

Retraction

Retracted: Innovative Construction of Reinforcement Learning Model for Information Fusion in Music Education

Security and Communication Networks

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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] Z. Zhang, "Innovative Construction of Reinforcement Learning Model for Information Fusion in Music Education," *Security and Communication Networks*, vol. 2022, Article ID 7367418, 12 pages, 2022.

Research Article

Innovative Construction of Reinforcement Learning Model for Information Fusion in Music Education

Zehui Zhang 

Education Department, Nanchang Normal University, Nanchang 330032, China

Correspondence should be addressed to Zehui Zhang; zhangzehuincnu@seu.edu.mk

Received 7 January 2022; Revised 26 February 2022; Accepted 10 March 2022; Published 23 June 2022

Academic Editor: Chin-Ling Chen

Copyright © 2022 Zehui Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the effect of modern music education, this paper applies the digital information technology of music resources to the construction of the music teaching system and derives two new types of semi-decision-making process reinforcement learning algorithms based on the Bellman optimality equation base on discrete time. Moreover, this paper uses the comparative research methods to obtain the Q-value learning curve of the incremental value iterative reinforcement learning algorithm based on the semi-Markov decision process and the incremental value iterative reinforcement learning algorithm based on the dichotomy to improve the fusion effect of music teaching resources. Finally, this paper combines the actual needs of modern music education to construct an intelligent music teaching model.

1. Introduction

Music language plays an important role in the music development of a country. The mother tongue is one or several languages that a person first contacts, learns, and masters [1]. The mother tongue is generally contacted from a young age and continues to be used in adolescents or later. Moreover, in a person's family or formal education, especially in the early stages, a considerable part of the knowledge is imparted through the mother tongue. The language of our country is mainly Chinese, and the music form is mainly Chinese folk songs and the five-tone mode of musical instruments [2]. However, the current universal music language in the world is mainly based on the Western education system, which puts a test for us to understand the world's music culture and creates difficulties for us to convey, learn, and understand the world's excellent music culture. Therefore, modern national educational institutions propose that music culture should inherit and carry forward China's excellent traditional culture, strive to learn the Western music culture knowledge, and integrate the advantages of the two aspects to establish a world cultural form with Chinese cultural heritage. This also puts forward higher requirements for our

country's music education in the world and also establishes the world status that our country's music education must achieve [3].

From a broad perspective, the ongoing music education in China is largely based on the level of teachers teaching students to learn, and teachers only understand the teaching of students simply listening to music. This traditional teaching mode has been going through for a long time. This kind of teaching method dominated by the transmission of knowledge is often easy to teach knowledge to death, and it is difficult for students to cultivate character and develop personality. Although I have learned a lot of music knowledge, it is easy to focus on music knowledge and ignore the deeper aspects of music.

At present, the overall development of China's music education industry is facing great challenges, and the trend of international music education cannot be responded quickly; knowledge updating lags behind; music education has not been solved at a higher level; the publication of academic works is extremely rare; China rarely participates in international music education cooperation, and the number of participants is also small. In addition, local traditional music teaching and teacher education in the

world, such as oriental music teaching, are considered to be vacant in many fields. It is undeniable that the cultural construction pattern of global integration is inevitable. It has made a positive response to the development of human science and technology, culture, and economy. The living space has been gathered and developed. Therefore, it is not profound for any modern country to abandon its own development or abandon its existing cultural traditions to talk about the development of the world pattern, and it will face many difficulties in the future. Chinese music education first needs to consider the actual background of China's digital music teaching and the development trend of current international music education, such as the mother tongue problem, the problem of education teachers, and the update of knowledge, which have very important historical significance for modern education in my country.

Due to the rapid development of multimedia technology, a large amount of information disseminated on the Internet is in different modalities. For example, for any web page, whether it is about sports, Weibo, or military politics, it will basically contain audio, video, and images, and any one. All modal data information is indispensable. The data information of these different modalities are all developed around the theme of the web page, so cross-modal analysis can be used, that is, the data of different modalities can be found in a certain way to find the correlation between them, so as to carry out comparative analysis. For example, when people listen to a piece of music, the music contains text lyrics that describe the content of the music. These audios and their corresponding text lyrics can be regarded as bimodal information, and they are related to each other in emotional expression. Especially in the era of big data and cloud computing sweeping the network and research in various fields, cross-modal analysis has become a hot spot in this field and has received more and more attention. Therefore, we need to continue to explore and analyze it.

The research and application of traditional single-modal data has been quite extensive, but in the era of multimedia data and big data becoming the mainstream, users' requirements for data are getting higher and higher, and the analysis of single-modal data can no longer keep up with the needs of the times, and people can no longer conduct broader research based on it. Because of the limitation of single modality data, the data information contained in it cannot be fully utilized by people and it cannot describe multimedia data more comprehensively. For multimodal data, they are often complementary to each other, and can express better results when characterizing objects.

The organization structure of this paper is as follows: The first part introduces the related concepts and background of music education and multimodal fusion analysis. The second part is the literature review part, which describes the main work of the music multimodal fusion analysis at home and abroad at present and the music multi-modal fusion analysis. The current situation and difficulties are introduced. The third part proposes an incremental value iterative reinforcement learning algorithm based on the needs of music education, and the algorithm is used to construct a multimodal fusion analysis model for music. The fourth part is

based on the support of the third part. The multimodal information fusion music education system is used for learning. The fifth part is to test the validity of the model in this paper by means of experiments. The conclusion part is a summary of the research results of this paper and puts forward the prospect.

The main contributions of this paper are as follows: (1) according to the iterative form of the Bellman optimality equation, a unified analysis framework for SMDP reinforcement learning algorithm is given, which can effectively promote the fusion of multimodal music information; (2) reinforcement learning under the average reward criterion algorithms to carry out related research, so that the music information fusion reinforcement learning algorithm can be more widely used in practical systems.

This paper uses the reinforcement learning model based on information fusion to innovate the way of music education, change the traditional music education model, and improve the actual effect of music education.

2. Related Work

Due to the rapid development of multimedia technology, a large amount of information disseminated on the Internet is of different modalities. For example, for any webpage, whether it is about sports, Weibo, or military politics, it basically contains audio, video, and images, and any one of them. The data information of each modal is indispensable. Moreover, the data information of these different modalities is developed around the theme of the web page, so cross-modal analysis can be used; that is, the data of different modalities can be found in a certain way to find the relationship between them, so as to conduct a comparative analysis. For example, when people listen to a piece of music, the music contains text lyrics that describe the music content. These audio and the corresponding text lyrics can be regarded as bimodal information, and they are related to each other in emotional expression [4]. Especially in the era when big data and cloud computing are sweeping the network and research in various fields, cross-modal analysis has become a hot spot in this field and has attracted more and more attention. Therefore, we need to continue to explore and analyze it. For cross-modal analysis, Cano [5] found the correlation between the two-modal data of text and image through the joint learning of the two-modal data and improved the accuracy of image search, text search, and multi-modal data search. In order to eliminate the duality of semantic expression of text information and improve the accuracy of expression, the text information is correlated with visual characteristics, and the text is expressed more semantically [6]. Dickens [7], based on the Probabilistic Semantic Model (PLSA), learns and trains multimodal information, obtains a hierarchical representation mechanism of multiple modal data, finds the correlation information of data between multiple modalities, and improves data search and query efficiency. Gonçalves [8] based on the multimodal LPP algorithm, trains the music and image sample sets to obtain a common low-dimensional subspace, in which the same sample and different modal data will be close together,

that is, multiple modal data. Relevance expression can be carried out, which eliminates the gap in semantic relevance between multiple modal data.

The research and application of traditional single-modal data has been quite extensive, but in the era when multimedia data and big data have become the mainstream, users have higher and higher requirements for data, and the analysis of single-modal data can no longer keep up with the needs of the development of the times. People can no longer carry out a wider range of research based on it. Because of the limitations of single-modal data, the data information it contains cannot be fully utilized by people, and it cannot describe multimedia data more comprehensively. For multimodal data, they are often complementary and interrelated and can express better results when portraying objects [9]. In the research on the automatic generation of multimodal fusion family music albums, by fusing two different modal data of image and music, it played a key role in the semantic understanding of music and images, which has a significant impact on the research results [10]. For multi-modal fusion, Gorbunova [11] compares music and images through the analysis and training of graphs and obtains good results in practical applications. Khulusi [12] based on the principle of Bernoulli distribution aimed to annotate images and texts. Magnusson [13] based on deep learning RBM (Restricted Boltzmann Machines) and DBN (Deep Belief Network) aimed to provide a common feature representation for multimodal data and apply it in various recognition fields. Partesotti [14] used the mutual mapping between text and image to find the mutual relationship between them, and then carried out text identification on the image. Scavone [15] proposed a supervised learning model, which is an improvement on the unsupervised model. It maps multimodal data to a subspace, in which the same type of data will be moved closer to each other. In addition, multimodal fusion has also been widely used in classification tasks and the purpose is to improve the accuracy of classification. Serra [16] classifies text, audio, and video features through a fusion model and has achieved good results in the experimental results. Tabuena [17] based on the mapping technology spatially mapped and classified multimodal data, which effectively proves the accuracy of multimodal fusion. Tomašević [18] uses the BOW model to generate file-level lyrics, which are then fused with the extracted audio and video features, and performs semantic classification. In short, multimodal fusion will play an increasingly important role in future research.

The working principle of the emotion model [19] is as follows: first extract emotion words from the emotion library, mainly including some words and sets, and then classify these words and describe the emotions hidden in music according to the classification results. Although this model can better express the semantic information of music, it is not suitable for the extraction of music audio features. Therefore, this model is not suitable for the study of music emotion judgment in this article.

3. Incremental Value Iterative Reinforcement Learning Algorithm Based on Music Education Needs

3.1. Markov Decision Process Reinforcement Learning. Markov decision process reinforcement learning can improve the effect of music information decision-making, so this paper first analyzes the Markov decision process reinforcement learning.

In the discrete-time Bellman optimality equation, since the optimal average return η^{u^*} is unknown, the value iteration algorithm cannot be directly obtained by the formula. Based on the above analysis, it can be seen that the optimal average return η^{u^*} needs to be estimated, and the incremental value iteration algorithm belongs to the direct estimation method. The core idea of the direct estimation method is to directly estimate the optimal average return η^{u^*} in the process of value iteration.

We use $\tilde{\eta}$ to represent the estimated value of the optimal average return η^{u^*} , and B_u and B_l represent the upper and lower bounds of the stay $\tau^u(i)$. Then, for any strategy $\mu \in \prod_s$ [20],

$$1 \leq B_l \leq \tau^u \leq B_u. \quad (1)$$

In the abovementioned formula,

$$r^u = \pi^u \tau^u. \quad (2)$$

It can be seen from the formula that the average return η^u under the semi-Markov decision process can be regarded as the ratio of the average return $\pi^u r^u$ and $\pi^u \tau^u$. The performance function of the embedded chain is set as follows:

$$f^u = \frac{1}{r^u \tau^u}. \quad (3)$$

Combining the abovementioned formula, we can get $\pi^u f^u = \eta^u$, so the embedded chain and the original Markov chain have the same average reward. This can be seen as the entire segment is concentrated on the embedded decision point, so the essence of the time-concentrated Markov decision process is to introduce a reward function into the embedded chain. Since the performance function f is not only related to the state but also related to the actions taken in the state, the performance function f cannot be solved directly. In order to be able to perform the iterative process of updating the strategy on the embedded chain, this topic defines the following performance function [21]:

$$f(i, a) = r(i, a) - \tilde{\eta} \cdot \tau(i, a). \quad (4)$$

The performance potential g^u can be calculated by using the following formula:

$$g^u = \max_{\mu \in \prod_s} \{r^u - \tilde{\eta} \tau^u + P^u g^u\}. \quad (5)$$

Subsequently, the optimal strategy u^* can be obtained using the following strategy update process:

$$\mu^* = \arg \max_{\mu \in \prod_s} \{r^\mu - \tilde{\eta} \tau^\mu + P^\mu g^\mu\}. \quad (6)$$

Through the repeated iteration of formulas (5) and (6), the abovementioned strategy iteration algorithm can be obtained by the optimal strategy under the original semi-Markov decision process. In the iterative process, as the strategy μ approaches the optimal strategy μ^* , the estimated value also continues to approach the optimal average return η^{μ^*} . Therefore, the performance function $f(i,a)$ will also keep changing during the process of strategy iteration.

The optimal strategy ϕ^* of the abovementioned time-focused Markov decision process can be obtained by the following formula [22]:

$$\phi^* = \arg \max_{\mu \in \prod_s} \{\Delta \eta^\mu\}. \quad (7)$$

Among them, there are as follows:

$$\Delta \eta^\mu = \pi^\mu \cdot (r^\mu - \tilde{\eta} \cdot \tau^\mu). \quad (8)$$

$\Delta \eta^\mu$ can be rewritten as follows:

$$\Delta \eta^\mu = (\eta^\mu - \tilde{\eta}) \cdot n^\mu. \quad (9)$$

It can be seen from formula (7) that the optimal strategy p is related to the estimated value i of the optimal average return. The optimal strategy is analyzed below.

Since the strategy ui is the optimal strategy under the original semi-Markov decision process, $\eta^{\mu^*} \geq \eta^{\phi^*}$. According to formula (9), we can obtain the following:

$$\Delta \eta^{\phi^*} = (\eta^{\phi^*} - \tilde{\eta}) \leq (\eta^{\mu^*} - \tilde{\eta}) \cdot \eta^{\phi^*}. \quad (10)$$

In addition, because the strategy ϕ^* is the optimal strategy of the Markov decision process in time concentration, $\Delta \eta^{\phi^*} \geq \Delta \eta^{\mu^*}$. Using formula (9), we can obtain the following:

$$\Delta \eta^{\phi^*} \geq (\eta^{\mu^*} - \tilde{\eta}) \cdot \eta^{\mu^*}. \quad (11)$$

Combining formulas (10) and (11), the following relationship can be obtained:

$$\frac{\Delta \eta^{\phi^*}}{n^{\phi^*}} \leq \eta^{\mu^*} - \tilde{\eta} \leq \frac{\Delta \eta^{\phi^*}}{n^{\mu^*}}. \quad (12)$$

If $\tilde{\eta} \geq \eta^{\mu^*}$, the following inequality can be obtained by formula (10):

$$\Delta \eta^{\phi^*} = (\eta^{\mu^*} - \tilde{\eta}) \cdot n^{\phi^*} \leq 0. \quad (13)$$

According to the abovementioned formula, the following relational formula can be obtained:

$$\Delta \eta^{\phi^*} \leq \tilde{\eta}. \quad (14)$$

If $\tilde{\eta} \leq \eta^{\mu^*}$, combining formulas (9) and (11), we can obtain the following:

$$\Delta \eta^{\phi^*} \geq \eta^{\mu^*} \geq 0. \quad (15)$$

By observing (13) and (15), the following conclusions can be drawn: when $\tilde{\eta} \geq \eta^{\mu^*}$, $\Delta \eta^{\phi^*} \leq 0$ is established, when $\tilde{\eta} \leq \eta^{\mu^*}$, $\Delta \eta^{\phi^*} \geq 0$ is established. In summary, $\Delta \eta^{\phi^*}$ can provide the correct search direction for the optimal average return. Subsequently, the estimated value η^{μ^*} of the optimal average return $\tilde{\eta}$ can be updated iteratively through one-dimensional search, and the process is as follows:

$$\tilde{\eta} \leftarrow \tilde{\eta} + H \cdot \Delta \eta^{\phi^*}, H = \frac{1}{B_\mu}. \quad (16)$$

In the abovementioned formula, H is the search step size of the estimated value $\tilde{\eta}$.

3.2. SMDP Incremental Iterative Reinforcement Learning Algorithm. The extended form of the SMDP incremental value iterative reinforcement learning algorithm is the same as the R-learning algorithm, which is derived from the discrete-time Bellman optimality equation. Therefore, the SMDP incremental value iterative reinforcement learning algorithm's state-action-value function update formula is the same as that of the R-learning algorithm, and the specific process is shown in Algorithm 1.

- (1) Initialize Q (state action to Q-value) expressed as any real number, usually make $Q_0(i, a) = 0, \forall (i, a) \in S \times A$.

The algorithm sets $t_m = t_n = 0$ and specifies $\varepsilon > \sigma > 0$ and initializes $\tilde{\eta}_0$ as any real number.

- (2) The algorithm calculates $Q_{t_m+1}(X_n, A_n)$ for each state $X_n \in S$:

$$Q_{t_m+1}(X_n, A_n) \leftarrow (1 - \alpha) \cdot Q_{t_m}(X_n, A_n) + \alpha \cdot \left(R(X_n, A_n) - \tilde{\eta} \cdot \tau(X_n, A_n) + \max_{a' \in A} Q_{t_m}(X_{n+1}, a') \right). \quad (17)$$

- (3) If $sp(Q_{t_m+1} - Q_{t_m}) \leq \sigma$, the algorithm executes step 4. $sp(Q)$ is the span of Q, as shown below:

$$sp(Q) \equiv \max_{i \in S} \max_{a \in A(i)} Q(i, a) - \min_{i \in S} \max_{a \in A(i)} Q(i, a). \quad (18)$$

Otherwise, the algorithm sets $t_m \leftarrow t_m + 1, n \leftarrow n + 1$ and executes Step 2.

- (4) The algorithm calculates $\Delta \eta^{\phi^*}$:

$$\frac{1}{2} \cdot \left[\max_{i \in S} \left(\max_{a \in A(i)} Q_{t_m+1}(i, a) - \max_{a \in A(i)} Q_{t_m}(i, a) \right) + \min_{i \in S} \left(\max_{a \in A(i)} Q_{t_m+1}(i, a) - \max_{a \in A(i)} Q_{t_m}(i, a) \right) \right]. \quad (19)$$

If $|\Delta\eta^{\phi^*}| \leq \varepsilon$, the algorithm executes Step 5. Otherwise, the algorithm updates $\tilde{\eta}$.

$$\tilde{\eta}_{t_m+1} \leftarrow \tilde{\eta}_{t_m} H \cdot \Delta\eta^{\phi^*}, H = \frac{1}{B_\mu}. \quad (20)$$

The algorithm sets up $t_n \leftarrow t_n + 1$, $t_m \leftarrow t_m + 1$, $n \leftarrow n + 1$ and returns to Step 2.

(5) For any state $i \in S$, the algorithm selects the optimal strategy μ^* through the following formula:

$$\mu^*(i) \in \arg \max_{a \in A(i)} Q_{t_m}(i, a). \quad (21)$$

Inspired by formula (16), since $\Delta\eta^{\phi^*}$ can provide the correct search direction for the optimal average return, we can use the dichotomy to replace formula (16) to estimate the optimal average return directly. The algorithm gives two initial estimates element and element, so that it satisfies the relationship $\tilde{\eta}_1 \leq \eta^{\mu^*} \leq \tilde{\eta}_2$. The algorithm uses $\tilde{\eta} \leq \eta^{\mu^*} \leq \tilde{\eta}_2 = (\tilde{\eta}_1 + \tilde{\eta}_2)$ as the initial estimate of the optimal average return $\Delta\eta^{\mu^*}$. If $\Delta\eta^{\phi^*} > 0$, the algorithm sets $\tilde{\eta}_1 \leftarrow \tilde{\eta}$; otherwise, the algorithm sets $\tilde{\eta}_2 \leftarrow \tilde{\eta}$. Using the abovementioned coin update method instead of formula (20) in algorithm 1, an iterative reinforcement learning algorithm based on the dichotomy of incremental values can be obtained. Due to the introduction of the dichotomy, the IVI reinforcement learning algorithm based on the dichotomy also has a higher optimization efficiency in the case of a large state action space. Algorithm flowchart is shown in Figure 1.

3.3. Reinforcement Learning Curve Analysis. This section gives the Q-value learning curve of the incremental value iterative reinforcement learning algorithm for the semi-Markov decision process and the incremental value iterative reinforcement learning algorithm based on the dichotomy. The simulation experiment results of these two algorithms are shown in Figures 1 and 2, respectively. The learning rate α of the incremental value iterative reinforcement learning algorithm and the incremental value iterative reinforcement learning algorithm based on dichotomy are both $1/(k(X_n, A_n) + 1)$, and the parameters s and α are 10^* and 10, respectively. The search step size of the incremental value iterative reinforcement learning algorithm is $H=0.1$, and the initial optimal average return estimate $\tilde{\eta}_0 = 4$. The initial estimates of the incremental value iterative reinforcement learning algorithm based on dichotomy $\tilde{\eta}_1$ and $\tilde{\eta}_2$ are $\tilde{\eta}_1 = 0$ and $\tilde{\eta}_2 = 6$, respectively, and the initial estimated value of the optimal average return η^{μ^*} is $\tilde{\eta} = 3$.

As can be seen from Figures 2(a) and 2(b), the IVI reinforcement learning algorithm and the dichotomy-based IVI reinforcement learning algorithm both converge within 40,000 iteration steps. Among them, the IVI reinforcement learning algorithm converges faster than the IVI reinforcement learning algorithm based on the dichotomy.

In this paper, the incremental value iteration algorithm is used to directly estimate the optimal average return. In addition, this section uses the SSP value iterative algorithm to directly estimate the optimal average reward, so as to obtain the SMDP random shortest path value iterative reinforcement learning algorithm. The specific process of the random shortest path problem is shown in Figure 3.

As can be seen from the abovementioned figure, the random shortest path problem sets the transition probability $p^{\mu(i)}(i, \xi)$ to a special state ξ to zero, while the transition probability $p^{\mu(i)}(i, \xi)$ in other states remains unchanged. Artificially introduce the absorption termination state t , and set the transition probability of any state i to the termination state t as $p^{\mu(i)}(i, \xi)$. The expected reward function $r_{SSP}^\mu(i)$ of the random shortest path problem under the state $i \in S$ is defined as follows:

$$r_{SSP}^\mu(i) = r^\mu(i) - \tilde{\eta} \cdot \tau^\mu(i). \quad (22)$$

The algorithm assumes that $g_\mu(i)$ represents the expected total reward of the random shortest path problem under strategy μ starting from state i . Subsequently, there is the following relationship:

$$g_\mu(i) - \sum_{j=1}^{\xi-1} p^{\mu(i, j)} \cdot g_\mu(j) = r^\mu(i, \mu(i)) - \tilde{\eta} \cdot \tau^\mu(i, \mu(i)). \quad (23)$$

Both sides of the abovementioned formula are multiplied by $\pi^\mu(i)$ to the left at the same time, and then they are accumulated and summed from $i=1$ to ξ to obtain the following formula:

$$\pi^\mu(i) \cdot g_\mu(i) = \pi^\mu(i) \cdot [r^\mu(i, \mu(i)) - \tilde{\eta} \cdot \tau^\mu(i, \mu(i))] = \Delta\eta^\mu. \quad (24)$$

From the abovementioned formula, the following relationship can be derived:

$$g_\mu(\xi) = \frac{\Delta\eta^\mu}{\pi^\mu(\xi)}. \quad (25)$$

The optimal strategy φ^* for the random shortest path problem can be obtained by the following formula:

$$\varphi^* = \arg \max_{\mu \in \prod_s} g_\mu(\xi). \quad (26)$$

The derivation process of the iterative update formula of the SSP value iterative algorithm coin is similar to the algorithm. Combining (2) and (9), we can see the following:

$$\Delta\eta^\mu = \pi^\mu \tau^\mu \cdot (\eta^\mu - \tilde{\eta}). \quad (27)$$

According to the optimality of the strategy μ^* , we know that $\eta^{\mu^*} \geq \eta^{\varphi^*}$. Combining formulas (25) and (27), we can obtain the following:

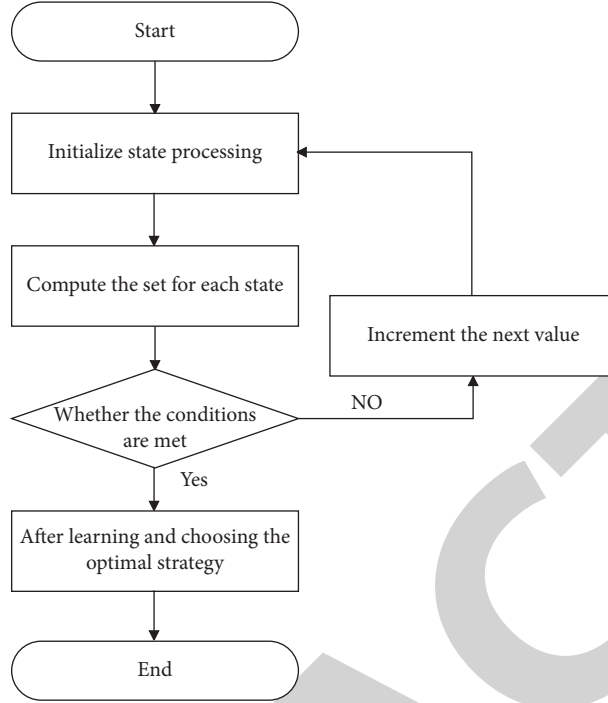


FIGURE 1: Algorithm flowchart.

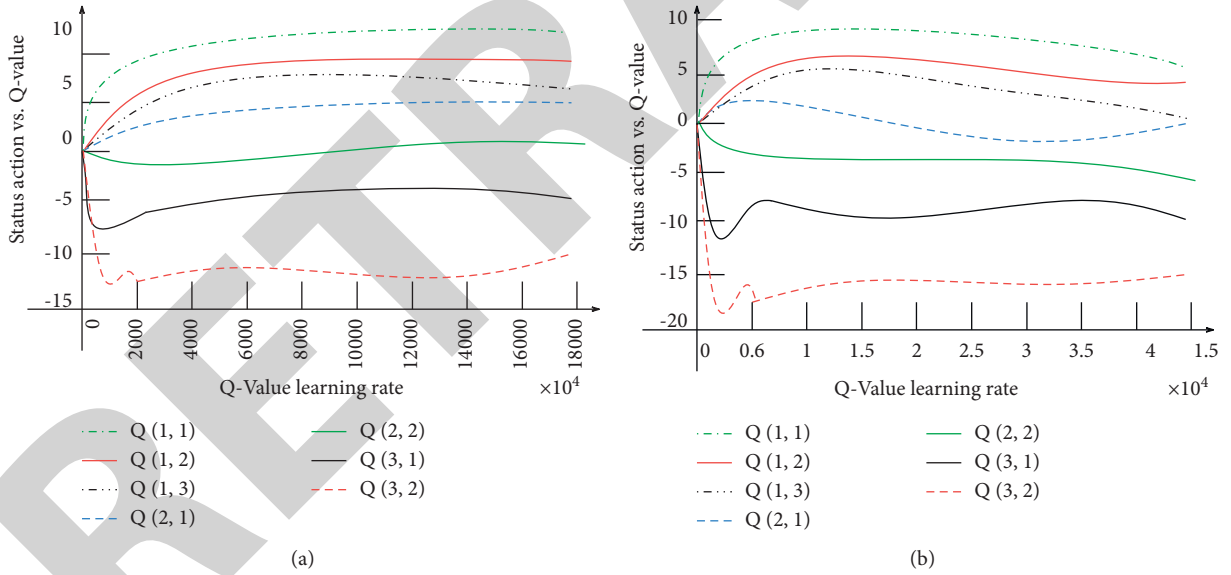


FIGURE 2: Learning curve. (a) Q-value learning curve of IVI reinforcement learning algorithm. (b) Q-value learning curve of IVI reinforcement learning algorithm based on dichotomy.

$$g_{\varphi^*}(\xi) = \frac{\Delta\eta^{\varphi^*}}{\pi^{\varphi^*}(\xi)} = \frac{\pi^{\varphi^*} \tau^{\varphi^*}}{\pi^{\varphi^*}(\xi)} \cdot (\eta^{\varphi^*} - \bar{\eta}) \leq \frac{\pi^{\varphi^*} \tau^{\varphi^*}}{\pi^{\varphi^*}(\xi)} \cdot (\eta^{\varphi^*} - \bar{\eta}). \quad (28)$$

According to the optimality of strategy φ^* , we know that $\Delta\eta^{\varphi^*} \geq \Delta\eta^{\mu^*}$. Similar to the above analysis, we can obtain the following:

$$g_{\varphi^*}(\xi) = \frac{\Delta\eta^{\varphi^*}}{\pi^{\varphi^*}(\xi)} \geq \frac{\Delta\eta^{\mu^*}}{\pi^{\varphi^*}(\xi)} = \frac{\pi^{\varphi^*} \tau^{\varphi^*}}{\pi^{\varphi^*}(\xi)} \cdot (\eta^{\mu^*} - \bar{\eta}). \quad (29)$$

Under the assumption of ergodicity, according to (28) and (29), it can be known that when $\bar{\eta} \geq \eta^{\mu^*}$, $g_{\varphi^*}(\xi) \leq 0$ holds, and when $\bar{\eta} \leq \eta^{\mu^*}$, $g_{\varphi^*}(\xi) \geq 0$ holds. In summary, $g_{\varphi^*}(\xi)$ can provide the correct search direction for the optimal average return η^{μ^*} . Therefore, the estimated value $\bar{\eta}$ of η^{μ^*} can be updated iteratively by the following formula:

$$\bar{\eta} \leftarrow \bar{\eta} + \gamma \cdot g_{\varphi^*}(\xi), \gamma = \frac{1}{B_\mu}. \quad (30)$$

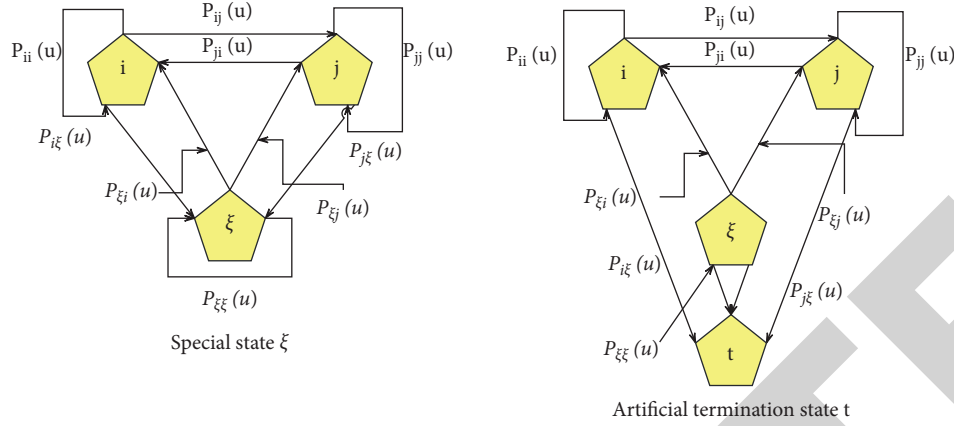


FIGURE 3: Example of a random shortestest path problem.

In the abovementioned formula, γ is the search step length of the estimated value $\tilde{\eta}$. Subsequently, this subject obtained the random shortestest path value iteration reinforcement learning algorithm under the semi-Markov decision process.

- (1) The algorithm initializes the Q table (state action versus Q-value) to any real number, and the algorithm usually sets $Q_0(i,a) = 0, \forall (i,a) \in S \times A$.

The algorithm sets $t_m = t_n = 0$ and specifies $\varepsilon > \sigma > 0$ and initializes $\tilde{\eta}_0$ as any real number.

- (2) For each state $X_n \in S$, the algorithm calculates $Q_{t_m+1}(X_n, A_n)$:

$$Q_{t_m+1}(X_n, A_n) \leftarrow (1 - \alpha) \cdot Q_{t_m}(X_n, A_n) + \alpha \cdot \left(R(X_n, A_n) - \tilde{\eta} \cdot \tau(X_n, A_n) + \max_{a' \in A} Q_{t_m}(X_{n+1}, a') \cdot I\{X_{n+1} \neq \xi\} \right). \quad (31)$$

Among them, $I\{\dots\}$ is the indicator function.

- (3) If $sp(Q_{t_m+1} - Q_{t_m}) \leq \sigma$, the algorithm executes step 4. Otherwise, the algorithm sets $t_m \leftarrow t_m + 1, n \leftarrow n + 1$ and executes step 2.
- (4) The algorithm uses $\max_{a \in A(i)} Q_{t_m+1}(\xi, a)$ to calculate $g_{\varphi^*}(\xi)$. If $|\max_{a \in A(i)} Q_{t_m+1}(\xi, a)| \leq \varepsilon$, the algorithm executes step 5. Otherwise, the algorithm updates $\tilde{\eta}$.

$$\tilde{\eta}_{t_m+1} \leftarrow \tilde{\eta}_{t_m} + \gamma \cdot \max_{a \in A} Q_{t_m+1}, \gamma = \frac{1}{B_\mu}. \quad (32)$$

We set $t_n \leftarrow t_n + 1, t_m \leftarrow t_m + 1, n \leftarrow n + 1$ and return to Step 2.

- (5) For any state $i \in S$, the algorithm selects the optimal strategy μ^* by the following formula.

$$\mu^*(i) \in \arg \max_{a \in A(i)} Q_{t_m}(i, a). \quad (33)$$

Since $g_{\varphi^*}(\xi)$ can provide the correct search direction for the optimal average reward η^{μ^*} , the dichotomy can also be used to replace formula (30) to directly estimate the optimal average reward η^{μ^*} . The algorithm gives initial estimates $\tilde{\eta}_1$ and $\tilde{\eta}_2$, if $g_{\varphi^*}(\xi) > 0$, the algorithm sets $\tilde{\eta}_1 \leftarrow \tilde{\eta}$, otherwise, the algorithm sets $\tilde{\eta}_2 \leftarrow \tilde{\eta}$. Using the abovementioned update

method instead of formula (32) in algorithm 2, an iterative reinforcement learning algorithm based on the dichotomy of SSP values can be obtained.

The Q-value learning curve of the SMDP random shortestest path value iterative reinforcement learning algorithm and the random shortestest path value iterative reinforcement learning algorithm based on the dichotomy are given, respectively. The simulation experiment results of the algorithm are shown in Figures 4 and 5, respectively. The simulation environment of these two algorithms is the same as that of the IVI reinforcement learning algorithm. The learning rate of SSP value iterative reinforcement learning algorithm α , parameters s and o , and the estimated value of the initial optimal average return are the same as the IVI reinforcement learning algorithm, and its search step $\gamma = 0.1$. The initial estimates $\tilde{\eta}_1$ and $\tilde{\eta}_2$ of the SSP value iterative reinforcement learning algorithm based on the dichotomy are the same as those of the IVI reinforcement learning algorithm based on the dichotomy. The special state in the random shortestest path problem is $\xi = 1$.

It can be seen from Figures 4 and 5 that the SSP value iterative reinforcement learning algorithm and the dichotomy-based SSP value iterative reinforcement learning algorithm both converged within 22,000 iteration steps. Among them, the SSP value iterative reinforcement learning algorithm converges faster than the SSP value iterative reinforcement learning algorithm based on the dichotomy.

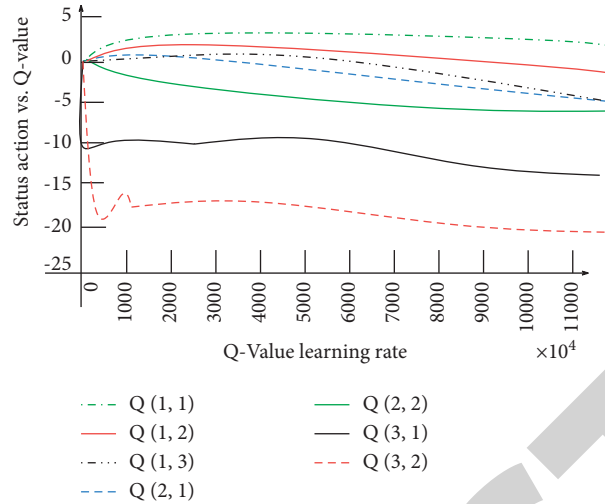


FIGURE 4: Q-value learning curve of SSP value iterative reinforcement learning algorithm.

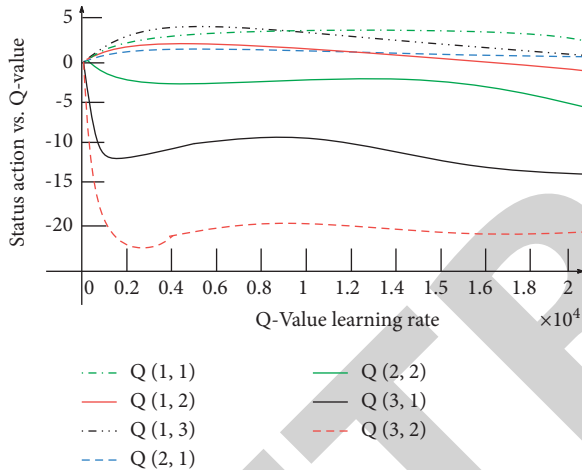


FIGURE 5: Q-value learning curve of SSP value iterative reinforcement learning algorithm based on dichotomy.

The SSP value iterative reinforcement learning algorithm has the fastest convergence speed and the best convergence performance. Compared with other semi-Markov reinforcement learning algorithms in this article, the SSP value iterative reinforcement learning algorithm has the highest efficiency in finding the optimal strategy. By comparing Figures 1 and 4, it can be seen that the SSP value iterative reinforcement learning algorithm has improved the convergence performance of the IVI reinforcement learning algorithm. The simulation results in Figures 4 and 5 verify the convergence of the SSP value iterative reinforcement learning algorithm and the SSP value iterative reinforcement learning algorithm based on the dichotomy. Although the effect of using the dichotomy to directly estimate the optimal average return η^{μ^*} is not as good as the SSP value iterative reinforcement learning algorithm. However, the iterative reinforcement learning algorithm of SSP value based on dichotomy compared with the semi-Markov reinforcement learning algorithm in this paper still has a large degree of improvement in convergence performance.

4. Music Education Innovation Model Based on Information Fusion Reinforcement Learning Model

The multimedia system for studying music teaching needs to meet the following basic requirements. First, it needs to build a rich music resource library, which is the core of the multimedia teaching system. Second, for students with a higher start, the system needs to provide high-quality music works for students to appreciate and learn, which will help improve students' music appreciation level. For students with a low start, the system needs to provide a variety of basic teaching resources to lay a solid foundation for students.

There are video resources, document resources, and picture resources in the music multimedia teaching system. The system needs to support online browsing of different types of music resources. In addition, it needs to provide ways to realize the unified browsing of video resources, document resources, and picture resources.

In addition, the system needs to support intelligent answering. Traditional teaching systems only support real-time interaction between teachers and students. This interaction mode is limited by time and space, and teachers cannot answer students' questions if they are not online. In order to solve this defect, the system introduces intelligent Q&A, which can solve most students' questions and help improve the quality of teaching.

Music resources must have a detailed classification system, which is helpful for subsequent location searches and allows students to quickly find music resources of interest.

Once the music resources stored in the traditional paper mode are damaged, they cannot be restored. However, the resources stored after informatization must have high reliability and be able to deal with database failures under normal circumstances. The basic requirements can be summarized from the above requirements, such as dynamic management of music resources and music resource

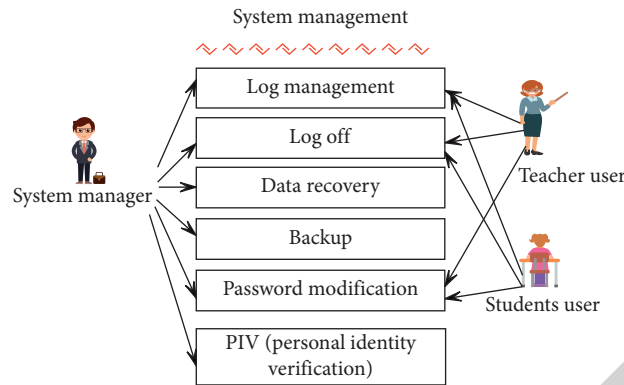


FIGURE 6: Use case diagram of system management module.

classification. At the same time, data backup and data recovery functions need to be introduced to save music resources. The abovementioned analysis is the basis for clarifying the functional requirements of the music multimedia teaching system.

The use case diagram of the system management module is shown in Figure 6.

The user can accurately locate the music resources in the system according to the resource keywords to save time. If the user searches for the resources one by one, a lot of time will be wasted. The use case diagram of the music resource management module is shown in Figure 7.

Homework management is mainly to examine students' mastery of each knowledge point. Teachers can post word-based homework online, and students can upload their answers online to the teacher for correction after completion. The use case diagram of the homework management module is shown in Figure 8.

The purpose of building a music multimedia teaching system in this subject is to realize the sharing of music resources. The physical structure of the system is shown in Figure 9.

The music multimedia teaching system is designed and implemented based on the J2EE platform. Developers can assemble commonly used codes in the system to form reusable codes so that they can be reused. The architecture diagram is shown in Figure 10.

The music multimedia teaching system is composed of five parts: music resource management, system management, homework management, interactive management, and basic data setting. The functional structure diagram is shown in Figure 11.

5. Test and Discussion

MATLAB-based expansion test analysis: the specific implementation process of the experiment is to lock the source of the data on the music platform and collect it based on the target of the task.

On the basis of the abovementioned analysis, the model proposed in this paper is verified, and the practical effects of the music education innovation system based on the

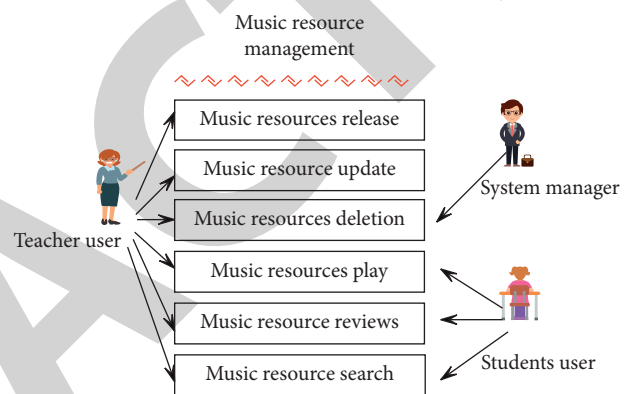


FIGURE 7: Use case diagram of music resource management module.

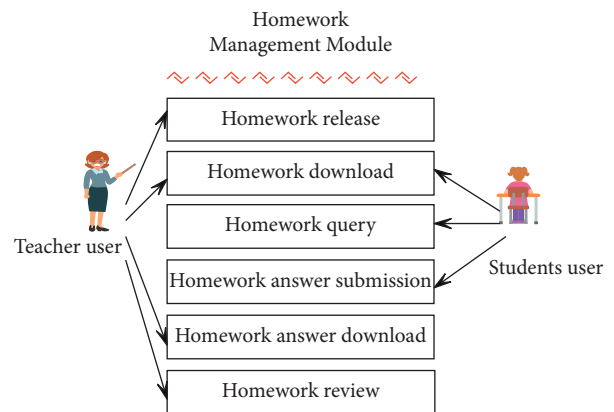


FIGURE 8: Use case diagram of homework management module.

information fusion reinforcement learning model are explored, and the information fusion and teaching effects of the system are, respectively, counted. The method proposed in this paper is compared with the literature in [20]. The results are shown in Tables 1 and 2.

From the abovementioned research, it can be seen that the music education innovation system based on the information fusion reinforcement learning model proposed in this paper has a good music resource information fusion effect and teaching effect.

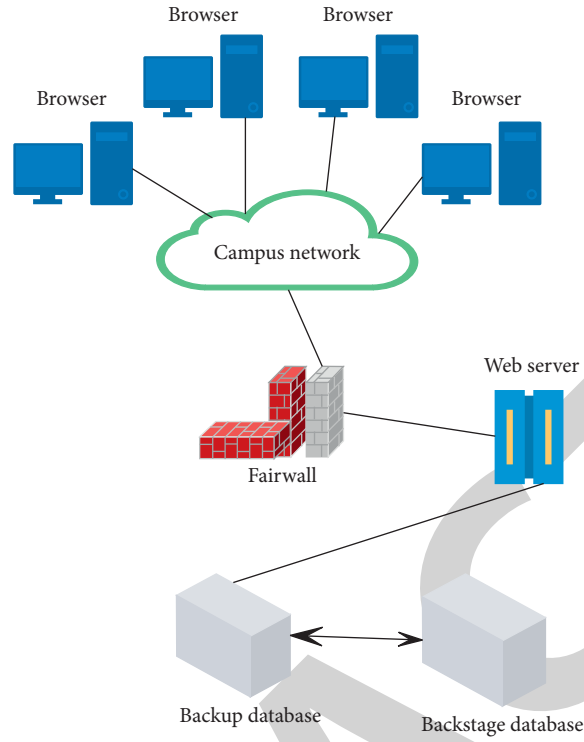


FIGURE 9: System topology diagram.

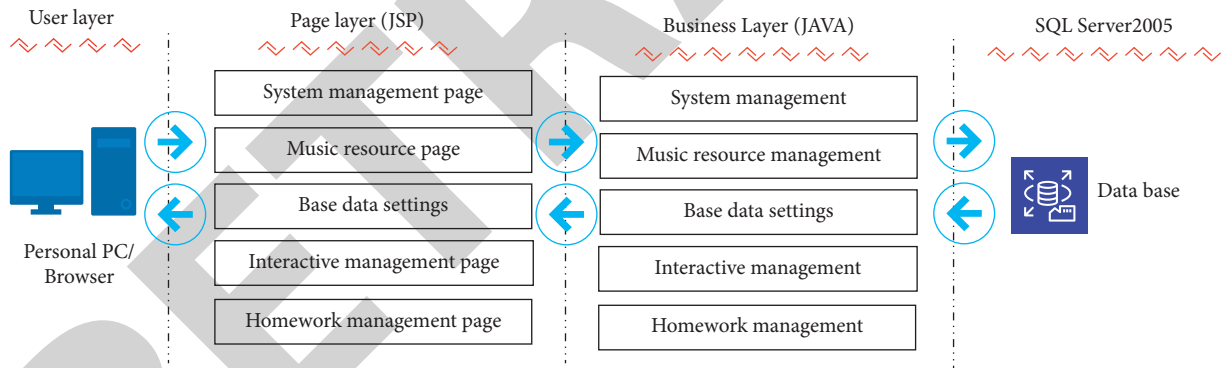


FIGURE 10: Structure of the system.

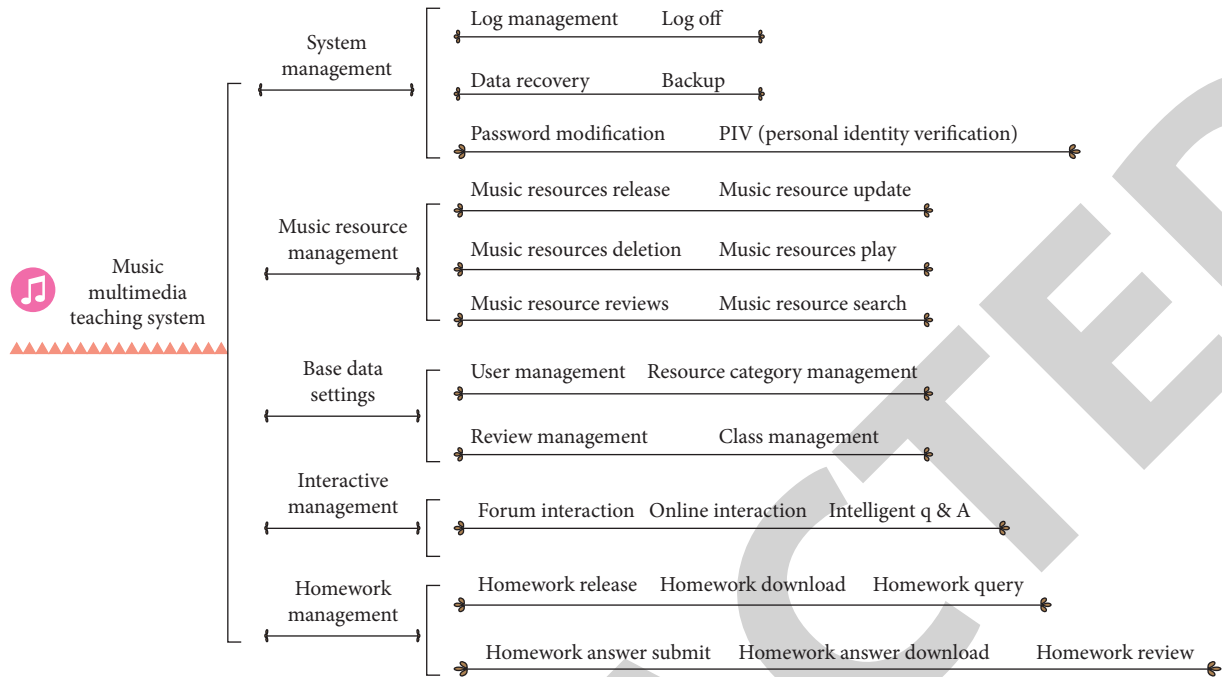


FIGURE 11: Functional structure diagram.

TABLE 1: Evaluation of the effect of music resource information fusion.

Number	The method of this paper	The method of [20]	Number	The method of this paper	The method of [20]
1	91.47	81.91	16	95.70	95.27
2	93.68	81.13	17	91.52	83.11
3	94.46	81.63	18	89.09	86.11
4	89.54	79.00	19	96.91	83.27
5	93.84	84.77	20	89.64	88.15
6	92.41	89.93	21	91.25	89.53
7	92.19	80.55	22	92.31	89.91
8	91.98	87.33	23	91.21	84.25
9	92.72	90.38	24	96.39	95.97
10	88.03	87.57	25	93.63	86.97
11	94.82	84.14	26	92.42	83.82
12	92.94	80.34	27	92.29	87.45
13	96.46	83.99	28	95.40	85.42
14	92.62	86.47	29	91.22	90.57
15	91.15	89.30	30	92.93	84.80

TABLE 2: Evaluation of the effect of music education innovation.

Number	The method of this paper	The method of [20]	Number	The method of this paper	The method of [20]
1	77.74	82.93	16	71.86	82.58
2	89.49	92.67	17	79.80	80.54
3	77.50	82.66	18	91.72	78.29
4	85.37	89.41	19	85.62	90.26
5	72.11	84.10	20	87.92	88.83
6	80.13	86.20	21	76.53	84.52
7	73.44	85.58	22	81.72	88.32
8	82.07	86.25	23	85.94	88.87
9	85.76	79.19	24	90.45	84.80
10	72.06	80.71	25	76.27	84.42
11	88.56	91.30	26	87.21	89.19
12	90.34	85.12	27	82.15	86.11
13	79.23	92.26	28	81.14	88.42
14	72.33	85.06	29	71.37	80.05
15	85.70	85.46	30	71.96	79.25

6. Conclusion

In modern music education, school education has always appeared as a dominant form and has played an important role. In terms of educational consciousness, music education has always been in a subordinate position, and it has often failed to arouse people's attention to the role played by music education's ability to influence. Moreover, music courses have always been marginalized courses or dispensable courses. This is also a kind of partial subject phenomenon caused by the shortcomings of the modern examination system, which has caused no high attention from students to parents to society. In addition, modern music teachers are generally incompletely equipped, which causes modern music education to treat this subject only from the perspective of entertainment. In addition, it has not attracted the attention of the state and society. This paper combines the reinforcement learning model based on information fusion to innovate the way of music education, change the traditional music education model, and improve the actual effect of music education. From the experimental analysis, it can be seen that the music education innovation system based on the information fusion reinforcement learning model proposed in this paper has a good music resource information fusion effect and teaching effect.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interests.

Acknowledgments

This study was sponsored by Nanchang Normal University.

References

- [1] A. Amendola, G. Gabbriellini, P. Dell'Aversana, and A. J. Marini, "Seismic facies analysis through musical attributes," *Geophysical Prospecting*, vol. 65, no. S1, pp. 49–58, 2017.
- [2] J. A. Anaya Amarillas, "Marketing musical: música, industria y promoción en la era digital," *INTERdisciplina*, vol. 9, no. 25, pp. 333–335, 2021.
- [3] B. Babich, "Musical "covers" and the culture industry: from antiquity to the age of digital reproducibility," *Research in Phenomenology*, vol. 48, no. 3, pp. 385–407, 2018.
- [4] F. Calegario, M. M. Wanderley, S. Huot, G. Cabral, and G. Ramalho, "A method and toolkit for digital musical instruments: generating ideas and prototypes," *IEEE Multi-Media*, vol. 24, no. 1, pp. 63–71, 2017.
- [5] E. Cano, D. FitzGerald, A. Liutkus, M. D. Plumbley, and F. R. Stöter, "Musical source separation: an introduction," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 31–40, 2018.
- [6] E. Costa-Giomi and L. Benetti, "Through a baby's ears: musical interactions in a family community," *International Journal of Community Music*, vol. 10, no. 3, pp. 289–303, 2017.
- [7] A. Dickens, C. Greenhalgh, and B. Koleva, "Facilitating accessibility in performance: participatory design for digital musical instruments," *Journal of the Audio Engineering Society*, vol. 66, no. 4, pp. 211–219, 2018.
- [8] L. L. Gonçalves and F. L. Schiavoni, "Creating digital musical instruments with libmosaic-sound and mosaiccode," *Revista de Informática Teórica e Aplicada*, vol. 27, no. 4, pp. 95–107, 2020.
- [9] I. B. Gorbunova, "Music computer technologies in the perspective of digital humanities, arts, and researches," *Opción*, vol. 35, no. 24, pp. 360–375, 2019.
- [10] I. B. Gorbunova and N. N. Petrova, "Music computer technologies, supply chain strategy and transformation processes in socio-cultural paradigm of performing art: using digital button accordion," *International Journal of Supply Chain Management*, vol. 8, no. 6, pp. 436–445, 2019.
- [11] I. B. Gorbunova and N. N. Petrova, "Digital sets of instruments in the system of contemporary artistic education in music: socio-cultural aspect," *Journal of Critical Reviews*, vol. 7, no. 19, pp. 982–989, 2020.
- [12] R. Khulusi, J. Kusnick, C. Meinecke, C. Gillmann, J. Focht, and S. Jänicke, "A survey on visualizations for musical data," in *Computer Graphics Forum*, vol. 39, no. 6, pp. 82–110, 2020.
- [13] T. Magnusson, "The migration of musical instruments: on the socio-technological conditions of musical evolution," *Journal of New Music Research*, vol. 50, no. 2, pp. 175–183, 2021.
- [14] E. Partesotti, A. Peñalba, and J. Manzolli, "Digital instruments and their uses in music therapy," *Nordic Journal of Music Therapy*, vol. 27, no. 5, pp. 399–418, 2021.
- [15] G. Scavone and J. O. Smith, "A landmark article on nonlinear time-domain modeling in musical acoustics," *Journal of the Acoustical Society of America*, vol. 150, no. 2, pp. R3–R4, 2021.
- [16] X. Serra, "The computational study of a musical culture through its digital traces," *Acta Musicologica*, vol. 29, no. 1, pp. 24–44, 2017.
- [17] A. C. Tabuena, "Chord-interval, direct-familiarization, musical instrument digital interface, circle of fifths, and functions as basic piano accompaniment transposition techniques," *International Journal of Research Publications*, vol. 66, no. 1, pp. 1–11, 2020.
- [18] D. Tomašević, S. Wells, I. Y. Ren, A. Volk, and M. Pesek, "Exploring annotations for musical pattern discovery gathered with digital annotation tools," *Journal of Mathematics and Music*, vol. 15, no. 2, pp. 194–207, 2021.
- [19] L. Turchet and M. Barthet, "An ubiquitous smart guitar system for collaborative musical practice," *Journal of New Music Research*, vol. 48, no. 4, pp. 352–365, 2019.
- [20] L. Turchet, T. West, and M. M. Wanderley, "Touching the audience: musical haptic wearables for augmented and participatory live music performances," *Personal and Ubiquitous Computing*, vol. 25, no. 4, pp. 749–769, 2021.
- [21] O. Y. Vereshchahina-Biliavska, O. V. Cherkashyna, Y. O. Moskvichova, O. M. Yakymchuk, and O. V. Lys, "Anthropological view on the history of musical art," *Linguistics and Culture Review*, vol. 5, no. S2, pp. 108–120, 2021.
- [22] L. C. S. Way, "Populism in musical mash ups: recontextualising Brexit," *Social Semiotics*, vol. 31, no. 3, pp. 489–506, 2021.