Optimization of Computer Aided Design System for Music Automatic Classification Based on Feature Analysis

Minmin Ge[1]*, Yan Tian[2] and Yifan Ge[3]

[1]College of Preschool Education, Jiangsu Second Normal University, Jiangsu 210000, China, geminmin61@163.com
[2]College of Preschool Education, Jiangsu Second Normal University, Jiangsu 210000, China, tianyan33@163.com
[3]School of Mechanical Electronic & Information Engineering, China University of Mining and Technology Beijing, Beijing 100083, China, sqt1900405071@student.cumtb.edu.cn

Corresponding author: Minmin Ge, geminmin61@163.com

Abstract. The automatic classification of music is essentially a speech signal recognition problem, which has always been paid attention and researched by people. Although with the development of speech recognition technology, many new methods have been applied to the field of music classification, but due to the diversity and uncertainty of music, it is still far from large-scale practical applications. This paper systematically studies the learning methods based on feature analysis and the principles, methods and techniques of music classification, improves the existing algorithms based on feature analysis, and tries to apply the integration based on feature analysis to the study of music classification. Finally, A large number of numerical experiments and performance tests were performed on the proposed algorithm. In the experiment, the results of different classifiers for different feature sets were simulated. The simulation results not only verified the final accuracy of music classification after using feature-based analysis integration. It has been greatly improved, and it also shows the advantages of the method based on feature analysis over other classifiers in classification problems.

Keywords: Music classification；ensemble learning； feature extraction
DOI: https://doi.org/10.14733/cadaps.2022.S3.153-163

1 INTRODUCTION

In recent years, computational music has achieved many practical results. The rapid development of digital encoding, digital compression, and digital storage of electronic musical instruments and music signals has promoted the popularization and application of digital broadcasting and multimedia, and has shown broad market prospects [1]. The research on computational music-related issues, such as music information retrieval, digital music management, etc., has aroused great interest in academia and industry due to its important academic significance and wide application prospects. In the music information retrieval system, users not only use song titles and humming to retrieve, but also propose retrieval based on genre, style and other information. The basic method of mining different genre patterns in digital music is classification. Several scholars have made a lot of efforts in the field of music classification and have conducted preliminary research on various issues involved in music classification.

The music classification problem not only has the characteristics common to various pattern classification problems, it can use various classic machine learning algorithms widely used in other fields; it also has its own characteristics, such as feature extraction, feature selection process and classifier performance, etc. There are specific requirements [2]. The number of digital features that can be extracted from music data is huge, often reaching hundreds of dimensions. If all model classifications are introduced, the classification effect will be affected. Moreover, among the hundreds of dimensional features, there are not only related features, but also a large number of irrelevant and redundant features. Their existence seriously affects the performance of the classifier. How to effectively select features for a specific classification is an important problem in the study of music classification.

Feature selection is to select the feature subset that optimizes a certain evaluation criterion from the input feature set. Since scholars Jin and Yang [3] have conducted research on feature selection, but at that time it was mainly done from the perspective of statistics and information processing, and the problems involved are usually the number of features. not much. In machine learning theory, a lot of research has been carried out on learning algorithms, but there are relatively few researches on feature selection. The main reasons are shown in the following:

1) The performance of many algorithms is negatively affected by irrelevant or redundant features. Existing research results show that the number of training samples required by most learning algorithms increases exponentially with the increase of irrelevant features. Therefore, feature selection can reduce computational complexity and improve classification accuracy, and help to find an easier-to-understand algorithm model.

2) The continuous emergence of large-scale data processing problems. The so-called large-scale, on the one hand refers to the huge number of samples, on the other hand refers to the high dimension of the characteristics of the sample. The development of data mining puts forward urgent requirements for the research of large-scale data processing, such as information retrieval, genetic analysis, etc. This is already an empirical "axiom" in the field of machine learning. Therefore, feature selection algorithms are urgently needed to reduce the dimensionality of data, and the feature selection of high-dimensional data also affects the existing The feature selection algorithm presented a severe challenge.

The current music classification research, due to its own characteristics, not only requires the study of classifiers, but also strengthens the study of feature selection methods in this field. This topic is based on this idea to study the feature selection algorithm in music classification and apply the algorithm to genre classification for verification. In the experiment, the selected digital features have a certain correlation with the specific music classification problem, which can provide very valuable information for the analysis of digital features. At the same time, mature algorithms in the field of music classification can be extended to other fields of data mining. Therefore, the research of this topic has propriety in promoting the progress of music information retrieval technology and the research and improvement of algorithms of classification.

2 RELATED STUDIES

After entering the 1990s, the research work of content-based music classification has made great progress, and the number of relevant papers published internationally has increased significantly.

The current music classification research framework contains two key contents: feature extraction and classifier selection. There have been many papers in the field of music classification that have discussed and researched these two issues.

George is a representative of genre classification research. He gave the most commonly used feature sets and their extraction methods in the literature. Three feature sets representing timbre texture, melody content, and pitch content are used in the article. A comparative study was done on the performance of the software, and an accuracy rate of 61% was obtained, which became the most groundbreaking work. Therefore, the literature is widely cited by other papers.

Alam, F. [5] proposed a new feature, using the sub-band energy to calculate the feature value method, in the genre classification problem, reached 74% accuracy. Dardard, F [6] proposed a set of new features based on wavelet transform, using the three-dimensional parameters of wavelet histograms for music classification, and achieved good results in experiments combined with other feature values. In Ntalampiras, S, [7] genre data set among the classifications, the highest accuracy of 78.5% is reached. Albornoz, E.M et al. [8] used wavelet coefficients as features to automatically classify the three types of music signals of rock music, jazz and piano music, and compared the distinguishing capabilities of different wavelet construction methods and classifiers for these three types of problems.

For the research on the application of classifiers in the field of music classification, which gives the influence of different parameters and kernel functions in eigenvalue classification methods on genre classification. Azam and Gavrilova [9] studied how to use the minimum message length criterion to automatically classify rock music and classical music. In their work, they compared an unsupervised learning method called "Snob" with three supervised classification methods: decision trees, decision graphs, and artificial neural networks. The results show that the supervised classification method is significantly better than non-supervised classification methods in this problem. Supervise classification methods.

In general, the more music categories involved in the research, the more difficult the research. Its content covers most western modern music and a part of classical music. Among them, the optimal classification level with ten music categories can only reach 61% average classification accuracy rate. Therefore, for real needs, the accuracy of music classification needs to be further improved.

3 MUSIC CLASSIFICATION BASED ON FEATURE SELECTION

3.1 The Concept of Feature Selection and Classic Algorithms

The classic feature selection can be regarded as an optimization problem. The key is to establish an evaluation standard to distinguish which feature combinations help to classify which feature combinations are redundant, partially or completely irrelevant.

The feature selection method is divided into two types: filter and wrapper. The evaluation function has nothing to do with the classifier, and the wrapper uses the error probability of the classifier as the evaluation function. The Filter method only uses the data set to evaluate the relevance of each feature (subset), and does not consider the subsequent learning algorithm. The non-parametric methods can also be used to calculate mutual information containing continuous-valued features: if a parameter is given other features and the condition is independent of the class label, the parameter is considered an irrelevant feature. The Wappers method uses learning algorithms to evaluate the quality of each feature. Compared with the Filters method, the Wappers method usually has a large amount of calculation, but it can obtain a higher accuracy rate.

The size of the data set can be measured from two aspects: the number of features n and the number of samples p, n and p may be very large, and the largeness of n often causes problems such as Curse of Dimensionality. This is already an empirical "axiom" in the field of pattern recognition. The purpose of selection is based on some The criterion selects the smallest feature subset so that tasks such as classification and regression can approximate or even better the effect before feature selection. The audio preprocessing and feature extraction subsystem diagram is shown in Figure 1.

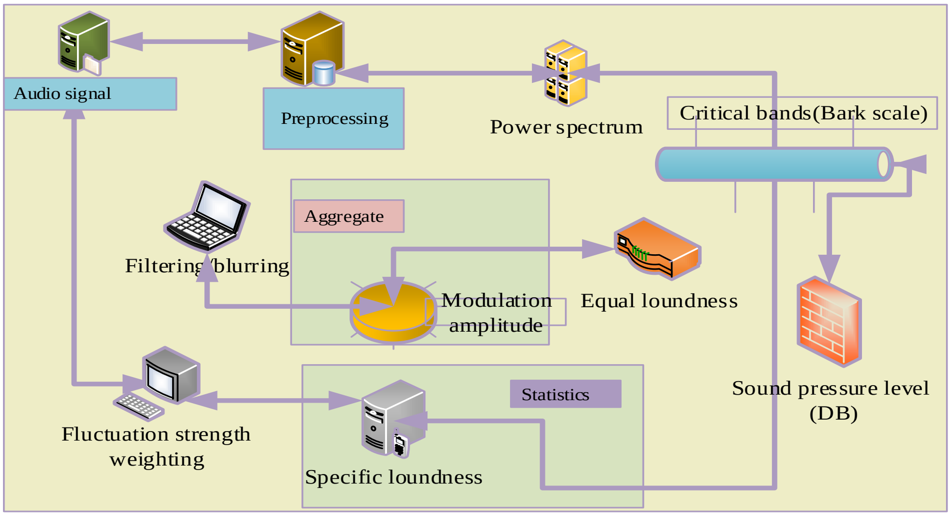

Figure 1: Audio preprocessing and feature extraction subsystem diagram.

3.2 The Basic Concept and General Process of Feature Selection

Feature selection definition can be expressed as: Given a learning algorithm, a data set , the data set *S* comes from an example space with *n* features $X_1, X_2, ..., X_n$, a category label *Y*, and an example space that conforms to the distribution *D*, X_{opt} is an optimal feature sub set which contains a feature subset that makes a certain evaluation criterion $C = C(L, S)$. It can be seen from the definition of feature selection that under the premise of a given learning algorithm, data set, and feature set, the definition of various evaluation criteria and the application of optimization techniques will constitute an important content of feature selection.

In practical applications, it is usually difficult to find the optimal feature subset. Many problems related to feature selection are found to be NP-hard problems. Finding the minimum feature subset that meets the requirements is an NP-complete problem. However, through the use of heuristic search algorithms, a good balance can be found between efficiency and quality, which is also the target of many feature selection algorithms.

For learning algorithms, effective feature selection can reduce the complexity of learning problems, improve the generalization performance of learning algorithms, and simplify learning models. In order to accurately estimate parameters or predict unknown samples, training samples increases exponentially with the number of features, which is the so-called Curse of Dimensionality. The performance of many learning algorithms is affected by irrelevant features and redundant features. For example, the nearest neighbor method needs to calculate the distance between samples, so whether it is redundant or irrelevant features, it will affect the distance calculation and will inevitably reduce the performance of the algorithm. And researchers found that the nearest neighbor classifier is easy to fall into the dimensionality disaster, especially when there are many irrelevant features. In addition, the performance of decision trees and neural networks is also affected by irrelevant or redundant features.

There are two types: Filter and Wrapper according to the combination with subsequent learning algorithms. The next chapter will introduce in detail. The main work of this paper is to introduce the feature selection process in the study of music classification, and to study the effect of feature selection on the music experimental data set. The feature selection process is shown in Figure 2.

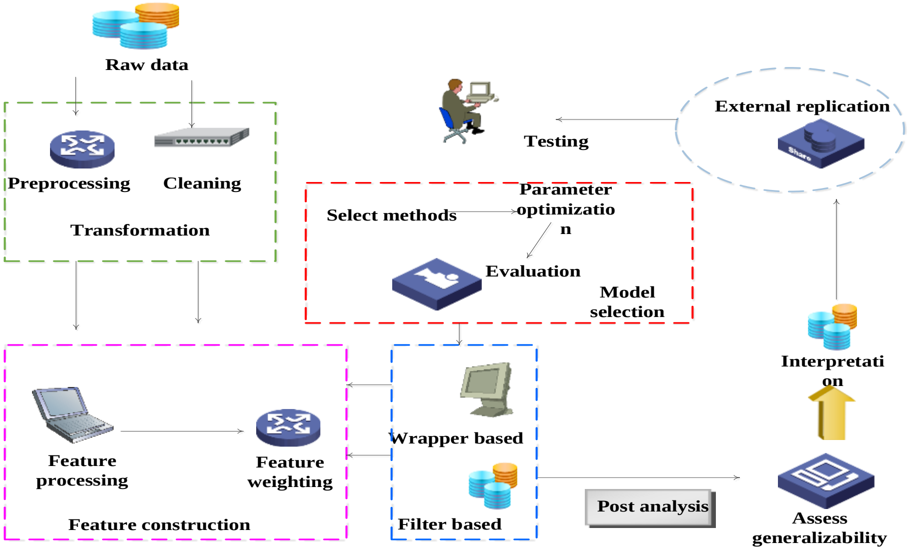

Figure 2: Feature selection process.

3.3 Extraction of Temporal Features

The most natural and direct method for signal analysis is to analyze time as the independent variable. The typical time domain characteristics of speech signals include short-term energy, short-term average zero-crossing rate, short-term autocorrelation function, and short-term average amplitude difference [10]. In this section, we mainly introduce the extraction methods of these temporal features.

Short-term energy is the energy of the speech signal changes significantly over time. Generally, the energy of the unvoiced part is much smaller than that of the voiced sound. The short-term energy of the signal can measure the amplitude changes. For signal $\{x_n\}$, short time average energy is defined as follows:

$$
\begin{aligned}
g^2(n-i) &= E/\sum_{i=1}^{n} x_i^2 \\
E &= \sum_{i=1}^{n} (x_i^2 f(n-i)) \\
&= x_n^2 f(n)
\end{aligned}
\tag{1}
$$

In the formula, $f(n)=g^2(n)$, E represents the short-term energy when the window function is started at the nth point of the signal, the unit impulse response of the linear filter is $f(n)$.

Short-term energy is mainly used for applications such as distinguishing unvoiced and voiced sounds, and distinguishing voiced and unvoiced segments. In the speech recognition system, it can also be used as a one-dimensional parameter in the feature to express the energy of the signal and the super-segment.

Short-term average zero-crossing rate (STAR) is the short-term average zero-crossing rate is the simplest feature in the time domain analysis of speech signals. For the discrete signals processed in the thesis, the STAR is essentially the number of times the sign of the signal sampling point changes. The STAR can reflect the spectral properties of the signal to a certain extent. A rough estimate of the spectral characteristics can be obtained through the STAR. The formula is:

$$\begin{aligned} g(n-i) &= 2Z / \sum_{i=1}^{n} | s(x_i) - s(x_{i-1}) | \\ Z \quad &= 1/2 \sum_{i=1}^{n} | s(x_i) - s(x_{i-1}) | \end{aligned} \tag{2}$$

The STAR can be used for speech signal analysis. Because the STAR can reflect the frequency to a certain extent, it generally has a lower zero-crossing rate in the voiced segment, and has a higher zero-crossing rate in the unvoiced segment. The STAR can be used to initially judge the unvoiced and voiced sounds. In addition, the STAR can be combined with the short-term energy to determine the location of the voice start and end points, that is, endpoint detection.

Short-term autocorrelation function is used to intercept a section of signal with a short-time window near the nth sample point of the signal, and the result of autocorrelation calculation is that

$$\begin{aligned} W(j) &= \sum_{i=1}^{n} | x_i g(x_{i+j}) | \\ &= \sum_{i=1}^{n} | x_i g(x_{n-i}) x_{i+j} g(x_{i-(i+j)}) | \end{aligned} \tag{3}$$

In the formula, *n* means that the window function is added from the nth point.

The short-term autocorrelation function has a very useful property: if the original signal $\{x_n\}$ is a periodic signal and the period is *T*, then its short-term autocorrelation function $\{W_j\}$ is also a periodic signal, and the period is also *T*. This property can be conveniently used to calculate the pitch period and pitch frequency parameters in the voiced signal.

4 ANALYSIS OF RESULTS

4.1 Design and Implementation of Computer-Aided Music Classification System

The purpose of the experimental system of this thesis is to complete the classification of music and provide the necessary preliminary classification for future music retrieval. According to the audio classification flowchart shown in Figure 3, the music classification system of this thesis is divided into three subsystems:

1. Music signal preprocessing and feature extraction subsystem;
2. Integrated learning and training subsystem;

3. Music classification subsystem.

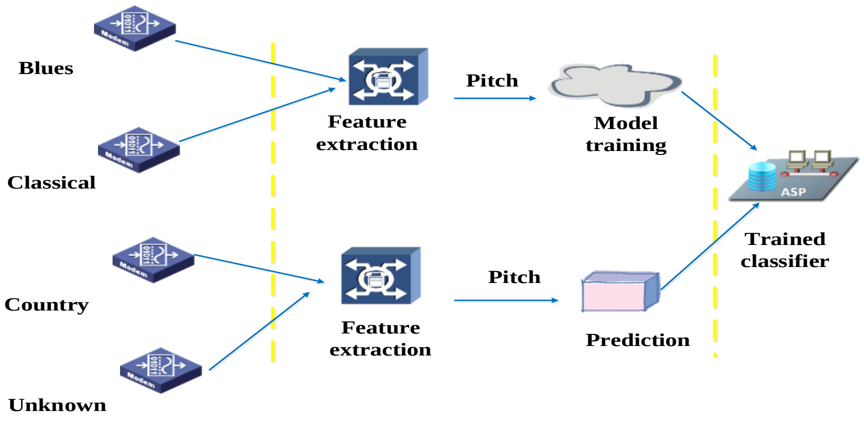

Figure 3: The implementation block diagram of the music classification system.

The music classification system is shown in Figure 3, which has two layers: training classifier and music classification: the upper layer first extracts sample music from the music database (used to train the music of the classifier's known category), and extracts the obtained classified music Music features based on content, and feature fusion. The music features as samples are used to train the classifier; the lower layer first numbers the original music to be classified (music of an unknown category used for classification) into the library, and performs feature extraction and fusion as well. , Input the obtained features into the trained classifier for classification, and finally classify the classified music according to the output result of the classifier and mark the class into the library. Below we specifically explain the three subsystems of this system.

First extract the music features of the training samples, enter the integrated learning training module, enter the training constraints according to the improved feature value-based classification integration algorithm, and then start the loop training to obtain a set of weighted weak classifiers, and the obtained weak classifiers are based on their weights Perform fusion to get the final classifier based on ensemble learning.

4.2 Comparison Between Classification Algorithms Based on Eigenvalues and Several Commonly Used Classification Algorithms

In the experiment of this paper, the music is first divided into non-overlapping audio segments with a time length of 1 second. It has been verified in the literature [11] that the accuracy of such a length is relatively high. The five categories of speech, pop music, folk songs, ancient instrumental music, and opera mentioned in the previous section are the classification categories of the audio segment in this one second. If two music categories appear on an audio segment at the same time, the thesis here specifies which audio takes a long percentage of time to consider which category the audio segment should be classified into, so that all audio segments have their own categories.

In the experimental design, this thesis first compares several commonly used classifier algorithms and algorithms based on eigenvalue classification to illustrate the effectiveness of eigenvalue classification algorithms; then the classification based on eigenvalue classification calculated with empirical parameters. The accuracy rate, the accuracy rate of the classification based on the feature value calculated by the optimal parameters calculated by the cross detection and the grid query is compared with the accuracy rate of the classification based on the ensemble learning algorithm based on the feature value classification, which shows the effectiveness of the improved algorithm, as shown in Figure 4.

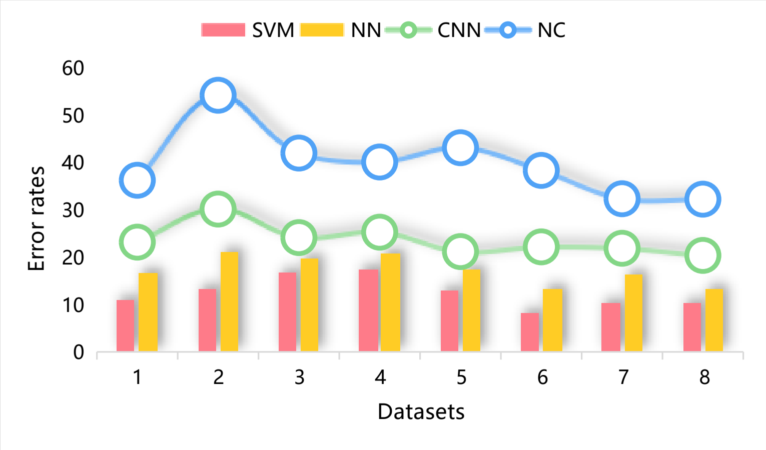

Figure 4: Comparison of error rates of four classifiers.

In this section, we will compare the algorithm based on feature value classification with other commonly used music classification algorithms. Literature has used sufficient experiments to compare SVM, NN (nearest neighbor), KNN (k-nearest neighbor), NC (nearest center), and the experimental results show the effectiveness of SVM.

The item with a check mark in the figure is the item with the lowest error rate in its class (the number in parentheses is the number of samples that are misclassified in the literature). It can be seen that the error rate of SVM is the lowest, which shows that the error rate is based on the effectiveness of feature value classification in music classification.

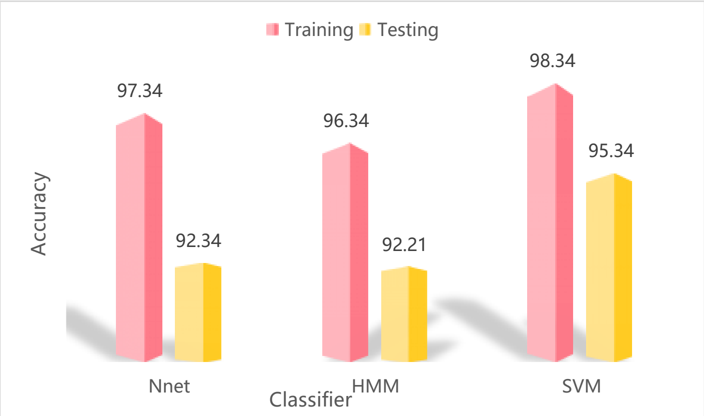

Figure 5: Comparison of three classification algorithms.

Figure 5 is a comparison of the three classification algorithms of Nnet, HMM and SVM, where Nnet stands for neural network classifier, HMM stands for hidden Markov classifier, and SVM stands for feature value classification classifier.

The training sample is the sample used to train the classifier, and the test sample is the sample used to detect the performance. For example, the SVM training sample means that there are how many samples for training the SVM classifier with a length of 1. The training accuracy rate is the percentage of correctly classified samples obtained by using the classifier obtained after training to distinguish the training samples. Similarly, the accuracy rate is the percentage of correctly classified samples obtained by using the trained classifier to distinguish the test samples. From the Figure 6, we can see the effectiveness of classification based on feature values in music classification.

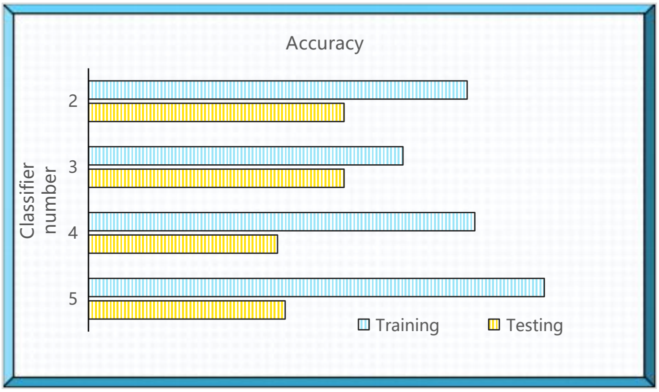

Figure 6: The classification results of the classifier based on the feature value classification under the best parameters.

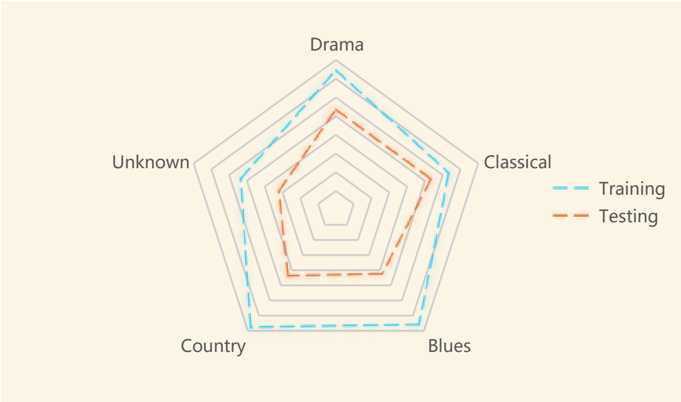

Figure 7: The ensemble learning classification results obtained by the computer-aided framework.

In summary, from Figure 7, it can be concluded that SVM and NN (nearest neighbor), KNN (k-nearest neighbor), NC (nearest center), neural network algorithm, hidden Markov algorithm, the classifier comparison of these five algorithms, based on feature values Classification is more effective in music classification algorithms.

5 CONCLUSION

The research of automatic music classification will become more and more important with the increasing number of music resources. Applying more effective classification methods and extracting more features that represent various characteristics of music will be the key to solving this problem. On the basis of reading a large number of documents, the author has done a detailed research on this issue. This article analyzes the audio short-term processing technology in depth, clarifies the audio hierarchical structure, and quotes the definition of audio structural units at different levels. The basic principles of audio classification are studied in detail. According to the audio hierarchical structure, the audio feature properties and extraction methods of audio frames and audio segments are discussed in detail. In-depth study of training algorithms based on feature analysis. The training algorithm based on feature analysis is the core problem to realize the feature-based classifier. The composition of the feature-based music classifier is discussed. The final experiment shows that the feature-based analysis method can be used as the final music classifier. This text has carried on the detailed introduction and design to each module in the system structure, at the same time puts forward some improvement schemes according to the concrete application, and realizes in the simulation experiment.

Minmin Ge, https://orcid.org/0000-0002-6756-9758
Yan Tian, https://orcid.org/0000-0003-1396-8216
Yifan Ge, https://orcid.org/0000-0001-5372-7922

REFERENCES

[1] Jiang H.: Computer-Assisted Interactive Platform Design for Online Music Teaching Based on Cloud Computing, International Journal of Engineering Intelligent Systems, 27(2), 2019. https://doi.org/10.1007/10.14733/cadaps.2021.S2.92-101
[2] Anders, F.; Hlawitschka, M.; Fuchs, M.: Automatic classification of infant vocalization sequences with convolutional neural networks, Speech Communication, 119, 2020, 36-45. https://doi.org/10.1016/j.specom.2020.03.003
[3] Jin, H.; Yang J.: Using Computer-Aided Design Software in Teaching Environmental Art Design, Computer-Aided Design and Applications, 19(S1):, 2021, 173-183. https://doi.org/10.14733/cadaps.2022.S1.173-183
[4] Rocha, B.-M.; Pessoa, D.; Marques, A.; Carvalho, P.; Paiva, R.-P.: Automatic Classification of Adventitious Respiratory Sounds: A (Un) Solved Problem? Sensors, 21(1), 2021, 57. https://doi.org/10.3390/s21010057
[5] Alam, F.; Danieli, M.; Riccardi, G.: Annotating and modeling empathy in spoken conversations, Computer Speech & Language, 5(1), 2018, 40-61. https://doi.org/10.1016/j.csl.2017.12.003
[6] Dardard, F.; Gnecco, G.; Glowinski, D.: Automatic classification of leading interactions in a string quartet, ACM Transactions on Interactive Intelligent Systems (TiiS), 6(1), 2016, 1-27. https://doi.org/10.1145/2818739
[7] Ntalampiras, S.; Ludovico, L.-A.; Presti, G.; Prato, P.-E.; Battini, M.; Cannas, S.; Palestrini, C.; Mattiello, S.: Automatic classification of cat vocalizations emitted in different contexts, Animals, 9(8), 2019, 543. https://doi.org/10.3390/ani9080543
[8] Albornoz, E.-M.; Vignolo, L.-D.; Sarquis, J.-A.; Leon, E.: Automatic classification of Furnariidae species from the Paranaense Littoral region using speech-related features and machine learning, Ecological Informatics, 3(8), 2017, 39-49. https://doi.org/10.1016/j.ecoinf.2017.01.004
[9] Azam, S.; Gavrilova, M.: Person's discriminating visual features for recognizing gender: LASSO regression model and feature analysis, International Journal of Biometrics, 9(4), 2017, 305-318. https://doi.org/10.1504/IJBM.2017.088248

[10] Zamora, W.; Calafate, C.-T.; Cano, J.-C.; Manzoni, P. Noise-sensing using smartphones: Determining the right time to sample. In Proceedings of the 15th International Conference on Advances in Mobile Computing & Multimedia, 12(1), 2017, 196-200. https://doi.org/10.1145/3151848.3151868

[11] Corrêa, D.-C.; Rodrigues, F.-A.: A survey on symbolic data-based music genre classification. Expert Systems with Applications, 60, 2016, 190-210. https://doi.org/10.1016/j.eswa.2016.04.008