ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2- 2026

Markov decision process modeling of vocational college students' artistic literacy improvement

Yangyijie Chen[1,*]

[1] The School of Film and Television Art, Hunan Mass Media Vocational and Technical College, Changsha, 410100, Hunan, China

SUMMARY: *In order to solve the problems of rough state identification, static teaching intervention and insufficient feedback utilization in the process of improving students 'artistic literacy in vocational colleges, this paper constructs a Markov decision process model for continuous teaching scenarios. Based on classroom behavior, work performance, process revision and reflection text, this study established the mapping relationship between student state representation, teaching action set and long-term reward function, and realized the optimization of intervention path through strategy learning. Experimental results show that the proposed model is superior to the empirical rule group, SVM group and MLP group in terms of average cumulative return, strategy stability and action coverage. After 12 weeks of teaching, students 'comprehensive artistic literacy score is increased to 76.4 points, which is 14.9 points higher than the initial level. This method provides a computable and iterative teaching decision path for the cultivation of artistic literacy.*

KEYWORDS: *Vocational colleges; Improvement of artistic quality; Markov decision process; Sequential teaching intervention*

1 Introduction

Under the realistic background of vocational education from "single skill training" to "comprehensive quality cultivation", the improvement of artistic quality is no longer just an auxiliary link in curriculum beautification or campus culture construction, but has gradually become an important dimension to measure students 'aesthetic perception, creative expression, emotional understanding and cross-situational communication ability. Different from the traditional art courses that emphasize knowledge imparts and techniques practice, the cultivation of artistic quality in vocational colleges emphasizes the continuous perception, judgment and expression of students in the real learning process, and emphasizes the linkage value between professional learning, personality development and career adaptation. Because of this, art teaching can no longer stay in the linear mode composed of fixed content, uniform rhythm and one-time evaluation, but should be oriented to students' differences, learning fluctuations and feedback loops, and establish a more dynamic and targeted training mechanism.

From the perspective of teaching practice, the development of students 'artistic accomplishment in vocational colleges often has obvious stage and state-dependent characteristics. The intensity of participation, the quality of work completion, the depth of aesthetic understanding and the initiative of creation of students in different learning cycles

*15874117301@163.com
https://doi.org/10.65102/is2026745 Published: 30 April, 2026

are not stable, and the learning experience of the former stage will directly affect the investment level and acceptance mode of the latter stage. If teaching is still organized by static stratification, experience grouping or post-hoc evaluation, it is easy to ignore the potential continuous evolution law in the formation process of artistic literacy, and it is difficult to make a more accurate judgment on "when to intervene, what intervention to use, and whether to improve after intervention". Especially under the condition of continuous expansion of digital teaching environment, classroom behavior records, platform interaction logs, work submission trajectories, reflective texts and stage evaluation results can be continuously collected, which provides a realistic basis for transforming the problem of artistic literacy improvement into a sequential decision-making task that can be calculated, analyzed and optimized.

Markov decision processes provide a modeling framework with explanatory power for this problem. It does not regard students' artistic literacy as a static score after a test, but understands it as a collection of changing states in a series of teaching situations. The measures given by the teacher or the system, such as resource recommendation, task arrangement, creation guidance, and feedback reinforcement, can be regarded as decision actions acting on the state transition. The enhanced engagement, deepening understanding, and improved expression that students present in subsequent performance can be further translated into reward signals that can be used to evaluate the effectiveness of the intervention pathway. Such a processing method is consistent with the characteristics of "observation, judgment, intervention, feedback and readjustment" of the education process itself, and also makes the cultivation of artistic literacy gradually move from experience-driven to data-supported strategy optimization.

It is worth noting that art education is not a training scene with clear standard answers and single paths, and students 'aesthetic experience and creative performance have strong individual differences and situational sensitivity. Therefore, the significance of introducing Markov decision process into the research on the improvement of artistic literacy in vocational colleges is not to replace teachers 'judgment with algorithms, but to sort out the key state variables that affect the development of artistic literacy with the help of computational models, identify the adaptation relationship of different teaching interventions in different student stages, and find a more robust improvement path in multiple rounds of interaction. Compared with traditional statistical analysis, which can only reveal the correlation of factors, this sequential modeling method pays more attention to the continuity of decision-making and long-term benefits in the process, so it is more suitable for dealing with the problems of delay effect, path dependence and individual heterogeneity common in art teaching.

Based on the above understanding, this paper defines the improvement of vocational college students 'artistic literacy as a dynamic decision-making problem oriented to long-term rewards, tries to construct a Markov decision process model composed of student state representation, teaching action set, state transition mechanism and reward function, and verifies its application effect by combining with digital teaching scene design experiments. This study not only focuses on the computational performance of the model in strategy selection, but also focuses on its supporting role in artistic participation, aesthetic understanding and creative performance improvement in real educational contexts. Through this research, this paper hopes to provide an analysis path for art teaching in vocational colleges with both educational interpretation and computational operability, and make the cultivation of art literacy further move from "experience adjustment" to a new stage of "combination of state perception and strategy optimization".

2 Related Research

In recent years, the research on artistic literacy and creative ability in higher education has shifted from the discussion of "whether the course content is rich" to the analysis of "whether students can form aesthetic judgment, creative expression and transfer and application ability in continuous participation". Ji and Chang pointed out that there is a significant correlation between the creativity level and creative performance of higher vocational students, and the autonomy motivation plays a key mediating role in it, which indicates that the improvement of artistic literacy cannot only depend on external indoctrination, but need to mobilize the internal drive of students [1]. Niu and Wu's research in online learning scenarios showed that learning environment, gender differences and behavioral investment would jointly affect creative performance, and the development of artistic ability had obvious context dependence [2]. Lijie et al. further proposed that the critical thinking and cognitive flexibility of learners in the AI era will affect their creative development, which makes the research on artistic literacy begin to connect with the analysis of learning process supported by computing [3]. Setiamurti and Kurniawati found through a systematic review that creativity cultivation in colleges and universities is moving from a single classroom activity to a comprehensive design across stages and tasks [4]. Rae emphasizes that the generation of creativity in higher education cannot be separated from the connected learning relationship, and the interaction network between students and between students and resources will change the quality of creative output [5]. Tam's research on visual arts courses pointed out that without process support in creative thinking teaching, students tend to stay on the surface to imitate and it is difficult to form stable expression ability [6]. Karunarathne and Calma also found in the research on the evaluation of creative thinking that the existing evaluation tools in colleges and universities generally have the problems of single dimension and insufficient dynamic tracking [7]. Putri et al. proposed that the effectiveness of creative teaching method does not completely depend on the teaching material itself, but more on whether the teaching intervention can match the stage state of students [8]. Starting from the long-term application of interdisciplinary artistic literacy methods, Leonido et al. proved that the cultivation of artistic literacy has the characteristics of continuous construction and delayed appearance [9]. These studies collectively show that artistic literacy is not a stable formation of one teaching, but a continuous evolution process influenced by learning state, environmental feedback and intervention methods.

With the advancement of education digitalization, learning analytics, artificial intelligence and adaptive teaching began to provide new technical paths for art education research. de Reizabal and Gomez pointed out that learning analysis technology is able to transform student behavior trajectories in music learning into interpretable data evidence, but multimodal information integration in art education is still insufficient [10]. Fredriksson et al. comprehensive analysis of music education research shows that teaching effect is often influenced by classroom interaction, task organization and reflection mechanism, and a single outcome index is difficult to explain real learning changes [11]. Merchan Sanchez-Jara et al. believe that the core value of AI-assisted music education does not lie in replacing teachers, but in improving the accuracy of state recognition, feedback organization and personalized support [12]. Li and Wang further pointed out that after AI enters music education, the recommendation, diagnosis and learning path generation capabilities are significantly enhanced, but their effective operation relies on reasonable modeling of student states and teaching objectives [13]. Zhang et al. systematic review shows that the current AI-driven music teaching research has gradually shifted from content generation to learning support and strategy optimization [14]. Carvalho et al. research on online teaching of performing arts shows that creative performance under complex teaching situations has strong volatility, and fixed teaching processes are difficult to maintain long-term effects [15]. Wang et al. meta-analysis also found that virtual technology has an overall promotion effect on students' creativity, but the effects of different intervention methods are quite different [16]. It can be seen that the improvement of artistic literacy requires a computational framework that can not only handle continuous feedback, but also optimize teaching actions for long-term benefits.

In this context, reinforcement learning and Markov decision process have gradually entered the field of educational modeling. Fahad Mon et al. pointed out that reinforcement learning is suitable for dealing with sequential intervention problems in educational scenarios, and its advantage lies in its ability to adjust subsequent strategies according to students' historical states [17]. Nickl et al. found that real-time adaptive scaffolding can improve learning performance, and the key lies in whether the system accurately identifies the current stage of the learner [18]. Faber et al. experiment on game-based learning further showed that dynamic scaffolds not only affect performance, but also change students' cognitive load and participation level [19]. Ouhaichi et al. proposed in the study of multimodal learning analysis that the design of education system should consider behavior, emotion and task completion information simultaneously, otherwise it is difficult to support refined intervention [20]. From the perspective of fairness and transparency, Jin et al. reminded that the decision model in the learning analysis system should have interpretability and student-centered characteristics [21]. Nguyen et al. have tried to use contextual bandit to generate learning sequences and proved the practical feasibility of sequential decision model in education recommendation [22]. Chen et al. research on reinforcement learning in heterogeneous environments shows that when learning objects are different, policy learning must consider the inconsistency of potential state distribution [23]. Liao et al. study on average return Markov decision process provided theoretical support for long-term gain-oriented strategy learning [24]. Table 1 summarizes the existing research vein.

Table 1: Main directions and shortcomings of related research

Research Direction	Representative References	Main Focus	Existing Limitations
Artistic Literacy and Creativity Development	[1]–[9]	Motivation, creative thinking, curriculum design, and artistic expression	Most studies focus on influencing factors, with limited dynamic decision-making modeling
Learning Analytics and AI-Assisted Art Education	[10]–[16]	Multimodal data, personalized feedback, and intelligent support	Mostly used for description and recognition, with less attention to long-term strategy optimization
Reinforcement Learning and Sequential Educational Intervention	[17]–[24]	State identification, dynamic scaffolding, and learning path generation	Dedicated research in the context of artistic literacy improvement remains insufficient

Synthesizing the existing results, it can be seen that art education research has fully revealed the multi-factor, stage and context of the formation of artistic literacy, and computational education research has also proved that sequential decision-making model can support personalized intervention. However, there is still no real connection between the two research paths. One kind of research is good at explaining what factors affect artistic literacy, but it is difficult to answer "what teaching actions should be taken in different states". The other kind of research can optimize the learning strategy, but rarely enter the task scenario with aesthetic, open and delayed feedback characteristics such as art literacy. Based on this gap, this paper models the improvement of students 'artistic literacy in vocational colleges as a Markov decision process, and integrates students 'state representation, teaching intervention action, state transition and long-term reward into a unified framework, in order to provide a modeling method with both educational interpretability and computational implementability for the improvement of artistic literacy.

3 Markov Decision process modeling method of vocational college students 'artistic quality improvement

The improvement of students 'artistic literacy in vocational colleges is not suitable to be treated as a static prediction task with one input and one output. Students' aesthetic perception, work understanding, creative input, classroom participation and reflective expression in art courses usually show the characteristics of continuous change, stage fluctuation and path dependence. The effect of a certain round of teaching intervention is often not completely observed at the same time, but will gradually emerge in subsequent exercises, work revision, classroom interaction and cross-task transfer. Because of this, this paper does not understand the improvement of artistic literacy as the fitting of the results of a single evaluation, but models it as a sequential decision-making process over time: the system selects the appropriate teaching action after continuously observing the students 'state, and the environment transitions according to the students 'reaction, and the improvement of artistic literacy is used as the reward signal. The overall modeling idea is shown in Figure 1.

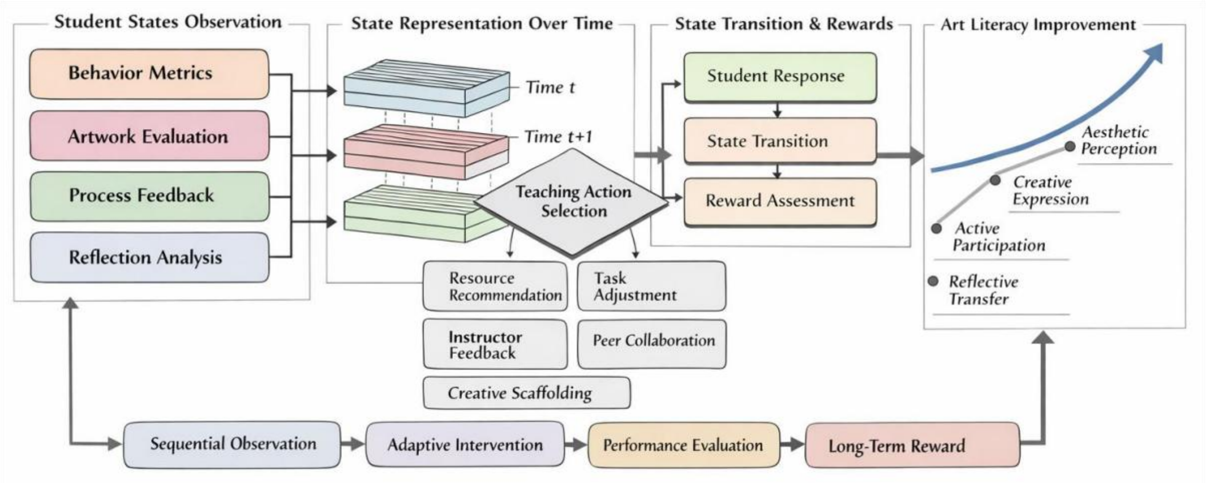

Figure 1: Markov decision process modeling framework for the improvement of students' artistic literacy in vocational colleges

At the formal level, this paper defines the problem as a finite-domain Markov decision process as follows.

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma, \mathrm{T}) \tag{1}$$

where S represents the state space of students' artistic literacy, A represents the action space of teaching intervention, P represents the state transition probability, R represents the reward function, $\gamma \in (0,1)$ is the discount factor, and T is the length of the decision cycle. The key of this definition is not the mathematical form itself, but that it can put "what state of development students are in", "what kind of support methods teachers or systems should adopt" and "whether short-term investment can be transformed into long-term improvement" into the same computational framework for unified analysis.

In order to avoid the simple compression of artistic literacy into a single score, this paper uses multi-source features to jointly represent student status. The student observation vector in the TTH learning cycle is denoted as follows.

$$x_t = [x_t^{(b)}, x_t^{(w)}, x_t^{(p)}, x_t^{(r)}] \tag{2}$$

$x_t^{(b)}$ is the behavior characteristics, including attendance, task completion rate, resource access frequency and interaction frequency. $x_t^{(w)}$ is the feature of the work, including the score of composition integrity, color harmony, rhythm grasp or performance coherence. $x_t^{(p)}$ is the process characteristics, including revision times, feedback response time and phase progress slope. $x_t^{(r)}$ is the reflective feature, which is derived from the semantic coding results of learning logs, text responses and teacher comments. Since the dimensions of different modal data are not consistent with the noise structure, this paper first standardizes the continuous variables:

$$z_{t,j} = \frac{x_{t,j} - \mu_j}{\sigma_j} \tag{3}$$

Here, μ_j and σ_j represent the mean and standard deviation of the JTH feature on the training set, respectively. After standardization, multi-source information such as behavior, works and texts are mapped to a comparable scale space, which provides a basis for subsequent state fusion.

Standardization alone is still not enough to describe the inherent differences in art learning, because the importance of each modality information in different stages is not the same. Some students showed more fluctuations in participation in the early stage, while others were mainly reflected in changes in the expression and reflection depth of the work. To this end, this paper introduces modal-level attention weights to adaptively aggregate different source features. For the m-th mode, the weight is defined as follows.

$$\alpha_t^{(m)} = \frac{\exp\left(u^\top \tanh(W_m z_t^{(m)} + b_m)\right)}{\sum_k \exp\left(u^\top \tanh(W_k z_t^{(k)} + b_k)\right)} \tag{4}$$

On this basis, the fused observation representation of the current time is formed as follows.

$$o_t = \sum_m \alpha_t^{(m)} W_m z_t^{(m)} \tag{5}$$

Considering that the improvement of artistic literacy has significant time continuity, this paper does not directly regard o_t as the final state, but further introduces a temporal recurrence unit to encode the learning process, so as to retain the cumulative impact of early intervention on later performance:

$$s_t = \mathrm{GRU}(o_t, s_{t-1}) \tag{6}$$

The resulting s_t contains both the explicit performance of the current learning cycle and the implicit inertia information in the previous learning trajectory. In other words, the state is no longer an isolated slice but an updatable learning profile. See Figure 2 for the mapping between student states and teaching actions. In the figure, the state space was divided into several intervals such as "insufficient participation type", "unstable expression type", "lagging understanding type" and "comprehensive improvement type", corresponding to different intensities and combinations of teaching intervention strategies.

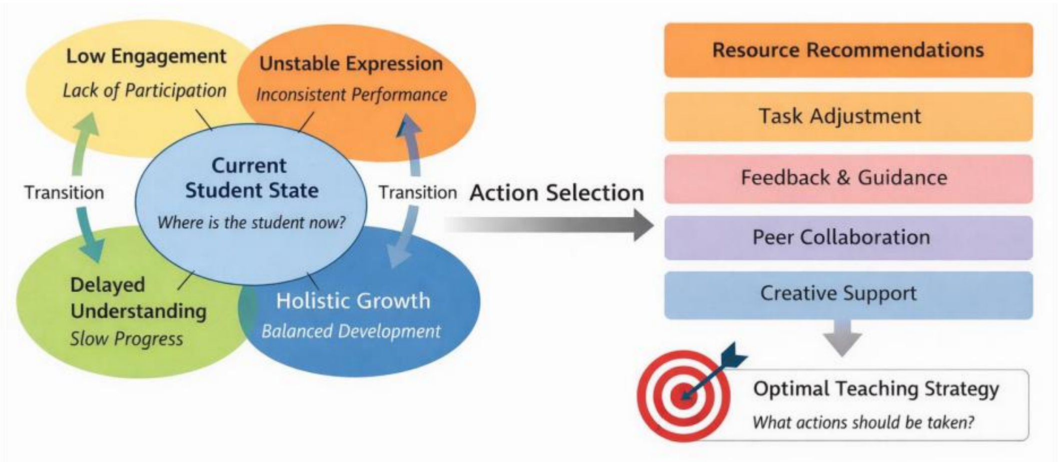

Figure 2: Mapping diagram between student state representation and teaching action

In terms of action space design, this paper does not treat teaching behaviors as abstract symbols, but maps them into executable combinations of teaching interventions. Specifically, action is composed of resource recommendation, task difficulty adjustment, teacher feedback density, peer collaboration mode and creation support strength, which can be expressed as follows.

$$a_t = [a_t^{(res)}, a_t^{(dif)}, a_t^{(fb)}, a_t^{(col)}, a_t^{(sca)}] \tag{7}$$

Among them, $a_t^{(res)}$ represents the resource push strategy, $a_t^{(dif)}$ represents the level of task difficulty, $a_t^{(fb)}$ represents the frequency and granularity of feedback, $a_t^{(col)}$ represents the organization mode of collaborative learning, and $a_t^{(sca)}$ represents the strength of the scaffold. This action design has two considerations: on the one hand, it retains the common real intervention form in art teaching, which is convenient for subsequent application. On the other hand, it can make the strategy learning not only output a single suggestion of "high or low", but also give a multi-dimensional linkage teaching regulation scheme.

In the design of reward function, if only the score of the final work is used as the only goal, the model is easy to bias the short-term visible results, and ignore the quality of students 'participation, the depth of understanding and the ability of continuous improvement. To this end, this paper defines the artistic literacy level as a multi-index weighting function:

$$L_t = \omega_1 q_t^{(a)} + \omega_2 q_t^{(c)} + \omega_3 q_t^{(e)} + \omega_4 q_t^{(m)} \tag{8}$$

Among them, $q_t^{(a)}$ represents the score of aesthetic perception dimension, $q_t^{(c)}$ represents the score of creative expression dimension,$q_t^{(e)}$ represents the score of artistic participation dimension, $q_t^{(m)}$ represents the score of reflection transfer dimension, $\omega_1, \omega_2, \omega_3, \omega_4$ are the weights of each dimension, and $\sum \omega_i = 1$ is satisfied. Based on this, this paper defines the immediate reward as:

$$r_t = \beta_1(L_{t+1} - L_t) + \beta_2(u_{t+1} - u_t) - \beta_3 d_t \tag{9}$$

Here, u_t represents the degree of learning engagement, d_t represents the cognitive load or the cost of teaching friction, and $\beta_1, \beta_2, \beta_3$ are the adjustment coefficients. This design means that the model does not encourage relying on high-pressure tasks in exchange for short-term score increases, but pays more attention to the balance between "literacy improvement, participation enhancement, and controllable burden".

Since the reward in art learning is delayed, and the effect of a particular round of intervention on the quality of the work may not be fully reflected until several weeks later, this paper does not use the single step reward as the only basis for judgment, but uses the long-term cumulative reward:

$$G_t = \sum_{k=0}^{T-t} \gamma^k r_{t+k} \tag{10}$$

where, the discount factor γ is used to balance the short-term and long-term returns. When γ is large, the model will pay more attention to the subsequent state improvement and sustainable development. When γ is small, the model focuses more on instantly visible lifting effects. Considering that art teaching in vocational colleges emphasizes stage growth rather than single competition results, this paper adopts a high discount setting in training, so that the strategy is more inclined to long-term gain. a_t the state transition level, the probability that a student enters a new state s_{t+1} after performing the action at is expressed as follows.

$$\mathcal{P}(s_{t+1} \mid s_t, a_t) = \Pr(S_{t+1} = s_{t+1} \mid S_t = s_t, A_t = a_t) \tag{11}$$

The transition relation is not given by artificial rules, but learned by history teaching samples. In other words, the system does not presuppose that "a certain type of students will inevitably improve under a certain type of intervention", but estimates the effect probability of different interventions on different states of people in the actual data. This not only retains the uncertainty in the education scene, but also enables the model to gradually approach the dynamic evolution mechanism in the real teaching process.

In the strategy solution, if the discrete state transition table is directly constructed, there will be a serious sparsity problem in the face of high-dimensional, multi-modal and continuously changing student state space. Therefore, in this paper, a deep action-value network is used to approximate the Q-function. Given a policy π, the action-value function is defined as follows.

$$Q^\pi(s_t, a_t) = \mathbb{E}_\pi[r_t + \gamma Q^\pi(s_{t+1}, a_{t+1}) \mid s_t, a_t] \tag{12}$$

The optimal policy can be further written as follows.

$$\pi^*(s_t) = \arg\max_{a \in \mathcal{A}} Q^*(s_t, a) \tag{13}$$

This means that in each learning cycle, the system chooses the teaching action that will bring the maximum long-term benefit from the current state, rather than making a local judgment based on the current work performance. Different from the traditional static recommendation, the policy output here will be dynamically updated with the change of state, thus forming a closed loop of "perception - decision - feedback - decision again". In the network training phase, this paper uses the target network to generate the temporal difference target value:

$$y_i = r_i + \gamma \max_{a'} Q\left(s_{i+1}, a'; \theta^-\right) \tag{14}$$

Here, θ^- denotes the target network parameters. The main network parameter θ is updated by minimizing the mean square error loss:

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N}(y_i - Q(s_i, a_i; \theta))^2 + \lambda\|\theta\|_2^2 \tag{15}$$

where λ is the coefficient of the regularization term, which is used to suppress model overfitting. Considering the time correlation between educational data samples, it is easy to cause estimation oscillation if training directly in the original order. To this end, this paper introduces an experience replay buffer in the implementation to scatter samples at different stages, and combines the target network delay update mechanism to reduce training instability. Figure 3 shows the process of policy learning and deployment. Starting from the history teaching data, the figure goes through state encoding, action estimation, reward transmission and parameter update, and the final output can be used for personalized intervention strategies in real classrooms.

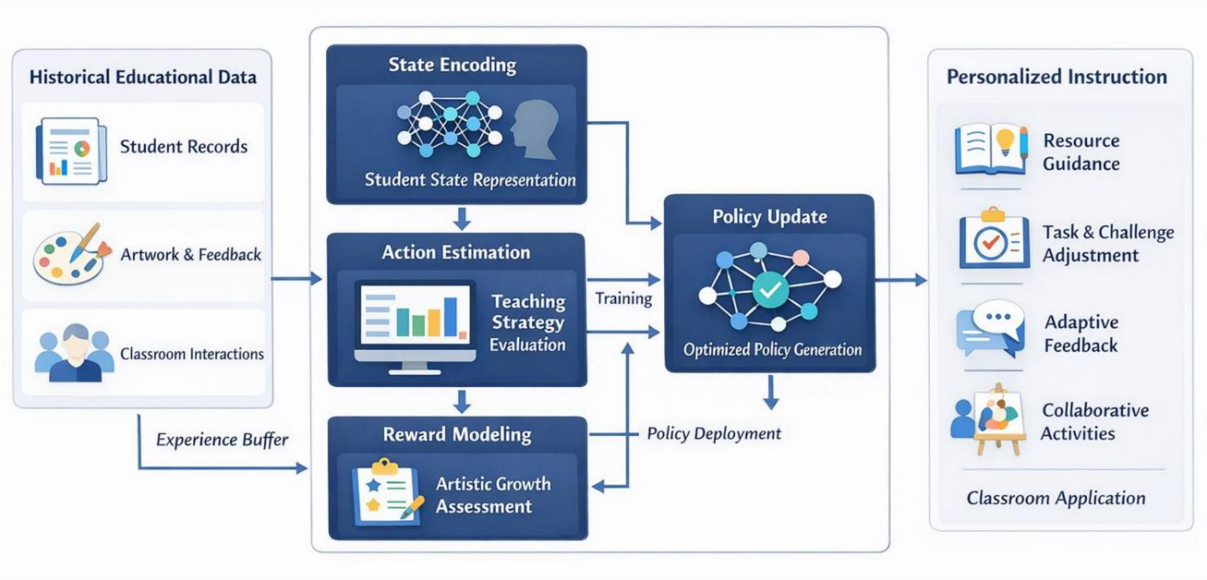

Figure 3: Flow chart of strategy learning and teaching deployment of Markov decision process

Relying only on action-value maximization, the strategy may still be too aggressive. For example, the system continues to assign high-intensity tasks to some middle level students in order to pursue faster score growth, but ignores the emotional stability and expression confidence in art learning. Therefore, in this paper, a policy smoothing constraint is added to the training so that the action changes in adjacent periods are not too drastic, so as to ensure the acceptability of the teaching rhythm. For two consecutive cycles of policy output, the smoothness term is defined as follows.

$$\Omega_t = \|\pi(s_t) - \pi(s_{t-1})\|_2^2 \tag{16}$$

The final optimization goal is as follows:

$$\min_{\theta} \ \mathcal{J}(\theta) = \mathcal{L}(\theta) + \rho \sum_{t} \Omega_t \tag{17}$$

Here, ρ is the smoothness constraint coefficient. This term does not cancel the strategy adjustment, but limits its mutation amplitude in adjacent stages, so that the model output is more in line with the realistic boundaries of the teaching organization of vocational colleges. In this way, the intervention recommendations given by the system are optimized oriented while maintaining the enforceability required by the educational scenario.

It should be pointed out that the Markov decision process model constructed in this paper does not intend to mechanically turn art education into an industrial process of "reward maximization", but tries to provide a computable description framework for the complex and open process of artistic literacy cultivation. Its core value lies in three points. Firstly, the model promotes the formation process of students 'artistic literacy from static result judgment to dynamic state recognition, making "what development range is the current" an analysiable object. Secondly, the model integrates resource recommendation, task adjustment, feedback organization and collaboration arrangement into the action space, which avoids the improvement of art teaching staying at the level of scattered experience. Thirdly, the model deals with the problem of delayed returns in art learning through long-term reward modeling, so that the system not only pursues the improvement of a certain assignment score, but also pays more attention to the overall improvement composed of continuous participation, gradual understanding and stable expression.

4 Experimental design for the application of artistic literacy improvement

In order to test the practicability of the Markov decision process model constructed in this paper in the scene of artistic literacy improvement in vocational colleges, the experimental design does not stay at the offline prediction accuracy level, but puts the model into the continuous intervention chain corresponding to the real teaching process, and forms a closed-loop verification from state recognition, action allocation, stage feedback to strategy update. The experiment focused on the effects of two levels. The first was whether the model could stably learn a better teaching strategy in multiple iterations. The second is whether the strategy can promote the continuous improvement of students 'artistic literacy in actual teaching tasks. Figure 4 shows the overall process of the experiment. In Figure 4, the platform first gathers multi-source data such as classroom behavior, work submission, process revision and reflection text, and then completes state coding and action matching. Then, the personalized teaching action is put into the next round of learning activities, and the strategy parameters are updated according to the feedback results.

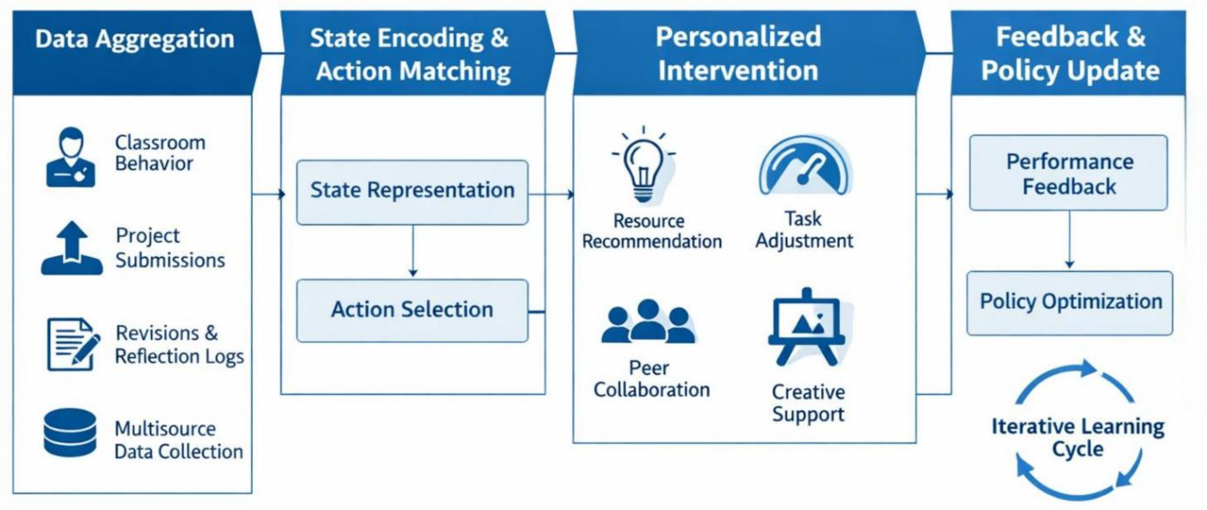

Figure 4: Flow chart of the experiment oriented to the application of artistic literacy improvement

The experimental subjects were 186 students majoring in film and television art, digital media art and music performance in a vocational college, covering two consecutive teaching cycles, a total of 12 weeks. In order to avoid the deviation caused by a single course content, the learning tasks were divided into five categories: work appreciation, imitation and reconstruction, theme creation, collaborative display and reflection expression. Each type of task contained a relatively stable evaluation dimension, which could cover the aesthetic perception, understanding and judgment, creative expression and transfer reflection in artistic literacy. The original data is composed of learning platform logs, classroom check-in records, work rating tables, teachers 'comments and students 'weekly Chronicles. After cleaning, a total of 2232 perimetric particle size samples were formed. Considering that the sequential decision model not only needs to retain the time correlation, but also needs to avoid the information leakage of the test set, this paper uses the data scheme divided by student dimension to divide the sample into training set, validation set and test set, and the proportion is set as follows.

$$N_{train}: N_{val}: N_{test} = 0.72: 0.08: 0.20 \tag{18}$$

This division means that 20% of the students are reserved as independent test objects, and 10% of the remaining samples are selected as the validation set for parameter selection and early stop control. After this processing, the model faces the trajectories of students not involved in training in the testing phase, which better reflects its generalization ability.

In order to make the experiment both conform to the law of art teaching and be able to support computational modeling, each week is considered as a decision step in this paper. In each time step, the system outputs an intervention action according to the student's state vector of the previous week, which is subsequently implemented into resource recommendation, task difficulty adjustment, feedback frequency adjustment, peer collaboration grouping or creation support configuration in this week's teaching. Different from the traditional static hierarchical teaching, this design retains the dynamic relationship of "intervention-response-re-intervention". If the comprehensive level of artistic literacy in week t is denoted as L_t, the weekly improvement between two adjacent time steps is defined as follows.

$$\Delta_t = L_t - L_{t-1} \tag{19}$$

This metric is used to describe the magnitude of immediate improvement of students in continuous learning. Considering the phenomenon that "short-term fluctuations are larger and long-term trends are more important" often exists in art learning, this paper further defines the cumulative improvement index:

$$H = \frac{1}{T}\sum_{t=1}^{T}\frac{\Delta_t}{1+\sigma_\Delta} \tag{20}$$

Here, σ_Δ is the standard deviation of the lifting in each week. The implication of this equation is that with similar average improvement levels, a more stable growth trajectory will be rated higher, thus avoiding the model from creating "false high improvement" through sharp fluctuations in individual weeks.

In terms of experimental grouping, this paper sets up four methods for comparison: empirical rule group, static classification group, supervised prediction group and the proposed model group. In the rule of experience group, teachers manually stratified students according to existing teaching experience, and used a fixed template to implement intervention. The static classification group used support vector machine to identify the categories of student states, and then assigned predefined teaching schemes according to the categories. The supervised prediction group used multi-layer perceptron to predict the next stage of artistic literacy level, and allocated resources according to the predicted value. The model group in this paper outputs the optimal action according to the Markov decision process. The purpose of this setting is not just to compare "who has a higher score", but to investigate whether different methods can maintain effective and stable teaching control ability in the face of continuous state changes. Table 2 shows the experimental environment with the core parameter Settings.

Table 2: Experimental environment and parameter Settings

Item	Configuration
Server Environment	Ubuntu 22.04, Intel Core i7-12700, 32 GB RAM
Development Framework	Python 3.11, PyTorch 2.2, Scikit-learn 1.5
State Encoding Method	Multi-source feature normalization + GRU-based temporal encoding
Decision Cycle	12 weeks, with weekly strategy updates
Batch Size	32
Learning Rate	2×10^{-4}
Optimizer	AdamW
Discount Factor (\gamma)	0.93
Experience Replay Capacity	5000
Early Stopping Patience	5
Baseline Methods	Experience rules, SVM, MLP, MDP

In the training phase, it used the offline historical interaction data to initialize the policy network, and then selected the optimal parameter combination through the validation set. During the experiment, it was found that if the discount factor was set too low, the model was more inclined to choose the action that could push up the score of the work in a short time, which was easy to cause the subsequent participation of some students to decline. If the discount factor is too high, the model will pay too much attention to the long-term returns, resulting in insufficient intervention intensity in the current period. Comprehensive verification results show that when γ=0.93, the performance of the model is more balanced between immediate improvement and long-term profit. In order to measure whether the action assignment truly reflects the personalized characteristics, this paper defines the action coverage index:

$$C = \frac{|\mathcal{A}_{\text{used}}|}{|\mathcal{A}|} \tag{21}$$

Here, $|\mathcal{A}_{\text{used}}|$ represents the number of action types actually invoked during the experiment, and $|\mathcal{A}|$ represents the total number of action libraries. If the value of C is too low, it means that the model reuses a few teaching actions for a long time, and its adaptability is weak. If C is moderate and synchronized with the state change, it indicates that the strategy has good discrimination ability.

In terms of application task design, the experiment did not understand artistic literacy as a single work score, but constructed real teaching situations through multi-task combination. The work appreciation task was used to observe students 'aesthetic recognition and style judgment ability. The imitation reconstruction task was used to investigate its formal understanding and detail reproduction ability. Thematic creation task was used to measure creative expression and structure organization ability. The collaborative display task was used to test the ability of communication cooperation and stage presentation. The reflective expression task is used to capture the level of interpretation, revision and transfer of the creative process. The five categories of tasks together form a capability chain from low to high and from acceptance to generation. In order to form a correspondence between intervention actions and task scenarios, the mapping structure of "state interval -- intervention module -- task carrier" is established in this paper, as shown in Figure 5. In Figure 5, students in different states entered into different intervention paths: the type with insufficient participation emphasized interest activation and low-threshold feedback, the type with unstable expression emphasized scaffold decomposition and stage demonstration, the type with lagging understanding added case comparison and text explanation support, and the type with comprehensive promotion introduced more open creation and cross-group collaboration.

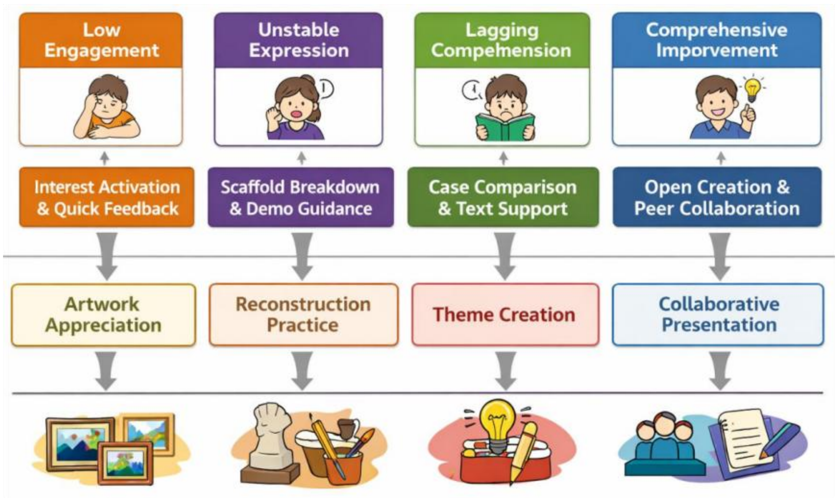

Figure 5: Mapping structure diagram between artistic literacy promotion task and intervention action

The design of the evaluation index also follows the parallel idea of "model performance" and "teaching effect". At the model level, the number of policy convergence rounds, the average reward, the action coverage and the decision stability of the test set were used to measure the learning quality. At the application level, comprehensive artistic literacy score, cumulative improvement index, task completion rate and participation retention rate were used to measure teaching effectiveness. In order to reflect the degree of volatility of the strategy in different weeks, this paper defines the decision stability as:

$$S = 1 - \frac{1}{T-1}\sum_{t=2}^{T}\frac{\|\pi(s_t) - \pi(s_{t-1})\|_1}{Z} \tag{22}$$

Here, Z is the normalization constant. The higher the value of S, the smoother the action changes of the model in the adjacent stages, and the more in line with the requirements of "progressive adjustment" in the teaching organization. The comprehensive score of artistic literacy was weighted by teacher evaluation, peer evaluation and system process index, of which teacher evaluation accounted for 0.5, peer evaluation accounted for 0.2, and process index accounted for 0.3. Such a design can reduce the bias caused by a single source of scoring and make the model effect closer to the comprehensive judgment in the real classroom.

5 Analysis of results

5.1 Analysis of performance results of Markov Decision process model

After the completion of training and validation, the model performance analysis is carried out under the framework of sequential decision making, instead of using single classification accuracy as the only judgment basis. The reason is that the improvement of students 'artistic literacy in vocational colleges has obvious stage continuity. Even if the model can give plausible intervention suggestions at a certain point, if it cannot maintain stable benefits in subsequent rounds, its teaching value is still limited. Based on this understanding, this section analyzes the Markov decision process model built in three levels: average cumulative return, loss convergence trend and test set strategy performance.

From the perspective of the training process, the model in this paper shows a faster revenue climbing speed in the first 10 rounds of updates, and the average cumulative return is increased from 0.22 to 0.67. After entering the 20th round, the return continued to increase but the slope gradually slowed down, and it basically completed convergence around the 31st round, and finally stabilized at about 1.05. In contrast, the average cumulative return of the MLP group in the same round only reached 0.79, the SVM group was 0.60, and the empirical rule group maintained near 0.28 for a long time, indicating that although the static classification or fixed rule can provide a certain degree of intervention reference, it is difficult to gradually revise the strategy in continuous feedback. Figure 6 shows the average cumulative return variation of the four methods over training rounds. It can be seen that the model in this paper still maintains stable gains in the middle and late stages, while the other three groups appear plateform in the early stages, which indicates that the Markov decision process has stronger optimization ability in dealing with the linkage relationship between "current state-subsequent actions-long-term benefits".

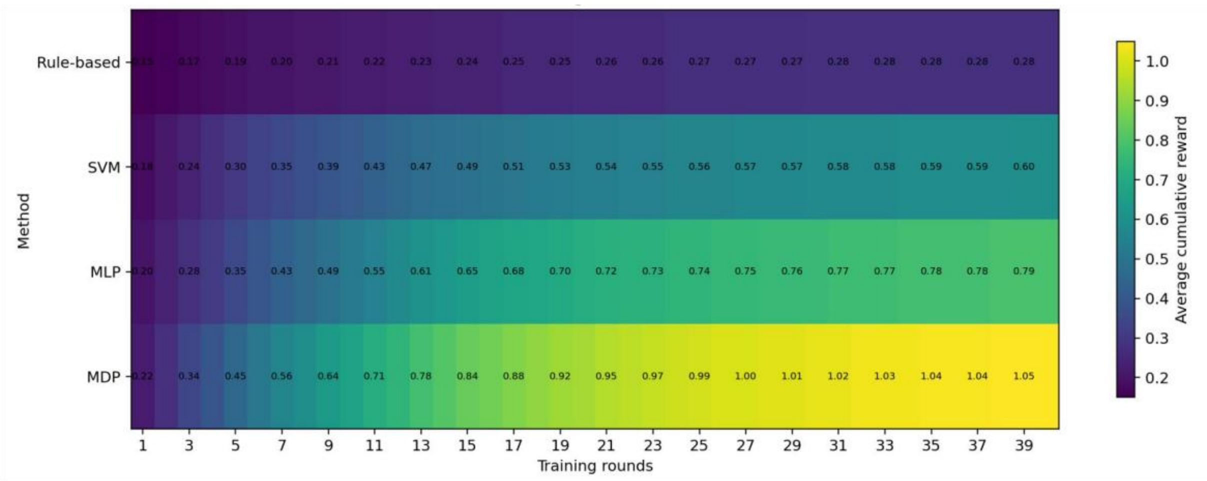

Figure 6: Thermal distribution plots of average cumulative returns for different methods

This advantage is not only reflected in the final number, but also in the way the policy is learned. The empirical rule group essentially relies on the teacher's preset intervention template and lacks real-time response to the state evolution. Therefore, when students change from "insufficient participation" to "unstable expression", the system is difficult to adjust the action synchronously. SVM group can complete the state division, but its output is still static matching driven by category, and it is difficult to deal with the fine-grained differences within the same category. Although supervised prediction is introduced in the MLP group, the optimization objective is mainly directed to the next stage result fitting, and there is no explicit constraint on the long-term returns. In contrast, the proposed model incorporates the potential benefits of several future stages into the estimation at each action selection, so it is more able to avoid local optima such as "short-term score rise and long-term participation decline".

From the perspective of the change of loss function, the training process of Q network is generally stable, and there is no obvious shock. The training loss continuously decreases from 1.24 in the initial stage to 0.17, and the validation loss decreases from 1.27 to 0.31, and maintains small fluctuations in the later stage, indicating that the experience replay and target network update mechanism effectively alleviates the instability problem caused by the high correlation of educational sequence samples. Figure 7 shows the variation trend of training loss and validation loss with the number of rounds. In the figure, the two curves declined rapidly in the first 15 rounds and then entered the slow convergence interval, and the validation curve was always higher than the training curve but did not diverge significantly, which means that although the model had a strong fitting ability, it did not form serious overfitting. Combined with the performance of the validation set, it can be further judged that the current parameter scale and regularization setting are reasonable, which can support the policy network to extract effective state information from multi-source art learning data.

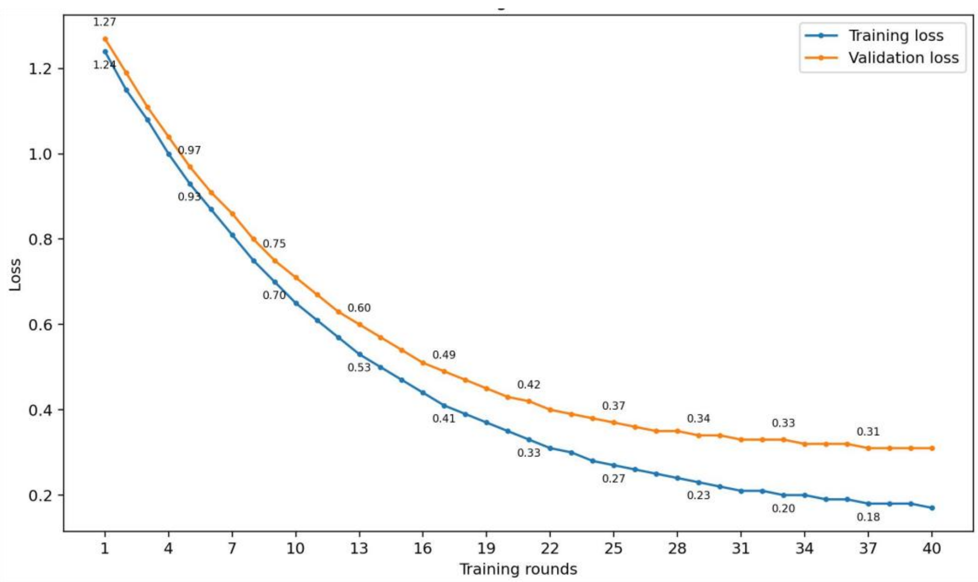

Figure 7: Training loss versus validation loss curve

It is important to note that model stability in educational scenarios cannot be judged by loss values alone. If the strategy output fluctuates too much between adjacent rounds, even if the average return is high, it will bring difficulties to the actual teaching organization. To this end, we further compare the average reward, policy stability and action coverage on the test set. The results show that the strategy stability of the empirical rule group is 0.71, and the action coverage is only 0.34, indicating that its action calls are highly concentrated and almost depend on a few intervention templates for a long time. The strategy stability of the SVM group was increased to 0.78, and the action coverage was 0.52, showing a certain ability to distinguish. The MLP group reaches 0.83 and 0.63, respectively, which can already respond to some state changes. The results of the proposed model on these three indicators are 1.05, 0.91, and 0.81, respectively. It not only has the highest average gain, but also achieves a better balance between the action call breadth and the output smoothness of adjacent stages. See Figure 8 for a related comparison.

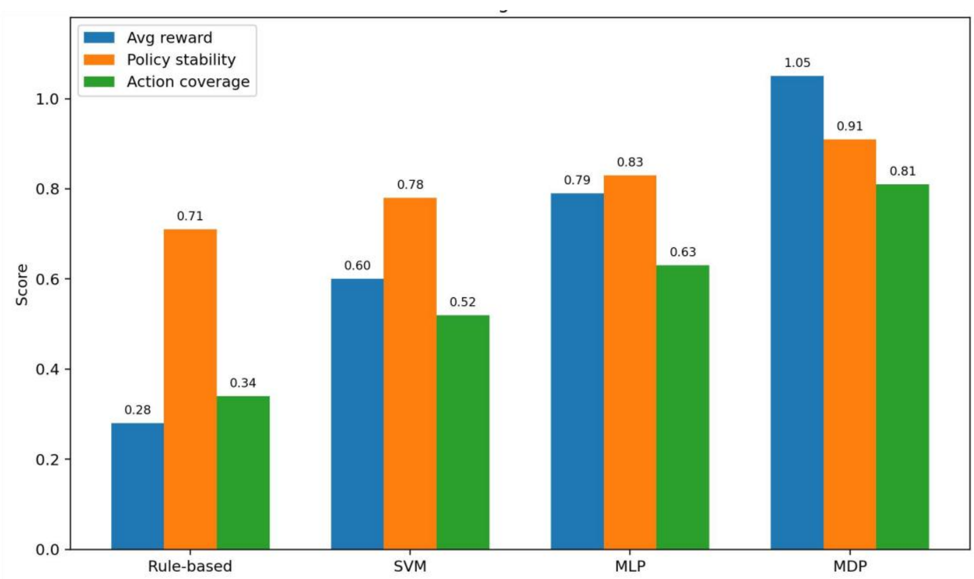

Figure 8: Comparison plot of test set performance metrics

The differences presented in Figure 8 have strong explanatory significance. The high action coverage indicated that the model did not compress all students into a few fixed intervention paths, but could call richer resource recommendation, feedback strength and scaffold combination according to different states. The high stability of the strategy indicates that although the model will adjust the action with the change of the state, this adjustment is not a drastic jump, but maintains the gradual nature required for the implementation of teaching. This is particularly critical for vocational art courses. The improvement of artistic literacy often relies on continuous experience and repeated revision. If the intervention strategy is frequently flipped, students are prone to produce adaptation burden, which will weaken their learning engagement. The model in this paper achieves high stability and high coverage on the test set at the same time, indicating that it not only learns "students in different states should be treated differently", but also learns "complete policy transfer within an acceptable rhythm".

5.2 Analysis of application results of artistic literacy improvement

After the Markov decision process model is deployed to the scene of artistic literacy improvement, its role is no longer limited to one-time hierarchical judgment, but to connect task placement, feedback regulation and stage evaluation into a continuous intervention chain through weekly granularity state update, and then answer the core question of "what kind of support is more appropriate for students in different states". The formation of artistic literacy often relies on continuous experience, repeated revision and gradual expression. If the teaching intervention is only based on static scores, it is easy to cause students with insufficient participation in the early stage to lag behind for a long time, and it is easy to make students with a certain expression ability stay in low-intensity tasks. In the application stage, the intervention actions are constantly corrected through weekly granularity status update, so that the teaching support can change synchronously with the students 'development trajectory.

From the perspective of the overall application effect, the proposed method is superior to the control method in the three aspects of comprehensive artistic literacy score, task completion rate and participation retention rate. Table 3 shows that the overall artistic literacy score of the rule of experience group at the end of 12 weeks was 65.5, which was 4.3 points higher than the starting point. The SVM group increased to 69.1, with an increase of 7.8 points; The MLP group reached 71.5, an increase of 10.1 points; The final score of the model group in this paper was 76.4, which was 14.9 points higher than the initial level. At the same time, the task completion rate of the proposed model group reached 92.6%, and the participation retention rate reached 89.7%, both of which were significantly higher than those of other groups. This shows that the sequential decision framework not only improves the outcome scores, but also improves students 'continuous engagement in the learning process. These kinds of process improvements are often more explanatory than a single review for an art course, because the growth in quality itself is based on steady practice, continuous feedback, and multiple rounds of revision.

Table 3: Comparison of the overall results of different methods in the application of artistic literacy enhancement

Method	Initial Overall Score	Final Overall Score	Improvement	Task Completion Rate / %	Participation Retention Rate / %
Experience Rule Group	61.2	65.5	4.3	78.6	74.2
SVM Group	61.3	69.1	7.8	84.7	80.5
MLP Group	61.4	71.5	10.1	87.9	84.8
Proposed Model Group	61.5	76.4	14.9	92.6	89.7

Further returning to the internal structure of artistic literacy, it can be found that the advantages of the model in this paper are not concentrated in a single dimension, but reflected in the collaborative improvement of multiple dimensions. Table 4 presents the pre-test results of the proposed model group in four aspects: aesthetic perception, creative expression, artistic engagement and reflective transfer. It can be seen that the dimension of art participation has the largest improvement, which increased from 62.6 to 78.9, an increase of 16.3 points. Creative expression increased from 60.4 to 75.8, an increase of 15.4; Aesthetic perception and reflective transfer also improved by 14.1 and 13.8 points, respectively. This result has strong pedagogical implications. In traditional art courses, it is often easier for teachers to directly observe the quality of works, but it is difficult to continuously track students 'participation status and reflection depth. After the behavior log, work revision record and text reflection were incorporated into the state representation of the model in this paper, it was easier for the system to identify students who "completed the task on the surface but had insufficient understanding" or "had high creative intention but unstable expression", and accordingly assigned teaching actions with different support strength, so the improvement was no longer limited to the work results. Rather, it extends to deeper literacy dimensions of engagement, understanding, and transfer.

Table 4: Pre-test results of each dimension of artistic literacy in the model group of this paper

Dimension	Pre-Test Score	Post-Test Score	Improvement
Aesthetic Perception	63.1	77.2	14.1
Creative Expression	60.4	75.8	15.4
Artistic Participation	62.6	78.9	16.3
Reflective Transfer	59.8	73.6	13.8
Overall Level	61.5	76.4	14.9

From the trend of weekly degree change, the comprehensive score curve of the proposed model group still maintained a stable rise in the middle and late period, while the empirical rule group had significantly slowed down after the sixth week. Although the SVM and MLP groups still increased, the slope was weaker than that of the proposed model group. Figure 9 shows the changes in the comprehensive artistic literacy scores of the four groups over the 12 weeks. It can be seen that the model group in this paper gradually opened the gap from the fourth week, and the advantage became more obvious after the eighth week. This phenomenon shows that the value of Markov decision process is not only reflected in "giving advice faster", but also in its ability to gradually accumulate more significant benefits in the later stage through small action adjustments in the early stage. For art teaching, this delayed improvement path is reasonable. Because aesthetic understanding, creative organization, and expressive confidence usually don't jump right after one or two tasks, but develop gradually with continuous feedback, peer collaboration, and multiple rounds of revision.

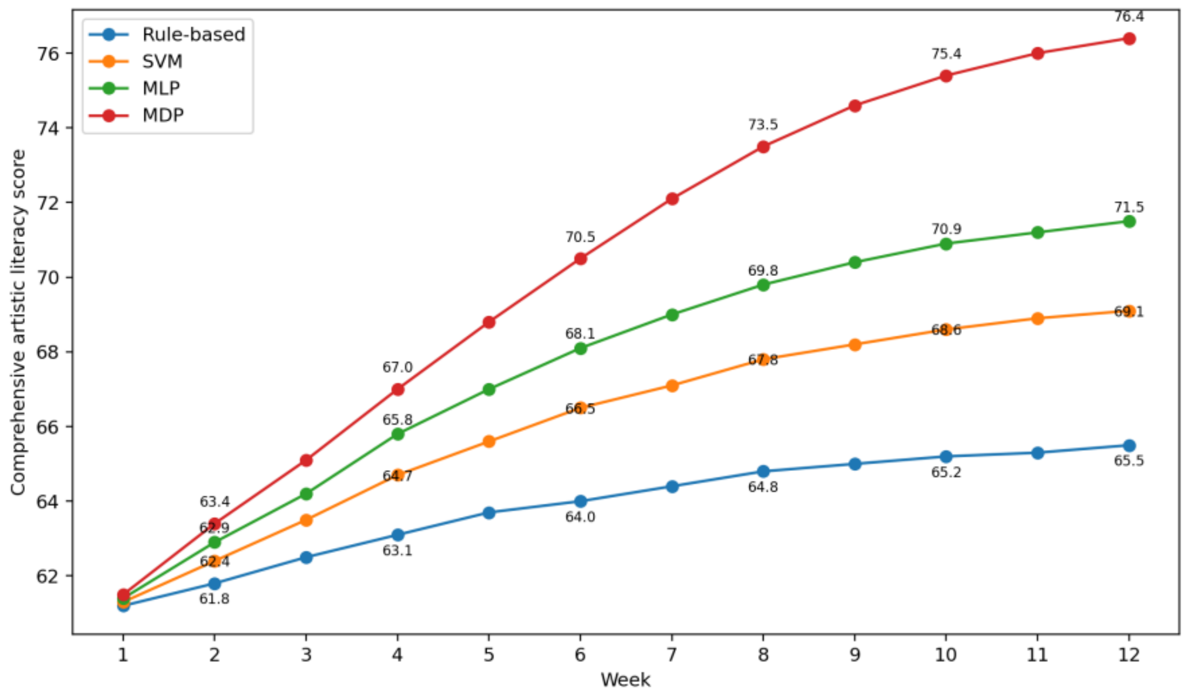

Figure 9: Trend chart of comprehensive artistic literacy score change in 12 weeks

On the whole, the effect of the model in this paper is better than that of static rules and supervised prediction schemes in the application of artistic literacy improvement. The reason is not mysterious. On the one hand, the model treats student development as a continuous state, rather than a fixed label after a grouping. On the other hand, it optimizes the long-term gain rather than the score of a certain task, so it can better balance the relationship between the maintenance of participation, the improvement of expression and the deepening of reflection. This also shows that the introduction of Markov decision process into the cultivation of artistic literacy is not only feasible at the algorithm level, but also has practical value at the level of teaching application. But will evolve with continuous feedback, peer collaboration, and multiple rounds of revision.

6 Conclusions

Based on the real scene of art teaching in vocational colleges, this paper transforms the education problem of improving students 'artistic literacy, which is originally difficult to quantify and has the characteristics of continuous evolution, into a computable Markov decision process modeling task. At the theoretical level, artistic literacy is no longer understood as a static result after a single evaluation, but defined as a dynamic process that is constantly changing in multi-dimensional states such as aesthetic perception, creative expression, learning participation and reflection transfer. At the method level, the closed-loop relationship between state representation, teaching action, reward feedback and strategy update was constructed based on multi-source learning data, so that the teaching intervention could gradually shift from experience driven to sequential optimization supported by data. The experimental results show that the proposed model performs better in terms of strategy convergence, decision stability and action coverage ability, and achieves more significant comprehensive gains in the application of artistic literacy improvement. This indicates that introducing long-term rewards into art teaching decisions not only helps to improve the quality of single task completion, but also promotes the gradual improvement of students in continuous participation, creation organization and aesthetic understanding. The value of this paper is not only to give an algorithm implementation scheme, but also to provide an analysis path for art courses in vocational colleges that can take into account both educational interpretation and computational operability. Through this modeling method, the improvement of students 'artistic literacy in vocational colleges is no longer regarded as a one-time scoring result, but transformed into a continuous decision-making process with perceived states, configurable actions and cumulative returns, so as to form a more stable computational connection between teachers 'judgment, students 'development trajectories and platform data.

Funding

This work was supported by Innovative Research on Ideological and Political Education in Art Vocational Colleges Based on AI, Research Project on Ideological and Political Education in Hunan Province in 2025, Project Number: 25A122.

References

[1] Ji L, Chang Y C. The Impact of Creativity on Creative Performance among University Students in Higher Vocational Education in China: The Mediating Role of Autonomous Motivation[J]. Journal of Pedagogical Research, 2024, 8(2): 296-309.

[2] Niu X, Wu X. Factors influencing vocational college students’ creativity in online learning during the COVID-19 pandemic: The group comparison between male and female[J]. Frontiers in Psychology, 2022, 13: 967890.

[3] Lijie H, Kun T, Leng C H, et al. The mediating effects of critical thinking on the motivation and creativity of Business English learners in the age of AI: Cognitive flexibility theory[J]. Thinking Skills and Creativity, 2024, 53: 101578.

[4] Setiamurti A, Kurniawati F. Fostering creativity in higher education institution: A systematic review (2018–2022)[J]. Open Education Studies, 2024, 6(1): 2022-0221.

[5] Rae J. Connecting for creativity in higher education[J]. Innovative Higher Education, 2023, 48(1): 127-143.

[6] Tam C O. Integrating creative thinking skills pedagogies into a higher education visual arts course[J]. International Journal of Art & Design Education, 2023, 42(1): 16-32.

[7] Karunarathne W, Calma A. Assessing creative thinking skills in higher education: deficits and improvements[J]. Studies in Higher Education, 2024, 49(1): 157-177.

[8] Putri I G A P E, Widdop Quinton H, Selkrig M. Reshaping teaching in higher education through a mandala of creative pedagogies[J]. Teaching in Higher Education, 2024, 29(8): 2075-2094.

[9] Leonido L, Pereira A, Mendes L, et al. Art education: Two decades of creation,

evaluation and application of an interdisciplinary method of artistic literacy[J]. Education Sciences, 2023, 13(6): 589.

[10] de Reizábal M L, Gómez M B. Learning analytics and higher music education: perspectives and challenges[J]. Artseduca, 2023 (34): 219-228.

[11] Fredriksson K, Zandén O, Wallerstedt C. Teaching and learning in music education–a meta-synthesis[J]. Music Education Research, 2024, 26(2): 193-204.

[12] Merchán Sánchez-Jara J F, González Gutiérrez S, Cruz Rodríguez J, et al. Artificial intelligence-assisted music education: A critical synthesis of challenges and opportunities[J]. Education Sciences, 2024, 14(11): 1171.

[13] Li P, Wang B. Artificial intelligence in music education[J]. International Journal of Human–Computer Interaction, 2024, 40(16): 4183-4192.

[14] Zhang Y, Fen B W, Zhang C, et al. Transforming Music Education Through Artificial Intelligence: A Systematic Literature Review on Enhancing Music Teaching and Learning[J]. International Journal of interactive mobile technologies, 2024, 18(18).

[15] Carvalho A, Vieira C, Santos I G, et al. The online teaching of the performing arts in higher education in a context of confinement as a creativity challenge[J]. Media Practice and Education, 2024, 25(4): 326-342.

[16] Wang Y, Liu W, Yu X, et al. The impact of virtual technology on students' creativity: A meta-analysis[J]. Computers & Education, 2024, 215: 105044.

[17] Fahad Mon B, Wasfi A, Hayajneh M, et al. Reinforcement learning in education: A literature review[C]//Informatics. MDPI, 2023, 10(3): 74.

[18] Nickl M, Sommerhoff D, Radkowitsch A, et al. Effects of real-time adaptivity of scaffolding: Supporting pre-service mathematics teachers' assessment skills in simulations[J]. Learning and Instruction, 2024, 94: 101994.

[19] Faber T J E, Dankbaar M E W, van den Broek W W, et al. Effects of adaptive scaffolding on performance, cognitive load and engagement in game-based learning: a randomized controlled trial[J]. BMC Medical Education, 2024, 24(1): 943.

[20] Ouhaichi H, Bahtijar V, Spikol D. Exploring design considerations for multimodal learning analytics systems: An interview study[C]//Frontiers in Education. Frontiers Media SA, 2024, 9: 1356537.

[21] Jin Y, Echeverria V, Yan L, et al. FATE in MMLA: A Student-Centred Exploration of Fairness, Accountability, Transparency, and Ethics in Multimodal Learning Analytics[J]. arXiv preprint arXiv:2402.19071, 2024.

[22] Nguyen L M D, Lin F, Chang M. Generating learning sequences using contextual bandit algorithms[C]//International Conference on Intelligent Tutoring Systems. Cham: Springer Nature Switzerland, 2024: 320-329.

[23] Chen E Y, Song R, Jordan M I. Reinforcement learning in latent heterogeneous

environments[J]. Journal of the American Statistical Association, 2024, 119(548): 3113-3126.

[24] Liao P, Qi Z, Wan R, et al. Batch policy learning in average reward markov decision processes[J]. Annals of statistics, 2022, 50(6): 3364.

[25] Brons A, Wang S, Visser B, et al. Machine learning methods to personalize persuasive strategies in mHealth interventions that promote physical activity: scoping review and categorization overview[J]. Journal of Medical Internet Research, 2024, 26: e47774.