Evolutionary Deep Learning for Sequential Data Processing in Music Education

Lin Jing
School Quanzhou Normal University; Quanzhou Fujian, 362000, China
E-mail: lj2771258@163.com

Keywords: sequential data processing, evolutionary deep learning, music education applications

Received: November 20, 2023

In response to the shortcomings of insufficient music structure, this article proposes a structured model based on motivational phrases and phrases. Starting from the composition structure of motivational phrases, deep learning techniques are used to learn composition. In the music generation model, a Scratch music generation model that can generate Pianoroll format music is constructed by using a generative adversarial network based on emotions and time structures. And use convolutional neural networks in the generator and discriminator to improve training speed. The effectiveness and practicality of the two algorithm models were verified through multiple comparative experiments and algorithm effectiveness experiments. This method achieves structural feature extraction of music by designing feature extractors at different music granularities. By designing feature expression functions at multi-scale music granularity, the music structure embedded in the music itself is incorporated into the reward function. Use forward backward propagation method to update the parameters of the model, and use dropout technique to improve the model's ability to resist overfitting. The test results show that the model has specific generalization ability, with an accuracy rate of 90%, and high recall and accuracy of the model. The experimental results show that this method can achieve better music generation results than the reward function method based on manual rules and before and after relationships. Solved the problem of lacking knowledge of music theory to propose rules, and compensated for the pain of insufficient utilization of music structure information in network models based on context.

Povzetek: Študija uvaja evolucijsko globoko učenje za glasbeno izobraževanje, ki uporablja GAN za generiranje glasbe in CNN za izboljšanje hitrosti učenja.

1 Introduction

Introducing music-disciplinary core literacy is an essential symbol of today's deepening music curriculum reform. Unlike the established music curriculum that focuses on the learning of music subject knowledge and music skills, the subject core literacy means that music teaching should shift from teaching content-based instruction to student development-based education [1]. In this context, highlighting students' subjectivity in the classroom and developing comprehensive music literacy has become an important development direction of music classroom teaching reform in the era of core literacy. For front-line teachers, how can they highlight the centrality of students in teaching, thus prompting classroom teaching changes and innovations? This is not only an important sign of classroom transformation but also a typical feature of reflecting a student-centred classroom. Therefore, deep learning around students' active participation and inquiry has become a meaningful way to transform classroom teaching today. Music education must move towards a new stage focusing on students' all-around development based on promoting their learning of basic music knowledge and skills [2]. Therefore, music teaching should go beyond the machine learning and training of specific knowledge and skills, emphasize the perception and experience of music, guide students to participate in activities such as music creation, music expression, and music understanding, and form a new type of learning based on independent inquiry and active construction. In a word, saying goodbye to the mechanical and passive knowledge of the past and moving towards in-depth and inquiry-based learning is an essential choice for music teaching in the new era to develop students' core literacy in the subject [3].

Promoting music curriculum reform and teaching innovation with deep learning is a crucial way to achieve the fundamental goal of moral education in music curricula in the current era of core literacy. To this end, this paper selects a new perspective of deep learning. It conducts a theoretical argument and practical exploration of how the high school music curriculum reflects a new ecology of student-centered classroom teaching. This paper is a theoretical demonstration and a practical exploration of how a new ecology of student-centered classroom teaching can be reflected in the high school music curriculum. To achieve this goal, the thesis starts from the perspective of the fun nature of the learning mode of the smart education platform. It introduces the game mode into the platform through deep reinforcement learning, which has received much attention in recent years, to increase students' interest in using the smart education platform without putting too much effort into the game, avoiding students' burnout in the face of the

unchanging learning mode, and thus motivating them to engage in more active learning activities [4]. To a certain extent, deep reinforcement learning is like the human learning process, which can be summarized as follows: human beings interact with the real environment through different perceptual organs to obtain a large amount of state information, which is processed by the brain to extract practical information, produce corresponding decision-making behavior, and make a judgment on the merits of the decision, and complete learning through this process of trial and error. In contrast, deep reinforcement learning obtains data information by interacting with the environment in the simulation environment created for it and outputs action data after processing by neural networks.

As the educational objectives of primary education in the new era are constantly updated and transformed, the classroom is shifting from a rigid one-way teaching style to a vibrant and conversational life classroom. The trend from teacher-oriented to student-oriented education is dismantling the dull classroom dominated by knowledge and lack of emotion and thought. A new type of dialogic classroom is quietly sprouting. Thus, achieving deep learning in high school music is the best way to solve the problem of superficiality in the current high school music classroom teaching process. This paper defines the unit concept of music discipline from the perspective of deep learning through literature analysis, case study analysis, and summary and induction methods. By reading and studying relevant literature, we familiarize ourselves with the status of research and learn from research experience to provide theoretical support for this paper. Secondly, through a case study of a junior high school music classroom in Changzhou, we explore the critical elements of "unit teaching" in music, develop ideas for unit teaching design, evaluate the effectiveness of research and future development trends, and try to explore teaching strategies and methods to improve the overall level of unit teaching in music classrooms. The proposed model has higher accuracy and stability in predicting, recognizing, and generating music sequences compared to the SOTA method. This is mainly due to the optimization of evolutionary deep learning algorithms, which enable the model to better learn and adapt to the complexity and diversity of music sequences. Although the SOTA method performs well on certain specific tasks, the proposed model exhibits stronger generalization ability in cross domain transfer. This is due to the consideration of more music features and contextual information in the design and training process of the model, which enables the model to better adapt to different music education and application scenarios. The proposed method has novelty in the following aspects:

This article combines evolutionary deep learning algorithms and applies them to music education, which is an innovation of this method. By utilizing the optimization ability of evolutionary algorithms, the performance and generalization ability of deep learning models can be further improved.

The proposed model has strong cross domain transfer ability and can adapt to different music education and application scenarios. This is due to the consideration of more music features and contextual information in the design and training process of the model.

3. Compared to some SOTA methods, the proposed model has a more concise structure and fewer parameters. This not only reduces the complexity of the model, but also improves the training efficiency and generalization ability of the model.

2 Related work

Abd Elaziz researched the launch of the core literacy research project by the Organization for Economic Cooperation and Development in China, which led to some important insights on constructing the correct definition and definition of core literacy in China [5]. She argues that the basic premise of core literacy selection must align with society's needs and personal visions, emphasizing harmonious communication between people and tools, people, and individuals, etc. The development of deep learning relies on the implementation of deep teaching in the classroom [6]. Deep teaching is a type of teaching that focuses on teaching knowledge to convey the meaning of knowledge and the values behind it [7]. The most important thing in achieving deep learning is constantly questioning whether core literacies are being implemented. According to Lu, authentic teaching is conducted when students understand why and how knowledge exists, how it is developed, and when the learning is integrated into their individual experiences [8]. Vrysis believes that deep learning requires attention to each student's real needs and non-intellectual factors such as interests, aspirations, ideals, ideologies, emotions, attitudes, and values in the development process [9]. However, the data-driven approach requires a large amount of data, and copyright issues limit the amount of data and make manual labeling efforts more inefficient. To address this problem, combining deep neural networks and re-refining salient features has proposed a method that has yielded promising results [10].

Deep reinforcement learning algorithms based on policy gradients have demonstrated the ability to solve problems for high-dimensional continuous problems, compensating for the shortcomings of value-based algorithms and significantly improving the applicability of deep reinforcement learning [11]. In addition to these algorithms, several other deep reinforcement learning algorithms are emerging, such as hierarchical reinforcement learning attempts to solve the problem of reward sparsity and reverse reinforcement learning algorithms, such as those for solving the problem of hard-to-get rewards during interaction with the environment [12]. These algorithms are constantly being improved, allowing deep reinforcement learning techniques to play an essential role in an increasing number of areas. Many excellent algorithms are still being proposed and applied in various fields [13]. Deep reinforcement learning is still developing rapidly, and there are still many challenges to be overcome, such as how to accelerate the training process more effectively, how to make the trained model more general, how to set a more accurate and reasonable

reward function, and how to choose the current strategy according to the longer-term return. Still, the fantastic achievements of deep reinforcement learning at this stage prove that deep reinforcement learning has a very broad [14]. With the development of hardware and algorithms, deep reinforcement learning will be able to play a more significant value [15].

The processing of sequential data in music education requires a large amount of annotated data, including timing information of notes, pitch information, rhythm information, etc. However, currently there are certain difficulties in obtaining and organizing annotated data, which require a lot of manpower and time. Meanwhile, the quality of annotated data can also have an impact on the training and performance of the model. The complexity of music sequences is high, requiring models to have high representation and learning abilities. However, current music education models based on deep learning often suffer from high model complexity, leading to long training time, high computational resource consumption, and also prone to overfitting. The sequence data in music education has diversity and complexity, and models need to have good generalization ability. However, current music education models based on deep learning often have certain limitations in terms of generalization ability, making it difficult to adapt to various complex music education scenarios.

The method studied in this article can achieve better music generation results than the reward function method based on manual rules and contextual relationships. Solved the problem of lacking knowledge of music theory to propose rules. Compensated for the pain of insufficient utilization of music structure information in network models based on the relationship between before and after.

3 Evolutionary deep learning sequence data processing methods analysis

Since the music relative loudness estimation task is derived from the music detection task by further classifying music events as foreground music events or background music events, it can be observed that the event categories of these two tasks naturally form a two-level hierarchical structure. Based on this observation, the joint task of music detection and music relative loudness estimation is highly relevant to the hierarchical classification problem [16]. Evolutionary algorithm is an optimization algorithm that draws inspiration from natural selection and genetic mechanisms in biological evolution. In deep learning, evolutionary algorithms can be used to optimize the parameters and structure of neural networks, thereby improving the performance of the model. Specifically, evolutionary algorithms can affect the training or structure of neural networks in the following ways. Evolutionary algorithms can be used to optimize parameters such as weights and biases in neural networks. In each iteration, the parameters of the neural network are evaluated based on the fitness function, and the parameters with higher fitness are selected for genetic operation, gradually optimizing the parameters of the neural network. Evolutionary algorithms can also be used to optimize the structure of neural networks, such as the connection method and number of layers of neurons. By simulating the genetic mechanisms involved in biological evolution, different network structures can be generated and evaluated under fitness functions. Select a network structure with better performance for genetic operations, in order to gradually optimize the structure of the neural network. For a segment on a time step, detecting events on both tasks can be constructed by classifying the segment into two hierarchical levels of event classes. The segments from the audio are time-series dependent, especially those that are temporally adjacent. This is because segments are short, but an event may last several seconds or minutes. This means that a series of adjacent segments in a period may belong to the same event class. Therefore, models need to be designed that can guarantee the continuity of an event and can model the temporal relationships between time steps. Recurrent neural networks have shown some advantages in modeling sequential data, so the same iterative structure as recurrent neural networks is used to improve the performance of continuous event detection in the study of this chapter.

$$Target Q = \frac{\gamma \max Q(s', a, \theta_i)}{r} \tag{1}$$

The input of the neural network of DQN is the observation, which is generally the state s. The deep neural network calculates the value function of each action under the input states, and then the c- greedy exploration strategy described in Section 1 is used to select one of the actions as the output. The process of updating the value function matrix by exploration can be described as follows: firstly, the observed value is obtained by observation, i.e., the current state s. The Agent brings the value of the value function Q (s, a) about each action and in state's according to the Q value stored in the value function matrix and then selects an action a from the steps according to the exploration strategy used and executes a. The environment at the next moment after the action is performed will change because based on this, the DQN updates the parameters of the value function matrix according to the obtained reward r and conducts the next round of iterative training until a sufficiently good value function matrix is obtained, and the structure of the DQN is shown in Figure 1.

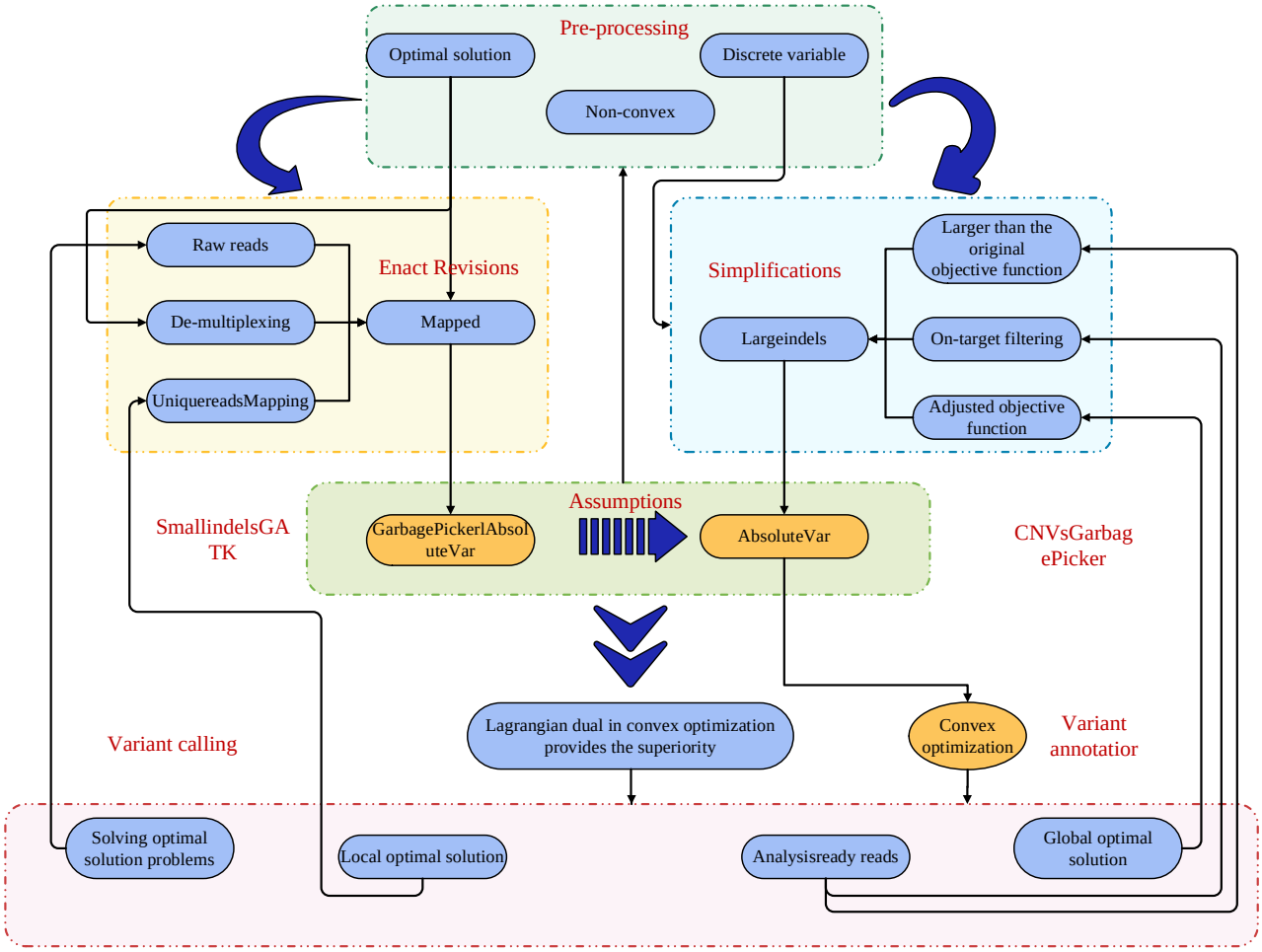

Figure 1: Sequence data processing method

However, applying deep learning directly to reinforcement learning is problematic, and two of these problems pose numerous difficulties in combining the two. Reinforcement learning Q-learning algorithms are updated iteratively by the payoffs at the current moment and the estimated value at the next moment, making the present Q strongly correlated with the future Q [17]. There are often discrepancies in the data, and this instability in the data leads to reduced validity of the data, which may fluctuate with each round of iterations as a result and will have an impact on subsequent iterations to the detriment of the algorithm's convergence. In addition, this type of task may also have problems such as reward delay; in reinforcement learning, the reward generated by the action may be reflected in a reasonably long period, while the deep learning method input and output is generally a direct mapping, the training of reinforcement learning is relatively much more difficult.

$$L_i(\theta_i) = E_{s,a,r,s^i}\left[\frac{\gamma\, max\, Q\,(s', a, \theta_i)}{rQ(s, a, \theta_i)^2}\right] \quad (2)$$

whereθ_i^- is the parameter of the Target Q-network for the ith iteration. It is the parameter of the Q-network after the ith iteration. The loss function of the DQN is a residual model that calculates the square of the difference between the actual value and the estimated value. Represents the estimated value, which is also used as the input to the neural network.

Sample data for deep learning are usually independently and identically distributed. Still, in reinforcement learning, the states as training samples are a sequence and they are highly correlated with each other. The value function and action value estimates are constantly optimized and updated as training proceeds. As the value function changes, the output actions also change continuously, leading to a changing distribution of the training samples. This strong correlation of the reinforcement learning data samples is incompatible with the nature of the data samples required for training by deep learning. In addition, there is also the problem of inefficient data use. Supervised deep learning algorithms mostly need a large amount of data as support to achieve good results, while reinforcement learning algorithms face generally sparse data tasks; after each iteration, the sample data used this time is directly discarded, then more interaction with the environment is needed to obtain samples to prepare for subsequent training.

$$\frac{\partial L_i}{} = E_{s,a,r,s^i}\left[\frac{\gamma\, max\, Q\,(s', a, \theta_i) - \Delta_{\theta_i} Q(s, a, \theta_i)}{Q(s, a, \theta_i)}\right] \quad (3)$$

The policy-based reinforcement learning algorithm can effectively solve the problems as mentioned earlier of value-based algorithms. In the task, the policy-based approach provides an approximate representation of the randomized policy by describing the policy r as a function containing parameter 0, i.e.:

$$\pi\theta(a \mid s) = p(a \mid s) / p(\theta) \tag{4}$$

By representing the strategy as a continuous function, the optimal strategy can be found by the optimization method of continuous functions; the most common way is gradient descent. The performance of strategy r is measured by the expected return of the action trajectory, and it is used as the optimization objective:

$$J(\theta) = \left[\sum_{a}^{n} Q(a \mid s)\, \pi(s, a)\right] \tag{5}$$

When using the value function-based approach for continuous tasks, the action space must be discretized, and the constant space must be simulated by other means, which leads to an increase in workload and a decrease in the accuracy of model training [18]. Using the policy gradient method, the optimal policy is obtained as a probability distribution or probability density function, which can handle both continuous and discrete state action space tasks, effectively compensating for the shortcomings of the value function-based method.

Like ordinary attention, the attention mechanism attached to a convolutional neural network needs to assign different weights to the data from some dimension, among which the more common one is the channel-based attention mechanism, as shown in Figure 2.

To extract the motives, we need from the massive music dataset, there is no relevant dataset in all music datasets, so for this paper, it is a process from 0 to 1 [19]. For this purpose, we need to manually annotate the existing dataset and then learn the process of motive extraction by the seq2seq model.

$$Accuracyj \frac{1}{D_{[s,s\prime]}{}_{s,s\prime}} \tag{6}$$

The music XML file is selected as the source of music data. Since Mozart's music has strong melodic mobility and apparent motives, the music XML file of Mozart's music is selected as the source of the music data set in this section. Manual annotation extracts the desired music data from the music XML files. And the beginning of the motive and phrase, the beginning note, the end note, and the end note are marked manually. A motive and a phrase are used as a set of data. The format of the data set is such that one row contains one motive sequence and one phrase sequence. All the data are classified into training and test sets in the ratio of 8 to 2.

4 Experimental design for music education application

Considering the characteristics of different modules, in order to verify the modeling ability of the model, this study combined the modules and conducted experiments. Firstly, this article conducted experimental research on the weights of the cosine loss function. The quantizer of the dataset itself is a module for shoppers. During the experiment, it was found that when the weight was set to Ralph, the quantifier of the model was closer to the dataset in terms of lime green than other weights, indicating that the distribution of notes in the measure was closer to the music dataset

The accuracy of a bidirectional LSTM model with single-layer and three-layer hidden layers was evaluated at 1000 iterations. Compared to accuracy, a bidirectional LSTM model with three hidden layers is more reasonable than a two-dimensional LSTM model with a single hidden layer. The selection of the number of hidden layers and corresponding nodes directly affects the reliability and accuracy of the model. Further analysis was conducted on the impact of different hidden layers and node numbers on the model. As a result of the analysis, a reasonable number of hidden layers and corresponding numbers of nodes were given to make the model optimal, i.e. minimizing the error. Before training, determine parameters such as small batch size, learning rate, and optimizer based on previous experience and relevant knowledge. In-depth learning goals are necessary to reflect the characteristics of deep learning. The plans are set to target the structural framework of knowledge learning that students already have and will have, as well as to develop students' transferability and implement the generation of students' overall literacy [20]. In this model, the mutual interpretation of teaching objectives and teaching evaluation serve as the two forward momenta of the spiral occurring in the process of deep learning in high school music, in which the teacher can adequately evaluate the learning phenomena reflected by students after each stage of learning. The results of their feedback evaluation are used to set the teaching objectives for the next session.

Instructional goals that point to deep learning in music should be oriented to learning outcomes and dynamic and reverse instructional goal design, in which we can progressively define instructional purposes by focusing on core music disciplinary literacy and around challenging transferable learning tasks. This is because only studies with authentic contexts and challenges can achieve the educational goal of enabling students to respond to real-world opportunities and difficulties rather than shallow verbal or written responses to limited prompts. For example, an authentic challenge in music is translating a complex set of instructions into a smooth and moving entire repertoire rather than just learning a bunch of notes [21]. Performing a particular piece of music (and appreciating others' performances) reflects a student's proficiency in the challenge; for example, in the Music and Drama module, the central authentic challenge is whether the student can perform or sing on stage with integrity and grace to role-play and form.

First, music education is a crucial way to implement aesthetic education. Students learn music to enhance their experience and perception of beauty, gradually enhance their interest in music learning and their desire for beautiful things in the process of aesthetic experience, and subconsciously improve their aesthetic level and moral sentiment in the rich and vivid music learning, to realize the role of educating people with beauty and beautifying the soul. Secondly, music learning emphasizes practice because music is a creative art. Therefore, students can form direct experience when participating in various kinds of music practice activities, transform this experience into acquired skills, and gain different music learning experiences by participating in different contexts and forms of music activities, which can also enhance students' imagination and creative thinking, as shown in Figure 3.

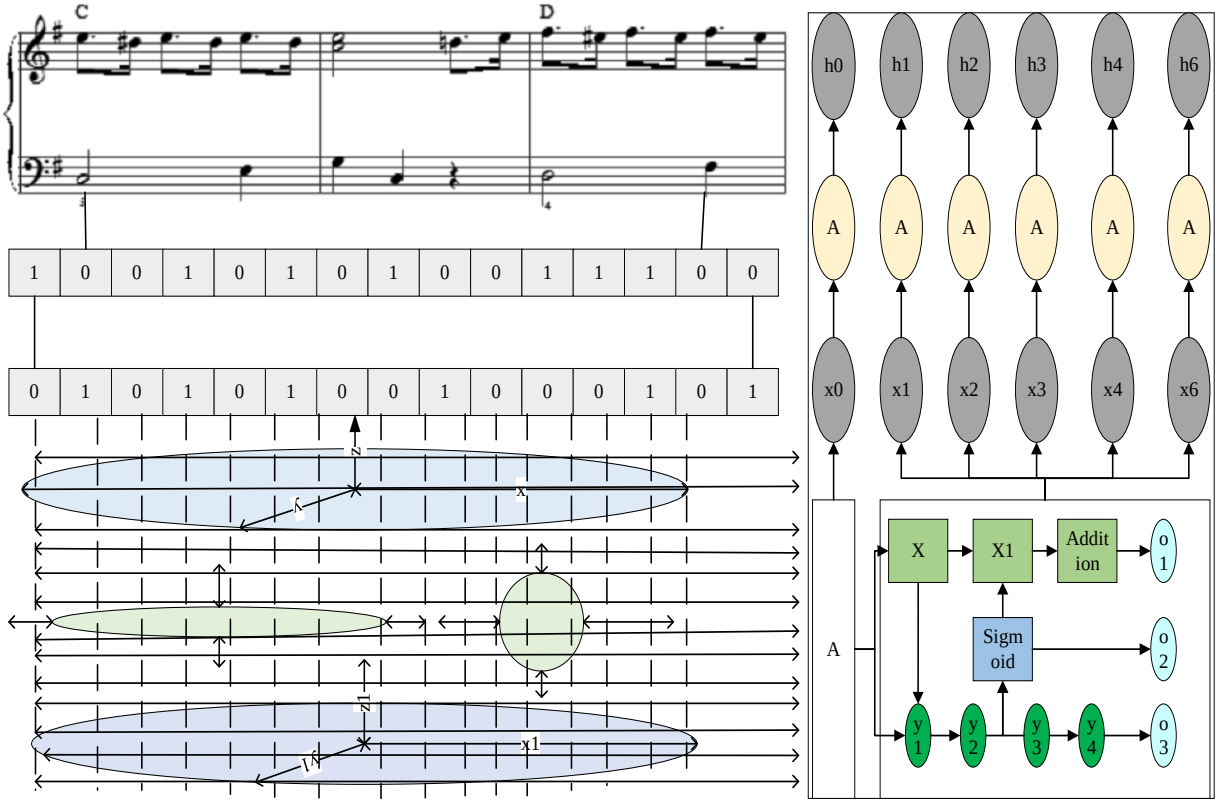

Figure 2: Deep learning sequence data processing framework

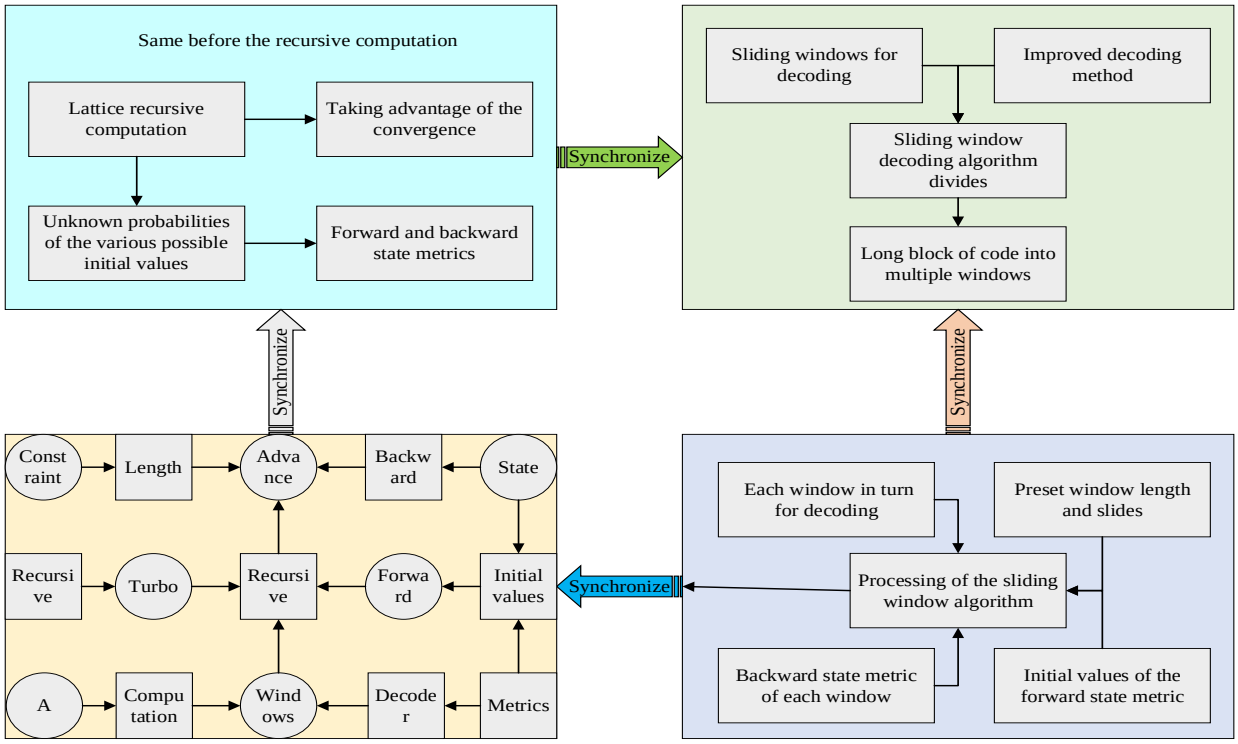

Figure 3: Structure of music teaching model

A sequence of notes is converted into a single solo thermal code and connected to the current structural feature, which is obtained from the previous music structural feature extractor and used as input to the reward model. The reward model consists of a single troubadour connecting a fully connected layer. Thus, the music generation process always incorporates thematic information into the note generation. A note and its probability are generated by the reward model based on the previous note sequence and the theme. The possibility of a note being predicted serves as the actual reward for the reward model [22]. All the 359 theme models contain a similar structure. By integrating different music structure information through theme weights, three music generation models capable of remembering additional music structure information are obtained, finally completing the music reward function model.

Considering the nature of the frivolous naive model, using the maximum likelihood probability principle to select notes leads to a high repetition rate of the generated notes. To solve this problem, this study employs a Boltzmann sampling approach to generate music. The model outputs the probability corresponding to each note and treats it as a polynomial distribution. The size of each note is generated according to the generation probability, and the probability is used as the criterion for sampling, i.e., the probability of predicting that note is used as the probability when generating music. In this example, this study uses a random sequence as the model starter bar. The first note generated is the first output obtained by the model after the input from the starter bar, and then the whole music sequence is generated sequentially, as shown in Figure 4.

In the process of practice, teachers can easily find that the teaching model is often misinterpreted as a means of pursuing efficiency by focusing on progress and conclusions, treating teaching as learning, progress as a task, and teaching materials as a curriculum [23]. Cooperative learning is often reduced to a classroom embellishment of finding answers together, and there is no effective communication, sharing, and division of labor, which is undoubtedly a waste of time and resources. Teachers' reflection is a focus before classroom practice activities through experience, sharing, and communication, to continue to develop their professional capacity and literacy. The breakthrough of reflection and the key to crossing the transformation barrier is the awareness of student learning, "How do students learn in? Teachers should keep exploring in depth from this thread what factors influence student learning and keep adjusting at the right time in this spiral process. To accomplish an excellent harmonic arrangement, one should know that the main reason for the harmonic arrangement process affecting the chord progression is the difference in style and what kind of chords are used to produce a specific effect.

It demonstrates the development process of its basic idea - the basic motive - in a constantly changing environment. These different aspects of the basic motive - its change and development - are shaped by the environment that arises from considerations of diversity, structure, expressiveness, etc. Whereas the arrangement in a photographic book is chronological, the motivic type is not, and its order is governed by the requirements of comprehensibility and musical logic [24]. The phrase is a higher form of construction than the section. It not only states an idea but also develops it immediately. The phrase

form is often used in the dominant themes of sonatas, symphonies, etc., but it is also applicable to smaller forms. The opening of the phrase already contains repetition; therefore, the subsequent section requires a more distantly varied motivic pattern.

5 Analysis of results

5.1 Performance analysis of evolutionary deep learning algorithm results

Music generation is like the problem of text generation in natural language processing, and recurrent neural networks or long and short-term memory recurrent neural networks are often used to solve this sequential problem. However, since the music generated using these methods only solves a sequence generation problem, there is no way to control the emotional tone of the generated music. Generative adversarial networks perform well in solving image generation problems, where the generator and the discriminator reach a Nash equilibrium, and the high-quality images generated by the generator will successfully "trick" the discriminator. Based on this idea, this paper designs a Scratch music generation model based on generative adversarial networks. During training, Scratch music with the same sentiment tone is used as training data for the discriminator, and the sentiment category is added to the feature vector for the generator to learn. When the discriminator cannot distinguish the raw training data from the Scratch music generated by the generator, the generated Scratch music will be labeled with the corresponding sentiment.

Due to the temporal nature of music, solving the music generation problem using generative adversarial networks is more complex than solving the image generation problem. When composers create music, they often describe music as a multi-level hierarchical structure: the beat and note values and pitches are considered as the smallest repetitive structure, and notes, chords, etc combine the music measures. A certain number of music measures are integrated into phrases; the variations of phrases are combined into movements, and finally, multiple movements are incorporated into complete music. The quality of music is directly affected by the temporal dependence between musical measures, so it is essential to model the temporal structure in the generative adversarial network. Moreover, generative adversarial networks are suitable for generating continuous data; for example, the target output of the video generation task is a constant video, as shown in Figure 5.

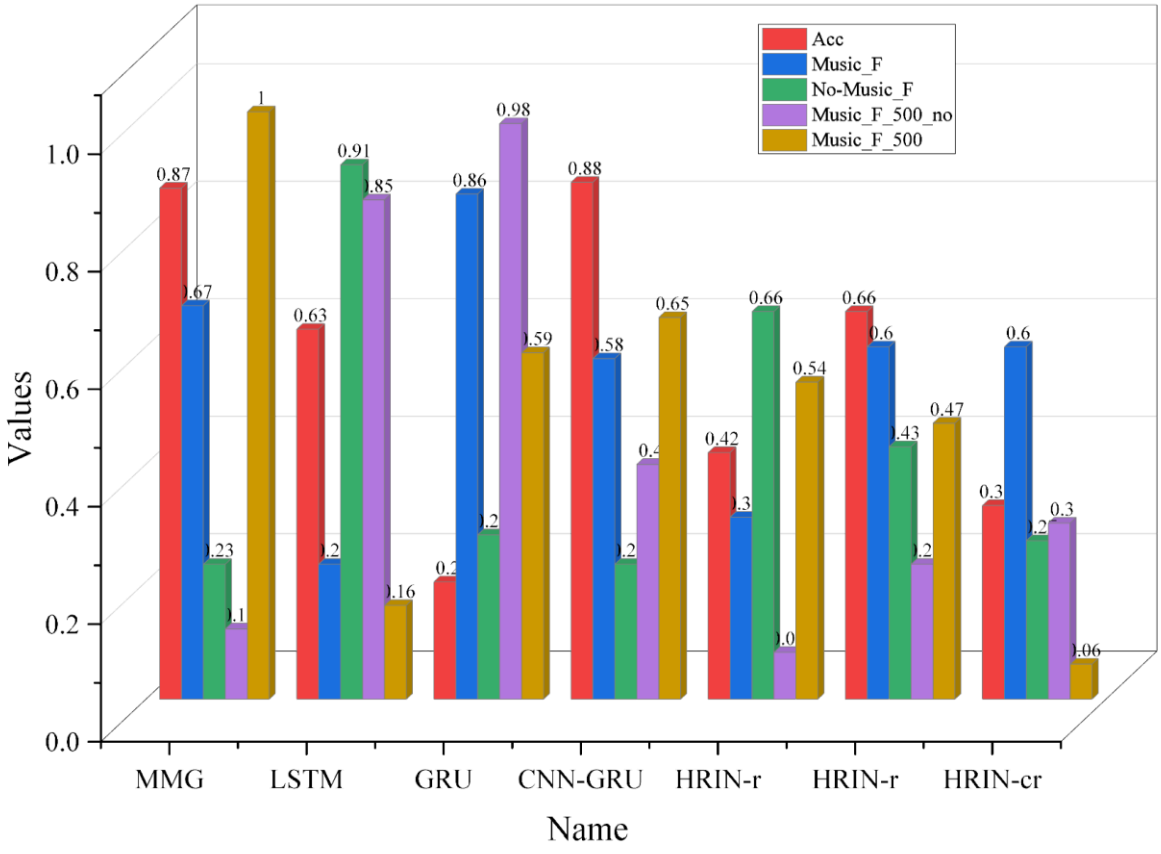

Figure 4: Experimental results of the baseline model on the music detection task

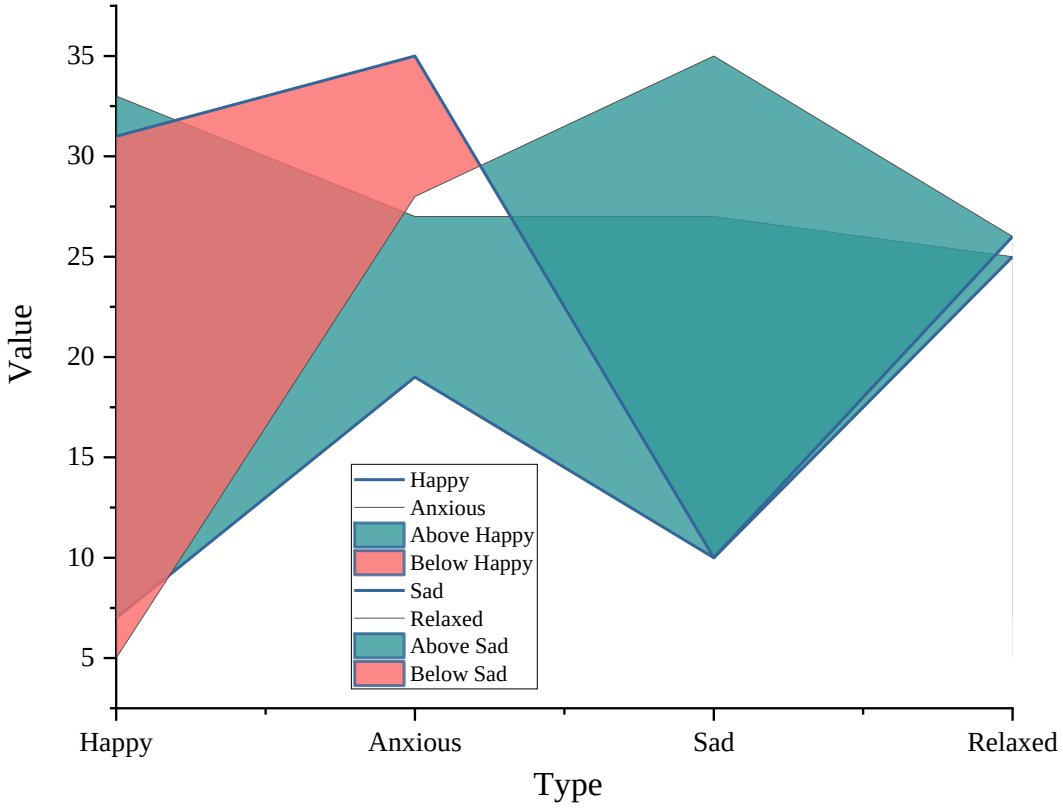

Figure 5: Graph of comparison of experimental results of music emotion recognition

From the figure, we can see that the Scratch music set generated using the Scratch music generation model has little difference in the performance of the Scratch music sentiment recognition model, with Precision, Recall, and F1-score reaching 71.0%, 70.8%, and 70.8%, respectively. However, they are lower compared to the Scratch music dataset. This is because the generated music is different from the original Scratch music dataset, and the Scratch music recognition model is trained using the constructed Scratch music dataset, which naturally performs better on the Scratch music dataset. However, the Scratch music generation model performs better than the other two models, and the Precision value of the data generated by the Scratch music generation model in the music emotion recognition task is much larger than the probability value of the randomness of the four classification problems, which proves the rationality of the design of the emotion-based GAN music generation model. The effectiveness of the Scratch music generation model based on sentiment design is illustrated.

Considering the characteristics of different modules, to verify the modeling ability of the model, this study combined the modules and launched experiments. First, this study conducted experiments on the weights of the cosine loss function. The quantifiers of the data set itself are the shopper's module. In the experimental process, it was found that when the weights were taken as the Ralphs, the quantifiers of the model were closer to the data set in terms of lime green than the other weights, which indicated that the note distribution in the bars was closer to the music data set. At the same time, the before-and-after similarity was also closer to the music data set, which meant that the music was closer to the music data set in terms of the before-and-after relationship, indicating that the overall before-and-after perception of the music was closer to the real music. This means that the music is closer to the music dataset in terms of the before-and-after relationship, indicating that the overall before-and-after perception of music is closer to the real music data.

As can be seen from the results, when the temporal structure generator is used, the percentage of pitch differences between bars of 8, 16, and 24 degrees or more is significantly lower than that of the model without the temporal structure generator and the other models. This is due to the presence of the temporal structure generator, which makes the generated measures arranged in a particular order, thus making the generated measures more coherent, while the music generated by the models that use the music measure generator alone or do not consider note coherence is more independent, and the measures are not related to each other, resulting in excessive pitch differences, as shown in Figure 6.

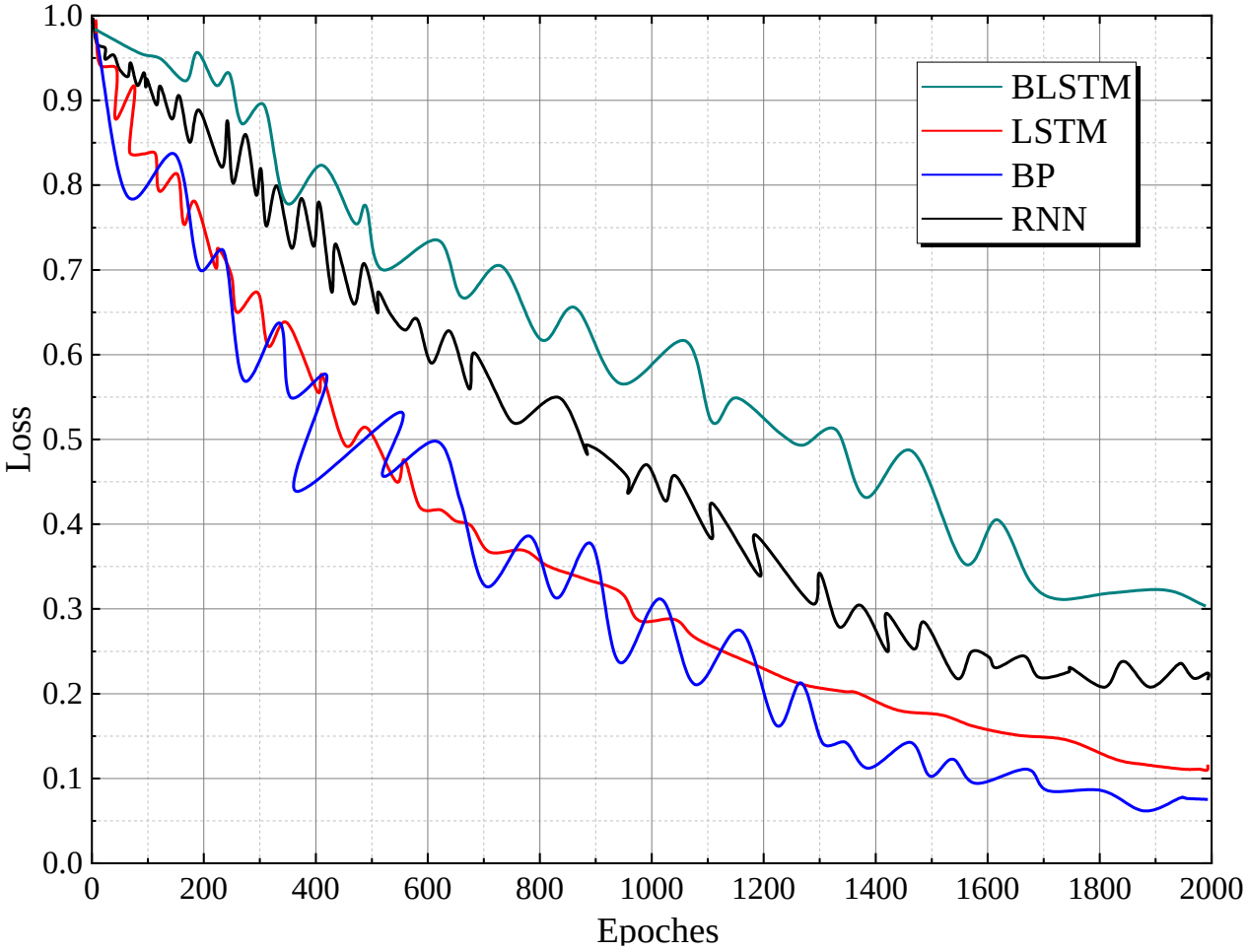

Figure 6: Loss of the four network models

The accuracy of the bidirectional LSTM model with single and three hidden layers is 73% and 90% at 1000 iterations, respectively. Compared with the accuracy, the bidirectional LSTM model with three hidden layers is more reasonable than the bidirectional LSTM model with single hidden layers. The accuracy of BP, RNN, LSTM, and bidirectional LSTM models with single hidden layers are 40%, 47%, 68%, and 73%, respectively. Three models are tested with no higher than 70% accuracy, and the single-layer bidirectional LSTM model has the highest accuracy.

As the choice of the number of hidden layers and the corresponding number of nodes directly affects the reliability and accuracy of the model. Further analysis is carried out regarding the degree of influence of different hidden layers and the number of nodes on the model. As a result of the analysis, a reasonable number of hidden layers and the corresponding number of nodes are given to make the model optimal, i.e., with minimum error. Before training, parameters such as mini-batch, learning rate, and optimizer are determined concerning previous experience and related knowledge. We can see that as the complexity of the model increases, the accuracy of the model also improves. The bidirectional LSTM model with three hidden layers has the highest accuracy, reaching 90%, while the bidirectional LSTM model with a single hidden layer has an accuracy of 73%. This indicates that increasing the number of hidden layers in the model can improve its performance. The confusion matrix can help us understand the prediction performance of the model on different categories. Through the confusion matrix, we can intuitively see the performance of the model on each category and which categories have the most accurate predictions. Accuracy is the proportion of samples predicted by the model to be true positive examples. Overfitting refers to the model performing well on training data but performing poorly on test data. This is usually due to the model being too complex, resulting in overfitting of the training data. To prevent overfitting, we can adopt some strategies, such as increasing the amount of data, using regularization, and reducing model complexity. Insufficient fitting refers to the poor performance of the model on both training and testing data. This is usually due to the model being too simple to capture the complex patterns of the data. To solve the problem of insufficient fitting, we can increase the complexity of the model or use a more complex model structure.

5.2 Music teaching application results

This process may involve the teacher's expertise, experience, knowledge, and understanding of the teaching environment and students. To assess the goals of deeper music instruction, teachers should first believe that all students can make appropriate progress, that students

understand the criteria for assessment, and that learning-oriented assessment activities will truly achieve deeper learning in music. Second, teachers should strive to design assessment criteria that are more operational and promote education in response to the necessarily demanding standards. Teachers design appropriate learning tasks and guide their completion in ways that, in turn, are extensions of classroom instruction rather than simple replications of curricular requirements, and learning goals and assessment tasks can maximize their pointing power if the process by which students complete the assessment is the same as the learning process. While the previous chapter's deep learning model for music reveals how deep learning develops in classroom instruction, a clear assessment of learning objectives for which guided assessment tasks is the initial point of the launch can save the time experience required to invest in the learning model.

Whether written or performance audition, teachers' elaborate design of evaluation situations should be objective and diverse, combining listening experience analysis and discussion in depth as grades. In addition to the innovation of the evaluation process, the deep combination of vocal teaching and instrumental teaching and pointing to the core literacy of music discipline should have complied with the curriculum in the evaluation session. Different musical language, artistic emotion, and artistic, social value should be examined, and the deep evaluation of experiential ability, analytical ability creative ability and cultural understanding should be conducted in connection with the characteristics of other disciplines, as shown in Figure 7.

Comparing the results of Experiment 1 and Experiment 2, it can be found that the addition of the pitch feature is not as significant as the rhythm for improving the effect. The reason for this is that the pitch feature already contains enough information about the structure of music; the pitch feature is the basic feature of music, in which the change of pitch represents the structure of music; the rhythm is also a critical auxiliary feature for music, which determines the properties of notes in time-based on the determination of the structure of music, and together with the pitch, it constitutes the structural feature of music. The pitch, on the other hand, represents the keynote of the music and represents only the overall sonic height of the music, not the structure of the music, and therefore has little effect.

The purpose of this part of the experiment is to verify the effect of the overall link, taking the MIREX05 dataset, which is an audio file and includes a variety of styles of music. This experiment first uses the method proposed in Chapter 3 to extract the main melody pitch and music rhythm and pitch, converting them to JSON format; after that, the method in Chapter 4 is used to calculate the similarity, and this model uses the BiLSTM with better effect plus the attention mechanism.

In this part of the experiment, to verify the similarity of the music structure, the MIREX05 dataset was used while the songs were split into several short and fixed-length music segments. Since the segments are short and have high similarity under the same song, this experiment slices each audio segment according to a length of 100. Therefore, in this part, the audio label in the dataset is set to the name of the song to which the clip belongs. In the experiment, the similarity Top1 and Top3 are used to measure the accuracy of the links. After inputting a sample, its similarity to other samples is calculated and ranked, where Top1 refers to the sample with the top accuracy rate, and Top3 refers to the sample with the top three accuracy rates. When the sample whose accuracy ranking meets the requirement has the same label as the calculated sample, the sample is considered to be correctly identified, as shown in Figure 8.

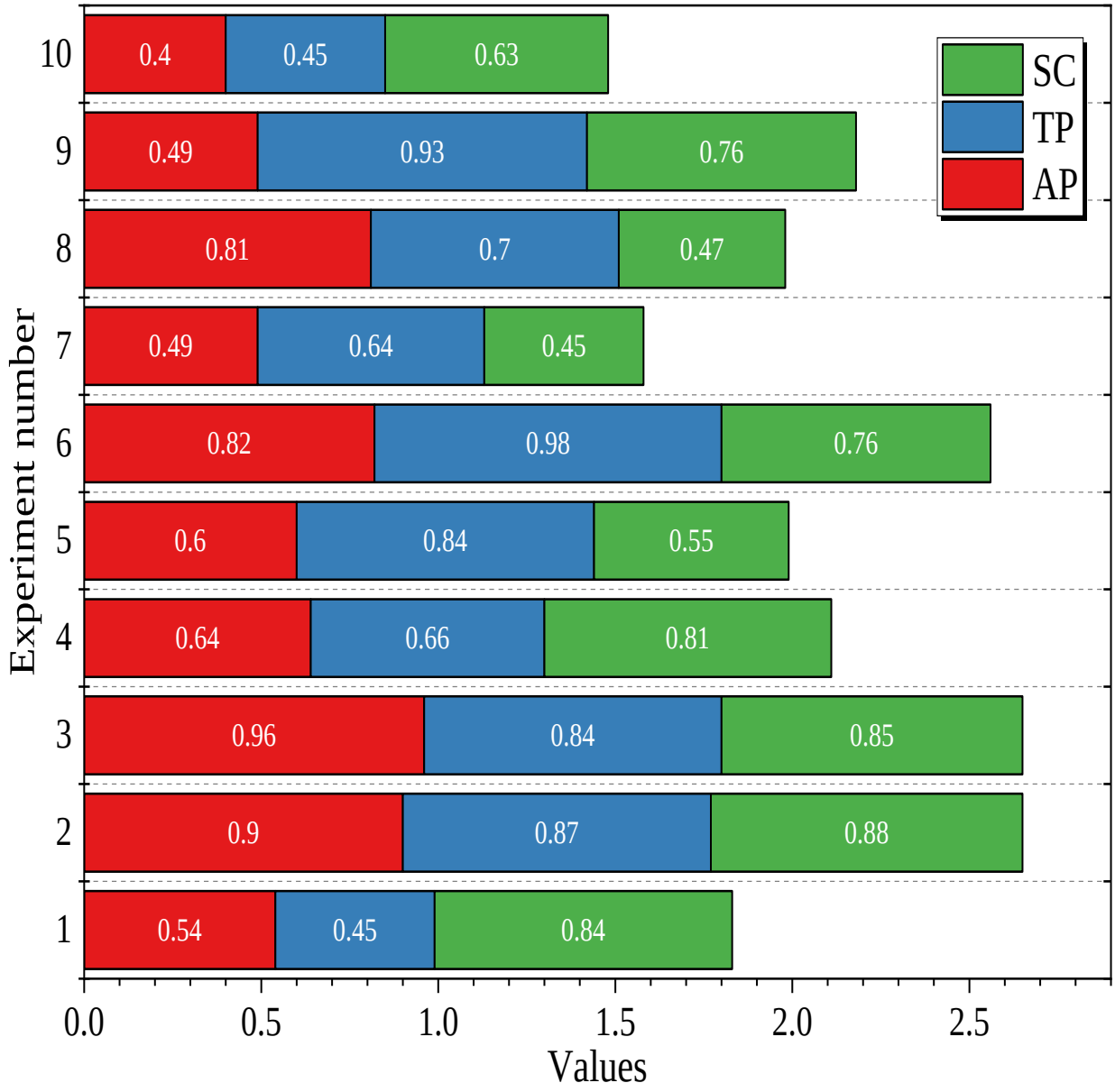

Figure 7: Experimental results of the model and similarity calculation formula

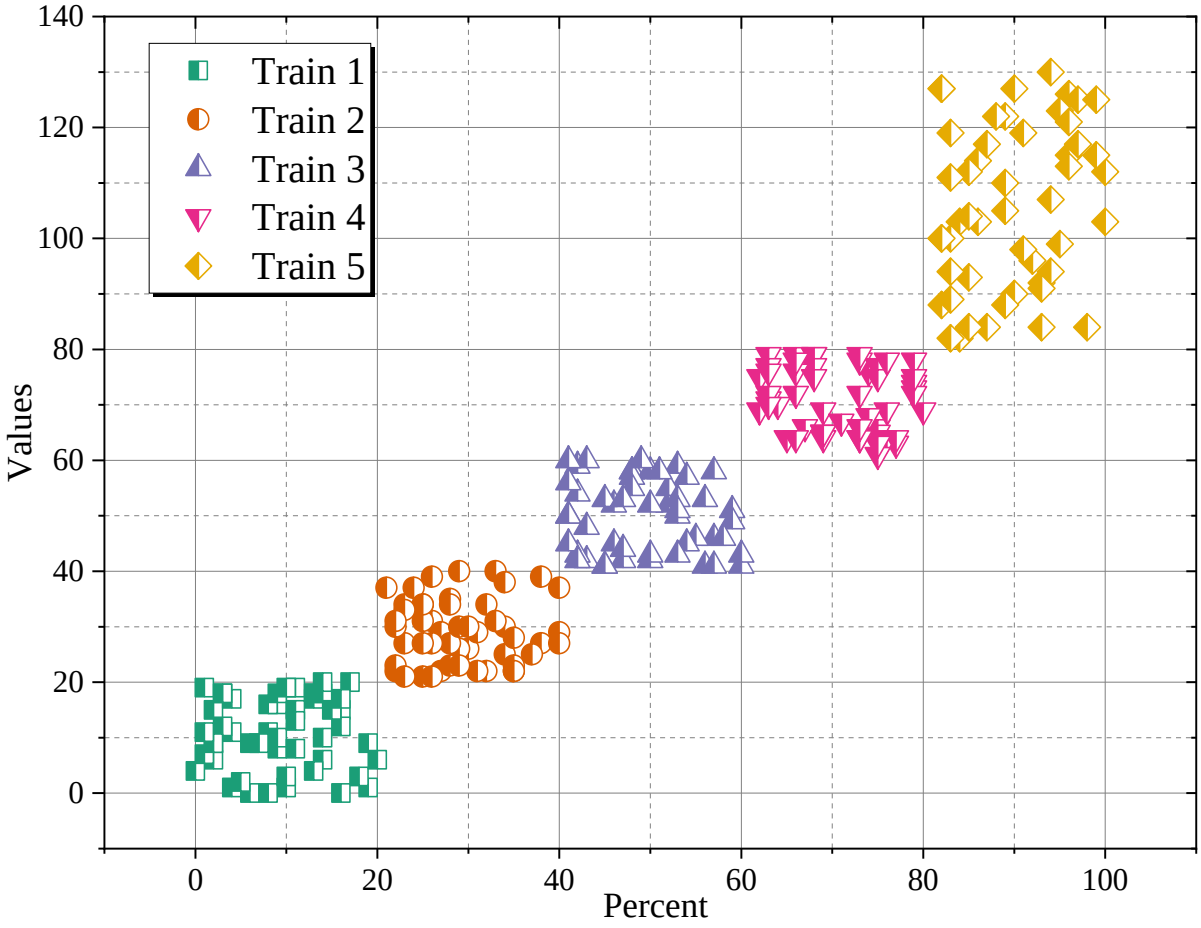

Figure 8: Experimental sample clustering

The experimental results show that the attention mechanism has a positive effect on improving the accuracy rate, while the accuracy rate performs better when the cosine similarity is used. For the contour coefficient, the attention mechanism has little effect. In contrast, the distance calculation formula has a greater effect, whereas the cosine similarity has the most apparent impact on the improvement of the contour coefficient. The second part verifies which musical features impact similarity detection more. The experiments show that rhythm has a greater impact on the experimental results when using pitch sequences as the base feature, and pitch impacts the results, but the effect is not as evident as the rhythm feature.

Music generation models can be used as part of teaching tools to assist music education. For example, students can use these models to generate their own music works, or create based on specific emotional or stylistic requirements. In addition, these models can also be used to analyze students' music works and provide feedback on their emotions, styles, and other aspects. The use of music generation models can enhance students' learning experience, enhance their interest and motivation in learning. By creating their own music works, students can better understand the composition and creative process of music, thereby deepening their understanding and appreciation of music. In addition, these models can also help students understand how music with different emotions and styles is encoded, thereby enhancing their music analysis and comprehension abilities. Although music generation models may be new technologies for some educators, many existing educational and research institutions are actively promoting and applying these technologies. In addition, with the development of technology, these models have become increasingly easy to use and understand. Educators can learn and master these technologies by participating in relevant training courses or seminars, in order to apply them to their teaching.

6 Conclusion

This paper is based on the actual teaching practice of middle school music classroom, through theoretical study and front-line teaching practice, and based on our own professional teaching experience and classroom teaching reflection, we focus the center of the paper on the exploration of large units of teaching in the music discipline. The teaching mode in line with deep learning can permeate the core literacy of music subjects into each teaching module, effectively addressing the implementation of students' musical ability and literacy and better promoting a deep understanding of music culture. This is also mainly related to the characteristic connotation of deep learning and its alignment with the values of core literacy. Whether from a temporal or a

content perspective, deep learning is extended and deepened. It is always conducive to developing students' abilities and literacies for lifelong learning in music and reaching the essence and core of the music discipline. The comparison experiment in this paper uses many data sets. Because of the specificity of the training data, experiments involving different input data can only change data sets. Therefore, this paper adds comparative experiments for similarity detection under the overall process. In this paper, the multi-stage and multi-level double helix form of the teaching model is chosen precisely because it conforms to the real contextual, practical activity and perceptual-experiential tendency path, from the acceptance stage to the participation stage to the migration stage and the specific practice process of the model, in which students and teachers take each other as subjects, and the bottom-up and real-time dynamic design of teaching objectives and teaching evaluation oriented to learning outcomes, and through the spiral of both development and interpretation from the bottom up through the music learning behavior.

Competing of interests

The authors declare no competing of interests.

Authorship contribution statement

Lin Jing: Writing-Original draft preparation, Conceptualization, Supervision, Project administration.

References

[1] X. Wang, S. Zhao, J. Liu, and L. Wang, "College music teaching and ideological and political education integration mode based on deep learning," *Journal of Intelligent Systems*, 31(1): 466–476, 2022. https://doi.org/10.1515/jisys-2022-0031

[2] G. Taranto-Vera, P. Galindo-Villardón, J. Merchán-Sánchez-Jara, J. Salazar-Pozo, A. Moreno-Salazar, and V. Salazar-Villalva, "Algorithms and software for data mining and machine learning: a critical comparative view from a systematic review of the literature," *J Supercomput*, 77: 11481–11513, 2021. https://doi.org/10.1007/s11227-021-03708-5

[3] J. Liu, S. Snodgrass, A. Khalifa, S. Risi, G. N. Yannakakis, and J. Togelius, "Deep learning for procedural content generation," *Neural Comput Appl*, 33(1): 19–37, 2021. https://doi.org/10.1007/s00521-020-05383-8

[4] M. AlQuraishi and P. K. Sorger, "Differentiable biology: using deep learning for biophysics-based and data-driven modeling of molecular mechanisms," *Nat Methods*, 18(10): 1169–1180, 2021. https://doi.org/10.1038/s41592-021-01283-4

[5] M. Abd Elaziz *et al.*, "Advanced metaheuristic optimization techniques in applications of deep neural networks: a review," *Neural Comput Appl*, 1–21, 2021. https://doi.org/10.1007/s00521-021-05960-5

[6] Q. Zhang, J. Lu, and Y. Jin, "Artificial intelligence in recommender systems," *Complex & Intelligent Systems*, 7: 439–457, 2021. https://doi.org/10.1007/s40747-020-00212-w

[7] A. Darwish, A. E. Hassanien, and S. Das, "A survey of swarm and evolutionary computing approaches for deep learning," *Artif Intell Rev*, 53: 1767–1812, 2020. https://doi.org/10.1007/s10462-019-09719-2

[8] J. Lu *et al.*, "Illustrating changes in time-series data with data video," *IEEE Comput Graph Appl*, 40(2): 18–31, 2020. https://doi.org/10.1109/MCG.2020.2968249

[9] L. Vrysis, N. Tsipas, I. Thoidis, and C. Dimoulas, "1D/2D deep CNNs vs. temporal feature integration for general audio classification," *Journal of the Audio Engineering Society*, 68(1/2): 66–77, 2020. https://doi.org/10.17743/jaes.2019.0058

[10] M. Abdel-Basset, H. Hawash, R. K. Chakrabortty, M. Ryan, M. Elhoseny, and H. Song, "ST-DeepHAR: Deep learning model for human activity recognition in IoHT applications," *IEEE Internet Things J*, 8(6): 4969–4979, 2020. https://doi.org/10.1109/JIOT.2020.3033430

[11] S. H. Lim, S. Kim, B. Shim, and J. W. Choi, "Deep learning-based beam tracking for millimeter-wave communications under mobility," *IEEE Transactions on Communications*, 69(11): 7458–7469, 2021. https://doi.org/10.1109/TCOMM.2021.3107526

[12] P. Gomathi, S. Baskar, P. M. Shakeel, and V. R. S. Dhulipala, "Identifying brain abnormalities from electroencephalogram using evolutionary gravitational neocognitron neural network," *Multimed Tools Appl*, 79: 10609–10628, 2020. https://doi.org/10.1007/s11042-022-13850-8

[13] Y. SATO, Y. HORAGUCHI, L. VANEL, and S. SHIOIRI, "Prediction of image preferences from spontaneous facial expressions," *Interdiscip Inf Sci*, 28(1): 45–53, 2022. https://doi.org/10.4036/iis.2022.A.02

[14] S. Bhaskaran and R. Marappan, "Analysis of collaborative, content & session based and multi-criteria recommendation systems," *The Educational Review, USA*, 6(8): 387–390, 2022. Doi: 10.26855/er.2022.08.009

[15] V. A. Vuyyuru, G. A. Rao, and Y. V. S. Murthy, "A novel weather prediction model using a hybrid mechanism based on MLP and VAE with fire-fly optimization algorithm," *Evol Intell*, 14: 1173–1185, 2021. https://doi.org/10.1007/s12065-021-00589-8

[16] I. Santos, L. Castro, N. Rodriguez-Fernandez, A. Torrente-Patino, and A. Carballal, "Artificial

neural networks and deep learning in the visual arts: A review," *Neural Comput Appl*, 33: 121–157, 2021. https://doi.org/10.1007/s00521-020-05565-4

[17] M. Littmann *et al.*, "Validity of machine learning in biology and medicine increased through collaborations across fields of expertise," *Nat Mach Intell*, 2(1): 18–24, 2020. https://doi.org/10.1038/s42256-019-0139-8

[18] E. Apostolidis, E. Adamantidou, A. I. Metsai, V. Mezaris, and I. Patras, "Video summarization using deep neural networks: A survey," *Proceedings of the IEEE*, 109(11): 1838–1863, 2021. https://doi.org/10.1109/JPROC.2021.3117472

[19] H. Ghanei, F. Manavi, and A. Hamzeh, "A novel method for malware detection based on hardware events using deep neural networks," *Journal of Computer Virology and Hacking Techniques*, 17(4): 319–331, 2021. https://doi.org/10.1007/s11416-021-00386-y

[20] J. Klinger, J. Mateos-Garcia, and K. Stathoulopoulos, "Deep learning, deep change? Mapping the evolution and geography of a general purpose technology," *Scientometrics*, 126: 5589–5621, 2021. https://doi.org/10.1007/s11192-021-03936-9

[21] I. A. Doush and A. Sawalha, "Automatic music composition using genetic algorithm and artificial neural networks," *Malaysian Journal of Computer Science*, 33(1): 35–51, 2020. https://doi.org/10.22452/mjcs.vol33no1.3

[22] L. Ma and B. Sun, "Machine learning and AI in marketing–Connecting computing power to human insights," *International Journal of Research in Marketing*, 37(3): 481–504, 2020. https://doi.org/10.1016/j.ijresmar.2020.04.005

[23] X. Wang, Y. Han, V. C. M. Leung, D. Niyato, X. Yan, and X. Chen, "Convergence of edge computing and deep learning: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, 22(2): 869–904, 2020. https://doi.org/10.1109/COMST.2020.2970550

[24] C. Gresse von Wangenheim, J. C. R. Hauck, F. S. Pacheco, and M. F. Bertonceli Bueno, "Visual tools for teaching machine learning in K-12: A ten-year systematic mapping," *Educ Inf Technol (Dordr)*, 26(5), 5733–5778, 2021. https://doi.org/10.1007/s10639-021-10570-8