International Journal of Frontiers in Sociology
ISSN 2706-6827 Vol. 3, Issue 1: 1-10, DOI: 10.25236/IJFS.2021.030101

Neural Network Technology in Music Emotion Recognition

Yun Liu*

College of Music, Jilin university of Art, Changchun 130021, China
**Corresponding author e-mail: 455746104@qq.com*

ABSRTACT. *Music plays an important role in human history, especially in the digital age. Now the number of music is growing exponentially, while the demand for music organization, classification and retrieval is increasing. The classification and retrieval based on music emotion is different from the traditional classification and retrieval method based on music text. It pays more attention to the emotional expression of creators and the unique characteristics of music in psychology. It is also an indispensable personalized demand for users. Therefore, more and more attention has been paid. This paper analyzes the current research situation of music emotion recognition at home and abroad, and summarizes the existing emotional models, data sets, music features, machine learning algorithms, system frameworks in the study of music emotion recognition. According to these characteristics, we choose to use machine learning to recognize music emotion. According to the experiment, the accuracy of the SVM algorithm is 88, the recognition rate of happy emotion is 90, and the recognition rate of anger is 75.*

KEYWORDS: *Music Emotion, Emotion Recognition, Emotion Model, Neural Network Technology*

1. Introduction

With the arrival of the big data era, the explosive growth of music resources and people are often under the dual influence of great pressure life, emotional retrieval of music has become a way for people to choose music resources quickly and effectively. The key to music emotion retrieval is music emotion recognition, that is, to predict music emotion. Human ability to appreciate music is innate. Concerts of different emotions make people have different emotions and further affect people's [1] of life.

Nowadays, people are in a fast moving living environment, music plays an important role in people's lives, not only can regulate people's mood, enrich people's lives, more and more people also begin to use music to relieve stress, adjust life and so on. At the same time, people's demand for music is no longer just entertainment,

ISSN 2706-6827 Vol. 3, Issue 1: 1-10, DOI: 10.25236/IJFS.2021.030101

more emotional resonance [2]. Along with the rapid development of modern computer technology and artificial intelligence technology, emotional computing (AffectiveComputing) has become a research point with broad development prospects in academic research circles. In the modern high-tech era, the computer has been fully integrated into people's lives, played an irreplaceable role. In order to maximize the role of computers in people's lives, researchers are committed to enabling computers to interact like people. Emotional computing is to enable the computer to observe, understand and produce all kinds of emotions like people, so that the computer can carry out natural and vivid interactive [3]. emotion recognition is the basis of emotion expression, emotional understanding, and emotional communication. with the development of big data, the number of digital music increases rapidly. how to quickly and accurately choose the desired music becomes a difficult problem. emotion classification of music becomes a popular trend in emotion research. The study of establishing a model to detect music emotion is called music emotion recognition. In recent years, the study of music emotion in the world a strong wind, music emotion recognition has a deep impact on people's lives. Nowadays, the retrieval of music equipment such as music player is mostly classified according to region, singing language, playing musical instrument, age and so on. It is an effective and practical method to classify and retrieve music classify music emotion [4] .

According to the research points of MER, i. e. music emotion representation model, music training database, music feature extraction and pattern recognition algorithm, this paper summarizes the development of this research. The music emotion of discrete model is realized by SVM, and finally the problems of MER are pointed out.

2. Related Concepts

2.1 Convolutional Neural Networks

After the sliding segmentation window splits a sample from the inertial data stream, the sample is directly input into the trained convolutional neural network. after a series of multiplication and addition operations and nonlinear functions, the recognition results are output. Here is no artificial feature extraction project, so researchers call this method end-to-end method [5].

In general, the previous layer or layer is a convolution layer, followed by a cascade of one or more full connection layers. Each convolution layer and full connection layer are followed by activation functions for nonlinear transformation. A pool layer is inserted after one or more convolution layers to reduce the amount of data in the intermediate results. LeNet input is a 32/32/3 picture. After the first layer of convolution layer and activation function, a 28/28/6 intermediate result is obtained. The researchers call this kind of intermediate result a feature map. The individual data in the feature map is called eigenvalue or activation value. after passing through the next convolutional layer, activation function, and pooling layer,

ISSN 2706-6827 Vol. 3, Issue 1: 1-10, DOI: 10.25236/IJFS.2021.030101

the shape of the feature map becomes 5*5*16. It should be noted here that the feature map of :5*5*16 will be straightened into a 400-dimensional vector and used as input to the first full connection layer. after passing through the following multilayer fully connected layers and activation functions, LeNet outputs a 10-dimensional vector. Above process is a brief introduction to the flow process of LeNet feature graph, which basically covers the basic concept of current neural network. Below will briefly introduce convolution layer, pool layer, full connection layer, activation function specific content [6, 7].

2.2 Musical Emotional Expression Model

First appeared, the most widely influenced model in the discrete model of musical emotion belongs to the study of 1936 Hevner. By analyzing and studying the artistic expression forms such as music, the Hevner defines the emotional attribute accurately, and uses the ring structure to express emotion. The ring structure is divided into eight sectors with typical connotations, each of which contains similar but different subclasses. By using this model, the bipolar (such as happiness / sadness) of musical emotion is highlighted, and the possible way of them is expressed in space across circles and the multi-level and multi-label nature of music emotion classification.

The Hevner model is widely used in the research of music field, and some researchers improve the Hevner emotion model. FarnsWorth recombine it into ten adjectives, and then Schubert recombine it into nine.

A discrete emotion model is based on Ekman emotion classification theory. Ekman emotional models are centered on basic or general emotions that are considered to have typical facial expressions and emotion-specific physiological characteristics. such as happiness, sadness, fear, disgust, anger, surprise, etc. Basic emotions can be found in all cultures, and they are usually associated with different physiological changes or emotional expression patterns. However, the concept of basic emotions has received many criticisms, the most obvious reason is that different researchers put forward different basic emotional [8].

Hevner and improvements to Hevner models, such as UHM9; expression of different basic emotions, such as 5 BE、4BE; improvements to AV models, such as AV4Q、AV11C、AV4Q-UHM9., to be discussed in the next section One of the advantages of the discrete dimension model is that it can be used directly in music information retrieval. The disadvantage is that it can not express the emotion contained in music properly, and it is a kind of model of expressing emotion with coarse granularity.

ISSN 2706-6827 Vol. 3, Issue 1: 1-10, DOI: 10.25236/IJFS.2021.030101

2.3 Music Training Data

Machine learning is a data-based algorithm, and the quality of training data determines the classifier. hence, the quality of training data sets is crucial in MER research.

For the establishment of training data set for MER research, there are two main aspects: music segment selection and emotion tagging. For the selection of music fragments, first of all, the number of music in the data set is more, and the types of music should be very rich (music style, musical instruments, languages, etc). Then the length of the music segment, generally select 30 s (the middle 30 s, chorus 30 s, etc.), for classical music only a few seconds.

For emotional tagging, first of all, we adopt the standard emotional model, which requires as many people as possible for each music segment, and the emotional annotator should consider the mixed selection with and without musical basis [9] .

However, the selection of music training data set should conform to its application in special application scenarios. For example, music emotion recognition with accompaniment and music emotion recognition with singing should be different in the selection of music fragments.

The difficulty of establishing training set includes two aspects: music segment selection and emotion tagging.

For the selection of music fragments to ensure the diversity of music samples, we should consider that the selected music samples cover as many kinds of emotion, language, singer, style and other factors as possible, and require that each sample can maintain a certain number, so as to reflect the characteristics of this kind of music.

The marking of music emotion is a difficult process. Because everyone has different understanding of music, different people have different feelings for the same concert. For that kind of music itself emotional expression is not very clear songs, people's experience will be different. At present, the commonly used methods can use special development tools to complete the whole tagging process, can also outsource research institutions or enterprises in related fields, or buy existing mature and abundant databases.

2.4 Foundation for Music

Music theory is the formation rule of music and the basis of MER. Also used in MER notes, keys, chords and other musical knowledge. The following is a brief introduction and of the common sense of music required by the MER [10].

Notes (note): symbols that record different sounds. According to the physical properties of sound, there are four properties of sound, namely, height (pitch), length (interval), intensity) and timbre. The note is represented by Arabic numerals, which are recorded as :1,2,3,4,5,6,7. According to the level of sound, the sound is divided

ISSN 2706-6827 Vol. 3, Issue 1: 1-10, DOI: 10.25236/IJFS.2021.030101

into seven levels, and its phonetic names are C、D、E、F、G、A、B. respectively

Height (Pitch): the degree of sound. Different notes have different pitch.

Sound intensity (Intensity): the size and intensity of a sound, also known as volume. The change of sound intensity in music works can cause the difference of playing intensity and is the musical characteristic that can affect music emotion.

Twelve mean law: because there is an audio difference between adjacent sounds, and it is not equal, the octave sound is divided into twelve equal parts according to the frequency and so on. Each first order is called that is, the second degree, and the second degree is two equal parts.

$$\frac{f_{i+1}}{f_i} = \frac{440\sqrt[12]{2^{i+1}}}{440\sqrt[12]{2^{i}}} = \sqrt[12]{2} \approx 1.0594630 \tag{1}$$

2.5 Correlation formula

(1)Spectrum

In signal processing, a signal can be regarded as a combination of frequencies. For audio signals, the spectrum can be obtained by discrete Fourier transform. The transformation formula

$$X_k = \left| \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \right|, k = 0,..., N/2 \tag{2}$$

(2)Spectrum centroid

The centroid of the spectrum is the midpoint of the energy distribution of the spectrum. Its formula is as follows:

$$SC_t = \frac{\sum_{n=1}^{N} P_t[n] * n}{\sum_{n=1}^{N} P_t[n]} \tag{3}$$

(3)Spectrum Rolling Point

The frequency at which a fraction k (0<k<1) below this frequency resides is often taken as 0.85 or 0.95 in practice. The formula is as follows:

$$\sum_{n=1}^{SR_t} P_t[n] = k * \sum_{n=1}^{N_t} P_t[n] \tag{4}$$

ISSN 2706-6827 Vol. 3, Issue 1: 1-10, DOI: 10.25236/IJFS.2021.030101

3. Application Design of Neural Network Technology in Music Emotion Recognition

3.1 Data sets and Emotional Representations

Data sets of this experiment are derived from the aforementioned AMG1608 data sets. The AMG1608 also uses the AV dimension model, each music segment only marks a pair of AV values (a pair of values between -1 and 1), just like the Cartesian coordinate system, AMG1608 expresses emotion as a point in the AV emotional space.

3.2 Feature Vectors

Based on AMG1608 public data set, A 72-dimensional feature vector s the first 50 frames of its 30 fragments is extracted. The eigenvalues include 20 Mel frequency cepstrum coefficients and their first derivative ,17 octave band signal intensities using a triangular octave filter bank and the ratio of these intensity values ,2 linear predictor coefficients capturing the spectral envelope of the audio signal ,9 spectral fluxes and spectral shape descriptors, and 4 shape statistics including the centroid of the music signal (centroid), Extension (spread), skewness (skewness) and kurtosis (kurtosis). This chapter uses different feature representations from Chapter 3. In chapter three, First, the rhythm, beat, chromatogram, constant Q chromatogram, normalized chromatogram, Mel plot, Mel cepstrum coefficient, rms, spectral centroid, spectral bandwidth, spectral contrast, spectral roll-down point polynomial coefficient, tonal centroid feature, zero crossing rate, harmonic component, percussion element, etc. are calculated by frame for 1 minute music, Then calculate its mean, standard deviation, variance. resulting in a 54-dimensional feature vector for each piece of music. In this chapter, the music fragment is processed according to the frame to obtain the 72 Witt eigenvalue, and then the 72 Witt eigenvalue of each frame is arranged in chronological order, and the first 50 frames are taken to form 3600 dimensional eigenvectors for experiment. The time information is preserved.

4. Experimental Results and Analysis

4.1 KNN Regression Experiments

Table 1. *KNN Regression Results*

Adjacent numbers	Number of parallel operations	Arouse mean error variance	Valence mean error variance
2	2	0.1323651006999807	0.10377276195336392
5	2	0.1054903134905482	0.07848791335418484
10	2	0.0942929031600554	0.07145362439382772
14	2	0.09366282503988543	0.07113340084690666

ISSN 2706-6827 Vol. 3, Issue 1: 1-10, DOI: 10.25236/IJFS.2021.030101

100	2	0.10058008336867362	0.07729751898632799
200	2	0.10269908771058184	0.07665647166316514

When the number of training sets is 900 and the number of tests is 100, the experimental results are shown in Table 1. Since this experiment is based on the affective space model of AV dimension MER, the judgment of prediction accuracy is obviously different from that of Chapter 3. Because the emotion prediction of dimension space is regarded as regression model and the predicted value is determined, the accuracy of the prediction is based on the mean error variance, that is, the mean square error.

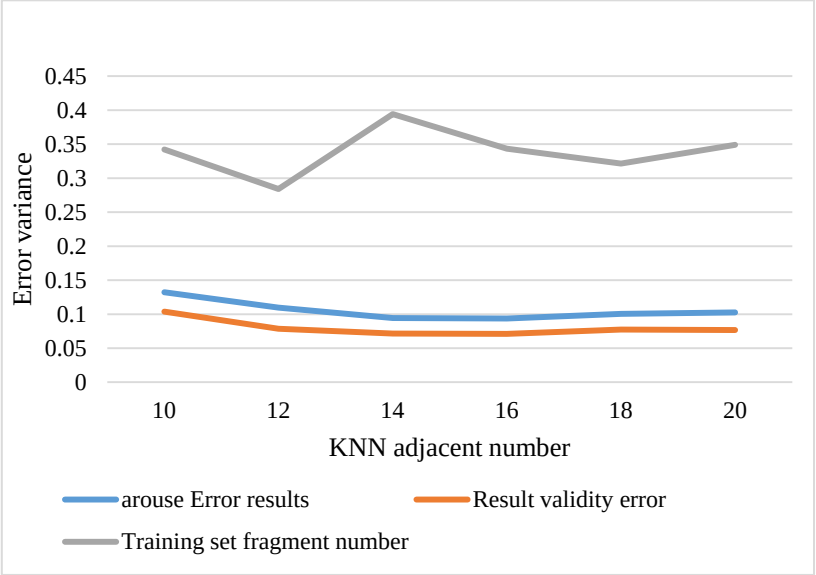

Figure 1. *Valence and Arouse Error Results and Classification Accuracy of KNN Algorithms*

KNN regression is obtained in the experiment, the performance reaches the best when the proximity number increases to a certain value. When exceeded, the performance is slightly reduced, but if the value of the nearest neighbor number is too large, there will be excessive aggregation, which will not belong to the class.

We can see from the figure that when k=4, the KNN algorithm can obtain the best classification accuracy, that is, when the k is the same as the number of classification, the prediction results are the best. Too much or too little is not good. When k=4 KNN however, the optimal accuracy of the algorithm is only 40.

ISSN 2706-6827 Vol. 3, Issue 1: 1-10, DOI: 10.25236/IJFS.2021.030101

4.2 Music Emotion Recognition of Discrete Model SVM Algorithm

Table 2. *Experimental Results*

kernel function	Number of training set segments	Number of test set segments	accuracy rate
polynomial kernel	35	30	46.7%
polynomial kernel	60	15	66.7%
polynomial kernel	65	8	87.5%
Linear kernel	65	8	87.5%
Sigmoid nuclear	65	8	25%
Radial basal nucleus	65	8	25%

It can be seen from the table that the accuracy of algorithm classification is increasing with the increase of the number of training sets. However, in the case of the same number of segment number, the selection of kernel function has a great influence on the classification accuracy of the algorithm, in which the linear kernel function of polynomial kernel function has a better recognition effect. The recognition effect of Sigmoid kernel and radial kernel is poor.

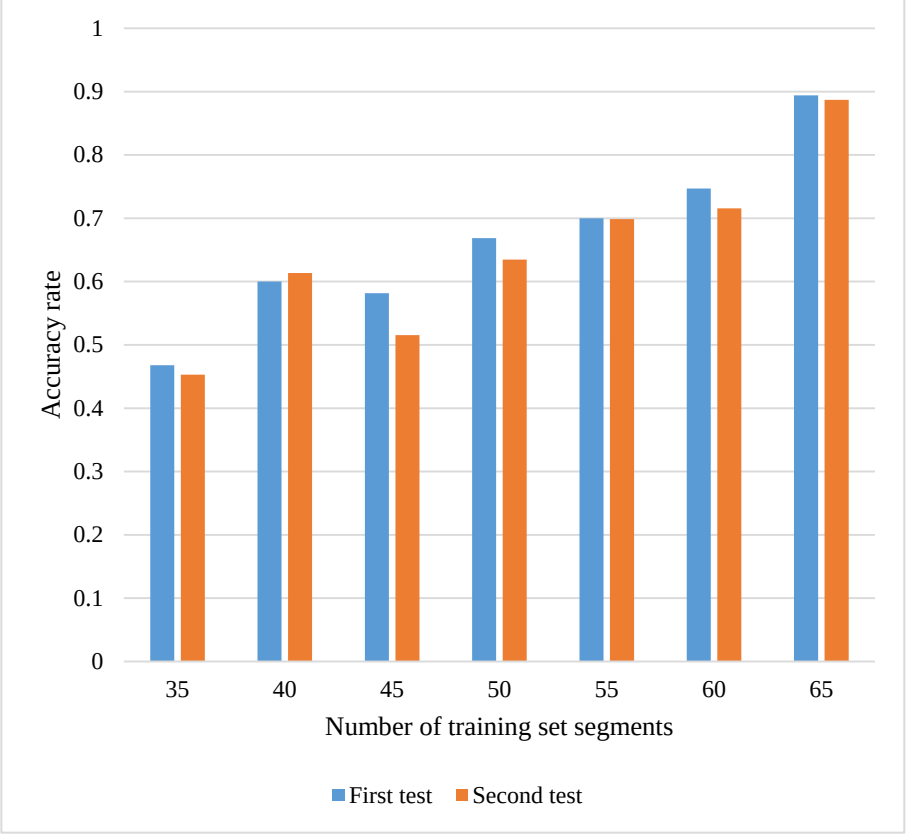

Figure 2. *SVM Algorithm Classification Accuracy*

ISSN 2706-6827 Vol. 3, Issue 1: 1-10, DOI: 10.25236/IJFS.2021.030101

The experimental results show that the classification accuracy of the SVM algorithm is increasing with the increase of the number of training sets from 30 to 65. This shows that the SVM algorithm needs a certain number of samples to achieve the best results, and that a sufficient number of training sets is of great significance to the accuracy of the SVM algorithm.

Table 3. *Test Set Recognition Results for 60 Training Fragments*

Test pieces	Result			
	anger	happy	sad	relax
anger	3	1	0	0
happy	1	4	0	0
sad	0	0	2	1
relax	1	0	1	1

It can be seen from the table that the recognition rate of relaxation and sadness is low, and the recognition rate of relaxation is the lowest, and the recognition rate of joy and anger is higher. The possible reason is that sadness and relaxation can not be clearly expressed through the characteristics of songs, and do not have strong emotional characteristics in the process of machine learning machine learning joy and anger are relatively easy to express emotions, with stronger emotional characteristics, And these emotions in different forms of songs have similar performance, so it is easier to distinguish.

5. Conclusion

This paper makes some research on the related aspects of music emotion recognition, mainly studies the following contents: the expression model of music emotion is an important part of music emotion recognition research. The discrete model and dimension emotion are analyzed, and the difference and relation between them are analyzed. The training set consists of two aspects: music segment selection and emotion tagging. The method of music emotion recognition based on K proximity regression: the KNN algorithm is studied, and the music emotion classifier based on KNN algorithm is realized, and the selection of adjacent numbers in the algorithm is studied emphatically. The method of music emotion recognition based on support vector machine: SVM algorithm is studied, and the music emotion classifier based on SVM algorithm is implemented. The selection of kernel function and the influence of sample number of training set in SVM algorithm are studied. The performance of the KNN algorithm is compared.

References

[1] Thammasan N , Moriyama K , Fukui K I , et al. Continuous Music-Emotion Recognition Based on Electroencephalogram[J]. Ice Transactions on Information & Systems, 2016, 99(4):1234-1241.

ISSN 2706-6827 Vol. 3, Issue 1: 1-10, DOI: 10.25236/IJFS.2021.030101

[2] Wang J C , Lee Y S , Chin Y H , et al. Hierarchical Dirichlet Process Mixture Model for Music Emotion Recognition[J]. IEEE Transactions on Affective Computing, 2015, 6(3):261-271.
[3] Goshvarpour A , Abbasi A , Goshvarpour A , et al. A novel signal-based fusion approach for accurate music emotion recognition[J]. Biomedical Engineering Applications Basis & Communications, 2016, 28(06):485-1026.
[4] Nalini N J , Palanivel S . Music emotion recognition: The combined evidence of MFCC and residual phase[J]. Egyptian Informatics Journal, 2016, 17(1):1-10.
[5] Panda R , Malheiro R M , Paiva R P . Audio Features for Music Emotion Recognition: a Survey[J]. IEEE Transactions on Affective Computing, 2020, PP(99):1-1.
[6] Dong Y , Yang X , Zhao X , et al. Bidirectional Convolutional Recurrent Sparse Network (BCRSN): An Efficient Model for Music Emotion Recognition[J]. IEEE Transactions on Multimedia, 2019, 21(12):3150-3163.
[7] Sutcliffe R , Rendell P G , Henry J D , et al. Music to my ears: Age-related decline in musical and facial emotion recognition[J]. Psychology and Aging, 2018, 32(8):698-709.
[8] Fukayama S , Goto M . Adaptive aggregation of regression models for music emotion recognition[J]. The Journal of the Acoustical Society of America, 2016, 140(4):3091-3091.
[9] Panda R , Rocha B , Paiva R P . Music Emotion Recognition with Standard and Melodic Audio Features[J]. Applied Artificial Intelligence, 2015, 29(4-6):313-334.
[10] Chin Y H , Hsieh Y Z , Su M C , et al. Music emotion recognition using PSO-based fuzzy hyper-rectangular composite neural networks[J]. IET Signal Processing, 2017, 11(7):884-891.