International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 6021-6035

https://doi.org/10.70517/ijhsa464439

Research on the internal logic and dynamic early warning system of sports ideological and political education based on data mining algorithm

Chengfeng Jiang[1,*]
[1] Physical Education Institute, Zhengzhou University of Industrial Technology, Zhengzhou, Henan, 451150, China
Corresponding authors: (e-mail: 13838232745@163.com).

Abstract The Internet of Things based on data mining algorithms has begun to be used in the early warning of dynamic systems.in recent years, the diversified application of data has developed rapidly, especially in data mining. With the help of data mining, we can select valuable data for research and analysis in huge data. It can be said that data mining is active in all walks of life. This paper focuses on the application of data mining algorithms in the internal logic and dynamic early warning system of sports ideological and political education. Sports ideological and political education plays an important role in colleges and universities. Sports ideological and political education usually includes logical elements such as sportsmanship cultivation, teamwork guidance, competitive ethics, and sports psychology counseling. Among them, the data is huge and mixed. If the effective data cannot be fully screened, it will not only bring defects to the work, but also have a negative impact on the management of the school. This paper investigates the topic of the internal logic of political and psychological learning with the help of data extraction and clustering-based analysis. Based on the data of assessment quantification table, this paper divides the obtained data into four attributes. It calculates that the scores of the four attributes are 0.6171, 0.5927, 0.536, and 0.5917 respectively, and then it is concluded that the management attitude of the counselors in the work assessment is at a high level. But there are certain problems with the management method. Combined with the actual situation, this paper sets early warning thresholds for different attributes and their index scores, and conducts timely early warning analysis on the related work of thinking and political study. Meanwhile, this paper also randomly selects 1000 assessment forms, using traditional methods and dynamic early warning systems respectively. It is concluded that the accuracy rate, time efficiency ratio and reliability performance ratio of the early warning system are 0.25, 0.3 and 0.45 higher than those of the traditional analysis method, respectively. Data mining under the Internet of Things has some important reference significance for the development of artificial intelligence.

Index Terms Digital mining algorithm, Sports ideological and political education, Cluster analysis, Dynamic warning, Sports behavior analysis

I. Introduction

Based on data mining algorithms.the diversified applications of data are gradually known, among which digital mining algorithms play an increasingly important role in daily life. It can even be said that data is born because of us, and we cannot do without it. In recent years, the national education policy has brought convenience to many students, and has also allowed them to go to the palace of universities. As the number of students and teachers increases, it has also overwhelmed the original education management system, especially in the area of the students' spiritual and technical knowledge. However, the existing sports ideological and political education management system is still in its infancy and has not yet formed a scientific and effective model. In physical education, students' fair competition, teamwork and other on-field behaviors and psychological states are direct manifestations of ideological and political literacy, but traditional methods cannot quantify these implicit data. The conventional means of interpretation simply dependent on the estimate of students and schools, or compares the scores. The results of the evaluation of this traditional method are often unfair and cannot fully and effectively evaluate the management effect of the counselors. In such a situation, we can use statistical methods to deal with the work assessment of the counselors.

Considering the above problems, the paper intends to use cluster analysis to analyze the quantitative table of the counselor's work assessment. In this process, it deeply mines large one number of the help of data mining algorithm, and makes full use of these data. It effectively overcomes the traditional analysis methods in the past,

and education of thought and politics of counselors fairly and impartially. It can also better understand the shortcomings of daily ideological and political education work, and it can bring guiding suggestions for follow-up work. It also brings positive significance to the development of education of thought and politics in schools.

Based on data mining algorithms.this paper mainly adopts digital mining algorithms and agglomeration of analysis. Through the digital mining algorithms, this paper fully mines each data. It is analyzing and studying these data and eliminating some invalid data. The data excavated are classified by the cluster analysis method in statistics. In this case, we mainly use K-means technique to agglomerate the data about the ideological and political education work. The employment of this procedure not only saves a whole lot of time and resource, but also greatly enhance the operating efficiency.The innovation in this paper is mainly to use the dynamic early warning system under data mining to perform data mining on the internal logic of ideology and politics.As an important carrier of ideological and political work, physical education can more directly reflect students' ideological dynamics through data such as sports performance, teamwork behavior and competition ethics. By exploring the inherent connections of these data, a more accurate evaluation and early warning mechanism for sports ideological and political education can be established.

II. Related Work

The Internet of Things Based on Data Mining Algorithms.It is not a simple data use. Data mining ranges from shallow data query to fully mining the internal effective information of data, so as to maximize the effect and bring help to daily life. Analyzing past data and establishing a dynamic early warning system allows companies to reduce risks. For example, while the efficient use of renewable energy, it also brings uncertain risks to the power system. Therefore, based on IS, Zhang Y used the ELM model to establish an early warning system to provide help to power companies [1]. Song W built a crude oil price early warning system using neural network model, and took the crisis index as the research index. It enables the constructed early warning system model to be able to accurately predict the future price trend of crude oil [2].In order to predict the risk value of the internal logic of ideological and political education, a data-driven approach is adopted. Inspired by visual qualitative analysis and quantitative analysis, Kong Y proposes a dynamic early warning method for ideological and political education. Predict future trends by predicting key parameters using strongly correlated variables[3]. The application of the early warning system brings early warning to enterprises and individuals, finds problems in advance and then solves them, reducing potential risks. However, the above early warning system does not capitalize on the obtained statistics, which makes the early warning system have certain errors. Now people can use data mining algorithms to deeply mine the potential the treasure of the database and capitalize on data to solve common problems in life.

Based on data mining algorithms.In recent years, more and more scholars around the world have applied data mining to real life. Xu L mainly studied data mining (PPDM) in reducing the risks brought by the data mining process and protecting his privacy while data mining [4]. Data mining algorithms are widely used in daily life. Chaurasia V has studied breast cancer prediction models with the benefit of data mining technology, and its accuracy is much higher than that of IBK and BF Tree methods [5]. Data mining can also influence the decision-making of company managers, avoid risks and bring benefits. Yan X S selected 1,800 data from a company's financial statements. Through data mining, he established corresponding models to predict and analyze stock returns and avoid risks, as well as analyze the company's past profits and losses [6]. In the past traffic accidents, statistical analysis was the main method, which was not only inefficient, but also did not bring convenience to the follow-up work. Lei and Lin analyzed traffic accidents through data mining, classified different accidents, and studied the correlation between time points and accidents, thereby reducing the accident rate and waste of public resources [7]. Data mining is the basis of the whole work, and it needs to go to the next step through cluster analysis. To put it simply, cluster analysis is to find similar correlations in a large amount of data, classify them into multiple categories, analyze different categories, and observe the feature differences of different categories. It finally analyzes some specific categories. Cluster analysis is widely used in various fields. Ilmarinen P used K-means cluster analysis to divide 171 asthma patients into 5 groups, and then took different methods for treatment. The results can be used to predict adult asthma patients and help develop personalized treatments, greatly reducing the difficulty of work and improving efficiency [8]. The data obtained by the data mining algorithm significantly increased the early warning system and brings great help to enterprises and individuals.

III. Method of Internal Logic and Dynamic Early Warning System of Ideological and Political Education

III. A. Data Mining Algorithms

Based on data mining algorithms.The mechanism of ideological and political education is a whole composed of various components based on a certain logical relationship. In-depth discussion of the inner logic of ideological and political education for college students is an inevitable requirement for grasping its inner essence in the theoretical thinking space and space vision. Along the research path of "logical starting point - logical evolution - logical meaning", starting from the three dimensions of germination base point, evolution process and influence realization, it reveals its internal connection: the basic contradiction of ideological and political education, the need for interaction between elements and the The communication practice is the starting point of the driving force, and the process goes from decentralization to agglomeration. The logical evolution from disorder to order, from endogenous to externalization, finally realized a comprehensive effect with institutionalization, benignity, normalization, and predictive control as the core.

Based on data mining algorithms.This paper mainly studies the internal logic of ideological and political education of data mining algorithms. Everything has a cause and an effect. In the diagnostic process of ideological and a political and educational work, it is recommended to apply the theory related to association rules[9]. It is the discovery of the existence of such association rules that makes the originally seemingly unrelated data connect. By borrowing this rule, it finds the coupling relationship between the data, so that a large amount of complex data can be followed regularly, and finally an effective conclusion is obtained.If there is an abnormality in the data, the system will terminate the data mining or the obtained data cannot determine the correlation between them.Data mining is a test method and calculation method for establishing a data mining model based on data. To build a model, the algorithm analyzes the data you provide and then finds specific patterns and trends. The algorithm uses the results of this analysis to determine optimal parameters to build a mining model. These parameters are then applied to the entire data set, resulting in possible models and specific statistics.

Data mining is a critical step in the entire process. It mainly uses methods such as association rules, cluster analysis and classification analysis. After choosing a method in the continuation, it also needs to choose the correct processing method. In general, in association rules, the Apriori algorithm is preferred. In cluster analysis, most of them use the K-means method, which can quickly draw effective conclusions [10].

Based on data mining algorithms.Data mining is the premise of data analysis. It is of special significance to apply this technology to ideological and political education. The evaluation system in the past could not draw conclusions effectively, and could only one-sidedly analyze the work of ideological and political education. It also cannot help counselors understand the level of ideological and political education of students. Now, with the help of data mining methods, under the original data, it deeply mines the internal relationship of the data. Through the cluster analysis method, it converts a large amount of data into clustering results, making the results more concise and clear, and then better helping the counselors to work [11]. The process of data mining is shown in Figure 1:

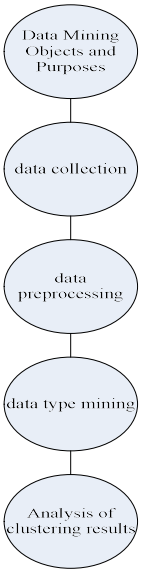

Figure 1: Data mining process based on Internet cluster

Data mining first needs to determine the object and purpose, then collect the data, preprocess the collected data, and then classify and mine according to the type of data, and finally get the result analysis through the clustering algorithm.

III. B. Cluster analysis

Based on data mining algorithms.In the data exploitation of the process, cluster analysis is the main method. It is to classify a large amount of data according to the similarity to form clusters with different characteristics. It not only reduces the workload of analysis, but also avoids the phenomenon that the results deviate from expectations due to similar data. Cluster analysis will form sets with different characteristics after data integration. It can visually express the application of cluster analysis in the form of scatter points, as shown in Figure 2:

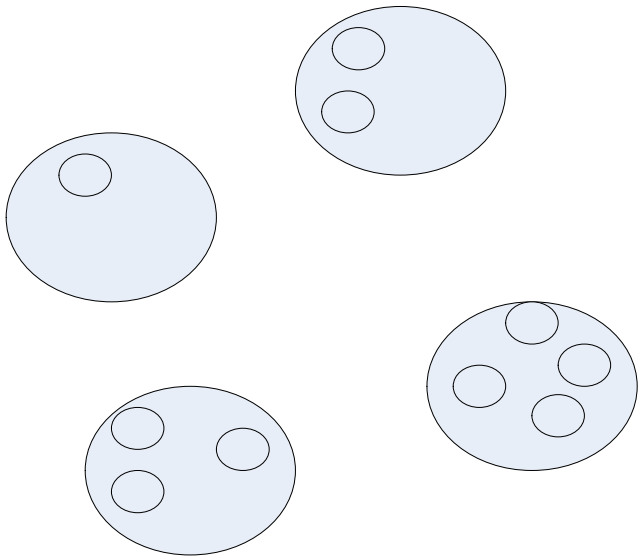

Figure 2: Schematic diagram of cluster analysis based on data mining algorithm

As can be seen from Figure 2, each small circle in the circle represents a similar case, and the large circle represents a cluster of different characteristics, which can clearly and clearly see the degree of aggregation between different data.

At the same time, cluster analysis can also be represented by a flowchart, as shown in Figure 3:

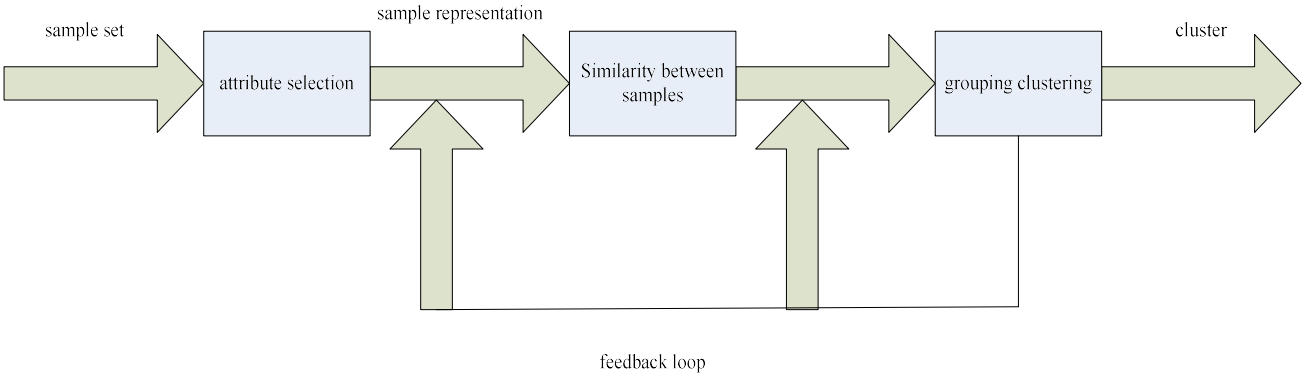

Figure 3: The process of cluster analysis based on data mining

In the cluster coherence analysis process, data is generally preprocessed, mainly two types of data, as follows:

(1) Data Matrix

Based on data mining algorithms.This paper selects P variables to describe all N target objects [12]. As far as this article is concerned, the management effect is measured by "hosting class meetings and college-level meetings + understanding students' class situations + student evaluations". It first determines the interval scale between data types, and then constructs this relationship into an $n \times p$ matrix:

$$\begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \cdots & \cdots & \cdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \tag{1}$$

(2) Dissimilarity matrix

Based on data mining algorithms.When objects in a particular dataset have such pairwise similarity that all objects can be equally well matched, the matrix formed by this dataset is called the heterogeneity matrix[13]. It can be expressed by the following matrix:

$$\begin{pmatrix} 0 & \cdots & \cdots \\ \cdots & 0 & \cdots \\ d(n,1) & d(n,2) & 0 \end{pmatrix} \tag{2}$$

In the above matrix, $d(i,j)$ is the specific quantification form of the dissimilarity between any two objects i and j of the selected N target objects, which are generally divided into the following cases:

a) When $d(i,j)$ tends to 0, it indicates that the two objects are similar.

b) When $d(i,j)$ tends to be extremely large, it indicates that there are differences between the two objects, and $d(i,j)=d(j,i)$ and $d(i,i)=0$ may also exist.

To further understand the above matrix, we will discuss the measurement method of $d(i,j)$:

The interval scale metric generally describes the similarity between two objects by calculating the distance between the two objects, also known as the Euclid distance [14]. The calculation formula is as follows:

$$\mathrm{d(i,j)} = \sqrt{\left|\mathrm{x_{i1}} - \mathrm{x_{j1}}\right|^2 + \left|\mathrm{x_{i2}} - \mathrm{x_{j2}}\right|^2 + \ldots + \left|\mathrm{x_{im}} - \mathrm{x_{jn}}\right|^2} \tag{3}$$

Among them, $\mathrm{i} = (\mathrm{x_{i1}}, \mathrm{x_{i2}}, \ldots, \mathrm{x_{in}})$ and $\mathrm{j} = (\mathrm{x_{j1}}, \mathrm{x_{j2}}, \ldots, \mathrm{x_{jn}})$ are n-dimensional data objects,d is the direct distance between two data objects, and Manhattan distance [15] can also be used. The calculation formula is as follows:

$$\mathrm{d(i,j)} = \left|\mathrm{x_{i1}} - \mathrm{x_{j1}}\right| + \left|\mathrm{x_{i2}} - \mathrm{x_{j2}}\right| + \ldots + \left|\mathrm{x_{in}} - \mathrm{x_{jn}}\right| \tag{4}$$

Regardless of formula (3) or (4), the following conditions should be satisfied:

(1)$\mathrm{d(i,j)} \geq 0$

(2)$\mathrm{d(i,j)} = 0$

(3)$\mathrm{d(i,j)} = \mathrm{d(j,i)}$

(4)$\mathrm{d(i,j)} \leq \mathrm{d(i,h)} + \mathrm{d(h,j)}$

With the improvement of the research method, the above two formulas have also been deformed and proved experimentally. The deformed formula is as follows, which is also called the Minkowski distance [16].

$$\mathrm{d(i,j)} = \left\|\mathrm{x_i} - \mathrm{x_j}\right\|_\mathrm{p} = \left(\sum_{\mathrm{k=1}}^{\mathrm{n}} \left|\mathrm{x_{ik}} - \mathrm{x_{jk}}\right|^\mathrm{p}\right)^{\frac{1}{\mathrm{p}}} \tag{5}$$

Among them, $\mathrm{p} \geq 1$. When p is 1, it is formula 4, and when p is 2, it is formula 3.K is the judgment coefficient, and p is the number of samples.

A binary variable is that each variable has two states, which can be represented by 0 for no and 1 for yes. If we continue to use formulas 3, 4 and 5 to measure binary variables, it may bring errors to the clustering results. Therefore, we must adopt another method to measure the dissimilarity of binary variables.

Binary variables generally have symmetric and asymmetric types. In short, if the meanings of 0 and 1 are equivalent, then it can be called a symmetric binary variable. Most scholars use Equation 6 to express the dissimilarity of two objects:

$$\mathrm{d(i,j)} = {(\mathrm{r}+\mathrm{s})}\Big/{(\mathrm{q}+\mathrm{r}+\mathrm{s}+\mathrm{t})} \tag{6}$$

If 0 and 1 of the resulting binary variable represent different meanings, it is called an asymmetric binary variable. Generally, the Jaccard coefficient is used to evaluate the dissimilarity of binary measures, and the formula is as follows:

$$\mathrm{d(i,j)} = {(\mathrm{r}+\mathrm{s})}\Big/{(\mathrm{q}+\mathrm{r}+\mathrm{s})} \tag{7}$$

Mixed-type measurement means that there is a large amount of data in a database, and there is not necessarily only a single variable relationship between the data, and there may be multiple types of variable combinations. In

order to better measure the similarity between these variables, the general situation is to group these variables into several groups, and then use cluster analysis on the variables of each group separately. However, this method is complicated and has certain limitations. So we can use another method to combine various types of variables into a dissimilarity matrix. And it maps all valid variables to the interval range of $[0,1]$.

Assuming that there are p mixed components in the database, we can define the difference between objects i and j by the following statement:

$$d(i,j) = \frac{\sum_{f=1}^{p} \delta_{if}(f) d_{if}^{(f)}}{\sum_{f=1}^{p} \delta_{if}(f)} \tag{8}$$

If the variable f is an interval-scale variable, then

$$d_{if}^{(f)} = \frac{|x_{if} - x_{jf}|}{max_h x_{hf} - mi_h x_{hf}} \tag{9}$$

h passes through all non-missing objects of f.

At the same time, if the variable f is a binary variable or a categorical variable, then $d_{if}^{(f)}$ is 0 when $x_{if} = x_{jf}$, and 1 otherwise. If the variable f is a proportional scale variable, there are two processing methods. The first is to perform logarithmic transformation on the variable, and then treat the processed data as an interval scale. The second is to treat f as a continuous ordinal variable, calculate r_{if}, z_{if}, and then treat z as interval-scaled data. This method is similar to the previous method, except that formula (8) normalizes the variables so that the values are mapped to $[0,1]$, so that even if the variable types are different, the dissimilarity between objects can still be measured.After data mining, the class situation and students' evaluation can be accurately evaluated, and the meaning between sentences and objects can be judged after data processing.

III. C. Algorithm Comparison

Comparison of algorithms based on data mining.There are many kinds of cluster analysis algorithms, such as single connection algorithm, K-means method and COBWEB algorithm. Now this article briefly introduces the single connection algorithm, COBWEB algorithm and K-means method.

(1) Single connection algorithm

It calculates two objects in a certain data, and groups them into a cluster according to the calculated distance. It repeats the above operations until all objects are in the corresponding cluster or the distance between the clusters exceeds a certain value, and the cluster analysis process ends. This method is a single-connection algorithm, and the calculation formula is as follows:

$$d_{min}(c_i, c_j) = min_{p \in c_i, p \in c_j} |p - p'| \tag{10}$$

$$d_{max}(c_i, c_j) = max_{p \in c_i, p \in c_j} |p - p'| \tag{11}$$

$$d_{mean}(c_i, c_j) = |m_i - m_j| \tag{12}$$

$$d_{avg}(c_i, c_j) = \frac{1}{n_i n_j} \sum_{p \in c_i} \sum_{p \in c_j} |p - p'| \tag{13}$$

(2) COBWEB algorithm

The COBWEB algorithm generally adopts non-complex incremental concept analysis. The input data points to the algorithm are represented using attribute-value pairs. The classification effect can be described as:

$$\frac{1}{n} \sum_{k=1}^{n} p(c_k) \left[\sum_i \sum_j p\left(A_I = \frac{v_{ij}}{c_k}\right)^2 - \sum_i \sum_j p(A_i = v_{ij})^2 \right] \tag{14}$$

n represents the node in the new partition $\{C_1, \ldots, C_2\}$ generated at a certain level of the classification tree. Dissimilarity also occurs in this classification, where the probability P can be expressed as an internal similarity by the following formula:

$$P\left(A_i = {}^{V_{ij}}\!/_{C_k}\right) \tag{15}$$

When P is larger, it can be seen from the above formula that the number of $A_i = V_{ij}$ is larger, then the probability of predicting the property is higher.

(3) K-means method

Algorithms based on data mining.In the current data research, the K-means method is generally used. The division criterion is that the data in the same set has a certain correlation, while the data between different sets has no similarity.

Because the relationship between the single-connection algorithm and the COBWEB algorithm in judging concepts is likely to lead to large differences in judgment, and the K-means algorithm is more accurate, the K-means algorithm is adopted, and there is no comparison between these two algorithms. Compare.

Suppose now that a data set containing M-tuples is to be randomly divided into K groups, each group is unique, $K < M$, and each group needs to have at least one data in these K groups. And this data is unique and cannot have any intersection with other groups, so repeatedly, K groups are finally obtained. It arithmetically averages the data in each group, and takes the average value, then the average value can fully have the corresponding characteristics of this group. This paper mainly uses the K-means method:

The central element of the K-means method is that we should first determine the clustered centroids before using the K-means method [17]. The general four-step K-means method:

Step 1: it randomly selects K objects, and sets these objects as the centroids of several clusters.

Step 2: It compares the remaining objects with the previously set K objects.

Step 3: According to the calculation results, the remaining objects are grouped into the clusters with the shortest distance.

Step 4: After all the remaining objects are summed up, the centroid of the new cluster is calculated here, and replicate the operation until the result of the standard function is converged[18]. The equation of the criterion function is as follows:

$$E = \sum_{i=1}^{k} \sum_{p \in c_i} \left| p - m_i \right|^2 \tag{16}$$

The formula for calculating the centroid is as follows:

$$\mu_j = \frac{\sum_{i=1}^{m} 1\left\{c^{(i)} = j\right\} x^{(i)}}{\sum_{i=1}^{m} 1\left\{c^{(i)} = j\right\}} \tag{17}$$

E represents the sum of squared errors calculated for all objects in space. And p refers to all the data in the cluster, m_i represents the centroid of the cluster c_i, and both p and m_i are multi-dimensional. The reason for using this criterion is to make clusters as independent and compact as possible to avoid similarities between clusters. The algorithm is shown in Table 1 below:

Table1: K-means algorithm flow based on data mining

Algorithm name: The k-means algorithm for partitioning, the center of each cluster is represented by the mean of the objects in the cluster.
Input: k: the number of clusters,
D: A dataset containing n objects.
Output: The set of k clusters.
method:
(1) arbitrarily select k objects from D as the initial cluster center;
(2)repeat
(3) According to the mean of the objects in the cluster, each object is assigned to the most similar cluster;
(4) Update the cluster mean, that is, calculate the mean of the objects in each cluster;
(5)until no longer changes

Using the previous method, repeated iterations and calculations to make the value of E smaller and smaller [19]. The smaller the E is, the higher the similarity of objects in the same cluster. At the same time, the larger the pair of data objects, the better the effect of the algorithm, and the conclusion will be in line with the ideal expectations.

In order to better reflect the research of K-means algorithm, the clustering step in $K = 2$ is now performed, and the results are shown in Figure 4 below:

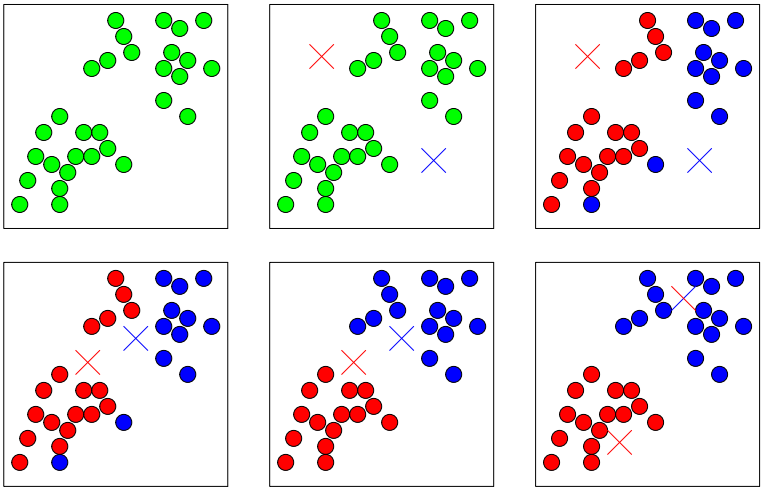

Figure4: Example of clustering steps of K-means algorithm based on data mining

In such a way to better display the superiority of the K-means algorithm, we will compare and analyze the scalability of the algorithm, the type of clusters and the efficiency of processing the problem. The results are as follows:

Table 2: Comparison of Clustering Algorithms Based on Data Mining

Algorithm	Scalability	cluster shape	Algorithmic efficiency
K-means	excellent	convex or spherical	high
Single join algorithm	excellent	convex or spherical	Low
COBWEB	generally	any shape	Low

In Table 2, we can see that the performance of K-means algorithm is outperformed by the other two algorithms, so K-means algorithm will be used to this paper to be analyzed and study the sequel.

IV. Experiment on Internal Logic and Dynamic Early Warning System of Ideological and Political Education

IV. A. Identifying Data Mining Objects

Based on data mining algorithm. This paper collects and organizes the 2020 "PE Teachers' Ideological and Political Work Assessment Form" of a certain university's School of Physical Education, randomly selects 120 assessment data related to sports ideological and political education, and sets relevant topics such as "Ideological and Political Education Work", "Party Style and Academic Style Construction", and "Daily Management and Psychological Education of Students". This paper also conducts in-depth mining of the data obtained from the problems, and draws effective conclusions. The conclusions achieved are being used to help counselors in the practice of thinking and polity in education.In order to ensure the accuracy of the data, multiple extractions are performed in this paper, and the final data is the mean value of the extraction.

IV. B. Data Acquisition and Preprocessing

(1) Data collection

From the Student Affairs Office of the school, we collected the "Counselor Work Assessment Quantitative Form" of the School of Big Data in 2020.

(2) Data preprocessing

The "Physical Education Teachers' Ideological and Political Assessment Form" mainly sets indicators such as "fair refereeing of competitions", "penetration of sports spirit into teaching", and "effectiveness of team conflict mediation". It is divided into five levels of "excellent, good, medium, qualified, and unqualified" based on multiple indicators such as training logs, game video analysis, and team members' mutual evaluation, as well as the counselors' work performance. In this paper, the collected 120 assessment forms are simply sorted, and 113 valid forms are finally obtained.

(3) Data conversion

Given that there are too many indicators in the work assessment quantification table, it is not conducive to subsequent analysis. In this paper, the 15 indicators in the 113 valid assessment forms are divided into four clusters according to the method of cluster analysis. They are named as "management attitude", "management ability", "management method" and "management effect", among which:

"Management attitude" = clear reward and punishment, adhere to the principle + establish a deep relationship between teachers and students + unity of knowledge and action, set an example + fairness and justice, honesty and self-discipline.

"Management ability" = grasp the situation of students' families, do a good job in helping them + take corresponding measures for special students, do not wear tinted glasses + do not condone students who violate discipline + actively organize activities that are beneficial to the body and mind + help students do praise and priority work .

"Management method" = carry out ideological education and service work for students + insist on weekly inspection of student dormitories and other issues + take the initiative to understand their own situation, enlighten students' ideological status + participate in morning running and morning self-study with students + student grants and scholarships.

"Management effect" = hosting class meetings and college-level meetings + understanding students' class status + student evaluation

In the entire assessment quantification table, each indicator is relatively abstract. First, each rating must be converted into a common data type for better analysis. We can map the assessment of the above five grades to the area of $[0,1]$, and with the help of formula 18, we can calculate the five values of the assessment rating (1, 0.75, 0.5, 0.25, 0). Doing this makes all variables have the same weight.

$$z_{if} = \frac{r_{if} - 1}{M_f - 1} \tag{18}$$

The value of the four attributes can be calculated by the arithmetic mean method. Now this paper selects one of the attributes for calculation, and the other attributes follow the calculation formula as follows:

$$Y = \frac{(x_1 + x_2 + ... + x_n)}{n} \tag{19}$$

Among them, x_n is the value of the four attributes after data conversion, n is a positive integer and .

IV. C. Algorithm Implementation

Based on data mining algorithms.In view of better understanding the situation of thought and polity education and intuitively solving the problem, the top 3 of 100 assessment quantified tables were freely selected in this paper as the sample criteria for this data mining. Then, this paper segmented the desired data points into the nearest clusters by adopting Euclidean distance. Finally, the point mean vector in each cluster is divided and used as the new centroid for recursive operation as follows.

Assuming that there is a data set $D = \{x_1, x_2 \ldots x_n\}$, the purpose is to find K clusters $(C_1, C_2, \ldots C_k)$ in it. Meanwhile, K=3 was taken in this study by analysis. The algorithm is thus programmed through VC++:

Taking the first 3 samples as cluster centers:

While the cluster Ck changes do

Form clusters:

For k=1,…n do

Ck= $\{X \in D \mid D(Rk, x) \le D(Rj, x)\}$, $D(Rj, x)$ is against all $j = 1, \ldots k, j \ne k$.

End;

re-cluster new centers;

For k=1,…n do

$Ck = Rk$ mean vector of interior points

End;

End.

At the same time, the flow chart of the algorithm is also introduced, as shown in Figure 5:

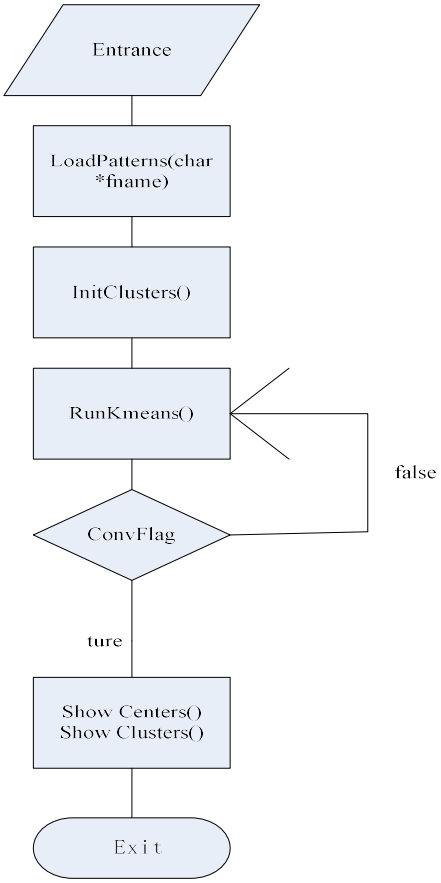

Figure 5: Flowchart based on data mining algorithm

IV. D. Data Clustering Mining Results

Through the preprocessing and transformation of the data, the specific data mining samples are finally obtained, as shown in Table 3:

Table 3: Based on the basic example of data mining

Sportsmanship training	Teamwork guidance	Competitive ethics education	Sports psychology influence
0.65	0.6	0.56	0.58
0.65	0.6	0.56	0.58
0.35	0.35	0.31	0.33
0.76	0.76	0.68	0.75
0.8	0.8	0.81	0.83
…	…	…	…

After getting the above results, what we expect is that the above data sample information can be subdivided into three aspects: excellent, good and unqualified. It can also use the K-means method in the cluster analysis to further obtain the distribution of the sample data in Table 3 and the proportional distribution of the three grades. It is better to find out the merits and deficiencies of counselors' ideological and political education.

V. Discussion

V. A. Clustering Results

Through the cumulative analysis of the last 100 different samples, the clustering results obtained are shown in Table 4:

Table 4: Clustering Results Analysis Based on Data Mining

	Management attitude	Management ability	Management Method	Management Effect	Number of Samples
Excellent	0.77	0.77	0.74	0.79	30
Good	0.61	0.57	0.54	0.56	62
Difference	0.31	0.31	0.28	0.3	8

Cluster 1 deducts 1 sample data through the analysis of data samples, leaving 29 samples. So, 29/100=0.29; subtracting 1 sample from cluster 2 (moderate), it is 61/100=0.61; and the third cluster (poor) is 7/100=0.07. Then the four attribute scores are calculated by the formula, formula (20) is as follows:

$$Y = x_1 \times n_i + x_2 \times n_2 + ... + x_n \times n_n \tag{20}$$

Among them, x_n is the rating score of each item, n_n is the sample size, and n is a positive integer.Some unsatisfactory conditions are found by drawing data, which will be deleted, and the final data will be illustrated in the form of graphs below.

At the same time, considering the centroid of this data mining, the average score of the above four items is:

Management attitude=0.77*0.29+0.61*0.61+0.31*0.0.07=0.6171

Management ability=0.77*0.29+0.57*0.61+0.31*0.07=0.5927

Management method=0.74*0.29+0.54*0.61+0.28*0.07=0.5636

Management effect=0.79*0.29+0.56*0.61+0.30*0.07=0.5917

In order to express the situation of the above four attributes more intuitively, the average algorithm of the above four attributes is now performed, and the results are as follows:

Average = (management attitude + management ability + management method + management effect)/4=(0.6171+0.5929+0.5636+0.5917)/4=0.5913 (four decimal places). This paper mainly compares the new dynamic early warning system and the traditional early warning system from the three aspects of accuracy, time efficiency ratio and reliability.The distribution of the five results in the graph is shown in Figure 6:

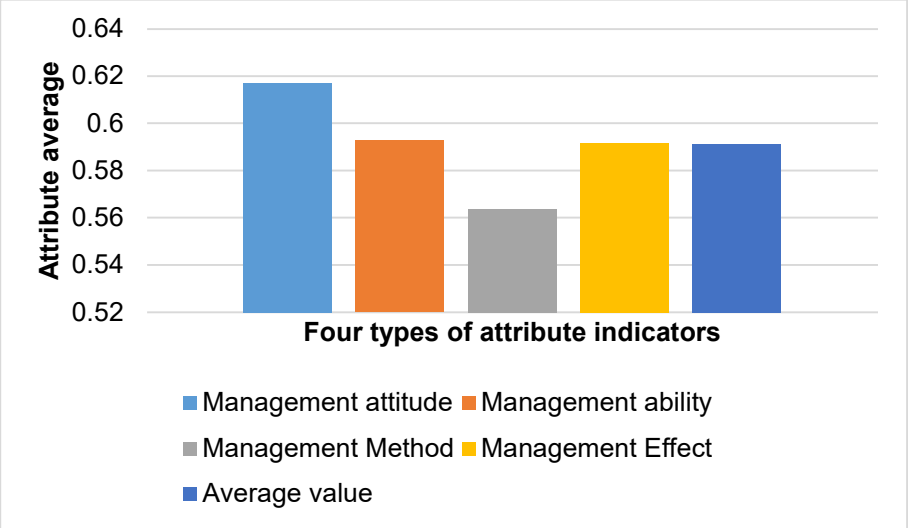

Figure 6: Attribute and Average Based on Data Mining Algorithm

Figure 6 shows that the level of ideological and political education work of counselors in this college is at a comparatively high level, which indicates that they work diligently in their daily ideological and political education work and actively deal for the problems in their work. Among the four different attributes, only management attitude exceeded 0.6, which was much higher than the other three attributes. It shows that in the ideological and political education work of the big data college, the management attitude of the counselors is of great help to the daily ideological and political education work. After taking the average of the four attributes, and comparing the four attribute values with the average, it was found that the management method was far below the average. It shows that there are certain problems in the management method of the big data college. By analyzing five indicators of management methods, the college's counselors can make improvements in their work. If ideological education work is carried out every weekend, students should be patriotic education, actively go to the student dormitory to

understand the situation, care about the physical and mental development of students, and guide them to face life and study positively. When students achieve excellent results, they should actively help students apply for scholarships and other jobs.

V. B. Dynamic Early Warning System

Based on data mining algorithms.As can be seen in Table 4, there are three rating criteria for each attribute, and there are certain differences in their scores. It can be seen from the score of each attribute calculated in combination with the above that the difference in the score of each rating will directly affect the final attribute score. Therefore, we should set up a dynamic early warning system that can find the cause and solve the problem in time. As mentioned above, we can see that the score of each attribute is high or low. First, we should set a warning threshold of 1 for the four attributes. When the score of an attribute is lower than the warning threshold of 1, an early warning will be issued, so that the counselor can make timely improvements for the early warning attribute. After combining the above attribute score calculation formula, it can be seen that each attribute is related to the score of its three rating criteria. Therefore, it can also set separate warning thresholds for the scores of the three rating standards in the early warning system. An early warning is issued when a certain rating criterion of its attribute is scored too high or too low [20]. If the excellent score is lower than the set warning threshold or the poor score is higher than the warning threshold, a corresponding warning will be issued according to the actual situation. The process of the dynamic early warning system for ideological and political education is now introduced:

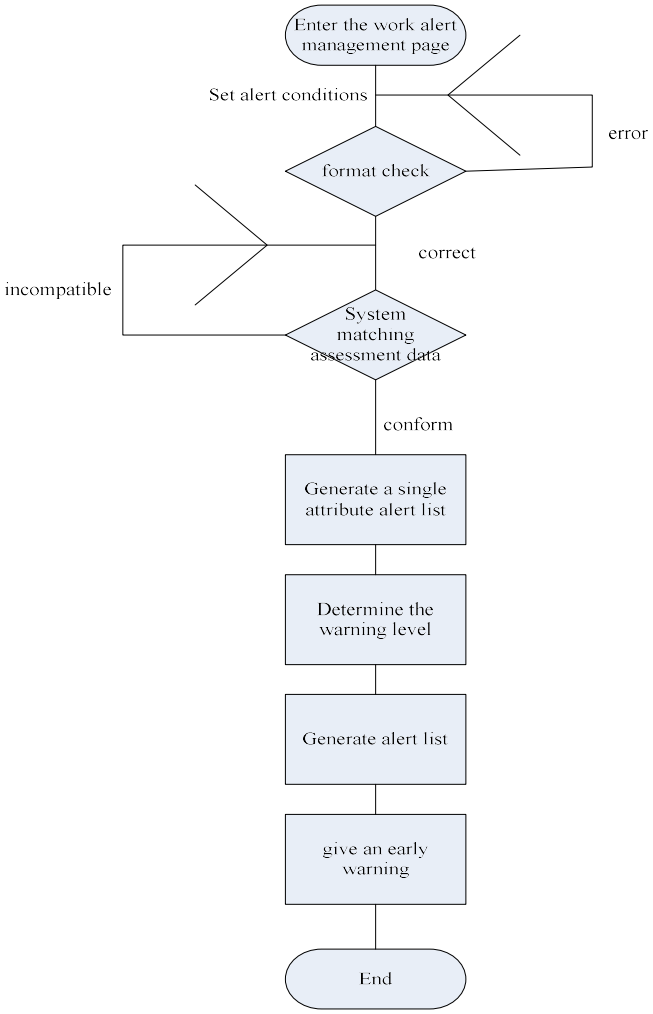

Figure7: Flow chart of early warning release based on data mining algorithm

As shown in Figure 7, physical education teachers or coaches can enter athlete behavior data, teamwork scores, competition ethics performance, etc. into the flow chart, and finally obtain the corresponding warning results, and adjust the training plan or ideological and political education strategy in time according to the warning results to

avoid delays in ideological and political education. Combining the above situation, it can be seen that the score of a certain scoring standard will affect the attribute score. In order to issue warnings more accurately, different warning levels should be set for it so that counselors can handle it according to the warning level. The warning conditions for sports ideological and political education need to be combined with the characteristics of the sports scene, as shown in Table 5:

Table 5: Comparison table of warning levels and conditions based on data mining algorithms

serial number	Warning condition	Warning level
1	Some athletes showed negative competition behavior	Level 3 Early Warning
2	Teamwork scores continued to decline	Level 2 Early Warning
3	Risk of collective sports ethics violations	Level 1 Early Warning

After the above attribute formula is calculated, it calculates the score results of the four attributes respectively. As can be seen, management approach scored lower than the other three attributes. In combination with Table 4, it can be seen that the scores of the three rating criteria of the counselor's management method are lower than the other three attributes, which directly makes the overall score at a low level, and the score of the management method exceeds 0.55. It shows that the management method is generally fine, but there may be problems with some of the scores of the management method. So we can use the early warning system flow chart to quickly know which rating has a score lower or higher than the threshold. Combined with the relevant data of the work assessment quantitative table, it gives the counselor an accurate judgment on the problem and solves the problem in a timely manner.

In order to better detect the effect of the early warning system, this paper selects the data of 1,000 copies of the "Counselor Work Assessment Quantitative Scale" in 2021, and sets up an experimental group and a control group. The experimental group used an early warning system, while the control group used traditional methods for analysis. After the experiment, it is concluded that the accuracy of the early warning system is much higher than that of the traditional analysis method. The experimental data are shown in the following Table 6:

Table 6: Data table of control group and experimental group based on data mining

	Control group	Test group
Accuracy(%)	70	95
Time efficiency ratio	0.6	0.9
Reliability performance ratio	0.5	0.95

To keep the histogram in perspective, we have harmonized the percentages in Table 6 with the values of the two indicators below, and the result is as follows:

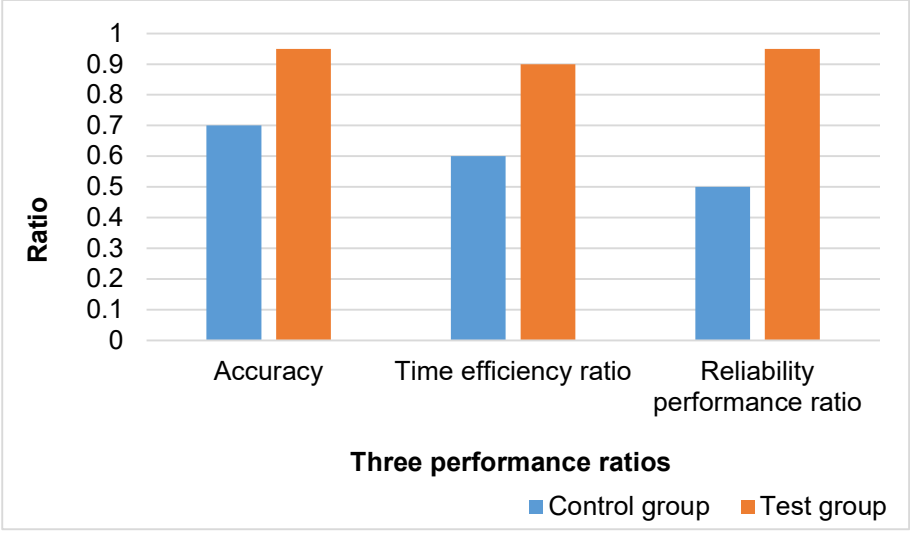

Figure 8: Comparison chart based on data mining algorithm

Combining the data in Table 6 and Figure 8, it can be seen that after the dynamic early warning system is adopted, its accuracy, time efficiency and reliability performance ratio are all larger than those of the traditional analysis method. It shows that the effect of this dynamic early warning system is remarkable.

VI. Conclusion

Based on data mining algorithms.In recent years, data mining algorithms have been increasingly sought after by researchers in various countries. Data mining can not only identify potential data, but also conduct integrated research on these data. Through data mining technology, it can not only increase corporate profits and avoid risks, but also greatly reduce the company's costs. Data mining is a vigorous industry, applied in all aspects of society. However, in the ideological and political education of the education system, the application is relatively small. The data of ideological and political education work is huge and complex, and there are many potential data, and ordinary processing methods cannot effectively use these data. It may even miss some important data. Now it can carry out data mining on the ideological and political education work of counselors, and make full use of the data, which can help the ideological and political education work. This paper mainly takes the ideological and political work assessment of physical education teachers and athletes' behavior data as examples to conduct data mining.In the cluster analysis method, this paper first filters and preprocesses the data, and then uses the K-means method for in-depth analysis. It draws some useful conclusions. After adopting the dynamic system early warning, it was found that the accuracy was improved by 25%, the time efficiency ratio was improved by 0.3, and the reliability was improved by 45%.Based on the excavated data, this paper conducts an early warning analysis of the ideological and political education work, and constructs an early warning flow chart. It is more accurate to obtain the situation between attributes and between attributes and attributes within attributes. This article then makes corrections according to the specific indicator early warning level, which greatly improves the ideological and political education work efficiency of counselors. The main work of this paper can be simply summarized into the following four points:

Processing and transformation of mined data:

(1) This article has collected 120 quantification tables for the work assessment of counselors, of which the number of indicators has reached 15, and the entire data is disorganized. In view of this situation, we first clustered the data into four categories, and then took the arithmetic mean of the entire data.

(2) Select K-means method

There are many algorithms in the cluster analysis method. After combining the data scale and the actual situation, the K-means method was finally selected.

(3) Analysis of the results of cluster analysis

After integrating the effective data, it is concluded that there are problems in the management method of the school, which brings accurate help to the counselors' ideological and political education work.

(4) Dynamic early warning system

The workload of thinking and Political Education is not only very complicated, but also has a lot of aspects. It should be handled by counselors in accordance with the urgency of the work and in different hierarchies. Before that, we can have a different hierarchy according to the importance of the issue. The higher the classification, the more urgent the matter should be. It is important to define priorities and solve problems quickly before things have gotten worse.

In this paper, data extraction is used with a cluster assessment, and an early warning system for thinking and policy education work is being developed based on the conclusions obtained. The apply of this system will provide an enlivening effect for the thinking and politic education work in varsity.Data mining algorithms play an important role in the Internet of Things environment, which is conducive to the rapid progress of artificial intelligence and science and technology.The dynamic nature of sports activities provides real behavioral data for ideological and political education. By analyzing sports performance, team interaction and on-field behavior, it is possible to more accurately capture students' ideological dynamics. This data mining method based on sports scenarios provides a new practical path for ideological and political education.

Funding

Ministry of Education Industry-University Cooperative Education Project, Project name: The promotion and application of smart sports in the teaching of physical education in application-oriented colleges and universities (Project number: 230802130044522).

Reference

[1] Zhang Y, Yan X, Zhao Y D, Intelligent Early Warning of Power System Dynamic Insecurity Risk: Toward Optimal Accuracy-Earliness Tradeoff. IEEE Transactions on Industrial Informatics, 2017, 13(5):2544-2554.

[2] Song W. Building an Early Warning System for Crude Oil Price Using Neural Network. Social Science Electronic Publishing, 2017, 14(2):79-109.

[3] Kong Y, Xie C, Zheng S. Dynamic Early Warning Method for Major Hazard Installation Systems in Chemical Industrial Park[J]. Complexity, 2019, 2019(11):1-18.

[4] Xu L, Jiang C, Wang J, Information Security in Big Data: Privacy and Data Mining. IEEE Access, 2017, 2(2):1149-1176.

[5] Chaurasia V, Pal S. A Novel Approach for Breast Cancer Detection using Data Mining Techniques. Social Science Electronic Publishing, 2017, 3297(1):2320-9801.

[6] Yan X S, Zheng L. Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach. Review of Financial Studies, 2017, 30(4):1382-1423.

[7] Lei, Lin, Qian, Data Mining and Complex Network Algorithms for Traffic Accident Analysis: Transportation Research Record, 2018, 2460(1):128-136.

[8] Ilmarinen P, Tuomisto L E, O Niemelä, Cluster Analysis on Longitudinal Data of Patients With Adult-Onset Asthma. The Journal of Allergy and Clinical Immunology: In Practice, 2017, 5(4):967-978.

[9] Berka P. Comprehensive concept description based on association rules: A meta-learning approach. Intelligent data analysis, 2018, 22(2):325-344.

[10] Atta-ur-Rahman, Da Sh S. Data Mining for Student's Trends Analysis Using Apriori Algorithm. International Journal of Control Theory and Applications, 2017, 10(18):107-115.

[11] Castillo-Rojas W, Medina-Quispe F, Vega-Damke J. New Visualization Scheme for Cluster Models in Data Mining. RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao, 2017, 2017(21):67-84.

[12] Feng X D, Xuming H E. Robust low-rank data matrix approximations. Science China Mathematics, 2017, 60(002):189-200.

[13] Rahman A, Oldford R W. Euclidean distance matrix completion and point configurations from the minimal spanning tree. Siam Journal on Optimization, 2018, 28(1):528-550.

[14] Li X, Liang W, Zhang X, A cluster validity evaluation method for dynamically determining the near-optimal number of clusters. Soft Computing, 2020, 24(12):9227-9241.

[15] X Zhang, Li R, Hou W, Research on Manhattan Distance Based Trust Management in Vehicular Ad Hoc Network. Security and Communication Networks, 2021, 2021(2):1-13.

[16] Xu D G, Zhao P L, Yang C H, A Novel Minkowski-distance-based Consensus Clustering Algorithm. International Journal of Automation and Computing, 2017, 14(1):33-44.

[17] Lee Y, Na J, Lee W B. Robust design of ambient-air vaporizer based on time-series clustering. Computers & Chemical Engineering, 2018, 118(OCT.4):236-247.

[18] Somme S. A Note on Probabilistic Convergence Defined by Unbounded Modular Functions and αβ-Statistical Convergence. Kragujevac Journal of Mathematics, 2021, 45(1): 127-138.

[19] Doan T T, Subhonmesh B, Hoa N D, Convergence of the Iterates in Mirror Descent Methods. IEEE Control Systems Letters, 2018, 3(1):114-119.

[20] YiSu, Yu Y Q. Dynamic early warning of regional atmospheric environmental carrying capacity. The Science of the Total Environment, 2020, 714(Apr.20):136684.1-136684.9.