International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 1423-1435

https://doi.org/10.70517/ijhsa463110

A Study on Mining the Potential Value of Association Rule Mining Technology in Civic Education Network Data Collection

Peipei Cheng[1,*]
[1] Nantong Institute of Technology, Nantong, Jiangsu, 226001, China
Corresponding authors: (e-mail: 17318828392@163.com).

Abstract Establishing moral character is the fundamental task of education and the foundation of colleges and universities. Combined with the characteristics and requirements of specialized disciplines, mining the potential value of class Civics education is the main method of Civics education. In this study, after collecting and pre-processing the civic and political network data of different colleges and universities, we use the TF-IDF algorithm and the LDA theme model to conduct multi-dimensional mining of the potential value of civic and political education themes, and then use the Apriori algorithm to make its potential value of the association rules visualized. The mining results show that there are 7 major themes of potential value themes of Civic and Political Education, which are Red Culture, Online Service, Competition, Real-time Exercise, Cultural Promotion, Moral Literacy, and Knowledge Learning. Based on the analysis results, corresponding management countermeasures are proposed from the three levels of spirit, education and practice, aiming to improve the management level of Civic and Political Education.

Index Terms civic politics network data, TF-IDF algorithm, LDA topic model, association rules, Apriori algorithm

I. Introduction

With the rapid development of science and technology, the times are changing, the rapid development of cloud computing, artificial intelligence, the Internet of Things and other technologies, so that the concept of “data” has quietly penetrated into various industries and fields, and the impact of “data” on people has become more and more extensive and diversified, and the field of ideological and political education in colleges and universities is no exception. The influence of “data” on people is also more and more extensive and diversified, and the field of ideological and political education in colleges and universities is no exception. The arrival of the era of big data has had a great impact on the teaching process, teaching evaluation, teaching methods and teaching effect of colleges and universities, and there are great challenges while promoting the development of high quality of ideological and political education [1], [2]. In this context, in order to realize the high-quality development of Civic and Political Education, it is necessary to realize the change of educators' concepts and the rise of skill level, and to be a Civic and Political Educator of Colleges and Universities who keeps up with the development of the times [3], [4].

However, the ideological dynamics of Civic Education as well as the teaching effect are more abstract, and it is difficult to effectively and directly quantify them by means of big data, and the main difficulty is reflected in the operability of the collection of data that can truly reflect the ideology of students [5]-[7]. Although big data technology claims to be able to quantify everything, it is never easy to quantify the human mind, and there is a certain degree of difficulty in mining data on students' ideological dynamics [8], [9]. Although it is relatively easy to obtain structural behavioral data such as students' academic performance, online learning, and campus consumption status, it is difficult to mine and control the ideological situation about students' behavioral motives, emotional changes and preferences [10]-[12]. Therefore, based on the characteristics of data mining and analyzing technology, we can deeply excavate the potential educational value and elements of ideology in Civic Education, so as to strengthen the quality of Civic Education teaching with the fundamental task of “cultivating morality and educating people” [13], [14].

In this paper, we firstly collect the online data of Civic and Political Education from different universities based on the data mining process, and pre-process the (unstructured) data. Further, this paper applies TF-IDF algorithm and LDA topic model to text mining of Civic and Political Education network data, extracts keywords and topics, and conducts visual display and in-depth analysis to analyze the potential value of Civic and Political Education from multiple dimensions, which goes beyond the previous limitation of analyzing only from the vocabulary level. Finally, the Apriori algorithm was used to mine the association rules of the potential value of Civic and Political Education, and visualized and analyzed in terms of support and other dimensions, which revealed the complex relationship between the potential values of Civic and Political Education, and provided a visualization tool for Civic and Political Education.

II. Civic education network data mining and pre-processing

II. A. Civic education network data mining process

Data mining is a complete process, the process of mining from large databases unknown, effective, practical application of knowledge and information, and use this knowledge for decision-making, etc. Practical applications of Chinese medicine data mining process is basically the same as this process, as shown in Figure 1.

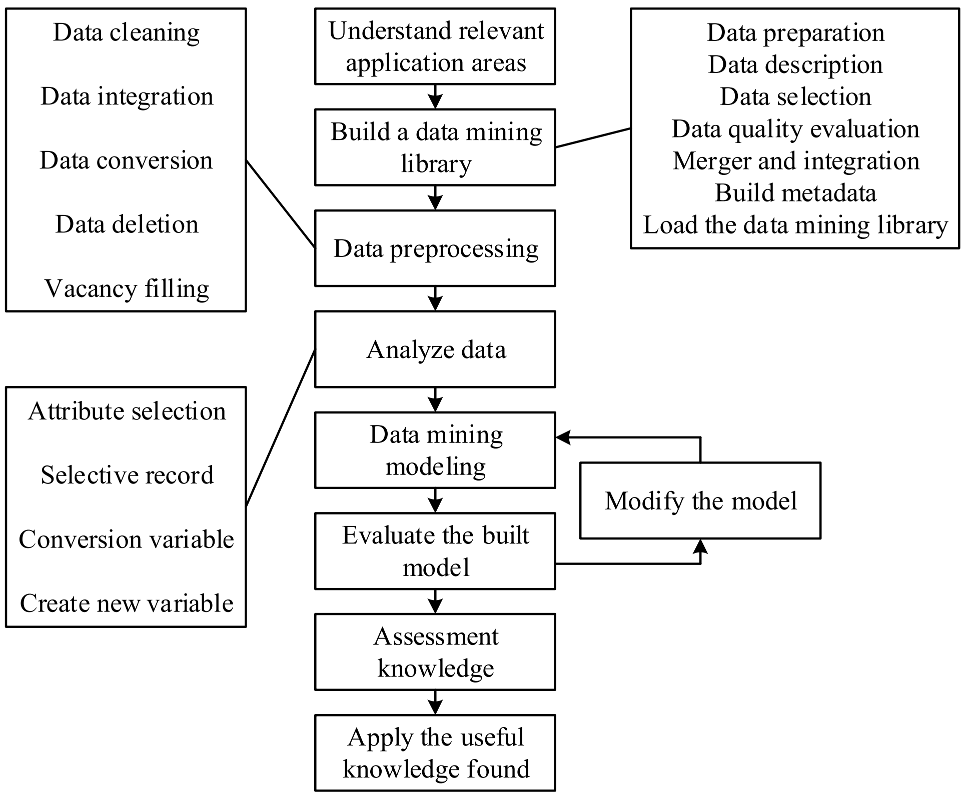

Figure 1: Data mining process

(1) Understanding the relevant application areas includes identifying the problems and objectives of Chinese medicine research, familiarizing with the mainstream methods for solving such problems understanding the corresponding medical domain knowledge, identifying the objectives of medical data mining and the evaluation criteria for the conclusions, and so on three aspects, in which the main emphasis is on the understanding of the medical business

(2) Establishment of data mining library: the data here can be extracted from the relevant Chinese medicine data warehouse, or you can establish your own number mining collection system for data collection, or you can directly read directly from external data sources.

(3) Data preprocessing: Due to the diversity, incompleteness, redundancy, heterogeneity and other characteristics of TCM data, it is necessary to carry out corresponding data preprocessing for various types of data before data mining. Data preprocessing methods mainly include data clearing, data conversion, data reduction, and filling of vacant values.

(4) Analyze the data: initially determine the number of batches of mining needs of the number of batches of attributes and record sets, and from the original number of plus set of small-scale sampling of the initial mining test, if necessary, the attributes should be transformed or create new attributes. This step is slightly different from general data mining in that, due to the diversity and incompleteness of medical data, a small-scale test of the selected dataset is required to determine its usability.

(5) Data mining modeling: this step includes the selection of mining algorithms, the determination of training and testing procedures, and the initial building of the model.

(6) Evaluating the built model: the built model is evaluated using appropriate evaluation methods, and if the model has a big difference with the actual system, the model needs to be improved until the model is closer to the actual system.

(7) Evaluating the knowledge obtained: the conclusions obtained from the excavation are interpreted in Chinese medicine and compared with the initial research objectives. If the results are unsatisfactory, trace back the possible errors less plunge in the mining process and find its solution.

(8) Applying the useful knowledge found: applying the knowledge obtained from the excavation to clinical practice. It should be emphasized here that out of the possibility of affecting the physical and mental health of patients, relevant medical experts must be involved in the process of TCM data mining application, mainly responsible for analyzing and interpreting the medical meaning of the data mining results.

II. B.Data acquisition

II. B. 1) Acquisition process

The data source of this paper is mainly the learning data of a university's Civics online course as well as coursework grades, which were collected from March 1, 2024 to June 15, 2024. The data acquisition in this paper is mainly through web crawler. The main scripting program language used is Python language. The basic steps of web crawling using Python include identifying the target, initiating the http request, obtaining the response response content, parsing the obtained web pages, and storing the obtained data.

The difficulties encountered in this process of crawling data and information are mainly the settings of the course website anti-crawler mechanism, the so-called anti-crawler mechanism is a series of technical means adopted by the website to prevent the crawler program (automated program) from over-accessing and crawling the website data. Its purpose is to protect the normal operation of the website, prevent illegal access to data, reduce server pressure, as well as maintain data security and privacy.

II. B. 2) Collection results

The form of the results of the data collection includes several fields: establishing correct values, promoting all-round development, enhancing civic awareness, realizing value perceptions, fostering the spirit of innovation, and cultivating practical skills.

II. C.Data pre-processing

II. C. 1) Structured data preprocessing

Due to the intersectionality of the data and the differences in the form of different data, the collected data are cleaned, de-emphasized and other operations are performed, and unified processing is carried out, and the collected data are set up with unified presentation fields for value, three views, spirituality, learning, development, awareness, and so on. The above use of Python's Scrapy package and Xpath positioning rules can successfully crawl the desired data, but due to the differences in the design of different websites, it is still necessary to carry out some pre-processing of data to facilitate the subsequent descriptive analysis. As the website mixes the spirit and three views fields together, the crawler is not able to directly distinguish between them, so this part of the job requirements field using Python to write a program to intercept the three views of the potential value of the keywords of the subsequent text content as the subsequent analysis.

II. C. 2) Preprocessing of unstructured data

In this paper, the dictionary-based segmentation method is selected because it is currently more widely used and the technology is more mature, which makes it simpler to implement. Due to the existence of a small number of specialized words in the three-view potential value field, the author needs to expand the dictionary. The principle of expansion is to expand the keywords that appear more frequently in the potential value of the three views into the dictionary because most of them belong to the professional vocabulary, using the jieba package in Python, the TF-IDF algorithm in it to obtain the keywords of each potential value of the three views, and then counting the word frequency, and the keywords with a frequency of more than 1,000 will be selected and expanded into the dictionary. After the word separation operation, we also need to carry out a deactivation operation, the so-called deactivation words are words without practical significance may recur, such as “the”, “had”, “well” and so on. We use the deactivation table to remove the deactivated words in the word division results, all the words appearing in the deactivation table will be removed, so that all the related data processing work has been completed.

III. Mining model of the potential value of Civic Education network data

III. A. Text feature extraction and topic modeling

III. A. 1) TF-IDF based potential value feature extraction

Text feature extraction is the operation of filtering out keywords that can represent the main information of the text from a large number of text words after text preprocessing and word separation. The usual method is to score all the feature items by using some evaluation function, and then sort the scores from high to low, and then select a few items with higher scores as feature items. Feature selection mainly includes methods such as TF-IDF, mutual information, information gain and chi-square statistics.

TF indicates the ratio of the number of times a word is found in the sample data to the total number of words in the sample data. By setting a certain threshold, words below the threshold can be filtered out, so that the text can be filtered out to a certain extent to compare the important feature words [15]. IDF denotes the logarithm of the ratio of the total number of messages to the number of messages containing certain words, and TF-IDF is calculated as shown in equation (1):

$$TF-IDF=TF\times\log\frac{N}{N} \tag{1}$$

In equation (1), N represents the total number of messages in the archive and N' represents the total number of messages containing words in the archive.

III. A. 2) LDA-based Thematic Modeling of Potential Value

Topic models are mainly used for semantic analysis and text mining problems in natural language processing, such as collection and categorization. Topic models are statistical models that group the semantic structure of content in an unsupervised learning manner. LDA based text classification is to be predicted text after the same preprocessing as the training corpus using the trained model, the document-topic matrix can be calculated, i.e., what is the probability that a document to be categorized belongs to which topic is carried out, and according to the probability that a document belongs to which topic is large, the document is classified under that topic.

LDA can calculate the probability of a document belonging to each topic and the occurrence of a certain word can be speculated on how likely the text belongs to a certain topic.LDA believes that each document in the document set is viewed as consisting of multiple topics, and each topic consists of many words. The process of constructing a document is to select a certain topic with a certain probability distribution, and then select a certain word with a certain probability from that topic, and keep repeating each word of the document until the whole document is generated. Document to topic and topic to word obey a certain polynomial probability distribution:

1) For each word in the current document, a topic number is randomly initialized.

2) Using the GibbsSampling formula, resample each word for its topic.

3) Repeat the above process until GibbsSampling converges.

4) Count the distribution of topics in the document, which is θ_d.

5) According to the topic distribution of the current document, the one with the highest probability is the topic of the document.

III. B. Relevance Mining of Civic Education Data Potential Value

III. B. 1) Concept of Association Rule Mining

Association rule mining is to find out the potential correlation from a large amount of data or objects, to discover frequent patterns or association rules based on feature words from a large-scale text set, and to mine out the connections implied in the textual information for regularity analysis. The related basic concepts are as follows:

(1) Transaction set: each document corresponds to a transaction, for this study each potential value of Civic Education is a transaction, all the documents form the transaction set $D=\{D_1,D_2,\cdots,D_n\}$.

(2) Transaction: a transaction corresponds to a document D_i, a transaction is a collection of items, and the items correspond to the value feature words in the document, i.e., $D_i=\{w_{i1},w_{i2},\cdots,w_{im}\},D_i\subseteq D$.

(3) Item: The item is the element in the transaction, that is, the value feature word obtained after the preprocessing of the potential value text of Civic Education.

(4) Item set: item set is a collection of a number of items, for $T=\{w_1,w_2,\cdots,w_m\}$ is the collection of all the risky feature words in the document set D, and any subset X of T is called an item set, and X is said to be a k-item set if $|X|=k$; a document D_i is said to contain the itemset X if $X\subseteq D_i$.

With the above basic concepts explained, the association rule can be defined as:

(5) Association rule: An association rule is an implication of the form $X\Rightarrow Y$. For the potential value of ideological and political education feature item sets X and Y, and $X\subset T,Y\subset T,X\bigcap Y=\varnothing$, the implicit formula $X\Rightarrow Y$

represents the rules that the item set X and Y occur at the same time, and X,Y are the preceding and posterior terms of the potential value association rule of ideological and political education $X \Rightarrow Y$, respectively.

III. B. 2) Principles of Association Rule Mining

Utilizing association rules to mine the implicit association relationships in the text of potential value of Civic Education can facilitate the management of Civic Education. The association rule $X \Rightarrow Y$ represents a kind of association, for a document D_i contains all the feature words in the item set X , then the document D_i also has an association relationship with the feature words in the item set Y , and the attributes of the association rule are usually described by the degree of support, the confidence and the degree of enhancement.

(1) Support

The support of item set X can be expressed by the probability of X appearing in the potential value document set D :

$$support(X) = P(X) = \frac{\left|\{D_i \mid D_i \subseteq D, X \subseteq D\}\right|}{|D|} \tag{2}$$

For the association rule $X \Rightarrow Y$, the probability that the item sets X and Y appear simultaneously in the document set D of the potential value of Civic Education is calculated, i.e., the ratio of the number of documents in the document set D that contain both the risk items X and Y to the total number of documents, reflecting the probability of simultaneous appearance of the potential value of Civic Education.

$$support(X \Rightarrow Y) = P(X \Rightarrow Y) = \frac{P(X \cup Y)}{P(T)} = \frac{\left|\{T \mid X \cup Y \subseteq T, T \subseteq D\}\right|}{|D|} \tag{3}$$

(2) Confidence level

The confidence level of association rule $X \Rightarrow Y$ is the ratio of the number of documents containing both itemsets X and Y to the number of construction risk documents containing itemset X , which is used to measure the credibility of the mined association rule on the potential value of Civic Education.

$$confidence(X \Rightarrow Y) = P(X \mid Y) = \frac{P(X \bigcup Y)}{P(X)} = \frac{support(X \bigcup Y)}{support(X)} \tag{4}$$

In the mining of potential value association rules for Civic Education, the minimum support degree and minimum confidence degree are initially set to determine the strength of potential value association rules for Civic Education.

(3) Enhancement degree

Enhancement degree is a reference index set in association rule mining, which is used to judge the effectiveness and importance of association rules, and can reflect the dependence relationship between the former and latter items of association rules.

$$lift(X \Rightarrow Y) = \frac{support(X \bigcup Y)}{support(X) \cdot support(Y)} \tag{5}$$

If the lift (lift) is greater than 1, it means that $X \Rightarrow Y$ is a valid strong association rule with high analytical value; if the lift (lift) is equal to 1, it means that the two itemsets X and Y are independent of each other; if the lift (lift) is less than 1, it means that $X \Rightarrow Y$ is an invalid strong association rule.

III. B. 3) Apriori algorithm

The Apriori algorithm first identifies all frequent sets that exceed the minimum support and minimum confidence, and generates association rules from the frequent sets that satisfy the conditions. The algorithm uses support to identify all frequent sets of items, then uses confidence to determine the strength of association rules, and combines with boosting to identify effective strong association rules [16]. In this paper, Apriori algorithm is used to mine the risky association rules implied in the text of potential value of Civic Education, and the steps of implementation are as follows:

(1) Construct the set of potential value affairs of Civic Education D , $D = \{D_1, D_2, \cdots, D_n\}$.

(2) Set the minimum support and minimum confidence, up to frequent k itemsets, and make the initial value $k = 1$.

(3) Traverse the text data of potential value of Civic Education to identify the influencing factors and generate the candidate 1-item set.

(4) Calculate the support degree of candidate 1-item set, and generate frequent $k-$ item sets that are greater than the minimum support degree.

(5)Determine whether the frequent $k-$ itemset is the empty set, if not then merge to generate $(k+1)-$ itemset, $k=k+1$; if it is the empty set then calculate the confidence of the frequent $k-$ itemset.

(6) Loop steps (3), (4) and (5) to generate all frequent k itemsets.

(7) Calculate the confidence of all frequent k itemsets, construct association rules, and judge the generation of all construction risk association rules with greater than minimum confidence.

(8) Calculate the enhancement degree of each rule, select the effective strong association rules with enhancement degree greater than 1, and analyze the association rules by combining with the actual meaning of the potential value of Civic and Political Education.

III. B. 4) Visualization of association rules for potential value of Civic Education data

In this paper, the potential value of ideology and politics education text preprocessing data mining, through the sample data repeatedly test debugging, set the minimum degree of support for 0.2 and the minimum degree of confidence for 0.4, so as to obtain more valuable analysis of the rules. By programming the output association rules and assessment data, filtering and sifting out the association rules of the potential value of Civic and Political Education with the support degree less than 0.2, confidence degree less than 0.4, and the enhancement degree less than 1. Because of the huge amount of data, some of the association rules are displayed as shown in Table 1.

Table 1: Research on the potential value association rules of the education

The associated rules are in the preceding paragraph	Correlate	After term	Support %	Confidence %	Degree of ascension
Thinking of politics	⇒	Mental value	35.922	65.372	2.197
Thinking of spirit	⇒	Conscious value	22.845	43.586	1.488
Thinking of education	⇒	Innovative value	28.492	52.839	1.697
Political management	⇒	Three value	41.894	65.932	1.722
Thinking of consciousness	⇒	Educational value	33.147	58.475	2.641
Courses, Studies	⇒	Management	44.683	55.851	2.783
Education, spirit	⇒	Consciousness	48.355	68.225	1.574
Thinking of practice	⇒	Competitiveness	27.753	51.984	1.189
Thinking morality	⇒	Personal cultivation	39.135	70.218	2.458
Thinking training	⇒	Social development	44.481	73.876	3.083

Through the mining of potential value textual data processing of civic education, it can be found that there is a certain correlation between the value factors, including two itemsets of risk association rules, according to the mining results of the above table of the potential value of civic education association rules interpretation and analysis, due to the huge data of the generated association rules, interpretation and analysis of a few examples of the potential value of the association rules are as follows:

(1) Rule - {Civic Learning}⇒ {Spiritual Value}, the rule support is 33.147%, confidence level is 65.372%, and enhancement is 2.197. The rule indicates that when conducting Civic Learning without seriousness, then it may result in the lack of spiritual value, which indicates that Civic Learning is insufficient for the lack of spiritual value to appear, then the spiritual value lack of emergence is 65.372%. For the spiritual value of management guidance, can be targeted to strengthen the management of Civics learning, then can be effective spiritual value of the situation.

(2) Rule - {Civic Awareness}⇒ {Educational Value}, the rule support degree is 31.879%, confidence level is 58.475%, and enhancement degree is 2.641. The rule indicates that if an individual's Civic Awareness is insufficient, then it may be accompanied by the absence of educational mechanism, and the possibility of the rule's latter item Civic Awareness appearing at the same time is 58.475%. Is 58.475%. For the present Civic and Political Awareness Management Guidance, it can be targeted to enhance the management training of educators and learners of Civic and Political courses to enhance the probability of Civic and Political Awareness.

Due to the huge potential value text data of civic education, there are associations between many value factors, and it is impossible to obtain the potential value association relationship accurately and quickly, so this paper adopts IBM SPSS Modeler 18.0 software to construct the potential value association matrix and generate the potential value association visualization network diagram, which can efficiently and intuitively identify and obtain the potential value association relationship, and thus make judgment and processing, as shown in Figure 2.

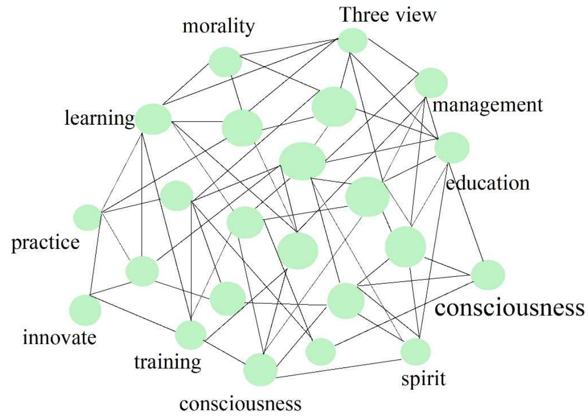

Figure 2: Visual network diagram of potential value association rules

The knowledge visualization method is used to draw a visual network diagram of the potential value association of ideological and political education, and the characteristic words of ideological and political education are used as nodes to represent the correlation between potential values by connecting lines, so as to improve the efficiency of identifying and reading the potential value association information. From the above figure, it can also be found that there is a correlation between many potential values, such as "practice" and "spirit", "innovation", "consciousness", "management", "learning", "morality", "three views" are related, and so on, the visual network diagram can be used to efficiently and accurately obtain the association, and take targeted measures to improve ideological and political education.

IV. Results of mining the potential value of Civic Education data

IV. A. Theme Discovery of Potential Value of Civic Education Data

IV. A. 1) Analysis of high-frequency words

Table 2: The statistical results of the high frequency word of the education potential value

Word	Frequency	Word	Frequency
Education	1876	Construction	722
Learning	1643	Three view	674
Practice	1487	Information	641
Dpirit	1465	Politics	626
Innovate	1137	Activity	587
Consciousness	1014	Patriotism	515
Develop	958	Courses	501
Party history	922	Society	497
Moral literacy	884	Liability	486
Training	836	Ideal	479
Management	784	Culture	446
Age	739	Disseminate	421

Count high-frequency words for the tweet title after jieba word segmentation, run python to count the title Chinese high-frequency words, because there are synonyms or similar words in the word segmentation, therefore, construct a synonym list for synonym merging, and in the process of word frequency statistics, the synonym word frequency in the document is merged and processed, and the sum of word frequencies is counted. After synonym merging, the top 24 words were obtained, as shown in Table 2. The table shows the top 24 words in the title. Among them, it can be seen that the top five words in frequency are "education", "learning", "practice", "spirit" and "innovation", among which, "education" appears the most frequently, 1876 times; The word "learn" came second with 1,643 times; The word "practice" ranked third, with 1,487 occurrences; The word "spiritual" ranked fourth most frequently, with 1,465 occurrences; The word "innovation" ranked fifth with 1,137 times.

The top 23 words with word frequency can be roughly divided into four levels according to the word frequency, which are the first level of the word frequency of more than 1000, the second level of the word frequency of 700-1000, and the third level of word frequency 400-700. According to this kind of division standard, there are 11 words at the first level, namely "education", "learning", "practice", "spirit", "innovation" and "consciousness", and these 6 words have a frequency of more than 1,000 words, which is the strongest high-frequency word team of the potential value of ideological and political education. There are seven words in the second level, which are twice as many as the number of words in the first level, namely "development", "party history", "moral literacy", "training", "management", "era" and "construction". There are 11 words in the third level, including "three views", "information", "politics", etc. Figure 3 shows the high-frequency word frequency chart of the potential value of ideological and political education, and the word frequency chart more intuitively shows the comparison of the word frequency of the top 23 words.

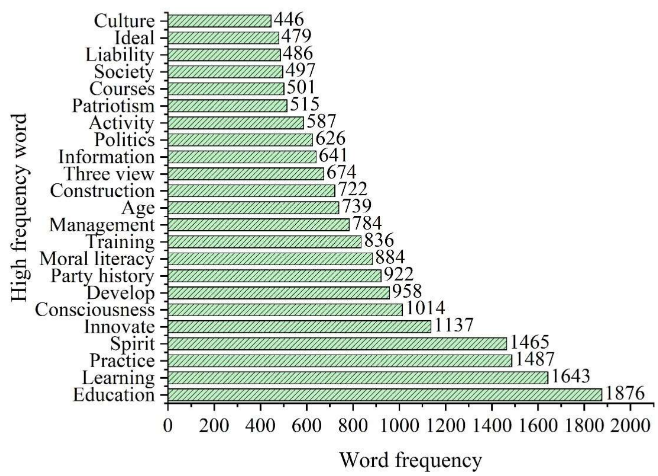

Figure 3: The frequency diagram of the high frequency word of the education

IV. A. 2) LDA topic modeling

Text-Rank model is mainly based on word-word co-occurrence for keyword extraction, the advantage of this model is that it can take into account the syntactic and semantic relationships between words, thus ensuring that the extracted feature words are more accurate; its disadvantage is that it needs to construct a text-unidirected graph, the computational complexity is high. In this paper, the TF-IDF model is invoked through python to extract document feature words from documents. Secondly, the number of topics is determined, and the methods to determine the number of document topics are confusion, consistency, and visualization methods. Figure 4 shows the theme consistency of Civic Education from 2020 to 2024. From the figure, it can be seen that the theme consistency curve has the highest point when the number of themes is 7, which indicates that it is most appropriate to extract the theme of the document and set the number of themes at 7 as inferred from the theme consistency data.

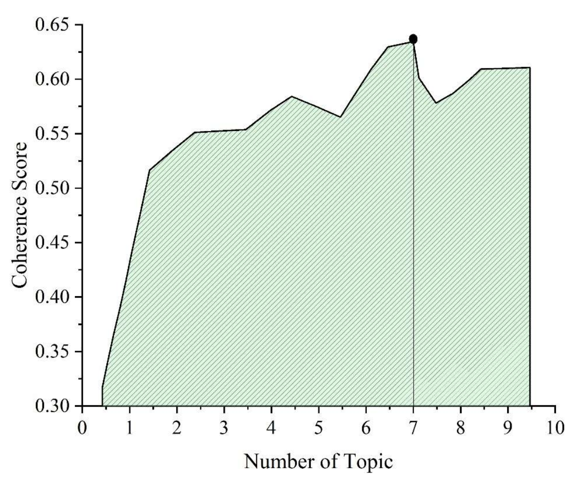

Figure 4: 2020-2024 theme consistency of education

The visualization method can more intuitively show the relationship between the themes and the degree of crossover, the less the themes and the themes cross each other indicates that the theme differentiation is higher, then the corresponding number of themes is more optimal. The visualization of the results of the themes of potential value of ideological education is shown in Figure 5, from which it can be seen that when the number of themes is 7, the independence between the themes is better and the crossover is small, then the number of themes is finally set at 7.

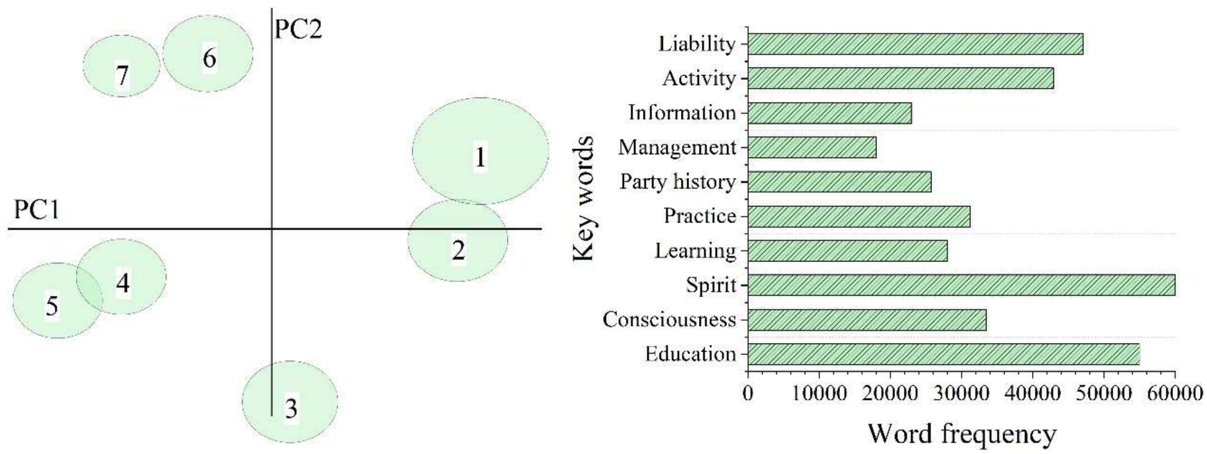

Figure 5: Visualization of the potential value of the education

IV. A. 3) Thematic category analysis

According to the LDA theme modeling to get the seven themes of potential value of Civic and Political Education, as shown in Table 3, the seven themes are obtained by summarizing the internal words of the themes. The seven themes of Civic and Political Education are obtained as Theme 1 Red Culture, Theme 2 Online Service, Theme 3 Competition, Theme 4 Real-time Exercise, Theme 5 Cultural Promotion, Theme 6 Moral Literacy, and Theme 7 Knowledge Learning.

Table 3: 2020-2024-word distribution of the education theme

Numbering	Theme	Topic key
1	Red culture	history, culture, spirit, time, thought, construction, party history and development
2	Online service class	books, lending, service, time, study, reservation, teacher, training
3	Race class	activities, participation, registration, theme, dissemination, course, education
4	Real-time practice class	ideal, research, culture, practice, construction, consciousness, innovation
5	Cultural promotion	reading, history, life, recommendation, time, books, communication, theme
6	Moral literacy	civilization, three views, cultivation, quality, morality, management, responsibility and patriotism
7	Knowledge learning class	courses, learning, video, data, information, time, ideal

IV. B. Correlation Analysis of the Potential Value of Civic Education Data

IV. B. 1) Association Rule Mining Based on Apriori Algorithm

By analyzing the factors of potential value of ideological and political education by association rules, the association law between these factors can be discovered in time. In the previous paper, the potential value of ideological and political education was mined and analyzed using the LDA theme model, which resulted in 7 major themes and 24 theme keywords. Using the mining results in the previous paper, the textual information in the potential value of ideological and political education is converted into structured data. In this paper, the potential value in each topic of ideology and political education is labeled, and the factors that appear are labeled as 1 and those that do not appear are labeled as 0, thus creating a structured Boolean dataset composed of “0-1”. This dataset provides data support for mining the potential associations between the latent values of Civic Education using association rule technology. Some of the basic data specifically used for mining are shown in Table 4.

Table 4: Association Rule Mining Base Dataset

Theme	Party history	Learning	Activity	Practice	Innovate	Disseminate	Patriotism	Time
1	1	0	1	0	0	1	0	1
2	0	0	0	1	0	0	0	1
3	0	1	0	0	0	0	0	0
4	0	0	0	0	1	0	1	1
5	1	0	1	1	0	0	0	0
6	0	1	0	0	1	0	1	0
7	1	0	0	1	1	0	1	0

In this section, based on Apriori algorithm, association rule mining is carried out on the potential value factors of Civic Education mined in the previous section.In Apriori algorithm, support and confidence are two important thresholds affecting the results of association rule mining, and the combination of thresholds suitable for this study is determined by using trial-and-error method and comparative analysis. On the basis of several trials, the minimum support degree was finally selected as 0.05 and the minimum confidence degree as 0.1, and a total of 98 effective association rules were obtained, and the distribution characteristics of the association rules in terms of the support degree and confidence index are shown in Figure 6.

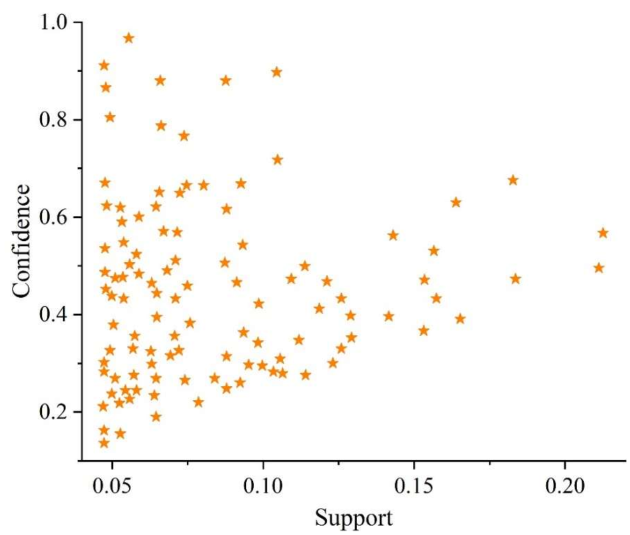

Figure 6: Association Rule Scatter plot

In this section, three key metrics of association rules are shown through scatter plots: support and confidence. In the scatter plot, the horizontal axis indicates the support degree and the vertical axis indicates the confidence degree. It is obvious from the figure that the support degree of the data in this paper is mainly concentrated between 0.05-0.1, accounting for 80.37% of all the rules; the confidence degree is more scattered, but mainly between 0.2-0.6, accounting for 91.42% of all the rules, and with the increase of the confidence degree, the number of rules shows an obvious downward trend; the enhancement degree is mostly between 1-2.6, accounting for all the rules 97.11% of all rules, and the rules with higher lift are mostly distributed in the region with higher confidence level. These results reveal the distribution characteristics and internal rules of the association rules in the data of this paper.

IV. B. 2) Analysis of the importance of potential value

This subsection takes the 98 strong association rules obtained from the final screening as the object of study, and conducts a preliminary statistical analysis of the causal factors involved in them. By counting the frequency of each factor's occurrence in the preceding and following items, in order to understand which are the key themes and key factors affecting Civic Education. The specific statistical results are shown in Table 5. From the table, it can be seen that spirit is the most important causal theme, accounting for 50.5% of the total frequency, followed by education, accounting for 19.3% of the total frequency, followed by practice, accounting for 18.4% of the total frequency, and finally consciousness, accounting for 11.8% of the total frequency. This shows that spirit is the most important cause of Civic Education, and the level and effect of education and practice are also important aspects that affect Civic Education, while consciousness has the least effect on Civic Education.

Table 5: The statistics of the education category and factors of the education

Categories	Numbering	Factor	Frequency	Tot	Proportion
Consciousness	A1	Liability	5	25	11.8%
	A2	Patriotism	4		
	A3	Culture	8		
	A4	Disseminate	8		
Education	P1	Courses	14	41	19.3%
	P2	Teacher	13		
	P3	Books	14		
Practice	M2	Training	25	39	18.4%
	M3	Competition	14		
Spirit	S1	Consciousness	10	107	50.5%
	S2	Develop	35		
	S3	Innovate	41		
	S4	Patriotism	21		

IV. B. 3) **High Support Association Rule Analysis**

In order to reveal the frequent associations between potential factors affecting coal mine fatalities and injuries, the 10 rules with the highest frequency of occurrence were selected from all the relevant rules, which were called high support association rules. The degree of support is the probability of these factors occurring at the same time. Table 6 demonstrates the antecedent, consequent, support, confidence, and lift of these rules. The support of these rules varies between 0.115 and 0.149, the confidence varies between 0.289 and 0.549, and the lift varies between 1.022 and 1.468. Figure 7 visualizes the support of these rules.

As can be seen from the figure, some of the important influences include responsibility (A1), patriotism (A2), curriculum (P1), books (P3), and competition (M3). There are various correlations between these factors, with the highest support being responsibility (A1) with a support of 0.149, implying a 14.9% probability of this factor occurring. Some other rules, culture (A3), teacher (P2), communication (A4), training (M2), awareness (S1), development (S2), innovation (S4), and patriotism (S5), also have high support, indicating a strong association between these factors.

Table 6: High support association rules

Serial number	Fore term	Back direction	Support	Confidence	Degree of ascension
1	A1	A2	0.149	0.549	1.422
2	A3	A4	0.149	0.389	1.422
3	A2	A1	0.137	0.394	1.022
4	A4	A3	0.137	0.351	1.022
5	P1	P2	0.125	0.522	1.162
6	P2	P1	0.125	0.289	1.162
7	P3	P1	0.125	0.417	1.025
8	P3	P2	0.125	0.319	1.025
9	M2	M3	0.121	0.451	1.468
10	M3	M2	0.121	0.393	1.468
11	S1	S2	0.121	0.488	1.213
12	S3	S4	0.115	0.661	1.492
13	A1	P2	0.115	0.268	1.492
14	M3	S1	0.115	0.477	1.239
15	S4	P3	0.115	0.311	1.239

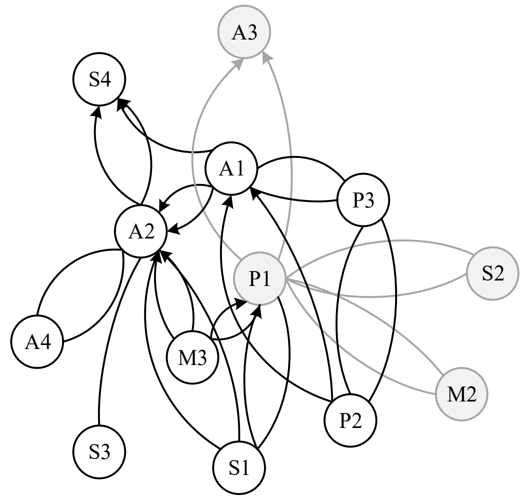

Figure 7: High support association rule diagram

V. Conclusion

The study extracts potential value features through TF-IDF algorithm, uses LDA model for text modeling, combines confusion degree and LDA technology to determine the number of themes, and summarizes theme identification and causal factors. Finally, Apriori algorithm is used to analyze the association between themes and factors, and targeted control strategies are proposed based on the analysis results. The research results, as follows:

(1) According to the theme consistency synthesis, the number of themes is determined as 7, and the LDA model is used to analyze the themes of the potential value of ideological and political education, which are red culture, online service, competition, real-time practice, cultural promotion, moral literacy, and knowledge learning.

(2) The correlation analysis by Apriori algorithm concluded that spirit is the most important causal theme among the themes affecting the potential value of Civic and Political Education, accounting for 51% of the total frequency, followed by education, accounting for 19% of the total frequency, practice, accounting for 18.4% of the total frequency, and finally consciousness, accounting for 11.8% of the total frequency.

Funding

Jiangsu Province Social Science Application Research Fine Project Special Topic of Ideological and Political Education in Colleges and Universities "Research on the Development of Digital Empment of Network Ideological and Political Education in Colleges and Universities" (Project Number: 24SZC-136).

References

[1] Wang, X. (2019, July). Characteristics and Application Ways of Online Education of Ideological and Political Theory Courses in Universities from the Perspective of Big Data. In International Conference on Frontier Computing (pp. 1626-1632). Singapore: Springer Singapore.
[2] Liu, Y. (2018). The Modern Ideological and Political Education from the Perspective of Big Data. Advances in Education, 8(3), 167-171.
[3] Li, H. (2025). Multicultural data assistance mining analysis for ideological and political education in smart education platforms using artificial intelligence. Wireless Networks, 31(1), 567-581.
[4] Fan, Y. (2025). Research on the Innovation Path of Network Ideological and Political Education in Colleges and Universities. Curriculum Learning and Exploration, 2(4).
[5] Zhou, W., & Yang, T. (2021, May). Application analysis of data mining technology in ideological and political education management. In Journal of Physics: Conference Series (Vol. 1915, No. 4, p. 042040). IOP Publishing.
[6] Xiaoyang, H., Junzhi, Z., Jingyuan, F., & Xiuxia, Z. (2021). Effectiveness of ideological and political education reform in universities based on data mining artificial intelligence technology. Journal of Intelligent & Fuzzy Systems, 40(2), 3743-3754.
[7] Yang, R. (2021, September). Construction and analysis of ideological and political education reform path based on data mining. In 2021 4th International Conference on Information Systems and Computer Aided Education (pp. 214-218).
[8] Shabani, N., Beheshti, A., Farhood, H., Bower, M., Garrett, M., & Alinejad-Rokny, H. (2023). A rule-based approach for mining creative thinking patterns from big educational data. AppliedMath, 3(1), 243-267.
[9] Shao, Z., Sun, H., Wang, X., & Sun, Z. (2020). An optimized mining algorithm for analyzing students' learning degree based on dynamic data. IEEE Access, 8, 113543-113556.
[10] Angeli, C., Howard, S. K., Ma, J., Yang, J., & Kirschner, P. A. (2017). Data mining in educational technology classroom research: Can it make a contribution?. Computers & Education, 113, 226-242.
[11] Yang, Y., Zhang, B., & Dong, Z. (2022, December). Based on Student Learning Data Management Platform a Framework and Strategies of Ideological and Political Elements Discovering in Computer Algorithm Course. In 2022 3rd International Conference on Modern Education and Information Management (ICMEIM 2022) (pp. 272-280). Atlantis Press.

[12] Zhu, Y., & Liu, M. (2022, October). Exploration on Blended Teaching of" Data Mining" from Perspective of Ideological and Political Education. In Proceedings of the 14th International Conference on Education Technology and Computers (pp. 58-63).
[13] Li, B. (2024). Evaluation System of College Ideological and Political Education Index Based on Data Mining Algorithm. Advances in Computer, Signals and Systems, 8(2), 12-21.
[14] Ling, C., Xiaowei, H., & Wenjuan, L. (2025). Evaluation Mechanism of Patriotism Education for College Students based on Big Data Analysis in the Network Era. International Journal of High Speed Electronics and Systems, 2540218.
[15] Sanghun Sul & Seung Beom Cho. (2024). Understanding people’s attitudes in IoT systems using wellness probes and TF-IDF data analysis. Multimedia Tools and Applications,83(35),82495-82514.
[16] Mahsa Hajihosseinlou,Abbas Maghsoudi & Reza Ghezelbash. (2024). Geochemical Anomaly Detection and Pattern Recognition: A Combined Study of the Apriori Algorithm, Principal Component Analysis, and Spectral Clustering. Minerals,14(12),1202-1202.