



Utilizing artificial intelligence to enhance ideological and ideopolitical education: a framework for adaptive learning systems

Taohong Zhang^{1,*}

¹ Department of Basic Teaching, Chizhou Polytechnic, chizhou 247000, Anhui, China

SUMMARY: *Ideological performance and the effects of ideological-political education are strongly associated with learners' knowledge status and learning characteristics. This paper proposes an adaptive tracking framework that regards learners as knowledge-distillation objects and uses artificial intelligence technology as an implementation tool. The framework predicts students' cognitive level through teacher-student cooperative extraction. On this basis, it captures prior knowledge from the teacher model and improves the predictive ability of the corresponding student model. Within this framework, the learning resource relationship model is divided into an association model, a sequence model, and a collaborative-filtering recommendation model, which are used to forecast learners' performance on new knowledge points through multiple parameters. A hybrid differential evolutionary algorithm combined with parameter estimation is adopted to construct a personalized recommendation model for Civics learning resources based on relational patterns. The model is applied to the experimental samples' learning process in ideological-political courses. With the support of personalized learning resources recommended by the model, the experimental group's scores increase by 16.83 points, and the highest satisfaction rating reaches 8.1. The results show that the personalized recommendation model for ideology and politics learning resources can work together with the adaptive learning framework, taking teachers and students as dual subjects. It helps optimize and enhance students' ideology through parameter characteristics of different knowledge points and by recommending the most compatible learning resources.*

KEYWORDS: *learning resources; ideology and politics; adaptive tracking; personalized recommendation; ideology; hybrid differential evolution*

1 Introduction

China has long placed great emphasis on ideological-political education and has regarded it as an important approach to consolidating national ideological security and promoting harmonious social development [1]. Ideology forms a key part of national security. It not only embodies the country's cultural soft power, but also serves as a spiritual pillar for maintaining regime stability, promoting social harmony, and guiding national development [2, 3]. Therefore, focusing on ideological-political education is both a long-standing tradition in China and an inevitable requirement for maintaining national security and advancing social progress in the new era. As an important way to spread mainstream ideology and shape social values, ideological-political education guides people to establish a correct worldview, outlook on life, and value system. It also strengthens their sense of identity with and belonging to mainstream ideology through systematic theoretical education and practical activities. At the

*U983867416@163.com

<https://doi.org/10.65102/is2026498>

same time, it plays an important role in identifying and correcting incorrect trends of thought in a timely manner, resisting the infiltration and interference of external ideologies, and safeguarding the purity of the state ideology as well as social stability [4–7]. In the context of the new era, facing a complex and changing international situation and domestic environment, it is necessary to pay closer attention to ideological-political education, continuously strengthen and improve teaching approaches, and enhance educational effectiveness and quality [8–10]. Only in this way can China effectively respond to various challenges and threats, ensure that national ideological security remains stable, and provide ideological support and spiritual strength for realizing the Chinese dream of national rejuvenation.

Artificial intelligence, as a typical representative of the fourth industrial revolution, has deeply changed the ways human beings think, live, and learn [11]. The deep integration of intelligent technology into educational practice has triggered a new wave of innovation in teaching forms. Therefore, the integration of artificial intelligence with ideological-political education should be actively promoted. At the same time, the reform and innovation of higher education should be advanced through the powerful support of artificial intelligence technology [12]. For example, personalized recommendation systems based on machine learning algorithms can effectively build multidimensional and three-dimensional teaching models. Big data analysis technology can also significantly improve the effectiveness of knowledge transfer. Intelligent technology empowerment not only enriches the connotation of the teaching process, but also helps cultivate students' intelligent-technology literacy and data-driven thinking ability [13–15]. Therefore, conducting in-depth research on AI-enabled ideological-political education and exploring its mechanism for enhancing ideology has important theoretical significance and practical value for improving the effectiveness of ideological-political teaching.

In this paper, we utilize the knowledge distillation model to soften the knowledge concepts and construct an adaptive Civics knowledge tracking framework containing a teacher model, a student model and a teacher-student cooperation prediction mechanism. The learning resources are then classified into association mode and sequence mode by mining their relationships, and the two relationship modes are formally defined. The collaborative filtering recommendation algorithm is selected to construct the scoring matrix between learners and recommended resources, calculate the similarity between learners, and obtain the prediction values of learners for each parameter of knowledge points. Sequentially design the differential coding strategy, parameter estimation, evaluation function, learning resource recommendation and differential operation to construct the personalized Civics learning resource recommendation model based on the relational model. Select experimental samples to construct experimental datasets, and for the adaptive tracking framework, determine the data sparsity adaptive processing scheme through different embedded performance data. Demonstrate the effect of knowledge tracking and cognitive level prediction of the framework, and further analyze the correlation between the global importance of knowledge and the degree of influence score. Estimate the relevant parameters of the experimental samples to the important knowledge points by using the personalized Civic and Political Learning Resources Recommendation Model based on the relational model, and construct the network of ideological and political knowledge points. A control model is selected to evaluate the learning resource recommendation performance of the personalized recommendation model in this paper.

2 Adaptive tracking framework based on knowledge distillation

2.1 Definition of the problem

A learner's historical answer record \mathcal{X} can be defined as a sequence consisting of t interactions: $\mathcal{X} = \{x_1, x_2, \dots, x_t\}$, $x_t = (e_t, r_t)$, where e_t denotes the exercise answered by the student at the t th moment, and the binary variable r_t denotes the performance of the learner's answer at the t th exercise as in equation (1):

$$r_t = \begin{cases} 1, & \text{Student correct answer } e_t \\ 0, & \text{Student incorrect answer } e_t \end{cases} \quad (1)$$

The knowledge tracking task can be formalized as a binary sequence prediction task, i.e., given a historical response record \mathcal{X}_{t-1} and a new exercise e_t , the goal of knowledge tracking is to track the changes in the knowledge state of the learner and to predict his or her response to answering the exercise e_t at exercise r_t , which can be expressed as $P(r_t = 1 | \mathcal{X}_{t-1}; e_t)$. Here the probability is approximated by learning a function f_Θ determined by the parameter Θ as in equation (2):

$$\hat{r}_t = f_\Theta(\cdot) \quad (2)$$

where $\hat{r}_t = P(r_t = 1 | \mathcal{X}_{t-1}; e_t)$, and (\cdot) denotes the features used by the model to predict the learner's response to the answer. In this study, concepts are considered as exercises and the knowledge distillation model aims to soften the knowledge concepts to extract more effective prior knowledge.

2.2 Distillation Knowledge Tracking Framework

In the core idea of the Distilled Knowledge Tracing Framework (DistilledKT), the teacher model can employ any state-of-the-art knowledge tracing model and is not constrained by the model architecture. Specifically, the teacher model employs a highly complex architecture to model learners' historical interaction sequences to estimate their mastery of each concept (i.e., logits). DistilledKT then softens the logits using a sigmoid function with temperature τ . The student model uses the same input features as the teacher model and has a shallower network depth and fewer model parameters. The student model loss function consists of both knowledge distillation loss and ternary loss components. Different from the previous mechanism of using the student model to make predictions alone, this paper designs a collaborative teacher-student prediction mechanism to make predictions of learners' cognitive levels from both the teacher model and the student model perspectives. In order to show the details of the framework more clearly, the teacher model and student model are discussed separately in the following.

2.2.1 Teacher model

In the DistilledKT framework, the core goal of the teacher model is to extract valuable prior knowledge from the training set to guide the prediction process of the student model.

This study adjusts the model's parameters by reducing the negative log-likelihood of the

student's response sequence during the teacher network's training phase. Therefore, the loss function based on cross-entropy minimization of the difference between predicted and real labels is examined in this study as follows in Equation (3):

$$\mathcal{L}_{teacher} = -\sum_{t=1}^{|\mathcal{X}|} (r_t \log \hat{r}_t + (1-r_t) \log(1-\hat{r}_t)) \quad (3)$$

where r_t denotes the true label of the t th training sample, \hat{r}_t is a scalar representing the probability that the teacher's model predicts that the learner will correctly answer exercise e_t , and $|\mathcal{X}|$ denotes the sequence of all interactions in the training set.

Next, this paper proposes the $KT_{model}(\cdot)$ function for mapping input sequences to non-normalized predictions. To simplify the representation, the logits of the teacher model (i.e., z^T) are defined in this paper as in Eq. (4):

$$z^T = KT_{model}(\mathcal{X}_t, \mathcal{W}_*) \quad (4)$$

where \mathcal{W}_* denotes the parameters of the training model. Taking the most classical DKT model as an example, z^T is calculated as in equation (5):

$$\begin{aligned} i_t &= \sigma(\mathcal{W}_{xi}x_t + \mathcal{W}_{hi}h_{t-1} + b_i) \\ f_t &= \sigma(\mathcal{W}_{xf}x_t + \mathcal{W}_{hf}h_{t-1} + b_f) \\ c_t &= f_t c_{t-1} + i_t \tanh(\mathcal{W}_{xc}x_t + \mathcal{W}_{hc}h_{t-1} + b_c) \\ o_t &= \sigma(\mathcal{W}_{xo}x_t + \mathcal{W}_{ho}h_{t-1} + b_o) \\ h_t &= o_t \tanh(c_t) \\ z^T &= \mathcal{W}_z o_t + b_z \end{aligned} \quad (5)$$

where i_*, f_*, c_*, o_*, h_* denote input gates, forgetting gates, memory units, output gates, and hidden states, respectively, and $\{\mathcal{W}_*, b_*\}$ denote learnable weights parameters and bias vectors.

Similar to most studies, the last layer of the teacher model is the output activation layer. However, unlike multi-label classification tasks in computer vision, this paper focuses on predicting learners' future performance scores. Therefore, the knowledge tracking task can be formulated as a binary classification problem. Therefore, sigmoid activation function is used in this study instead of softmax. sigmoid activation function is used to map logits to the probability of a concept being answered correctly as in equation (6):

$$P(r_t = 1 | \mathcal{X}_{t-1}; e_t) = \frac{1}{1 + e^{-z^T}} \quad (6)$$

where \mathcal{X}_{t-1} and $r_t = 1$ denote the input data and the corresponding labels, respectively.

Next, a sigmoid activation function for temperature τ is applied to adjust the ‘‘softening degree’’ of the model output to obtain the corrected logits z^T as in Eq. (7):

$$P_{\tau}(r_t = 1 | \mathcal{X}_{t-1}; e_t) = \frac{1}{1 + e^{-\frac{z^T}{\tau}}} \quad (7)$$

where the temperature τ is used to control the scaling of z^T , thus helping the model to obtain a “softened” probability distribution. Although other more advanced temperature functions may be more efficient than Eq. (7), the focus of this study is on the knowledge tracking problem, so a relatively simple scaling form is chosen here.

2.2.2 Student model

The student model aims to improve its capacity to predict learners' present and future cognitive states by integrating the insights obtained from the examination of the instructor model. Specifically, the prediction skills of the latter are anticipated to be improved by creating DT-Loss and TS-Prediction, two methods for integrating the softened previous information into the student model.

(1) DT-Loss

Existing loss functions for knowledge tracking models only focus on the difference between the probability distribution predicted by the model and the true label. In this study, the DT-Loss loss function is proposed, which not only measures the accuracy of the model prediction, but also considers the similarity between samples. Specifically, in this study, the joint knowledge distillation loss (\mathcal{L}_{KD}^S) and the triplet loss ($\mathcal{L}_{triplet}^S$) are used to design the objective function of the student model. The overall loss of the student model can be expressed by Equation (8):

$$\mathcal{L}_{student}^S = \theta \mathcal{L}_{KD}^S + (1 - \theta) \mathcal{L}_{triplet}^S \quad (8)$$

where \mathcal{L}_{KD}^S denotes the loss function of the distillation model, $\mathcal{L}_{triplet}^S$ denotes the ternary loss function, and the weight parameter θ is used to balance the \mathcal{L}_{KD}^S and $\mathcal{L}_{triplet}^S$ importance in the training phase of the model.

The knowledge distillation loss function \mathcal{L}_{KD}^S of the student model under the guidance of the teacher model can be expressed as Eqs. (9)-(11):

$$\mathcal{L}_{KD}^S = \gamma \mathcal{L}_{hard}^S + (1 - \gamma) \mathcal{L}_{soft}^S \quad (9)$$

$$\mathcal{L}_{hard}^S = -\sum_{t=1}^{|\mathcal{X}|} (r_t \log \hat{r}_t^S + (1 - r_t) \log (1 - \hat{r}_t^S)) \quad (10)$$

$$\mathcal{L}_{soft}^S = \sum_{t=1}^{|\mathcal{X}|} D_{KL} \left[\left(\frac{1}{1 + e^{-\frac{z_t^T}{\tau}}} \right) \left\| \left(\frac{1}{1 + e^{-\frac{z_t^S}{\tau}}} \right) \right] \quad (11)$$

where $\frac{1}{1 + e^{-\frac{z_t^T}{\tau}}}$, $\frac{1}{1 + e^{-\frac{z_t^S}{\tau}}}$ denote the soft targets generated by the teacher model and the student model, respectively, and the weighting parameter γ is used to balance the importance of \mathcal{L}_{hard}^S and \mathcal{L}_{soft}^S in the model training phase. Similarly, the logits z^S of the

student model can be calculated according to Equation (12):

$$z^S = KT_{model}(\mathcal{X}_t, \mathcal{V}_*) \quad (12)$$

where \mathcal{V}_* denotes the parameters corresponding to the student model. It is important to note that while the student model and the instructor model have the same backbone design, their parameter settings differ.

Ternary loss is a loss function for learning embedding representations, which is commonly applied in similarity learning tasks. Specifically, consider a ternary constraint (x_a, x_p, x_n) consisting of an anchor sample x_a , a positive sample x_p , and a negative sample x_n , where (x_a, x_p) has the same label and (x_a, x_n) has different labels (i.e., $r_{x_a} = r_{x_p} \neq r_{x_n}$). The ternary loss function is designed to make the anchor samples closer to the positive samples and farther away from the negative samples. The model's generalizability will be enhanced in this way. Equation (13) provides the formula for the ternary loss function:

$$\mathcal{L}_{triplet}^S = \sum_{(a,p,n) \in \Omega} \left[\left\| P_t^T(x_a) - P_t^S(x_p) \right\|_2^2 - \left\| P_t^T(x_a) - P_t^S(x_n) \right\|_2^2 + m \right]_+ \quad (13)$$

where $P_t^T(\cdot)$ and $P_t^S(\cdot)$ denote the outputs of the teacher model and student model, respectively. Where Ω is the indicator set of the sample, $\|\cdot\|_2$ denotes the Euclidean distance between the outputs of the teacher model and the student model, and $[\cdot]_+$ denotes that when the value of Eqn. (13) is greater than 0, $\mathcal{L}_{triplet}^S$ is equal to the value of the loss when Eqn. (13) has a value less than 0, $\mathcal{L}_{triplet}^S$ is equal to 0. Thus, minimization of $\mathcal{L}_{triplet}^S$ enables further differentiation of different knowledge concepts between correct and incorrect answers, thus improving the generalization ability of the model.

(2) Teacher-student cooperative prediction mechanism

The teacher model and the student model were the two perspectives from which the teacher-student cooperative prediction framework was employed in this study. While the student model integrates both the a priori information obtained from the teacher model and ternary loss that takes into account the variances among the predicted labels, the teacher model obtains the a priori knowledge via training the predicted label and real label. As a result, this paper develops a robust prediction function without any additional computational cost. The overall prediction function of the DistilledKT model is represented as in Equation (14):

$$\hat{r}_{tu} = \beta \hat{r}_{tu}^{(1)} + (1 - \beta) \hat{r}_{tu}^{(2)} \quad (14)$$

where $\hat{r}_{tu}^{(1)} = \text{sigmoid}(z^S)$, $\hat{r}_{tu}^{(2)} = \text{sigmoid}(z^T)$, and the hyperparameter β regulates both the instructor model's and the student model's contributions to the output outcomes prediction. It is simple to outperform the teacher model and enable DistilledKT to "beat the blue" by tightly connecting the predictions between the teacher model and the student model.

3 Recommendation of Personalized Civics Learning Resources Based on the Relationship Model

3.1 Relationship Model of Civic Learning Resources

Domain Topic (DT), Structure (S), Learning Resource Type (LRT), Format (F), Interactive Type (IT), and Author (AT) in the Civic Learning Resource Description Ontology are defined as the context information of the learning resource. Each contextual information corresponds to an attribute feature of the Civics learning resource, and the learning resource can be categorized according to that attribute feature.

There are three relationship models to analyze and define the relationship model of learning resources at the context level:

- (1) DomainTopic's association relationship
- (2) Sequential relationship of DomainTopic
- (3) Multiple relationships between DomainTopic and other contextual attributes.

The above three kinds of relationships reveal different teaching-related information, such as aggregation rules between different learning resources, semantic and structural information between learning resources, and learning sequence rules between learning resources according to different learning tasks.

While the link between learning materials and user preferences may be determined from the behavior analysis of the user when interacting with these materials, the study of the relationship between learning materials in Civics is based on the use data supplied by the user.

By mining the relationship of Civics learning resources, two relationship patterns can be found: the association relationship pattern and the sequence relationship pattern.

3.2 Relationship Schema Definitions

Civic Learning Resource Relationship Definition:

Definition 1: For all $x \in L$, the probability of x appearing in a user session S is expressed as equation (15):

$$F(x) = \frac{|\{s | s \in S \wedge x \in L_s\}|}{|S|} \quad (15)$$

For $X = \{x_1, x_2, \dots, x_n\}$, $X \subseteq L$, the probability of the set X appearing in session S is expressed as equation (16):

$$FS(X) = \sum_{i=1}^n F(x_i) \quad (16)$$

For $x, y \in L$, the probability that x and y occur simultaneously in session S is denoted as equation (17):

$$F(x, y) = \frac{|\{s | s \in S \wedge x \in L_s \wedge y \in L_s\}|}{|S|} \quad (17)$$

For $x, y \in L$, the probability of x appearing in front of y in session S is denoted as equation (18):

$$F(x \rightarrow y) = \frac{|\{s \mid s \in S \wedge x \in L_s \wedge y \in L_s \wedge st_x < st_y\}|}{|S|} \quad (18)$$

Definition 2: Let $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_n\}$ for two LO sets, $x, y \in L$. Given X , the association relation between Y and X is defined as equation (19):

$$C(Y|X) = \frac{\sum_{i=1}^n \sum_{j=1}^n F(x_i, y_j)}{\sum_{i=1}^n F(x_i)} \times \text{WeightingFactor}(Y) \quad (19)$$

where $\text{WeightingFactor}(Y)$ is defined as the weighting value, expressed as equation (20):

$$\text{WeightingFactor}(Y) = \frac{\sum_{j=1}^n -\log F(y_j)}{|S|} \quad (20)$$

As described in Definition 1, the association relationship between LO sets X and Y is asymmetric, i.e., $C(Y|X)$ is not equal to $C(X|Y)$. According to the asymmetric *co-occurrence* analysis method, the less frequently occurring lexemes should have a higher priority level, and more effective experimental results can be obtained by using asymmetric weight values and coefficients. Apply this method to the LO relation definition.

Definition 3: Let $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_n\}$ be two LO -sets, and the sequence association of $X, Y \subseteq L$, X and Y is defined as equation (21):

$$C(X \rightarrow Y) = \frac{\sum_{i=1}^n \sum_{j=1}^n F(x_i \rightarrow y_j)}{\sum_{i=1}^n F(x_i)} \times \text{WeightingFactor}(Y) \quad (21)$$

The weighting value $\text{WeightingFactor}(Y)$ here is defined as in definition (1).

Based on the basic correlations defined above, the following definitions of correlation patterns and sequence patterns are given:

3.2.1 Definition of the association model

The association patterns are categorized into basic and preferred associations.

Basic association relations are defined based on $DT \times DT$ two-dimensional space as follows:

Definition 4: Let $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_n\}$ are subsets of L and $X, Y \subseteq LT$, and let θ be the threshold. There is an underlying correlation between X and Y , denoted as $X \Leftrightarrow Y$, defined as equation (22):

$$X \Leftrightarrow Y, \text{ iff } C(Y|X) > \theta \quad (22)$$

Binary preference relations: $DT \times S$, $DT \times LRT$, $DT \times F$, $DT \times IT$, $DT \times AT$

Ternary preference relations: $LT \times \{S, LRT\}$, $LT \times \{S, F\}$, $LT \times \{S, IT\}$, $LT \times \{LRT, F\}$, $LT \times \{LRT, IT\}$, $LT \times \{F, IT\}$.

Quadratic preference relations: $LT \times \{S, LRT, F\}$, $LT \times \{S, LRT, IT\}$, $LT \times \{LRT, R, IT\}$.

Five-element preference relation: $LT \times \{S, LRT, F, IT\}$

Define the ternary $LT \times \{S, LRT\}$ as an example as follows:

Definition 5: Let $X \subseteq LT$, $S = \{s_1, s_2, \dots, s_m\}$, $IT = \{IT_1, IT_2, \dots, IT_m\}$. The preference association relation on $X \times \{S, LRT\}$ is defined as equation (23):

Preference association relations are defined on the space of $LT \times \{othercontexts\}$ and represent combinations of preferences where contextual attributes differ. An example of possible combinations is as follows:

$$Prefer((S, IT) | X) = \left\{ \begin{array}{l} (s_i, it_j) | FS(X \cap s_i \cap it_j) \\ = MAX_{s \in S, it \in IT} FS(X \cap s \cap it) \end{array} \right\} \quad (23)$$

3.2.2 Sequence pattern definitions

Sequence patterns are defined based on $LT \times LT$.

Definition 6: Let $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_n\}$ are subsets of L , $X, Y \subseteq LT$, and let φ be the threshold. There exists a simple sequence relation between X and Y denoted as $X \Rightarrow Y$, defined as equation (24):

$$X \Rightarrow Y, \text{ iff } C(X \rightarrow Y) > \varphi \quad (24)$$

Definition 7: Multiple sequence relations are derived from simple sequence relations.

(1) to-N sequence relations

For $X, M \subseteq LT$, if $M = \{m_i | m_i \subseteq LT \wedge X \Rightarrow m_i\}$, then it is said that there is a 1-to-N sequence X to M relationship.

(2) N-to-1 sequence relation

For $X, M \subseteq LT$, if $M = \{m_i | m_i \subseteq LT \wedge m_i \Rightarrow X\}$, then there exists a 1-to-N sequence relation from X to M . .

In N-to-1 sequences, there are two cases called AND and OR.

1) Let φ be the threshold value

$$\text{N-to-1AND: } M = \{m_i | m_i \subseteq LT \wedge C(m_i \rightarrow X) \geq \varphi\}$$

$$\text{N-to-1OR: } M = \{m_i | m_i \subseteq LT \wedge C(m_i \rightarrow X) < \varphi\}$$

2) Continuous relationship

Let $LP = \{X_1, X_2, \dots, X_m\}$, $X_i \subseteq LT$, if $X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_m$, X_1 to X_m is said to be a continuous sequence relation, while LP is said to be a learned sequence.

3.3 Personalized Recommendation of Civics Learning Resources

3.3.1 Parameter estimation

The user-based collaborative filtering approach of recommendation predicts learner-related parameters such as medium choice, content preference, learning capacity, and hotness of knowledge points.

(1) Construct learner-recommended resource rating matrix

Using equation (25), create the learner's rating matrix for the knowledge points they have acquired.

$$R(m,n) = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1n} \\ \dots & \dots & \dots & \dots \\ R_{m1} & R_{m2} & \dots & R_{mn} \end{bmatrix} \quad (25)$$

R_{mn} denotes the rating obtained by the m th learner when learning the n th knowledge point.

(2) Calculate the similarity between learners

Equation (26) illustrates how learners are sorted according to their degree of similarity using the Pearson's distance as the criteria.

$$Sim(l_1, l_2) = \frac{\sum_{k \in R_{mm}} (R_{l_1 k} - \overline{R_{l_1}})(R_{l_2 k} - \overline{R_{l_2}})}{\sqrt{\sum_{k \in R_{mm}} (R_{l_1 k} - \overline{R_{l_1}})^2} \sqrt{\sum_{k \in R_{mm}} (R_{l_2 k} - \overline{R_{l_2}})^2}} \quad (26)$$

l_1 and l_2 represent learners numbered 1 and 2, respectively, $R_{l_1 k}$ represents the rating value obtained by the l_1 th learner when learning the k th knowledge point, and $\overline{R_{l_1}}$ and $\overline{R_{l_2}}$ represent the mean value of the ratings obtained by learners l_1 and l_2 when learning all the knowledge points.

(3) Estimation of learners' parameter values for new knowledge points

After obtaining the similar learner set l_u , the learners' ratings $P_{l_1, m}$ for unlearned knowledge points are predicted. If several similar learners' ratings for the same knowledge point can be obtained from the neighbor set, the rating given by the learner who has the greatest similarity to the current learner is adopted, as shown in equation (27).

$$P_{l_1, m} = \overline{R_{l_1}} + \frac{\sum_{l_u \in N} (Sim(l_1, l_u)(R_{l_u m} - \overline{R_{l_u}}))}{\sum_{l_u \in N} Sim(l_1, l_u)} \quad (27)$$

where $\overline{R_{l_1}}$ represents the mean of the ratings obtained by the learners, $Sim(l_1, l_2)$ is the similarity between the learners, and $R_{l_2 m}$ is the rating obtained by learner l_2 when learning the m th knowledge point.

3.3.2 Personalized Recommendation Algorithm Process

Figure 1 illustrates the process of creating a customized recommendation system for civics education materials.

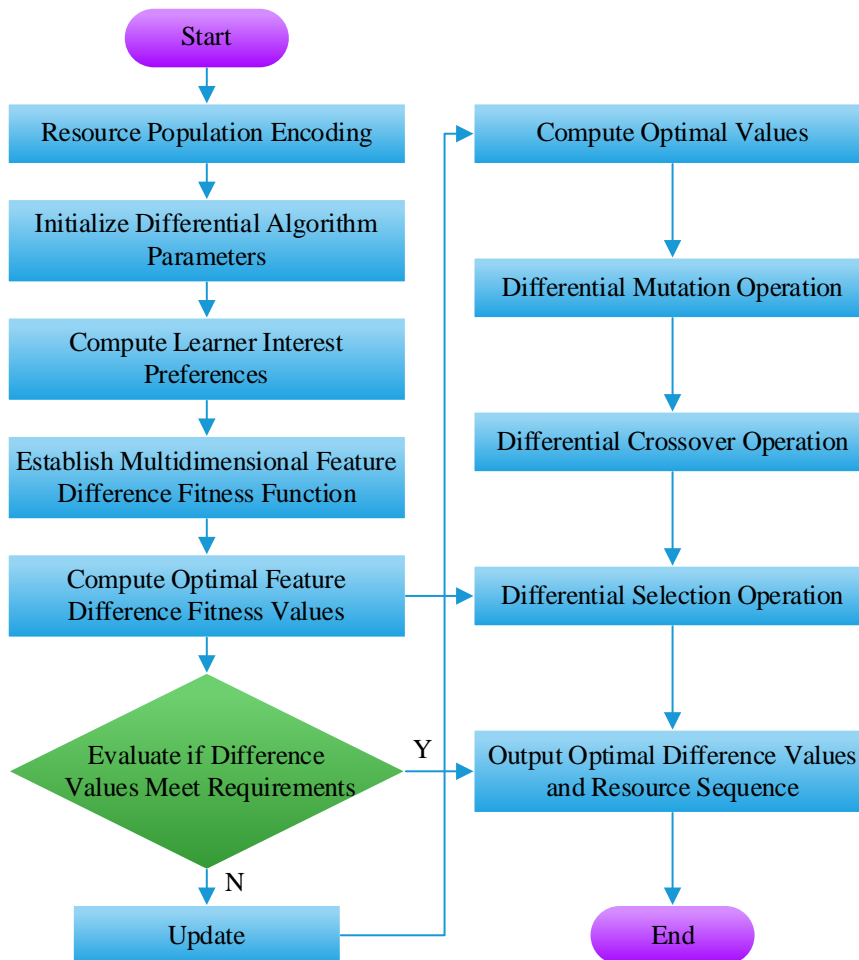


Figure 1: Algorithm flow

(1) Differential coding strategy

1) Learning resource population coding

Learning resources are encoded using a binary approach, defining $X = \{X_1, X_2, \dots, X_n\}$ as a sequence of learning resources as in equation (28).

$$X_n = \begin{cases} 1 & rand > 0.5 \\ 0 & rand < 0.5 \end{cases} \quad (28)$$

Rand is a random value, if the random value is greater than 0.5, X_n is equal to 1 and vice versa.

2) Initialization of the difference algorithm

The scaling factor, iteration times, and fitness-function accuracy of the differential algorithm are initialized.

(2) Evaluate the difference recommended learning resources

1) Evaluate the feature difference value

According to the population value obtained after updating, the fitness function is introduced to recalculate the characteristic difference value. Then, whether the recalculated characteristic difference value satisfies the requirements is judged. If it does not meet the requirements, this step is repeated until the condition is satisfied.

2) Recommended learning resource sequence

The minimum feature difference value is identified, and the learning resource sequence corresponding to this minimum value is recommended.

(3) Differential variation operation

The mutation strategy of the differential algorithm varies according to different situations. A common strategy is to randomly select several individuals from the population, synthesize their vector differences, and then mutate the target individuals after scaling. To better fit the characteristics of the discrete differential algorithm and accelerate the evolution of the population toward the optimal value, this study improves the standard differential evolution strategy. Specifically, two indicators are considered at the same time: the optimal solution of the group and the optimal solution of the individual. This makes the mutation operation of the differential algorithm more targeted. By using a mapping function, the search process is converted from continuous intervals to the interval (0,1). The corresponding expression is shown in equation (29):

$$\begin{aligned}
 t_{i,j}^{k+1} &= x_{i,j}^k + F(p_{ij} - x_{i,j}^k) + F(g_{ij} - x_{i,j}^k) \\
 T(t) &= \left| \frac{2}{\pi} \arctan(t_{i,j}^{k+1}) \right| \\
 V_{ij} &= \begin{cases} 1 & T(t) > \text{random} \\ 0 & T(t) < \text{random} \end{cases}
 \end{aligned} \tag{29}$$

where F is a vector scaling factor and $T(t)$ is a mapping function that maps real-valued functions to (0,1). p_{ij} is the individual optimal value and g_{ij} is the group optimal value.

(4) Differential crossover operation

Based on the initialized upper and lower limits of crossover probability, the crossover probability is dynamically adjusted according to the population iteration process. Then, crossover operations are performed on each population according to the crossover probability value. At the early stage of the algorithm, the crossover variance rate of the population is relatively large. Therefore, the population has a higher probability of crossover variation, which helps increase population diversity. At the later stage of the algorithm, the crossover variation rate becomes smaller. As a result, the population has a lower probability of crossover variation, which accelerates population convergence. In this way, the next generation of the population differs from the previous one while maintaining the diversity of the algorithm population, as shown in equation (30):

$$\begin{aligned}
 CR &= ((1-t/T)(CR_{\max} - CR_{\min}) + CR_{\min}) \\
 U_n &= \begin{cases} V_{ij}^k & \text{rand} > CR \\ 0 & \text{otherwise} \end{cases}
 \end{aligned} \tag{30}$$

where CR denotes the crossover probability value, CR_{\max} represents the maximum crossover probability, and CR_{\min} refers to the minimum crossover probability. T indicates the total number of iterations, while t denotes the current iteration number.

(5) Differential selection operation

After the initialized population generates a new sub-population through mutation and crossover, the fitness function value of the new population is compared with that of the previous generation. The individual with the better fitness value is retained in the next

generation, as expressed in equation (31).

$$X_n = \begin{cases} V_{ij}^k & f(U_i) \leq f(X_i) \\ X_i & \text{otherwise} \end{cases} \quad (31)$$

4 Adaptive-based Ideological and Political Learning Effectiveness Enhancement

A total of 240 students in the second year of college within the same major of a university are selected as experimental samples, to which test papers containing a total of 60 questions are distributed, and the answer sheet data are collected and combed to form an experimental dataset. The application effect analysis of adaptive tracking framework based on knowledge distillation and the feasibility verification of personalized civic learning resources recommendation based on relational model are carried out sequentially.

4.1 Optimization and application of an adaptive tracking framework

4.1.1 Sparsity Adaptive Processing

The experimental dataset is divided into sub-datasets with sparsity of 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400 and 1500 according to the shortest sequence length. The performance of the adaptive tracking framework based on scalar embedding and vector embedding is tested on all the sub-datasets respectively is shown in Fig. 2.

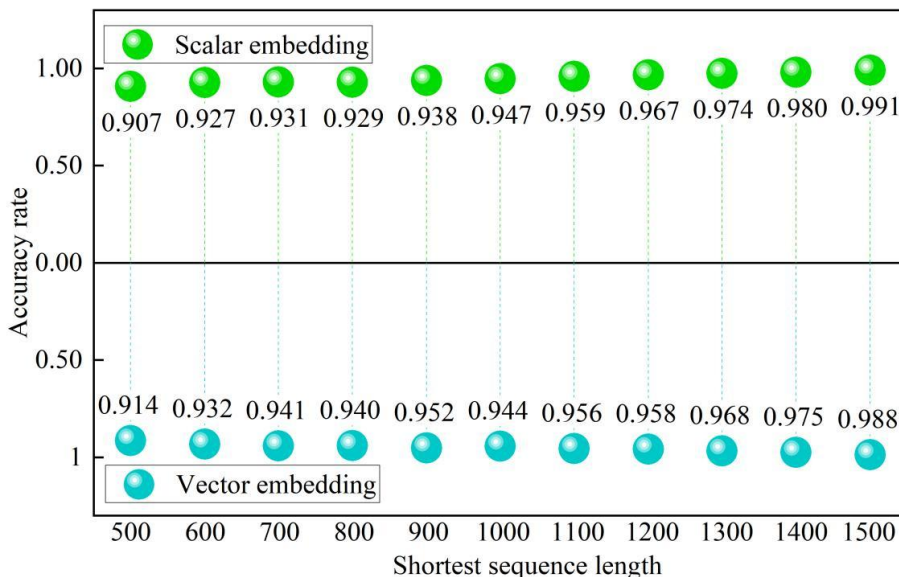


Figure 2: A comparative experiment between scalar embedding and vector embedding

The performance of scalar embedding and vector embedding is negatively correlated with the degree of data sparsity, and with the shortest sequence length of 1000 as the cutoff point, when the data sparsity is less than 1000, the vector embedding performance (0.914~0.952) outperforms the scalar embedding performance (0.907~0.938), whereas when the data sparsity is greater than or equal to 1000, the scalar embedding performance (0.947~0.9991) outperforms vector embedding (0.944~0.988) performance. Accordingly the adaptive tracking framework based on knowledge distillation uses scalar embedding if the data sparsity is large

and vector embedding if the weak data sparsity is small when dealing with learner data.

4.1.2 Visualization of knowledge tracing

A sample of 25 randomly chosen civics practice questions (1–25) was used in the experiment; each practice question had five knowledge points (B1–B5). The (0,1) interval is used to calculate the likelihood that the learner will properly answer the question. The model developed in this research and its knowledge tracking findings are visualized in Figure 3 below. From question number 13 forward, we can observe that the student has a high level of comprehension of B2 Civics knowledge points inside the (0.8,1.0) interval. As for the B1, B3-B5 knowledge points, it shows a clear rising pattern, indicating that the student has improved his knowledge point mastery after continuous knowledge point review and training during answering the practice questions. It can be seen that the adaptive tracking framework in this paper is able to track students' mastery of each knowledge point in detail, which assists in recommending personalized Civics learning resources and learning guidance in practice.

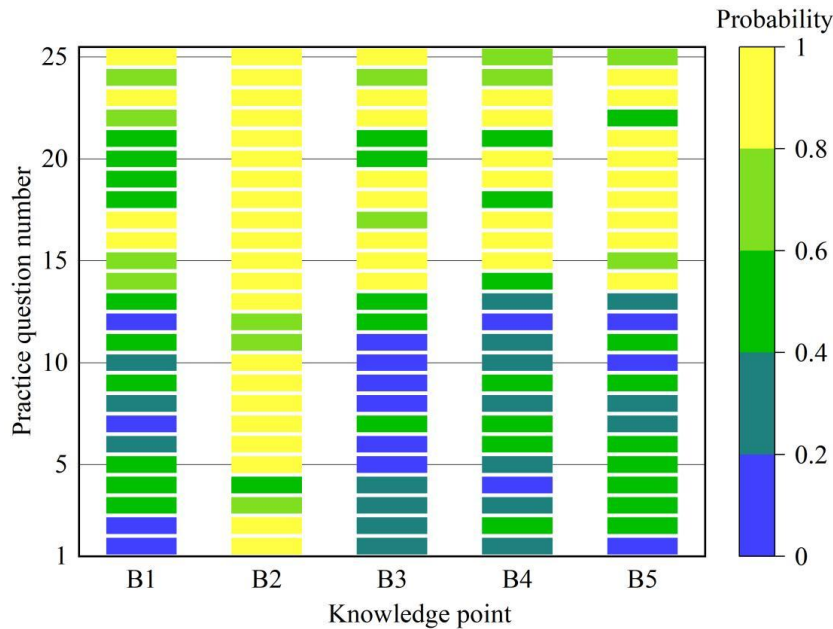


Figure 3: Knowledge tracking results

4.1.3 Cognitive level

The mastery level of the 240 experimental samples (001-240) is summarized in Table 1. Even for the same learner, the mastery level of different learning points is susceptible to a large gap due to the different focus and content of different learning points. For example, Student 010's mastery of knowledge points B1-B4 is above 0.7000, while the mastery of knowledge point B5 is only 0.2026.

Table 1: The degree of mastery of sample knowledge points

Sample	B1	B2	B3	B4	B5
001	0.9317	0.4569	0.0178	0.8272	0.622
002	0.3658	0.5219	0.1866	0.5623	0.3455
003	0.2314	0.6893	0.8934	0.6778	0.4601
004	0.3364	0.5623	0.9912	0.0462	0.3096
005	0.8335	0.3382	0.1302	0.4411	0.7501
006	0.3452	0.6066	0.5581	0.7045	0.188
007	0.6172	0.2259	0.1027	0.6916	0.4821
008	0.4001	0.0372	0.8389	0.7484	0.5954
009	0.5753	0.6305	0.1293	0.4043	0.4018
010	0.7158	0.8273	0.7934	0.8212	0.2026
...
240	0.6061	0.4779	0.1379	0.8328	0.5938

4.1.4 Knowledge Global Importance and Impact Score Levels

Taking the subjective questions numbered 6-10 on the experimental dataset as an example, the relationship between the global importance of knowledge acquired based on the adaptive tracking framework and the degree of influence on the score is shown in Figure 4. Taking topic 6 as an example, answering this topic requires a total of five knowledge points, C1, C8, C3, C10 and C12, and the correlation coefficients between the level of knowledge mastery and the actual scores of the five knowledge points are, in order, 0.5120, 0.5190, 0.5453, 0.5651 and 0.6120, with the higher correlation coefficients indicating that the higher the degree of the knowledge point's influence on the students' scores on the topic. The more important the knowledge, the greater the correlation coefficient between the level of mastery and the actual score, a trend that is consistent with general cognitive laws and further justifies the adaptive tracking framework.

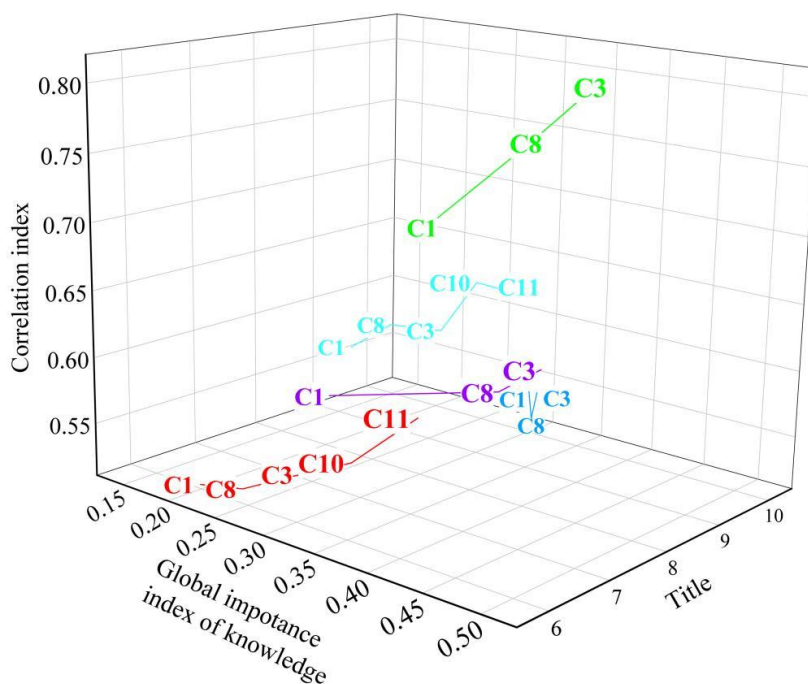


Figure 4: Correlation coefficient performance

4.2 Operational Performance of Personalized Civics Learning Resources Recommendation Model

4.2.1 Parameter estimation for learners

A randomly selected sample of students, based on this paper's recommended model to calculate their content preference, learning ability and learning enthusiasm for a total of 10 important ideological and political knowledge points C1-C10, a total of three parameters are shown in Table 2. For the convenience of analysis, the three parameters are set to take the value of the interval for the (0,10], the higher the value is, the stronger the degree. Taking knowledge point C1 as an example, students' content preference is 4 and learning ability is 3, which are both at a low level, so the enthusiasm for learning is only 2. For knowledge point C5, the content preference is 8 and learning ability is 9, so the enthusiasm for learning reaches 10.

Table 2: User attention calculation based on the forgetting function

Knowledge point	Content preference	Learning ability	Learning enthusiasm
C1	4	3	2
C2	5	6	6
C3	7	8	7
C4	6	6	6
C5	8	9	10
C6	7	8	7
C7	8	7	8
C8	2	3	2
C9	5	5	6
C10	6	6	5

4.2.2 Construction of a network of knowledge points

The entire structure of the knowledge point network is depicted in Fig. 5, which is based on the recommendation model of customized civics learning resources and the total of 50 crucial ideological and political knowledge points found in the test question pool in the experimental sample. The nodes of the network are knowledge point labels, and the connected edge weights are initialized to the knowledge point similarity matrix, and the connection between knowledge points can be preliminarily probed from the overall structure of the network and local features. First, from the knowledge point network, it can be observed that the knowledge points with high importance (dark-colored nodes, C1-C10) are located in the central area of the knowledge point network, in the following order: (C1) moral norms, (C2) cultural intermingling, (C3) social functions, (C4) value concepts, (C5) people's masses, (C6) laws, (C7) productivity, (C8) order, (C9) globalization, (C10) cultural confidence, and the connections between these more important knowledge points are also stronger. Secondly, the knowledge points with relatively low importance (light-colored nodes, C11-C50) are in the fringe area of the network, and their distribution is also more dispersed.

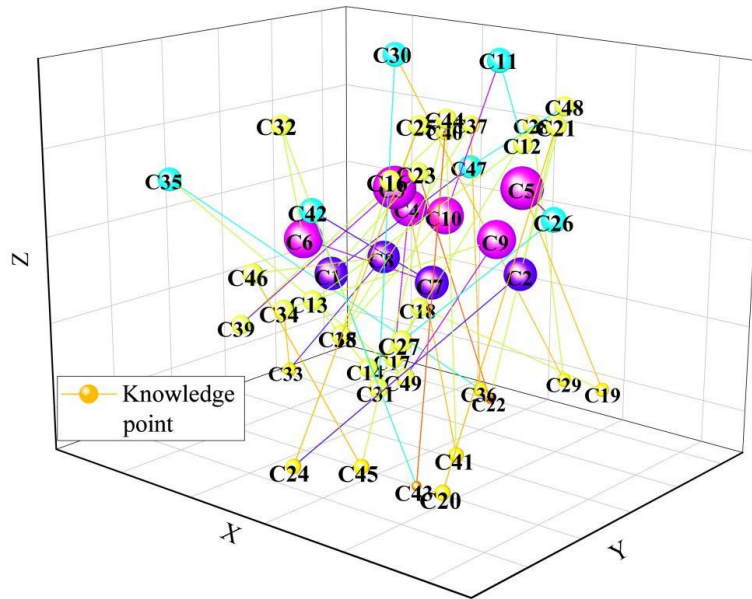


Figure 5: The network structure of some knowledge points

4.2.3 Recommendations for Civics Learning Resources

The 240 samples were divided into 8 groups, each containing 30 students each. A total of seven personalized recommendation models (D1) User-based CF, (D2) SVD, (D3) AutoRec, (D4) deepFM, (D5) Node2vec, (D6) NR, (D7) VL2v are selected as the control, and (D8) this paper's model to expand the ideological and political learning resources recommendation of students in the independent groups, respectively, as shown in Table 3. Ideological and political performance and student satisfaction scores (out of 10 points) reflect the model's learning resources recommendation effect, the experimental cycle is a semester.

Table 3: The effect of the learning resource recommendation model

Model	Experimental stage and score			Satisfaction score		
	Before	After	Δ	First	Second	Third
D1	73.19(6.6)	78.57(5.76)	5.38	6.6	6.2	6
D2	74.28(8.7)	78.16(7.16)	3.88	6	5.8	5.4
D3	70.32(7.4)	75.38(6.26)	5.06	7	6.8	6.4
D4	70.17(6.3)	75.09(5.56)	4.92	6.8	6.6	6
D5	78.87(8.1)	83.07(7.36)	4.2	6	5.8	5.6
D6	74.27(6.4)	80.47(5.56)	6.2	6.4	6.2	6
D7	73.23(7.5)	82.62(6.76)	9.39	7.0	6.6	6.2
D8	72.98(6.7)	89.81(5.06)	16.83	8.1	7.9	7.8

Before the experiment the performance of the students in the groups corresponding to each model was in the (70,79) range with no significant difference. After the model assistance, each group of students' performance and the group gap have different degrees of improvement. Among these, the model in this research helps students achieve a mean improvement in their ideological and political scores of up to 16.83, which is much higher than those of the other seven control models (3.88~9.39). Additionally, (D8) shows the real performance of the customized ideological and political resources suggestion model, with the maximum student satisfaction score using this model in this article being 8.1, which is also the highest of the eight models.

5 Conclusion

This paper proposes a knowledge distillation-driven adaptive knowledge tracking method for predicting students' ideological and political cognitive level, which integrates scalar embedding and vector embedding to adaptively handle sparse data, has the prediction logic of positive correlation between the global importance of knowledge and the degree of impact scores, and is capable of tracking the mastery and cognitive level of students on different knowledge points in detail. The learning resources relationship model is divided into association mode and sequence mode, combined with collaborative filtering recommendation algorithm to form a personalized Civic and Political Learning Resources Recommendation Model, and the model is used to construct a network of ideological and political knowledge points based on the experimental sample dataset. Students assisted by the personalized ideological and political learning resources recommendation model in this paper not only improved their ideological and political scores by 16.83 points compared with the preexperimental scores, but also gave the highest satisfaction score of 8.1.

In the field of curriculum education and construction on ideology and politics, the educator must catch up with the progress and apply modern technologies, including artificial intelligence technology, to enhance teaching efficiency. Through the adaptive tracking framework, learners' ideological dynamics, learning characteristics and learning demands are grasped, and personalized ideological and political learning resources recommendation model is used to target students' ideological puzzles and value guidance. In this process, we should also focus on the topic of ideology and politics, pay constant attention to the identity dialog between educators and learners, and adjust the teaching output and dissemination of ideology and politics according to the dialog situation. Taking the learner as the main body, realize the scientific construction, correction and improvement of the learner's thought system.

References

- [1] Liu, X., Xiantong, Z., & Starkey, H. (2023). Ideological and political education in Chinese Universities: structures and practices. *Asia Pacific Journal of Education*, 43(2), 586-598.
- [2] Slawotsky, J. (2021). The fusion of ideology, technology and economic power: Implications of the emerging new United States national security conceptualization. *Chinese Journal of International Law*, 20(1), 3-62.
- [3] Kuznetsov, M. A. (2020). National-cultural ideology as a factor social and economic development of the Russian Federation. *Revista San Gregorio*, (38), 53-64.
- [4] Zixuan, P. (2022). Literature Review on Intelligent Media of Ideological and Political Education. *Academic Journal of Humanities & Social Sciences*, 5(5), 10-17.
- [5] Luo, W., & Qin, Z. (2023). A Research on the Individual Value of Ideological and Political Education in Colleges and Universities in the New Era: Taking Core Value Education as an Example. *Adult and Higher Education*, 4, 120-125.
- [6] Chen, H., & Zhang, Y. (2024). Exploration of the value of current ideological and political education in the context of great ideological and political views. *Journal of Art, Culture and Philosophical Studies*, 1(1).

- [7] Han, X. (2023). Innovative methods of ideological and political education: the combination of theory and practice. *Adult and Higher Education*, 5(20), 29-35.
- [8] Zhu, G., & Tao, T. (2021). Strategies for ideological and political education in colleges and universities from the perspective of general education. *Academic Journal of Humanities & Social Sciences*, 4(11), 18-23.
- [9] Li, D. (2024). The Pathways to Improving the Quality of Higher Education Teaching from the Perspective of Ideological and Political Education. *Journal of Modern Education and Culture*, 1(4).
- [10] Wang, J. (2021). Analysis of challenges and countermeasures of ideological and political education in colleges and universities in the new era. *J. High. Educ. Res*, 2(6), 35.
- [11] Singh, S. K., Sharma, S. K., Singla, D., & Gill, S. S. (2022). Evolving requirements and application of SDN and IoT in the context of industry 4.0, blockchain and artificial intelligence. *Software Defined Networks: Architecture and Applications*, 427-496.
- [12] Tang, C. (2023). Innovation of Ideological and Political Education Based on Artificial Intelligence Technology with Wireless Network. *Eai Endorsed Transactions on Scalable Information Systems*, 10(6).
- [13] Ma, F. (2025). Learning behavior analysis and personalized recommendation system of online education platform based on machine learning. *Computers and Education: Artificial Intelligence*, 8, 100408.
- [14] Veldkamp, B., Schildkamp, K., Keijsers, M., Visscher, A., & de Jong, T. (2021). Big data analytics in education: big challenges and big opportunities. *International perspectives on school settings, education policy and digital strategies: A transatlantic discourse in education research*, 266.
- [15] Zhang, X., & Chen, L. (2021). College English smart classroom teaching model based on artificial intelligence technology in mobile information systems. *Mobile information systems*, 2021(1), 5644604.