Research on Recognition and Classification of Folk Music Based on Feature Extraction Algorithm

Xi Wang
Henan Polytechnic, Zhengzhou, Henan 450046, China
E-mail: xi33n9@yeah.net

Keywords: folk music, feature extraction, music classification, support vector machine

Received: December 8, 2020

In this study, the feature extraction algorithm for folk music was analyzed. The features of folk music were extracted in aspects of time domain and frequency domain. Then, a support vector machine (SVM) was selected to identify and classify folk music. It was found that the performance of SVM was the best when σ^2 was 26 and C was 4; the recognition rate of using only one feature was inferior to that of using all features; the highest recognition rate of SVM was 92.76%; compared with back propagation neural network (BPNN) and decision tree classification method, SVM had a higher recognition rate. The experimental results show the effectiveness of SVM, which can be applied in practice.

Povzetek: V tem študentskem članku je predstavljena klasifikacija glasbe s pomočjo metod umetne inteligence.

1 Introduction

As an art form, music can express people's thoughts, feelings, and life style and has a role in promoting people's emotion and spirit. With the improvement of human living standards, music has become more and more popular. With the development of science and technology, more and more people have tended to enjoy music through the Internet. Therefore, finding out music which users want to listen to from a massive amount of music has become more and more important, and the recognition and classification of music have attracted more and more extensive attention. Huang et al. [1] improved the hidden Markov model (HMM) using an artificial neural network (ANN). The application of the improved HMM in practical music classification found that HMM had a fast calculation speed but a poor classification performance, ANN had a good classification performance but a high computational complexity. The combination of them could improve the recognition rate of HMM by 4% - 5% while maintaining the same calculation speed as HMM. Abidin et al. [2] recognized a Turkish music data set, SymbTr, with ten machine learning algorithms, and found that the performance of the algorithms was between 82% and 88%. Rao et al. [3] studied chord recognition. Pitch Class Profile features were extracted from raw audio and recognized by spare representation. Through the experiment on MIREX09, it was found that the method had robustness to Gaussian white noise. Iloga et al. [4] studied the genre classification of music, designed a sequential pattern mining method, and carried out experiments on GTZAN. They found that the accuracy of the method was 91.6%, which was more than 7% higher than the existing classifiers. Chinese folk music refers to the music played by traditional instruments, which has high artistry and nationality [5], but there is little research on its recognition and classification. Therefore, this study took folk music as the research subject, carried out feature extraction in aspects of time domain and frequency domain, established a feature database, and then identified and classified folk music with a support vector machine (SVM), and verified the reliability of the method through experiments. The present study contributes to the realization of the automatic classification of folk music and the improvement of retrieval efficiency.

2 Folk music and feature extraction

2.1 Folk music

Folk music includes instrumental music, songs, opera, etc. Musical instruments play a very important role in folk music, which can be divided into four categories, as shown in Table 1.

Wind instruments	Xiao, Suona, Lusheng, Xun, pan flute, etc.
Plucked stringed instrument	Chinese lute, moon lute, Guqin, kayagum, Zheng, Konghou, etc.
Percussion instruments	Collected bronze bells, wooden fish, bronze drum, long drum, gong, etc.
String instruments	Erhu, Xiqin, horse head string instrument, Leiqin, etc.

Table 1: Folk music instruments.

Musical instruments can be solo or ensemble, and different combinations of musical instruments will form different styles of instrumental music. For example, the music played by percussion instruments has a strong

rhythm and rich timbre; the music performed with string instruments has a delicate style and simple and elegant style; the music played with wind instruments, and string instruments tends to be light and lively; the music played with wind instruments and percussion instruments is joyful and enthusiastic.

2.2 Music feature extraction

To identify and classify folk music, it is necessary to extract the features of folk music. Music is composed of many monosyllables. In psychology, sound includes the following four characteristics:

(1) pitch: pitch refers to people's feeling of the frequency of sound, determined by the number of vibration of an object;
(2) sound duration: sound duration refers to the duration of a note, which is determined by the duration of the vibration;
(3) sound intensity: sound intensity refers to the loudness that people feel, which is determined by the vibration amplitude;
(4) timbre: timbre refers to people's perception of sound quality, which is determined by the material, structure,. and shape of the sound body.

In the recognition of folk music, timbre is the main feature because music is played by different instruments. Timbre is a short-term feature, which can be extracted from the following three aspects.

2.2.1 Time-domain characteristics

Time-domain characteristics aim at the characteristics of the audio signal waveform. The time-domain features selected in this study are as follows.

(1) Short-time average energy (STE): it is used for reflecting the change of music signal amplitude. It refers to the average energy of the signal in the short-term audio window. For a short-time frame with a window length of N, suppose that the signal value of the n -th sampling point is $x(n)$, the window function is represented by $w(n-m)$. For the m -th frame, its STE can be expressed as:

$$E(m)=\frac{1}{N}\sum_{m}\big(x(n)w(n-m)\big)^2$$

(2) Zero crossing rate (ZCR): it refers to the number of times a signal waveform passes through the zero point in a frame. For the m -th frame, its ZCR can be expressed as:

$$ZCR(m)=\frac{1}{2}\sum_{m}|sgn[x(n)]-sgn[x(n-1)]|w(n-m)$$

where sgn stands for the sign function,

$$sgn=\begin{cases}1, x(n)\geq 0\\ 0, x(n)<0\end{cases}$$

2.2.2 Frequency-domain characteristics

Audio contains a lot of information, which needs to be obtained in the frequency domain analysis. The frequency domain features can be obtained by converting the signal to the frequency domain through Fourier transform. The features selected in this study are as follows.

(1) Spectrum centroid (SC): it refers to the characteristic quantity of the spectrum center of a signal. Fourier transform is represented by $F(\delta)$, $\delta\in(g,h)$, and the maximum and minimum values of frequency are represented by g and h respectively. Then SC can be expressed as:

$$SC=\frac{\sum_{\delta=g}^{h}\delta|F(\delta)|^2}{\sum_{\delta=g}^{h}|F(\delta)|^2}$$

(2) Spectrum energy (SE): it refers to the frequency domain energy of the signal, which can be expressed as:

$$SE=\sqrt{\frac{1}{h-g}\sum_{\delta=g}^{h}|F(\delta)|^2}$$

(3) Mel frequency cepstrum coefficient (MFCC) [6]: it refers to the cepstrum characteristics at Mel frequency, which has 13 dimensions. Suppose that the frequency of the music signal is f, then its Mel frequency is:

$$f_{mel}=2595\times log_{10}\left(1+\frac{f}{700}\right).$$

3 Support vector machine-based classification algorithm

SVM is a machine learning method [7], which has significant advantages in a small sample and nonlinear field and has been successfully applied in many fields, such as speech recognition [8] and image classification [9].

Suppose that in Euclidean space R^d, the training sample is $\{(x_1,y_1),(x_2,y_2),\ldots,(x_N,y_N)\}$ $(y\in\{+1,-1\})$, the linear discriminant function is $g(x)=wx+b$, and the classification plane equation is $wx+b=0$, where w refers to the hyperplane normal vector, and b refers to the offset. To separate the samples correctly, the problem can be expressed as:

$$min\frac{1}{2}\|w\|^2$$
$$y_i[(wx_i)+b]-1\geq 0$$

In the case of inseparable linearity, relaxation variable λ and penalty factor C are introduced. Then the above equation is transformed into:

$$min\frac{1}{2}\|w\|^2+C\sum_{i=1}^{N}\lambda_i$$
$$y_i[(wx_i)+b]\geq 1-\lambda_i$$

The Lagrange function is introduced to solve the above equation. Lagrange coefficient is set as a_i, then the optimal classification function is:

$$f(x)=sgn\left\{\sum_{i=1}^{N}a_iy_ik(x_i,x)+b\right\}$$

For any unclassified sample x, the result of classification can be obtained by calculating $f(x)$. $k(x_i, x_j)$ represents the kernel function. In SVM, the commonly used ones are:

(1) linear kernel function: $K(x_i, x_j) = x_i \cdot x_j$;

(2) polynomial kernel function: $K(x_i, x_j) = [(x_i \cdot x_j) + 1]^d$, where d is an adjustable parameter;

(3) RBF kernel function: $K(x_i, x_j) = exp\left(-\frac{\|x_i - x_j\|^2}{\sigma}\right)$, where σ is an adjustable parameter.

In SVM, the RBF kernel function is the most commonly used and has the best performance; therefore, this study uses RBF kernel function. In SVM, the values of kernel function parameter σ and penalty parameter C have a great influence on the results [10], which needs to be determined in the experiment.

4 Experimental analysis

4.1 Folk music data set

The folk music was downloaded from the Internet and then converted to the WAV format of a single channel with a sampling frequency of 16 KHz by GoldWave software. The music file was processed by slicing by CoolEdit software and divided into 10 s segments. The final data sets obtained are shown in Table 2.

Song	Types of folk music	Number
"Notturno in the Fisherboat", "Jackdaw Playing in the Water"	Zheng	116
"Ambush on All Sides", "Zhaojun Going Out of the Frontier"	Chinese lute	121
"Lofty Mountains and Flowing Water", "White Snow In Sunny Spring "	Guqin	164
"Journey to Suzhou", "Partridges Flying"	Bamboo flute	138
"Hundreds of Birds Worshipping the Phoenix", "A Flower "	Suona	97
"The Moon Over a Fountain", "The Song of Burying Flower"	Erhu	167

Table 2: Data sets of folk music.

Features were extracted from the obtained data set, including 13-dimensional MFCC features and four one-dimensional features. The average value and standard deviation were taken, then each segment obtained 36-dimensional features. Then 80% of the features were selected as the training set, and 20% as the testing set.

4.2 Experimental results

Firstly, two parameters of SVM need to be determined. Two hundred of samples were selected. and determine the value of parameters through the cross test, as shown in Tables 3 and 4.

C	Recognition rate/%
2^{-1}	90.11
2	91.23
2^2	93.87
2^3	92.18
2^4	92.09
2^5	91.63
2^6	91.29
2^7	90.88
2^8	90.64
2^9	89.72
2^{10}	88.33

Table 3: The influence of the value of C on the recognition rate when the value of σ^2 takes 2.

σ^2	Recognition rate /%
2^0	88.64
2^1	89.72
2^2	90.07
2^3	91.22
2^4	92.08
2^5	93.09
2^6	93.87
2^7	93.06
2^8	92.18
2^9	91.53
2^{10}	90.27

Table 4: The influence of the value of σ^2 on the recognition rate when the value of C takes 4.

It was seen from Tables 3 and 4 that the recognition rate of SVM was the highest when $C = 4$ and $\sigma^2 = 2^6$. Therefore, $C = 4$ and $\sigma^2 = 2^6$ were selected as the optimal parameters for the experiment.

The influence of feature selection on the results was compared. The selected features were time-domain, frequency-domain, and time + frequency-domain features of folk music. The results are shown in Figure 1.

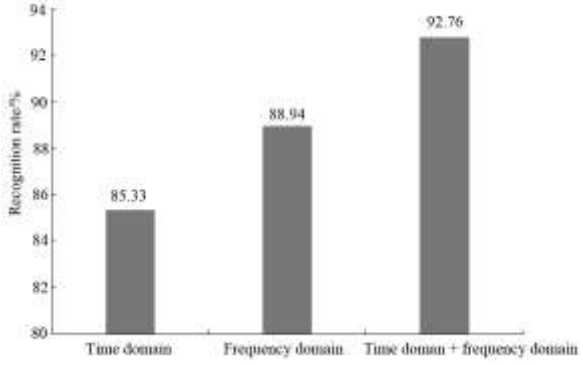

Figure 1: The influence of feature selection on the recognition rate.

It was seen from Figure 1 that the recognition rate of SVM was 85.33% when only the time domain features of folk music were selected and was 88.94% when only the frequency domain features were selected, and the increase of 4.23% might be due to the more feature dimensions contained in the frequency domain; when all the features were used for recognition, the recognition rate of SVM was 92.76%, which was 8.7% and 4.3% higher than the time domain and frequency domain. It was found that the recognition effect of SVM was good when all the features were used.

The recognition performance of SVM for different types of folk music is compared, and the results are shown in Figure 2.

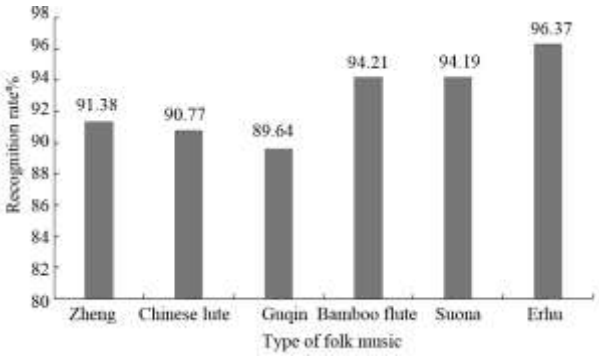

Figure 2: Recognition effect of different types of folk music.

It was seen from Figure 2 that SVM had the highest recognition rate for erhu, 96.37%, which might be because there was only one kind of string instrument, i.e., erhu, in the folk music data set studied in this study, which was significantly different from other types of folk music. The recognition rate of SVM was 91.38%, 90.77%, and 89.64% for Zheng, Chinese lute, and Guqin, which might be because the three instruments were slightly similar and more difficult to recognize.

To further verify the recognition performance of SVM, BP neural network (BPNN) [11], decision tree [12], and SVM were compared by the same folk music data set. The results are shown in Figure 2.

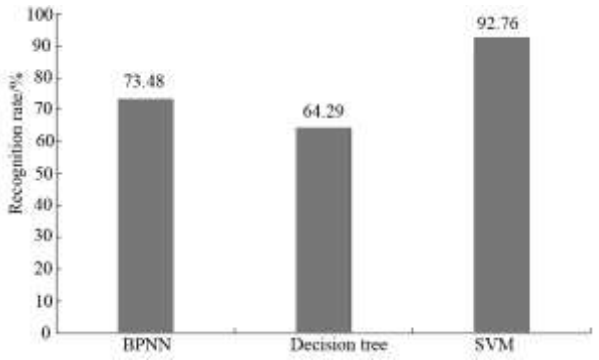

Figure 3: Comparison of recognition effects of different algorithms.

It was seen from Figure 3 that the recognition rates of the three algorithms were 73.48%, 64.29%, and 92.76%, respectively, and the recognition rate of SVM was 26.24% higher than that of BPNN and 44.28% higher than that of the decision tree. The results showed that SVM had significant advantages in the classification and recognition of folk music.

5 Discussion

The current research on music recognition and classification includes the classification of genres [13], musical instruments [14], emotions [15], composers, and so on. Through the identification and classification, users can quickly and accurately retrieve the music they want to hear, and it is also more convenient to manage the music. With the development of technology, music recognition and classification has made great progress, and more and more machine learning methods have been applied, such as hidden Markov, decision tree, nearest neighbor, etc. [16]. In this study, SVM was used for classifying folk music.

In the identification and classification of folk music, this study extracted the time-domain and frequency-domain features to form the folk music data set and then used the SVM method for classification. In the experiment, to obtain the optimal parameters of SVM, this study analyzed the influence of different values on the results by the cross-check method, and then the obtained optimal parameters were used for the next step of the experiment. The results showed that the recognition rate of SVM was higher when more comprehensive features were selected. In folk music recognition, when using time-domain and frequency-domain features, the recognition rate of SVM reached 92.76%. In recognizing different types of folk music, the recognition rate of SVM for erhu was the highest (96.37%), while the recognition rates of three plucked instruments were relatively low. In comparison with other methods, this study selected BPNN and decision tree for comparison. It was seen from Figure 2 that the recognition rate of SVM used in this study was significantly higher than the other two methods, which indicated that SVM had a better performance in the recognition of folk music.

Although some achievements have been made in this paper, further research is needed. In future work, we will:

(1) further study the selection of features;
(2) further improve the classification performance of SVM;
(3) perform experiments on a more extensive data set.

6 Conclusion

In this study, the method of feature extraction was analyzed for the recognition and classification of folk music, SVM was selected as the classifier, and a data set was established for experimental analysis. The results demonstrated that:

(1) the selection of parameters had an influence on the result of folk music recognition;
(2) when all the features were used, the recognition rate of SVM was the highest (92.76%);

(3) SVM had the highest recognition rate for erhu, reaching 96.37%;
(4) compared with BPNN and decision tree, SVM had a significantly higher recognition rate.

References

[1] Huang W, Zhang YT. (2020). Application of Hidden Markov Chain and Artificial Neural Networks in Music Recognition and Classification. *ICCDE 2020: 2020 The 6th International Conference on Computing and Data Engineering*, pp. 49-53.

[2] Abidin D, Özacar T, Ozturk O. (2018). Using classification algorithms for Turkish music makam recognition. 6, pp. 377-393. https://doi.org/10.15317/Scitech.2018.139

[3] Rao Z, Feng C. (2018). Sparse representation classification-based automatic chord recognition for noisy music. *Journal of Information Hiding and Multimedia Signal Processing*, 9, pp. 3400-409.

[4] Iloga S, Romain O, Tchuente M. (2018). A sequential pattern mining approach to design taxonomies for hierarchical music genre recognition. *Pattern Analysis & Applications*, 21, pp. 3363-380.

[5] Xie CY. (2015). *Research on the Development of National Music in the New Media Era*, International Conference on Education. Atlantis Press.

[6] Lalitha S, Geyasruti D, Narayanan R, Shravani M. (2015). Emotion Detection Using MFCC and Cepstrum Features. *Procedia Computer Science*, 70, pp. 329-35. https://doi.org/10.1016/j.procs.2015.10.020

[7] Abdiansah A, Wardoyo R. (2015). Time Complexity Analysis of Support Vector Machines (SVM) in LibSVM. *International Journal of Computer Applications*, 128, pp. 975-8887. https://doi.org/10.5120/ijca2015906480

[8] Bhavan A, Chauhan P, Hitkul, Shah RR. (2019). Bagged support vector machines for emotion recognition from speech. *Knowledge Based Systems*, 184, pp. 104886. https://doi.org/10.1016/j.knosys.2019.104886

[9] Gao L, Li J, Khodadadzadeh M, Plaza A. (2015). Subspace-Based Support Vector Machines for Hyperspectral Image Classification. *IEEE Geoscience & Remote Sensing Letters*, 12, pp. 349-353. https://doi.org/10.1109/LGRS.2014.2341044

[10] Rosales-Pérez A, Gonzalez J A, Coello CAC, Escalante HJ, Reyes-Garcia CA. (2015). Surrogate-assisted multi-objective model selection for support vector machines. *Neurocomputing*, 150, pp. 163-172. https://doi.org/10.1016/j.neucom.2014.08.075

[11] Wang J, Yan WQ. (2016). BP-Neural Network for Plate Number Recognition. *International Journal of Digital Crime & Forensics*, 8, pp. 34-45.

[12] Kumar R, Singh B, Shahani DT, Chandra A. (2015). Recognition of Power Quality events using S-transform based ANN classifier and rule based decision tree. *IEEE Transactions on Industry Applications*, 51, pp. 1249-1258.

[13] Costa YMG, Oliveira LS, Silla CN. (2017). An Evaluation of Convolutional Neural Networks for Music Classification Using Spectrograms. *Applied Soft Computing*, 52, pp. 28-38. https://doi.org/10.1016/j.asoc.2016.12.024

[14] Giannoulis D, Klapuri A. (2013). Musical Instrument Recognition in Polyphonic Audio Using Missing Feature Approach. *IEEE Transactions on Audio, Speech, and Language Processing*, 21, pp. 1805-1817. https://doi.org/10.1109/TASL.2013.2248720

[15] Bai J, Luo K, Peng J, Shi J, Wu Y, Feng L, Li J, Wang Y. (2017). Music Emotions Recognition by Machine Learning With Cognitive Classification Methodologies. *International Journal of Cognitive Informatics and Natural Intelligence*, 11, pp. 80-92. https://doi.org/10.4018/IJCINI.2017100105

[16] Nasridinov A, Park YH. (2014). A Study on Music Genre Recognition and Classification Techniques. *International Journal of Multimedia & Ubiquitous Engineering*, 9, pp. 31-42. https://doi.org/10.14257/ijmue.2014.9.4.04