www.cadanda.com
Computer-Aided Design AND APPLICATIONS

Music Style Recognition Method Based on Computer-Aided Technology for Internet of Things

Jun Zhao[1], Marianne Zhao[2], Xueping Yang[3], Xiaojie Li[4] and Zixi Chen[5,*]

[1]Department of Music, Zhengzhou Preschool Education College, Zhengzhou 450099, China, drzhaojun@hotmail.com
[2]Department of Music, Zhengzhou Preschool Education College, Zhengzhou 450099, China, mlaumets@gmail.com
[3]Department of Music, Zhengzhou Preschool Education College, Zhengzhou 450099, China, yxpdy2010@126.com
[4]Department of Music, Zhengzhou Preschool Education College, Zhengzhou 450099, China, lixiaojiemn@126.com
[5]Department of Music, Chongqing University of Art and Science, Chongqing 402160, China, 20040081@cqwu.edu.cn

Corresponding author: Zixi Chen, 20040081@cqwu.edu.cn

Abstract. The computer automatic genre analysis of music has great theoretical significance and practical value for the retrieval, classification, distribution and research of music works. At present, the research results of using computer to identify music style are relatively few. The common scheme is to segment the music, judge the genre characteristics of each segment respectively, and judge the genre of the whole music by synthesizing the results of each segment. In view of the problems that the relationship between music styles is unclear and confusing due to the manual differentiation of music styles, and some songs are difficult to divide their styles manually, the use of music melody features can help computers identify music styles, and analyze the emotions expressed by songs based on the song lyrics data, so as to divide their style attribution. A music style prediction model is established by using the computer-aided technology for the Internet of things. The model is used to mine the frequent patterns of the melodies of different styles of music and test the style recognition of music. By processing the data of music signals, different music is classified into a certain style, which can help users quickly find the music of related styles and realize more effective management of music database.

Keywords: Internet of things; computer aided design; music style; intelligent identification; database.
DOI: https://doi.org/10.14733/cadaps.2023.99-109

1 INTRODUCTION

Nowadays, the rapid development of Internet technology and digital multimedia technology provide a very important way of transmission and good storage medium for music, and promote the communication and development in the field of music. With the development of the Internet, the main media of pop music has gradually changed from traditional radio and record to network download and network radio. Network radio stations need to automatically recommend and play music according to the known preferences of listeners. Chen and Li [1] mentioned that in recent years, various music styles have emerged in endlessly, which fully demonstrates the personality of music creators. As everyone's favorite music may span several styles, which are very different and need to be treated separately, it is necessary to explore how to distinguish music styles. In pop music, traditional style concepts include pop, country, jazz, rock, rhythm and blues, new century and other categories. They can be divided into many sub categories, and some sub categories can even be further subdivided. Each song can only be manually labeled with style. This approach has many shortcomings: the relationship between some categories is not clear, some categories are too rough or fine, some category labels are not recognized, and some music attribution is controversial or difficult to plan. Therefore, with the increasing number of music, it is very important to classify and manage the generated music. At present, the commonly used music retrieval methods are content-based retrieval methods, such as classification based on music name, designated composer, music age, etc. As the carrier of emotion, music retrieval research based on music emotion classification has very important objective significance. In a certain scene, matching music with emotion can better set off the atmosphere and enhance the appeal. For example, before going to bed, play a piece of soothing and tender music; When running, play a passionate and rhythmic music; When having a picnic outdoors, play a light and quiet music; When couples propose, play a sweet and beautiful music, and so on. The emotion-based retrieval method can just make up for the needs of users for music emotion. Chen et al. [2] mentioned that multimedia video also needs to make real-time music according to the emotion transmitted by the video content. Many factors lead to the difficulty in extracting music signal features, which leads to the difficulty in improving the recognition accuracy of music style. Therefore, recently, the recognition of different music styles has received extensive attention and has been developed rapidly. Different performers will also make full use of different skills to form different musical styles. Many studies have analyzed the relationship between performance skills and music style from the perspective of music, but few from the perspective of signal processing and computer recognition. If the computer can recognize different playing skills, it will be very convenient to analyze the music style from the perspective of playing skills. Therefore, how to use computer to recognize these features and to analyze and recognize music signals is a very meaningful research direction. This not only helps people understand the essence of music more deeply, but also can be applied in music teaching and searching. Based on this, people can classify and search the online music resources to obtain the required resources, which greatly facilitates the dissemination and exchange of music. For music lovers who learn to play musical instruments, this classification method will be very convenient to learn. It can be seen that with the popularization of the Internet and the continuous development of network applications, the importance of digital music retrieval and recommendation is self-evident in the face of huge user groups and massive data. Music classification is an important field of music retrieval.

2 PRELEVANT THEORIES AND TECHNOLOGIE

2.1 Music Style Classification

In this era of digital music, music is divided into many genres, each of which has its own representative works with distinctive characteristics. Therefore, it is not easy to predict the style of a given song according to its lyrics or music. With the development of social life, culture and art are developing regularly. Especially in the continuous enrichment and organization of music melody,

rhythm, harmony, polyphony, structure, musical form, orchestration, etc. Music styles often change with the changes of the times. This change is continuous. Therefore, the music styles of one period and another period will overlap. Sometimes there is no strict style division. However, from a familiar point of view, music mainly includes classical music and pop music. Classical music has influenced generations with its profound historical connotation. It has stable structure, diverse techniques, perfect harmony, and expresses people's emotions in different historical periods. It is precisely because of the emotions and humanistic feelings contained in classical music that it can still have a foothold in today's fast food entertainment culture. Hajj et al. [3] mentioned that there are several important periods in the development of classical music: Renaissance, Baroque, Vienna classical music, Romanticism, national music, Impressionism and electronic music. There are many kinds and styles of pop music. Before 1980, the main types of pop music were blues, jazz, country music and rock music. After 1980, with the development of multimedia, music styles continued to evolve in depth, mainly including rap, hip hop, R&B and new age. Music style is a concept with rich connotation. Different kinds of music styles are ever-changing, and the research angles are also diverse. Music style classification can be said to be a branch of music information retrieval. Before music signal retrieval, the song must be characterized by certain symbolic features. The same is true of the style classification of music. Music signal consists of timbre, pitch, rhythm and other factors. Before the music signal is classified, it is necessary to select the audio features that conform to the music characteristics, so that the extracted feature parameters can accurately express the sound quality, melody, rhythm and other characteristics of music. The music quality characteristics analysis includes the time domain and frequency domain analysis of the audio signal texture, and the computer intelligent understanding of the music texture characteristics. Gardini et al. [4] mentioned that the rhythm characteristic of music is an important time sequence characteristic of music, which reflects the dynamic rhythmic information in music. It is a kind of characteristic with high discrimination among music species. Music rhythm analysis includes accurate positioning of each beat in a song, prediction and correction of subsequent beats. The melody features of music show the outline of a musical tune, and reflect the basic situation of the style and emotion of a song. Figure 1 shows the music computing research tasks and their interrelationships.

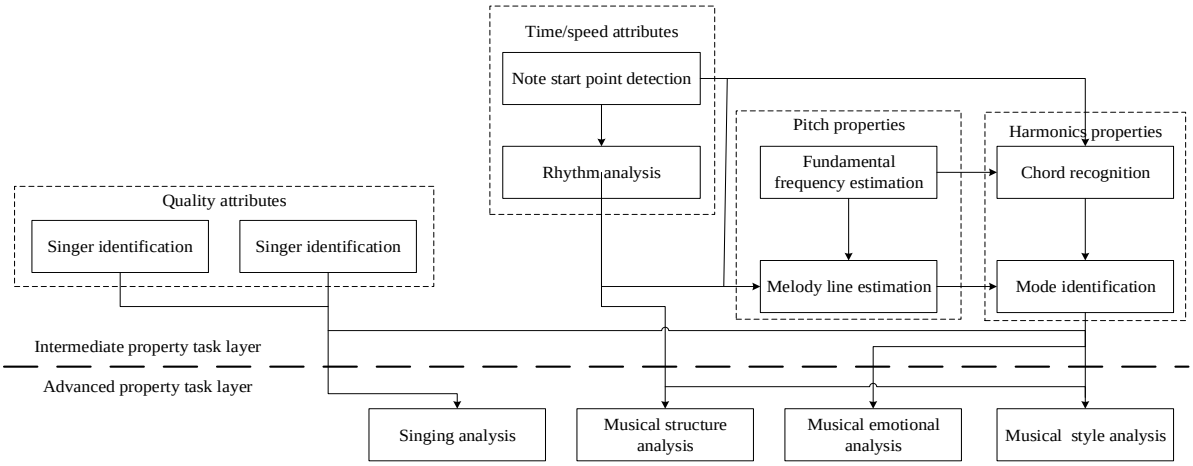

Figure 1: Music computing research tasks and their interrelationships.

It can be seen from the figure that the typical attribute tasks of music computing generally include: Note starting point detection, rhythm analysis, fundamental frequency estimation, melody line estimation, chord recognition, mode recognition, singer recognition, and instrument recognition. The above subtasks belong to the time / speed attribute, pitch attribute, harmony

attribute and sound quality attribute of music. It can be seen that the research tasks with the above characteristics can cover the analysis methods of all the form elements of music audio. Hakanpaa et al. [5] mentioned that through the research on these attribute tasks, we can establish intelligent analysis methods for the form elements of music from various angles. In addition, the typical high-level tasks of music computing include singing analysis (performer identification, lyrics identification, voice part identification, etc.), music structure analysis, emotion analysis, style analysis, etc.

2.2 Music Theory Knowledge

Music is a kind of sound art, which is produced by the regular vibration of the vocal body. The various elements of music language cooperate with each other and have the ever-changing expressive force. When each element changes, the music image will change to varying degrees. It is difficult to express a song accurately with one or several symbols or eigenvectors. Music is more abstract to concrete people and things as well as vivid life reality. It does not have a specific reflection on life like TV dramas, art or literature. The carrier of music is music, while harmony, mode, rhythm, melody, musical form and other elements constitute music. Kroher et al. [6] mentioned that the premise and foundation of music signal processing is the full analysis of music signal. Music signal has many characteristics. In addition to text information such as singers, there are also characteristics based on music content signal itself, such as sound quality, pitch, rhythm and so on. At present, music signal feature extraction includes sound quality feature, melody feature and rhythm feature.

a. Melody. Melody is the soul of music, the foundation of music formation, and it expresses the image and thought of music. Melody is a complex of many musical elements, including pitch and rhythm. Pitch line is also called melody line, which indicates the movement direction of melody. If the melody goes up, it goes up; to the low point, it is downward; when one or several tones in the melody repeat continuously, it is homophonic repetition; when the melody is repeated up and down with a certain sound as the axis in a small interval, it is carried out in a circular manner.

b. Interval. Interval can represent the distance relationship between the pitches of two notes. The low and middle note of two notes is the root note, and the other is the crown note. The interval number of lines between two tones in an interval is the degree of the interval. On the staff, each interval or line is a degree, and the adjacent lines or intervals represent a second interval, that is, if there are several intervals or lines, they are called several degrees. The number of intervals refers to the number of half tones and whole tones in an interval. 1/2 represents half tones and 1 represents whole tones. The degree and number of intervals determine the nature of an interval.

c. Harmony. Harmony and melody are similar to the front and back of coins. The chord of harmony is expressed as vertical lines, while the melody is expressed as horizontal lines. The functional system of harmony includes the strength, stability and harmony of the chord, which will affect the tightness and strength of the music rhythm. In addition, harmony also includes four color functions: thick, light, thick and thin, which makes it play an important role in tonal music or non-tonal music.

d. Chords. Chord refers to the vertical combination of three or more pitches. When superimposed, it can be carried out according to the relationship of three degrees or non-three degrees. With the development of harmony, many musical chords use the non-three-dimensional superposition relationship. Compared with the three-dimensional superposition relationship, the non-three-dimensional superposition relationship has more colorful chords. All the sounds of a chord have names. In the staff, from bottom to top, the first sound is called 'root sound', the second sound is called three, the third is called five, and the fourth is called seven, and so on.

e. Mode. The height of the tonic is the tone of the scale. The height of the tonic is represented by a key sign. The key sign is usually in the form of a rising sign and a falling sign, which is marked after the clef of the staff. The sequence of tones orderly arranged according to the specific interval relationship is called the mode; the arrangement of tones from high to low in a mode is

called a mode scale. The organic integration of mode and tonic pitch forms tonality, that is, the synthesis of harmonic modes. In the music, the continuous change and contrast of the key, mode and tonality set off the different emotions and colors of the music.

f. Musical form. In music creation, music form and its content complement each other. Music content determines music form, and music form makes music content more prominent. The creators will be restricted by the existing musical forms, but they will also have bold innovations. The musical form can be divided into spatial structure and temporal structure. The spatial structure is called 'texture', and the temporal structure is called 'musical form'.

The current music style system can only divide the general categories of music, but cannot divide the subdivided music genres. It is undeniable that the appreciation of music is the subjective feeling of human beings [7]. In some cases, it is difficult for human beings to distinguish some specific music genres, and there will be overlap between individual music genres. These factors determine the difficulty of music genre learning and automatic division. The types of music are complex. Different people have different emotional preferences for music because of their personality, environment and current mood. Sometimes, because of complex environmental factors, people prefer to listen to music with negative emotional characteristics when they are depressed, resulting in more depression. Different people have different preferences for the style of music. Some people like rock and some people like folk songs.

3 MUSIC FEATURE EXTRACTION ALGORITHM

3.1 Signal Preprocessing and Feature Extraction

The implementation of music style classification system involves relevant machine learning theories. Machine learning is to try to make the machine imitate the human brain, find the internal relationship and hidden rules, and use the learned knowledge to make inference and prediction on the future data. Common music classification systems are designed according to different genres. For example, music can be divided into dance music, jazz, lyric, Chinese folk music, rock and other styles. No matter how to classify music, the music classification system needs to include two modules: signal feature extraction and learning classification. Figure 2 is the overall framework of the vocal analysis method. It can be seen from the figure that, firstly, the music samples need to be preprocessed, and the useful information in the music signal is filtered out through signal processing technologies such as framing and windowing, so as to remove the noise and other useless information; Secondly, the feature vectors which can represent the characteristics of music style are extracted as the input parameters of the next classification model; Finally, according to the relevant knowledge of machine learning, the music classification model is trained and established, which can be used to classify the unknown music data [8].

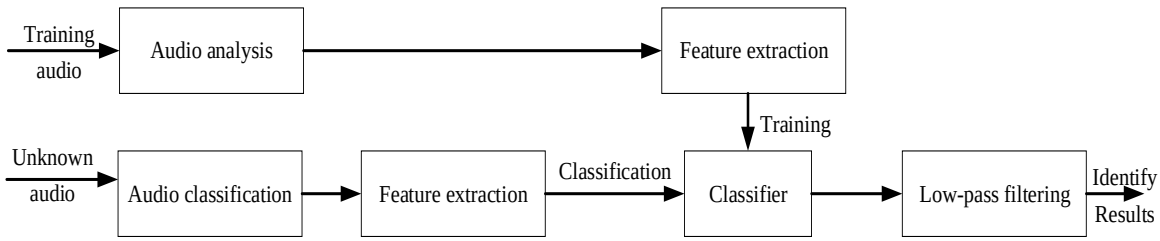

Figure 2: Overall framework of the vocal analysis method.

In music signal feature extraction. First, we need to use related technologies to preprocess the music signal to form a unified format. General preprocessing includes anti-aliasing filtering, pre-emphasis, digitization, windowing and framing. Because the music and songs stored on the Internet have usually been digitized, they can only be processed by pre-emphasis and windowing.

Secondly, the music signal is mixed with people's singing sound and musical instrument playing sound, resulting in the music signal spectrum is different from the ordinary voice signal. Therefore, by observing the frequency spectrum of the music signal, we can see that the waveform is not smooth, and the high-frequency part accounts for a very heavy proportion. Therefore, before analyzing the music signal, the pre-emphasis technique can be used to remove the noise, flatten the frequency waveform and strengthen the high-frequency part. Thirdly, it is found that the music signal will change with the passage of time, that is, it sometimes becomes unstable. However, for the part between 10ms and 30ms, the spectrum waveform is basically stable, which can be regarded as a short-term and stable process. Therefore, the characteristics of music can be found by cutting this part of the signal. Usually, this part of music signal is cut into several segments called frames, and then it is easy to analyze with correlation signal processing technology. Generally, the framing process is achieved by adding windows to a signal window of limited length. At the same time, adding windows can prevent large peak values in the frequency domain. They are the necessary information of speech and an important feature of music recognition [9].

3.2 Incremental Music Style Classifier based on Support Vector Machine

The current computer analysis technology of music genre has made great progress, but it also has some defects that need to be improved, including: (1) music segment segmentation affects the genre classification of music, and the current music segment segmentation technology itself is a difficult problem; (2) The classification accuracy is not high; (3) The classification operation time is long. Therefore, it is necessary to make further research and improvement on the computer analysis technology of music style and genre to improve its practicality. At present, for the research field of content-based music information retrieval technology, which is a practical technology driven by goals, due to its newness and complexity, its main research focuses on the analysis and recognition technology of various music audio features, attributes and elements; For the back-end music retrieval index establishment and query methods, due to the common retrieval technology has developed a very mature and effective method, so little attention has been paid. At the same time, it also reflects the key of content-based music information retrieval technology. As a common pattern recognition tool, the core idea of support vector machine is to reconstruct the features to be classified in a higher dimensional feature space, so as to obtain a classification representation more suitable for the input features in a space different from the original classification space. In this paper, the theory of active learning is applied to the realization of incremental music style classification system, and a music style recognition method based on Gaussian mixture model (GMM) is proposed, which has two modules: training and recognition. In the training part, firstly, preprocessing and feature extraction should be carried out for the training segments to obtain the frame-based classification feature vector [10]. The initial values of the style model parameters based on GMM are determined by clustering analysis of the characteristic frames of the training data, and then the expectation maximization (EM) algorithm is used to train the style model parameters, and then the model parameters of each style music are obtained. In the recognition part, the difference between the feature vector of the segment to be classified based on the frame and the style model is calculated, the likelihood ratio between the feature frame and each type of preset style model is calculated, the segment to be classified is classified into the style corresponding to the model with the greatest similarity, and the likelihood ratio is expressed in the radar chart using the song style vector, and the style discrimination results and visual analysis are given based on the style vector. Figure 3 is the overall framework of the music style analysis method.

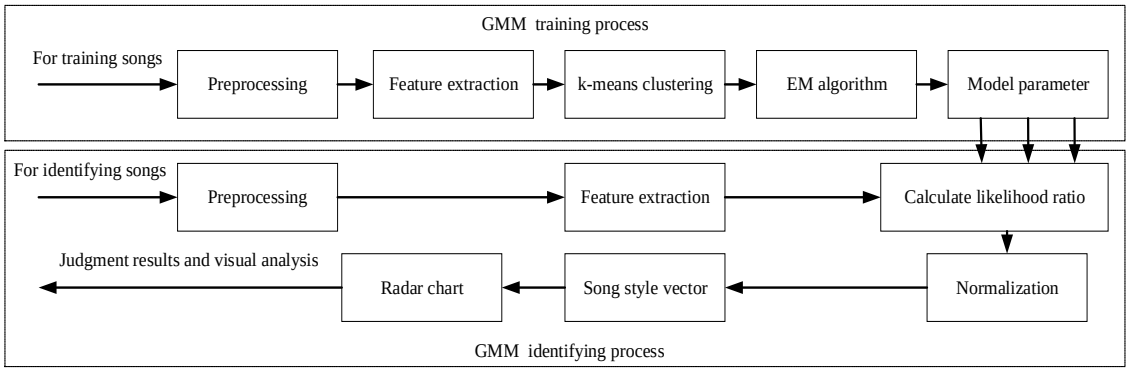

Figure 3: Overall framework of the music style analysis method.

4 DESIGN AND IMPLEMENTATION OF MUSIC STYLE RECOGNITION SYSTEM

The experimental data of this paper consider the five categories of commonly accepted data about the perception of music style, including lyric, jazz, dance music, rock and folk music. Figure 4 is the signal characteristic diagram of different music. It can be seen from the figure that the average deviation and standard deviation of spectrum contrast constitute a 24-dimensional feature vector of music clips. This method can catch the melody more accurately when dealing with classical and light music, which have clear melody levels and seldom use percussion instruments; But in the music of heavy metal style such as rock, the performance is average; The result of jazz music is lower than expected, mainly because the live version of the singing often improvisational tone shift and noisy background sound, which has a great interference to the recognition work.

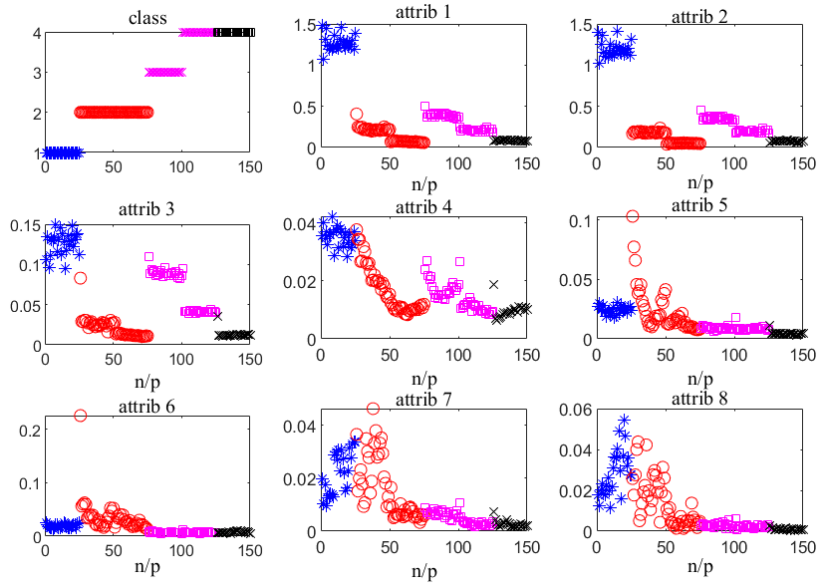

Figure 4: Signal characteristics of different music.

This analysis method can find the performance method used in the music, and also has certain recognition ability for the same bow method for a long time, which basically meets the needs. However, the problem lies in the poor ability to identify bows that occur in a short time, and because of the existence of the wrong judgment rate, it is difficult to distinguish whether certain bows are really used or whether they are misjudged because of segmentation or prediction. This problem can be remedied by more data training and more intelligent segmentation. Figure 5 shows the detailed classification results of 10 seconds music clips. It can be seen from the figure that adding more training data during training and further optimizing the algorithm and features will help to improve the accuracy of prediction. In the segmentation, we can consider using the change characteristics of time domain and frequency domain to segment, so it is easier to segment the standard bow segments, so as to improve the accuracy. Obviously, it is unscientific to evaluate the pros and cons of the melody discovery technology proposed in this paper only from the perspective of objective experimental statistical results, because music, as an art form, is divorced from people's subjective perception, that is, it has no meaning of existence. However, the method proposed in this paper is based on the subjective cognition of simulators to music, so it is more necessary to evaluate the experimental results from the subjective sense of listening.

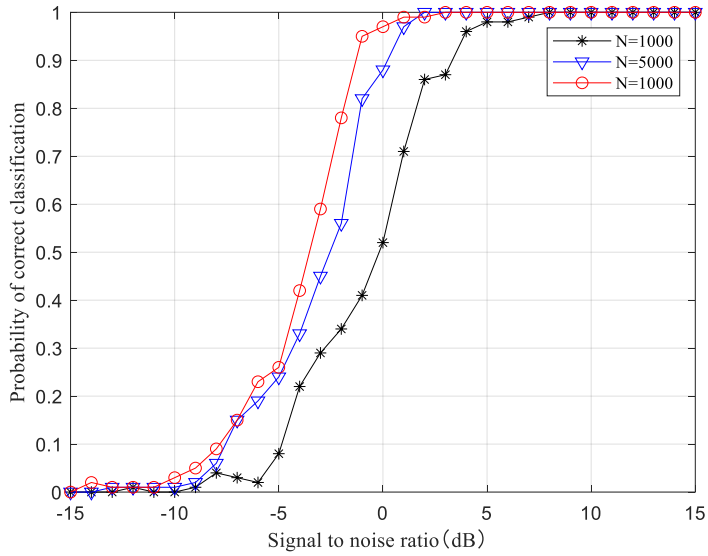

Figure 5: Detailed classification results of 10 seconds music clips.

The forward propagation process of music style recognition and generation algorithm is to input the processed multi-dimensional vector to the input layer, then through the interpretation layer, the genre analysis layer, and finally through the linear layer. In the process of calculation, the input vector is transferred step by step through the calculation flow. When calculating the output of each neuron, the matrix containing the strength of music features is finally obtained, and then the matrix generated by the training and the original matrix are used for operation. In the process of forward propagation, when comparing the difference of matrix values, a threshold is set before this. If the difference is less than the threshold, it means that the propagation is effective. If the difference is greater than the threshold, it means that the parameters in the model are still not optimal, and back propagation is still needed to update and adjust the parameters. Figure 6 shows the detailed classification results of the whole music. It can be seen from the figure that the average classification accuracy rate of the whole music is 90.8%, 82.3% after 10 seconds of editing is higher, and the classification error rate of each kind of music is reduced, which further shows that the spectral contrast feature can effectively improve the effect of music classification.

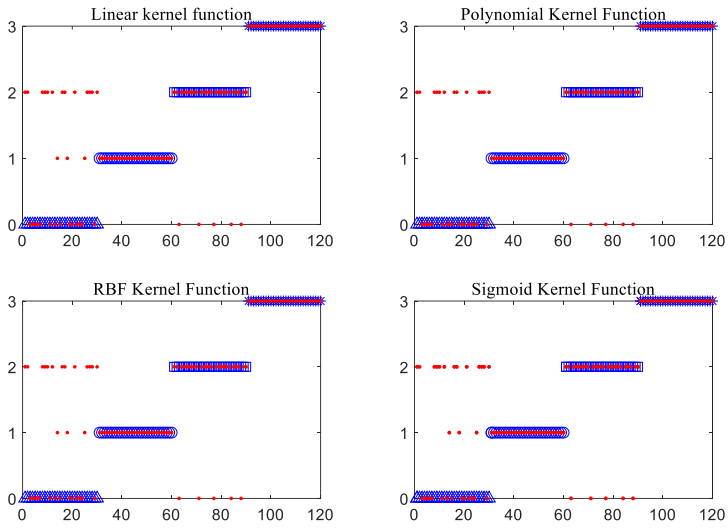

Figure 6: Detailed classification results of the whole song.

Figure 7 is a statistical chart of the accuracy of classification methods. It can be seen from the figure that the accuracy of music style classification has been relatively improved after the introduction of fuzzy membership function, that is, the support vector machine has enhanced the generalization ability through fuzzy classification.

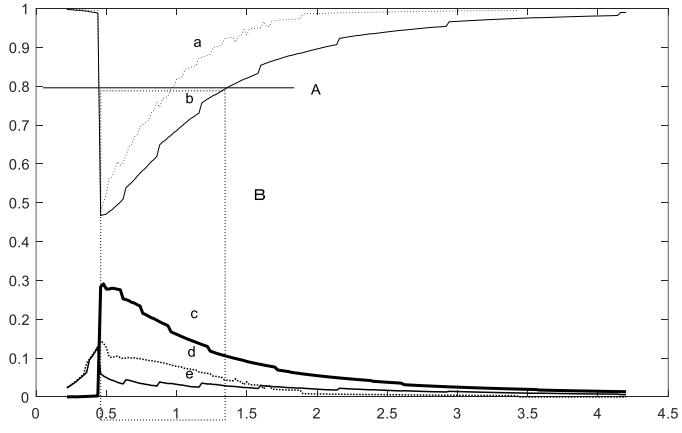

Figure 7: Statistical chart of classification method accuracy.

At the same time, it can be seen from the table that some music classification accuracy is relatively high and some are relatively low. Among them, jazz music is relatively difficult to distinguish, and its main erroneous classification is lyric and Chinese folk music style. The reason may be that jazz music has complex and changeable rhythms and rhythms. It is usually the improvisation of musicians, and there are great changes in various parts of the music. Rock music is relatively easy to distinguish, which shows that rock music is very different from other music in characteristics. Dragged down by rock and jazz music, the overall accuracy of halftone recognition is 61%, and the recall rate is only 67%. There are many misunderstandings between the two kinds of music, and the most obvious problem is the rock music with complex orchestrations and intense rhythms; Jazz is greatly influenced by the looseness of the singer's interpretation. The recognition

result of light music is good, which is directly related to the stable orchestration, slow rhythm and prominent melody.

5 CONCLUSION

With the advent of the Internet era, multimedia data has shown a blowout growth. In the music field closely related to people's life and entertainment, we urgently need a tool to help people quickly and accurately complete the retrieval of music tracks and improve the user experience. The emergence of music style classification technology can meet people's basic needs and has very important practical significance. However, in the traditional music classification system, on the one hand, it is difficult for the model to collect a complete music training set that can reflect the distribution of all samples, and the generalization ability and stability of the model will decrease with the passage of time; On the other hand, directly combining the huge incremental data with the original sample training set will lead to the defects of long training time and large storage space consumption. Music style reflects the general basic characteristics of music works and is the basis of music appreciation, analysis and research.

With its urgent application needs and broad market prospects, music information retrieval and intelligent music teaching and creation have become two research hotspots in the field of music information processing, and also an effective combination of audio signal processing technology and the oldest culture and art communication carrier in theory and practice. It can be predicted that the computer-aided technology based on the Internet of things is applicable to many fields, such as music data management, analysis, music service provision, etc., which can better meet the needs of computer analysis technology for music style and genre, and provide a technical foundation for the development of music style and genre analysis system in the future.

With the development of digital music analysis technology, more and more new music classification technologies will appear, and the original classification methods will be constantly adjusted. The development trend of music style analysis technology is predicted as follows:

a. New category detection. New category detection means that the classifier can recognize a new music style, that is, when the classifier detects a new music style, it can automatically mark the music as a new music style.

b. Multi class tags. Multi category tagging refers to marking a piece of music into multiple styles. With the development and integration of music. A piece of music may integrate multiple styles and classify music with multiple labels. It can effectively solve the problem of vague and ambiguous music style, and it is more in line with the actual situation.

6 ACKNOWLEDGEMENT

This work was supported by 2022 Henan Province Teacher Education Curriculum Reform Key Research Project: Innovation and Practice Research on the Localisation Path of Finnish Higher Normal Music Education Curriculum in Chinese Universities No.2022-JSJYZD-046.

Jun Zhao, https://orcid.org/0000-0002-0475-8383
Marianne Zhao, https://orcid.org/0000-0001-8809-6841
Xueping Yang, https://orcid.org/0000-0002-4659-8509
Xiaojie Li, https://orcid.org/0000-0001-9866-518X
Zixi Chen, https://orcid.org/0000-0003-0172-0598

REFERENCES

[1] Chen, G.-B.; Li, S.-J.: Network on chip for enterprise information management and integration in intelligent physical systems, Enterprise Information Systems, 15(7), 2020, 935-950. https://sci-hub.et-fine.com/10.1080/17517575.2020.1720828

[2] Chen, C.; Ma, J.-X.; Sun, P.: Firmware code instrumentation technology for internet of things-based services, World Wide Web-Internet and Web Information Systems, 24(3), 2020, 941-954. https://sci-hub.et-fine.com/10.1007/s11280-020-00838-3
[3] Hajj, N.; Filo, M.; Awad, M.: Automated composer recognition for multi-voice piano compositions using rhythmic features, n-grams and modified cortical algorithms, Complex & Intelligent Systems, 4(1), 2018, 55-65. https://sci-hub.et-fine.com/10.1007/s40747-017-0052-x
[4] Gardini, E.; Ferrarotti, M.-J.; Decherchi, S.: Using principal paths to walk through music and visual art style spaces induced by convolutional neural networks, Cognitive Computation, 13(2), 2021, 570-582. https://sci-hub.et-fine.com/10.1007/s12559-021-09823-y
[5] Hakanpaa, T.; Waaramaa, T.; Laukkanen, A.-M.: Emotion recognition from singing voices using contemporary commercial music and classical styles, Journal of Voice, 33(4), 2019, 501-509. https://sci-hub.et-fine.com/10.1016/j.jvoice.2018.01.012
[6] Kroher, N.; Diaz-Banez, J.-M.: Modelling melodic variation and extracting melodic templates from flamenco singing performances, Journal of Mathematics and Music, 13(2), 2019, 150-170. https://sci-hub.et-fine.com/10.1080/17459737.2019.1610194
[7] Lou, H.-L.: Design of college English process evaluation system based on data mining technology and Internet of things, International Journal of Data Warehousing and Mining, 16(2), 2020, 18-33. https://sci-hub.et-fine.com/10.4018/ijdwm.2020040102
[8] Mertala, P.: Young children's perceptions of ubiquitous computing and the Internet of Things, British Journal of Educational Technology, 51(1), 2020, 84-102. https://sci-hub.et-fine.com/10.1111/bjet.12821
[9] Xie, J.; Yang, Y.: IoT-based model for intelligent innovation practice system in higher education institutions, Journal of Intelligent & Fuzzy Systems, 40(2), 2021, 2861-2870. https://sci-hub.et-fine.com/10.3233/jifs-189326
[10] Zhang, J.-Y.: Reform and innovation of artificial intelligence technology for information service in university physical education, Journal of Intelligent & Fuzzy Systems, 40(2), 2021, 3325-3335. https://sci-hub.et-fine.com/10.3233/jifs-189372