Research on the Algorithm System for Generation and Creation by Computer Artificial Intelligence and Support Vector Machine

Xuan Wu
Xiamen Huaxia University, Fujian, Xiamen 361000, china

*Corresponding author: wuxuan121667403@icloud.com

Abstract. Algorithmic composition is also called automated composition. It is an attempt to use a specific form of process. Composers make full use of computers to carry out music creation and reduce their access. In this paper, based on the standard support vector machine (SVM) learning neural network, the least square support vector machine (LS-SVM) is combined with the recurrent neural network, and a new least square support vector machine learning neural network is proposed. The article realizes the efficient end-to-end multi-dimensional sound wave time series generation model Music-coder, through which the music style music of the famous singer Jay Chou is generated, and the quantified similarity with the real Jay Chou music data set reaches a maximum of 97.73%. The project in this paper shows that intelligent algorithm as a composition tool for music generation and creation is an effective music production program and will bring new development to music production.

Keywords: Computer, artificial intelligence, support vector machine, song creation, algorithm system.

1. Introduction
When using computer-aided algorithms to develop composition design, it is necessary to increase the degree of attention to knowledge rules, because knowledge rules are a static form. When the knowledge system is established, it is necessary to ensure that the knowledge system is static. According to the rules and standards given by the computer, develop composition, use the rhythm and law of Sarabande dance, use the rhythm of minuet and waltz, use Gavotte, etc., the emotional expression, musical chord rhythm, law and rhythm characteristics of the knot. To show the richness and diversity of rhythm and emotion, to achieve the goal of composition flexibility and richness, and to show composition styles with different rhythms and strengths.

Algorithmic composition or automatic composition is a study that attempts to use a certain formal process to minimize the degree of interpersonal interaction when people (or composers) use computers to create music [1]. Although the research on computer composition system is very active abroad. However, this technology requires close cooperation between computer experts and musicians. It is still in its infancy in the Chinese music industry. Studying the problem of algorithmic composition on the one hand allows us to understand and simulate the way of thinking of composers in the specific process of music creation, on the other hand, the different forms of music created by the composition
program developed based on the research technology of algorithmic composition Works can also be entertaining. With the popularization and development of computer technology, algorithm composition will become an important auxiliary method in music creation.

At present, artificial intelligence technology is widely used in social production fields such as industry and agriculture at home and abroad. Algorithm composition (computer composition) technology in artificial intelligence music creation technology abroad is in the research integration stage. Research in this area is very active, but as this technology requires close cooperation between computer experts and musicians, it is still in its infancy in China. The application of artificial intelligence music creation technology in the domestic creative industry is very rare. The music creation of the domestic creative industry is still based on traditional manual creation [2]. The generation of a music segment requires the cooperation of one or more personnel and undergoes many corrections before it can be generated. The current methods that can be applied to artificial intelligence music creation include: algorithmic composition method using Markov conversion table, algorithmic composition method based on rule-based knowledge base system, algorithmic composition method based on music grammar, algorithmic composition method based on artificial neural network, Algorithm composition method of genetic algorithm. Various composing algorithms have their own advantages and problems. There are many problems in the research of algorithmic composition. The most representative ones are: how to intervene in the process of human-computer interaction to make music creation more effective and creative; the knowledge expression of music, how to establish the surface structure of music Correspondence with the deep logic of music, whether the knowledge expression mechanism is practical and effective, and whether the cost of construction is high; the quality evaluation of the works generated by the system requires the establishment of a large library of formal rules, and the deduction of an accurate formal rule requires a large number of music clips of the same style. Mutual verification, the deduction process is very cumbersome and complicated. The purpose of this article is to study an algorithm that can well meet the needs of the creative industry for audio rapid creation.

2. Description of the work of this article

Through the first chapter, we have seen that corresponding to different music materials and types and tasks, different generative models and solutions can be selected. In this article, we mainly realize the generation of music of a specific composer genre through a neural network model. Since the carrier of music is sound, the data of the sound type can be regarded as a fixed-length time series data [3]. Based on this, we chose a neural network structure LSTM that can handle time series data types to implement a music generator through a supervisory mechanism with additional constraints, and perform music generation, and achieved excellent music generation effects. The main workflow of Music-coder music generator is shown in Figure 1. The main work of this paper is:

(1) Through the application of an LSTM deep learning model, the music generation of the specific composition is realized. (2) The similarity calculation between the generated music collection and the original music collection is carried out, and a good effect is achieved in the distribution of the generated music characteristics on the two indicators of similarity. (3) Double-blind monitoring of the generated music data by real musicians and composers, with an average pass rate of over 80%.
3. LS-SVM classification learning neural network

Given a classification training set \((z_i, y_i), i=1, \cdots, N\), where \(z_i \in \mathbb{R}^n\) is the training sample, and \(y_i \in \{-1, +1\}\) is the class corresponding to the sample, its classification decision surface can be expressed as \(f(z) = w^T \varphi(z) + b\), where \(w\) is the weight matrix, \(b\) is the offset, and \(\varphi(\cdot)\) is the distance from the input space to Non-linear mapping of feature space. LS-SVM classification learning is to solve the following constrained optimization problem:

\[
\begin{align*}
\min_{w,b,e} J(w,e) & = \frac{1}{2} w^T w + \gamma \sum_{i=1}^{N} e_i^2 \\
\text{s.t.} \quad y_i[w^T \varphi(x_i) + b] & = 1 - e_i, i=1, \cdots, N
\end{align*}
\]

(1)

To solve this problem, the Lagrange function can be introduced:

\[
L(w,b,e,\alpha) = \frac{1}{2} w^T w + \gamma \sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N} \alpha_i \{y_i[w^T \varphi(x_i) + b] - 1 + e_i\}
\]

(3)

Where: \(\alpha_i\) is the Lagrange multiplier. Obtain the optimal conditions of the problem by obtaining partial derivatives for each parameter, and subtracting \(w\) and \(e_i\) to obtain:

\[
1 - b y_i - \sum_{j=1}^{N} \alpha_j q_{ij} - \gamma \alpha_i = 0
\]

(4)

\[
\sum_{i=1}^{N} \alpha_i y_i = 0
\]

(5)

Where: \(q_{ij} = y_i y_j K_{ij} = \varphi(z_i)^T \varphi(z_j)\) is defined as the kernel function. If the kernel function satisfies the Meecer condition and the symmetric matrix \(Q = [q_{ij}]\) is positive definite, then the problem is an optimal convex problem, that is, it has only one global solution. This paper uses neural network to
solve the LS-SVM classification problem. The proposed neural network model is described by the following dynamic equation:

\[
\dot{\alpha}_i = \frac{\partial L}{\partial \alpha_i} = \sum_{i=1}^{N} a_i y_i \tag{6}
\]

\[
\alpha_i = \frac{\partial L}{\partial \alpha_i} = 1 - b y_i - \sum_{j=1}^{N} a_j q_{ij} - \gamma \alpha_i \tag{7}
\]

It is easy to see from the dynamic equation: the dynamic system (6), (7) satisfies the optimization conditions (4), (5) at the equilibrium point, that is, the proposed neural network satisfies KKT at the equilibrium point Conditional. In this way, when the proposed dynamic network converges to the equilibrium point, the LS-SVM problem can be solved. Equations (6) and (7) can be implemented using a recurrent neural network as shown in Figure 2.

![Figure 2. Network topology diagram of LS-SVM classification learning problem.](image)

The network structure can be easily implemented by analogy circuit hardware. Figure 3 shows the circuit implementation corresponding to the \( i \) Lagrange multiplier in the neural network, where \( v_{\alpha_i} \) corresponds to the topological structure \( \alpha_i \); \( v_b \) in Figure 2; corresponds to the offset \( b \); \( \gamma R_0 \); corresponds to the feedback structure of the integrator; \( R_0 |q_{ij}| \) corresponds to In the part of the connection weight \( q_{ij} \); the circuit uses multiple input linear integrators to realize the addition and integration links.

![Figure 3. Circuit implementation of neural network Lagrange multiplier \( \alpha_i \).](image)
As shown in the figure above, because the operational amplifier works in a linear state, it satisfies "virtual short and virtual off". It can be derived from this:

\[ \tau v_a = 1 - \sum_{j=1}^{N} v_{ij} q_{ij} \]  
\[ \frac{1}{\tau} \frac{dv_a}{dt} = v_b - v_a \]  

(8)

In the formula: \( \tau = R_0 C \). If \( \tau = 1 \) is taken, the circuit can realize the dynamic equation (7). The dynamic equation (6) can be realized by a simple linear integrator. From equation (8), it can be concluded that in terms of value, the positive and negative of \( v_a, v_b \) and \( q_v \) and can be reflected by \( \pm v_{ij} \). As for the entire circuit, if there are N training samples, N+1 operational amplifiers and N(N+3) connection resistors are required, and the adjustment of the LS-SVM penalty factor \( \gamma \) is realized by adjusting the resistor \( \gamma R_0 \).

4. System Design

4.1. LSTM structure

The key to LSTM is the cell state. This conveyor belt structure passes directly through the entire process chain with only some small linear interactions [4]. The information carried above can be easily transferred to the next computing unit without changing. At the same time, LSTM has the ability to add or delete information to the cell state. This ability is controlled by a gate structure. They are composed of a Sigmoid neural network layer and an element-level multiplication operation, which can selectively let information pass. A complete LSTM has 3 gate structures to protect and control the cell state, as shown in Figure 4.

![Figure 4. LSTM model structure.](image)

LSTM uses two gates to control the content of the cell state \( c \), one is the forget gate, which determines how much of the cell state from the previous moment is retained to the current moment. The other is the input gate, which determines how much of the network's input at the current moment is saved to the unit state. LSTM uses output gates to control how much of the unit status is output to the current output value of LSTM. \( W \) is the weight matrix of the calculation process, \( \text{tanh} \) is the activation function, \( C \) is the unit state, \( f \) is the forget gate, \( h \) is the hidden information, and \( x \) is the input information at a specific time.

4.2. Supervisor mechanism of Music Coder structure

The supervisor mechanism is mainly to check the intermediate production sequence results of LSTM (intermediate calculation music sequence), and eliminate sequences that do not meet the supervision...
rules (music feature constraints, mainly configured as pitch constraints and musical style constraints). The mechanism will enable the input of a single music sequence of the LSTM to achieve a specific generation effect without being too long. This mode improves the processing speed and efficiency of LSTM, thereby making the generation training process converge faster, as shown in Figure 5.

![Figure 5. Supervisor mechanism.](image)

We used 4 cards of Nvidia Titan XPGPU deep learning server in the process of artificial intelligence model training, carried out 10,000 iterations of training, optimized the model accuracy rate by controlling the model loss function and SoftMax function, and finally made the model efficiency more than 90%. The entire model training took 35 hours. Based on the deep learning model trained on Data Set 1 and Data Set 2, we get mixed-track music and single-track music [5]. Through music trimming, the single-track is combined as a supplement to the overall music, and finally the generated music is obtained.

5. Generate music result analysis

5.1. Quantitative analysis-generating music similarity calculations

This paper uses the calculation of the similarity between the generated music and the original music to quantitatively evaluate the music generation effect. By calculating the similarity between each generated music and each music in the original music set, the corresponding similarity of the one with the highest similarity is selected as the final similarity. Based on the trained generative model, we generate 50 music, each about 2-3 minutes, the generation speed is within 10s, and the Pearson similarity is selected to calculate the similarity between the generated music data set and the test data set (original music) and effect evaluation. From the original 100 pieces of real music, 50 pieces are randomly selected as the seed set S, numbered S1, S2...S50, and similar calculation benchmarks are performed [6]. For the generated music, 30 pieces of generated music are randomly selected from 50 pieces of generated music as generating set G, numbered G1, G2...G30, and the similarity calculation with the seed set is performed, and the similarity of the seed music set S itself is calculated as a reference. Calculate the maximum similarity, minimum similarity and average similarity corresponding to the G music set respectively, and the statistics of the detection results are shown in Table 1.

| Detect music collection | Maximum similarity | Minimum similarity | Average similarity |
|-------------------------|--------------------|--------------------|--------------------|
| G/G                     | 100%               | 0.11%              | 60.64%             |
| S/S                     | 100%               | 30.37%             | 82.80%             |
| G/S                     | 97.73%             | 8.89%              | 60.13%             |
5.2. Qualitative analysis-generating music expert evaluation statistics

In addition to the above quantitative similarity calculation, we also evaluate the effectiveness of our music generation through tests by artificial domain experts [7]. We conduct cross-testing by experts in the music field on the generated data set, by randomly mixing the seed music set S (S music set contains 50 songs) and the generated music data set G (G music set contains 50 songs), as the overall test set T, Numbered as T1, T2...T100. We invited 3 music professionals, namely, music students from the Conservatory of Music of Minzu University of China, Haidie Music Producer, and Ruili Fashion Musicians, numbered A, B, and C in sequence, and judge whether the music in Test T is original or not through audition. Music, the test results are shown in Table 2. The music participants completed a random mix of the original music collection to generate a total of 100 music. Each tester listened to 100 music and judged whether each piece of music is Jay Chou's original music. The test results are shown in Table 2.

| Tester ID | Detection accuracy rate of S music collection | G music collection detection pass rate | T music collection detection pass rate |
|-----------|---------------------------------------------|--------------------------------------|---------------------------------------|
| A         | 100%                                        | 76%                                  | 88%                                   |
| B         | 100%                                        | 78%                                  | 89%                                   |
| C         | 100%                                        | 86%                                  | 93%                                   |
| Mean      | 100%                                        | 80%                                  | 90%                                   |

6. Conclusion

The algorithm discussed in this paper, based on the system constructed on this basis, the quality of the music generated by it will ultimately require manual identification, correction, and selection. Because the "knowledge" about music in these systems is not like all other methods, it is "inferred" by people based on a certain music theory or from people's works. The research and development of Markov's random process self-organizing evolutionary algorithm composition system will bring revolutionary development to the creative industry Ruisi. The introduction of artificial intelligence creation in the creative industry will not only bring industrial benefits and work efficiency improvements for enterprises and individuals, but also the industry provides new creative methods and ideas in the field of creative creation, and the artificial intelligence of the creative industry is bound to become a development method of the world's creative industry.

References

[1] Sturm, B. L., Ben-Tal, O., Monaghan, Ú., Collins, N., Herremans, D., Chew, E., ... & Pachet, F. Machine learning research that matters for music creation: A case study. Journal of New Music Research, 48 (1) (2019) 36-55.
[2] Gorbunova, I. B., & Morozov, S. A. Music Computer Technologies in Teaching Students with Visual Impairments in Music Education Institutions. Journal of Critical Reviews, 7 (19) (2020) 3274-3284.
[3] Loughran, R., & O’Neill, M. Evolutionary music: applying evolutionary computation to the art of creating music. Genetic Programming and Evolvable Machines, 21 (1) (2020) 55-85.
[4] Studley, T., Drummond, J., Scott, N., & Nesbitt, K. Evaluating Digital Games for Competitive Music Composition. Organised Sound, 25 (1) (2020) 75-88.
[5] Thorogood, M., Fan, J., & Pasquier, P. A framework for computer-assisted sound design systems supported by modelling affective and perceptual properties of soundscape. Journal of New Music Research, 48 (3) (2019) 264-280.
[6] Alieva, I. G., Gorbunova, I. B., & Mezentseva, S. V. Music computer technologies as a worthwhile means of folklore studying, preserving and transmission. Utopía y praxis latinoamericana: revista internacional de filosofía iberoamericana y teoría social, 7 (6) (2019) 118-131.
[7] Benedict, C., & O’Leary, J. Reconceptualizing “music making:” Music technology and freedom in the age of neoliberalism. Action, Criticism, and Theory for Music Education, 18 (1) (2019) 26-33.