www.cad-journal.net
Computer-Aided Design and Applications

Optimization Algorithm in Computer-Aided Multi-Voice Music Arrangement and Collaborative Design

Bing Tian[1] and Jie Tian[2]

[1]School of Music and Dance, Zhengzhou University of Science and Technology, Zhengzhou 450000, China, music.tian_bj@zit.edu.cn
[2]School of Education and Music, Hainan Vocational University of Science and Technology, Hainan, 571126, China, musictjie@163.com

Corresponding author: Jie Tian, musictjie@163.com

Abstract. The objective of this research is to devise an optimization algorithm grounded in CAD (Computer-Aided Design) that can automatically refine diverse components in multi-voice music. This aims to attain an integrated and harmonious music structure. Additionally, this study aims to investigate the potential of this optimization algorithm in enhancing the collaborative process and the outcome of multi-voice music creation. Experimental findings reveal that, in contrast to alternative arrangement methods, this optimization algorithm swiftly identifies the globally optimal solution. As a result, multi-voice musical compositions exhibit a notably elevated level of originality and intricacy while preserving a coherent style. Moreover, this algorithm demonstrates commendable adaptability and reliability, rendering it suitable for multi-voice music arrangement endeavours across varying genres and complexities. This exploration not only introduces fresh perspectives and techniques to the realm of multi-voice music composition but also contributes positively to the advancement and evolution of musical technology.

Keywords: Computer-Aided Design; Genetic Algorithm; Multi-Voice Music Arrangement; Collaborative Design
DOI: https://doi.org/10.14733/cadaps.2024.S26.172-186

1 INTRODUCTION

With the continuous development of the times and the advancement of technology, people's pursuit of musical life is also constantly improving. Computer music has also developed tirelessly with the continuous growth of the music industry. Computer music is a type of digital music that combines computer application technology with music technology. It is a technology that includes various aspects such as arrangement, recording, mixing, and mastering. Computer music gradually entered China during the reform and opening up period and has been continuously applied to various fields of China's music industry in the past thirty years of development. In the digital age, music recognition technology has developed rapidly, but traditional music recognition methods often rely on known audio signals and score information. However, for music works that use unknown musical symbols as carriers, such as ancient scores or unconventional notation methods, traditional recognition methods are clearly ineffective. Therefore, exploring the use of the two-dimensional properties of unknown music symbols for neurological music recognition has become a research hotspot in the current field of music technology [1]. Computer music production technology almost encompasses all categories of music, from computer music analysis, computer music software, and computer music hardware systems to various technological fields such as computer music production technology under artificial intelligence today. The development of computer music production technology in China has also gone from almost blank since the reform and opening up to a hundred flowers blooming today. Whether it is outstanding performance in the application field or continuous innovation in software and hardware research and development, the development of music production technology in China is now in a world-class position. Andrea and Zahra [2] have studied computer music production technology for many years and witnessed the development of computer music in China over the past decade. By studying the iterative updates of software and hardware, the characteristics of music production technology in different periods in China, and the different characteristics of recording and mixing in audio production in different periods, we can streamline the development of computer music production technology in China. It summarizes its years of learning and provides a reference and outlook for future development. It is particularly important to explore how China can quickly transition from an almost blank industry to a world-class position. To explore new paths for future learning and work, better develop China's computer music production technology, and provide better assistance for learners. So it is important to summarize past experiences and clarify the development process of computer music production technology in China.

With the rapid development of digital media, audio technology, and artificial intelligence, the amount of music data continues to increase, and correspondingly, research on it is also increasing. The essence of music is the carrier of human emotional expression, and how to achieve more accurate recognition of music emotions has become our focus. Baro et al. [3] fused multimodal music features based on deep learning, which combines different types of information in music information. It combines the continuous and discrete emotional features of music to optimize the emotional recognition performance of the model. Calilhanna et al. [4] constructed the WLDNN_SAGAN model based on previous research. Optimize the GAN module of WLDNN_GAN, add a self-attention module, optimize the weight size of music signal input, and achieve more efficient and accurate music emotion recognition. It adopts the Schenkel analysis method to extract the most representative emotional segments in the movement and input them as the main melody vector. Based on the foundation of Chapter 3, optimize the MFCC features and combine them with RP by weighting to represent more complete and comprehensive music-emotional features. Input the fused music information into a digital network and conduct horizontal comparative experiments on models such as WGAN (Wasserstein GAN), MCCLSTM, and MCCBL. It is found that the model has the highest accuracy when the number of MFCC and PLP features extracted from the continuous emotion space of the input features is 1:1, and the number of main melody features extracted from the discrete emotion space is 1:1. This can prove that multimodal feature input has a positive effect on music emotion recognition.

The emergence of CAD technology has opened up new avenues for the arrangement and collaborative design of multi-voice music. In the field of music education, the introduction of CAD information technology is bringing revolutionary change, which has greatly promoted the modernization and efficiency of music education. Fu et al. [5] discussed the revolution of music education based on CAD information technology and focused on the application of analytic hierarchy process (AHP) and TOPSIS in teaching optimization. CAD information technology plays an important role in music education. It can not only assist teachers in music creation and composition but also help students better understand music theory and skills. Through CAD software, teachers can easily draw music scores and analyze music structures, while students can learn music knowledge and practice performance skills through an interactive interface. In addition, CAD information technology can also realize distance teaching and online learning, break the time and space constraints of traditional music education, and let more people enjoy high-quality music education resources.

By incorporating cutting-edge algorithms and models, CAD technology can precisely control and refine musical elements, ushering in automation and intelligence in music creation. The automatic music classification system realizes the automatic classification and arrangement of music works by extracting and analyzing the characteristic information in music. However, in practical applications, such systems still face many challenges, such as the accuracy of feature extraction, the efficiency of classification algorithms and so on. Therefore, it is of great practical significance to optimize the computer-aided design system of music automatic classification based on feature analysis. Ge et al. [6] used advanced signal processing technology and a machine learning algorithm to reduce dimension and denoise the extracted features. This can not only reduce the amount of calculation, and improve the real-time performance of the system, but also reduce the impact of noise on the classification results. By using a multi-core processor or cloud computing platform, the processing speed and scalability of the system can be significantly improved. Gorbunova and Plotnikov [7] chose music datasets and emotion recognition models during the domain definition stage. In the music feature extraction stage, the main focus is on extracting relevant features that are useful for emotion recognition. During the emotion recognition and classification stage, emotion labels were predicted. The two parts mentioned in the article that have the greatest impact on the accuracy of music emotion recognition are the extraction of music emotion features and the establishment of emotion recognition models. In the early days, research on music emotions often used a single music emotion feature or based on traditional machine learning models. Doing so would make the results less generalizable, as the way each piece of music expresses emotions is not necessarily the same. So when performing another music dataset recognition task, it is necessary to re-extract, which is very inefficient. Moreover, the accuracy of recognition will be greatly reduced, so the key breakthrough in improving the accuracy of music emotion recognition lies in music perception.

The primary objective of this investigation is to address the optimization challenges encountered in the realm of multi-voice music arrangement and collaborative design. Initially, the study presents an overview of the research context and its aims. Subsequently, a comprehensive elaboration on the underlying principles of the optimization algorithm and its integration within computer-aided multi-voice music arrangement is provided. Following this, the algorithm's performance is rigorously evaluated through simulation experiments, and the results obtained are subjected to a thorough analysis. Ultimately, the research findings are summarized, and potential avenues for future exploration are outlined.

The innovation of this article mainly includes the following aspects:

Algorithm application innovation: This article applies the optimization algorithm to the field of multi-voice music arrangement, which breaks through the limitation of traditional manual arrangement and realizes the automatic and intelligent music arrangement process. This innovation not only improves the efficiency of music arrangement but also finds a better arrangement scheme through the global search ability of the algorithm, which improves the overall quality of music works.

Experimental design innovation: This article designs a series of rigorous simulation experiments to verify the effectiveness of the optimization algorithm in multi-voice music arrangements. During the experiment, not only music works with different styles and difficulties are considered, but also a variety of comparison methods are set up to comprehensively evaluate the performance of the algorithm. This innovative experimental design ensures the objectivity and reliability of experimental results and provides strong support for the application of the algorithm.

Innovation by combining theory with practice: This article not only analyzes the application principle and realization process of optimization algorithm in multi-voice music arrangement from the theoretical level but also combines theory with practice through simulation experiments to verify the actual effect of the algorithm. This innovative research method makes the research results of this article have both theoretical value and practical guiding significance, which provides a useful reference for researchers in related fields.

Innovation of music creation: By introducing an optimization algorithm, this article provides a brand-new solution for multi-voice music creation. This innovative way of music creation not only improves the efficiency of creation but also provides a broader creative space and source of inspiration for composers. At the same time, this innovative way of music creation is also helpful to promote the popularization and development of music art, so that more people can enjoy wonderful music works.

2 RELATED WORK

In the field of music audio restoration, generative adversarial networks can be used to learn the mapping relationship between raw audio and compressed audio. By training generative adversarial networks, Lattner and Nistal [8] learned to recover the details and features of the original audio from compressed audio. The key to this method lies in building a suitable network structure and selecting the appropriate loss function and optimization algorithm for audio processing. Used for training generative adversarial networks. Then, by constructing the network structure of the generator and discriminator, and defining the loss function and optimization algorithm, the network is trained and optimized. During the training process, the generator continuously attempts to generate more realistic audio data, while the discriminator strives to distinguish whether this data is real or generated. By continuously iterating and adjusting parameters, a generative adversarial network model that can effectively recover compressed audio can be obtained. The automatic accompaniment technology based on computers realizes the automatic generation of music accompaniment through the algorithm, which provides more possibilities for music education. In the process of collaborative design, the application of optimization algorithms can further improve the effect of education. Li [9] discussed the optimization algorithm of computer-based automatic accompaniment in the collaborative design of music education and analyzed its application and advantages. Automatic accompaniment technology intelligently analyzes music through computer algorithms and generates qualified accompaniment according to preset rules and patterns. In music education, automatic accompaniment technology can provide students with real-time accompaniment support and help them better understand and feel the rhythm, melody, harmony and other elements of music.

The application of computer CAD technology in the field of music creation is increasingly widespread, providing a new visual tool and means for music arrangement. The technology of computer-aided visual music arrangement, through the combination of computer-aided design software and music creation software, realizes the visualization of music arrangement and brings unprecedented creative experience to music creators. Lima et al. [10] by transforming musical elements into visual graphics and images, music creators can intuitively understand the structure, rhythm and melody of music, so as to better compose music. Human cognition is the process of integrating and learning multiple types of information, and in the process of perceiving a scene, it receives signals from various aspects including vision, hearing, touch, and smell. Further fusion processing is performed to extract multimodal features from music data, by extracting sound wave information as audio and note information of the music itself. Further machine learning is closer to the form of human understanding of the world. Nowadays, multimodal feature fusion is widely used in the field of deep learning. Many achievements have been made in fields such as natural language processing and image processing, but the involvement in music recognition is not yet very deep. So this article focuses on multimodal music feature extraction and then constructs a music emotion model based on deep learning to study music emotion recognition. This method has significant advantages over traditional machine learning and extracting single music features, as it significantly improves the accuracy and efficiency of recognition and classification [11]. As an efficient and intuitive design tool, CAD-aided layout design has gradually been introduced into the layout of music appreciation courses. Pei and Wang [12] analyzed the advantages, implementation steps and significance of CAD-aided layout design of a music appreciation course based on network resources. It aims to explore new ways to improve the teaching quality of music appreciation courses. Through online music libraries, video tutorials, interactive platforms, etc., students can access various types of music works anytime, anywhere, and broaden their musical horizons. At the same time, network resources can also provide real-time updated music information and comments, so that students can timely understand the music trends and enhance their interest in learning.

The application of automatic translation technology in optical music recognition can not only realize the efficient recognition of music scores but also realize the automatic conversion of music scores between different languages, which provides great convenience for music cross-cultural communication. Ríos et al. [13] discussed the coding steps of applying automatic translation to optical music recognition. Before applying automatic translation to optical music recognition, it is necessary to preprocess the music score image. The information that music can convey can be divided into two categories: acoustic information and note information. As a composer's positive creation, note information contains a lot of emotional information, but acoustic information can also contain many key information. It can be used as auxiliary music information for research and application or can be separately applied for music emotion recognition. For music information, the more primitive the information contains, the more complete the information content. The transformation of information generally accompanies a decrease in information content, so selecting appropriate audio features has become a top priority. Rocamora et al. [14] selected MFCC coefficients and PLP coefficients. Because the unique cepstral-based extraction method of MFCC coefficients is more in line with human auditory principles. It is also the most common and effective audio feature extraction algorithm. The PLP coefficient is due to its better robustness against MFCC coefficients. Tan and Yang [15] discussed the principle, application and advantages of music arrangement collaborative design based on a computer three-dimensional auxiliary system, and looked forward to its future development trend. By constructing a three-dimensional space model, the computer-aided three-dimensional system presents the music elements in a three-dimensional way, so that music creators can intuitively observe and operate the music structure. In music arrangement, the three-dimensional auxiliary system can display musical elements such as notes, rhythm and harmony in the form of three-dimensional graphics or animation, so that the creator can more clearly grasp the overall layout and detail changes of music.

The amount of data required for discrete emotion recognition is relatively small, but during the sampling and recognition process, the chord and rhythm information contained in the music itself will be ignored. The emotional features obtained lose their musicality, and how to make the extracted features better reflect the musicality of one's own work becomes the key. However, continuous emotion recognition suffers from overly complex datasets and complex and massive computational processes, and effective computation becomes the key. In summary, combining two recognition methods can make us more effective in achieving emotion recognition [16]. The advantage of kernel density estimation is that it does not need to make any assumptions about the distribution of data, so it can adapt to music score images with different fonts and formats. With the rapid development of information technology, music recognition technology has become a research hotspot in the field of digital music. Especially for the recognition of large-scale multi-mode piano music, its complexity and challenges make the research in this field more meaningful. Yang et al. [17] discussed the principle, application and future development trend of large-scale multi-mode piano music recognition based on market fingerprint. The piano music recognition system based on market fingerprint needs to preprocess the large-scale piano music library, including audio signal acquisition, noise reduction, segmentation and other steps. Then, the market fingerprint technology is used to extract the characteristic fingerprint of each piano music, and the corresponding fingerprint database is established. In the identification process, the system will compare the piano music to be identified with the characteristic fingerprint in the market fingerprint database, and determine the attribution of the music works by calculating the similarity.

These studies provide an important theoretical and technical basis for the automatic arrangement of multi-voice music. However, although researchers have made some achievements, there are still some problems and challenges in computer-aided multi-voice music arrangement. For example, the existing algorithms and models are often difficult to deal with the complex and changeable multi-voice music structure; At the same time, the communication and coordination problems in the process of multi-person collaborative creation also need to be solved urgently. Therefore, this study aims to conduct in-depth research and exploration of these problems and challenges.

3 THEORETICAL FRAMEWORK

Multi-voice music is a musical form composed of multiple independent voices. Each voice has its unique melody line, rhythm and harmony functions, which are intertwined and synergistic, and together form the overall structure and style of music. In multi-part music, the relationship between parts can be harmonious or comparative, which depends on the composer's creative intention and the needs of music performance.

Within the framework of GA, genetic operators such as selection, crossover, and mutation collaborate to influence each model within the population. As a result, the offspring population experiences an exponential growth in the number of models characterized by lower orders, reduced definition distances, and average fitness levels exceeding that of the parent population. This phenomenon can be mathematically described as an increase in favourable patterns replicating exponentially with the progression of generations, influenced by genetic operations:

$$M\ h,t+1 \geq M\ h,t \cdot \frac{f\ h,t}{\bar{f}\ t}\left[1-p_c\frac{\delta\ h}{l-1}-p_m\cdot o\ h\right] \tag{1}$$

Among them, N_{new} $f(h,t)$ is the average fitness of the N_{new} t generation model N_{new} h; N_{new} $\bar{f}(t)$ is the average fitness of the N_{new} t generation population; N_{new} p_c is the probability of hybridization; N_{new} p_m is the mutation probability; N_{new} $M(h,t)$ is the sample number of the N_{new} t generation model N_{new} h; N_{new} l is the number of binary digits.

In the process of multi-part music arrangement and collaborative design, the parameters of each part (such as pitch, rhythm, dynamics, etc.) can be coded into chromosomes, and GA can be used to optimize the combination of these parameters. By defining appropriate fitness functions (such as harmony, complexity, etc.), the advantages and disadvantages of each individual can be evaluated, and excellent individuals can be selected for genetic operation, thus generating a better next-generation population. Finally, after several generations of evolution, we can get a set of optimized multi-voice music parameter combinations, which makes the music reach the optimal state in the overall structure and auditory effect.

4 REALIZATION OF THE OPTIMIZATION ALGORITHM

In computer-aided multi-voice music arrangement, the implementation process of the optimization algorithm designed in this article is as follows:

(1) Algorithm selection

Firstly, according to the characteristics and requirements of the problem, this section selects the appropriate optimization algorithm. In this study, GA is used as the main optimization tool. GA has the advantages of strong global search ability and easy parallelization and is suitable for dealing with complex multi-voice music arrangement problems.

The objective optimization modelling formula is as follows:

$$f\ x = \sum w_i g_i \tag{2}$$

Where w_i, g_i respectively represent the weight of the i constraint condition and the objective function value. The larger the objective function value, the better the satisfaction of the corresponding constraints.

(2) Coding and initialization

Next, it is necessary to encode each element of multi-voice music so that GA can handle it. The coding method can be selected according to the characteristics of specific problems. In this article, binary coding is used. After coding, a group of initial solutions are randomly generated as the initial population of GA.

(3) Fitness evaluation

In GA, the fitness function is used to evaluate the quality of each solution. In the problem of multi-voice music arrangement, the fitness function can be designed according to the evaluation index defined above. In this article, the indexes of harmony and complexity are weighted and summed to get a comprehensive score as fitness value. For the maximization problem, the fitness function is usually defined as:

$$f\ x = \frac{\text{Objective function}\ \ \text{x}\ \ \text{-minimum value}}{\text{Maximum-minimum value}} \tag{3}$$

Where x is the chromosome and the objective function is the function to be optimized?

(4) Genetic manipulation

GA progressively refines the solutions within the population by employing genetic operations: selection, crossover, and mutation. Selection identifies the individuals who are the most fit for propagation to the subsequent generation. Crossover generates novel individuals by swapping genetic material between two existing ones. Mutation randomly alters the genetic components of individuals, introducing diversity into the population. These genetic operations act synergistically on the population's solutions, driving them towards increasing optimality. The algorithmic framework is illustrated in Figure 1.

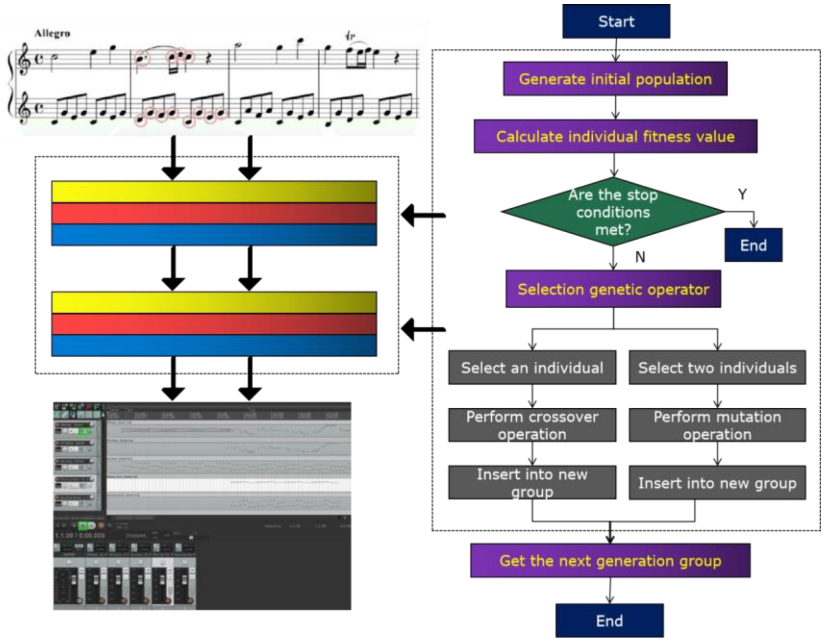

Figure 1: Algorithm model diagram.

The convergence of GA mainly depends on the design and effect of the crossover operator. The Crossover operator is a key operation in GA that combines the individual information of two parents to produce offspring, which is very important for the algorithm's global search ability and convergence speed. Non-uniform linear crossover is a common crossover method in GA, especially when dealing with floating-point coding. When the non-uniform linear crossover operator generates offspring, it does not simply take the middle value of the parent or mix it in a fixed proportion but determines the intersection according to some nonlinear change. To achieve an effective search, this adaptation typically relies on evolutionary algebra or alternate metrics to modulate the intensity of crossover. Initially, this ensures robust global exploration, gradually shifting focus towards meticulous fine-tuning to enhance the algorithm's convergence precision. In the context of non-uniform linear crossover, the genetic value of offspring resulting from crossover is computed using a specified formula:

$$\begin{cases} x_1' = r_1 x_1 + (1 - r_1) x_2 \\ x_2' = r_2 x_2 + (1 - r_2) x_1 \end{cases} \tag{4}$$

Among them, $r_1 \in (0,1), r_2 \in (0,1)$ are randomly generated.

$$P_c = \frac{1}{2 + 0.8 \ln G} + \varphi \tag{5}$$

Where G is an evolutionary algebra; φ is the convergence limit of crossover probability. For a given population of size n:

$$P = \{a_1, a_2, \cdots, a_n\} \tag{6}$$

Among them, the fitness value of the individual a_j is $f(a_j)$. The selection probability is:

$$p(a_j) = \frac{f(a_j)}{\sum_{i=1}^{n} f(a_i)}, \quad j = 1, 2, \cdots, n \tag{7}$$

Significant disparities in the fitness values among individuals within a population result in a substantial surge in the probability ratio between the fittest and least fit individuals during the selection process, exhibiting an exponential growth pattern. Therefore, in the next generation, the best individual will have a higher chance of survival and reproduction, while the worst individual may face a situation in which the chance of survival is greatly reduced or even completely deprived. The algorithm curve is shown in Figure 2.

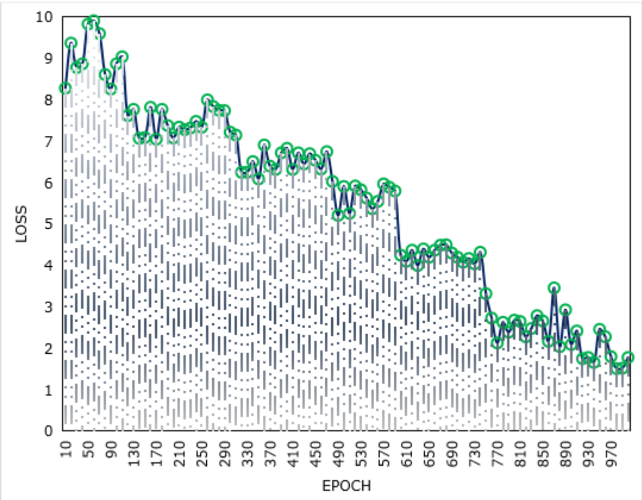

Figure 2: Algorithm graph.

In order to make individual selection more systematic, the sorting selection method is introduced. This method first sorts all individuals in the population according to their fitness values from high to low, forming an ordered sequence. Then, according to the pre-designed probability distribution scheme, different selection probabilities are given to each individual in the sequence. In the ranking and selection method, the individual's selection probability is closely related to his position in the ranking of fitness values, and individuals with high fitness values are usually given a higher selection probability. The calculation formula is as follows:

$$p\ x_j\ =\frac{1}{n}\left(\eta^{+}-\frac{\eta^{+}-\eta^{-}}{n-1}\ j-1\ \right),\qquad j=1,2,\cdots,n \tag{8}$$

Following the selection process, the anticipated count of individuals possessing the optimal fitness function value is denoted as η^{+}, whereas those with the poorest fitness function value are represented by η^{-}. Meanwhile, the expected numbers for the remaining individuals are organized based on the principles of arithmetic progression.

(5) Termination conditions and result output

The termination conditions of GA can be set according to the characteristics and requirements of the problem. Common termination conditions include reaching the maximum number of iterations and finding a solution that meets the accuracy requirements. When the algorithm meets the termination condition, this article will output the optimal solution for the current population as the result of a multi-voice music arrangement. This result can be a complete musical work or an optimized combination of musical parameters for the composer to create further and edit.

5 SIMULATION EXPERIMENT AND RESULT ANALYSIS

5.1 Experimental Process

Simulation experiments were conducted in this study to assess the efficacy and advantage of the optimization algorithm in computer-assisted multi-voice music arrangement. Music emotion space recognition based on continuous signals treats music data as analog signals; that is, music is viewed as a continuous state. This type of information is usually subjected to periodic sampling, but quantization processing can result in certain quantization noise and distortion. In order to reduce the impact of quantization noise and distortion on music emotion recognition, data preprocessing has become an indispensable part. This article will adopt preprocessing methods such as pre-emphasis, windowing, and framing to make music information have short-term stationarity. The preprocessed sound wave information cannot directly obtain music emotions, so the final step before sending music samples into the classifier is to extract emotional features from the music samples. The emotional feature extraction adopted in this article is MFCC and PLP features, which are used as inputs for the constructed WLDNN. The fusion of the two features is carried out in the high-dimensional space of WLDNN to preserve the original music features to the greatest extent possible. At the same time, the preprocessed music data is more representative, laying the foundation for the subsequent recognition part. Finally, sentiment recognition was performed in the GAN network and compared horizontally with current mainstream sentiment recognition models, resulting in the final VA predicted regression value.

Music emotion space recognition is based on discrete signals, treating music data as digital signals. When music data is viewed as a discrete state, its musical theory characteristics become particularly prominent. The chord information of the music itself can provide us with very effective information for emotional analysis. The analysis method adopted in this article is the Schenkel analysis method. Extract the most representative emotional segments from the movement using the Schenkel analysis method and input them into the improved digital network. Add self-attention mechanism to the previous GAN network structure, construct SAGAN, and conduct horizontal comparative experiments in the selected music database.

5.2 Experimental Result

After a series of simulation experiments, we obtained a wealth of experimental results and data. These data include the optimal solution, fitness value, running time and other key indicators of each experimental group. In order to display the experimental results more intuitively, we also use charts and visualization tools to process and display the data.

Figure 3 shows the running time of GA.

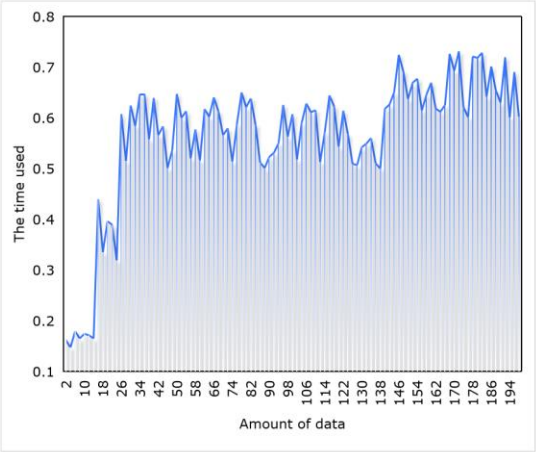

Figure 3: Running time of GA.

A piece of music can be divided into many frames, and each frame of speech can correspond to a spectrum after fast Fourier transform calculation, reflecting the relationship between frequency and energy; that is, the amplitude of different frequencies is different. A spectrogram should reflect the relationship between all frequencies and energy, and the corresponding image is often a large one. In order to obtain sound features that are suitable for final music emotion recognition, it is usually necessary to input them into a Mel scale filter group and transform them into Mel spectra. Figure 4 shows the convergence speed of GA.

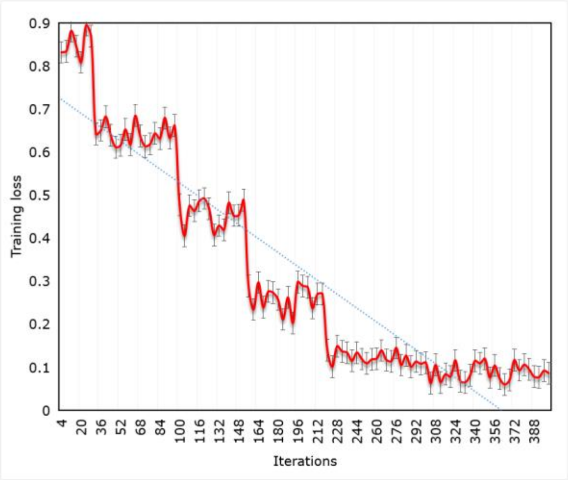

Figure 4: GA convergence rate.

In addition, this article also uses the random arrangement method and the rule-based arrangement method to compare with the optimization algorithm arrangement method in this article. Figure 5 shows the running time comparison of different methods.

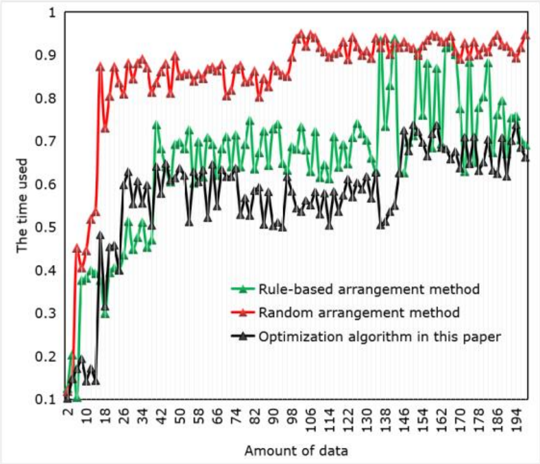

Figure 5: Comparison of running time of different methods.

The results that the running time of the optimization algorithm in this article is shorter than that of the random scheduling method and the rule-based scheduling method. Because of its randomness, the random arrangement method takes a long time to find a satisfactory solution when searching the solution space, so the running time is relatively long. Although the rule-based arrangement method is arranged according to certain rules, the formulation and implementation of rules involve complex logical operations, which leads to unsatisfactory running time. The optimization algorithm in this article can find a better arrangement scheme in a short time through a reasonable search strategy and efficient calculation method, which significantly improves the operation efficiency.

Figure 6 shows the comparison of the convergence speed of different algorithms.

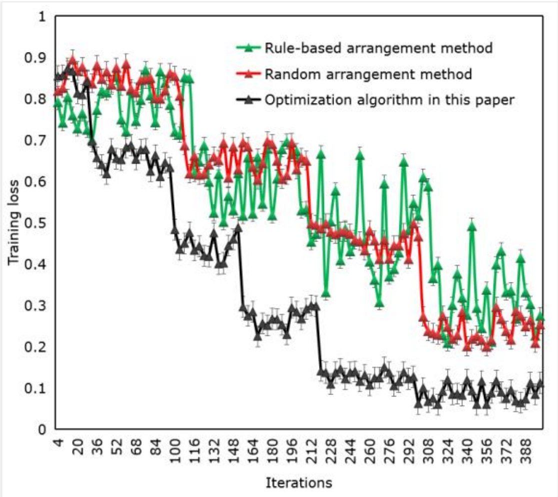

Figure 6: Comparison of convergence speed of different algorithms.

The findings indicate that the optimization algorithm presented in this article exhibits a notably swifter convergence rate compared to both the random and rule-based arrangement methods. The random approach is characterized by inconsistent convergence due to its inherent unpredictability. On the other hand, the rule-based method is constrained by the design of its rules, often resulting in a tendency to settle for local optima, thus slowing down convergence. In contrast, the optimization algorithm featured in this study employs a comprehensive search strategy, continuously refining the quality of solutions through genetic operations. This allows it to swiftly converge towards the globally optimal solution, explaining its superior convergence speed.

Table 1 shows the iterative records of the multi-voice music arrangement of the optimization algorithm.

Iterations	*Optimal solution description*	*Fitness value*
15	Initial arrangement, based on original data, without optimization.	80.2
25	The pitch of the melody part is adjusted, and the appearance of discordant intervals is reduced.	85.5
35	Optimize the harmony structure, and increase the diversity of chords and the fluency of conversion.	92.3
45	A complex rhythm pattern is introduced to improve the rhythm contrast between multiple voices.	95.8
55	Comprehensive optimization: the timbre distribution is adjusted, which enhances the overall harmony and innovation.	98.6

Table 1: Iterative recording of multi-voice music arrangement based on optimization algorithm.

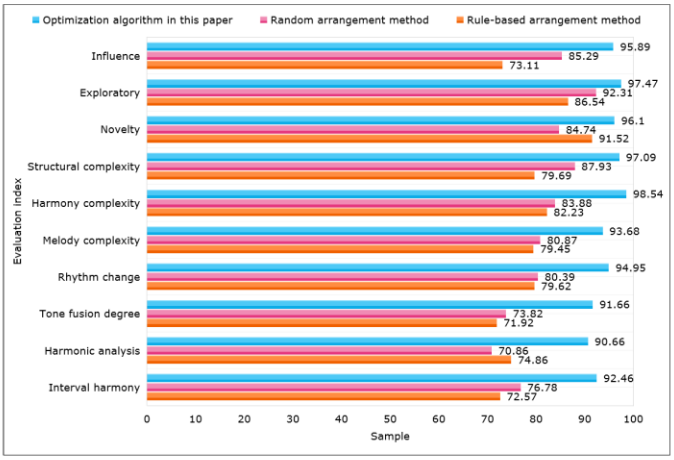

Figure 7: Experimental results of evaluation indexes with different arrangement methods.

Among them:

The "Iterations" column indicates the running times of the algorithm, and it is optimized step by step from the initial arrangement.

The column "Optimal solution description" details the specific optimization adjustments made by the algorithm in each iteration, which is aimed at improving the harmony, complexity and innovation of multi-voice music arrangement.

The fitness value column records the fitness value calculated after each iteration, which is used to quantitatively evaluate the quality of the arrangement scheme. The higher the fitness value, the better the quality of the arrangement scheme.

The experimental results of evaluation indexes of different arrangement methods are shown in Figure 7.

6 CONCLUSIONS

The main content of this article is to establish a neural network model for emotion recognition and related applications by extracting multimodal music features from music arrangement samples. Extract music features from both discrete and continuous emotion models and integrate them into the constructed neural network model. Improving the music emotion recognition model through three steps: preprocessing, feature extraction, and model optimization, thereby improving the accuracy of recognition. At the end of the article, an intelligent music system was designed, embedded with emotion recognition and music generation modules. The main work done in this article is as follows:

(1) Recognition based on continuous emotional space, treating music data as analog signals. When processing this type of information, it is usually cyclically sampled, but quantization processing can bring a certain degree of quantization noise and distortion. Therefore, preprocessing is an indispensable part. This article will adopt preprocessing methods such as pre-emphasis, windowing, and framing to make music information have short-term stationarity. The preprocessed sound wave information does not have emotional characteristics, so the last step before feeding the music sample into the classifier is to input the music sample

Extracting emotional features.

(2) The emotional feature extraction adopted in this article is MFCC and PLP features, which are used as inputs for the constructed WLDNN. Fusion of two features in the high-dimensional space of WaveNet to preserve the original music features to the greatest extent possible. The preprocessed music data is more representative, laying the foundation for the subsequent recognition part. Finally, sentiment recognition was performed in the GAN network and compared horizontally with current mainstream sentiment recognition models, resulting in the final VA predicted regression value. The music arrangement model constructed in this article has shown good performance in emotion recognition, but there are still some shortcomings. The future prospects for improving music emotion analysis are as follows;

The model constructed in this article has a complex structure and relies on the volume of the original dataset and the accuracy of music sentiment annotation, resulting in poor adaptability of the model. The goal of future development is to improve the model and have good adaptability to situations with poor sample quality.

The ratio ratio ratio used in the fusion of music featured in this article needs further improvement and optimization. The evaluation system for the results in this article is too singular, only judging based on traditional performance indicators, and cannot specifically demonstrate the rationality of the internal construction of the model. In the following work, we need to optimize the model further, find models that can ensure good adaptability for different music datasets and have a wider range of applications, and improve the evaluation mechanism to conduct more comprehensive and accurate evaluations and comparisons of the models.

Due to a lack of experience in building intelligent music platforms, optimization is needed in terms of user usage and interface aesthetics. In terms of user experience, some functions can be added, such as emotion recognition based on the user's recent music listening, to determine the user's preference for music emotions at this time and to recommend songs related to emotions accordingly

Bing Tian, https://orcid.org/0009-0009-3508-1772
Jie Tian, https://orcid.org/0009-0004-7258-8429

REFERENCES

[1] Alfaro, C.-M.; Valero, M.-J.-J.: Exploiting the two-dimensional nature of agnostic music notation for neural optical music recognition, Applied Sciences, 11(8), 2021, 3621. https://doi.org/10.3390/app11083621
[2] Andrea, P.; Zahra, A.: Music note position recognition in optical music recognition using convolutional neural network, International Journal of Arts and Technology, 13(1), 2021, 45-60. https://doi.org/10.1504/IJART.2021.115764
[3] Baro, A.; Riba, P.; Calvo, Z.-J.; Fornés, A.: From optical music recognition to handwritten music recognition: a baseline, Pattern Recognition Letters, 123(5), 2019, 1-8. https://doi.org/10.1016/j.patrec.2019.02.029
[4] Calilhanna, A.: Ogene Bunch music analyzed through the visualization and sonification of beat-class theory with ski-hill and cyclic graphs, The Journal of the Acoustical Society of America, 148(4), 2020, 2697. https://doi.org/10.1121/1.5147469
[5] Fu, Y.; Zhang, M.; Nawaz, M.; Ali, M.; Singh, A.: Information technology-based revolution in music education using AHP and TOPSIS, Soft Computing, 26(20), 2022, 10957-10970. https://doi.org/10.1007/s00500-022-07247-w
[6] Ge, M.; Tian, Y.; Ge, Y.: Optimization of computer-aided design system for music automatic classification based on feature analysis, Computer-Aided Design and Applications, 19(S3), 2021, 153-163. https://doi.org/10.14733/cadaps.2022.S3.153-163
[7] Gorbunova, I.-B.; Plotnikov, K.-Y.: Music computer technologies in education as a tool for implementing the polymodality of musical perception, Musical Art and Education, 8(1), 2020, 25-40. https://doi.org/10.31862/2309-1428-2020-8-1-25-40
[8] Lattner, S.; Nistal, J.: Stochastic restoration of heavily compressed musical audio using generative adversarial networks, Electronics, 10(11), 2021, 1349. https://doi.org/10.3390/electronics10111349
[9] Li, Y.: Application of computer-based auto-accompaniment in music education, International Journal of Emerging Technologies in Learning (IJET), 15(6), 2020, 140. https://doi.org/10.3991/ijet.v15i06.13333
[10] Lima, H.-B.; Santos, C.-G.-R.-D.; Meiguins, B.-S.: A survey of music visualization techniques, ACM Computing Surveys, 54(7), 2021, 1-29. https://doi.org/10.1145/3461835
[11] Maba, A.: Computer-aided music education and musical creativity, Journal of Human Sciences, 17(3), 2020, 822-830. https://doi.org/10.14687/jhs.v17i3.5908
[12] Pei, Z.; Wang, Y.: Analysis of computer-aided teaching management system for music appreciation course based on network resources, Computer-Aided Design and Applications, 19(S1), 2021, 1-11. https://doi.org/10.14733/cadaps.2022.S1.1-11
[13] Ríos, V.-A.; Esplà, G.-M.; Rizo, D.; Ponce, L.-P.-J.; Iñesta, J.-M.: Applying automatic translation for optical music recognition’s encoding step, Applied Sciences, 11(9), 2021, 3890. https://doi.org/10.3390/app11093890
[14] Rocamora, M.; Cancela, P.; Biscainho, L.-W.-P.: Information theory concepts applied to the analysis of rhythm in recorded music with recurrent rhythmic patterns, Journal of the Audio Engineering Society, 67(4), 2019, 160-173. https://doi.org/10.17743/jaes.2019.0003
[15] Tan, B.; Yang, F.: Dance movement design based on computer three-dimensional auxiliary system, Computer-Aided Design and Applications, 19(S3), 2021, 108-118. https://doi.org/10.14733/cadaps.2022.S3.108-118

[16] Wang, Q.; Zhou, L.; Chen, X.: Kernel density estimation and convolutional neural networks for the recognition of multi-font numbered musical notation, Electronics, 11(21), 2022, 3592. https://doi.org/10.3390/electronics11213592

[17] Yang, D.; Goutam, A.; Ji, K.; Tsai, T.-J.: Large-scale multimodal piano music identification using marketplace fingerprinting, Algorithms, 15(5), 2022, 146. https://doi.org/10.3390/a15050146