Music Emotion Feature Recognition based on Internet of Things and Computer-Aided Technology

Chaode Yang[1*] and Qingxun Li[2]

[1]Department of Art and Sports, Huanghe S&T University, Zhengzhou 457000, China,
yangchaodeabc@163.com
[2]Academy of Music, Tianjin College of Media & Arts, Tianjin 300000, China,
hnzzll1976@163.com

Corresponding author: Chaode Yang, yangchaodeabc@163.com

Abstract. As an important part of human life, music can convey emotion and regulate the emotions of listeners. Emotion is one of the essential features of music, and the relationship between music and emotion has become the subject of many academic studies. At present, with the rapid development of information technology and artificial intelligence, music emotion recognition has made rapid progress and become one of the important research directions in the field of digital music. Aiming at the problem of poor classification effect of musical emotion caused by the monotony of Support Vector Machine (SVM) projection space, this paper proposes an optimized SVM model for music feature emotion recognition. The new method can not only improve the accuracy of music emotion classification, but also improve the running speed and interpretability of the model. At the end, the practicality and reliability of the new approach are verified by public classification data sets and real music emotion data sets. This paper proposes an optimized SVM model for music feature emotion recognition. The new method can not only improve the accuracy of music emotion classification, but also improve the running speed and interpretability of the model. Finally, the practicality and reliability of the new approach are verified by both the public classification data sets and real music emotion data sets.

Keywords: Music feature emotion recognition; SVM; Computer aided technology; Optimization algorithm
DOI: https://doi.org/10.14733/cadaps.2022.S6.80-90

1 INTRODUCTION

At present, information technology is developing rapidly. Digitalization, as the core of information technology, promotes the vigorous development of various emerging technologies, among them, the power in the middle of the audio digitization become the digital trend Audio is to digitize sound and save the result, and music is an important part of audio, in recent years, digital music has attracted a large number of scholars to carry out research on it. Compared with traditional music, the advantages of digital music are mainly reflected in the production, storage, transmission and retrieval. Digital signal processing technology reduces the cost of music storage, and the application and popularity of the Internet promote the transmission of digital music with audio retrieval technology With the development of technology and the rapid growth of music data, the traditional retrieval based on text content is gradually difficult to meet the needs of users, the retrieval based on audio content is gradually rising, the research of digital audio content has gradually become the hot spot of audio digital technology.

But for music, its essential character is music emotion, music is more information behavior research pointed out that emotion is one of the most important standards people use in music retrieval based on emotional music retrieval is the core of music emotion recognition, as one of the important research direction of digital music, music is a multidisciplinary integration field Sense of identification research covers music psychology, music acoustics Audio signal processing In the field of natural language processing and machine learning, and other content and more than hundreds of related research results show that the different audience in music expression of emotion judgment so often have consistency with high accuracy to be can for emotion recognition.

Nowadays, driven by the two huge waves of Internet and information technology, multimedia industry has achieved unprecedented vigorous development. More and more forms of music are readily available, making it more difficult for people to choose their favorite music. The traditional music push mechanism adopts the method of user feedback. Moreover, due to the number and scale of users and subjective reasons, the evaluation results may not be in line with the public's aesthetic standards. If mature automatic music emotion recognition method is adopted, the efficiency of push will be greatly improved. In addition, music emotion recognition can not only be applied to the field of music push, but also to music creation. Music emotion recognition technology plays a feedback role in music creation. If the automatic music recognition technology is used to effectively discover the connection between music and emotion, and this connection is applied to the composition of music, it can effectively shorten the composition cycle of music.

Nowadays, music emotion recognition research progress is rapid, and successfully applied in various fields, including music emotion retrieval music emotion recognition of the performing arts, smart space design, such as music emotion recognition technology has paved the way for the road in music visualization research field, it shows great research prospects and important application value [1]. There are more and more researchers in music emotion recognition field, and the resulted works are updated quickly. But because of the complexity of music emotion information, music emotion recognition is a long-term research goal.

2 RELATED STUDIES

Musical emotion is a kind of psychological process, including various human emotional factors generated in the process of interaction between people and music. The purpose of music emotion recognition is to analyze the audio characteristics related to music file data and music emotion by computer, and to construct emotion classification model according to certain emotional psychological model, so as to establish corresponding relationship between music and various emotions and provide services for the demanders. Since the 1980s, more and more scholars have been engaged in emotional recognition of music. Katayose et al. published an article on emotional recognition of piano music in 1988, which for the first time took melody, chord, rhythm and other music-related factors as an important measurement standard to evaluate musical emotion, and extracted the audio features related to the above factors to carried on the classification, inspired later scholars.

The recognition of musical emotion is a pattern recognition process which requires the correct mapping of data from original musical feature space to new emotional space. In the research of musical emotion recognition, many foreign research institutions have carried on this research. Qin et al. [2] adopted the continuous regression method to predict the change point of English emotion. Daube et al. [3] studied a variety of acoustic features, including low-level features, melodic tones for musical emotion identification. Rajaee et al. [4] and Buettner et al. [5] combined the deep two-way LSTM model with extreme learning machine, and made real-time emotional state prediction for music fragments. Moscato et al. [6] proposed a bidirectional convolutional recursive sparse network, whose input is the spectrum map of music. In addition, the emotional recognition of music is not limited to the original audio, and all the methods mentioned above only use the original music for classification. Song et al. [7] proposed a multi-modal assessment model of music emotion, which not only considered the original sound of music, but also added lyrics of music and online users' comments on music as reference standards to classify features not limited to audio features, but also features from texts.

Research on affective information processing and artificial emotion started late in China. In the late 1990s, China listed the content related to affective computing as a key project in the Project guide of national Natural Science Foundation of China. Since then, the domestic research on musical emotion has gradually deepened and made some progress. Zhang et al. [8] studied the multi-modal music emotion classifier including the audio bottom feature and song text feature, and in 2011 studied the music sorting algorithm based on two-dimensional emotion model. Yang. [9] studied multi-label music emotion recognition technology based on evidence theory and semantic cell model based on timbre and rhythm characteristics, and applied principal component analysis to feature dimension reduction. Bhavan et al. [10] studied the song emotion recognition algorithm based on spiral model, extracted emotional features such as tone strength and speed, and adopted hierarchical K-nearest Neighbor classifier to construct spiral emotion model to realize the classification of song emotion. After that, the authors designed a neural network model combining attention mechanism with long and short-term memory network for music emotion recognition, and completed music recommendation with convolutional cyclic neural network.

At the present stage, the classification of musical emotions is still in its infancy, although there are breakthroughs in general emotional cognition of music in certain areas. But still no an algorithm in the face of the cover more style, more than cultural background music of the data sets satisfying, ignore the music the limitation of data source, machine learning by increasing the number of music training can expand the range of emotion recognition, in the face of small style attribute clear music, music feature classification method is still based on manual design is worth considering. In addition, music of similar emotions can also be divided into more finely differentiated emotions. The classification of such fine-grained emotions is also a thorny challenge, and the classification scope and separation granularity are still problems to be overcome in the future of music emotion classification. This paper proposes an optimized ELM model for music feature emotion recognition, which can not only improve the accuracy of music emotion classification, but also improve the running speed and interpretability of the model. The practicality and reliability of the proposed method are also verified by different datasets.

3 EMOTION FEATURE RECOGNITION BASED ON OPTIMIZED SVM MODEL

3.1 SVM Model

The purpose of establishing the music emotion classification model is to map the characteristic data of music emotion to be classified into the known four basic emotions through the classification model. The automatic recognition and classification process of emotion generally has two stages: The first stage is the training stage, in which the labeled training set is input emotion multi-classifier, and the relevant parameters of the classification model are determined by predicting the marked data. The second phase of testing phase, namely the selection of the new music pieces as a test set, using the trained classification model to predict, comparing predicted results with the actual category, and the statistical recognition as one of the evaluation model of performance indicators At present, the common classification methods are naive Bayesian classification decision tree (NBC), Artificial neural network (ANN), k-nearest neighbor (KNN), and Support Vector Machine (SVM), however, as the extracted musical emotional features are complex and non-linear, SVM classification algorithm is used In addition, a number of research results show that SVM based on Radial Basis Function (RBF) has better performance than traditional NBC or KNN classifier. Compared with ANN, THE classification process of SVM is more transparent and can avoid falling into local minimum, which is suitable for small data sets.

The schematic diagram of SVM is shown in Figure 1.

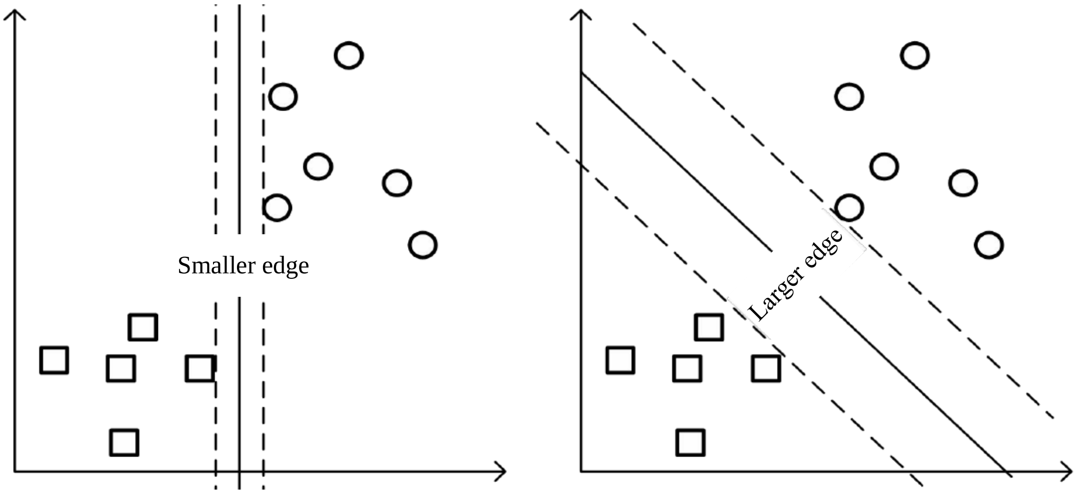


Figure 1: Schematic diagram of large and small edge hyperplanes.

For linearly separable SVM, the classifier divides the samples into two categories, as shown in Equation (1):

$$\begin{cases} \vec{\omega} \cdot \vec{x}_i + b \geq 1, & y_i = 1 \\ \vec{\omega} \cdot \vec{x}_i + b \leq -1, & y_i = -1 \end{cases} \tag{1}$$

Where $\vec{\omega}$ is Weight vector, b is bias.

Transform the constraint of the minimum value into the constraint of the maximum value, and obtain the optimization objective and constraint conditions, as shown in Equation (2):

$$\begin{cases} \min \frac{\| \vec{\omega} \|^2}{2} \\ \text{s.t. } y_i \left(\vec{\omega} \cdot \overrightarrow{x_i} + b \right) \geq 1, i = 1, 2, \cdots, l \end{cases} \tag{2}$$

The Lagrange function is defined as shown in Formula (3), and then the optimization problem can be transformed into a dual problem described as shown in Formula (4):

$$\Phi\left(\vec{\omega}, b, \alpha_i\right) = \frac{1}{2} \| \vec{\omega} \|^2 - \sum_{i=1}^{l} \alpha_i \left[y_i \left(\vec{\omega} \cdot \overrightarrow{x_i} + b \right) - 1 \right], i = 1, 2, \cdots, l \tag{3}$$

$$\begin{cases} \max Q(\vec{\alpha}) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \left(\overrightarrow{x_i}, \overrightarrow{x_j} \right) \\ \text{s.t.} \quad \sum_{i=1}^{l} \alpha_i y_i = 0, \quad \alpha_\text{i} \geq 0 \end{cases} \tag{4}$$

Where α_i is the Lagrange coefficient. The optimal classification function finally obtained is shown in Equation (5):

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} \alpha_i^* y_i \left(\overrightarrow{xx_i} \right) + b^* \right) \tag{5}$$

3.2 Optimized SVM Model based on PSO

Particle swarm optimization (PSO) algorithm was firstly given by Kennedy and Eberhart after simplification motivated by computer simulation of the simple alternating behavior of birds foraging. The velocity and position update formulas of particles are shown in Equation (6):

$$\begin{aligned} V_{id}^{k+1} &= \omega V_{id}^{k} + c_1 r_1 \left(P_{id} - X_{id}^{k} \right) + c_2 r_2 \left(N_{id} - X_{id}^{k} \right) \\ X_{id}^{k+1} &= X_{id}^{k} + V_{id}^{k+1} n \end{aligned} \tag{6}$$

Where c_1 and c_2 are the learning factor, r_1 and r_2 are uniform random number within the range of [0,1], ω is inertia weight. The right side of equation (6) is divided into three different parts. The first one is the inertia part, which represents the habit of particles to the previous velocity. The second part is the classification part, which represents the reflection of the particle on its optimal position in history. The last part is the social part, which reflects the optimal group position shared between particles. In order to improve the ability of PSO algorithm to balance global and local search, this paper adopts the PSO algorithm of inertia weight with adaptive weight, whose value is related to the objective function and fitness function, and the formula of inertia function is:

$$\omega = \begin{cases} \dfrac{\left(\omega_{\max} - \omega_{\min}\right)\left(F\left(X_i\right) - F_{\min}\right)}{F_{\text{avg}} - F_{\min}} + \omega_{\min}, & F\left(X_i\right) \leq F_{\text{avg}} \\ \omega_{\max}, \quad F\left(X_i\right) > F_{\text{avg}} \end{cases} \tag{7}$$

Where $\omega_{\max}$, $\omega_{\min}$ are the maximum and minimum inertia coefficients, F is the fitness function of the particle, F_{avg} and $F_{\min}$ are the minimum and average value of the fitness function.

4 RESULTS ANALYSIS

4.1 Model Performance Analysis

The experimental environment information is shown in Table 1. From 258 pieces of music, 200 pieces were selected as the training set, and 50 pieces were used for each type of music. Another 58 songs were used as the test set, including 12 songs of excitement, 15 songs of tension, 11 songs of sadness and melancholy, and 20 songs of relaxation and tranquility. In this section, the one-to-one support vector machine multi-classification model is compared with the one-to-many support vector machine multi-classification mode complete binary tree support vector machine multi-classification model.

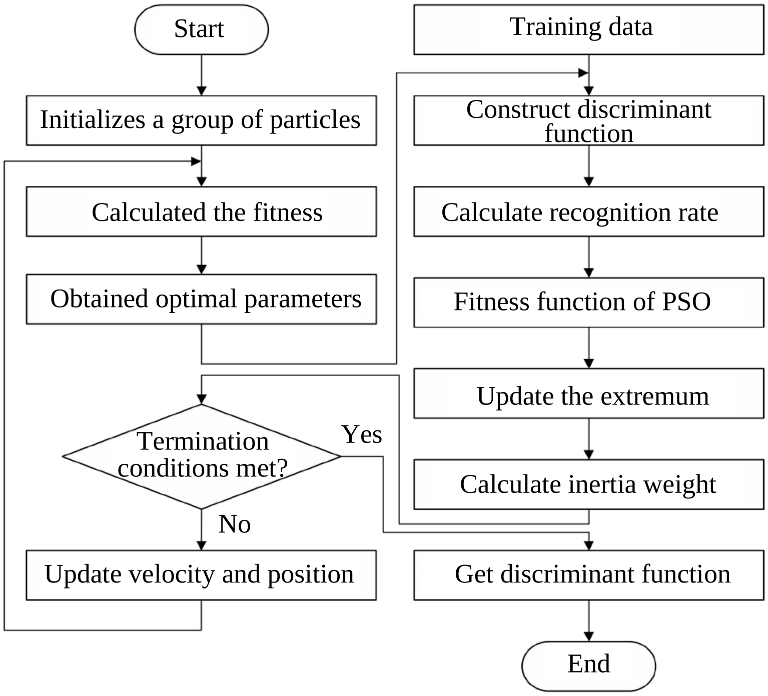

Start
Initializes a group of particles
Calculated the fitness
Obtained optimal parameters
Termination conditions met?
Yes
No
Update velocity and position
Training data
Construct discriminant function
Calculate recognition rate
Fitness function of PSO
Update the extremum
Calculate inertia weight
Get discriminant function
End

Figure 2: Flow chart of music emotion feature recognition.

Indicators	Performance parameter
Operating System Version	9.21
Bits in your system	7.99
internal storage	6.23
processor	32.24

Table 1: Music emotion feature recognition result.

The penalty parameter and kernel parameter are determined by empirical method. Then, PSO model is carried out to optimize the parameters of complete binary tree SVM globally, which further improves the recognition rate of music emotion.

In order to evaluate the performance of the algorithm, the ten-fold cross-validation method is carried out verify, which is a standard method to measure the accuracy of the learning performance on a specific data set. The original dataset is randomly classified as five partitions, where the class representation is roughly the same as the attributes in the full data set. During each run, one partition is selected for validation, while the rest is used to train the model. After that, this process is repeated ten times to ensure each partition is trained only once. Classifier property is also assessed by calculating the ratio of the number of correct classifications to the total number of instances. Table 2 shows the comparison of classification results of four support vector machines with multiple classifiers.

SVM	Training time(s)	Testing time(s)	Training classification accuracy (%)	Testing classification accuracy (%)
one-to-one	9.21	0.019	80.24	76.54
one-to-many	7.99	0.008	76.23	70.21
binary tree	6.23	0.006	79.81	73.55
PSO binary tree	32.24	0.005	85.17	79.82

Table 2: Music emotion feature recognition result.

In terms of classification time, time is related to the number of classifiers and the calculation of parameters. Among them, the complete binary tree algorithm model uses the least classifier and the shortest time. However, after optimization with PSO algorithm, the training time is slightly longer than that of the complete binary tree SVM algorithm because of the increased calculation of parameter solving. However, due to the small number of classifiers, it takes the shortest time to test the optimal parameters of classification, so it is better than other algorithms.

From the point of view of classification accuracy, the accuracy is related to the solution of optimal classification hyperplane to the standard method of input features. With the introduction of PSO algorithm, the optimal solution of penalty parameter and kernel parameter is found through several iterations, which improves the accuracy and intelligence of model parameter adjustment. Experimental results show that the PSO optimized complete binary tree SVM multi-classification algorithm has the highest classification accuracy and is obviously better than the other three algorithms.

The SVM classification model contains three classifiers, and the PSO algorithm is used to optimize three groups of parameters. The population size of the PSO algorithm was set as 30, and the position and velocity of the initial particle were set as 0. The number of iterations is set as 100, the coefficient of local acceleration factor is set as 1.5, and the coefficient of global acceleration factor is set as 1.7. Through the method of five-fold cross verification, the recognition rate of five-fold average cross verification is obtained after five iterations, which improves the reliability of the algorithm and reduces kernel function parameter error. Thus, the validity and practicability of the proposed method are proved from another point of view.

Figure 3 shows the convergence performance of particle swarm optimization fitness for the complete binary tree support vector machine after optimization. With the increase of the number of evolutions, the average fitness is 75%-90%, which verifies the optimization performance of PSO and illustrates the influence of the change of the model's kernel parameters on the accuracy of the classification model. To sum up, the complete binary tree SVM algorithm optimized by PSO is superior to the other three algorithms, with the shortest test time, the highest classification accuracy and high performance. It can be seen from several SVM multi-classification models that the one-to-one and one-to-many models often need more classifiers, so when there are more classification categories, the classification of the SVM algorithm is better than the other three algorithms will increase significantly.

Figure 4 shows the process of the three algorithms describing the final convergence of the test function after 500 iterations. The abscissa denotes the iterations number and the ordinate is the average of 30 experiments. For the single-peak function, the number of iterations required for

MBAS to approach convergence is less than 150, while the PSO algorithm needs about 200. BAS algorithm runs about 300 times before converging.

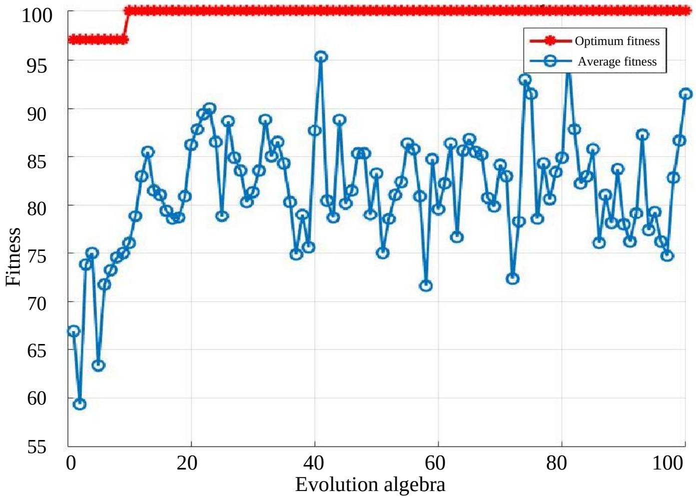


Figure 3: Convergence curve of fitness.

In addition, Figure 4 shows the convergence process of three algorithms. It can be seen from the figure that the convergence of sub-multi-peak function experiment is observed intuitively. It can be seen that the convergence speed of PSO algorithm is slightly faster than that of MBAS algorithm, and BAS algorithm slowly tends to converge. Thus, the validity and practicability of the proposed method are proved from another point of view.

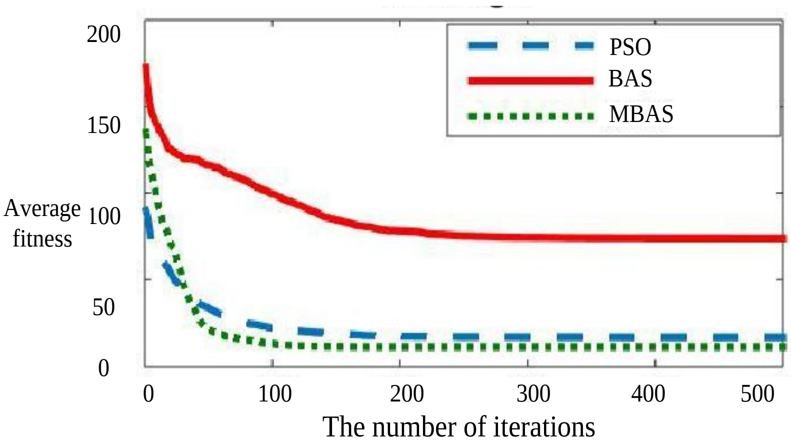


Figure 4: Three algorithms in the convergence process.

4.2 Music Emotion Feature Recognition Experiment based on SVM

This section focuses on the comparison of the time performance of three SVMS (one-to-one, binary tree, and PSO Binary Tree). Figure 5 and Figure 6 show the comparison of the training time of three SVMS in three musical emotion classification data sets respectively. On the left side of each group of graphs is the training time with MuStd feature, and on the right side is MOSC feature.

In general, the training time of MuStd feature was slightly longer than MOSC feature, because the dimension of MuStd feature was higher than that of MOSC feature one-to-one, and the training time of three data sets was much higher than that of binary tree and PSO Binary Tree, especially in the Song's dataset, whose training time exceeds 200 seconds, the one-to-one solution requires a lot of iterative operations and kernel function construction. PSO Binary Tree was better than BAS- binary tree and one-to-one except the training time sound-track.

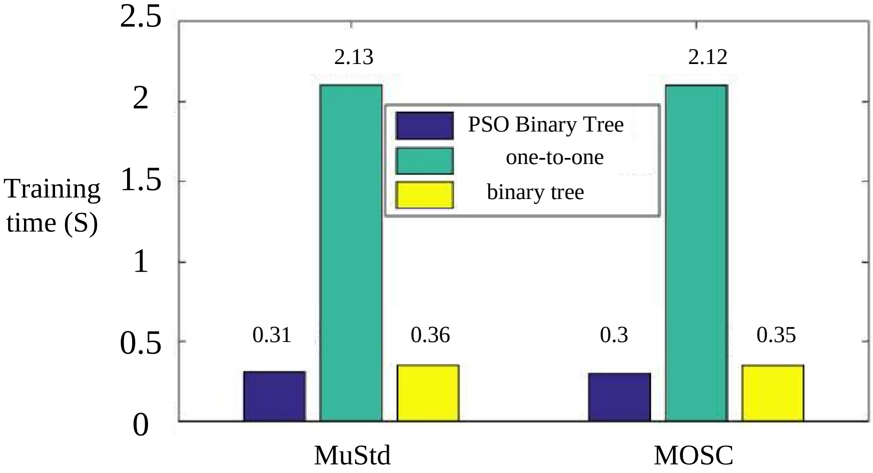


Figure 5: Training time of three SVMS in Sound-track.

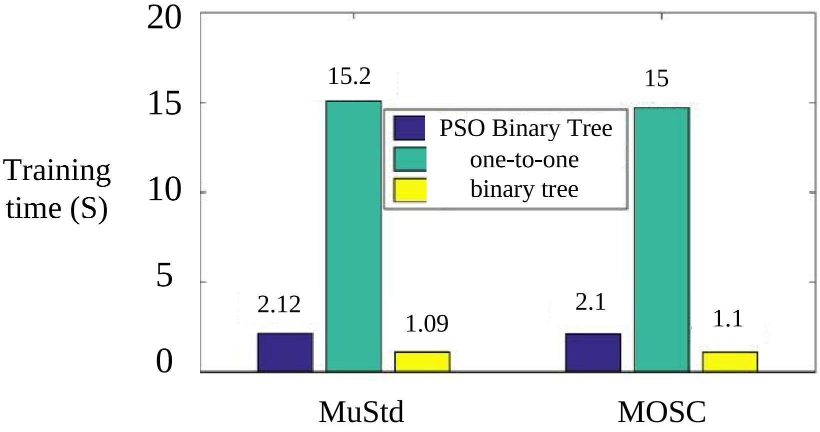
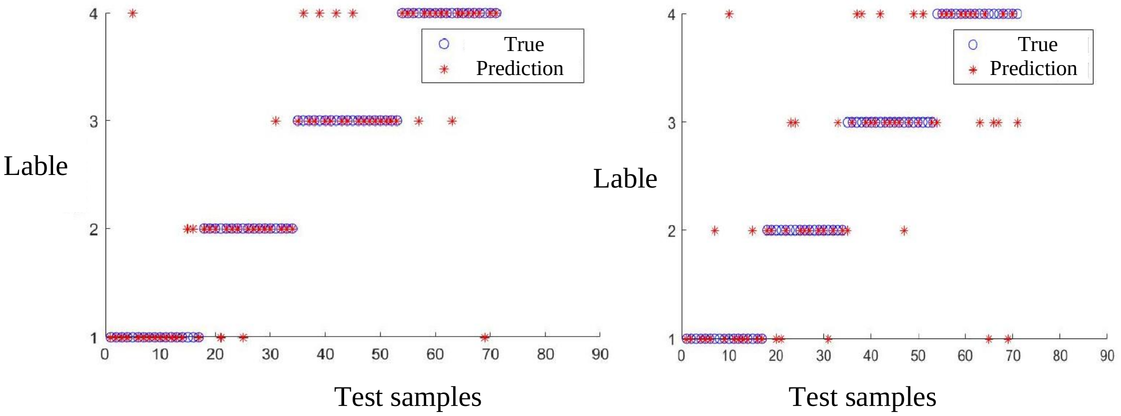


Figure 6: Training time of three SVMS in MIREX-modified.

To verify the effect of the classification model, a comparison was made between the proposed multi-classification model of music emotion based on complete binary tree SVM optimized by PSO and the music emotion classification model of BP neural network used in literature. The standardized music feature vector was taken as the input and the four kinds of emotions were taken as the output. (0,1,0,0) (0,0,1,0) (0,0,0,1) correspond to excited enthusiastic nervous angry sad melancholy and relaxed quiet.

According to Kolmogorov theorem, combined with extracted musical emotion features, BP neural network structure is designed as 7 15 4, that is, the input neurons is set as 7, the hidden neurons is set as 7, and the output layer neurons is set as 4.



(a) (b)

Figure 7: Test sample identification results.

In order to make the results of the output layer within the interval of [0,1], the training function is Levenberg-Marquardt BP, the learning function is learngdm, the performance function is mse, and the maximum number of cycles is 2000. The establishment of BP neural network was completed on the Python platform, and the recognition accuracy rate was 69%, as shown in Figure 7, which were the test sample classification results of the multi-classification model and BP neural network in this paper respectively.

In Figure 7, values 1, 2, 3 and 4 respectively correspond to the recognition results of two different algorithms: emotion 1(excited and enthusiastic), emotion 2 (nervous and angry), emotion 3(sad and melancholy) and emotion 4(relaxed and quiet). The multiple classification algorithm in this paper (a) has the largest number of samples of the four emotions accurately identified, which is superior to BP neural network (b) has higher generalization stability, higher recognition rate and practical value.

5 CONCLUSION

With the continuous development of modern information technology, the research on audio digitalization is also deepening the use of computer for music emotion recognition and the model is combined with a number of fields, gradually becoming a hot topic of research this paper first through the study of digital music foundation, combined with relevant literature, put forward some improvements on the feature analysis technology of music Then, according to the data scale, the emotion recognition model of music is established and optimized, and verified by relevant experiments. Finally, the model is applied to the generation of stage lighting control method, and effective reference is put forward for the design of stage lighting action. In this paper, the music emotion recognition model is studied, and the main work is as follows:

According to the emotional psychological model common in psychology, combined with the size of the experimental data, four kinds of basic emotions are used as the classification result, which is an important basis for establishing the emotional classification model. It avoids the ambiguity of emotional language and reduces the workload of system development. After that, the music emotion classifier is constructed according to the four kinds of basic emotions. By comparing several commonly used classification algorithms features, choose the support vector institutions built music emotion classifiers This article discussed several common support vector machine (SVM) classification model, selection of full binary tree support vector machine (SVM) over class model for music emotion classification using particle swarm algorithm for parameter optimization The final experiment design, Several support vector machine multi-classification models are compared, and the experimental performance of the emotion recognition model established are compared. It shows that the model established in this paper has a good performance in the experimental accuracy classification time.

Chaode Yang, https://orcid.org/0000-0002-5487-3881
Qingxun Li, https://orcid.org/0000-0002-7204-7348

REFERENCES

[1] Hizlisoy, S.; Yildirim, S.; Tufekc, Z.: Music emotion recognition using convolutional long short-term memory deep neural networks, Engineering Science and Technology, an International Journal, 24(3), 2021, 760-767. https://doi.org/10.1016/j.jestch.2020.10.009

[2] Qin, R.; Zhou, C.; Zhu, H.: A music-driven dance system of humanoid robots, International Journal of Humanoid Robotics, 15(05), 2018, 1850023. https://doi.org/10.1142/S0219843618500238

[3] Daube, C.; Ince, R.; Gross, J.: Simple acoustic features can explain phoneme-based predictions of cortical responses to speech, Current Biology, 29(12), 2019, 1924-1937. https://doi.org/10.1016/j.cub.2019.04.067

[4] Rajaee, M.; Hoseini, S.; Malekmohammadi, I.: Proposing a socio-psychological model for adopting green building technologies: A case study from Iran, Sustainable cities and society, 45, 2019, 657-668. https://doi.org/10.1016/j.scs.2018.12.007

[5] Buettner, R.; Sauer, S.; Maier, C.: Real-time prediction of user performance based on pupillary assessment via eye trackin, AIS Transactions on Human-Computer Interaction, 10(1), 2018, 26-56. https://doi.org/10.17705/1thci.00103

[6] Moscato, V.; Picariello, A.; Sperli, G.: An emotional recommender system for music, IEEE Intelligent Systems, 36(5), 2020, 57-68. https://doi.org/10.1016/j.cub.2019.04.067

[7] Song, T.; Zheng, W.; Lu, C.: MPED: A multi-modal physiological emotion database for discrete emotion recognition, IEEE Access, 7, 2019, 12177-12191. https://doi.org/10.1016/j.jestch.2020.10.009

[8] Zhang, J.; Yin, Z.; Chen, P.: Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review, Information Fusion, 59, 2020, 103-126. https://doi.org/10.1016/j.inffus.2020.01.011

[9] Yang, X.; Dong, Y.; Li, J.: Review of data features-based music emotion recognition methods. Multimedia systems, 24(4), 2018, 365-389. https://doi.org/10.1007/s00530-017-0559-4

[10] Bhavan, A.; Chauhan, P.; Shah, R.-R.: Bagged support vector machines for emotion recognition from speech, Knowledge-Based Systems, 184, 2019, 104886. https://doi.org/10.1016/j.knosys.2019.104886