Retraction

Retracted: Practical Exploration on the Construction of Theoretical Courses of Composition Technology in the Age of Artificial Intelligence

Mobile Information Systems

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] L. Luo, "Practical Exploration on the Construction of Theoretical Courses of Composition Technology in the Age of Artificial Intelligence," *Mobile Information Systems*, vol. 2022, Article ID 3099312, 14 pages, 2022.

Research Article

Practical Exploration on the Construction of Theoretical Courses of Composition Technology in the Age of Artificial Intelligence

Luo Luo ⓘ

Music Department, Xi'an Shiyou University, Xi'an 710065, Shaanxi, China

Correspondence should be addressed to Luo Luo; luoluo@xsyu.edu.cn

The theory of composition technology is to study the basic knowledge and skills of musicology, instrumental music, composition, and composition technology, including harmony, polyphony, musical form, orchestration, and so on, and then to analyze, create, and edit music. The current theory of composition technology has resulted in the phenomenon of relatively single form. The current method is the traditional way of composing music through the creation of composers. The defect is that various elements cannot be integrated together, and the meaning of music cannot be perfectly presented. In order to solve these problems, this paper proposes the use of recurrent neural network algorithm and backpropagation algorithm in artificial intelligence algorithm. It aims to study how to innovatively integrate the composition technology theory course with the current network technology. And it utilizes recurrent neural networks in artificial intelligence to help design part of the analysis of musical characteristics, through the evaluation of the music effect generated by automatic composition. The results show that the accuracy of note prediction obtained by the automatic composition method on the basis of objective evaluation is 81.93%, 90.15%, and 92.62%, respectively, on Top1, Top2, and Top3, which basically meet the current basic requirements for composition technology theory.

1. Introduction

Music is an ancient and time-honored sound art that has always played an important role in human development. From the chimes of the slave era in China to the various styles of music works in the current era, all of them reflect its long history. Therefore, the relationship between music and human beings is inseparable. In recent years, the development of music has a trend towards popularization, and the professionalism of human work composition is too strong to meet people's growing demand for music appreciation. At this time, a composition method that can generate music efficiently and quickly is needed, so the automatic composition method emerges as the times require. Automatic composition methods can generally be divided into two key parts: the representation of musical features and the design of the composition model.

This paper draws on the word vector model for semantic analysis of natural language processing and proposes a method for generating note feature vectors based on contextual semantic coding. It also solves the problem that the traditional automatic composition uses the form of one-hot encoding to represent the musical features, which leads to the inability to describe the contextual semantic information of the musical features. Through comparative experiments, it is shown that using the note feature vector proposed in this paper to represent musical features for automatic composition work can effectively improve the accuracy of note prediction. This paper proposes an automatic composition model based on Bi-GRU network and self-attention mechanism, which solves the problem that the traditional automatic composition model is difficult to teach the timing information of music and flexible dependency information due to the limitation of network structure. In the composition model of this paper, the Bi-GRU network can learn not only the forward-dependent information of music, but also the backward-dependent information of music due to

its bidirectionality, which can better describe the timing information of music. The self-attention mechanism flexibly adjusts the dependencies between musical features by assigning different weights and highlights key information in the process of note prediction. It makes up for the limitation of the recurrent neural network in the exploration of dependent information, which makes the connection between the notes become weaker and weaker due to the short-term memory, and optimizes the automatic composition model.

At present, in the automatic composition, the recognized technologies are mainly genetic algorithm, artificial neural network, deep learning, and reinforcement learning. Basically, these technologies have been rapidly and widely used in the field of automatic composition with the development of computer technology. In the design process of the automatic composition model, an automatic composition model based on deep learning is proposed. Through the in-depth study of recurrent neural network and self-attention mechanism, this paper adopts Bi-GRU network and self-attention mechanism as an automatic composition model according to the characteristics of the input MIDI note sequence. Because of its bidirectional structure, the Bi-GRU network can capture not only the forward dependency of the note sequence but also the backward dependency of the note sequence and is better at dealing with the prediction problem of longer sequences. The self-attention mechanism flexibly expresses the connection between the notes by controlling the attention weight between the notes and highlights the important information between the notes, so that the final result is a better expression of the note sequence prediction.

2. Related Work

The theory of composition and composition technology mainly studies the basic knowledge and skills of musicology, instrumental music, composition, and composition technology, including harmony, polyphony, musical form, and orchestration, so as to conduct music analysis, creation, editing, and so on. And the following are the different expressions of composition technology in various fields. Tran et al. [1] provided a new perspective to understand the value outcomes and innovation process of the "fuzzy front-end" (FFE) stage of product innovation by introducing the theory of music composition [1]. Through the research on the characteristics and connotation of folk music, Zhang [2] sorted out how to strengthen the learning and promotion of folk music theory in school music education. On the basis of the combination of music education and national music, a more practical and perfect music teaching method and system had been constructed [2]. Zhaleko [3] believed that composers wanted to impress audiences and achieved commercial success [3]. Outsomichalis [4] proposed that contemporary currents of thought have largely denounced mass theory and a series of past dichotomies. Instead, there was a tendency towards mixed, all-encompassing, and nonanthropocentric models [4]. Moseley [5] extended an analysis of its circular practice in relation to a broader organic view of nature and found in the loop composition a match of the often incomprehensible but omnipotent laws imagined in the world around it [5]. However, the above researches only stay in the theoretical part, and the practicality is not too strong.

Artificial intelligence is abbreviated as AI. It is a new technical science that studies and develops theories, methods, technologies, and application systems for simulating, extending, and expanding human intelligence. The following is the application of artificial intelligence in various fields. Dolezel [6] introduced a specific neural network-based decision procedure that can be considered for any network traffic processing controller based on traffic characteristics [6]. Musulin et al. [7] discussed the applicability of machine learning (ML) and evolutionary computation (EC) methods focused on regressing the epidemiological curve of COVID-19 and outlined the usefulness of existing models in specific domains [7]. Corradini et al. [8] discussed the prospects and advantages and highlighted the pitfalls and caveats, which were outlined in several recent studies attempting to use AI methods to classify clinically meaningful PCa and indolent lesions [8]. Dukhi et al. [9] suggested that the use of machine learning in this area of health research was still novel. By using machine learning methods, weighted assessments of cross-linkages between anemia risk indicators would help policymakers identify priority areas for intervention in participating countries [9]. It was already evident that Cross [10] created unique images through the rapid development and improvement of artificial intelligence and deepfakes. And they argued that the adoption of these new technologies required a reconsideration of current preventive messaging, as the utility of reverse image search may become somewhat redundant in the future [10]. However, the current artificial intelligence research on composition technology theory still did not get rid of the definition and thinking based on traditional composition theory courses. The field of artificial intelligence was currently limited by technology and cannot exert its own characteristics and lacked in-depth analysis and exploration of the functionality of artificial intelligence.

3. Artificial Intelligence Neural Network Algorithm

3.1. Deep Neural Networks. The structure of shallow neural network is simple, the network parameters are few, and it is easy to fall into the local optimal situation, so it is difficult to obtain useful information from massive datasets [11]. The hidden layer of the actual deep neural network is always more than 3 layers, and the number of hidden layer nodes in each layer is usually relatively large. The input layer inputs the feature values of the corresponding signal, such as speech features and image features. There are nonlinear activation functions at the nodes of the hidden layer, commonly used hyperbolic tangent, sigmoid, and ReLU functions. The setting of the output layer function is related to the purpose of the network. The regression task generally selects the linear function as the output function, and the multi-classification task generally selects the softmax function as the output function. For an L-layer deep neural network, in order to simplify the notation, the framework formula is

$$y^l = f(z^l)$$
$$= f(W^l y^{l-1} + b^l). \tag{1}$$

Among them,

$$z^l = W^l y^{l-1} + b^l, \tag{2}$$

where $y\prime, w\prime, b$ are the excitation vector, activation vector, weight matrix, and deviation coefficient matrix, respectively. The schematic block diagram of the deep neural network is shown in Figure 1.

3.2. Network Training Based on Supervised Learning. The output loss function will be calculated according to the objective function, and then the parameters will be updated through the backpropagation algorithm proposed in 2013 [12]. For speech enhancement tasks, the minimum mean square error is generally used as the loss function, which is defined as

$$E = \frac{1}{N} \sum_{n=1}^{N} \left\| \widehat{y}_n - y_n \right\|_2^2 = \frac{1}{N} \sum \left\| (x, W, b) - y_n \right\|_2^2. \tag{3}$$

Among them, y_n and $\widehat{y}_n$ represent the target feature and output feature of the network, respectively, and W and b are the weight parameters of the network. x_n is the input feature of the network, and n is the speech frame. N represents the parameter of minibatch training, which represents how many frames of data are processed at the same time in the same batch. The addition of minibatch will allow the training set to extract N sample data each time and use the stochastic gradient descent method to update the parameter matrix of the deep neural network, thereby improving the training efficiency. For a deep neural network with L layers, by using y^L to represent the output, the formula can be written as

$$E = \frac{1}{N} \sum_{n=1}^{N} \left(y_n^L - y_n \right) \left(y_n^L - y_n \right). \tag{4}$$

The weight parameters W and b of the network are updated with backpropagation:

$$W_{t+1}^l \leftarrow W_t^l \leftarrow a\Delta W_t^l,$$
$$b_{t+1}^l \leftarrow b_t^l \leftarrow a\Delta b_t^l. \tag{5}$$

Among them, W_T^L and b_t^l are the network parameters after the tth minibatch update, and a is the learning rate of the network, that is, the step size of each update. It is assumed that the output of the deep neural network adopts a linear activation function; the weight matrix gradient of the output layer is

$$\frac{\partial E}{\partial W_t^L} = \frac{\partial E}{\partial z_t^L} * \frac{\partial z_t^L}{\partial W_t^L}$$
$$= e_t^L * \frac{\partial (W_t^L y_t^{L-1} + b_t^L)}{\partial W_t^L} \tag{6}$$
$$= 2(y_t^l - y)(y_t^{L-1})^T.$$

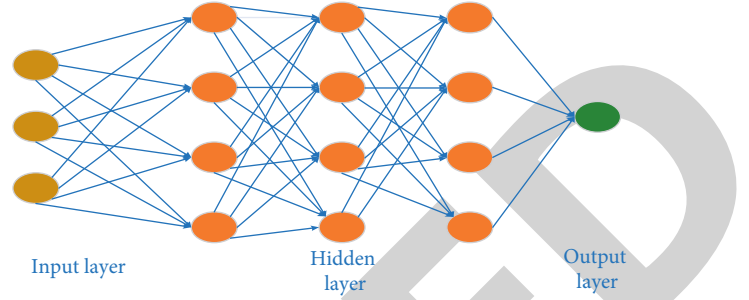Figure 1: Schematic diagram of deep neural network.

Among them,

$$z^l = W^l y^{l-1} + b^l. \tag{7}$$

Then, the vector gradients of the error and bias of the output layer are

$$e_t^l = \frac{\partial E}{\partial z_t^L}$$
$$= \frac{\partial \left((y^L - y)^T (y^L - y) \right)}{\partial z_t^L}$$
$$= 2(y_t^L - y),$$
$$e_t^L = \frac{\partial E}{\partial z_t^L}$$
$$= 2(y_t^L - y). \tag{8}$$

Similarly, for the middle hidden layer, it can be deduced that

$$\frac{\partial E}{\partial W_t^l} = \frac{\partial E}{\partial y_t^l} * \frac{\partial y_t^l}{\partial W_t^l}$$
$$= e_t^l \frac{\partial (W_t^l y_t^{l-1} + b_t^l)}{\partial W_t^l}$$
$$= \left[e_t^l * f(z_t^l) \right] (y_t^{l-1})^T,$$
$$\frac{\partial E}{\partial b_t^l} = \frac{\partial E}{\partial y_t^l} * \frac{\partial y_t^l}{\partial b_t^l} \tag{9}$$
$$= e_t^l \frac{\partial (W_t^l y_t^{l-1} + b_t^l)}{\partial b_t^l}$$
$$= e_t^l * f'(z_t^l),$$

where e_t^l is the error of the lth layer and $f'(z_t^l)$ is the first derivative of the activation function. If the sigmoid function is used as the activation function of the hidden layer of the network, the corresponding first derivative is

$$\sigma'\left(z_t^l\right) = \left(1 - \sigma\left(z_t^l\right)\right) * \sigma\left(z_t^l\right)$$
$$= \left(1 - y_t^l\right)y_t^l. \tag{10}$$

The backpropagation formula of the top layer and the propagation formula of the hidden layer are

$$e_t^{L-1} = \frac{\partial E}{\partial y_t^{L-1}}$$

$$= \frac{\partial E}{\partial z_t^L} * \frac{\partial z_t^L}{\partial y_t^{L-1}}$$

$$= e_t^L \frac{\partial\left(W_t^L y_t^{L-1} + b_t^L\right)}{\partial y_t^{L-1}}$$

$$= e_t^L\left(W_t^L\right)^T,$$

$$e_t^{l-1} = \frac{\partial E}{\partial y_t^{l-1}} \tag{11}$$

$$= \frac{\partial E}{\partial y_t^l} * \frac{\partial y_t^l}{\partial y_t^{l-1}}$$

$$= e\frac{\partial\left(W_t^l y_t^{l-1} + b_t^l\right)}{\partial y_t^{l-1}}$$

$$= \left[e_t^l * f'\left(z_t^l\right)\right]\left(W_t^l\right).$$

It plays a decisive role in the automatic composition model of the automatic composition technology, and the use of network training can help the composition process of the automatic composition model to play a role, laying the basic framework structure of the automatic composition technology.

3.3. Recurrent Neural Network Framework.

The framework of a recurrent neural network is a class that takes sequence data as input. Recurrent neural networks recurse in the evolutionary direction of the sequence, and all nodes are connected into a chain.

Similar to the multilayer feedforward neural network, the standard RNN consists of an input layer, a hidden layer, and an output layer. The network is input from the input layer, and after passing through the hidden layer, the output layer is output. Unlike multilayer feedforward neural networks, RNN allows adding connections within layers in addition to interlayer connections. The intralayer connection makes the input of the current moment contain not only the input of the network itself, but also the output of the previous moment. Therefore, a multilayer feedforward neural network can only map from an input to an output vector, while an RNN can in principle map from the entire history of previous inputs to each output. Its typical structure is shown in Figure 2.

By unfolding the RNN according to the timeline, it can be seen from Figure 2 that each moment of the RNN contains the information of several previous moments, and the links within the layer allow the RNN to accumulate in the time domain. Chained features reveal that RNNs are inherently related to sequences and lists and are more suitable for processing machine learning tasks related to sequence data [13].

RNN forward propagation is the same as a multilayer feedforward neural network with a single hidden layer. If the input X of sequence length is T, then the input layer contains I input units. For an RNN with H hidden units and K output units, the forward propagation is represented by

$$a_k^t = \sum_{h=1}^{H} w_h k b_h^t,$$

$$a_h^t = \sum_{i=1}^{l} w_{ih} x_i^t + \sum_{h=1}^{H} w_{h'h} b_h^{t-1}. \tag{12}$$

Backpropagation uses the BPTT algorithm to pass back the accumulated residual from the last moment, which is expressed by the following formula:

$$\delta_h^t = \theta\left(a_h^t\right)\left(\sum_{k=1}^{K} \delta_k^t w_{hk} + \sum_{h=1}^{H} \delta_h^{t+1} w_{hh'}\right),$$

$$\delta_j^t = \frac{\partial L}{\partial a_j^t},$$

$$\frac{\partial L}{\partial w_{ij}} = \sum_{t=1}^{T} \frac{\partial L}{\partial a_j^t} * \frac{\partial a_j^t}{\partial w_{ij}} \tag{13}$$

$$= \sum_{t}^{T} \delta_j^t b_i^t.$$

In the above formulas, a is the value calculated by the summation, b is the value calculated by the activation function, θ is the activation function, t is the time, w is the weight of the connection between different nodes, L is the loss function, a is the gradient, the subscripts h and k represent the hidden layer and the output layer, respectively, and i and j represent the serial numbers of the two nodes connected between the layers [14].

In RNN, the output of the current moment is obtained by the output of the previous moment and the input of the current moment. Therefore, RNN has a memory function for historical information and a strong ability to model time series data. However, because the chain rule is used to update the weights during backpropagation, when the time series is too long, it will lead to the problem of gradient explosion or echelon disappearance. So the researchers made improvements based on standard RNN to control the length of memory history information.

While the standard RNN can control the length of memory history information, the automatic composition model part used in the automatic composition scheme can help it expand the capacity of the dataset, so that the breadth of automatic composition technology can be improved.
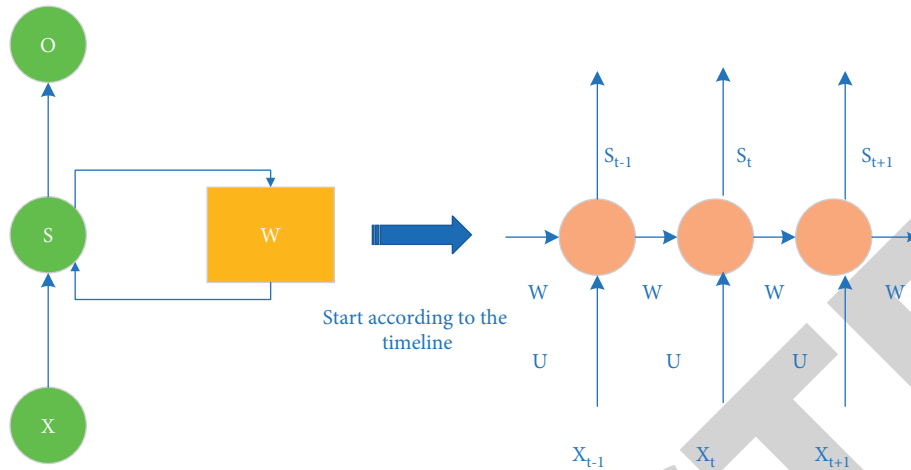
Figure 2: Typical RNN structure.

4. Exploration of Composition Technology Theory Courses

4.1. Musical Features Resolve. The representation of musical features is a key part of automatic composition, and the quality of musical feature representation directly determines the effect of subsequent automatic composition. In the previous researches on automatic composition, researchers mainly focused on choosing which composition model to improve the effect of composition and paid little attention to adopting a better representation of musical features that could also affect the effect of composition.

Previously, researchers viewed automatic composition as a sequence prediction problem. Therefore, they simply and directly use one-hot encoding to represent the sequence of notes extracted from the music to train the generated music [15]. In order to solve this problem, this paper uses the word vector model in the field of natural language processing to optimize the representation of music features. The data input into the composition model is no longer a note sequence represented by one-hot encoding, but a note sequence represented by a note feature vector with contextual semantic information generated after neural network pre-training. In this way, since the note feature vector itself has contextual semantic information, the accuracy of note prediction can be improved. The brief flowchart of note feature vector construction is shown in Figure 3.

The original MIDI music files are complex, and most of them have multiple tracks and a lot of additional music information. In the process of automatic composition, all the music information of multiple tracks is not needed. In this paper, only the note information on the main melody track is used as the dataset source for the composition model. Therefore, first of all, this article uses Python's MIDI library Mido to extract the required information from the original complex MIDI music files, including the pitch of the note, the start time and end time of the note sound, the strength, and the playing instrument to form a note information matrix. Then, based on the information of the note information matrix, the outline algorithm is used to extract the

main melody information and generate a simple single-track MIDI main melody music file. Finally, based on the MIDI main melody music file and using the word vector model to train to generate note feature vectors with context information, the number of note feature vectors obtained is equal to the note types in the dataset; that is, each note corresponds to a unique note feature vector with contextual information [16].

After obtaining the note feature vector, it is also necessary to use the note feature vector with context information to represent the note sequence. At this point, the main melody note sequence represented by the note feature vector can be used as the dataset for the subsequent automatic composition model to train and generate music.

4.2. Note Feature Vector Generation Method Based on Contextual Semantic Coding

4.2.1. Network Structure Design. This paper adopts the word-hopping algorithm, and its network structure adopts a feedforward neural network structure including input layer, projection layer, softmax layer, and output layer, as shown in Figure 4.

The input of the input layer is the one-hot encoding representation of the central note in the binary data group, so the dimension V of the input layer is 129, which means that the input notes are represented by the one-hot encoding of 129 dimensions. The dimension D of the note feature vector generated in this paper is 129, so the dimension of the output value of the projection layer is 129. It is assumed that the actual note type in the binary note dataset is K; a $K * 129$-dimensional note feature matrix will eventually be obtained, which means that the note feature vector of K notes is a multidimensional vector because of the output of the projection layer. Therefore, it is necessary to add a softmax function layer as the activation function after the projection layer and select the position with the highest probability in the multidimensional vector to output as the pitch number of the actual predicted note for subsequent network loss calculation [12].
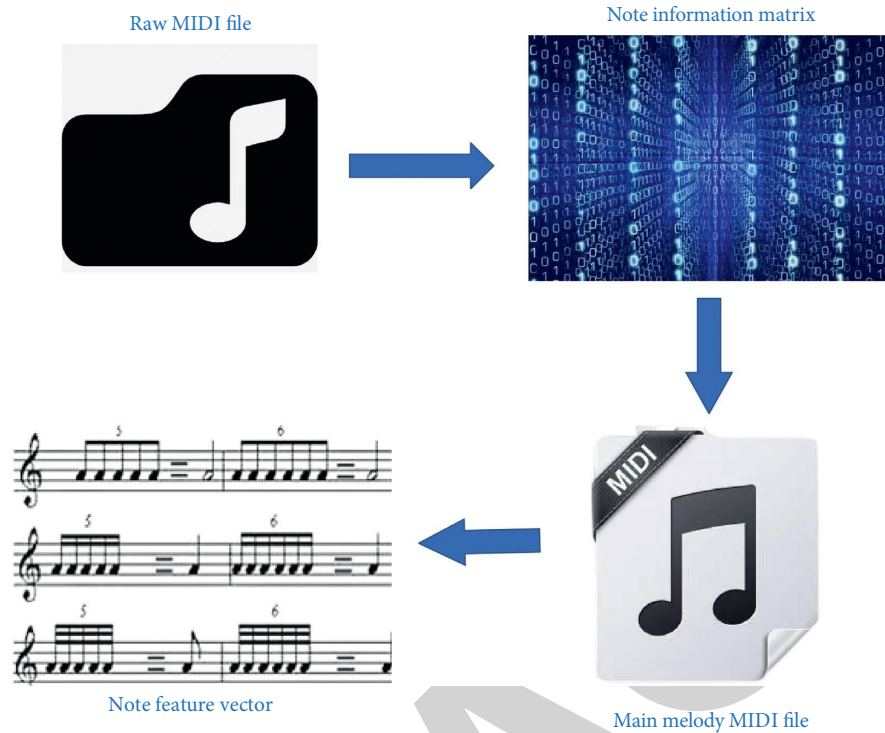
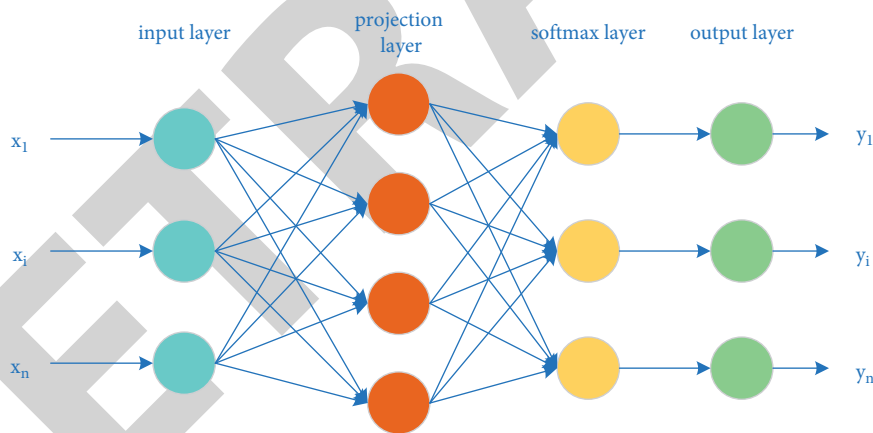FIGURE 3: The specific process of note vector construction.

FIGURE 4: Schematic diagram of the note feature vector generation network.

4.2.2. Generation of Note Feature Vector and Representation of Note Sequence. The generation of note feature vectors is the training process of the model. Continuous iterations are performed through the model's loss function to calculate the error between the actual output value and the target note in the binary note dataset. The loss function is optimized by batch gradient descent to make the model converge. Finally, the note feature vector with context information is obtained from the projection layer parameters. Each note generates a 129-dimensional note feature vector with contextual information.

4.3. Automatic Composition Model Based on Bi-GRU Network and Self-Attention Mechanism. In the above section, we investigated recurrent neural networks and self-attention

mechanisms and found that they can work well on sequence prediction problems. Therefore, this section proposes an automatic composition model based on Bi-GRU network and self-attention mechanism.

4.3.1. Design Thinking. Automatic composition is essentially a note sequence prediction problem, and recurrent neural networks are good at solving this type of problem. The automatic composition model in this paper selects a variant of the recurrent neural network, the gated recurrent unit network. In addition, music is an expressive art with a certain melody and rhythm that fluctuates with time. Usually, it is necessary to perceive the timing information of music and associate it with the contextual semantics of music in order to accurately analyze musical works.

Therefore, this paper chooses the structure of bidirectional recurrent network to deeply describe the contextual timing information of music. When generating the current note, not only the influence of the note information at its historical moment but also the influence of the note information at the future moment will be considered. Based on the above analysis of the network structure, only the bidirectional gated recurrent unit network can meet the requirements.

In the process of music generation, notes are affected by context information to different degrees, and this cannot be flexibly realized only by the structure of recurrent neural network. Therefore, the self-attention mechanism is introduced to focus on the important information in the context and ignore the secondary information, which makes the note generation more reasonable and accurate [17]. By adding a self-attention mechanism, the internal dependency features of the note sequence can be deeply explored by automatically assigning different probability weights to the note features in the context. It can also better express note sequences and improve the accuracy of note sequence prediction generation and the rationality of music generation.

4.3.2. Model Structure.
The entire automatic composition model structure based on Bi-GRU network and self-attention mechanism is shown in Figure 5.

Figure 5 is the structure diagram of the automatic composition model. It can be seen from Figure 5 that the entire composition model can be divided into three parts: the input layer, the hidden layer, and the output layer from the large module. The input and output mode selects the synchronous many-to-many mode; that is, the input note sequence and the output note sequence are more than one note and are of equal length.

The input data of the input layer is the main melody note sequence represented by the note feature vector with context information formed in the third chapter of this paper. It mainly includes two parts: one is the acquisition of the MIDI main melody note sequence, and the other is the acquisition of the note feature vector with context information. This part of the work has been described in detail in the third chapter of this paper. The final main melody note sequence is represented by this note feature vector with contextual information, which optimizes the representation of musical features in automatic composition, thereby improving the accuracy of note prediction and enhancing the effect of automatic composition [18].

The hidden layer is mainly divided into the implementation of the Bi-GRU network and the self-attention mechanism, in which a neural network training technique Dropout is added. Dropout is a method to increase the generalization ability of a network. It temporarily deactivates a part of the units in the middle layer according to a set probability and avoids overfitting by setting the output value of the unit to O to make it not work. The model in this paper adds a Dropout layer to the Bi-GRU network and the self-attention mechanism. Bi-GRU stores the timing dependency information of the input sequence through its internal bidirectional structure, namely, $H = h$, and learns from the
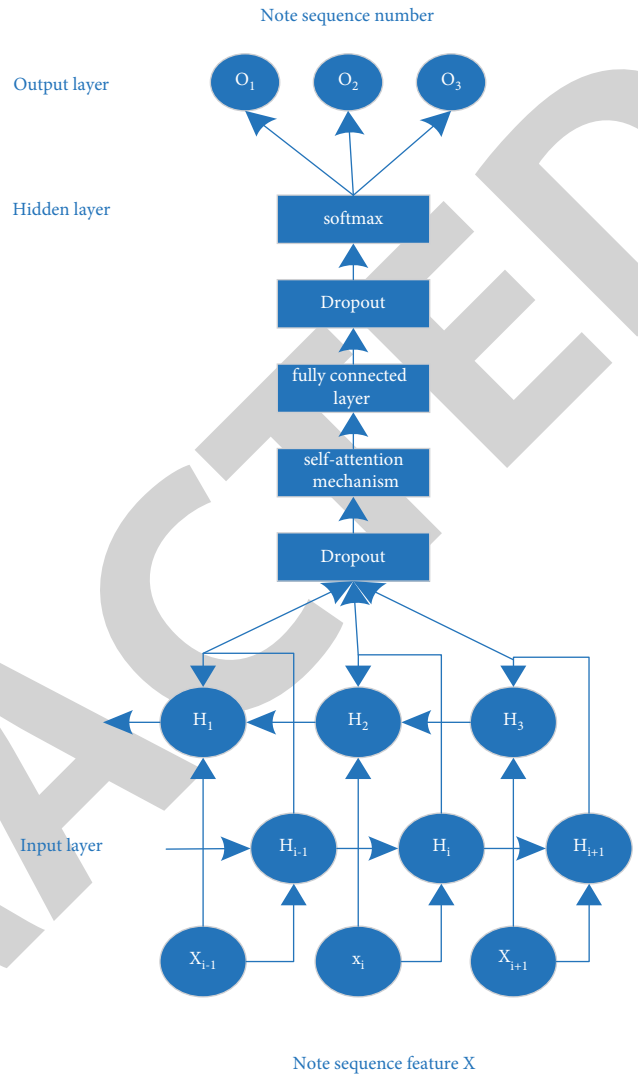

Figure 5: Schematic diagram of the automatic composition model.

future and historical directions to calculate the impact on the current state. Then, the output of the Bi-GRU network is used as the input of the self-attention mechanism layer, and the self-attention calculation is performed to obtain the corresponding attention weight probability distribution, so as to obtain the representation of the note sequence vector after adding the self-attention mechanism. Finally, the fully connected layer is input to the output layer for output.

4.3.3. Composition Process.
After continuous training and learning, the automatic composition model eventually gets a good note prediction model that converges. By using this model, a fixed-length sequence of notes can be generated, that is, the composition process.

The note prediction model also needs to input a segment note sequence with a fixed length of n, and the first input note sequence is randomly selected from the test set. And the next fixed-length n-segment note prediction sequence predicted by the note prediction model is output. Then, the outputted sequence of musical note predictions is used as

input to perform the next prediction again, and so on repeatedly until the preset length of the generated musical segment is generated, forming an automatic composition work. The above is the whole process of the automatic composition model to generate a piece of music [19]. In the test set, each time a segment note sequence with a fixed length of n is selected as input, a piece of music will be generated through the note prediction model, so that multiple pieces of music can be generated. The entire composition process is shown in Figure 6.

5. Resolve of the Results of the Automatic Composition Model

5.1. Model Evaluation. In order to better evaluate the music effect generated by automatic composition, this paper adopts objective and subjective evaluation methods. Both evaluation methods are based on comprehensive data analysis of the circulatory nervous system in artificial intelligence and selection of judges. The effect of automatic composition is evaluated from the perspective of data through objective experimental model conclusions, and the subjective ability of human beings to appreciate music is also considered through subjective online audition and evaluation. The two methods jointly ensure the accuracy and vividness of the music generated by the automatic composition in this paper [20].

5.1.1. Experimental Data. In order to achieve better experimental results, this paper obtained a total of 525 MIDI music files of the same style from the Internet. After the main melody extraction, 510 qualified MIDI main melody music files were obtained. All MIDI music files are in 4/4 time, the tempo is 120 beats per minute, and a total of 33 kinds of notes are found. The experiments in this chapter need to be completed in the server due to the large amount of data. Table 1 shows the specific configuration of the server.

5.1.2. Automatic Composition Experiment. The automatic composition of this paper adopts the method of comparative experiment. The difference between the comparative experiments is mainly in the representation of musical features and the choice of composition model. A mainstream automatic composition scheme is given in this paper—firstly, one-hot encoding is used as the musical feature to represent the main melody note sequence, and a separate Bi-GRU network is selected as the composition model. This is Experiment 1, which is used as a benchmark experiment for comparison. Secondly, the note feature vector of this paper is used as the musical feature to represent the main melody note sequence, and a separate Bi-GRU network is selected as the composition model, which is Experiment 2. Finally, the note feature vector of this paper is used as the musical feature to represent the main melody note sequence, and the Bi-GRU network and the self-attention mechanism are selected as the composition model. This is Experiment 3, which is the automatic composition scheme proposed in this paper. Through the comparison of the results of the above three groups of experiments, the advantages of the automatic
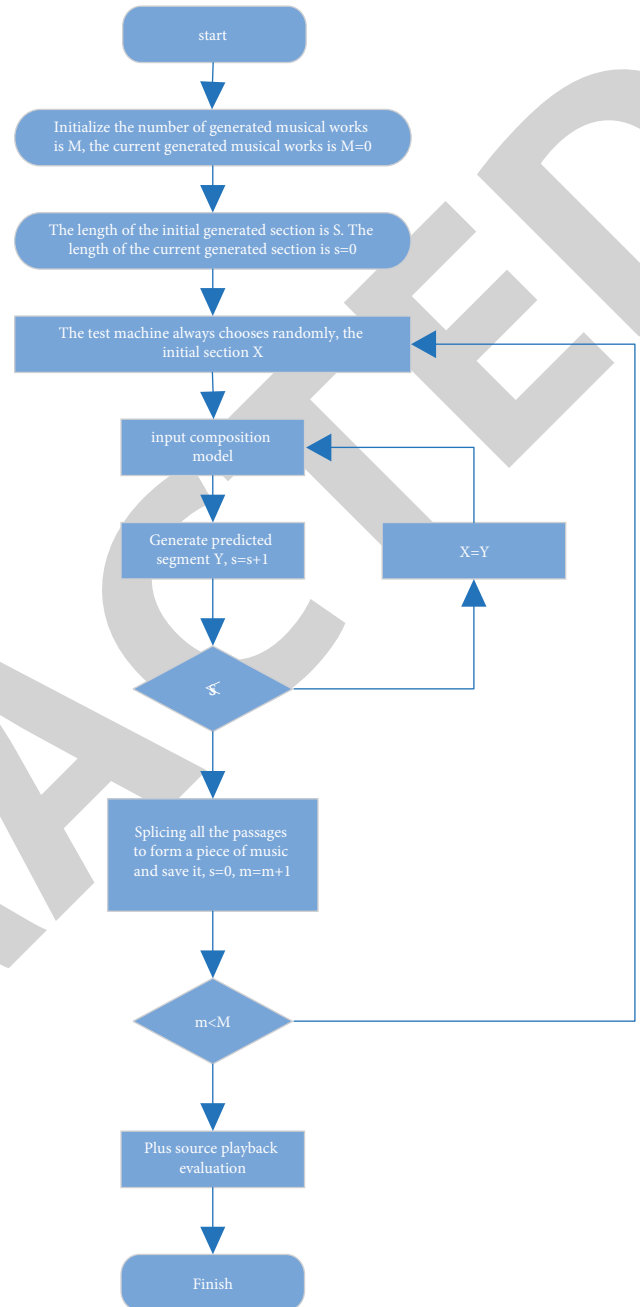

Figure 6: Flowchart of the automatic composition model.

Table 1: Server configuration.

Configure	Content
CPU	Intel (R) Xcon (R) CPU E5-2678 v3@2.50 GHz
RAM	251 GB
Graphics processor	TITAN RTX
Operating system	Ubuntu 18.04.2 LTS

composition scheme proposed in this paper are demonstrated. Based on the above, a brief description of the comparative experiments is shown in Table 2.

TABLE 2: Brief description of the experimental setup for the automatic composition comparison.

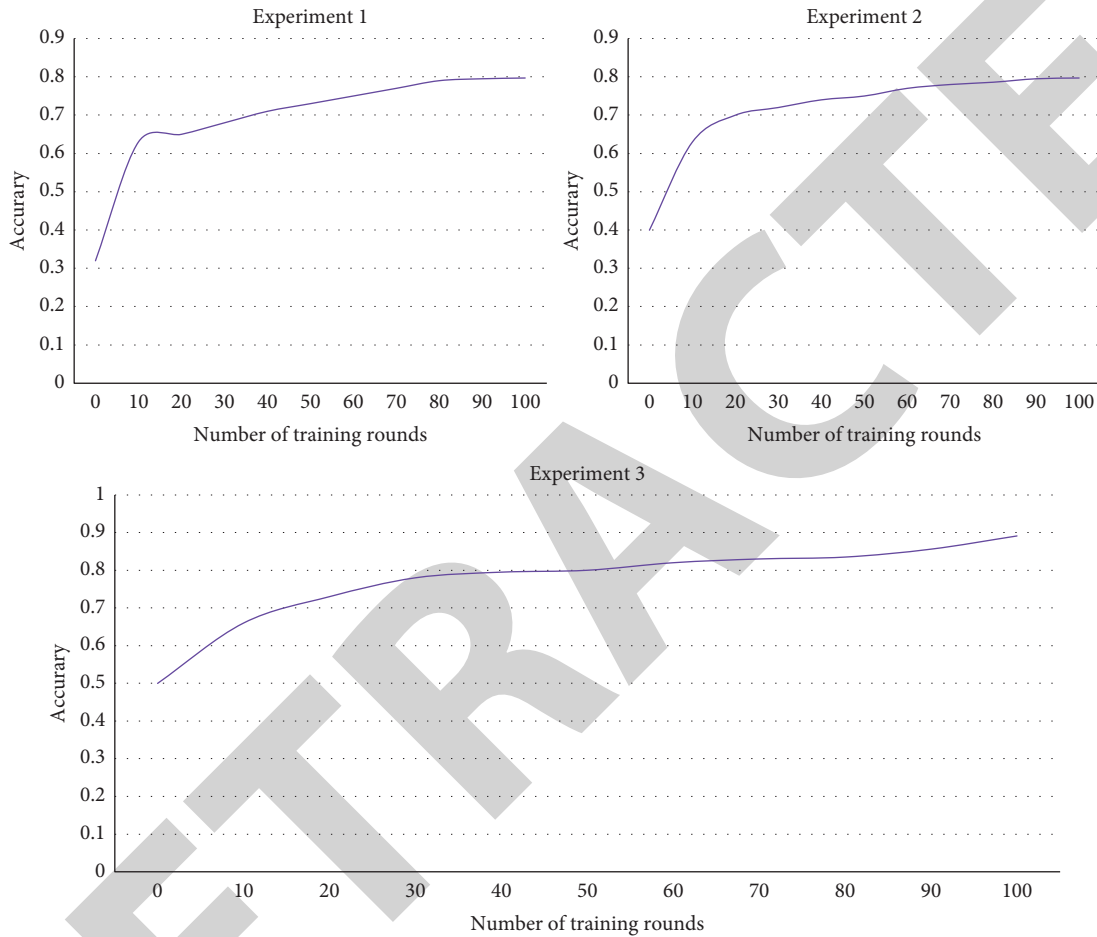| Experiment number	Feature extraction and representation	Composition model
1	One-hot encoding represents a sequence of notes	Bi-GRU + dense
2	The note feature vector represents the sequence of notes	Bi-GRU + dense
3	The note feature vector represents the sequence of notes	Bi-GRU + self-attention + -dense

FIGURE 7: Validation set Top1 Accuracy comparison curve.

This paper uses Python programming language and selects the deep learning framework of PyTorch to conduct the above three groups of automatic composition experiments based on deep learning. The main steps of the experiment are as follows: first, the MIDI main melody note sequence is obtained, and the specific steps have been introduced in detail in the third chapter. Each syllable contains 16 notes. Then, a dataset is formed according to two syllables as a piece of music, and the dataset is divided into training set, validation set, and test set, and the parts that do not overlap are independent of each other. The one-hot encoding representation is used in Experiment 1, the note feature vector representation proposed in this paper is used in Experiments 2 and 3, and then the Auto Compose model is selected. The Bi-GRU network is used in Experiments 1 and 2. Both the input and output layers have 129 neurons. The number of nodes in the hidden layer is the number of 32

notes contained in a sequence of musical notes; that is, the length of each input sequence of musical notes is 32. The number of neurons in the hidden layer is 256, and softmax is selected as the activation function. In Experiment 3, in addition to the above Bi-GRU network layer, there is a self-attention mechanism layer. The dimension of the Query and the Key is 64, and the dimension of the Value is 129. At the same time, in order to prevent overfitting, a 30% dropout layer is added after the Bi-GRU network and the self-attention mechanism layer. Then, the sequence of musical notes represented by different features in the training set is input into the model for training. During training, the cross-dimension function is used as the loss function to calculate the error, the Adam algorithm is used for optimization, the learning rate is set to 0.001, and the training is performed for 100 rounds. At the same time, the validation set is used to evaluate the performance of the network model for note
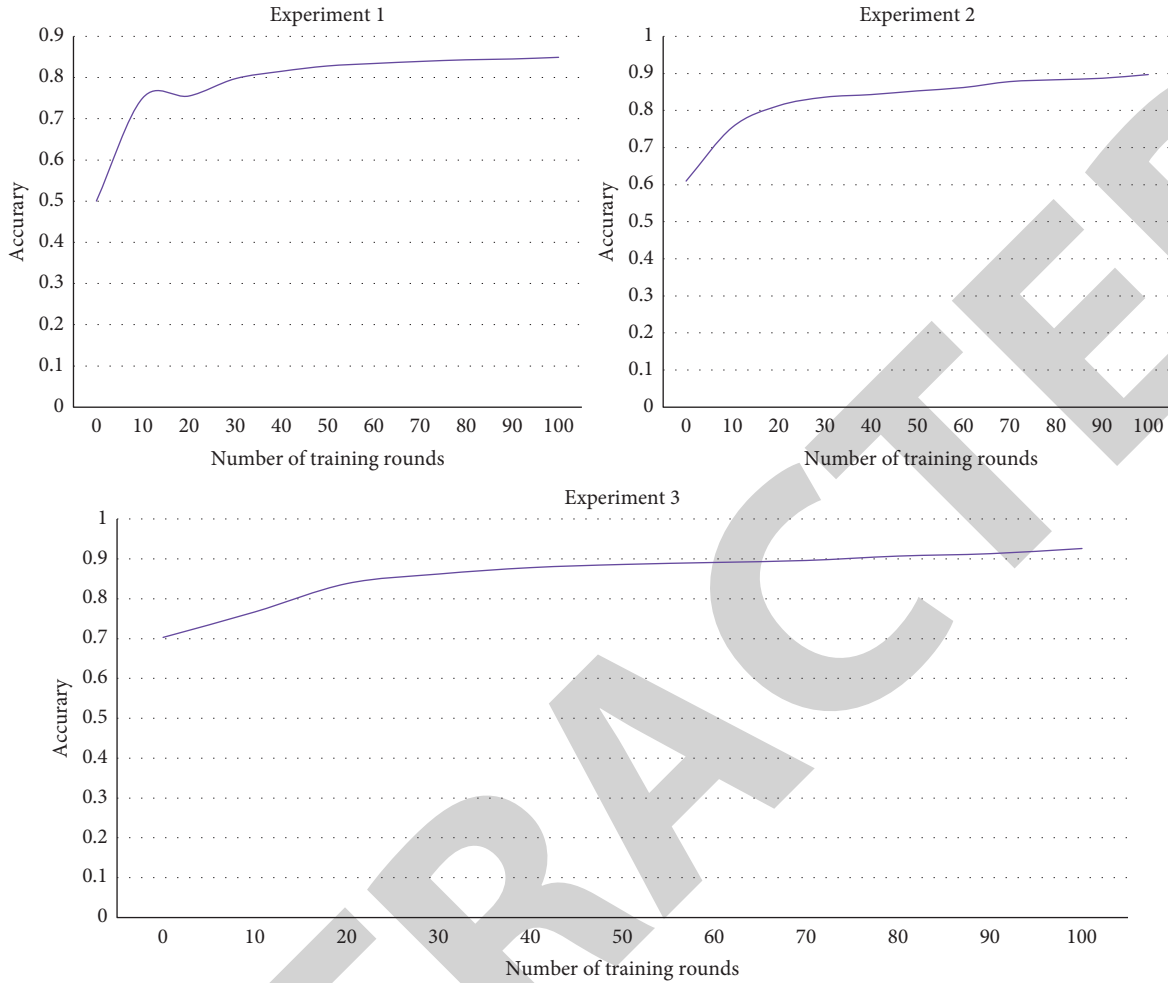
Figure 8: Validation set Top2 Accuracy comparison curve.

prediction, and the note prediction model with the best performance is obtained and saved. Finally, the number of composition sections is set to S. A sequence of notes of a passage is randomly selected from the test set as the input to the optimal note prediction model. The music segments of S are generated in sequence to form an automatically composed work, and in the same way, multiple automatically composed works can be obtained.

5.2. Evaluation of Composition Effects

5.2.1. Objective Evaluation of Composition Effect. That is, it is evaluated directly from the results of the composition model's note predictions. It uses the antipropagation algorithm in the artificial intelligence algorithm to accurately analyze the accuracy, making the results more objective. In this paper, the accuracy of note prediction is selected as the evaluation standard, and $N(i, j)$ is defined to represent the real note, and number i is predicted to be the number of notes with note number j. The overall note prediction accuracy is the ratio of the number of correct notes predicted to the total number of notes, that is,

$$\text{Accuracy} = \frac{\sum_i N(i, i)}{\sum_i \sum_j N(i, j)}. \tag{14}$$

In order to enrich the objective evaluation indicators, this paper will further refine the accuracy rate:

(1) Top1 Accuracy: when the predicted maximum probability note number is the same as the real note number, it is judged that the prediction is correct.

(2) Top2 Accuracy: when the predicted maximum first two-probability note number contains the real note number, it is judged that the prediction is correct.

(3) Top3 Accuracy: when the predicted top three-probability note number contains the real note number, it is judged that the prediction is correct.

After the automatic composition comparison Experiments 1, 2, and 3, the Top1 Accuracy, Top2 Accuracy, and Top3 Accuracy of the note prediction accuracy of the validation set obtained in this paper are shown in Figures 7–9.

By comparing Top1 and Top2, it can be concluded that different representation methods of musical features affect the final note prediction accuracy. The automatic
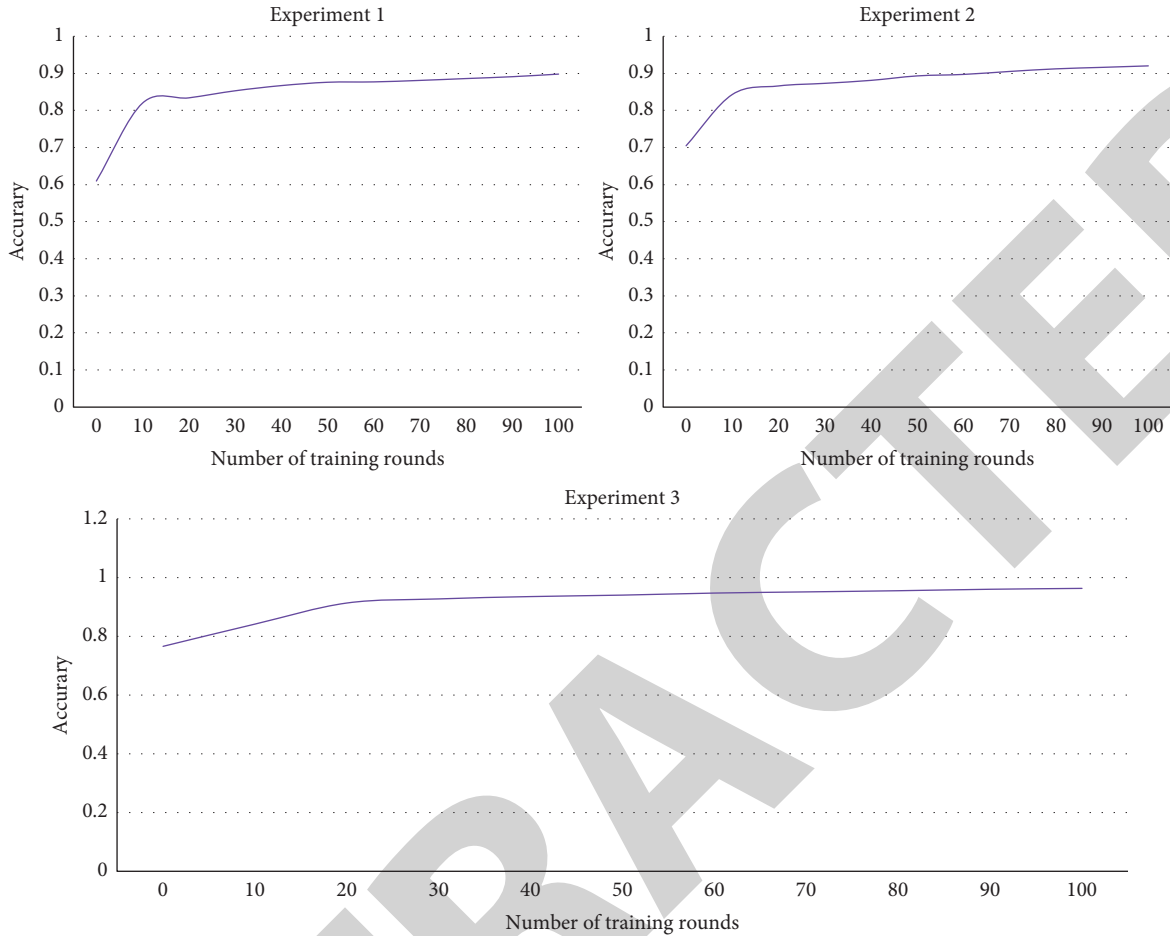
Figure 9: Validation set Top3 Accuracy comparison curve.

composition models selected by Top1 and Top2 are consistent, and the dimensions of their input data are also consistent. The only difference is that the Top1 input is a sequence of notes represented using a simple one-hot encoding. The Top2 input is a sequence of notes represented by note feature vectors with contextual semantic information generated in this paper [21]. Whether comparing the best note prediction accuracy that the model can achieve or comparing the curve changes of note prediction accuracy during the entire training process, Top2 is better than Top1.

By comparing Top2 and Top3, it can be concluded that the addition of the self-attention mechanism can effectively improve the accuracy of the final note prediction. The representation methods of the music features selected by Top2 and Top3 are consistent, and both use the note feature vectors with contextual semantic information generated in this paper. The difference is that the composition model selected by Top2 is only Bi-GRU network, while the composition model selected by Top3 is a combination of the Bi-GRU network and the self-attention mechanism [22]. It can be seen from this that the self-attention mechanism is good at mining the salient information between different notes in the note sequence that is beneficial to note prediction. By giving different weights to the notes in the note sequence, it

Table 3: System development tools for automatic action curve online evaluation.

Features	Content
Front-end programming language	JavaScript
Front-end framework	Vue 2.9.6
Backend programming language	Java
Background frame	Spring boot 2.2.6
Database software	MySQL5.7.21

Table 4: Music information to be rated.

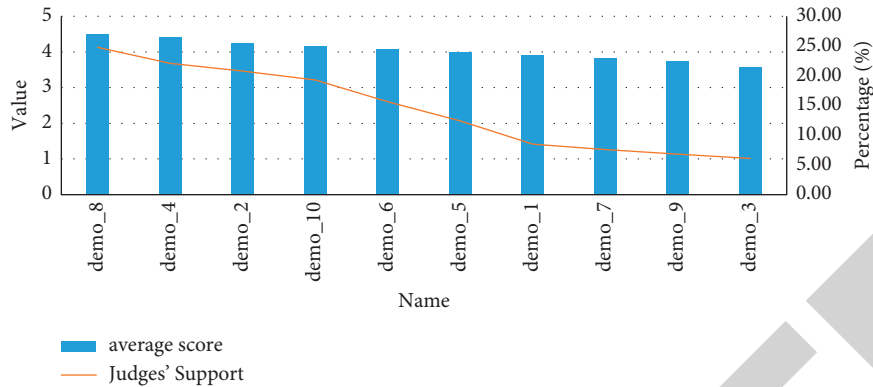| Numbering	Name	Way of composing
1	demo_1	Man-made song
2	demo_2	Automatic composition
3	demo_3	Automatic composition
4	demo_4	Man-made song
5	demo_5	Automatic composition
6	demo_6	Automatic composition
7	demo_7	Man-made song
8	decmo_8	Man-made song
9	demo_9	Automatic composition
10	demo_10	Man-made song

FIGURE 10: Statistical chart of online information evaluation ranking and the percentage of judges' support.

highlights the important information between the notes and ignores the secondary information, so that the final result is a better representation of the note sequence prediction.

By comparing Top1 and Top3, it can be concluded that both effective musical feature representation methods and suitable automatic composition models can improve the accuracy of note prediction. The method of Top1 is a general method used by most automatic composition or intelligent music generation and can also achieve certain experimental results. By adopting a more optimized music feature representation method and automatic composition model, Top3 improves the experimental effect of Top1 and improves the accuracy of note prediction. Top3 optimizes them separately from the representation of musical features and the design of automatic composition model, which are two key parts of automatic composition, and achieves good experimental results.

Based on the above statistics, the proposed method uses the note feature vector with contextual semantic information to identify the note sequence. The automatic composition scheme that selects the composition model of Bi-GRU network and self-attention mechanism is effective, and the accuracy of note prediction can basically meet the requirements in the field of automatic composition.

5.2.2. Subjective Evaluation of Composition Effect. In the subjective evaluation part of the composition effect, music lovers are invited to make corresponding scores according to their subjective feelings after the audition. Perceptual evaluation of the effect of the music generated by automatic composition can ensure the vividness and audibility of the music generated by the automatic composition scheme in this paper. The programming languages used by the automatic action curve evaluation system are shown in Table 3.

The standard of scoring is shown in Table 4, which adopts a five-point system. Five points are very satisfied, 4 points are satisfied, 3 points are relatively satisfied, 2 points

are average quality, and 1 point is poor quality. The way of producing the musical works for evaluation is divided into human work and automatic composition. Among them, 5 human work songs are obtained from the training set; 5 automatic compositions are generated by the automatic composition scheme Experiment 3 of this paper, and 12 seconds of audio are intercepted respectively. The specific work information is shown in Table 4.

A total of 15 music lovers are invited to participate in the online audition assessment. Among them, 10 are teachers and students majoring in music, and 5 are amateurs with certain musical literacy. The average score for each piece of music is calculated based on the scores they submitted, and the final result is shown in Figure 10.

From Figure 10, the following conclusions can be drawn:

(1) The music score of human works is generally higher, and the score can fluctuate around the center value of 4 points. This phenomenon shows that because the composer has accumulated a lot of musical literacy, the composition effect is generally better and the feeling to the audience tends to be stable.

(2) The music demo_2 generated by the automatic composition scheme in this paper also scored a high score of 4.25, ranking third, and the difference between the scores of the first two works is not very big. This phenomenon shows that, for listeners, it is difficult to distinguish whether the music is produced by human work or the automatic composition of the text. Moreover, the music effect produced by the automatic composition of this paper gives the audience a better subjective feeling.

(3) The scores of the music generated by the automatic composition scheme in this paper are not very stable. There are demo_2 with a high score of 4.25 and demo_3 with a minimum score of 3.58. This phenomenon shows that there are some differences in the quality of music produced by automatic

composition, and there is still room for further optimization.

6. Conclusions

This paper uses note feature vectors with contextual semantic information to represent musical features and chooses Bi-GRU network and an automatic composition model with self-attention mechanism. In the two key modules of the representation of music features and the design of automatic composition model, an optimization scheme is given, and the effectiveness of the automatic composition scheme in this paper is verified by experiments. In order to evaluate the effect of this automatic composition scheme, this paper proposes two evaluation methods: objective and subjective [23]. The objective evaluation is based on the accuracy of note prediction. The method of comparative experiment is used to optimize the representation of music features and the design of the automatic composition model. The two methods jointly ensure the accuracy and vividness of the music produced by the automatic composition in this paper. However, the automatic composition in this paper only studies the main melody of a single track, and a good piece of music should also have chord accompaniment. In the follow-up research, after the main melody is generated, appropriate chord accompaniment can be added to the main melody to form a relatively complete music. Therefore, the research on chord accompaniment should be added in the follow-up practical research. After the automatic composition technology is not limited by the computing performance of hardware such as servers, the follow-up research can expand the dataset according to the idea of this paper. If the depth of the automatic composition model is increased, the automatic composition method will satisfy people's rising demand for music appreciation in the future.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] M. K. Tran, C. Goulding, and E. Shiu, "The orchestra of ideas: using music to enhance the 'fuzzy front end' phase of product innovation," *Journal of Business Research*, vol. 85, pp. 504–513, 2018.

[2] H. . Zhang, "Explore the importance of strengthening folk music theory learning in music lessons in hohhot[J]," *Advances in Higher Education*, vol. 3, no. 2, p. 171, 2019.

[3] D. M. Zhaleko, "The value transformation and semantic space of kitsch in the music practice at the turn of the 20th and 21st centuries," *Problems of Interaction Between Arts Pedagogy and the Theory and Practice of Education*, vol. 56, no. 56, pp. 213–237, 2020.

[4] M.K. K Outsomichalis, "Ad-Hoc Aesthetics: context-dependent composition strategies in music and sound art," *Organised Sound*, vol. 23, no. 1, pp. 12–19, 2017.

[5] B. Moseley, "Cycles in webern's late music," *Journal of Music Theory*, vol. 62, no. 2, pp. 165–204, 2018.

[6] P. Dolezel, F. Holik, J. Merta, and D. Stursa, "Optimization of a depiction procedure for an artificial intelligence-based network protection system using a genetic algorithm," *Applied Sciences*, vol. 11, no. 5, p. 2012, 2021.

[7] J. Musulin, S. B. Egota, D. Tifani, I. Lorencin, and E. Markova-Car, "Application of artificial intelligence-based regression methods in the problem of COVID-19 spread prediction: a systematic review," *International Journal of Environmental Research and Public Health*, vol. 18, no. 8, pp. 1–39, 2021.

[8] D. Corradini, L. Brizi, C. Gaudiano et al., "Challenges in the use of artificial intelligence for prostate cancer diagnosis from multiparametric imaging data," *Cancers*, vol. 13, no. 16, p. 3944, 2021.

[9] N. Dukhi, R. Sewpaul, M. Derrick Sekgala, and O. Olawale Awe, "Artificial intelligence approach for analyzing anaemia prevalence in children and adolescents in brics countries: a review," *Current Research in Nutrition and Food Science Journal*, vol. 9, no. 1, pp. 01–10, 2021.

[10] C. Cross, "Using artificial intelligence (AI) and deepfakes to deceive victims: the need to rethink current romance fraud prevention messaging," *Crime Prevention and Community Safety*, vol. 24, no. 1, pp. 1–12, 2022.

[11] K. W. A. Mahardika, H. Santosa, and N. W. Ardini, "Transformasi lagu "kacang Dari" ke dalam chamber music," *Gondang Jurnal Seni dan Budaya*, vol. 4, no. 2, p. 96, 2020.

[12] K. C. Leonard, F. Hasan, H. F. Sneddon, and F. You, "Can artificial intelligence and machine learning Be used to accelerate sustainable chemistry and engineering?" *ACS Sustainable Chemistry & Engineering*, vol. 9, no. 18, pp. 6126–6129, 2021.

[13] A. Jeavons, "What is artificial intelligence?" *Research World*, vol. 2017, no. 65, p. 75, 2017.

[14] R. O. Vaishya and K. V. abhishek, "Review on micro-machining of pyrex glass using wire-ECDM," *Journal of Antimicrobial Chemotherapy: A Journal of Composition Theory*, vol. 13, no. 8, pp. 56–65, 2020.

[15] Y. C. Wang, X. Xu, A. Hajra et al., "Current advancement in diagnosing atrial fibrillation by utilizing wearable devices and artificial intelligence: a review study," *Diagnostics*, vol. 12, no. 3, p. 689, 2022.

[16] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: a review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.

[17] J. Licato and Z. Zhang, "Correction to: evaluating representational systems in artificial intelligence," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 1463–1493, 2019.

[18] J. Jiao, H. Pan, C. Chen, T. Jin, Y. Dong, and J. Chen, "Two-stage lesion detection approach based on dimension-decomposition and 3D context," *Tsinghua Science and Technology*, vol. 27, no. 1, pp. 103–113, 2022.

[19] M. Tosi and J. Reis, "Understanding the evolution of a scientific field by clustering and visualizing knowledge graphs: [J]," *Journal of Information Science*, vol. 48, no. 1, pp. 71–89, 2022.

[20] D. A. Marunova, N. V. Pronina, A. G. Kalmykov, D. A. Ivanova, and G. A Kalmykov, "The transformation stages of organic matter from the tutleim formation rocks

with respect to its maceral composition," *Moscow University Geology Bulletin*, vol. 77, no. 1, pp. 88–100, 2022.

[21] G. Dimauro, V. Bevilacqua, and L. Pecchia, "Bioelectronic technologies and artificial intelligence for medical diagnosis and healthcare," *Electronics*, vol. 10, no. 11, p. 1242, 2021.

[22] N. Biamonte, "Online music theory in music theory online," *Zeitschrift der Gesellschaft für Musiktheorie [Journal of the German-Speaking Society of Music Theory]*, vol. 13, no. 2, pp. 297–309, 2016.

[23] Y. O. Serdiuk, "Chamber music works by amanda maier in the context of European romanticism," *Problems of Interaction Between Arts Pedagogy and the Theory and Practice of Education*, vol. 56, no. 56, pp. 121–135, 2020.