

Personal Music Emotion Analysis Based on BP Neural Model

Wen Yaxin

Beijing Film Academy, Beijing, China, 100026

Keywords: music emotion analysis; BP neural network; music feature model

Abstract: Music as an art is an indispensable part of human life. From ancient times to the present, music has always been a special channel for humans to express their emotions with rich expression style and delicate emotional classifications. Thus, the emotional analysis of music has become an important issue. With the advent of the information age, the analysis of musical emotions through computers has begun to enter people's horizons. Music emotion analysis model can be generally divided into three parts, music feature model, music emotion model and classification cognitive model. The music feature model consists of features extracted from music. This paper proposes some new features. The emotional model mainly describes the classification of emotions. The BP neural network model is selected for this classification cognitive model.

1. Introduction

Analyzing the emotions in music through computers has gradually entered people's perspective. With the advent of the information age, the multimedia field including music has developed rapidly. BP neural-based multimedia information processing technology is an important research hotspot in this field, especially after the introduction of animation, audio, video and other dynamics. Multimedia technology has greatly enriched computer expressions of human emotions. Emotional computing is a highly integrated technical field. Applying it to the music field will involve psychology, cognitive science, computational science, music theory, and so on. Music emotion analysis is of great application value in many fields. Content based affective analysis will break through the previous forms of music classification based on the label form, making the restrictions of the audience smaller, earlier or more new, and the incomplete music fragments no longer exist. Secondly, classification can also be applied to music recommendation. Content based recommendation can be adjusted according to the preferences of the individual to meet the human differences in the understanding of music characteristics. Besides, the study of music emotion analysis will also help to promote the development of other research fields. The research on recognition technology will promote its reverse process, that is, the development of automatic music generation technology and the breakthrough of automatic composing, and the research of music emotion will promote the progress of artificial intelligence.

2. Musical Emotion Model

2.1 Basic emotional characteristics of music.

First is subjectivity. The author's musical works reflect their psychological emotions and are the subject of the author's sustenance of emotions, which is highly subjective. As an aesthetic object, music is not a clearly recognizable object. It does not have a strict concept of logic, and it is often ambiguous and embarrassing. Due to cultural, environmental, and personal differences, each person's understanding of music is different. On the other hand, differences in mindset and mood can also cause deviations in the understanding of music. Creators use different creative techniques and tactics to form different genres and styles; listeners have different interpretation methods for different schools and styles, and thus have different emotional experiences. Second is hierarchical. There are many levels in the process of cognitive music. Styles, emotions, patterns, and musical forms are the

highest levels. To get into the highest level, it must first pass through lower cognitions, such as knowledge of rhythm, speed, and strength. The understanding of space-time features such as rhythm, speed, and intensity is based on acoustic characteristics, such as the timbre and pitch of some notes. Therefore, the emotional perception of music can generally be divided into multiple levels. Third is the objectivity. Music emotion is not only subjective, but there are also uncertainties and definite aspects in musical image. The musical form of the work also imposes certain restrictions on the emotional response and imagination of the viewer. All kinds of fun judgments have universal validity, and this universality provides a certain existence basis for relatively consistent aesthetic evaluation standards.

2.2 Musical emotional psychological model.

The Thayer model is a more common musical emotion model. As shown in Fig. 1, the Thayer model is a first quadrant of two-dimensional coordinates, the horizontal axis represents tension, and the vertical axis represents energy. The Thayer model is a continuous classification model of emotions.

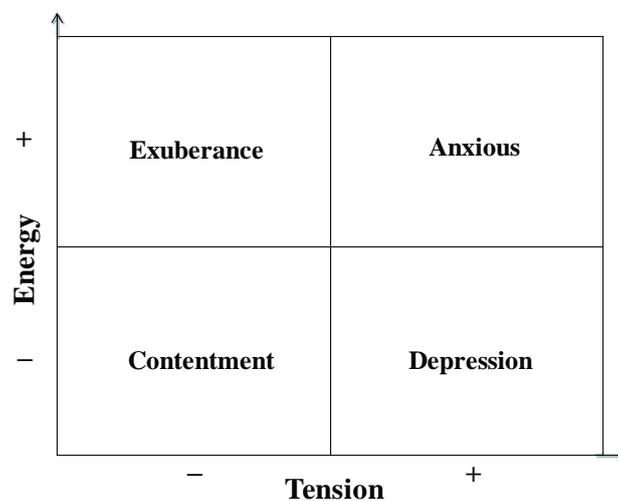


Fig.1 Thayer model

The horizontal axis coordinate pressure is from relaxed to happy and anxious, reflecting the individual's subjective internal psychology; the vertical coordinate energy is from calm to active, reflecting the individual's physiological vitality.

2.3 Classification cognitive model.

The classification cognition model takes the feature vector output from the feature model as input, and classifies the music data according to the emotion classification defined by the emotion model. This model is dominated by the algorithm, and the object selection is mainly pattern recognition algorithm.

3. Pattern Recognition Algorithm

3.1 Naive Bayes classification.

The definition of the Naive Bayes classification is: Set $x = \{a_1, a_2, \dots, a_m\}$ be an item to be categorized, and a is a feature attribute of x ; A collection of categories $C = \{y_1, y_2, \dots, y_n\}$; Calculation $P(y_1|x), P(y_2|x), \dots, P(y_n|x)$, where $P(B/A)$ is the conditional probability of occurrence of event A under event B ; if $P(y_k|x) = \max \{P(y_1|x), P(y_2|x), \dots, P(y_n|x)\}$ then $x \in y_k$. When using the Naive Bayes algorithm, if the data set satisfies the independence assumption, the classification accuracy is high. In addition, the algorithm does not have a classification rule output.

3.2 Neural network algorithm.

The artificial neural network (ANN) algorithm simulates the function and composition of a neural network in living organisms, and is an adaptive system composed of a large number of simple basic elements - neurons interconnected. Artificial neural networks need to be learned according to certain criteria before they can be formed. The characteristics of the artificial neural network are: (1) It has preliminary self-adaptive and self-organizing capabilities. After the acquired learning and training, different functions can be developed; (2) Generalization functions: For untrained samples, there is good predictive ability and control ability. (3) Nonlinear mapping capability: It does not require a thorough understanding of the system, but at the same time it can achieve the mapping relationship between input and output, simplifying the design difficulty; (4) high degree of parallelism.

3.3 BP neural network.

BP (Back Propagation) network is a kind of artificial neural network. As its name implies, it trains the network by back-propagating the results. It is a multi-layer feed-forward network. The BP network is a supervised machine learning method, which can obtain the corresponding algorithm model by training the existing input and output mapping relationships, and it is not necessary to understand the explicit training set mapping rules before training. The reverse propagation of BP networks depends on the weights and thresholds of each layer of network. When the weight and threshold are adjusted to make the result error less than the set standard, the network is convergent.

The structure of the BP neural network is generally divided into three parts: the input layer, the hidden layer, and the output layer. The input layer is responsible for receiving the input information of the user. It is necessary to determine the number of neurons according to the actual situation. The input layer receives information and passes it to the hidden layer. The hidden layer is responsible for the processing of information, and is the center of the whole network. The structure of the hidden layer is more complex. It can contain only one layer, and it can also contain many layers, and the number of neurons in each layer needs to be set. When the hidden layer is processed, the information is passed to the output layer and the result is displayed to the user from the output layer. The number of the neurons in the output layer is also determined according to the actual classification. When the output of the output layer and the expected output of the sample are more than a standard, the training enters the backpropagation stage, from the output layer to the hidden layer, and then from the hidden layer to the input layer, and the weights of each layer are corrected in this process. After the completion of the reverse transmission, the new training data are transferred forward again, and the process of the positive and reverse cycle is the process of the continuous adjustment of the network, that is, the learning process of the neural network. Fig 2 shows the structure model of BP network.

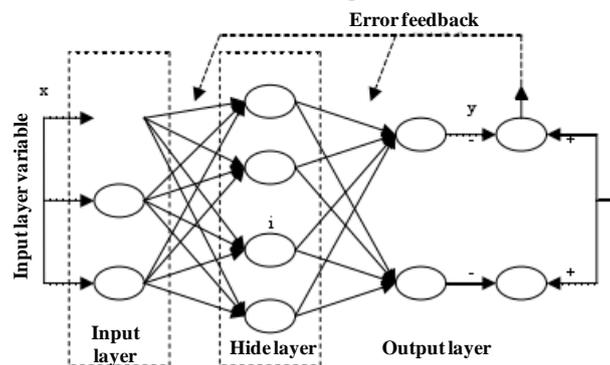


Fig.2 BP network structure model

4. Music Emotion Cognition Model Based on BP Neural Network

Application of BP neural network in cognitive model. The number of elements in the input layer of the BP neural network depends on the number of features of the sample feature vector. Four features were extracted in the melody area method, and 17 features were extracted from the audio analysis method. A total of 21 features were extracted, so the input layer had 21 elements.

The number of units in the output layer depends on the number of sample classifications. The output results can be expressed in two ways. If the total number of categories is m , then the first method uses numbers 0 to 1 in m bits, each bit representing the degree of relevance of the music to an emotional type, 1 being fully correlated and 0 being completely irrelevant. Since the Hevner affective ring model has divided emotions into eight categories, so the output layer has eight elements. The classification is defined as {vitality, enthusiasm, happiness, lightness, lyric, lyric, yearning, sad, sacred}, if the music is completely related to life, then the output layer output vector is {1,0,0,0,0,0,0,0}.

The hidden layer can be set up as a multi-layer or a layer. However, continuous functions in arbitrary closed intervals can be approximated by a network of intermediate layers. The determination of the number of elements in the hidden layer is a complex task, which is related to the requirements of the processing problems, the number of units in the input layer and the output layer. If the number of neurons is too small, the accuracy is not enough. If too much, the training time will be very long, the error may not be the best, and the correction may be excessive. According to experience, and after many experiments, the relationship between the input layer, intermediate layer, and output layer is N , $4N-4$, and M . Therefore, the number of hidden layer neurons selected in this paper is 80.

The initial value selection of the parameters is also very important for the network to converge at the end. In general, the initial values of the weights are set to be relatively small, sets all initial weights to 0. In the result correction based on personal preference, when it is necessary to train the network repeatedly, it is defined that the initial parameters used in each retraining after the first training network are set as parameters when the previous model training converged.

For the empirical choice of the learning rate, too large learning rate leads to the instability of the learning, and the too small value leads to too long training time. The adaptive learning rate achieves a reasonable high rate and reduces the training time by ensuring stable training. Therefore, in order to ensure the stability of training at first, a smaller learning rate value is selected, and the range of the general learning rate is between 0.01 and 0.8.

The transfer function of the middle layer is a tangent sigmoid function, and the transfer function of the output layer is a sigmoid function. Its formulas are:

$$\tan \operatorname{sig}(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (1)$$

$$\operatorname{sig}(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

The images are showed as bellow:

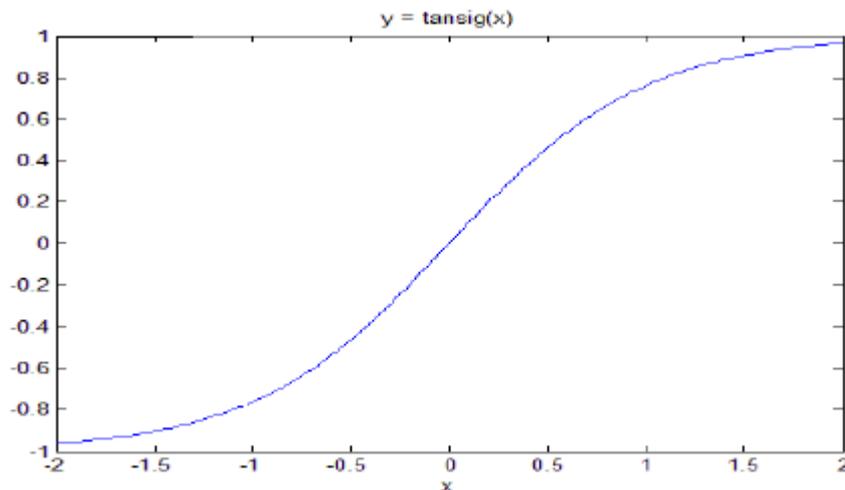


Fig. 3 S-Type tangent function

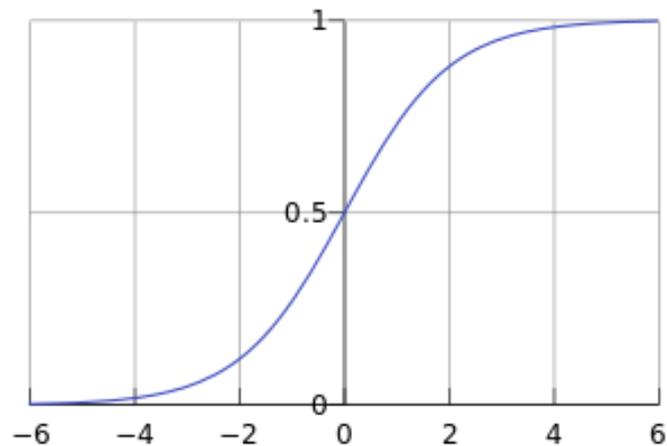


Fig. 4 S-Type function

The maximum number of training cycles is 1000 and the maximum error is 0.0001.

5. Neural Network Based Personality Correction Model

According to the characteristics of the neural network, a method for personality correction of the model results is proposed. Musical emotions are subjective. Growth, life habits, personality, and even different moods can lead to slight deviations in the understanding of musical emotions. Therefore, the output of the model does not have a correct result standard in absolute terms. Based on this, the neural network is used to repeat the training characteristics. Specific steps are as follows:

- 1) The data is fixedly divided into a training set, a test set, and initialization parameters.
- 2) Train a BP neural network model.
- 3) Play the music in the test set to the tester and let them select a dominant emotion, a secondary emotion, and compare it with the results given by the model. After playing 10 songs continuously, if the correct rate is higher than a ratio, the model can be trained and the training is over; if the correct rate is lower than a certain ratio, record the current parameter value and go to step 4.
- 4) Update the 10 songs tested according to the new information provided by the user, put them into the training set, and exchanges 10 songs. Set the parameter value to the initial value when the model is retrained. Return to step 2 to continue execution.

The correctness criteria in step 3 can be modified, which can be generally set at 70% or more. After training, the number of music can be adjusted according to the actual size of the database. In the correct standard, adjustments can be made due to the concept of dominant emotions and secondary emotions. For example, you can only allow the main emotions to be correct, or allow the user to select one of the primary and secondary emotions, and so on.

6. Summary

The relationship between music art and science is very complicated. It involves not only expressing problems but also deep cognitive and aesthetic issues. Emotion is the essential feature of music, and the research of musical emotion cognition model and interactive technology is an important part of music content understanding, natural harmony and human-computer interaction. The classification cognitive model is the core of the whole model. The feature model and the affective model all serve it: the feature vector output of the feature model is the input of the classification cognitive model, and the sentiment model defines the criteria of the output sentiment classification. In the cognitive model of classification, we use a relatively strong adaptive BP neural network. According to the characteristics of BP neural network and the characteristics of music emotion, it also proposes a result correction method based on personal preference.

References

- [1] Picard R.W. Affective computing: Challenges[J], International Journal of Human Computer studies,2013,59(1-2):55-64.
- [2] Li Jianjun, Ding Zhengsheng, Zhang Haiyan, Analysis of Common Discriminant Classification Methods [J], Journal of Xi'an University of Science and Technology, 2017, 27(1):138-142.
- [3] Zhou Yexin, Analysis of Musical Emotion [J], Artistic Baijia, 2005(4): p91-94.
- [4] Feng Yazhong, Zhuang Yueting, Pan Yunhe, A Heuristic Hierarchical Method for Retrieving Music by Singing[J], Computer Research and Development, 2014,41(2):333-339.
- [5] Lai Guoming, Yang Shengyun, Content-Based Music Retrieval Research[J], Computer Engineering and Design, 2006, 27(23):4468,4528.
- [6] Doraisamy S, Ruger S, Robust Polyphonic Music Retrieval with N-grams[J], Journal of Intelligent Information Systems, 2013, 21(1):53-70.
- [7] Peng Qiong, Zhi Jun, Research on Key Technologies of Computer Automatic Recognition of Musical Emotion[J], Electroacoustic Technology, 2008(4):35-38.