Hindawi
Journal of Mathematics
Volume 2024, Article ID 9820467, 1 page
https://doi.org/10.1155/2024/9820467

Retraction

Retracted: Research on Automatic Classification Method of Ethnic Music Emotion Based on Machine Learning

Journal of Mathematics

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Z. Wu, "Research on Automatic Classification Method of Ethnic Music Emotion Based on Machine Learning," *Journal of Mathematics*, vol. 2022, Article ID 7554404, 11 pages, 2022.

Hindawi
Journal of Mathematics
Volume 2022, Article ID 7554404, 11 pages
https://doi.org/10.1155/2022/7554404

Research Article

Research on Automatic Classification Method of Ethnic Music Emotion Based on Machine Learning

Zijin Wu

College of Humanities and Management, Guilin Medical University, Guilin 541199, China

Correspondence should be addressed to Zijin Wu; wuzijin@glmc.edu.cn

Received 21 November 2021; Revised 7 December 2021; Accepted 11 December 2021; Published 4 January 2022

Academic Editor: Naeem Jan

With the development of the country's economy, there is a flourishing situation in the field of culture and art. However, the diversification of artistic expressions has not brought development to folk music. On the contrary, it brought a huge impact, and some national music even fell into the dilemma of being lost. This article is mainly aimed at the recognition and classification of folk music emotions and finds the model that can make the classification accuracy rate as high as possible. The classification model used in this article is mainly after determining the use of Support Vector Machine (SVM) classification method, a variety of attempts have been made to feature extraction, and good results have been achieved. Explore the Deep Belief Network (DBN) pretraining and reverse fine-tuning process, using DBN to learn the fusion characteristics of music. According to the abstract characteristics learned by them, the recognition and classification of folk music emotions are carried out. The DBN is improved by adding "Dropout" to each Restricted Boltzmann Machine (RBM) and adjusting the increase standard of weight and bias. The improved network can avoid the overfitting problem and speed up the training of the network. Through experiments, it is found that using the fusion features proposed in this paper, through classification, the classification accuracy has been improved.

1. Introduction

With the development of information and multimedia technology, people's demand in the entertainment industry has increased. Particularly, the increase in demand for music has promoted the vigorous development of the music market. As of June 2019, the number of online music users in my country reached 583 million, an increase of 29.76 million from the end of 2018, accounting for 68.8% of the total Internet users. Music has become an indispensable part of our lives. Folk music takes root under the nourishment of local cultures. It contains the characteristics of local cultures and presents various forms of expression. It combines the excellent and unique cultural connotations of various places and uses stage performances and other performance methods to show the people's pursuit of various aspects of living conditions, lifestyles, and ideals of life. Through its performance, it reveals the cultural ecology of each place, shows the unique culture of each place, and constitutes a vivid road map of life. Folk music is an important cultural resource. It contains people's understanding of life, views of things, and evaluation of characters. It is of great significance to the study of the humanities and ecology of the region. Through folk music, we can have a preliminary understanding of the ideology and life beliefs of the people in the region. Folk music is the embodiment of the core values of the Chinese nation, and it contains the essence of the traditional Chinese culture, which is a pursuit of morality and beauty, and is an important part of our cultural core competitiveness. The in-depth study of folk music is an interpretation of the life, world, and values of drama creators. Through the performance of historical figures and the recurrence of historical events, folk music adopts a way of expression that is easy for people to accept, expressing the mainstream values of the people of our country, showing our traditional virtues, and condensing our cultural core competitiveness.

Emotion in music [1] is an important attribute of music, and the use of music emotion is reflected in all aspects of our lives. When scoring film and television works, a series of music works are often created because of their specific themes and emotional needs. This music not only brings the auditory enjoyment to the audience, but also a suitable soundtrack which helps to express the emotion of film and television works and describe the story. In terms of psychotherapy, the right music can soothe the patient's inner trauma. The selection of these music often does not care about the style and age but pays more attention to the expressions of emotions in the music. All this shows that people are more or less inseparable from the influence of music emotions. It is this unique attribute of music that makes music spread widely. And folk music also has the basic attributes of emotional expression, so it is necessary and meaningful to carry out recognition and classification research on the emotion of folk music. The motivation behind carrying out this research work includes several factors, such as the need for cultural protection, the support of national policies, the emotional needs of people, and the use of science and technology. These factors make the research on the identification and classification of folk music emotions not only necessary, but also meaningful and feasible. In fact, the recognition and classification systems are basically the same, whether it is the recognition and classification of the folk music emotions or the recognition and classification of popular music emotions. In the rest of the paper, Section 2 highlights some of the related literatures. In Section 3, the proposed methods are presented, involving DBN and SVM. In Section 4, the experimental analysis has been carried out to verify the strength of proposed methods. Finally, the conclusion is given in Section 5.

2. Related Work

With the rapid development of economy and society, audio resources are becoming more and more abundant, and people's demand and taste for music are gradually improving. Some methods of categorizing music are gradually emerging. Early music classification methods originated in the 1990s and were mainly used to distinguish the differences between speech, music, and environmental sounds. This is easier than distinguishing music styles. As the field of music classification continues to innovate, many pioneering research results have been produced. The subject of music style classification has gradually obtained more progress and results in practice and theory [2–4]. Music classification task is currently an important research direction based on content music retrieval, and it has received extensive attention from Music Information Retrieval (MIR) at home and abroad. In the MIR community, since 2004, most of the advanced tasks of the Music Information Retrieval Evaluation Exchange (MIREX) competition held every year are related to music classification, including music sentiment classification and genre classification. In the review of music classification in 2011 [5], through the investigation of music classification task, the current music classification task model is usually disassembled into two structures: feature extractor and classifier. Audio feature extraction is the most critical step in music classification. Whether the relevant features of the classification target can be extracted from the complex music signal is the key to improving the accuracy of the final classification. From the perspective of music comprehension, audio features can be divided into low-level, intermediate-level, and high-level features. Low-level characteristics refer to timbre characteristics and time characteristics. Tone characteristics capture the tonal quality of sounds related to different instruments, such as zero-crossing rate, spectral centroid, and Mel Frequency Cepstrum Coefficient (MFCC). The time feature captures the change and evolution of timbre over time, such as average, variance, and covariance. Low-level features can generally be directly obtained through signal extraction and processing techniques, which are easy to extract, and show good performance in almost all music classification tasks. Therefore, the current low-level features are important audio features of the audio classification system. But these low-level features are the basic attribute values of the signal, which is not closely related to people's perception. Therefore, compared with low-level features, intermediate features such as rhythm, pitch, and harmony are more in line with the musical attributes that people understand. These features can also be further extracted from low-level features through technology. High-level features currently refer to content-based semantic tags, such as genre, mood, genre, etc., and the attributes of intermediate features are more in line with people's direct perception of music. But this is an abstract concept because it is difficult to obtain directly from low-level and intermediate features.

In 2002, Tzanetakis and Cook [6] began to focus on the classification of music genres and provided a GTZAN data set containing 10 genres, which laid the foundation for future research on music genre classification. This article solves the problem of automatic music genre classification by using three sets of audio features (timbre, pitch, and rhythm). This research laid the foundation for music classification and labeling. Many researchers devote themselves to the study of manual extraction of combined features. MFCC is a single-tone low-level feature widely used in genre classification. Compared with other complex feature combinations, only MFCC features can produce a good classification effect. Currently, in content-based music classification, the common choices of classifiers are *k*-nearest neighbor [7], Support Vector Machine (SVM) [8], and Gaussian Mixture Models (GMM) classifier [9]. With the development of technology, such as logistic regression [10], Artificial Neural Network (ANN) [11], decision tree [12], Linear Discriminant Analysis (LDA) [13, 14], Gaussian mixture model, hidden Markov [15, 16], and other more advanced models have also been used for different music tasks. In recent years, deep learning has achieved remarkable results in both image and text fields, and more and more researchers have begun to explore the application of deep learning in the field of music. Convolutional neural network, as a typical neural network in deep learning, has begun to be widely used in music classification and labeling tasks. Unlike traditional classifiers that use complex music feature combinations, using Convolution Neural Network (CNN) as a classifier can use simple input features, such as MFCC, spectrogram, and audio signals. MFCC is still a lossy structure as an effective audio input form in low-level features, and the completely lossless original audio does not show better performance than the spectrogram. So, the spectrogram retains more information than MFCC. However, the dimensionality is lower than that of the original audio, and the information in the audio signal is fully utilized, which is more suitable for the classification task with CNN as the classifier [17]. In 2009, Kim et al. [18] first applied CNN to improve the accuracy of music genre and artist classification. In 2016, the CNN network designed by Choi et al. [19] showed excellent performance on the task of music classification. However, the spectrogram is not exactly the same as the traditional image. To some extent, the convolution of the spectrogram along the frequency axis seems to be unreasonable in the interpretation of music. Therefore, in 2014, Dieleman [20] introduced the 1D-CNN structure to deal with the spectrogram in the music category.

In the development process, convolutional neural networks have derived a variety of structures to greatly improve the performance of image recognition. The residual network ResNet [21] launched by He et al. in 2015 has made great innovations in convolutional neural networks. By introducing shortcut connections to solve the problem of network degradation, deeper networks can be effectively trained. Since ResNet was put forward, the variant networks of ResNet have emerged in an endless stream, each with its own characteristics, and the network performance has also been improved to a certain extent. The dense connection network DenseNet [22] realizes feature reuse through dense connection, which not only absorbs the advantages of ResNet, but also improves the model effect and has better performance. In 2018, Kim et al. [23] used the original audio signal as input and successfully applied the ResNet structure to automatic music labeling. Other commonly used deep learning models such as convolutional neural networks are usually based on large-scale label data. Due to the lack of current music emotion classification data, in addition to using enhanced data, the concept of transfer learning proposed in recent years can effectively solve such problems. Transfer learning is a new machine learning method, and it is a hot research topic now. There are extensive researches on images and texts. The parameters trained in the source domain are transferred to the target domain so that a small amount of sample data can be effectively learned. In 2017, Choi et al. [24] used the CNN method for source domain training, which use the trained parameters in other music classification tasks by transfer learning.

3. Method

In the study of the emotional classification of folk music in this paper, the establishment of a musical emotional model must first be combined with the expression characteristics of the music itself and related psychology knowledge. Because there are currently no publicly available data sets on folk music, it is necessary to collect and sort out music data sets before classification. And before obtaining the music features required for classification, it is also necessary to perform preprocessing operations on the music fragments. Because the deep belief network can perform feature extraction on sample data through autonomous learning, it can be well suited for tasks that require highly abstract and complex features. And compared to the convolutional neural network that is good at processing two-dimensional images, it can handle one-dimensional data well, and music signals are typical three-dimensional data. In order to achieve a better classification effect of folk music emotions, consider using the DBN structure model to extract music features and achieve classification. In this section, in the research on the recognition and classification of folk music emotions, through the exploration of the DBN network training and fine-tuning process, the DBN and SVM algorithms are combined to achieve classification.

3.1. Music Sentiment Classification Based on DBN and SVM Algorithm

3.1.1. Classification Model Based on DBN and SVM. Considering that DBN can learn more abstract and comprehensive features that characterize music attributes, this paper combines the algorithms of DBN and SVM to achieve classification. The deep belief network is used to extract the features, and then the SVM classifier is used for classification. That is to say, the output of the last hidden layer of DBN is used as the input of the SVM classifier to build a new classification model. First, the extracted music features are used as the input of the DBN network model. DBN obtains the music features extracted by DBN by learning the input feature data. Then, use the feature data extracted by DBN to train the SVM classification model. Finally, the test samples are classified through the trained classification model to realize the recognition and classification of folk music emotions. Therefore, the new classification model abstracts the input features of the network to obtain high-level features that are more conducive to classification and can retain the original features, while also having the advantages of the SVM classifier.

Figure 1 is a flow chart of using a combined algorithm to achieve emotional classification of folk music. As can be seen from the figure, the function of the algorithm is divided into two categories: feature extraction and classification. First, input the extracted folk music features into DBN. Each RBM in the DBN network is used to learn the original features layer by layer, and the output of the last hidden layer of the DBN is extracted as the audio feature for classification. Then, the classifier is trained by using the training samples corresponding to the feature samples extracted by DBN. Then, use the trained SVM classifier to test the test sample data, and finally get the accuracy of the classification.

In the research on the recognition and classification of folk music emotion based on the combined algorithm of DBN and SVM, the main factors that affect the classification result are (1) input feature parameters in the DBN network; (2) features extracted by the DBN network structure model; and (3) the training of the classifier when the SVM classifier implements mult-classification. The classifier used in the study of music sentiment classification in this chapter is still SVM, so the next step is to study the DBN network structure model.

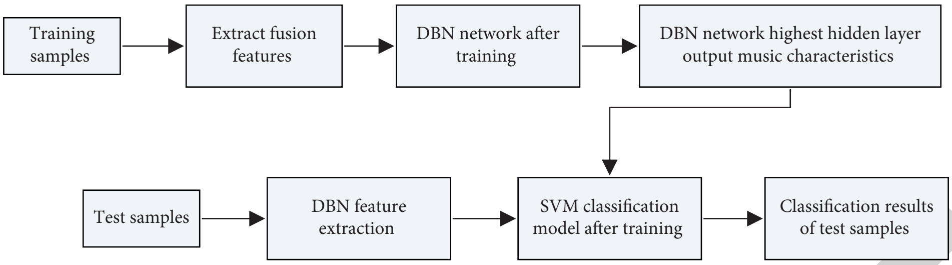
Training samples
Extract fusion features
DBN network after training
DBN network highest hidden layer output music characteristics
Test samples
DBN feature extraction
SVM classification model after training
Classification results of test samples

Figure 1: The flow chart of DBN and SVM combined algorithm to achieve classification.

3.1.2. The Training Process of DBN. In extracting the features of the audio signal using the deep belief network, the fusion feature of folk music fragments is used as the original feature input of the deep belief network, and the original input feature is only this one. When it is input into the entire network through the input layer, through the learning of the first RBM, the output of the hidden layer is used as the input of the second RBM visible layer. The hidden layer is mainly to ensure that the input feature vector is mapped to different feature spaces to retain as much feature information as possible. By decomposing and reconstructing the output features of the previous layer, more abstract new features can be used to realize the characterization of the original features. After multiple RBM learning, the output of the last hidden layer is the feature extracted by the DBN network. Each node of the hidden layer in the network uses the output of the visible layer as the input of the current layer. The activation function can be used to express the functional relationship between the input and output of the two layers of nodes. In this chapter, the sigmoid function is mainly used in the construction of the DBN structure model. The function expression is as follows:

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{1}$$

The training process of the deep belief network is mainly divided into the unsupervised pretraining process of forward propagation and the supervised reverse tuning process of back propagation. In the experiment using DBN to achieve feature extraction, pretraining is to implement unsupervised training for each RBM network separately. And it is necessary to ensure that after the original features of folk music have been mapped many times, the final output abstract high-level features can still contain as many original features as possible.

In fact, the most important thing in RBM training is to find the increments of weights and biases. The specific pretraining process of RBM is as follows:

Step 1. Initialize the weight W and bias a and b. Among them, both a and b are set to 0, and the initial value of W is close to 0; set the initial learning rate of the network, and set it to 0.001 based on experience.

Step 2. Input the characteristics of various music emotion sample data into the DBN network. The input of the experiment in this chapter is the fusion feature extracted in this article.

Step 3. The input original music feature is used as the input of the visible layer of the DBN network, and the input data sample is used as the initial state of the visible layer.

Step 4. According to the initial state of each node in the visible layer, the binary state of all hidden layer nodes can be obtained through the forward propagation of the network.

Step 5. When the state of each node in the hidden layer is determined, calculate the binary state of the visible layer.

Step 6. Steps 4 and 5 are a reconstruction of the visible layer, using the log-likelihood function to obtain partial derivatives of the weights and biases to obtain the corresponding likelihood gradients.

Step 7. Calculate the increment of the weight and the offset to update the corresponding offset value $m' = m + \Delta m, n' = n + \Delta n$, weight in the DBN network $W' = W + \Delta W$.

DBN training is based on the RBM training process. When the RBM in the network is trained layer by layer from bottom to top to the highest layer, then the optimal bias and weight corresponding to each RBM will be obtained, but this kind of optimal is only a local optimal. If the optimization of each RBM is only considered separately, and the overall network is not considered, the final feature extraction may fall into a local optimal situation. In the process of extracting the characteristics of folk music, in the reverse tuning part, according to the label information of the training samples, this paper uses the network weights and bias optimal values obtained in the pretraining stage as input to achieve self-topping.Finally, by comparing the size of the error between the target output and the actual output, according to the predetermined error range, it is determined whether the tuning should be terminated.

3.2. Classification Model Based on Improved DBN and SVM. Similar problems often arise when using deep learning network models for classification testing. That is, when the classification test is performed on the data test sample, the prediction accuracy of its classification is very low. However, in the model training process, the prediction accuracy of the data training sample classification is actually relatively high.

Considering the fact that it may be due to the overfitting problem caused by the lack of training data samples during the training process, overfitting is a common problem in many machines learning. This article is in the process of using the classification model based on the combination of DBN and SVM to classify folk music. Since the RBM in the DBN structure model can prevent overfitting of training to a certain extent, the overall classification result is not bad. However, consider the fact that the overfitting problem may make the trained model almost ineffective. In order to further prevent overfitting and alleviate the problem of low prediction accuracy brought by overfitting to the network, in this paper, a “Dropout” layer is added to each RBM in the DBN network to make the model’s generalization ability stronger. When using neural networks to solve practical problems, they often encounter the problem of long network training time. In the process of using DBN to extract the abstract features of folk music, this paper frequently encounters the problems of long training time and slow convergence speed. In order to alleviate the network pressure caused by time complexity, the momentum coefficient is introduced.

3.2.1. Dropout Layer. “Dropout” means that when the network model is being trained, some nodes are not functioning and do not participate in the learning process of features, and we can temporarily think that these non-functioning nodes do not exist. Because they are only considered temporarily that they do not exist, these nodes may play a role in subsequent training, so the weight matrix corresponding to these nodes still needs to be retained. Simply put, it is to stop certain neural nodes from working during the forward propagation stage of the neural network, which avoids the network from relying too much on some local features and effectively avoids the problem of overfitting. “Dropout” is currently used extensively in fully connected networks. In this article, when improving the DBN model, in each RBM of the DBN, “Dropout” is added between the visible layer and the hidden layer. Compared with the previous network, after adding “Dropout,” the training process of DBN will have a little change. Suppose the probability of the node working after joining “Dropout” is α. Then, in the training model stage, first, randomly select $(1-\alpha)$ times the node in the visible layer and temporarily delete it. Then, the input data samples are forwarded and back-propagated in the network; this step is the same as the training process of RBM. Then, update the weights and biases corresponding to the neural nodes that are not deleted, and finally restore the deleted neural nodes.

Figure 2 shows the corresponding node information of the visible layer and the hidden layer in the forward propagation process of the network before and after adding “Dropout.” It can be seen that before adding “Dropout,” the calculation formula for the visible layer to transfer the output value to the hidden layer is as follows:

$$\begin{aligned} p_i^{(n+1)} &= W_i^{(n+1)} q^{(n)} + c_i^{(n+1)}, \\ q_i^{(n+1)} &= f\left(p_i^{(n+1)}\right). \end{aligned} \tag{2}$$

By the adoption of “Dropout” network, the calculation formula becomes

$$\begin{aligned} \tilde{q}^n &= s^{(n)} * q^n, \\ p_i^{(n+1)} &= W_i^{(n+1)} \tilde{q}^n + c_i^{(n+1)}, \\ q_i^{(n+1)} &= f\left(W_i^{(n+1)} \tilde{q}^n + c_i^{(n+1)}\right), \end{aligned} \tag{3}$$

where n represents the number of layers of the network, which is the visible layer in the network of this article. $n+1$ denotes the hidden layer. $q_i^{(n)}$ is the node of the visible layer, $\tilde{q}_i^{(n)}$ is the node with the addition “Dropout”; $s^{(n)}$ is a random 0, 1 vector with “Dropout” added, and the value of the vector indicates the existence state of the corresponding node. Generally, it is generated by Bernoulli’s function $s_j^{(n)} \sim$ Bernouli $p(\alpha)$, where α is the probability of neural nodes working in the network. In this article, the value of p after adding “Dropout” is mainly obtained based on experience and experiments. The working probability p for the input layer node is set to 0.9, and the working probability for other hidden layer nodes is set to 0.5.

3.2.2. Momentum Coefficient. The momentum coefficient is mainly used to update the weight and bias in DBN. In RBM, the weight and bias are determined by first setting the initial state of the visible layer. After forward and reverse calculations, the binary states of all nodes in the hidden layer and the visible layer are obtained, and the visible layer is reconstructed. Then, by maximizing the log-likelihood function, the partial derivatives of the weights and biases are obtained, respectively, and the likelihood gradients corresponding to the weights and the biases are obtained. Finally, use formulas to calculate the increments of weights and biases to complete the update. In the update of the relevant parameters W, m, and n of the DBN network, this paper increases the dependence on the offset increment and weight increment produced by the previous iteration. The improved formula for weight and bias is as follows:

$$\begin{aligned} \Delta W_{ij(t+1)} &= \eta\left(x_i y_{j\text{data}} - x_i y_{j\text{model}}\right) + \lambda \Delta W_{ij(t)}, \\ \Delta m_{(t+1)} &= \eta\left(x_{i\text{data}} - x_{i\text{model}}\right) + \lambda \Delta m_{i(t)}, \\ \Delta n_{j(t+1)} &= \eta\left(y_{j\text{data}} - y_{j\text{model}}\right) + \lambda \Delta n_{j(t)}, \end{aligned} \tag{4}$$

where t represents the t-th iteration, which, respectively, corresponds to the tth increment of the weight and bias. η is the learning rate, and λ is the weight corresponding to the increase dependence, which is in the range of 0 to 1. Because every time the weight and bias are updated, it will affect the performance of the network. By increasing their dependence on the original increment, the weights and biases can be accelerated to change in the direction they changed in the

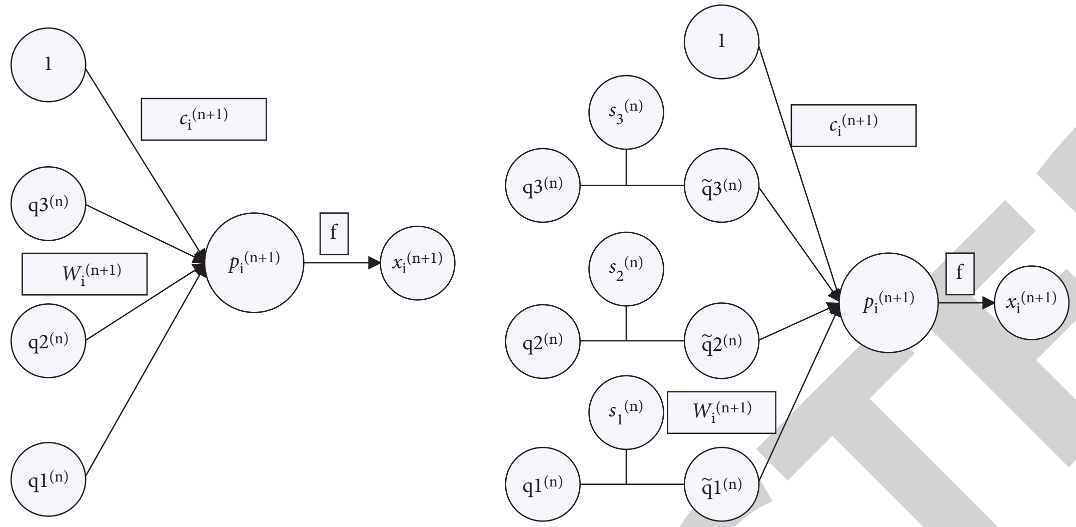

Figure 2: Corresponding information before and after adding the “Dropout” node.

Table 1: Specific information of the music clips used in the experiment.

Number	Music emotion category	Number of training clips	Number of test clips	Number of verify clips
1	Happy	600	200	200
2	Woeful	600	200	200
3	Eager	600	200	200
4	Peaceful	600	200	200

previous iteration. These speeds up the convergence process, thereby alleviating the problem of long network training time. When improving DBN in this paper, it is found through parameter tuning that when λ is set to 0.1, the overall performance of the network is improved, so this paper sets the weight λ to 0.1.

4. Experiment and Analysis

In this experiment, in order to ensure the fairness of classification, 1000 pieces of music are randomly selected for each kind of music emotion for experiment. Table 1 lists the number of music fragments used in the study of music sentiment classification in this article.

The emotion model in this paper uses participation and value orientation to describe emotions as a whole. The music emotion is divided into four parts, and adjectives are added to each part to describe the corresponding emotion. But just using these 3 words cannot mark all the musical emotions. In order to describe music emotions more comprehensively, four similar adjectives with the same dimension range are added to each category. These adjectives are all from the Hevner’s emotional ring model. See Table 2 for specific categories and descriptions.

The experimental results directly compare the classification performance of each classification algorithm under different feature extraction methods; that is, analyze the recognition and classification effect and misjudgment rate of the corresponding music emotion category under different algorithms.

This chapter mainly studies the classification of music emotion based on DBN extracted features. And through the improvement of DBN, the classification effect of the combined algorithm of DBN and SVM is improved. Therefore, the experiment mainly discusses from two aspects of the original feature extraction method and the classification algorithm. The original feature extraction methods mainly include MFCCs feature parameters and the fusion feature parameters proposed in this paper. The classification algorithm involves the combination of SVM, DBN and SVM, and the improved combination of DBN and SVM. Through the combination of features and classification algorithms, comparative experiments are realized. Table 3 shows the comparative experiments of using different classification algorithms to achieve classification under different feature extraction methods.

It can be seen from the table that when the classifiers are all SVMs, the accuracy of using MFCCs feature parameters to identify and classify folk music emotions is only 52.5%, which is because of the diversity of music features. If only considering the problem of music emotion classification from the inverse frequency domain of music, the extracted music features are relatively monotonous and do not have good distinguishability. In contrast, the abstract features obtained by learning the feature parameters of MFCCs using deep belief networks can more accurately recognize music emotions. This is because DBN can learn independently and extract abstract features that can better distinguish the emotions of ethnic music.

Table 2: Descriptive words corresponding to text sentiment classification.

Number	Music emotion category	Descriptive words
1	Happy	Longing, awe-inspiring, lyrical
2	Woeful	Heavy, dark, miserable
3	Eager	Satisfied, calm, gentle
4	Peaceful	Passionate, humored, victorious

Table 3: Classification accuracy of different classification algorithms.

Extract features	MFCCs (%)	Fusion feature (%)
SVM	52.5	58.6
DBN + SVM	68.3	76.5
Improved DBN + SVM	72.1	79.4

When the classification algorithm combining DBN and SVM is used, and the fusion feature is input into the DBN network as the original feature, the classification accuracy rate reaches 76.5%. This not only proves the ability of DBN to learn features, but also shows that the fusion features proposed in this paper can be effectively used for the classification of folk music emotions. Particularly, after making improvements to DBN, when the fusion feature is used as the network input, the classification accuracy has increased by 2.9%. Mainly because this article adds Dropout to each RBM, overfitting can be prevented, and the update method of weights and biases is changed, which makes their update direction clearer, improve the efficiency of determining weights and biases, and realize the optimization of DBN.

Figure 3 is based on MFCCs feature parameters, using SVM to achieve the classification accuracy and error of folk music emotion recognition classification. It can be seen from the figure that only using MFCCs features for classification, the accuracy of music sentiment classification is very low, and the classification misjudgment rate is mostly relatively high. In particular, the misjudgment rate of music clips that peaceful emotion is up to 66%, and the overall best recognition and classification rate is only 62%.

Figure 4 is the classification result of folk music emotion recognition classification based on fusion features. It can be seen from the figure that compared to extracting MFCCs features for classification, the overall accuracy of classification using fusion features has increased, and the misjudgment rate of emotion categories has also decreased. The overall best classification accuracy is for music samples in an eager mood, with a classification accuracy of 68%.

Figure 5 shows the classification accuracy and error of the folk music emotion recognition classification based on the MFCCs parameter characteristics, using the classification algorithm combining DBN and SVM. It can be seen from the figure that compared to only using the SVM classification model for classification, the classification accuracy of the model that uses the DBN to learn features is improved. And it showed a trend of overall increase in classification accuracy and overall decrease in misjudgment rate. This shows that DBN has a high learning ability for the original feature data of folk music input into the network.

Figure 6 is based on improved DBN and SVM combined algorithm classification, using MFCCs feature parameters as the original network input, to achieve the classification accuracy and error of folk music emotion recognition classification. It can be seen from the figure that, compared to using the abstract features of DBN for classification, the accuracy of classification using the abstract features learned by the improved DBN network has not changed much overall. However, the classification accuracy of calm and desire emotions has improved.

Figure 7 is the classification accuracy and error of folk music emotion recognition classification based on the fusion features proposed in this paper and using the classification algorithm combining DBN and SVM. It can be seen from Figure 7 that, compared to only using the MFCCs feature parameters as the input features of the DBN network, the fusion feature more comprehensively represents the emotion of the national music itself. Particularly, for music clips with eager emotions, the classification accuracy rate reaches 83%, and the probability of misjudgment of eager emotions has decreased. And compared with other combinations of feature extraction and classification algorithms, the probability that music clips of other emotions are misjudged as anxiety emotions become lower.

Figure 8 is the classification accuracy and error of folk music emotion recognition classification based on the fusion features proposed in this paper and using the improved DBN and SVM combined classification algorithm. It can be seen from the figure that compared to the combination of other classification algorithms and features, when the fusion features proposed in this paper are used, the overall classification accuracy is the best through the algorithm based on the combination of improved DBN and SVM. Among them, the classification accuracy of calm emotions is the highest, reaching 87%, and the classification rate of desire emotions has also increased a lot. It further illustrates the efficiency of the fusion feature proposed in this paper and also proves that this paper improves the effectiveness of the DBN network by adding "Dropout" to the DBN network and using momentum coefficients to adjust the increment of weights and biases.

Through the overall analysis of the six simulation results, it is found that among all the classification models, the classification of folk music with eager emotions has the best

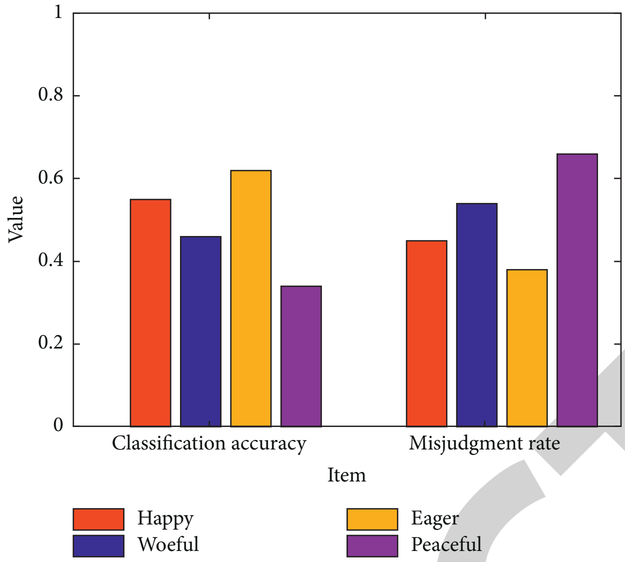

Figure 3: Comparison of classification results based on the characteristics of MFCCs.

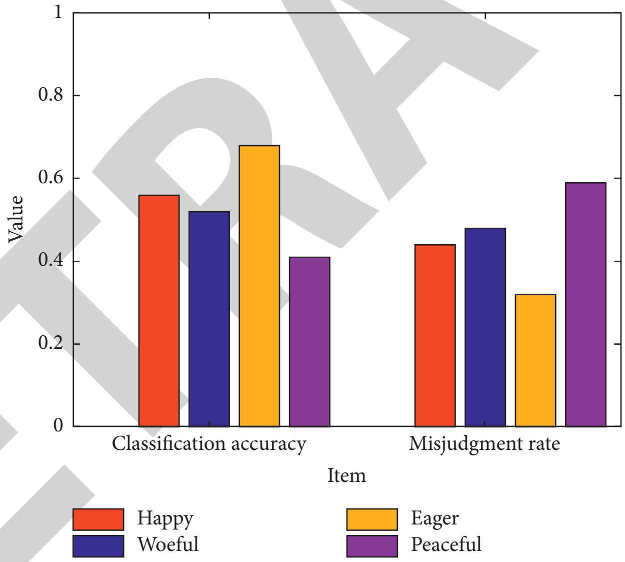

Figure 4: Comparison of classification results based on fusion features.

effect, and the classification of music with peaceful for emotions has the worst effect. The analysis shows that in the music emotion model of this article, peaceful represents a high degree of participation. But generally woeful or happy music can also indicate a high degree of participation to a certain extent. Therefore, there may be a slight deviation in the classification process of music fragments. And eager emotions represent low participation, and relatively speaking, the discrimination is high, so the overall classification accuracy is also high. Although the accuracy of music classification of peaceful for emotion is relatively low, the fusion feature is extracted, and then the fusion feature is used as the input of DBN for learning, and finally the DBN is improved. Every time the feature extraction method and the improvement of the classification algorithm are improved, the classification accuracy has been improved. It illustrates the superiority of using DBN to extract features to achieve classification, the effectiveness of improved DBN network, and the feasibility and efficiency of fusion feature classification.

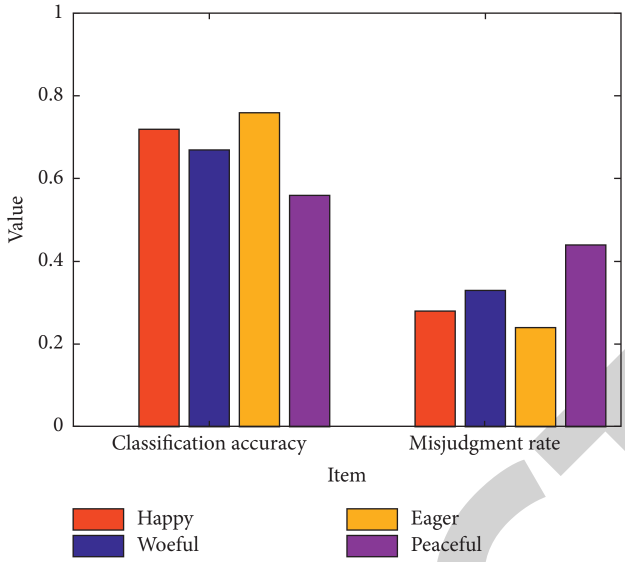

Figure 5: Classification result of MFCCs feature parameters used as DBN network input.

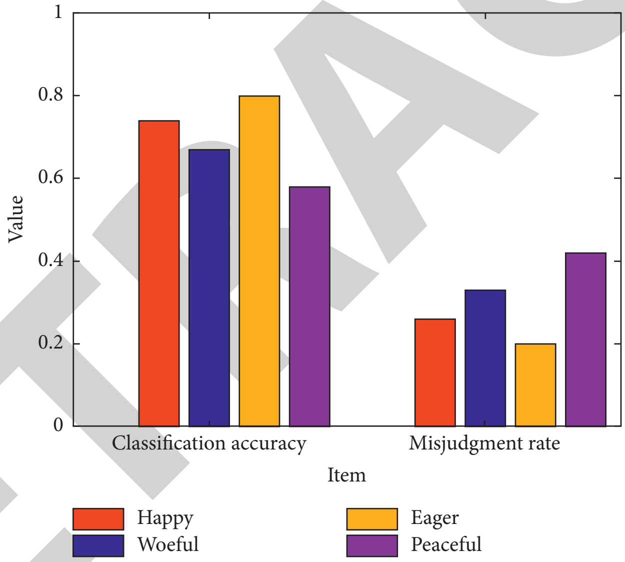

Figure 6: Classification result of fusion feature used as DBN network input.

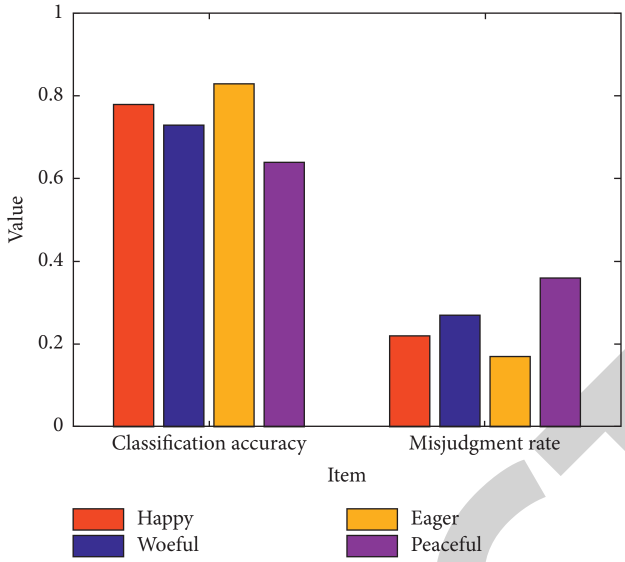

Figure 7: Classification result of MFCCs parameters used as improved DBN input.

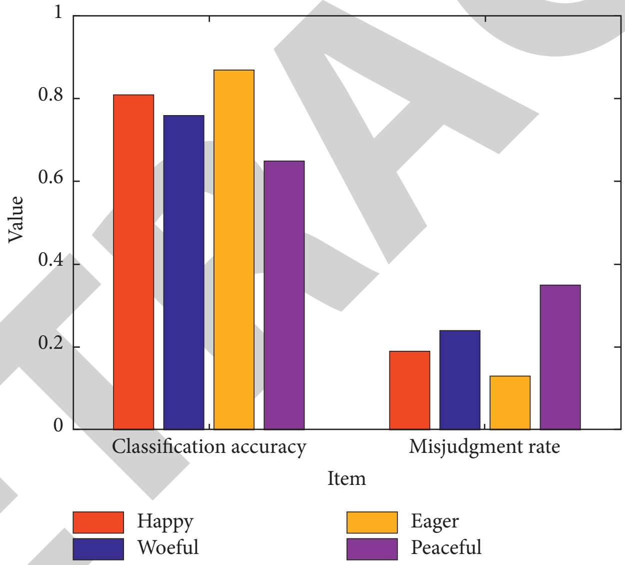

Figure 8: Classification result of fusion feature used as improved DBN network input.

5. Conclusion

With the steady development of the national economy, various art forms have also emerged in large numbers. As a result, the protection and inheritance of folk music has suffered a huge impact and faces the danger of loss. How to use science and technology to realize the digital protection and research of national music is a problem that needs to be solved urgently. On the other hand, with the maturity of deep learning related technologies, deep learning has been used in various aspects such as feature extraction, classification, and recognition, and achieved good results. This article tries to combine the research of folk music with deep learning knowledge. First, customize the fusion features of music and extract them, and use the features obtained from network learning as the input of the classifier to realize the emotional classification of ethnic music. The research of this article has realized the digital protection of national music resources to a certain extent and has achieved certain results in the research of emotion classification. Therefore, the study of folk music in this article is not only an attempt, but also a breakthrough. In the research on the classification of folk music emotions, this paper uses fusion features as network input, uses deep belief networks to extract features, and improves the update standard of weights and biases by adding Dropout to the restricted Boltzmann machine. Through experiments, it is found that the accuracy of ethnic music sentiment classification based on DBN extracted features is better than the classification results of traditional classification algorithms. And the use of improved DBN to extract features for classification not only ensures the improvement of classification accuracy, but also reduces the risk of overfitting. At the same time, using the fusion feature as the network input greatly improves the overall classification performance of the classification model, and finally the classification accuracy of folk music emotion reaches 79.4%.

Data Availability

The data sets used during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declares that he has no conflicts of interest regarding the publication of this paper.

References

[1] P. Vuilleumier and W. Trost, "Music and emotions: from enchantment to entrainment," *Annals of the New York Academy of Sciences*, vol. 1337, no. 1, pp. 212–222, 2015.

[2] T. Li, M. Ogihara, and Q. Li, "A comparative study on content-based music genre classification," in *Proceedings of the International Acm Sigir Conference on Research & Development in Information Retrieval*, ACM, Toronto, Canada, July 2003.

[3] E. Benetos and C. Kotropoulos, "Non-negative tensor factorization applied to music genre classification," *IEEE Transactions on Audio Speech and Language Processing*, vol. 18, no. 8, pp. 1955–1967, 2010.

[4] G. Wen, J. Tuo, and L. Jiang, "Audio feature extraction for classification using relative transformation," in *Proceedings of the International Conference on Audio*, Lund, Sweden, September 2012.

[5] Z. Fu, G. Lu, K. M. Ting, and D. Zhang, "A survey of audio-based music classification and annotation," *IEEE Transactions on Multimedia*, vol. 13, no. 2, pp. 303–319, 2011.

[6] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[7] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[8] R. O. Duda and P. E. Hart, *Pattern Classification*, Wiley, New York, NY, USA, 2000.

[9] F. Morchen, A. Ultsch, M. Thies, and I. Lohken, "Modeling timbre distance with temporal statistics from polyphonic music," *IEEE Transactions on Audio Speech and Language Processing*, vol. 14, no. 1, pp. 81–90, 2006.

[10] J. Shen, J. Shepherd, and B. Cui, "A novel framework for efficient automated singer identification in large music databases," *ACM Transactions on Information Systems*, vol. 27, no. 3, pp. 181–211, 2009.

[11] A. Meng and J. Shawe-Taylor, "An investigation of feature models for music genre classification using the support vector classifier," in *Proceedings of the International Conference Music Information Retrieval*, London, UK, September 2005.

[12] I. Mierswa and K. Morik, "Automatic feature extraction for classifying audio data," *Machine Learning*, vol. 58, no. 2-3, pp. 127–149, 2005.

[13] T. Li, M. Ogihara, and Q. Li, "A comparative study on content-based music genre classification," in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, Toronto, Canada, July 2003.

[14] C. H. Chang-Hsing Lee, J. L. Jau-Ling Shih, K. M. Kun-Ming Yu, and fnm Hwai-San Lin, "Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features," *IEEE Transactions on Multimedia*, vol. 11, no. 4, pp. 670–682, 2009.

[15] J. Marques and P. J. Moreno, *A Study of Musical Instrument Classification Using Gaussian Mixture Models and Support Vector Machines*, Vol. 4, Compaq Corporation, Cambridge Research laboratory, , Cambridge, UK, 1999.

[16] X. Shao, C. Xu, and M. S. Kankanhalli, "Unsupervised classification of music genre using hidden markov model," *Unsupervised classification of music genre using hidden Markov model*, 2004.

[17] L. Wyse, "Audio spectrogram representations for processing with convolutional neural networks," 2017, https://arxiv.org/abs/1706.09559.

[18] T. Kim, J. Lee, and J. Nam, "Sample-level CNN architectures for music auto-tagging using raw waveforms," *IEEE International Conference Acoustics, Speech, and Signal Processing*, vol. 1, pp. 366–370, 2018.

[19] K. Choi, G. Fazekas, and M. Sandler, "Automatic tagging using deep convolutional neural networks," 2016, https://arxiv.org/abs/1606.00298.

[20] S. Dieleman and B. Schrauwen, "End-to-End learning for music audio," *IEEE Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6964–6968, 2014.

[21] K. He, X. Zhang, and S. Ren, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition*, IEEE Computer Society, Seattle, WA, USA, August 2016.

[22] G. Huang, Z. Liu, and V. Laurens, "Densely connected convolutional networks," 2016, https://arxiv.org/abs/1608.06993.

[23] T. Kim, J. Lee, and J. Nam, "Sample-level CNN architectures for music auto-tagging using raw waveforms," 2017, https://arxiv.org/abs/1710.10451.

[24] K. Choi, G. Fazekas, and M. Sandler, "Transfer learning for music classification and regression tasks," 2017, https://arxiv.org/abs/1703.09179.