Management of Ideological and Political Education of College Based on Data Mining

Ge Chen[1,a], Kangsheng Hu[2,b], Chaofeng Huang[3,c]*

[a]Earrychen@163.com, [b]289981534@qq.com, [c]*Corresponding author: Earrychen@163.com

[1]National University of Defense Technologies Changsha, China

[2]Shijiazhuang Campus of Army Engineer University Shijiazhuang, China

[3]National University of Defense Technology Changsha, China

Abstract—The evaluation method of work assessment quantification table is suitable for vertically measuring the mastery degree of ideological and political education administrators to the students they manage. The information contained in this method is obviously insufficient, and there are some defects in the objectivity and accuracy of the evaluation results. It is invalid and inappropriate to evaluate ideological and political educators according to traditional analysis methods. This paper takes the management of ideological and political education in colleges and universities as the research object, combines data mining technology, adopts the idea of cluster analysis, and uses k-means algorithm to automatically analyze and mine a large number of evaluation data in the "work evaluation quantification table", and verifies it with real data, Experiments show that this method can effectively overcome the shortcomings of traditional analysis methods.

Keywords: Data Mining; Educational management; K-means algorithm

1 INTRODUCTION

Education of politics and ideology education are receiving an increasing amount of focus from both society and educational institutions as a result of advancements made in the creation of ideologies and political systems. Especially in 2017, curriculum ideology and politics were highly valued by the central government, and the important curriculum of ‘Opinions on deepening the Education System Reform and Mechanisms’ began to rise from local curriculum and curriculum reform exploration to national practice exploration and got successful results.

In 2018, the nationwide promotion of the ideological and political curriculum has further deepened the process and effect of the ideological and political curriculum promotion, and the ideological and political curriculum has also become a research hotspot, and a feasible optimization strategy is proposed. In the context of artificial intelligence and big data, the emergence of cloud computing, and intelligent algorithms, a revolution is taking place in the educational field, where traditional educational methods are constantly being, methods of high-

ITEI 2022, December 23-25, Harbin, People's Republic of China

DOI 10.4108/eai.23-12-2022.2329086

tech education. The intellectual and political framework of the curriculum should be coupled with artificial intelligence and big data foundation.

This research recommends utilizing recent information technologies and data mining to effectively leverage the political and ideological knowledge of educational advisors and students, so as to strengthen the ideological and political work of colleges and universities through Internet-based reasoning.

2 MODEL BASED CLUSTERING METHOD

Cluster analysis is a process of grouping data sets of fictitious or physical data objects. Its purpose is to make data sets of similar data objects. These data object groups generated during cluster analysis are called clusters, which are a collection of data objects. Objects in the same cluster have high similarity to each other, while objects in different clusters are different from each other. After clustering analysis, researchers can determine the dense and sparse range, thereby finding the overall architecture framework and meaningful correlation between various attributes.

The model-based clustering method tries to optimize the fit between the given data and the basic mathematical model. It is often based on the assumption that the data are generated according to the potential probability distribution. The most classical model-based clustering method is the COBWEB algorithm proposed by Fisher.

COBWEB algorithm (widely adopted non-complex incremental concept analysis method). The input data points of the algorithm are represented using attribute-value pairs. The algorithm builds hierarchical clusters by using a classification tree approach. In this way all data points in this tree represent some concept, containing the probabilistic case of this concept, and the statistical sum of all objects classified to this node. The probabilistic case represents the concept probability as well as the conditional probability of the form

$$P(A_{ij} = V_{ij} / C_k) \qquad (1)$$

Where $A_i = V_{ij} / C_k$ refers to an attribute value pair (i.e., the i-th attribute takes its j-th value likelihood) and C_k refers to the concept class. A new division is generated by a sibling node on some layer of such a tree. The algorithm uses heuristic estimation to calculate the categorization utility, which guides the process of building the classification tree. The categorization utility can be described as:

$$\frac{1}{n}\sum_{k=1}^{n} p(C_k)\left[\sum_i \sum_j p\left(A_i = \frac{V_{ij}}{C_k}\right)^2 - \sum_i \sum_j p\left(A_i = V_{ij}\right)^2\right] \qquad (2)$$

Where n represents the number of nodes, concepts or "categories" in the new partition $\{C_1, C_1, \ldots C_n\}$ generated at a certain level of the classification tree. Similarities within and differences between categories of classified effectiveness feedback:

The probability $P(A_i = V_{ij} / C_k)$ represents similar situations within a class. The larger the P value, the more members of the shared attribute value pair, and the greater the probability that the attribute value pair belongs to a class member.

Probability $P(A_i = V_{ij})$ refers to the difference between classes. The larger the P is, the fewer objects sharing this attribute - value pair's comparison class, and the more likely it is to predict this attribute value pair belongs to class prediction.

COWEB operates by incrementally adding objects to the classification tree. It first updates the count according to an appropriate path in the tree, and queries the "best host" or node of this object. This decision is to temporarily put the object into all nodes and calculate the classification effect of the result division. Among them, the specific positioning of the best classification effect can be regarded as a good host of objects[1].

3 MINING STEPS OF CLUSTERING TECHNOLOGY

Cluster analysis divides physical data objects into sets of data objects to partition similar data. Groups of data objects generated by data analysis are called clusters. The object data in the cluster has a high degree of similarity.

- The first step is using data mining to clarify the data. The very first phase in the data mining process is one of the most important and requires a very clear description of leap questions. This is done to establish the goal of data mining. The study challenge can be anticipated via data mining, even though it is unable to forecast the conclusion. A process of data mining that is carried out in a haphazard manner will result in erroneous results.

- The second step is data acquisition. This process is arduous and takes a significant amount of time. Only a portion of the information may be obtained from direct observation, while the remaining can be obtained through research.

- The third step is data pre-processing. The technique consists of changing the collected data set into an analyzable data model, which is constructed in accordance with the algorithm. The constraints that an algorithm places on a data model can vary greatly from one algorithm to the next.

- The fourth step is data clustering mining. The purpose of mining data using clustering is to organize the model of the data into collections of items that share characteristics with one another. The procedure primarily entails the data model input process, as well as the selection of the clustering method and its subsequent implementation.

- The fifth step is cluster result analysis. The study of the clustering data mining results obtained from the class that has numerous group attributes is the purpose of this process.

- The sixth step is to apply the knowledge. The process entails incorporating the useful information gleaned from the research into the education that the instructor of management receives. This is done so that the instructor may make use of the conclusions to promote teaching management and construct management strategies effectively [2].

Figure 1 is a flowchart of the data mining process

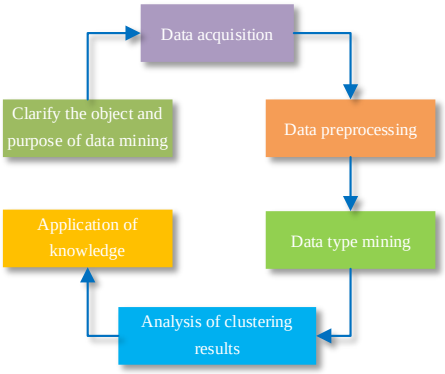

Figure 1. The data mining process

4 THE KEY ALGORITHM OF DATA MINING -- K-MEANS ALGORITHM

The following is an outline of the basic steps involved in clustering analysis: The input of clustering analysis consists of application data that must be processed, which are then aggregated into a sample set. The sample sets are preprocessed and selected or extracted based on the sample data objects' properties. The sample representation is obtained based on the data type of the sample data set and the requirements of cluster analysis. The cluster is then categorized according to sample similarity. In general, cluster analysis is a molding process. Based on the similarity of the samples, the data objects inside the result set are extremely similar. Various outcomes are concentrated in the data. The similarity between objects is minimal. The final clustering analysis returns the sample clusters obtained by different clustering algorithms[3].

Figure 2 demonstrates the cluster analysis method.

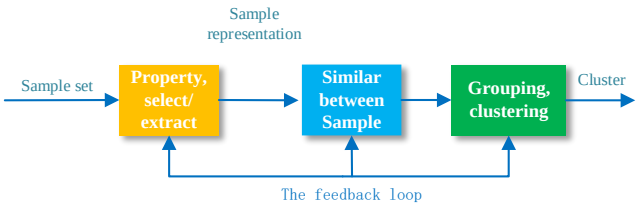

Figure 2. Clustering Process

What is called the "k-meter algorithm" means that you can average the distance between objects in each group to make a center or center get that distance. The basic actions are as follows: first, choose an object with a K center or a complete one; Secondly, according to the above steps, the distance between the remains and a set of K objects is calculated, and the results are concentrated

in the shortest unit. Finally, the weight or focus of the new cell can be calculated using this algorithm. Repeat performing this step until the default function has attained the minimal value of the target function. Throughout many cases, the include expression is utilized by the default function

$$E = \sum_{i=1}^{k} \sum_{p \in c_i} |P - m_i|^2 \quad (3)$$

Here, E represents the degree of error 2 in the size of all objects in space. The P are all meeting points. m_0 is the center of information, and in this case, P and m_i are multi-dimensional.

The goal of the algorithm is to minimize E by constantly repeating the calculation. The smaller the object, the smaller the distance and the higher the similarity. The more data, the more efficient the algorithm.

The flow of the algorithm implementation is shown in figure 3.

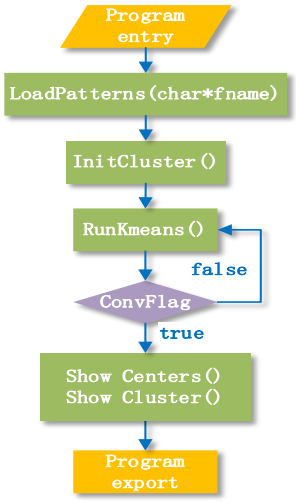

Figure 3 Flow chart of algorithm K-Means

The function known as LoadPatterns (char×fname) is used in the K-means method to load the sample data information into the program from dat, read out the pertinent information, and transform the data in the file into a sample Patten array. This is done by reading out the relevant information[4].

The cluster center is first initialized using the method InitClusters, and then it is set to be the first K data in the data sample. This is done so that the beginning cluster center can be determined.

The algorithm's main program, Run K-Means(), compares the distance between all items and the center of each cluster. The item is then subdivided into the cluster that is closest to the cluster's center, and the cluster's center is computed based on the new division. When this happens, if the cluster center has not moved, the clustering procedure is complete.

Function ShowCenters represents the cluster center defined by the algorithm. Function ShowCenters also provides the identifier of the sample provided by the algorithm. ConvFlag is used as an identifier to determine whether or not the cluster is complete.

5 THE IMPLEMENTATION OF CLUSTER ANALYSIS IN THE ADMINISTRATION OF IDEOLOGICAL AND POLITICAL EDUCATION

5.1 Data collection

Fifty quantitative evaluation tables of teachers' work in the faculty of Materials and Chemical Engineering were selected.

The first is the attitude of management. The management attitude mainly includes reward and punishment at work, treating each student objectively; Getting along well with classmates; Well said, setting a good example; Honesty, and discipline [5].

Second, managerial competence. The so-called management ability refers to the capacity to deal with problems in a certain employment, which is generally comprised of the following: a strong ability to organize and manage work, and the ability to efficiently organize pupils to actively engage in school; Accurately comprehend and handle the circumstances of low-income students, and do well in shifts, loans, and other demanding tasks; Student disciplinary infractions are dealt with in a fair and equal manner; Actively organize students to deliver a quality review; Accurately evaluate the conditions of student special leaders and actively help students[6].

Thirdly, management methods. Management technique is the specialized approach to managing students. These include at least three weekly visits to the course or home; Insist on frequent health inspections of student residence halls; Communicating with more than 30% of students every semester; Comply with rules while handling scholarship and bursary-related tasks; Consider the student's life, studies, etc., provide effective ideological direction, and comprehend the student's thinking fallacies; Create a QQ group and a WeChat platform to improve student participation and communication.

5.2 Data Conversion

The first step involves calculating the conversion of numeric data for four distinct properties. Modifying data using sequence variables primarily entails utilizing procedures that are difficult to objectively evaluate. The data can be evaluated objectively if just scores are considered. Thus, "good" and "bad" can be divided into "not good" and "not good". 0.75% of the evaluation was conducted using evaluation methods for various evaluation objectives. Example of management impact = 3(Participation in a seminar + report of the results of the post-school evaluation).

5.3 Data mining

Through data cleansing and preprocessing, as demonstrated in Table 1, it is possible to collect specialized data mining samples.

Table 1 Basic illustration of data mining

Management attitude	Ability to manage	Management effect	Management methods
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
…	…	…	…

5.4 Algorithm implementation

The k-mean algorithm was used to analyze 120 sample data: 3 standard samples representing better, medium and worse respectively and 117 sample data obtained from the quantitative work appraisal form and after data transformation. All the sample data contained four categories of attributes, namely: management attitude, management ability, management method and management effect; data mining clustering was performed on these aspects, and the initial k-value was set to 3. The final mining results are shown in Table 2.

Table 2 Analysis of pooling results

	Attitude	Ability	Methods	Effect	Samples
Good	0.77	0.77	0.74	0.79	36
Medium	0.61	0.57	0.54	0.56	74
Poor	0.31	0.31	0.28	0.30	10

According to the above results, the final proportion distribution range of data samples included in each cluster is as follows:

The first cluster (good) has 36 data samples in total, one predefined standard sample is deleted, and 35 data samples remain, accounting for 35/117=30%.

Cluster 2 (medium), there are 74 samples in total. After deducting one standard sample, there are 73 samples left, accounting for 73/117=62%.

For the third cluster (poor), there are 10 data samples in total. One predefined standard sample data is deleted, and there are 9 data samples left, accounting for 9/117=8%.

The overall scores of individual items are weighted due to the unequal number of samples. The overall scores of the four categories of items are

$$\begin{bmatrix} 0.77 & 0.61 & 0.31 \\ 0.77 & 0.57 & 0.31 \\ 0.74 & 0.54 & 0.28 \\ 0.79 & 0.56 & 0.3 \end{bmatrix} \times \begin{bmatrix} 0.30 \\ 0.62 \\ 0.08 \end{bmatrix} = \begin{bmatrix} 0.6002 \\ 0.5760 \\ 0.5480 \\ 0.5752 \end{bmatrix} \quad (4)$$

Therefore, the above research indicates that the Faculty of Materials and Chemical Engineering students have an ideological and political education score of over 0.5, which corresponds to the high school level. The managerial attitude has the highest value of the four values, indicating that

it is a strong preference for the ideological and political activities of university advisers. All students and those with the lowest grades demonstrate some degree of management strategies, content, and chemistry teachers from diverse departments making extra changes should be more than 30% of students each semester to discuss, visit dormitories, and so on, in order to deepen feelings between students and better comprehend their minds. In addition, WeChat and other new information tools have been ingeniously integrated into this paper, making it an additional vital communication tool for enhancing the connection between teachers and students.

6 CONCLUSION

This paper focuses on the application of clustering method in data mining to the management of ideological and political education in colleges and universities. This paper firstly analyzes the data of "work evaluation quantification form" of college counselors. Then, according to the process of data mining, various data pre-processing is carried out. Finally, the C++ process of clustering mining was implemented by using the k-means algorithm in the division method. Further analysis shows the results of clustering, and finally some conclusions are drawn, which are of practical guidance to improve the management of managers and can effectively overcome the defects and shortcomings of traditional analysis methods. This paper still needs in-depth research in optimizing algorithms to improve the computational speed of mining large-scale data and studying the fusion of heterogeneous data from multiple sources.

REFERENCES

[1] Han J, Kamber M, Fan Ming, Meng Xiaofeng, Data Mining: Concepts and Technologies [M] Beijing: China Machine Press, 2001.

[2] Li Jun. General model of data mining system implementation[J]. Journal of Daqing Petroleum Institute, 27(3): 35-37.

[3] Jacob Kogan, Introduction to Clustering Large and High—Dimensional Data[M], Cambridge University Press, 2006.

[4] Bu Yuanyuan, Guan Zhongren. Research based on K-means clustering algorithm[J]. Journal of Southwest University for nationalities, 2009, 01:1-3 [16].

[5] Stephen W Draper, Margaret I Brown. Increasing interactivity in lectures using an electronic voting system. Journal of computer assisted learning, 20(2):81–94, 2004.

[6] Dewan, F. Lin, D. Wen, M. Murshed, Z. Uddin. A Deep Learning Approach to Detecting Engagement of Online Learners (IEEE International Conference on Internet of People, Guangzhou, 2018).