Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 9256586, 11 pages
https://doi.org/10.1155/2022/9256586

Research Article

Study on Music Emotion Recognition Based on the Machine Learning Model Clustering Algorithm

Yu Xia[1] **and Fumei Xu**[2]

[1]*School of Aviation Services and Music, Nanchang Hangkong University, Nanchang 330063, Jiangxi, China*
[2]*School of Music, Jiangxi Normal University, Nanchang 330067, Jiangxi, China*

Correspondence should be addressed to Yu Xia; xylxm2022@njust.edu.cn

Received 23 August 2022; Revised 13 September 2022; Accepted 19 September 2022; Published 11 October 2022

Academic Editor: Baiyuan Ding

In recent years, the explosive growth of online music resources makes it difficult to retrieve and manage music information. To efficiently retrieve and classify music information has become a hot research topic. Thayer's two-dimensional emotion plane is selected as the basis for establishing the music emotion database. Music is divided into five categories, the concept of continuous emotion perception is introduced, and music emotion is regarded as a point on a two-dimensional emotional plane, together with the two sentiment variables to determine its location. The artificial labeling method is used to determine the position range of the five types of emotions on the emotional plane, and the regression method is used to obtain the relationship between the VA value and the music features so that the music emotion classification problem is transformed into a regression problem. A regression-based music emotion classification system is designed and implemented, which mainly includes a training part and a testing part. In the training part, three algorithms, namely, polynomial regression, support vector regression, and k-plane piecewise regression, are used to obtain the regression model. In the test part, the input music data is regressed and predicted to obtain its VA value and then classified, and the system performance is considered by classification accuracy. Results show that the combined method of support vector regression and k-plane piecewise regression improves the accuracy by 3 to 4 percentage points compared to using one algorithm alone; compared with the traditional classification method based on a support vector machine, the accuracy improves by 6 percentage points. Music emotion is classified by algorithms such as support vector machine classification, K-neighborhood classification, fuzzy neural network classification, fuzzy K-neighborhood classification, Bayesian classification, and Fisher linear discrimination, among which the support vector machine, fuzzy K-neighborhood, and the accuracy rate of music emotion classification realized by Fisher linear discriminant algorithm are more than 80%; a new algorithm "mixed classifier" is proposed, and the music emotion recognition rate based on this algorithm reaches 84.9%.

1. Introduction

Music is an artistic experience and an entertainment method that expresses people's thoughts and emotions and reflects real life by using music as the medium and carrier of expression [1]. So far, the main way people perceive music is still through hearing, but the expression of music and the transmission of emotional information is not limited to acoustic situations [2]. In modern society, people sing, dance, and experience music with the help of high-tech sound, light, electricity, etc., to achieve wonderful audio-visual effects and emotional interaction, such as music evenings and music fountains. It can be seen that from ancient times to the present, these behaviors reflect people's desire to interact synchronously through multisensing modes such as hearing, vision, and touch, perceive the pitch, loudness, duration, and timbre of musical sounds, and experience the rhythm, melody, harmony, and timbre of music [3]. Computers cannot "hear" music like humans but they can extract content features such as spectrum, rhythm, and mel-frequency cepstrum coefficients (MFCC) by performing audio processing on music. Classification and retrieval of music are known as content-based music retrieval technology. However, due to the complex and unknown

mechanism of human perception of music and the generation mechanism of emotion, pure machine classification makes subjective music emotion classification more difficult because of its objectivity, and emotion classification has not achieved the same accuracy as style classification. Exactly which features can express musical emotion more accurately, how to automatically analyze the extracted features, and which classification method can obtain higher accuracy are all issues worthy of study.

At present, with the improvement of people's material living standards year-by-year, people have put forward higher requirements for spiritual life. The pursuit of material civilization and spiritual civilization is the internal driving force of social progress. The way of information dissemination is moving from the era of "multimedia" to the era of "all media" [4]. It is the multimedia that disseminates information with images but pursues the experience of "all-media" information based on "audio-visual-tactile" multisensing channels. People with normal hearing function hope to experience music synchronously and interactively through the "listening-visual-tactile" multisensing channels, which can further increase the immersion of perceiving music [5]. People with hearing disabilities cannot "hear" music but they also desire to experience the emotion of music, receive music education, and perform music. The results of modern psychological research have shown that music perception is a cognitive activity coordinated by multiple sensory systems and is not limited to hearing. With appropriate stimulation, vision and touch can also perceive music through synesthesia [6]. In 2007, the research results of Edward et al. showed that the parts of the brain used by hearing-impaired people to process tactile information are the same as those used by normal people to process auditory information, which means that hearing-impaired people can use the tactile sensation of body skin to perceive music and can experience music like a normal person. These research results have laid a physiological and psychological foundation for the related research on "tactile auxiliary or alternative auditory perception of music" [1]. At present, a simple music player dedicated to the deaf has appeared on the market, which converts the rhythm of the music into vibration and emits light through LEDs. The color and brightness of the light change with the rhythm, and you can experience music through vibration or light changes. These music players do not yet meet the needs of hearing-impaired groups.

Technology is based on two or three perceptions of hearing, vision, and touch to experience music synchronously and interactively, involving the intersection and integration of various disciplines and technologies, including musicology, aesthetics, art, cognitive psychology, psychophysics, mechanical engineering, multimedia technology, signal processing, pattern recognition, intelligent control, virtual reality, instrument science, and other related disciplines [7]. At present, the related technologies of multisensory interactive music experience based on "listening-visual-tactile" have become a research hotspot at home and abroad. Related research has been carried out [8].

Music is the language and art of emotion, and emotion is the essential feature of music. Modern psychological research shows that the nonsemantic organizational structure of music vibrating with sound waves has a direct isomorphic relationship with human emotions and will activities [9]. Emotional experience is a series of emotional reactions caused by appreciating works, and the expression and perception of emotions are increasingly important as a means of human-human interaction and human-computer interaction [10]. Scholars at home and abroad have proposed a variety of emotion modeling methods. One type of view holds that emotion is composed of discrete basic emotions, the most representative being the OCC (Ortony, Clore, Collins) emotion model established by Ortony et al. and the Hevner emotion ring model proposed by Hevner et al. Behaviors, views on things, etc., define 22 basic emotions. The Hevner model is divided into 8 categories and uses 67 adjectives to describe musical emotions [11]. The other is the emotional model based on dimensional space theory, which considers that emotions are distributed in a certain space composed of several dimensions, a specific emotion can be mapped to a specific position in a continuous space, and the similarities and differences between different emotions can be measured according to the distance from each other in the dimensional space; different emotions are not independent, but continuous, which can achieve smooth conversion [12]. The most representative dimensional emotional model is the three-dimensional emotional model proposed by Wundt in 1907, another three-dimensional emotional model established by Plutchik in 1980, and the two-dimensional annular emotional model proposed by Russell in 1980 [13]. Russell's two-dimensional ring emotion model divides emotion into two dimensions, namely, pleasure and arousal. Pleasure is divided into positive and negative poles, and motivation is divided into low intensity and high intensity. In 1989, Thayer proposed a two-dimensional energy-stress model based on the Russell emotional model. The energy dimension is consistent with the arousal in the Russell model; the stress dimension represents valence [14]. Valence is a physiological or psychological pleasure response to external stimuli, with positive and negative directions. It can be seen that both the Thayer effective model and the Russell affective model reflect two aspects of valence and motivation and can map emotion to the "valence-incentive" emotional plane, so it can also be called "valence-incentive" [15]. The third type is the emotional model based on cognitive mechanisms, such as the EM emotional model, the Roseman emotional model, the EMA emotional model, and the salt and pepper emotional model [16]. The Hevner emotion ring model, the Russell emotion model, and the Thayer emotion model are widely used in music emotion recognition; based on the Hevner emotion ring model [17], the music emotion can be classified and the music emotion type corresponds to the adjective in the Hevner emotion ring model; based on the Russell or Thayer two-dimensional emotional model of music, the specific emotion of music can be mapped as a point in the emotional plane of "valence-incentive" (the horizontal axis corresponds to the effect value, the vertical axis corresponds to the incentive value), and the regression and classification of music emotions can be achieved by applying machine learning methods [18].

Due to differences in cultural background, age, gender, personality, and musical preferences, as well as the influence of external factors such as the perceived environment, people's descriptions of musical emotions to the same song may vary from person to person, and the adjectives that describe musical emotions themselves have a considerable degree of ambiguity [19]. Due to the subjectivity and ambiguity of music emotion, it is a very challenging task to accurately identify music emotion [20]. Another difficulty in identifying music emotion based on audio signal features is that there is a semantic level gap between music features and emotional cognition, so it is very difficult to accurately identify music emotion. The psychological process of music cognition can be described in four levels as follows: the physical layer, perceptual layer, musical layer, and semantic layer [21]. After comprehensively perceiving or judging all the characteristics of the physical layer, perceptual layer, music layer, and semantic layer and then identifying the emotional cognition of music through reasoning and thinking, it will be affected by people's cultural background, age, gender, personality, music preferences, and perceived environmental factors [22]. Since 2000, music emotion recognition has become a hot research topic at home and abroad. From MIDI format symbol music to audio format music emotion recognition, from western classical music to modern pop music, researchers have done a lot of work. The methods of music emotion recognition mainly include emotion classification and emotion regression [23].

Based on the Hevner emotional ring model and the Russell two-dimensional emotional model, the music emotion is classified. The more the classification, the lower the recognition rate of the music emotion. The researchers applied the machine learning algorithm to classify the music emotion into cheerful, angry, sad, and calm. Based on the method of audio signal processing, extract the energy, melody, harmony, time domain, frequency domain, and other dimensional features of music, through machine learning methods, including the support vector machine, the Gaussian mixture model, the neural network, and the K-nearest neighbor algorithm categorize the emotion of music [24]. Another way of music emotion recognition is to regress music emotion based on the Russell or Thayer emotion model through the regression method; the emotion of music corresponds to a point in the "valence-motivation" emotion plane [25]. Emotion, which can accurately calculate the effective value and incentive value corresponding to the music emotion, overcome the shortcomings of the classification method that is not precise enough to identify the music emotion and can track the changes of the music emotion in the "valence-incentive" emotional plane [26].

In our study, extract the energy, melody, time domain, frequency domain, and harmony of the five dimensions of perceptual music, and use the machine learning-based method to treat music emotion recognition as a regression problem. To realize the recognition of music emotion, a new classification algorithm is proposed to improve the recognition rate of music emotion. The mapping relationship of the emotional attributes of music in the emotional plane of "valence-motivation" is discussed, and the schemes of music emotional regression and classification are proposed, respectively; the characteristics of music energy, beat, time domain, frequency domain, and harmony are extracted. The method of machine learning realizes the regression and classification of music emotion and proposes a new music emotion classification method "Hybrid Classifier," which improves the emotion recognition rate, gives the experimental results of emotion recognition for each regression and classification method, and gives the results of emotion recognition. The experimental results are discussed and analyzed.

First, the mapping relationship of music emotional attributes in the emotional plane of "valence-motivation" is discussed, and the schemes of music emotional regression and classification are proposed, respectively. A variety of machine learning methods are used to achieve music emotion regression and classification, and a new music emotion classification method "Hybrid Classifier" is proposed to improve the emotion recognition rate. Finally, the emotion of each regression and classification method is given. The experimental results are identified, and the experimental results are discussed and analyzed.

2. Data Source and the Method

This paper selects the music emotion database from MediaEvaP4ˆas the material for music emotion recognition research. MediaEval is an organization dedicated to providing test standards for the evaluation of new algorithms for multimedia access and retrieval. Members of the organization research topics such as speech recognition, multimedia content analysis, music, and audio analysis, user information analysis, and audience emotional responses. The music emotion database mainly contains 1000 English music files in mp3 format downloaded from Free Music Archive (http://freemusicarchive.org/, FMA) [27]. The files are numbered from 1 to 1000. However, due to some redundancy in the initial collection process, a list of files to filter out redundancy is provided in the database annotation file. After filtering out the redundancy, there are 744 music files in the actual database, among which 619 and 125 music files are marked for training and testing of the music emotion regression model, respectively [1]. The length of each music file in the database is 45s, and the sampling rate is 44100 Hz. The annotation file contains information such as the song name of each song and also contains the static and dynamic excitation and valence values of 2 Hz in the range $[-1, 1]$ given by 10 participants, and standard deviation information is based on MediaEva music emotion database; the paper carried out the music emotion recognition research according to the framework of Figures 3 and 4 [28].

Computer-based music emotion recognition mainly refers to the use of modern signal processing technology to achieve music feature extraction and machine learning methods to achieve music emotion regression or classification [29]. Commonly used machine learning methods include supervised, semisupervised, and unsupervised machine learning methods [30]. At present, semisupervised and unsupervised methods still have the disadvantage of

unsatisfactory recognition effect. Most of the studies reported in the literature use supervised machine learning methods to realize music emotion recognition [31]. Therefore, the discussion on computer music emotion recognition in this study is based on the supervision of machine learning. A typical computer music emotion recognition model framework is shown in Figure 1. The model framework mainly includes two parts, namely, model training and unknown type of music emotion prediction. In the model training, the music signal of known music emotion is converted into one-to-one correspondence features through signal preprocessing and feature extraction; c and emotion labels [32]. The training data set of $v(x) > 0$; the machine learning algorithm trains the classification or regression model with the minimum classification error or minimum mean square error as the objective function for the input training data set. In the stage of emotional prediction of an unknown type of music, it mainly predicts the emotional attributes of music; after the music signal of unknown emotional type undergoes signal preprocessing and feature extraction similar to that in model training, a test data set JC containing only music feature vectors is generated, and then input x into the regression or classification model generated in the model training phase to achieve the prediction output of music emotion. Analysis of the computer music emotion recognition model framework shows that the music emotion attribute of the training data needs to be predicted in the model training, so the music emotion subjective scoring method is often used in the preparation of the training data set [33]. At the same time, due to the individual uniqueness and subjectivity of music emotion, the "training data set" obtained by the subjective evaluation method is used for the training of the music emotion recognition model, which can effectively generate a music recognition system with individual preferences.

Music signal preprocessing and feature extraction are the first steps towards realizing music emotion recognition [34]. This step mainly includes music signal framing, signal windowing function processing after framing, feature calculation in each sliding window, and original feature space projection or feature selection for dimensionality reduction [35]. The music signal is a continuous time-series nonstationary signal, so the signal needs to be framed. Since the dynamic annotation of the MediaEval music emotion database gives the mean and standard deviation of V and A of 2 Hz, that is, an annotation is given every 0.5 s from the beginning of the music file 15 s, the sliding length of each frame is 0.5 s, 50% window length overlapping rectangular windows to frame each music file in the database [36]. Feature extraction and statistics are performed on the music signals in each frame, respectively, and a local feature data set corresponding to the dynamic V and A annotations in the database can be obtained. Further statistical processing of 60 frames of data in each file can obtain the global feature data set corresponding to the static V and A annotations of the entire music file.

Windowing the music signal is to do a dot product directly with the window function $W_M(n)$ on the music time series as follows:

$$W_M(n) = \begin{cases} W(n), & 1 \le n \le M, \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Commonly used window functions include rectangular window, triangular window, Hanning window, Hamming window, and Gaussian window. The Gibbs effect and spectral leakage will appear when the rectangular window is used to truncate the signal for spectral analysis. The rest of the window functions can be better improved in the spectral analysis. The paper adopts the Hanning window function, as shown in equation (2), to realize the window processing of the music short-time frame signal.

$$\omega(n) = 0.5\left[1 - \cos\left(\frac{2\pi n}{(M+1)}\right)\right]. \tag{2}$$

The Hanning window of 1024 points (~22 ms) is superimposed on the signal, respectively, and the windowed signal can be further processed, such as short-time Fourier spectrum, and Mel frequency cepstral coefficient (MFCC) calculation.

After the feature extraction is performed on the music signal in this paper, the total dimension of the obtained feature data set (all feature spaces) is 548 dimensions.

On one hand, it is difficult to find an accurate regression model or classification plane to identify music emotion in high-dimensional space, and on the other hand, it will greatly increase the computational complexity of the algorithm. Therefore, this study also discusses the spatial projection based on principal component analysis and the relief-based algorithm. Feature selection features the dimensionality reduction method.

Principal component analysis (PCA) is an effective method for processing, compressing, and extracting information in samples based on a variable covariance matrix, which can effectively reduce the number of features containing noise or redundancy and is a common dimensionality reduction method. The core idea of PCA is to project the n-dimensional features into mutually orthogonal k-dimensional ($k < n$) features under the principle of preserving the maximum variance based on the assumption that the signal has a large variance and the noise has a small variance. The k-dimensional feature is also called a pivot. Let the sample feature matrix $X = \{x_1, \ldots, x_n\}$, x_i is a column vector. First, matrix B is obtained by subtracting the mean of each column by column; second, the covariance matrix C of B is computed as follows:

$$\begin{aligned} C &= E[B \otimes B] \\ &= \frac{1}{N} \sum B \cdot B^*. \end{aligned} \tag{3}$$

Third, the eigenvalues V and column eigenvectors D of matrix C are computed as follows:

$$[V, D] = eig(C). \tag{4}$$

Fourth, we sort V and D, calculate the contribution rate and cumulative contribution rate of each eigenvalue in V

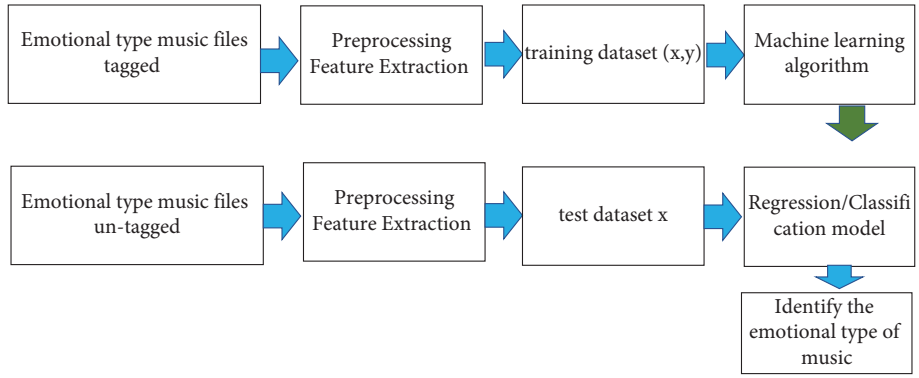

Figure 1: Computer music emotion recognition model framework.

after sorting, and select the rearranged k-column eigenvectors $W_{n\times k}$ corresponding to the eigenvalues whose cumulative contribution rate is greater than the threshold T; finally, the final PCA dimensionality reduction data is $Y = B * W$.

In this study, the transformation matrix W whose cumulative energy threshold T of the first k eigenvalues in V is greater than 90% is selected for eigenspace PCA dimension reduction. The original data set containing 548-dimensional features was dimensionally reduced using the PCA dimensionality reduction method, the threshold was set to 90% of the energy, and the final feature dimension after dimensionality reduction was 139-dimensional.

The Relief algorithm was first proposed by Kira and Rendell in 1992, and now it generally refers to a series of algorithms including Relief, ReliefF, and RReliefF. The Relief algorithm is one of the commonly used feature selection weight algorithms, which has the advantages of high operating efficiency and no restrictions on data types. The core idea of the algorithm is to assign weights related to categories to features. Based on the ability of features to distinguish close-range samples, the features with larger weights are finally selected to form a feature subset to represent the original feature set, and the classification is discarded. Small-weight features have less contribution. The Relief algorithm randomly selects a sample from the training set D and then follows the distance metric of formula (5) to find the nearest neighbor sample H from the samples of the same class and R and find the nearest neighbor sample M from the samples of different classes from R.

$$d = \frac{1}{2}\left(||R - M|| - \|R - H\|\right). \tag{5}$$

Then, we update the weight of each feature according to the rule 6 as follows: if the distance between the sum and M is less than the distance between the sum and M, it means that the feature is beneficial to distinguishing the nearest neighbors of the same and different classes and then increase the feature weight; conversely, if the distance between a feature and M is greater than the distance between M and M, indicating that the feature has a negative effect on distinguishing the nearest neighbors of the same class and different classes, the weight of the feature will be reduced. The above process is repeated m times, and finally, the average weight of each feature is obtained. The larger the weight of the feature, the stronger the classification ability of the feature, and the weaker the classification ability of the feature. The iterative formula for the weights of the algorithm process is as follows:

$$w[A] = w[A] - \frac{\operatorname{diff}(A, R_i, H)}{m} + \frac{\operatorname{diff}(A, R_i, M)}{m}, \tag{6}$$

where A is the dimension scalar of the feature, and $\operatorname{diff}(A, I_1, I_2)$ is defined as follows:

$$\operatorname{diff}(A, I_1, I_2) = \begin{cases} 1 \text{value}(A, I_1) = \text{value}(A, I_2), \\ 0, \text{otherwise}. \end{cases} \tag{7}$$

The improved Relief algorithm uses the average of the k nearest neighbors to replace the single nearest neighbor in the weight iteration, which reduces the influence of noise on the weight to a certain extent and makes the final feature selection more accurate. The weight update formula of the Relief algorithm is as follows:

$$w[A] = w[A] - \frac{\sum_{j=1}^{n} \operatorname{diff}\left(A, R_i, H_j\right)}{k \times m} + \sum_{C \neq \text{class}\left(R_i\right)} \frac{p(C)}{1 - p\left(\text{class}R_i\right)} \frac{\sum_{j=1}^{k} \operatorname{diff}\left(A, R_i, M_j(C)\right)}{k \times m}, \tag{8}$$

where $p(C)$ is the prior probability of each class estimated from the training set; $1 - p(\text{class}R_i)$ represents the sum of the probabilities of misclassification.

The RReliefF algorithm uses probability to express the rules that clearly distinguish between two classes. This probability can be estimated by establishing a model that predicts the relative distance between the two classes. The iterative formula for the weights of the RReliefF algorithm can be expressed as follows:

$$w[A] = \frac{P_{\text{diff}(C)|\text{diff}(A)} P_{\text{diff}(A)}}{P_{\text{diff}(C)}} - \frac{\left(1 - P_{\text{diff}(C)|\text{diff}(A)}\right) P_{\text{diff}(A)}}{1 - P_{\text{diff}(C)}}, \tag{9}$$

where

$$\begin{aligned} P_{\text{diff}(A)} &= P(\text{different value of A}|\text{mearest instances}), \\ P_{\text{diff}(C)} &= P(\text{different value of A}|\text{mearest instances}), \\ P_{\text{diff}(C)|\text{diff}(A)} &= P(\text{diff.prediction}|\text{diff.valuue of A andnearest instances}). \end{aligned} \tag{10}$$

In this study, the Relief feature selection algorithm is used to perform feature selection on the 548-dimensional original feature set. We select the ones with a weight greater than 0 and finally determine that the feature dimensions used for the training and prediction of the regression model of music emotional effect value (V) and incentive value (A) are 276 dimensions and 258 dimensions, respectively.

Regression is an analytical method used to predict changes in unknown dependent variables by determining the correlation between dependent variables and some independent variables, establishing regression equations, and adding extrapolation. The existing regression theories can be used to predict the V and A values of music in the emotional plane of music signal features.

Let X_i be the feature set of a certain piece of music after signal preprocessing and feature extraction and y_i be the emotional attribute of music (V or A value); the process of training an optimal regressor $r()$ is to give N inputs $(X_i, y_i), i \in \{1, 2, \ldots, N\}$, to achieve the smallest mean square error e between the predicted output and y_i as follows:

$$e = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - r\left(X_i\right)\right)^2. \tag{11}$$

Since the V and A values are real numbers in the range of $[-1, 1]$ in the emotional coordinate space, two regression models can be established according to the existing regression theory to predict the V value and the A value. In the regression model, the true values of V and A can be obtained by subjective scoring. Due to the use of the public database of MediaEval, the annotated V and A values are regarded as true values in the paper; Shi is the feature set after feature extraction, with a total of 548 dimensions. However, in order to obtain the best regression effect, specific analysis must be carried out for specific problems [37]. Therefore, in this paper, algorithms such as multivariate adaptive regression, support vector regression, and radial basis function regression are used to realize the V and A value regression of music emotion, respectively, and the optimal regression algorithm and regressor are selected by comparison.

3. Results and Discussion

3.1. Music Emotion Regression Model Training and Prediction Scheme. In music emotion recognition, regression and classification are the two main methods. Unlike classification, which only needs to distinguish emotions such as joy, anger, sadness, and calmness, the goal of music emotion regression is to identify more accurate music emotions. It regards the emotional plane as a continuous space and identifies V-A through the regression model. The emotional state is represented by each point in the emotional plane (find out the mapping between music and specific coordinate positions in the V-A emotional plane).

The training and prediction framework of the emotional regression model is shown in Figure 2. The framework uses machine learning methods to achieve regression model training and then can predict the V and A values of music emotions. Specifically, it includes training and evaluating the model through the training data under the rules of the regression algorithm and using the trained regression model parameters for the music emotion prediction of the test data; the V and A values of the predicted music emotion with the manually calibrated V and A values are compared to test the accuracy of music emotional regression.

Multivariate adaptive regression splines (MARS) is a high-dimensional data regression method with good generalization ability proposed by Friedman of Stanford University in 1991 for nonlinear problems. Its goal is to predict a continuous output variable from a large number of independent samples; support vector machine (SVM) was first proposed by Vapnik, and he further developed a series of machine learning algorithms that can realize nonlinear mapping of output feature vector to high-dimensional feature space. When SVM is used for classification problems, it is called support vector classification (SVC). When SVM is used for regression problems, it is also called support vector regression (SVR). Unlike SVC, the purpose of SVR is to find a function that can achieve the smallest deviation e from the true value y of the input training data set. At the same time, the function should be as flat as possible. Radial basis function regression (RBFR) can be regarded as a surface fitting problem in a high-dimensional space, and it is an effective regression method.

In this study, the mean absolute error (MAE), the regression value accuracy (A_c), and the sentiment classification accuracy (A_c of classification) are used as the evaluation criteria for the music sentiment classification system based on three different regression methods.

The mean absolute error is the average of the absolute values of the squares of the differences between all observations and the mean. The mean error makes the dispersion absolute value, so that the errors will not be canceled by positive and negative, so compared with the mean error, the mean absolute error can better reflect the error between the predicted value and the observed value. Value range for

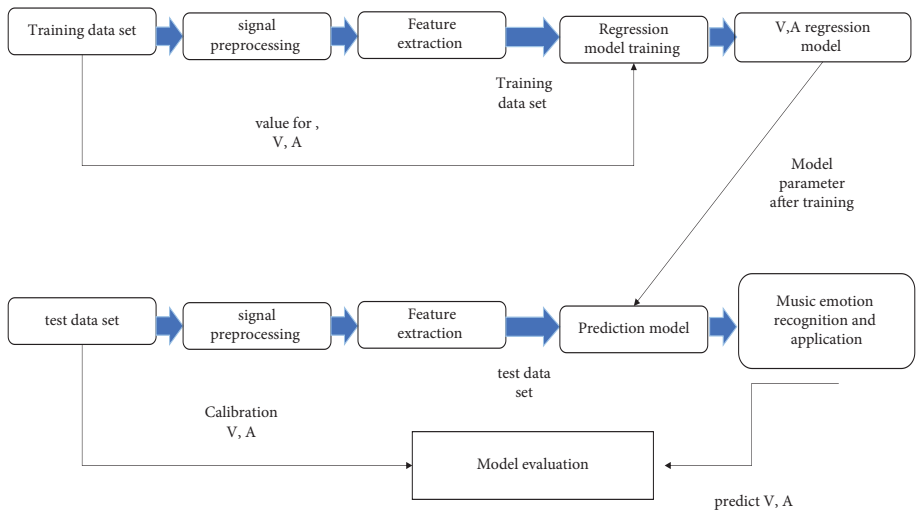

Figure 2: Music emotion regression model training and prediction framework.

accuracy of regression values (Ac of Arousal, Ac of Valence) can be divided into three (Table 1).

In order to check the regression effect of the VA value separately, the VA value is divided into three categories according to the value range of the VA value, and the accuracy of the regression value is determined by the accuracy of the category. If the predicted value is in the same range as the observed value, the classification is correct; otherwise, it is wrong.

The accuracy of sentiment classification is determined by whether the predicted value and the observed value are in the same sentiment category and is used to observe the classification effect of the entire classification system. The experimental evaluation results are shown in Table 2.

Figure 3 shows the distribution of the manually labeled values and predicted values of RBFR regression V and A. It can be seen that the predicted values have obvious regionality.

From Table 2, it can be concluded that the effectiveness of nonlinear regression (support vector regression and RBFR) is significantly improved compared with linear regression (polynomial regression), which indicates that there is an obvious nonlinear relationship between music feature vectors and emotional variables. Compared with RBFR, support vector regression has higher prediction accuracy for the arousal value, and the latter has higher prediction accuracy for the valence value, which may be related to the different decision relationship between different sentiment variables and eigenvectors.

Based on this, since the regression models of the VA values are obtained separately, there is no correlation between the two. In this study, support vector regression and RBFR are combined in the subsequent experiments, and support vector regression is used to obtain the arousal regression model, respectively. RBFR obtains the valence regression model and then observes the classification accuracy of the music emotion classification system, which has a certain improvement compared with the two methods alone, as shown in Table 3.

Table 1: Arousal, valence value range.

	Valence		Arousal
Cluster 1	[−0.6,0.6]	Cluster 1	[0.4,1]
Cluster 2,4	[0.2,1]	Cluster 2,5	[0,0.6]
Cluster 3,5	[-1,-0.2]	Cluster 3,4	[−0.6,0]

3.2. Comparison of the Results of the Regression Method and the Pattern Recognition Method. After the comparison and analysis of the efficiency of the three regression models, a regression-based music emotion classification system is finally formed by the combination of support vector regression and RBFR regression. In order to verify the effectiveness of the system, the text is also classified on the same music database using the support vector machine method to classify the music emotion.

When SVM is used for classification, the method of "one pair and the rest" is adopted, and 5 classifiers need to be trained. Each classifier distinguishes the current training category from other categories. When testing, the probability that the input test data belongs to each category is calculated and its maximum probability as the category of the data is taken. The comparison of the accuracy obtained by SVM classification and the regression method is shown in Table 4.

From Table 4, it can be concluded that the accuracy of the regression method is increased by 14% compared with the SVM method, which proves the effectiveness of the regression method. In addition, it can be found from the above table that the accuracy rates of categories 2, 3, and 4 are higher than those of categories 1 and 5, indicating that the feature vectors of categories 2, 3, and 4 can better express this category, and clearly compare it with other categories. The categories are distinguished, while the features 1 and 5 are not clear enough, and it is easy to divide them into other categories. Further research is needed on the selection of features.

Table 2: Efficiency of various regression algorithms.

	MAE of valence	MAE of arousal	Ac of valence (%)	Ac of arousal (%)	Ac of classification (%)
MARS	0.2826	0.2776	61.4	70	56
SVM	0.2333	0.2173	71	82	63
RBFR	0.1987	0.1803	72.4	80	64

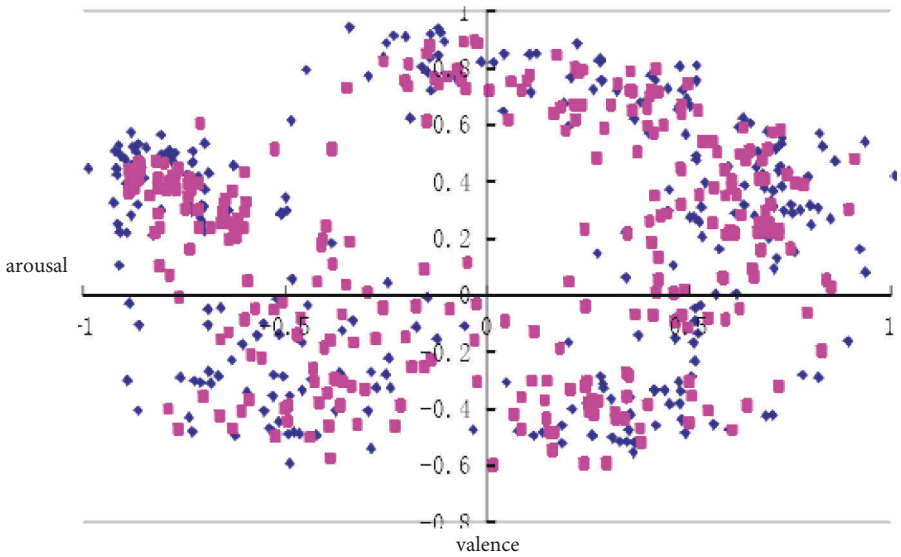

Figure 3: The distribution of the manually labeled values and predicted values of RBFR regression V and A. Pink rectangle is the prediction value and blue rectangle is manually labeled values.

Table 3: Comparison of RBFR combined with support vector regression and used alone.

	Ac of valence (%)	Ac of arousal (%)	Ac of classification (%)
SVR	71	81	63
RBFR	72	80	64
Combined	72	81	68

3.3. Analysis of the Experimental Results of Music Emotion Classification. The MediaEval database used in this study does not clearly mark the "music emotion type," but the V and A values of the music emotion attribute are marked in detail. The study defines the emotion type of music according to the coordinate quadrant of the mapping point of V and A in the "V-A emotion plane." The emotion type of music defines four following categories: I, II, III, and IV, representing cheerfulness, anger, sad, and calm musical emotions, respectively. At the same time, in order to overcome the influence of subjective scores, in the MediaEval database, the samples whose coordinates determined by the V and A values in the annotation are less than 0.05 from the origin are excluded; 125 sample data are randomly selected as the test sample set during the experiment, and the remaining data are used as the training sample set was independently repeated 10 times. Similar to music sentiment regression, this study performs PCA dimensionality reduction and Relief feature selection on all the extracted music features, and conducts sentiment classification experiments in all feature spaces, PCA feature spaces, and Relief feature spaces, respectively.

In this study, support vector machines, fuzzy neural networks, K-neighborhood, fuzzy K-neighborhood, Bayesian, linear discriminant analysis, and the proposed hybrid classification algorithm were used to train sentiment classification models, respectively. SVM adopts RBF kernel function, and relevant parameters are determined by optimization; K parameter in KNN is 8; FKNN adopts Gaussian function as fuzzy function; Bayes classification and LDA classification are realized by MATLAB built-in functions, respectively. At the same time, the paper also conducts model training and testing on the hybrid classifier proposed in this paper as shown in Figure 4, and each independent classifier adopts the above corresponding configuration. Finally, the experiment gives the results of music sentiment classification by multiple classifiers and compares the results of the hybrid classifier and the independent classifier.

After PCA dimension reduction, the feature dimension is 139 dimensions. The Relief feature selection algorithm is used to select the original data; if the weight is greater than 0.01, the feature dimension used for classification model training and prediction is finally determined to be 166 dimensions. The classification experiment results are shown in

Table 4: Comparison of SVM classification and regression classification.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Sum
Total songs	80	80	80	80	80	400
SVM correct number	32	41	50	56	28	207
Regression correct number	40	52	61	69	40	262
SVM accuracy rate	40%	51%	63%	70%	35%	52%
Regression accuracy number	50%	65%	76%	86%	50%	66%

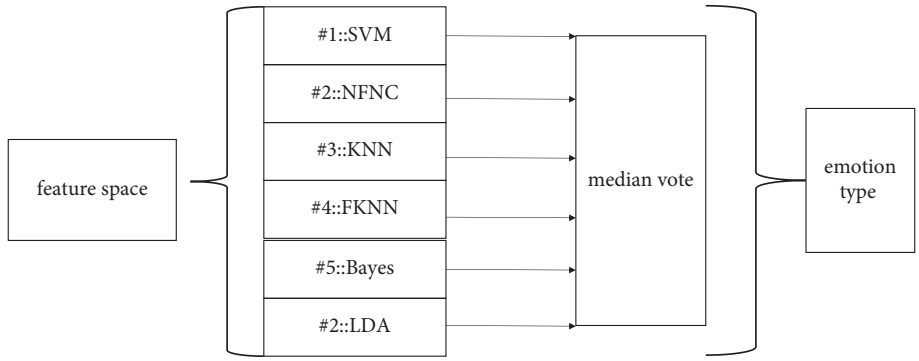

Figure 4: Mixed classifier structure diagram.

Table 5. It can be seen from the experimental results that the hybrid classifier model proposed in this paper has achieved the best music emotion recognition effect in all feature spaces, PCA feature spaces and Relief feature spaces, and its recognition accuracy is 84.9%, 83.4% and 80.2%, respectively. The experimental results show that using the hybrid classifier proposed in this paper can improve the accuracy of music emotion classification and further reduce the risk of misclassification. The proposed hybrid classifier method is effective.

By comparing the experimental results, it can be found that the classification effects of different classifiers in the three feature spaces have certain differences; music emotion classification also has the characteristics based on special cases, so it is necessary to select the optimal classifier and feature space to realize music emotion recognition task. The experimental results further show that by comparing the results on different feature spaces, it can be seen that the classification accuracy on the PCA feature space has increased or has only a small decrease in classification accuracy compared with the use of all features, so the PCA feature space is more suitable for music emotion [38]. Classification: comparing the recognition results of different classifiers on the same feature space, in addition to the proposed hybrid classifier with the best results, support vector machines, fuzzy K-neighborhood, and linear discriminant analysis all have the best performance in music emotion classification tasks. Good classification ability, in all feature spaces and PCA spaces, the recognition accuracy of these algorithms is greater than 80% under the experimental conditions and MediaEval database music samples. The results of one-time classification using the hybrid classifier on all feature sets show that most of the errors are in the classification between adjacent quadrants. This problem should be paid enough attention in the future music emotion recognition.

Table 5: Experimental results of music emotion classification (average accuracy, %).

	KNN	Bayes	LDA	NFNC	FKNN	SVM	Hybrid
ALL	62	69	80.4	79.3	83	83	85
PCA	62.8	76	81.6	58	82.2	81.1	83.5
Relief	66.8	59	77.4	69	50.3	79.3	80.2

4. Conclusions

In this study, the music emotion recognition was discussed and the mapping relationship of music emotion attributes on the "V-A emotion plane" was analyzed. The focus is on the use of music feature extraction and machine learning methods to achieve music emotion recognition.

Based on the MATLAB signal processing toolbox, sound description toolbox, and music information retrieval toolbox, this paper extracts the features related to music emotion (including energy, rhythm, harmony, time domain, and spectrum and other features). The dimension is 548, and the dimension of music feature space is reduced by the method of principal component analysis space projection and Relief feature selection.

Based on the emotional "valence-incentive" model, this paper applies machine learning algorithms such as multivariate adaptive regression spline method, support vector regression, radial basis function regression, random forest regression, and regression neural network, respectively. In the Relief feature space, the optimal regression results were achieved based on the support vector machine regression method and the random forest algorithm, respectively. The emotional valence and the incentive value R^2 for the statistical values are 29.3% and 62.5%, respectively, which are better than the results reported in the literature.

Based on intelligent algorithms such as support vector machine classification, K neighborhood classification, fuzzy neural network classification, fuzzy K neighborhood classification, Bayesian classification, and Fisher linear discrimination, this paper analyzes music in all feature spaces, PCA feature spaces, and Relief feature spaces, respectively. Emotions are classified; among them, the correct rate of music emotion classification realized by support vector machine, fuzzy K neighborhood, and Fisher linear discriminant algorithm is more than 80%; combined with the above intelligent algorithm, a hybrid classifier is proposed, which includes six independent subclassifiers and median voting decision algorithm implemented by the above intelligent classification algorithm; based on the hybrid classifier, in each feature space, the best classification results are achieved on all feature spaces and the recognition accuracy of music emotion on all feature spaces and PCA feature spaces is as high as 84.9% and 83.4%, respectively.

This study implements a music emotion classification system. Compared with traditional methods, the classification performance has been improved to a certain extent, but there are still areas for improvement in the process of learning and experimentation. First of all, although the music emotion database is established according to the emotion classification standard proposed at the International Conference on Music Information Retrieval, the standard is based on English songs. The description of categories in English may be biased in Chinese understanding. The important thing is that there is no broad mass base, and it is only established by students from the same school. Due to the small number of people, the obtained music library does not represent the will of the majority of people in a certain sense. In future experiments, it may be possible to collect the power of everyone on the Internet to build a more unified and convincing music emotion classification library. Using the regression idea to solve the problem of music emotion classification has its advanced nature. The regression algorithm has been researched maturely in a certain sense, but for music characteristics and emotional variables, which is insufficient. In this paper, only the support vector regression and k-plane segmentation regression algorithms are selected, which are more suitable for this system. In the future, further research and experiments can be carried out to obtain better results. In terms of feature selection, this paper selects the cepstral and spectral feature parameters of MFCC and RASTA-PLP to characterize music fragments based on the previous research and does not do further research. For musical emotion, due to its ambiguity and subjectivity, as well as the complexity of the process of human perception of emotion, what factors constitute the difference of musical emotion remains to be studied, and only these two types of characteristics cannot fully express its characteristics. In terms of feature selection, more experiments and screening are needed to obtain more accurate feature expressions in terms of musical emotion.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Acknowledgments

The study was supported by the Humanities and Social Science Research Planning Fund Project of the Ministry of Education in China 2020 Research on "Environmental Music" Cultural Mission (20YJA760100).

References

[1] R. W. Picard, "Affective computing: challenges," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 55–64, 2003.

[2] S. Zhu and Y. Liu, "Automatic artist recognition of songs for advanced retrieval," *Journal of Shanghai Jiaotong University*, vol. 13, no. 5, pp. 513–520, 2008.

[3] T. Beierholm and P. Baggenstoss, "Speech music discrimination using class-specific features," vol. 2, pp. 379–38217, Cambridge, UK, August 2004.

[4] D. Liu, N. Zhang, and H. Zhu, "Computer aided design system for developing musical fountain programs," *Tsinghua Science and Technology*, vol. 5, no. 6, pp. 612–616, 2003.

[5] Y. H. Yang, C. C. Liu, and H. Homer, "Music emotion classification: a fuzzy approach," *IEEE Transactions on Software Engineering*, vol. 25, no. 1, pp. 70–93, 2000.

[6] W. Alicja, S. Piotr, and W. Zbigniew, "Multi-label classification of emotions in music," *Advances in Intelligent Systems and Computing*, Springer, vol. 35, no. 10, Berlin, Germany, 2006.

[7] K. Hevner, "Expression in music: a discussion of experimental studies and theories," *Psychological Review*, vol. 42, no. 2, pp. 186–204, 1935.

[8] K. Hevner, "Experimental studies of the elements of expression in music," *American Journal of Psychology*, vol. 48, no. 2, pp. 246–268, 1936.

[9] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980.

[10] J. A. Russell, A. Weiss, and G. A. Mendelsohn, *Journal of Personality and Social Psychology*, vol. 57, no. 3, pp. 493–502, 1989.

[11] A. Hanjalic and L. Xu, "Affective video content representation and modeling," *IEEE transaction of Multimedia*, vol. 16, no. 1, pp. 143–154, 2005.

[12] M. D. Korhonen, D. A. Clausi, and M. E. Jernigan, "Modeling emotional content of music using system identification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 36, no. 3, pp. 588–599, 2006.

[13] F. Sally and W. Manfred, "Sample compression, learn ability, and the vapnik-chervonenkis dimension," *Machine Learning*, vol. 2, no. 2, p. 213, 1995.

[14] B. Parisa, G. Amin, and A. Mojtaba, "Support vector regression based determination of shear wave velocity," *Journal of Petroleum Science and Engineering*, vol. 3, no. 11, pp. 15–17, 2014.

[15] J. Dunik, M. Simandl, and O. Straka, "Unscented kalman filter: aspects and adaptive setting of scaling parameter," *IEEE Transactions on Automatic Control*, vol. 57, no. 9, pp. 2411–2416, 2012.

[16] O. Mangasarian and N. Arbitrary, "Arbitrary-norm separating plane," *Operations Research Letters*, vol. 24, no. 1-2, pp. 15–23, 1999.

[17] C. Chang and C. Lin, "LIBSVM: a library for support vector machines," *ACM transactions on intelligent systems and technology*, vol. 2, no. 3, pp. 389–396, 2001.

[18] K. Yamawaki and H. Shiizuk, "Synesthesia and common recognition concerting music and color," *Journal.-of.Systems.and.Control.Engineering*, vol. 220, no. 1, pp. 735–742, 2006.

[19] Z. Eitan and I.. Rothschild, "How music touches: musical parameters and listeners' audio-tactile metaphorical mappings," *Psychology.of.Music*, vol. 39, no. 4, pp. 449–467, 2011.

[20] R. Anders, "Emolion rendering in music range and characteristic values of seven musical variables," *Cortex*, vol. 47, pp. 1068–1081, 2011.

[21] M.-K. Shan, F.-F. Kuo, M.-F. Chiang, and S. Y. Lee, "Emotion-based music recommendation by affinity discovery from film music," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7666–7674, 2009.

[22] Y. H. Yang and H. H. Chen, "Ranking-based emotion recognition for music organization and retrieval," *IEEE Transactions on Audio Speech and Language Processing*, vol. 19, no. 4, pp. 762–774, 2011.

[23] Y. C. Lin, Y. H. Yang, and H. H. Chen, "Exploiting online music tags for music emotion classification," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 7S, no. 1, pp. 1–16, 2011.

[24] K. J. Zhang and S. Q. Sun, "Web Music Emotion Recognition Based on Higher Effective Gene Expression programming," *Neuro Computing*, vol. 105, pp. 1–7, 2012.

[25] K. Hevner, "The affective value of pitch and tempo in music," *American Journal of Psychology*, vol. 49, no. 4, pp. 621–630, 1937.

[26] B. E. Boser, "A training algorithm for optimal margin classifiers," *Proceedings of Annual Acm Workshop on Computational Learning Theory*, vol. 5, pp. 144–152, 2008.

[27] C. B. Moon, H. S. Kim, H. A. Lee, and B. M. Kim, "Analysis of relationships between mood and color for different musical preferences," *Color Research & Application*, vol. 39, no. 4, pp. 413–423, 2014.

[28] W. Xin, W. Li, and X. Lingyun, "Comparison and Analysis of Acoustic Features of Western and Chinese Classical Music Emotion Recognition Based on V-A Model," *Applied Sciences*, vol. 12, no. 12, 2022.

[29] W. N. Wang, Y. L. Yu, and S. M. Jiang, "IEEE Systems, Man, and Cybernetics Society," in *Proceedings of the 2006. SMC '06. IEEE International Conference on*, pp. 3534–3539, Taipei, Taiwan, 2006.

[30] E. Schmidt and Y. Kim, *Projection of Acoustic Features to Continuous Valence-Arousal Mood Labelsvia regression*, 2009.

[31] G. Jacek, "Music emotion recognition using recurrent neural networks and pretrained models," *Journal of Intelligent Information Systems*, vol. 57, no. 3, 2021.

[32] C. Laurier, O. Lartillot, and T. Eerola, "Exploring Relationships between Audio Features and Emotion in music," in *Proceedings of the 7th Triennal Conference of European Society for Cognitive Sciences of Music*, Jyväskylä, Finland, 2009.

[33] R. Dannenberg, "SMERS: music emotion recognition," in *Proceedings of the 10th International Society for Music Information Retrieval Conference, ISMIR 2009*, pp. 651–656, Kobe, Japan, October 2009.

[34] Y. H. Yang, Y. C. Lin, Y. F. Su, and H. H. Chen, "A regression approach to music emotion recognition," *IEEE Transactions on Audio Speech and Language Processing*, vol. 16, no. 2, pp. 448–457, 2008.

[35] Y. H. Yang and H. H. Chen, "Machine recognition of music emotion: a review," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 3, no. 3, p. 1, 2012.

[36] M. You, J. Liu, G. Z. Li, and Y. Chen, "Embedded feature selection for multi-label classification of music emotions," *International Journal of Computational Intelligence Systems*, vol. 5, no. 4, pp. 668–678, 2012.

[37] X. Zhu, Y. Y. Shi, H. G. Kim, and E. Ki-Wan, "An integrated music recommendation system," *IEEE Transactions on Consumer Electronics*, vol. 52, no. 3, pp. 917–925, 2006.

[38] H. Wang, "Research on the application of wireless wearable sensing devices in interactive music," *Journal of Sensors*, vol. 2021, pp. 1–8, Article ID 7608867, 2021.