A Design of College Ideological and Political Education Management System Based on Data Mining Technology

Jinhui Zhao, Jiahui Wang

Hebei Agricultural University, Baoding 071000, China

Abstract

Data mining, one of the key steps in database knowledge discovery, refers to the process of searching hidden information from massive data resources by selecting various algorithms to analyze the information data. In the context of the swift development of big data, the significance of data mining technology has also been fully developed. In the field of education, a ideological and political education management system based on data mining technology can effectively analyze the substantial amount of ideological and political education management information and consequently provide an essential reference for the development of ideological and political education management. However, the application of data mining technology in the management system for ideological and political education in colleges and universities is still in the exploratory stage, and there still exist great differences in technology and application. For this purpose, it is of vital practical significance to study the application of data mining technology in the ideological and political education management in colleges and universities.

Keywords: Data Mining Technology, Ideological and Political Education, Education Management System.

1. RESEARCH OVERVIEW

1.1 Research background

As information technology constantly advances, database management system is increasingly applicable and the amount of data to be processed is also enlarged day by day, which brings both opportunities and a series of challenges to the development of data mining technology. On the other hand, with the continuous expansion of college enrollment in China, the number of undergraduates has reached a peak and the quantity of teachers is expanding at the same time, which produces massive information and data in the education and administration database. Based on its own actual conditions, each university sets up a corresponding education management system that generally has indirect operation and powerful functions, effectively analyzes the data of ideological and political education (IPE) in colleges and universities and makes up for the loopholes in the education management system. However, since the application of data mining technology in college IPE management is still in the exploratory stage, traditional manual analysis and calculation are still adopted in a number of aspects and it is difficult to exert the vital significance of the IPE management system. Therefore, it is necessary to conduct an in-depth study on data mining technology and integrate it with college IPE management system, which provides a reference for IPE management and effectively improves the level of college IPE management.

1.2 Literature review

IPE management is a vital component of the college education management database and also an important approach to promote the normal development of ideological and political work and ensure the quality of ideological and political management. How to carry out IPE management accurately, scientifically, fairly and impartially and to improve the level of college IPE management is one of the key issues that industry experts and scholars attach great importance to. With the continuous reform and development of the management system of higher education, especially after counselors undertake IPE and management of students, the evaluation methods of IPE in colleges and universities have been transformed from the traditional ways such as meeting and voting to applying emerging technologies such as data mining, reflecting the overall trend of quantitative development (Liao et al., 2011). At the earliest, clustering analysis is a key statistical analysis. With the continuous progress of machine learning, clustering analysis is gradually evolved into a key example of non-guided learning as well as an important pretreatment step of data mining technology. These algorithms are then processed on the basis of
generated classes, which can effectively enhance the efficiency and level of data mining techniques. For this purpose, clustering analysis has received extensive attention and numerous research perspectives, such as pattern recognition, data analysis, image processing, market research, etc., and has possessed a broad development prospect and an application market (Huang, 2017). The traditional evaluation method of college IPE management is mainly based on quantitative evaluation on managers’ performance appraisal, which is more suitable for horizontally comparing the management level of IPE administrators to students, but is largely defective in case of comparing the ability and level of different IPE managers or measuring the difference between a certain IPE manager and the average level of all IPE managers. The application of data mining technology is able to unearth the key hidden information from abundant data and is of great significance for evaluating the working performance of college IPE administrators (Qi and Zhang, 2017).

2. CLUSTERING ANALYSIS OF DATA MINING TECHNOLOGY

2.1 Overview of clustering analysis technology

Clustering analysis technology refers to the process of dividing and assembling imagined or physicalized data and its core purpose is to form a data set with similar data objects. Specifically, a group of data objects is also referred to as a cluster. Individuals in the same cluster tend to have relatively high similarity, and individuals in different clusters tend to have significant differences. After clustering analysis, researchers can eventually conclude the relevance between the overall framework and various attributes on basis of the characteristics of denseness and coefficients (Feng and Yin, 2017). Clustering analysis technology has the following features (Figure 1),

Figure 1. Main Features of Cluster Analysis

The first feature refers to scalability. The traditional clustering analysis algorithms mainly target small data sets. However, with the continuous enlargement of the database scale, the traditional clustering analysis algorithms produce some discrepancies, but the extension on basis of it can consequently meet the needs of the database (Yang et al., 2016).

The second feature involves the ability to handle different types of attributes. In the context of rapid social development, the content of the traditional clustering analysis algorithms tends to expand. Many other types or attributes of data can still adopt clustering algorithms, such as binary type, classification type, ordinal type, mixed type, or the mixture of the above types of data.

The third feature is to identify the clusters of arbitrary shape. Clustering analysis algorithm is based on distance measurement, where individuals in each cluster have a certain similarity and the differences among different clusters are significant in terms of shape, density and other aspects.
Fourthly, susceptible input parameters are determined. When a number of clustering algorithms are applied, it is necessary for users to design certain susceptible parameters in advance, such as the number of generated clusters, and these input parameters will have a profound impact on the final result. Accordingly, higher requirements for the professional abilities are put forward. If deviation occurs to the parameters input by users, the accuracy of clustering algorithm is difficult to be guaranteed.

The fifth feature is the ability to deal with noisy data. In actual processing of data, more outliers or isolated points are usually generated. The analysis of these data will have a negative impact on the quality of the analysis. To this end, clustering analysis algorithm often need to isolate and eliminate these low quality data to ensure the quality of analysis.

Sixthly, after adding new data to many clustering algorithms, overall re-calculation is required, and the new data cannot be directly added into the existing data set. Some other clustering algorithms need to input data information in a certain order. If the inputting order deviates, the result would be greatly different (Wang and Wang, 2016).

The seventh feature is the high-dimensional nature. Humans are able to process information in a three-dimensional environment. In fact, a large quantity of information and data is higher-dimensional. Accordingly, clustering analysis algorithm is required to possess the characteristic of high-dimension so as to meet the more diversified data analysis demands.

The eighth feature is the limitation of constraints. In practice, clustering algorithm is limited by external constraints. For example, innovation and entrepreneurship activities require investigation on market environment, relevant policies, moral quality, honesty, rent, etc., which are also the constraints (Cui and Zhao, 2016).

The ninth feature concerns interpretability. Users’ core idea of applying clustering algorithms is to obtain valuable information through the key clustering algorithm and accordingly provide a reference for its own activities. Hence, the results of clustering analysis algorithm should be explanatory and available.

2.2 Data types in clustering analysis

It is assumed that a severe analysis needs to deal with n data objects and the n objects refer to individuals of many different attributes and different types. Hence, the following two types of matrices can be constructed:

The first type is data matrix, which concerns n objects. Variable p is selected for description and is constructed as a matrix, namely the $n \times p$ matrix. The matrix is:

$$
\begin{bmatrix}
  x_{11} & x_{12} & \ldots & x_{1p} \\
  x_{21} & x_{22} & \ldots & x_{2p} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{n1} & x_{n2} & \ldots & x_{np}
\end{bmatrix}
$$

(1)

The second type is dissimilarity matrix. The similarity between each two objects of the n objects is judged and constructed into the $n \times n$ matrix. The matrix is indicated as below:

$$
\begin{bmatrix}
  0 & d(2,1) & 0 & \ldots & 0 \\
  d(3,1) & d(3,2) & 0 & \ldots & 0 \\
  \vdots & \vdots & \ddots & \ddots & \vdots \\
  d(n,1) & d(n,2) & \ldots & 0
\end{bmatrix}
$$

(2)

where $d(i, j)$ refers to the similarity between two objects: i and j. In general, $d(i, j) \geq 0$. When the degree of similarity between i and j is high, the value of $d(i, j)$ is small and infinitely approaches 0. When the degree of similarity between i and j is low, the value of $d(i, j)$ infinitely approaches positive infinity (Gao, 2016). Therefore, $d(i, j)=d(j, i)$ and $d(i, j)=0$. 
In practice, the application range of a dissimilarity matrix is more extensive. However, in some special cases, data is described by data matrix. Thus, a data matrix needs to be transformed into a dissimilarity matrix. The method is illustrated in the following part.

2.2.1 Interval-scaled variable

The most classic calculation method of interval-scaled variable is Euclidean distance calculation method, and its definition formula is as follows:

$$d(i,j) = \sqrt{(x_{i1}-x_{j1})^2 + (x_{i2}-x_{j2})^2 + \cdots + (x_{in}-x_{jn})^2}$$  \hspace{1cm} (3)

In addition, another classic distance measurement method is Manhattan distance method. The definition formula is:

$$d(i,j) = |x_{i1}-x_{j1}| + |x_{i2}-x_{j2}| + \cdots + |x_{in}-x_{jn}|$$  \hspace{1cm} (4)

Both formulas must satisfy the following conditions. First, $d(i,j) \geq 0$, and i and j are nonnegative values. Secondly, $d(i,j) = 0$, indicating that the distance between themselves is zero. Thirdly, $d(i,j) = d(j,i)$, namely the symmetry of the function presented between i and j. Fourthly, $d(i,j) \leq d(i,h) + d(h,j)$, demonstrating that the distance between i and j is less than or equal to the distance where other object k reaches object j (Ma, 2016).

In terms of the above two formulas, Minkowski put forward its promotion formula, and the definition formula is as below:

$$d(i,j) = \|x_i - x_j\|_p = (\sum_{k=1}^{n}|x_{ik} - x_{jk}|^p)^{1/p}$$  \hspace{1cm} (5)

where $p \in [1, \infty)$. This method highly generalizes Euclidean distance calculation and Manhattan distance calculation. When p is 1, $d(i,j)$ represents Manhattan distance calculation method. When p equals 2, $d(i,j)$ represents Euclidean distance calculation method (Wang and Xu, 2016).

2.2.2 Binary variables

In the perspective of binary variables, each variable contains two states. One state is 0, namely nonexistence or empty, and the other state is 1, namely existence or occurrence. Under this method, if the traditional function formula is still adopted to compute interval-scaled variables, certain errors would inevitably occur. Hence, a study on the dissimilarity of binary variables is a necessity. It is assumed that each weight is the same in the binary vector. The following table is derived:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>q</td>
<td>r</td>
<td>Q + r</td>
</tr>
<tr>
<td>0</td>
<td>s</td>
<td>t</td>
<td>S + t</td>
</tr>
<tr>
<td>Total</td>
<td>Q + s</td>
<td>R + t</td>
<td>p</td>
</tr>
</tbody>
</table>

In terms of binary dissimilarity, the dissimilarity between i and j is usually expressed by the following formula:

$$d(i,j) = (r + s)/(q + r + s + t)$$  \hspace{1cm} (6)

If the importance, i.e. the weight, reflected by 0 and 1 is inconsistent, the binary variables can be called asymmetric binary variables. For example, in college IPE management, 1 indicates when managers have a high management level and students’ ideological and political level is raised to a certain extent. 0 refers to when the management level of managers is low and the IPE level of students is not improved and even is reduced (Xu, 2016). Asymmetric binary variable dissimilarity can be calculated using the jaccard coefficient, and its formula is as follows:
\[ d(i, j) = \frac{(r + s)}{(q + r + s)} \]  

For example, when colleges conduct surveys on IPE management, among information of student’ names, gender, age, hobby, grades, etc., only gender is a binary variable; name is the object identification symbol; the rest are asymmetric binary variables (Wu and Zhang, 2017).

3. COLLEGE IPE MANAGEMENT SYSTEM BASED ON DATA MINING TECHNOLOGY

3.1 The application of data mining technology in the CCP construction in colleges and universities

CCP construction is a fundamental basis for colleges and universities to carry out IPE. An outstanding CCP construction cadre team serves as a battle bulwark and cultivates more outstanding student CCP members with high ideological and moral values and firm political positions. In the work of CCP construction, the selection of CCP cadre personnel in colleges and universities becomes even more prominent. To this end, data mining techniques can be employed to analyze teachers’ teaching ability, teaching level, work enthusiasm, quality level and other indexes and to further select qualified talents with various acceptable indexes. In the meantime, it is also possible to excavate and educate the activists who join CCP through data mining technology, thereby ensuring the advancement and purity of CCP in colleges and universities (Zhang, 2014).

3.2 The application of data mining technology in dormitory management

Data mining technology can effectively analyze students’ study-rest time, living habits, learning habits and other indexes, which provides important a reference for scientific dormitory allocation and can satisfy each student’s personalized accommodation and living environment to the maximum extent. At the same time, students’ access to the Internet has always been a key issue that has plagued the dormitory management. Excessive online surfing exerts a negative impact on students’ learning and living. However, students should not be blindly opposed to using the Internet. Managers should understand students’ purposes of the Internet, deeply analyze their psychological conditions, help students get rid of Internet addiction and engage in learning with more enthusiasm.

3.3 The application of data mining technology in student employment

Promoting student employment is the focus of college teaching management. Influenced by many subjective or objective factors, the employment rate of college graduates is low and finding jobs is difficult. Enterprises also face the problem of shortage of talents. The main reason is the major gap between talents cultivated by colleges and the demands of enterprises. Therefore, through data mining technology, we can analyze the demand direction of talents for enterprises as well as each student’s IQ, EQ, employment preference, working ability, spirit of adventure, innovative IQ, physical condition, etc. By combining these two aspects of consideration, a more reasonable teaching mode is designed for students so as to train more excellent compound high-quality talents for enterprises, which is of great significance to realizing students’ individual value and promoting the all-round development of enterprises and the society (Liu and Ding, 2015).

3.4 The application of data mining technology in comprehensive evaluation

Comprehensive evaluation can intuitively reflect students’ learning level and ideological and ethical conditions and clearly identify the defects of the current IPE management in schools, which is of vital significance for the development of targeted training programs and the promotion of the overall level of student learning and social development. Therefore, the application of data mining technology first requires tracking information and enterprise feedback of graduates. Besides, the actual self-development of students is combined to conduct an in-depth comprehensive evaluation through data mining technology, so as to understand the defects of student IPE management, to propose countermeasures, and to effectively improve the level of college IPE management. In addition, by means of clustering analysis, correlation analysis and other algorithms, we are able to identify the hidden information in the complex data, provide a decision basis for college IPE management and drive the leap-forward development of the college IPE management system (Fan and Han, 2015).

REFERENCES


Huang J. (2017). Using Internet thinking to innovate the basic path of Ideological and political education for young officers and soldiers, Journal of Xi'an Political College, 30 (01), 71-75.


