal accuracy was 91.3% for arousal and 91.1% for valence, are produced by the drilled SVM classifier making use of beta frequency band characteristics [10]. The accuracy of cross-validation findings in Table 1 and Table 2 clearly demonstrate that the SVM performs better than KNN and ANN. Using KNN, the biggest data of classification accuracy was 86.75% in the earlier work [9]. Using PCA in this investigation resulted in the retrieved features being uncorrelated. Furthermore, selecting the right kernel (RBF) significantly improved the cross-validated accuracy [10]. Based on previous research findings，they prepare three kinds of classifiers in order to find the most accurately machine of the analysis emotions of the music composition.

4. Limitations and prospects

From the point of view, it is supposed that if researchers just rely on the machine to analyze the emotions of the music composition are not enough, because the machine is too blunt and not comprehensive enough for emotional analysis. Just according to some change frequency and brain waves are not enough. The reason is that one cannot rule out the composer's intention. For example, there is a song named “Farewell book” which from China. This is a pure music. It is supposed that when everyone first listens to this melody they will all think this song is a high mood song, because it has brisk pace and it seems that people are walking along the path of the field and enjoying the beautiful scenery. Whereas, you will never think that this music’s true situation is the son and mother went to the park and looked at the flowers for the last time. Soon after，the mother died. In conclusion, according to this, the machine will just recognize happy mood about the people who just first listen to this song. This is where these machines are still flawed. The suggestion is before recognize the emotion of the music one needs to know the background of the music. Personally, this is something that machines can never do. Besides, in the future, It is expected that the machines for recognizing human moods are getting better and better，more and more humane.

5. Conclusion

To sum up, the current findings have greatly increased the understanding, and thanks to the information now available of the emotions analysis of the music composition. This research has a deep study of this project. At first, it is just thought that it only related some personal feelings. However, the analysis needs to use some professional machines to obtain. According to the analysis, the different results based on various models have been demonstrated. It is meaningful for me to do such a study and this in itself is very meaningful. It can provide a better understanding of the composers, it will let us have a very good

DOI: 10.54254/2755-2721/68/20241415

understanding and experience of music. Overall, these results shed light on guiding further exploration of music emotion recognition based on machine learning.

References

[1] Lazzarini Victor. The Development of Computer Music Programming Systems. 2013 Journal of New Music Research vol 42(1) pp 97-110.
[2] Manning P. Electronic and computer music. 2013 Electronic and computer music. Oxford University Press.
[3] Wu Y. On computer music and computer composition. 2008 Computer sound engineering composing electronics. Shanghai Conservatory of Music.
[4] Han D, Kong Y, Han J, et al. A survey of music emotion recognition. 2022 Frontiers of Computer Science vol 16(6) p 166335.
[5] Yang Y and Homer H C. Machine recognition of music emotion. 2012 CM Transactions on Intelligent Systems and Technology (TIST) vol 3(3) pp 1-30.
[6] Scherer K R. Which Emotion Can be Induced by music? What Are the Underlying Mwchanisms? And How Can We Measure Them? 2004 Journal of new music research vol 33(3) pp 239-251.
[7] Hu X, Stephen D J and Andreas F E LYRIC TEXT MINING IN MUSIC MOOD CLASSIFICATION 2009 American Music vol 183(5) pp 2-209.
[8] Zhang J, Yin Z, Chen P, et al. Information Fusion 2020 Information Fusion vol 59 pp 103-126.
[9] Domínguez-Jiménez J A, Campo-Landines K C, Martínez-Santos J C, et al. Biomedical Signal Processing and control. 2020 Biomedical signal processing and control vol 55 p 101646.
[10] Bazgir O, Mohammadi Z and Habibi S A H. Emotion Recognition with Machine Learning Using EEG Signals. 2018 25th national and 3rd international iranian conference on biomedical engineering (ICBME).