

<https://doi.org/10.70517/ijhsa464385>

Big Data Monitoring Model for Ideological and Political Education Effectiveness Based on High-Performance Computing

Pingping Long^{1,*}, Zeng Wang¹ and Xu Jiang¹

¹ College of Marxism, Chongqing Vocational and Technical University of Mechatronics, Chongqing, 402760, China

Corresponding authors: (e-mail: 15215065384@163.com).

Abstract The rapid development of information technology provides new opportunities for educational reform, and the application of big data technology in the field of education is becoming more and more widespread. Aiming at the problems of insufficient precision and inefficient data processing in the monitoring of the effect of ideological education, this study constructs a model for monitoring the effect of ideological education in a big data environment based on high-performance computing support. Methodologically, the improved particle swarm algorithm is used to optimize the K-means clustering algorithm, the inertia weight is improved by introducing the Sigmoid function, the position update formula is optimized by combining the time weight, and the weighted Euclidean distance is applied to improve the clustering accuracy. Then a complete monitoring system containing data collection, preprocessing and analysis is constructed, and MapReduce framework is applied to realize high-performance big data computing. The results show that using 14,620 student civic education data from a university in 2024 for validation, the improved algorithm converges to a steady state after 20 iterations, the distribution of students' grades is normally distributed within the interval of 70 to 100 points, and the best effect is achieved when the number of clustering categories is set to 14. Through 3-dimensional feature analysis, the students were successfully classified into three categories, namely, high scoring segment, middle segment and low scoring segment. The study shows that the model can effectively improve the accuracy and efficiency of monitoring the effect of civic education, providing scientific basis and technical support for realizing personalized civic education.

Index Terms High Performance Computing, Big Data Environment, Civic Education, Effect Monitoring, Clustering Algorithm, Accuracy

I. Introduction

In recent years, with the rapid development and popularization of information technology, big data technology gradually penetrates into various fields and plays a great role in the field of education [1]. Ideological and political education is an important way to cultivate students' correct worldview, outlook on life and values, and it is an important part of college education, while the effect of ideological and political education is the lifeline of the school, which is related to the growth and future of the students, as well as the reputation and development of the school [2]-[4]. In order to ensure the effect of ideological and political education, the monitoring of the effect of ideological and political education using big data technology has become a hot issue at present [5].

The core of monitoring the effect of ideological and political education lies in accurately and timely assessing the effectiveness of each link in the education process [6]. Traditional education quality monitoring methods usually rely on questionnaire surveys, classroom observation and other methods, but these methods are often affected by subjective factors, and the timeliness and comprehensiveness of the data are also relatively lacking [7], [8]. In contrast, big data technology provides a brand new solution. Big data technology can cover a large amount of students, schools and a variety of data in the education process, including students' academic performance, learning behavior, classroom participation, teachers' teaching methods, the use of teaching resources and so on [9]-[12]. By analyzing these massive data, a more comprehensive and objective understanding of the current status of education quality can be obtained, and potential problems and trends can be identified [13], [14].

This study firstly clarifies the direction of data-driven education mode change by analyzing the application value of big data technology in civic and political education. Secondly, in-depth research on high-performance computing technology and improved clustering algorithm is conducted to design an algorithm optimization scheme applicable to the characteristics of educational data. Then, a complete monitoring system containing data collection, preprocessing, analysis and feedback is constructed to realize the comprehensive monitoring of the effect of students'

civic education. Finally, the effectiveness and practicability of the model is verified through experiments to provide scientific basis and technical support for the precise implementation of Civic and Political Education in colleges and universities.

II. Development of political education in a big data environment

II. A. Advantages of Big Data to Optimize the Civic Education Environment

(1) Crossing time and space limitations

In the context of big data, Civics education can be taught regardless of time and place, creating any location in the school as an environment for Civics education. It realizes that students become the main body of learning, and also promotes students to develop self-management and self-education consciousness.

(2) Real-time dynamic feedback

The application of big data effectively solves the problems encountered by ideological education in the network era.

Colleges and universities in the face of a large amount of data information should take stock of the situation, the application of big data on the environment of civic education to carry out a full range of monitoring, its monitoring platform into a dynamic, the hot news of current affairs can be the first time to grasp the technical support conditions for optimizing the environment of civic education in colleges and universities, can quickly improve the quality of teaching and efficiency of civic education.

(3) Provide precise service

Civic and political education in colleges and universities is a kind of activity with social nature, which encompasses data and is equally limited by data.

The application of big data in the teaching of civic education in colleges and universities makes the teaching mode and method more diversified, further optimizes the environment of civic education, and the changes in students' thoughts and behaviors can be recorded through various platforms, so that the teachers can accurately grasp the status of the students' thoughts and changes in their behaviors.

Big data technology can monitor and analyze students' thoughts and behavioral changes in an all-round way from multiple environments, such as family, society, school and network, and form an accurate dynamic map. Understand the individual differences of students and create an environment of ideological education that is closer to students' thoughts. It provides convenient conditions for teachers to make corresponding decisions and judgments for students, which is more conducive to the teachers' work of civic education.

(4) Expanding the scope of audience

The biggest advantage of the Internet is that it is not limited by time and space, which can give full play to the influence of Civic and Political Education, make education more fair, and accelerate the speed of information dissemination. At the same time, the application of big data and new media will play a better role in the propaganda of Civic and Political Education.

II. B. Initiatives of Big Data to Optimize the Civic Education Environment

(1) Changing the philosophy of education management

(2) Establishing a shared data platform

(3) Meeting the needs of students of the times

(4) Adopting advanced education methods

(5) Build a data service platform

(6) Digging deep into the resources of Civic and Political Education

(7) Standardize the procedure of using big data

(8) Setting up accurate and relevant contents

II. C. Precision of ideological education in big data environment

II. C. 1) Connotation of precision in ideological and political education

Precision of ideological and political education is not only the response of ideological and political education to the rapid development of society in the era of big data, but also an effective solution to the problems of ideological and political education in colleges and universities, such as "flooding" and insufficient relevance. A deep understanding of the connotation and characteristics of the precision of ideological and political education is the prerequisite and foundation for exploring the path of precision of ideological and political education in colleges and universities [15].

The connotation of precision in ideological and political education is to achieve the correct matching of education supply and education demand under the guidance of precise thinking, with the goal of establishing morality and cultivating people in colleges and universities, following the laws of ideological and political education and the general law of students' growth and success. Specifically, "fine" means that the education supply should be refined,

and the content and methods of education can be correctly matched. "Accurate" means to accurately identify the education object, and to make efforts to understand and grasp the personality characteristics and real needs of the education object. "Fine" and "quasi" are organically unified, and the two achieve "supply and demand balance". "Modernization" shows that the precision of ideological and political education is not achieved overnight, but is based on the educational goals, identifying the needs of educational objects, providing educational content and methods, and constantly approaching the goal in the practice of education and teaching.

II. C. 2) Big Data Technology + Precision of Civic Education

Through the predictive function of big data, the analysis captures the hidden correlation value between a large amount of data, providing parameters for ideological and political education work. The predictive analysis of big data makes the personality characteristics and hidden features of the education object to be displayed to a greater extent, and the dynamic evaluation and feedback system enables educators to adopt more precise supply methods to buttress the personality needs of the education object and optimize the precision system of ideological and political education in colleges and universities.

(1) Emphasize data acquisition and dig deep into the relevance of data

The prerequisite for the implementation of the precision of ideological and political education in colleges and universities is the need for a large amount of basic data, which requires that colleges and universities must pay attention to the collection and acquisition of data on campus, standardize the format of the data, formulate a perfect data system, broaden the sources of data, realize data sharing, and pay attention to the connection between the data, which is a necessary condition for the universities and colleges to use big data to carry out the precision of ideological and political education.

(2) Construct an assessment and feedback system to guarantee the effectiveness of ideological and political education

Evaluation is a basic part of the process of ideological and political education, and is also one of the basic ways to feedback information on ideological and political education.

With the emergence of big data technology, it is not only imperative but also feasible to build an intelligent and dynamic ideological and political education assessment and feedback system, and the precision of ideological and political education assessment and feedback has a direct impact on the results of education.

(3) Sound supervision system and proper protection of data privacy

In the actual education process, ideological and political educators should adhere to the bottom line thinking, be cautious, and do a good job of data security protection. The precision of ideological and political education will often use the personal data of education targets, which will involve personal data and other issues.

II. C. 3) High-performance computing modeling techniques

(1) High Performance Computing Technology

High-performance computing (HPC) is a high-performance computing system consisting of multiple computer clusters interconnected by a high-speed interconnection network, which is a collective term for modern high-performance computers [16], [17].

Based on this computing paradigm, computer systems can perform better on computational tasks with high parallelism and high throughput rates. In the field of computer science, HPC is defined as a high-performance computer architecture based on parallel processing technology. HPC is a multilevel system, and its hierarchical structure mainly includes several parts such as storage system, microprocessor system, communication system and application software. Among them, the storage system is the most important component in HPC, providing data storage functions for microprocessors and various application software. The microprocessor is the core of the HPC and bears the main function of handling complex computing tasks. The communication system is responsible for the communication work between different computers, thus ensuring that the HPC can operate efficiently in a complex environment. Application software is an important tool for HPC system to process tasks, which can efficiently process various data.

The main component of the high-performance computing model technology in the big data environment is the data center, which mainly includes three parts, namely, hardware infrastructure, software infrastructure, and data storage and processing platform.

High-performance computing model technology in big data environment mainly stores and processes massive data, and combines with high-performance computing model technology in big data environment to realize high-efficiency analysis and processing of massive data generated within the enterprise.

(2) High-performance big data computing framework

High-performance big data computing is actually the embodiment of distributed computing or parallel computing, distributed computing is a computer science, the main object of study is distributed systems. A distributed system

is a hardware and software system composed of multiple computers in an interoperable network environment. This system cooperates with each other to accomplish common computing tasks. Distributed computing, i.e., operational tasks completed on a distributed system, splits tasks requiring complex and large numbers of calculations into several subtasks in a certain way, which are computed by multiple computers in a distributed, parallel, or combined manner.

(3) Data Processing Algorithm-K-means Algorithm

K-Means is a typical contact clustering algorithm. The sample similarity is determined based on the proximity of the distance between the data coordinate points, and the set X containing n data samples is categorized into K classes (K clusters), with higher similarity of data points within each cluster.

Assuming that the data set $X = (x_1, x_2, \dots, x_n)^T$ contains K initial clustering centers, denoted by $C = (C_1, C_2, \dots, C_k)$, calculate the different distances between data objects and clustering centers. Using the Euclidean metric formula $d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$, calculate the distance between each data sample, the Euclidean distance between the cluster centers C_k , and configure the data samples close to the cluster centers C_i into the corresponding clusters, which is calculated as shown in the following equation:

$$d(X, C_i) = \sqrt{(x_{i1} - C_{i1})^2 + (x_{i2} - C_{i2})^2 + \dots + (x_{in} - C_{ij})^2} \quad (1)$$

where C_{ij} clusters the j th attribute value of the center C_i , and x_n denotes the n th data item. The great

likelihood estimation method is used to establish $SSE = \sum_{i=1}^i \sum_{X \in C_i} |d(X, C_i)|^2$ objective function to compute the minimized sum of squares of errors for the whole data set.

The particle fitness within the cluster is calculated for different data similarities, setting the data particle position x_i , the updated position $x'_i = x_i + v_i$, the update rate v_i , and the learning factor of the i th cluster c_1, c_2 , then the data location update rate can be expressed as $v'_i = v_i + c_1 * rand(k_{ibest} - x_i) + c_2 * rand(G_{ibest} - v_i)$. Finding and updating the cluster centers is done in an iterative manner until it stops when there is no change in the cluster centers within the cluster.

The basic idea of the PSO algorithm is to consider each particle as an individual with position and velocity, and each particle converges towards the overall optimum of the cluster and the individual optimum during the iterative process, by constantly searching and updating the optimum value.

In the data search space of the d dimension, if there are n particles forming a community, the i particle can be denoted as $X_i = (x_{i1}, x_{i2}, \dots, x_{id})$, and its velocity can be denoted as, $V_i = (v_{i1}, v_{i2}, \dots, v_{id})$, i The optimal position of the particles that is currently searched will be used as the individual optimum:

$$P_{best} = (P_{i1}, P_{i2}, \dots, P_{id}) \quad (2)$$

Eq, $i = 1, 2, \dots, n$.

For the whole PSO population, the currently searched optimal position can be denoted as the PSO population optimal position:

$$G_{best} = (P_{g1}, P_{g2}, \dots, P_{gd}) \quad (3)$$

Eq, $g = 1, 2, \dots, n$.

In PSO, if the individual optimal position and the population optimal position are obtained, the particle swarm velocity and position update expressions in PSO are:

$$v'_{id} = wv_{id} + c_1r_1(p_{id} - x_{id}) + c_2r_2(p_{gd} - x_{id}) \quad (4)$$

$$x'_{id} = x_{id} + v_{id} \quad (5)$$

where w is the inertia weights. c_1 and c_2 are learning factors, and r_1 and r_2 are random numbers between $[0, 1]$.

To obtain the optimal solution of the algorithm, the sum of squared errors is used as the fitness function, i.e., the objective function is:

$$f(x) = \sum_{i=1}^k \sum_{x_j \in C_i} d(x_j, C_i)^2 \quad (6)$$

where $d(x_j, C_i)$ is the distance from particle x_j to the center of the particle swarm C_i , the smaller the value of the objective function, the better the algorithm works.

Inertia weight is one of the important parameters in intelligent optimization algorithm. The linearly decreasing inertia weights are:

$$w = w_{\max} - \frac{w_{\max} - w_{\min}}{t_{\max}} t \quad (7)$$

where w_{\max} is the maximum value of inertia weights. w_{\min} is the minimum value of inertia weight. t is the current iteration number and t_{\max} is the total selection number.

The inertia weights are negatively correlated with the number of iterations in Eq.

To optimize the traditional linear weight decreasing speed imbalance, the inertia weights are improved based on the characteristics of the S-shaped growth trend of the Sigmoid function.

Definition 1: The expression of the improved inertia weights is:

$$w = w_{\max} - \frac{w_{\max} - w_{\min}}{a} \quad (8)$$

where $a = 1 / (1 + e^{-3x})$ is the improved Sigmoid function variant.

Since the PSO algorithm converges slowly, its later speed can be improved according to the characteristics of the Sigmoid function, and it is found by many experiments that the optimization effect of the induced inertia weights is better when $a = 1 / (1 + e^{-3t})$. The improved inertia weight formula can effectively balance its search ability and the convergence speed of the particles in the iterative process of PSO algorithm.

Definition 2: Introducing time weights into the position update formula of the traditional PSO algorithm, the improved position update is:

$$x'_{id} = x_{id} + (10 - \frac{w_{\max} - w_{\min}}{t})^n v_{id} \quad (9)$$

where x'_{id} is the position of the particle after the update. x_{id} is the current position of the particle before updating, v_{id} is the current velocity of the particle, and the size of n adjusts the range of the particle's position.

The similarity between samples is measured by Euclidean distance, the smaller the distance the more similar the samples are. There is a set of n sample points $X = \{x_1, x_2, \dots, x_n\}$, the distance between two sample points x_i and x_j is defined as the Euclidean distance, where x_i is an m -dimensional vector, and the two m -dimensional vectors $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ and the Euclidean distance between $x_j = (x_{j1}, x_{j2}, \dots, x_{jm})$ is:

$$d(x_i, x_j) = \sqrt{\sum_{s=1}^m (x_{is} - x_{js})^2} \quad (10)$$

In order to improve the accuracy of calculating the similarity between two different vectors, the Euclidean distance is weighted, and the cumulative difference between two different vectors and the similarity between the corresponding elements are considered in order to effectively improve the stability of the K-means clustering results and the clustering accuracy.

In order to increase the degree of differentiation between data attributes, the weights of different dimensional data are obtained as:

$$\omega_{ip} = \frac{x_{ip}}{\frac{1}{n} \sum_{i=1}^n x_{ip}} \quad (11)$$

The p th component in the i th sample point of x_{ip} . The average of the p th component of each sample point in the denominator sample set. The introduction of weights not only does not change the calculation of Euclidean distance in the traditional K-means algorithm, but also increases the degree of differentiation between data features.

The formula for calculating the weights for similarity between corresponding elements is:

$$\omega_{im} = \frac{e^{-|x_i - y_i|/\sigma}}{M} \quad (12)$$

The formula sets $\sigma = 1, m_i = e^{-|x_i - y_i|/\sigma}, M = \sum m_i$.

Definition 3: The weighted Euclidean distance between two sample points x_i and x_j is denoted as:

$$d(x_i, x_j) = \sqrt{\sum_{s=1}^n \omega_i (x_{is} - x_{js})^2} \quad (13)$$

where $\omega_i = (\omega_{ip} + \omega_{im}) / \omega_{ip}$ are the improved weights, ω_{ip} are the weights of the data of different dimensions of the sample points in the computation of the distance, ω_{im} are the weights for the weights for computing the similarity of the corresponding elements, and x_{is} and x_{js} are the data values of the sample point x_i and the sample point x_j in the s th dimension of the vector space, respectively.

The new Euclidean distance formula is obtained after introducing improved weights in Eq. (9). Eq. (13) considers both the cumulative difference between two vectors and the similarity between corresponding elements when calculating the similarity of two vectors, which avoids one-sidedness and thus improves the stability and accuracy of the clustering results.

Based on the improved PSO particle swarm algorithm, K-Means clustering algorithm for IoT data mining, the coding scheme for different data particles' positions and update rates are proposed, which are set as:

$$Z = \{z_1, \dots, z_i, \dots, z_d\}, V = \{v_1, \dots, v_i, \dots, v_d\}, i = 1, 2, \dots, k \quad (14)$$

In order to improve and enhance the search efficiency of the algorithm, in order to ensure that the individual extreme value of the data particles is similar to the global extreme value and the global search is optimal, the Euclidean distance function is used to define the dispersion value of the particles of the data mining, which is calculated by the following formula:

$$fit(ind) = \frac{1}{1 + \sum_{i=1}^k \sum_{X_i \in C_i} D(X_i, B_j)} = \frac{1}{1 + \sum_{i=1}^k \sum_{X_i \in C_i} \|X_i - B_j\|} \quad (15)$$

B_i is the clustering center, X_i is the data particles clustering C_j , k is the number of clustering centers, and $fit(ind)$ is the sum of the dispersion (distance) of the IoT data particles to clustering centers, and the derivation results in the value of the fitness of the data particles, that is, the greater the sum of the data particles' discretization (distance), the smaller the fitness value. The specific IoT data mining execution process is as follows:

Step1: Data particle population initialization. Set the initial clustering center of $C = (C_1, C_2, \dots, C_k)$ according to the IoT numbered dataset of $Z = \langle z_1, \dots, z_i, \dots, z_d \rangle, V = \langle v_1, \dots, v_i, \dots, v_d \rangle$, and use the N data text codes of the dataset as the position codes of the data particles. According to the above initialization particle update rate, repeat M times to generate a data particle cluster with population number M .

Step2: Based on the K-Means clustering algorithm, divide M data particles into N data sets, calculate the distance between data particles and clustering centers in the data sets, and divide the cluster list according to the distance between the data samples and the cluster center points. Calculate the fitness value of the data particles according to the above equation, and update the individual and global extreme values of the data particles.

Step3: Observe whether the global extreme value changes within a specific number of iterations, if it changes, utilize the cloud mutation operator to carry out the particle cluster mutation operation, otherwise jump to Step4.

Step4: Adjust the inertia weights, update the data particle velocity and location according to the above formula.

Step5: Judge whether the algorithm reaches the maximum number of iterations or meets the convergence conditions, if yes, go to Step6, otherwise return to Step2.

Step6: Output the maximum fitness of data particles and the clustering result of K-Means algorithm as the final result.

II. D. Construction of monitoring system for students' civic and political education in colleges and universities

Based on the many application values of big data in the civic education of colleges and universities, the construction of a scientific monitoring system for the effect of civic education of college and university students has become a prerequisite for the implementation of accurate civic education.

(1) Data Collection

Data collection is a key step in building a monitoring system for the effect of students' civic education in colleges and universities, which is divided into two parts: on-campus and off-campus. On campus, the learning management system records learning information from multiple dimensions, course grades reflect learning achievements, classroom performance reflects learning attitude, and online learning behavior shows independent learning ability and interest. In terms of life data, campus card consumption data reflects consumption habits, dormitory management data reflects self-discipline and interpersonal interactions, and clubs and volunteer service data reflects interests and teamwork abilities. Student management data includes reward and punishment records, academic records and mental health counseling records, which can comprehensively present students' academic, moral and psychological status. Off-campus, social media platform data such as WeChat, Weibo, Tik Tok and other social media platforms are collected through legitimate technology, and sentiment analysis and theme model mining are utilized to judge students' emotional tendencies and understand the dynamics of students' thoughts. Public opinion monitoring tools are utilized to pay attention to online public opinion related to universities and college students, providing a basis for responding to public opinion and carrying out ideological education.

(2) Data Preprocessing

Data pre-processing is the basis for building a monitoring system for the effect of students' ideological education in colleges and universities. It mainly covers the three core aspects of data cleaning, integration and transformation. Data cleaning removes these interfering factors and improves data quality by means of rule-based cleaning and cluster cleaning. Data integration, on the other hand, integrates student data scattered in multiple systems with different format structures, solves the problem of data conflict, and forms a unified view. Data transformation, on the other hand, is aimed at the differences in the data outline and value range of different data sources, standardization, normalization, and also feature extraction conversion, so that the data meets the requirements of the algorithm.

(3) Data Analysis Methods

Data analysis is the core link of monitoring the effect of ideological education of students in colleges and universities, and in-depth excavation of the information behind the data can provide a scientific basis for ideological education.

Cluster analysis is to classify students according to their ideological and behavioral characteristics, and group similar students into one class, so as to formulate differentiated strategies for civic education, commonly used K-Means clustering, hierarchical clustering and other algorithms.

(4) Monitoring system construction

Based on the data collection, pre-processing and analysis methods, a comprehensive and scientific monitoring system for the effect of students' civic education in colleges and universities is constructed.

III. Application of improved clustering algorithms in the effectiveness of civic education

III. A. Algorithm implementation in a high-performance big data computing framework

The complete implementation of the optimization algorithm based on improved particle swarm and K-means clustering requires constant iteration, so the project for this experiment has five files FirstCenter.java, KMapper.java, KMeans.java, KReduce.java, and NewCenter.java as a way of achieving a complete fusion of the optimization algorithm process for improving particle swarm and K-means clustering.

First the files need to be passed into HDFS. In FirstCenter.java set the number of categories for clustering to 14 and initialize the centroids. Finish reading in KMapper.java and calculate the distance from each data point to the center point location. If the distance to the corresponding centroid is the maximum, the data point is classified to that centroid. The Squared Distance method is implemented in KMeans.java, which calls an internal method of the MLUtils class to quickly calculate the distance.

NewCenter.java mainly accomplishes the function of updating the centroid file, reading the parameters line by line, converting them to string types, and determining whether they should be stopped. Continuously deletes the old center point file based on the new data from the iteration, making sure that the old file does not exist before the new file is written to the input path.

Finally, configure DFS Locations and set the parameters for the final inputs and outputs, and package it to run on the command line, so that you can view the jobs running in your browser.

III. B. Comparison of Algorithms

The experimental data were used from the entire student body of a university in the year 2024. Preprocessing of the raw data produced 14,620 records. The processed data had a total of 3-dimensional features. The data were analyzed and compared according to these 3-dimensional features. The optimization algorithm based on improved particle swarm + K-means clustering and K-means algorithm were used to process the experimental data using Matlab respectively.

The convergence comparison of the two algorithms is shown in Figure 1. The optimization algorithm based on improved particle swarm + K-means clustering converges at a steady rate after the number of iterations reaches near 20. As can be seen from the figure, under the same conditions, the K-means algorithm converges quickly, but it is easy to fall into the local minimum. And the improved PSO particle swarm algorithm + K-means clustering algorithm global optimization ability is better than the K-means algorithm, and because the particle swarm algorithm does not exist in the degradation phenomenon of stochastic optimization, the convergence is smoother, and there is a faster convergence speed.

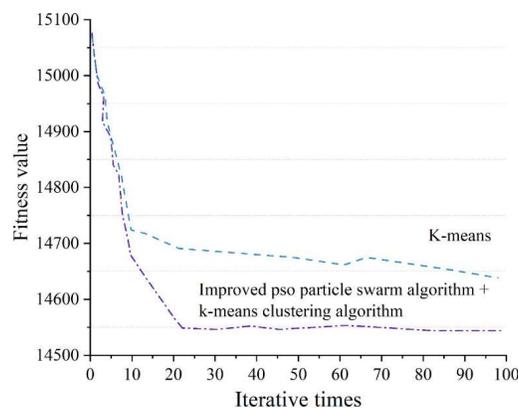


Figure 1: Convergence of two algorithms

III. C. Monitoring and Analyzing the Effectiveness of Civic and Political Education

III. C. 1) Summary of course grades

The distribution of students' Civics raw grades in colleges and universities is shown in Figure 2. It can be seen that it approximates a normal distribution, and the distribution of all course grades is between 70 and 100 points, showing a trend of wide in the middle and narrow at both ends, indicating that the difficulty of the students' final exams is moderate and the differentiation is good.

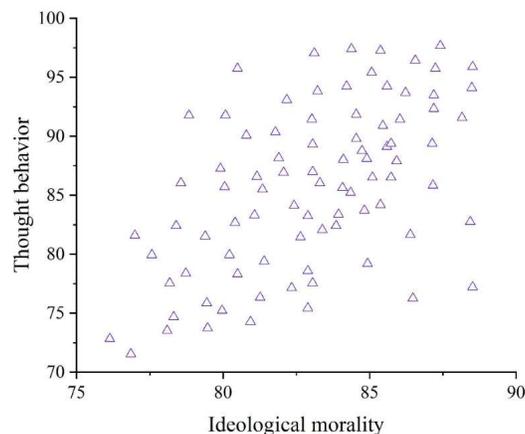


Figure 2: The student's original performance distribution

All student scores were categorized into three categories, with the first category having scores <70 , which is in the low sub-band. The second category with a score of 70-80 belongs to the middle band. The third category with scores >80 belongs to the high segment.

Analysis found that the proportion of students in the first category is closer to the proportion of students in the third category, and the total of the two is close to half of the total number of students, and the results of the comprehensive assessment are also in line with the law of distribution of the middle large and the two small, with objectivity and rationality.

III. C. 2) Effectiveness analysis of improved PSO+K-means algorithm

The clustering distribution results and centroids are shown in Figure 3. As can be seen from the figure, students located in the first category achieved higher results in both dimensions (moral and ethical and intellectual behavior levels). Students located in the middle category have a bias or are more average in both courses. Students in the third category, on the other hand, had less satisfactory results in both courses. Compared with the simple percentage conversion of GPA in weighted ranking, the results of K-means clustering algorithm with improved particle swarm algorithm are less affected by subjective factors and better reflect the macro ability ranking of students in the group.

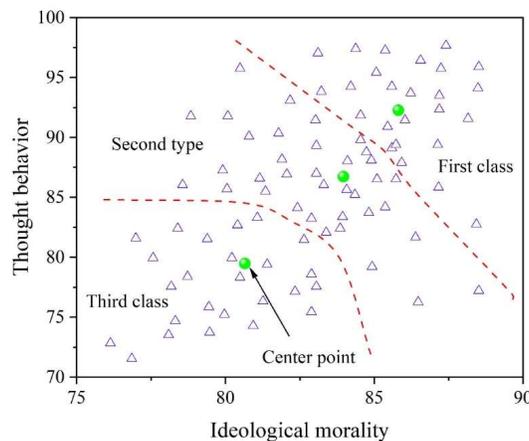


Figure 3: Cluster distribution results and center points

In order to accurately understand the students' mastery of the Civics course in each segment, the value of K was adjusted upward to obtain the distribution of students' performance with different clustering results. When $K=4$, the clustering distribution results and center points are shown in Figure 4.

Except that there is no significant difference between the centroids of the first and the last class, the student grades located in the middle class obviously show different data distribution tendencies.

When $K=4$, there is no significant change in the achievement of the level of thought and behavior at the center point of the vertical coordinate, and the achievement of the course of thought and morality at the center point of the horizontal coordinate is significantly higher. This indicates that the basic Civics and Political Science courses within the evaluation of this grade distribution have better differentiation and have better effect on the comprehensive evaluation of students' ability.

The relevant scores of the final examination of Civics and Political Science in 2024 are selected as data samples, and the optimization algorithms based on improved particle swarm and K-means clustering are used to perform cluster analysis on the percentile scores of the level of ideology and morality and ideological behavior of the Civics and Political Science courses respectively, and to establish a set of performance analysis system suitable for the teaching effect of Civics and Political Science education. The student Civics performance analysis system focuses on data mining, and by collecting and analyzing a large number of grades and training the model, it synthesizes the grades to accurately and comprehensively assess the overall learning situation of students' Civics.

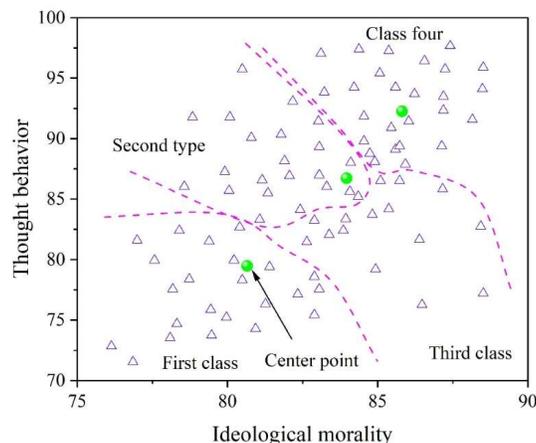


Figure 4: When K= 4, cluster distribution results and center points

IV. Conclusion

This paper verifies the effectiveness of the improved algorithm in educational data analysis by constructing a model for monitoring the effect of Civic and Political education in a big data environment supported by high-performance computing.

The experimental results show that the processed data has 3-dimensional characteristics and can comprehensively reflect the students' Civics and Political Science learning status. The fusion of the improved PSO particle swarm algorithm and K-means clustering algorithm effectively improves the clustering accuracy and convergence stability, and avoids the local optimal trap compared with the traditional K-means algorithm. When the number of clustering categories K value is adjusted to 4, the algorithm is able to identify the differentiated characteristics of the student group more accurately, which provides data support for the formulation of personalized education strategies. The high-performance computing implementation based on MapReduce framework significantly improves the processing efficiency of massive educational data and meets the demand for real-time monitoring.

The construction of the monitoring system realizes the whole-process management from data collection to analysis and feedback, and forms a complete portrait of students' civic and political status through multi-dimensional data fusion. The results of the cluster analysis show that the student population exhibits obvious stratification characteristics, with students in the high, middle and low segments showing different distribution patterns in the dimensions of ideology and morality and behavioral level. This finding provides a scientific basis for the implementation of stratified teaching and precise intervention.

The model lays the foundation for universities to build an intelligent education management system while improving the accuracy of monitoring the effect of ideological education. In the future, the data source can be further expanded and the performance of the algorithm can be optimized to promote the development of Civic and Political Education in the direction of digitalization and intelligence.

References

- [1] Bamiah, S. N., Brohi, S. N., & Rad, B. B. (2018). Big data technology in education: Advantages, implementations, and challenges. *Journal of Engineering Science and Technology*, 13(Special Issue on ICCSIT 2018), 229-241.
- [2] 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.
- [3] Lu, J. (2017). Ideological and political education in China's higher education. *East Asian Policy*, 9(02), 78-91.
- [4] CHen, F. (2015, October). Study on Ideological and Political Education. In *International Conference on Education, Management and Information Technology* (pp. 243-247). Atlantis Press.
- [5] Dong, Y. (2023). Teaching quality monitoring and evaluation in higher education through a big data analysis. *International Journal of Emerging Technologies in Learning (Online)*, 18(8), 61.
- [6] Bazhenov, R., Bazhenova, N., Khilchenko, L., & Romanova, M. (2015). Components of education quality monitoring: Problems and prospects. *Procedia-Social and Behavioral Sciences*, 214, 103-111.
- [7] Vollenbroek, W., Jagersberg, K., de Vries, S. A., & Constantinides, E. (2014, November). Learning education: an 'educational big data' approach for monitoring, steering and assessment of the process of continuous improvement of education. In *2014 European Conference in the Applications of Enabling Technologies, ECAET 2014*.
- [8] Zhang, Y., & Gao, J. (2022). Research on classroom teaching quality evaluation and feedback system based on big data analysis. *Scientific programming*, 2022(1), 7870113.
- [9] Yang, Y. (2022, May). Construction of Monitoring System for Student Growth Evaluation under the Background of Big Data. In *2022 International Conference on Information System, Computing and Educational Technology (ICISCET)* (pp. 98-101). IEEE.



- [10] Ben Shahr, T. H. (2017). Educational justice and big data. *Theory and Research in Education*, 15(3), 306-320.
- [11] Kalim, U. (2021). The growing role of big data in education and its implications for educational leadership. *International Journal of Research and Innovation in Social Science*, 5(01), 257-262.
- [12] Chen, G., & Jiang, Y. (2024, November). Application of Data Mining Technology in the Teaching Quality Monitoring of Ideological and Political Courses in Colleges. In *Proceedings of the 2024 3rd International Conference on Artificial Intelligence and Education* (pp. 477-481).
- [13] Li, Y., Li, P., Zhu, F., & Wang, R. (2017, August). Design of higher education quality monitoring and evaluation platform based on big data. In *2017 12th International Conference on Computer Science and Education (ICCSE)* (pp. 337-342). IEEE.
- [14] Wang, H. (2023). Teaching quality monitoring and evaluation using 6G internet of things communication and data mining. *International Journal of System Assurance Engineering and Management*, 14(1), 120-127.
- [15] Wang Yuefen & Ma Chaoqun. (2022). Precision Strategy of Ideological and Political Education Using Big Data Analysis in Online Behavior Monitoring Environment. *Journal of Environmental and Public Health*, 2022, 2394668-2394668.
- [16] Sushama Seetaram Shinde & Aparna Pande. (2025). High-Performance Computing-Based Brain Tumor Detection Using Parallel Quantum Dilated Convolutional Neural Network. *NMR in biomedicine*, 38(6), e70035.
- [17] Fernando Almeida & Edet Okon. (2025). Assessing the impact of high-performance computing on digital transformation: benefits, challenges, and size-dependent differences. *The Journal of Supercomputing*, 81(6), 795-795.