International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 4312-4327

https://doi.org/10.70517/ijhsa464279

# Research on optimization of ideological and political education content dissemination path based on association rule mining algorithm in big data era

**Juan Zheng**[1,*]
[1] School of Marxism, Henan Open University, Zhengzhou, Henan, 450046, China
Corresponding authors: (e-mail: zjalice625@126.com).

**Abstract** Ideological and political education dissemination is crucial to the ideological and political education of college students, and its effect directly affects the effectiveness of college students' ideological and political education. Firstly, the traditional association rule mining algorithm, in the process of ideological and political education content dissemination path optimization existing problems are described. In order to solve this kind of problem, on the basis of traditional association rule mining algorithm, transaction matrix and user interest degree are introduced, and finally the design of association rule mining algorithm based on transaction matrix and user interest degree is completed, and the algorithm is used to explore the current situation of ideological and political education content dissemination. When the number of ideological and political education resources is 50, 100 and 200 respectively, this paper's algorithm has high-quality user satisfaction, which verifies that the introduction of transaction matrix and user interest degree on the traditional association rule mining algorithm is more conducive to the development and construction of ideological and political education content dissemination work.

**Index Terms** association rules, ideological and political education, user interest degree, transaction matrix

## I. Introduction

Human society is in the process of transforming from industrialization to digitalization and intelligence, and this transformation is driving fundamental changes in the entire socio-economic structure and technological systematization, with new modes of production, lifestyle, organization, and education constantly emerging, and human history is undergoing a profound structural change [1], [2]. All walks of life are constantly utilizing big data, all of which bring opportunities to ideological and political education and also produce major challenges [3]. Ideological and political education is a unique way of working to realize the governance of the country, and ideological and political education in colleges and universities, as an important branch of ideological and political education, is an important means to cultivate college students' outlook on life, worldview and values [4], [5]. In the face of new situations and new problems, ideological and political education must comply with the requirements of the big data era, carefully study the internal mechanism of the role of big data, make full use of the favorable conditions brought by the big data era for educational and practical activities, and promote the innovation of ideological and political education [6], [7].

Big data has a very broad application prospect in the field of ideological and political education. On the one hand, big data makes the content of ideological and political education more accurate, through the mining and analysis of big data on the thought and behavior information of the education object, it can make ideological and political education pay more attention to the group specificity, and improve the educational pertinence and accuracy [8]-[10]. On the other hand, big data provides support for the construction of ideological and political education content, and with the help of big data analysis tools, it can realize the coordination of resources of the educational subject, the interactive use of innovative types, etc., which provides the basis for the innovation of ideological and political education content [11], [12]. Educators should learn to use big data well, constantly in the process of utilizing big data technology, improve the educators' ability and literacy to mine and analyze big data, and then gradually enhance the effectiveness of education, more deeply grasp the application and development of the law of ideological and political education big data, so that the big data can play a greater role in promoting the dissemination of the content of ideological and political education.

This paper first describes the data mining technology definition and method pairs, learns about several commonly used mining algorithms, and in view of the limitations of the research content, develops a research program for ideological and political education content dissemination based on association rule mining algorithms. It is found that the traditional association rule mining algorithm has a series of problems in the ideological and political education content dissemination path optimization project, such as low efficiency of rule extraction and multiple cyclic access to transactions. Under the premise of changing without changing the research subject, the content dissemination of ideological and political education based on association rule mining algorithm is optimized by introducing transaction matrix and user interest degree. Finally, relevant research data are synthesized to verify the application analysis of the research scheme of this paper.

# II. Optimization of ideological and political education path

## *II. A. Data mining techniques*

### II. A. 1) Overview of data mining

Data mining is an information processing technology that analyzes and extracts from a large amount of data the hidden and unknown information that is valuable for decision-making [13]. Data mining is also the process of using algorithms to discover potentially valuable, regular relationships, and instructive information in data from large amounts of fuzzy, noisy, and random data. Data mining utilizes cross-cutting techniques from several fields, on the one hand, it mainly comes from sampling and hypothesis testing in statistics, on the other hand, it mainly involves machine learning, artificial intelligence, pattern recognition and other knowledge technologies [14]. There are some others including signal processing technology, visualization technology, database technology and other knowledge fusion of various fields. The value of the data is explored by analyzing the prediction of the trend of the future development of things and events. The development of data mining technology in recent years has promoted changes in the education industry, gradually changing from the previous traditional teaching management model to personalized education and new teaching evaluation system. How to utilize big data to address the valuable information that exists in the management of education and teaching, so a new field of educational data mining (EDM) has been formed. The object of educational data mining is students and teachers, and educational data mining has certain uniqueness in data sources, data characteristics, and application of results, etc. EDM utilizes the application data of educational informatization to help schools solve the problems in the teaching of Civics, such as assisting school administrators to make decisions, helping teachers to improve the teaching mode of Civics, and helping students to improve the efficiency of Civics learning and so on.

### II. A. 2) Data mining process

Data mining technology has developed a more mature process with continuous development, and the data mining process is shown in Figure 1. Data mining is knowledge discovery in data (KDD), there are also data mining as a basic step in the knowledge discovery process. This paper agrees with the latter point of view, and defines data mining as a step of knowledge discovery, then the process of knowledge discovery mainly has the following steps: data cleaning, data integration, data selection, data change, data mining, pattern evaluation, and knowledge representation.

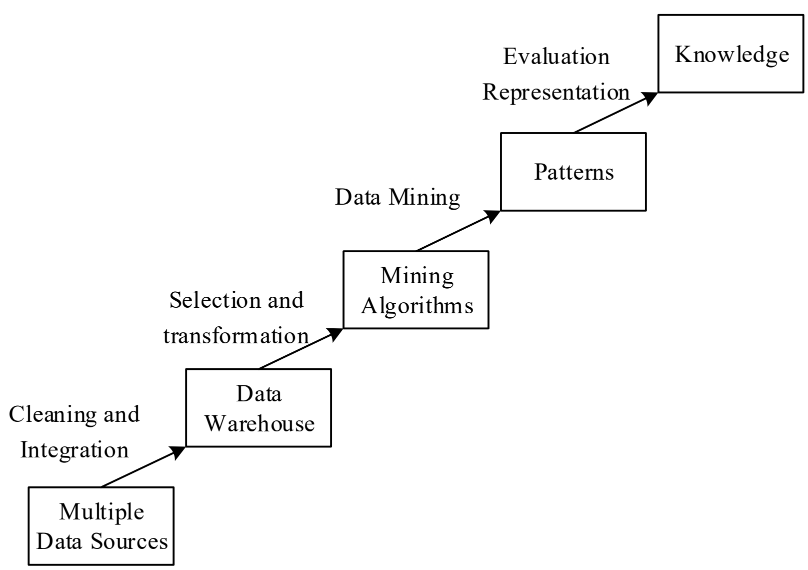



Figure 1: Data mining process

Data cleansing mainly solves the problem of data inconsistency by deleting redundant and noisy data, filling in missing data and modifying erroneous data. Since the collected raw data may have incorrect data, missing fields, null values and so on due to the application system, manual operation, etc., the results obtained from processing and analyzing such raw data are bound to be unreliable. Because in order to improve the quality of data mining results, to find the real value of the information must be cleaned up before data mining, to ensure data integrity, correctness, uniqueness and consistency.

Data integration is the combination of data from multiple data sources stored together in a unified way. Since the data comes from various application systems, there are great differences in the format and type of data, and data integration is the combination of these data such as database data, web page data, log data, etc. into a common database. In data integration there are inconsistencies, redundancy and other problems, such as attributes representing a unified concept may have different names in different databases, solving these problems will help to improve the speed and accuracy of mining in the subsequent mining.

The order of the two steps, data selection and data transformation, can sometimes be switched. Among them, data selection is to select the data related to the analyzing task from the database, selecting the right size of data can reduce the time of data mining, while selecting the accurate data related to the purpose of the research can get the mining results that match the research object. Data transformation is to transform and unify the data that has been selected and processed into a form suitable for data mining. Commonly used transformation methods are aggregation, normalization, discretization and so on. Aggregation is to summarize the data, normalization can scale the data up and down in equal proportions, such as putting the data between 0-1. Discretization can categorize the data. Transformation of data can be the exposure of certain characteristics of the data in order to better go for data mining.

The four steps of data cleaning, data integration, data selection and data transformation are generally referred to as the data preprocessing phase, which is an important step before data mining. The purpose of data preprocessing is to improve the quality of the mined data and to provide accurate, complete and standardized data for subsequent mining in order to improve the efficiency of the data mining algorithms and the accuracy of the results.

#### II. A. 3) Data mining methods

Data mining tasks can usually be categorized as predictive and descriptive. Predictive tasks refer to predicting unknown data type information based on known information, such as classification analysis, regression analysis, outlier analysis, etc. Descriptive tasks refer to finding hidden patterns in summarized data, such as clustering analysis, correlation analysis, sequential pattern mining, and visualization mining. According to the goal of data mining in a knowledge discovery process, multiple mining methods are often used to get the desired knowledge.

#### II. A. 4) Common tools for data mining

With the rapid development of data mining technology, resulting in the emergence of many classic data mining tools, many famous database companies have developed data mining tools, the main common types of tools are as follows:

(1) SPSS Modeler

SPSS Modeler was previously developed by the famous company SPSS, which was acquired by IBM in 2009.SPSS Modeler is an open data mining tool, which has won two British SMART innovation awards. It can not only carry out the entire data mining process, including data acquisition, conversion, modeling and knowledge assessment, etc., but also supports the data mining industry standard CRISP-DM. SPSS Modeler has a lot of object-oriented extension module interface, so that users can develop their own algorithms and run.

(2) Weka

Weka is an open source software developed by Waikato University, its full name is Waikato Intelligent Analytics Environment (Weka), is based on the Java environment, a collection of various machine learning algorithms and data preprocessing tools. It is able to perform data preprocessing, classification, clustering, regression, correlation analysis, etc. in a visual interactive interface, which is a widely used open source data mining tool.

(3) DBMiner

DBMiner is a data mining tool developed by Simon Fraser University in Canada, which integrates data mining and relational databases together, and is able to complete the analysis of association rules, classification rules, evolutionary knowledge analysis and other functions, and proposes an interactive data mining query language DMQL.

(4) MineSet

MineSet is a multi-task data mining tool jointly developed by Standford University and SGI, with a variety of data mining algorithms and visualization functions. It allows users to directly see the valuable information of the mined data and can directly read the data directly from a variety of databases.

In addition to this, there are a number of data mining tools, such as SAS Enterprise Miner developed by SAS, Microsoft's SQL Server Analysis Services and so on have a wide range of applications.

***II. B.Civic content dissemination based on association algorithms***

The association rule mining technique mainly analyzes the user's online transaction database and mines the association rules latent in the transaction data set by means of data mining. Because most of the users' behaviors have a certain purpose, they contain many association patterns. The process of dissemination of Civics content based on association algorithm has four steps: construction of transaction database, association rule mining, dissemination candidate set, and dissemination of associated items. The advantage of the algorithm of association rules is that its algorithm has mature theoretical support, and the rule mining process only needs to analyze the combination of Civic content in the transaction, and does not need to deal with user rating data information. However, it also has some disadvantages: the scale of data in the dissemination of Civics content is usually very large, and it is necessary to carry out complex conversion quantization and other operations on large-scale data, and secondly, the transaction in the association rule mining process is accessed cyclically many times, and it is necessary to traverse the database of the transaction many times, and the efficiency of rule extraction is low. To address the above described drawbacks, an association rule mining algorithm based on transaction matrix and user interest degree is proposed to improve the existing association rule mining algorithms from two phases of frequent item mining and association rule mining without reducing the number of interesting rules, with the objectives of eliminating the redundant set of candidate items, improving the efficiency of searching for frequent items, and reducing the number of redundant rules, and optimizing the path of dissemination of Civic and political education content.

**II. B. 1) Description of the problem**

The Apriori algorithm converts the problem of mining frequent item sets into an iterative discovery process. Let there be a transaction database $D=\{T_i|i=1,2,\cdots,m\}$ that records $m$ transactions and $n$ item items, and a set of item sets $I=\{I_j j=1,2,\cdots,n\}$, and any transaction $T_i$ in $D$ corresponds to a subset $I_r$ of $I$. The concepts related to the algorithm are as follows:

Support (Sup): let the set of item items composing the rule $r$ be $I_r \subseteq I$, and the support of $I_r$ on $D$ is equal to the percentage of the transactions containing $I_r$ to all the transactions in $D$, i.e., the probability of the occurrence of $I_r$ in all the transactions, as shown in Equation (1):

$$Sup(I_r)=\frac{\left\|\{T_i|I_r \subseteq T_i, i=1,2,\cdots,m\}\right\|}{\|D\|} \tag{1}$$

Confidence (Conf): for rules $r$, $r$ defined as in $D$ and $I$ as in $I_i \to I_j$, the confidence on database $D$ is the ratio of the number of transactions containing both $I_i$ and $I_j$ to the number of transactions containing $I_i$, as shown in Equation (2):

$$Conf\left(r: I_i \to I_j\right)=\frac{Sup\left(I_i \cup I_j\right)}{Sup(I_i)} \tag{2}$$

Frequent itemset: $D$ The set of all $I$ non-empty subsets of $D$ that satisfy the specified minimum support ($Sup_{\min}$) is the frequent itemset.

**II. B. 2) Establishment of a transaction matrix**

Transaction matrix: let there be a database of recorded transactions $D$ and a set of item items $I$. A two-dimensional matrix $T$ of $m \times n$ is built from $D$ as shown in equation (3):

$$T=\left(T_1,T_2,\ldots,T_m\right)'=\begin{pmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{m1} & \cdots & t_{mn} \end{pmatrix}, t_{ij}=\begin{cases} 1 & I_j \in T_i \\ 0 & I_j \notin T_i \end{cases} \tag{3}$$

where the $i$ th row of $T$, $T_i$, represents the $i$ th transaction, and if the $i$ th transaction, $T_i$, contains the item item $I_j$, then $t_{ij}=1$. Otherwise, $t_{ij}=0$. Typically, $T$ is a sparse matrix.

For example, the process of transforming the corresponding transaction matrix for a transaction database is shown in Figure 2. A transaction database contains 7 transactions as well as 7 different item entries. The transactions are represented by the set $T=\left\{T_i \middle| i=1,2,\ldots,7\right\}$, and each transaction $T$ has a unique identifier denoted as TID.The set of item items is $I=\{A,B,C,D,E,F,G\}$. The transaction database can be represented as $D=\left\{\{A,B,C,D,G\},\{B,E\},\{A,C,E,G\},\{A,B,E,F,G\},\{A,C,D\},\{A,F\},\{A,B,C,D\}\right\}$. Then, the matrix $M$ is the obtained transaction matrix.

Transcation Database D

| TID | Items |
|---|---|
| 1 | *A,B,C,D,G* |
| 2 | *B,E* |
| 3 | *A,C,E,G* |
| 4 | *A,B,E,F,G* |
| 5 | *A,C,D* |
| 6 | *A,F* |
| 7 | *A,B,C,D* |

→

Transcation Matrix M

| TID | Item *A* | Item *B* | Item *C* | Item *D* | Item *E* | Item *F* | Item *G* |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 2 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 3 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 4 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| 5 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| 6 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 7 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |

Figure 2: Example of transforming a database to Transaction Matrix

Absolute support: for the transaction matrix $T$, the absolute support of the item term $I$ is the sum of the $j$th column vectors of $T$, as shown in equation (4):

$$SupA\left(I_j\right)=\sum_{i=1}^{m} t_{ij}, i=1,2,\ldots,m \tag{4}$$

Support: for the transaction matrix $T$, the support of the item term $I_j$ is equal to the ratio of the absolute support of $I_j$ to the number of transactions $m$, as shown in equation (5):

$$Sup\left(I_j\right)=SupA\left(I_j\right)/m=\sum_{i=1}^{m} t_{ij}/m, i=1,2,\ldots,m \tag{5}$$

### II. B. 3) Fast Frequent Item Mining

(1) Frequent 1-item mining

Let the minimum support be $Sup_{\min}$, and the candidate 1-item set $C_1$ is the set of item items $I$. First, the absolute support $\operatorname{Sup}A\left(I_j\right)$ of the item item $I_j$ is calculated, and then the support $\operatorname{Sup}\left(I_j\right)$ of each item of each candidate 1-item set is calculated according to Eqn. and compare it with the minimum support $Sup_{\min}$, i.e., the frequent 1-term set $L_1$ is obtained, as shown in Equation (6):

$$L_1=\left\{SupA\left(I_j\right)\geq Sup_{\min}, j=1,2,\ldots,n, P_1\leq n\right\}, L_1\subseteq I \tag{6}$$

Second, remove the columns in $T$ where the vector sum $\sum_{i=1}^{m} t_{ij}<m\cdot Sup_{\min}$, i.e., remove the columns in which the item items with less than minimum support $I_j'$ are located, and $I_j'$ is defined as shown in Equation (7):

$$
\begin{aligned}
& I'_j \in \overline{L}_1 = \left\{ I'_j \middle| Sup\left(I'_j\right) < Sup_{\min}, j = 1,2,\ldots,n \right\} \\
& L_1 \cup \overline{L}_1 = I \ And \ L_1 \cap \overline{L}_1 = \varnothing
\end{aligned} \tag{7}
$$

Finally, remove the rows where the sum of vectors $\sum_{j=1}^{n} t_{ij} < 2$ in the matrix $T$ as shown in Eq. (8):

$$
\overline{T} = \left\{ T_i \middle| \sum_{j=1}^{n} t_{ij} < 2, i = 1,2,\ldots,m \right\} \tag{8}
$$

Clearly, there is $\overline{T} \subseteq T$. Let the number of rows, columns of $T$ be $m_1\left(m_1 \le m\right)$ and $n_1\left(n_1 \le n\right)$, respectively, and the number of frequent 1-items be denoted as $P_1, T$, and $T$ as $T_1$.

Setting the minimum support $Sup_{\min}$ equal to 0.4, we can find that the support of the item item $F$ is $2/7$ (less than $Sup_{\min}$), so we delete the column where the item item $F$ is located, i.e., the sixth column of data. After that, we find that the sum of elements in the sixth row is less than 2, so we continue to remove the sixth row of data. Finally, the frequent 1-item set is obtained, denoted as $L_1 = \{A,B,C,D,E,G\}, T$, the number of rows and columns of $T$ is $m_1 = n_1$, the number of frequent 1-items $P_1 = 6$, and the transaction matrix $T$ is transformed into $T_1$.

(2) Frequent 2-terms mining

First, an $m_2$-order Hermite matrix $H$ is constructed based on $T_1$, as shown in Eq. (9):

$$
\begin{aligned}
H = T' \cdot T &= \begin{pmatrix}
\sum_{i=1}^{m_2} t'_{1i} t_{1i} & \sum_{i=1}^{m_2} t'_{1i} t_{2i} & \cdots & \sum_{i=1}^{m_2} t'_{1i} t_{m_2} \\
\sum_{i=1}^{m_2} t'_{2i} t_{1i} & \sum_{i=1}^{m_2} t'_{2i} t_{2i} & \cdots & \vdots \\
\vdots & \cdots & \ddots & \vdots \\
\sum_{i=1}^{m_2} t'_{m_2} t_{1i} & \cdots & \cdots & \sum_{i=1}^{m_2} t'_{m_2} t_{m_2}
\end{pmatrix} \\
&= \begin{pmatrix}
\sum_{i=1}^{m_2} t_{i1} t_{i1} & \sum_{i=1}^{m_2} t_{i1} t_{i2} & \cdots & \sum_{i=1}^{m_2} t_{i1} t_{m_2} \\
\sum_{i=1}^{m_2} t_{i2} t_{i1} & \sum_{i=1}^{m_2} t_{i2} t_{i2} & \cdots & \vdots \\
\vdots & \cdots & \ddots & \vdots \\
\sum_{i=1}^{m_2} t_{m_2} t_{i1} & \cdots & \cdots & \sum_{i=1}^{m_2} t_{m_2} t_{m_2}
\end{pmatrix}
\end{aligned} \tag{9}
$$

$H$ has the following characteristics:

Then, the matrix $H$ is decomposed into three parts: a diagonal matrix $Y$, an upper triangular matrix $X = \left(x_{ij}\right)$ and a lower triangular matrix $X'$. Where $X$ and $X'$ are symmetric matrices to each other as shown in Equation (10). Then, a matrix $G = \left(g_{ij}\right)$ is constructed from the matrix $X$, as shown in Equation (11):

$$H=\begin{pmatrix} \sum_{i=1}^{m_2} t_{i1}t_{i1} & \sum_{i=1}^{m_2} t_{i1}t_{i2} & \cdots & \sum_{i=1}^{m_2} t_{i1}t_{im_2} \\ \sum_{i=1}^{m_2} t_{i2}t_{i1} & \sum_{i=1}^{m_2} t_{i2}t_{i2} & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ \sum_{i=1}^{m_2} t_{im_2}t_{i1} & \cdots & \cdots & \sum_{i=1}^{m_2} t_{im_2}t_{im_2} \end{pmatrix}=\begin{pmatrix} & & X \\ & Y & \\ X' & & \end{pmatrix} \tag{10}$$

$$G=g_{ij}=g_{ji}=\begin{cases}1 & g_{ij}\geq m\cdot Sup_{\min} \quad i\neq j \\ 0 & g_{ij}< m\cdot Sup_{\min} \quad i\neq j \\ 0 & i=j\end{cases} \tag{11}$$

It is easy to see that the frequent 2-term set $L_2=\{C_2 \mid g_{ij}=1, i, j=1,2,\cdots,P, i\neq j\}$ can be obtained from the matrix $G$. Finally, the rows of the matrix $T_1$ with element sum less than 3 are removed, denoted as $T_2$.

(3) Frequent $K-$-term mining $K>2$

After generating the set of candidate $K-$items $C_k$, let the number of candidate $k$-items in $C_k$ be $q_k$, and arrange the candidates in $C_k$ into a $q_k$-dimensional vector $Q_k$ according to the subscripts of frequent 1-items. Construct a $m'\times n$ -dimensional matrix $U^k$ from $Q_k$ and $T_k$. When $q_k>m$, $m'=q_k$. When $q_k<m$, $m'=m$, in which case $U^k$ consists of $C_k$ and rows of zero vectors as shown in equation (12):

$$U^k=\left(U_1^k,U_2^k,\cdots,U_m^k\right)'=\left(C_k,C_k,\cdots,C_k,0,\cdots,0\right)'=\left(Q_k,0,\cdots,0\right)' \tag{12}$$

Let the $m$-dimensional matrix $W=(w_{ij})_{m\times m}=U^k\cdot T_k'$, and define the special function $f(w_{ij})$ as shown in Eq. (13):

$$f(w_{ij})=\begin{cases}1 & w_{ij}=k \\ 0 & w_{ij}<k\end{cases}, i, j=1,2,\cdots,m \tag{13}$$

For matrix $W$, $f(W)$ is shown in equation (14):

$$f(W)=f(U^k\cdot T_k')=\begin{pmatrix} f\left(\sum_{i=1}^{n} u_{1i}t_{1i}\right) & \cdots & f\left(\sum_{i=1}^{n} u_{1i}t_{mi}\right) \\ \vdots & \ddots & \vdots \\ f\left(\sum_{i=1}^{n} u_{mi}t_{1i}\right) & \cdots & f\left(\sum_{i=1}^{n} u_{mi}t_{mi}\right) \end{pmatrix} \tag{14}$$

$$i, j=1,2,\cdots,m$$

After that, construct a $m$ dimensional column vector $v$ from $W$, as shown in equation (15). $v(v_i>0)$ can get the support of each candidate $K-$ term $Sup(C_k)=v_i/m;$, $v_i>0$, compared with the minimum support degree $Sup_{\min}$, that is, you can get the frequent $K-$ itemset $L_k$, and the number of elements in the $L_k$ is represented by $P_k$. Remove elements in the matrix $T_k$ and rows less than $(k+1)$, and name the modified matrix $T_{k+1}$. Namely:

$$v=(v_i)_{m\times 1}=\left(\sum_{i=1}^{m} f(w_{1i}),\sum_{i=1}^{m} f(w_{2i}),\cdots,\sum_{i=1}^{m} f(w_{mi})\right)' \tag{15}$$

For $K = K+1$, if the number of elements $P_{k-1}$ contained in the frequent $(K-1)-$ itemset $L_{k-1}$ is $P_{k-1} < 2$, i.e., the candidate $K$-itemset $L_{k-1}$ generated by utilizing the frequent $(K-1)-$ itemset $L_{k-1}$ is $L_k = \varnothing$, then the The algorithm stops. Otherwise, continue mining for frequent $(K+1)-$ itemsets.

#### II. B. 4) Measurement of interest

The rationality of the interest degree measure of association rules directly determines the quality of the rules, which in turn affects the quality of propagation. Interest degree: for any rule $R: A \to B$ in the transaction database, Piatetsky-Shapiro defines its interest degree as the probability of simultaneous occurrence of $A$ and $B$ in the transaction database minus the product of the probability of occurrence of $A$ and the probability of occurrence of $B$, i.e., the famous PS formula, as follows:

$$I(R) = P(A,B) - P(A)P(B), I \in [-0.25, 0.25] \tag{16}$$

$$Crit(R) = \begin{cases} 1 - P(A,\overline{B}), & I(R) \geq 0 \\ 1 - P(A,B), & I(R) < 0 \end{cases}, Crit \in [0,1] \tag{17}$$

Obviously, when $P(A,B) = P(A) = 1$, $Crit(R)$ has maximum value 1. When $P(A) = 1$ and $P(A,B) = 0$, $Crit(R)$ has minimum value 0.

User index: let the minimum support that the user is most interested in be $SupC_{\min}$, so that $(SupC_{\min})^{\mu} = 0.5$, then $\mu$ is called the user index, which is computed as shown in Eq. (18):

$$\mu = \log_{SupC_{\min}} 0.5 \tag{18}$$

From the nature of Piatetsky-Shapiro formula, which is the traditional interest measure for association rules, it can be seen that when the support degree is equal to 0.5, the interest degree of association rules has a maximum value of 0.25. However, in practical applications, considering that different users are very likely to be interested in the rules with the support degree lower than 0.5, for example, the inverse rule. Therefore, it is a feasible solution for users to set the support degree by themselves.

#### II. B. 5) Contextualization

Context is defined as the situation, environment or scene information, such as time, weather and location, that describes the event or the person's behavior at the time of the event. Context is divided into five categories: environmental context that describes the characteristics of the user's environment, user context that describes the user's physiological and psychological characteristics and state, behavioral context that describes the user's tasks and behaviors, social context that describes the user's interpersonal relationships and other social characteristics, and spatiotemporal context that describes the temporal and spatial attributes.

#### II. B. 6) Considering the interest level of a situation

Context weight: For the item set $I_i \in I$ in the transaction database $D$, its context attribute $S_i$ is represented by the quaternion $(Os, Sh, W, Dct)$, where $Os$ denotes the Civics course, $Sh$ denotes the Civics content, $W$ denotes the dissemination path, and $Dct$ denotes the dissemination effect. Let the user context be $S'$, for any rule $R: A \to B$, $A, B \in I$ in the transactional database, the definition of context weight of $R$ is shown in Eq. (19), which embodies the dependency between $A$, $B$ under the context $S'$. Namely:

$$\begin{aligned} &Ctx(R) = \mathrm{Re}l(S', S_A) \cdot \mathrm{Re}l(S', S_B) \\ &\mathrm{Re}l(S', S_I) = \big((Os' \cap Os_I) + (Sh' \cap Sh_I) + (W' \cap W_I) + (Dct' \cap Dct_I)\big)/n \end{aligned} \tag{19}$$

where $\mathrm{Re}l(S', S_I)$ is the correlation between the context $S'$ and the item item $I$ context attribute. When the correlation between $S'$ and the item item is 1, it means that the two are related, i.e., the context $S'$ is a sufficient condition for the appearance of the rule $R$. When $0 < S' < 1$, whether rule $R$ holds or not is determined by context $S'$. When the correlation degree is 0, it indicates that there is no connection between the two, i.e., the context $S'$ denies the existence of rule $R$.

Improving the degree of interest: for any rule $R: A \to B$ in the transaction database, the degree of interest is defined in this paper as shown in Equation (20):

$$I(R) = \left(P(A,B) - P(A)P(B)\right)^{\mu} \cdot Crit(R) \cdot Ctx(R)$$
$$V \in [-1,1], I \in [-0.25, 0.25] \tag{20}$$

As can be seen from Eq. (20), the newly defined interest measure formula integrates the effects of rule validity as well as the user's subjective preference on the rule's interest measure. The formula is in line with the three principles of defining the measure of rule interest proposed by Piatetsky-Shapiro. Meanwhile, the range of values of the formula is the same as that of the original PS formula, and the interval of interest degree values is symmetric about the zero point: for any rule $R$, $I(R) \in [-0.25, 0)$ when its premise and conclusion are positively correlated. When independent, $I(R) = 0$. When negatively correlated, $I(R) \in (0, 0.25]$. Thus, the improved interest degree formula (20) maintains the following advantages of the original interest degree formula:

(1) The interval of interest degree is $[-0.25, 0.25]$, and the interval of interest degree for positive and negative correlations is symmetric.

(2) The absolute value of the interest degree of a rule is small regardless of whether the support or confidence of the rule is too large or too small.

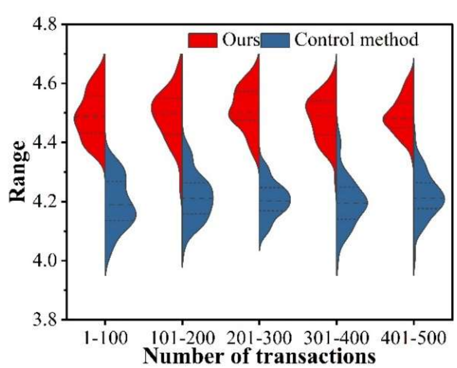
# III. Empirical analysis of ideological and political education content dissemination

## *III. A. Validation Analysis of Association Rule Mining Algorithm*

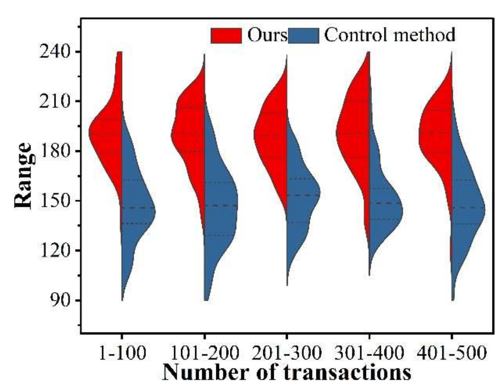
### III. A. 1) Satisfaction analysis

In order to verify the effectiveness of the association rule mining algorithm in this paper, the association rule mining algorithm based on transaction matrix and user interest level is compared with the traditional association rule mining algorithm. Conclusions are drawn by analyzing user satisfaction. Subjective satisfaction is obtained by averaging the user's ratings of ideological and political education content distribution paths, and objective satisfaction is the average of the user's browsing time of ideological and political content.

In order to realize the comparison of the number of different ideological and political education contents, it is necessary to set up different transaction databases for rule updating, and the rule updating runs the algorithm of this paper according to all the transactions in the transaction databases to replace the previously generated rules, and the rules are updated once for every 200 access records. In the beginning stage of propagation, the rules have not been generated yet, so the first 200 transactions are propagated by traditional association propagation. Only after the mining algorithm in this paper, it makes the ideological and political education dissemination efficiency and satisfaction significantly improved. This experiment is tested by setting the number of 50, 100 and 200 respectively, and the test results are shown in Fig. 3, Fig. 4 and Fig. 5, where (a)~(b) are subjective satisfaction and objective satisfaction respectively. When the number of ideological and political education resources is 50, 1, 2 in the two tests due to the small number of transactions, this paper's method (association rule mining algorithm based on transaction matrix and user interest degree) has no obvious advantage over the control method (traditional association rule mining algorithm). In the later tests, with the increasing number of transactions, the user satisfaction of this paper's method increases, while the control method does not have much change, which indicates that the ideological and political education content dissemination path and user satisfaction are improved after adding the transaction matrix and user interest degree on the basis of the traditional association rule mining algorithm. Such a direction is verified in the number of 100 and 200 tests, the change is more obvious, and the user satisfaction are rapidly increasing.

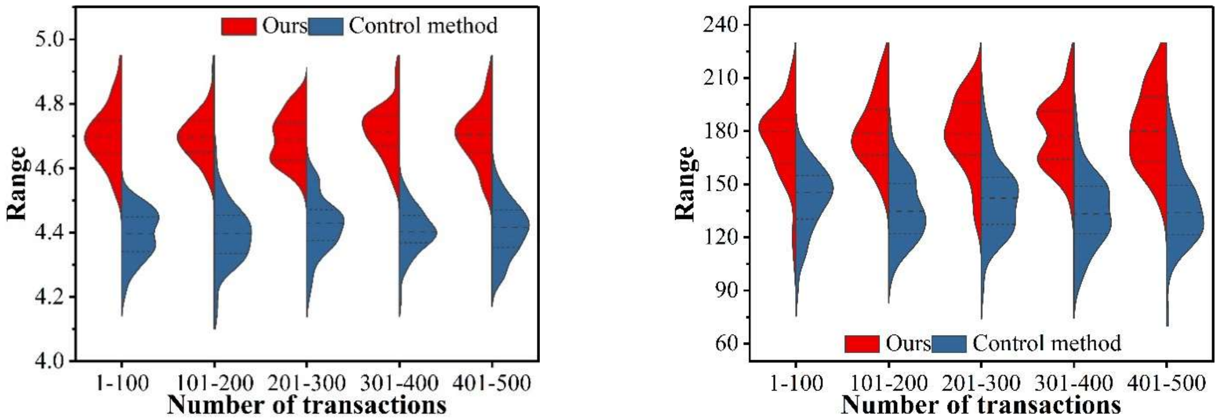



(a)Subjective satisfaction



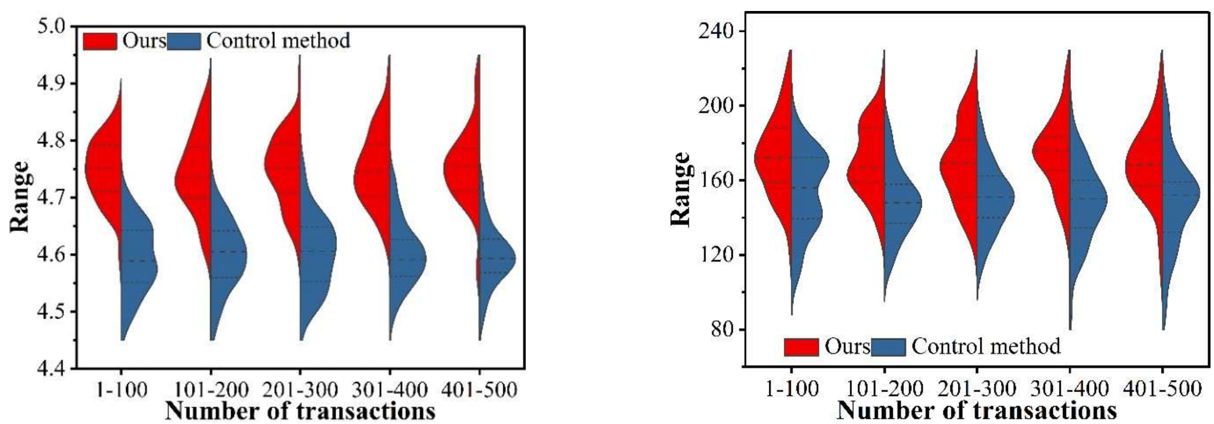

(b)Objective satisfaction

Figure 3: The number of test results is 50




(a)Subjective satisfaction (b)Objective satisfaction

Figure 4: The number of test results is 100




(a)Subjective satisfaction (b)Objective satisfaction

Figure 5: The number of test results is 200

**III. A. 2) Algorithm performance analysis**

In order to verify the correctness of the algorithm and to compare its execution efficiency with the traditional Apriori algorithm, simulation experiments are conducted in this paper. These experiments include comparing the execution efficiency of the two algorithms for different minimum support, testing the execution efficiency of the improved algorithm for different sizes of data, testing the propagation effect of the algorithm for different groups of data, and evaluating the performance of the two algorithms reasonably while obtaining the experimental results.

The test database is Mushroom Database, Mushroom is characterized by a dense distribution of frequent itemsets, which can produce a large number of frequent itemsets even with a large support. This database has 8124 records, the number of attributes i.e. items is 119, 23 attributes such as cap color, neck shape, color, whether poisonous or not of mushrooms are recorded, i.e. the average length of things is 23.Mushroom is a commonly used database to test the association rules.Mushroom dataset has a size of 465KB, which is imported into an Access database for operation.

Test environment: Pentium IV 2.93GHz, 1.0G RAM, Windows XP. The Apriori algorithm implemented in C++ language is ported to Visual Studio.net 2005 platform.

(1) Comparison of Execution Efficiency under Different Support Levels

The execution efficiency of the two algorithms, Apriori algorithm and the improved algorithm, for different degrees of support are experimented separately. The results of the experimental tests are shown in Fig. 6, from which it can be seen that the execution efficiency of the improved algorithm is higher than that of the Apriori algorithm for different degrees of support. And the superiority of the execution efficiency of the improved algorithm compared to the Apriori algorithm will be more obvious when the support degree is small.

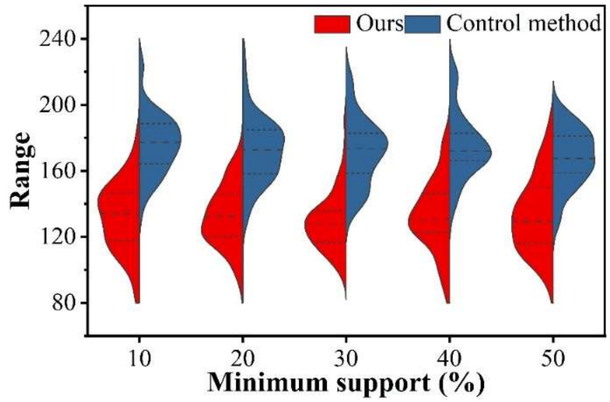



Figure 6: Two algorithms run time on Mushroom

(2) Data experiments for algorithms of different sizes

We take the minimum support of 50% for the growing number of transactions as the test experimental data, and increase the number of transactions from 1 to 500 to experiment on the execution efficiency of the algorithm. The experimental test results are shown in Fig. 7. With the increasing number of transactions, the execution efficiency of this paper's algorithm grows almost linearly, which indicates that both this paper's algorithm and Apriori algorithm have good stability.

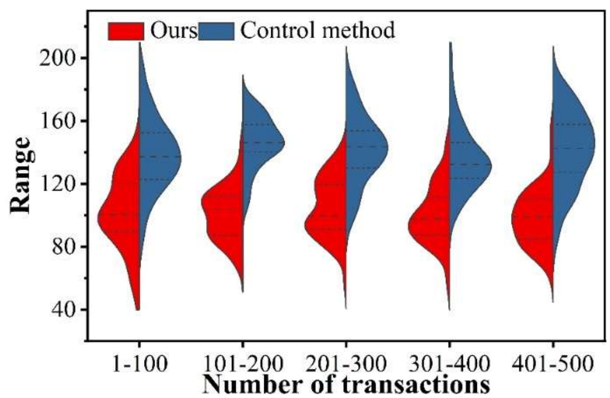



Figure 7: Compare execution times under different transaction records

(3) Comparison of dissemination effect

Experiments were conducted to carry out two kinds of ideological and political education content dissemination effect, and under the same experimental conditions, the ideological and political education content dissemination effect of the method of this paper was verified to be higher. When the target user is a new user, multiple rounds of experiments are conducted to verify the quality of the dissemination effect of ideological and political education content when the new user.

The proximity between the predicted rating of the dissemination effect and the real rating given by the user is measured by the average absolute error, which is calculated by the following formula:

$$MAE = \frac{\sum_{i=1}^{n} |p_i - q_i|}{n} \tag{21}$$

where $n$ represents the number of scoring data, $p_i$ denotes the predicted scoring of the dissemination effect, and $q_i$ is the real scoring.

The formula for utilizing F1 as a comprehensive evaluation index of checking completeness and checking accuracy is specified as follows:

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{22}$$

where P represents the check accuracy rate and R denotes the check completeness rate.The larger value of F1 represents the worse effect of ideological and political education content dissemination.

The experimental dataset is cut by the five times cross-validation method to obtain the test set and training set in the experiment, whose number is five, of which there are about 6,517 data in the training set and 2,785 data in the test set.

In the state of the same experimental environment and experimental conditions, the algorithm is tested for the average absolute error and F1 value by different test sets and training sets. The average absolute error experimental data is shown in Fig. 8, and the F1 value experimental data is specifically shown in Fig. 9. According to the analysis of the data in Fig. 8 and Fig. 9, compared with the traditional association rule mining algorithm, the F1 value and MAE value of this paper's method (association rule mining algorithm based on transaction matrix and user's interest degree) are both significantly lower, which indicates that the ideological and political education content dissemination effect is better.

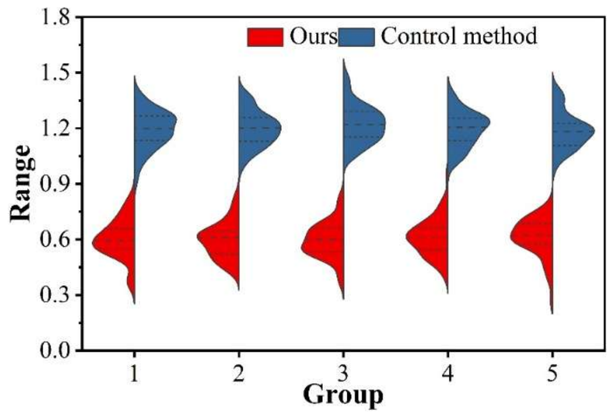



Figure 8: Mean absolute error experimental data

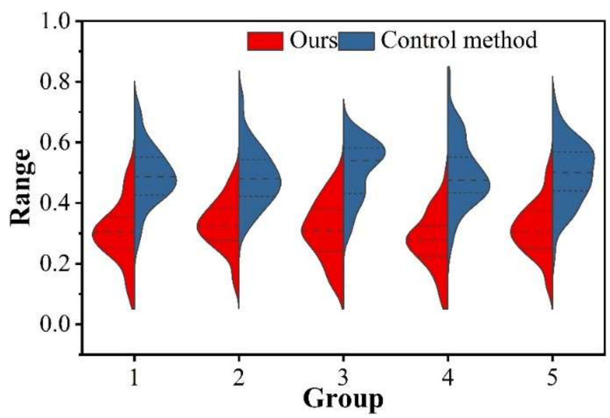



Figure 9: F1 value experiment data is specific

In the state of the same experimental environment, experiments are conducted on five sets of training sets and five sets of test sets. In the experiment, the new user is set as the target user, and the method of this paper is utilized to disseminate the Civics theory resources to the target user, and the experimental data of the average absolute error in the experiment is shown in Fig. 10, and the experimental data of the F1 value in the experiment is shown in Fig. 11. According to the data analysis in Fig. 10 and Fig. 11, when the target is a new user, the F1 values of the five groups of this paper's algorithm are 1.073, 1.082, 0.983, 0.997, 0.951, respectively, and the corresponding MAE values are 0.085, 0.087, 0.089, 0.106, 0.0901, which indicates that this paper's method can better realize the ideology and politics of the new user-oriented education content dissemination for new users.

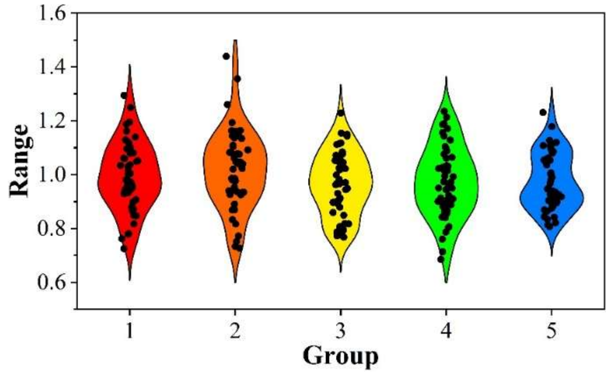



Figure 10: The average absolute error in experimental data

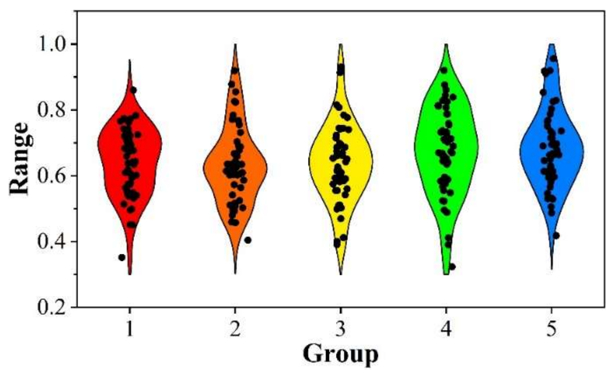



Figure 11: F1-number experimental data in the experiment

### *III. B. Correlation Analysis of Ideological and Political Education Communication Paths*

The three most important steps in correlation analysis algorithm for data mining are data collection, data preprocessing and data analysis.

#### III. B. 1) Access to data

In this paper, a questionnaire is used to obtain the basic evaluation of college students on the ideological and political education dissemination path based on the association rule mining algorithm, which contains a total of 47 questions covering different aspects such as learner factors, instructor factors, environmental factors, etc. The questionnaire is conducted in the form of a web page. The questionnaire was conducted in the form of a web page, and the resultant data of the survey was exported through the Excel document format, with the first row storing the title of each question, and the data below it being the index value of the options. In order to reduce storage redundancy, the content of each option was stored centrally in other documents.

#### III. B. 2) Data pre-processing

The questionnaire data in Excel should first be processed by adding the code of the question in front of the index value of the options of each question, so as to distinguish the same index value of different questions, and then imported into RStudio through the Excel toolkit of R language. The questionnaire data should first be converted into transition data format before correlation analysis, so the data need to be converted into List format first, and then converted into transaction format.

#### III. B. 3) Data analysis

Transaction format data can begin to carry out the association analysis, in this paper association rule mining algorithm, we set the minimum support for 0.2, the minimum confidence level of 0.9. see that the rules that meet the requirements of more than 160,000 rules, of which the length of 2 rules are only 75 rules, the length of 3 more than 2,000 rules, the length of the rules of more is too complex and redundant, do not serve as the analysis of the main target. We modified the parameters of the algorithm so that the maximum number of association rules is 3, which means that only association rules such as A->B and A&B->C are exported. Then the analysis results are arranged according to the degree of enhancement, and the following rules are obtained, and the results of the association analysis of ideological and political education content dissemination paths are shown in Table 1. The above are the first 15 rules that have been ranked in descending order of the degree of enhancement, among which the largest degree of enhancement reaches 3.739.

Table 1: Correlation analysis of propagation paths

| N | Rule A | Rule B | Support | Confidence | Lift |
|---|---|---|---|---|---|
| 1 | {Question 1,1 question 7, 0} | {Question 8,0} | 2.066 | 0.918 | 3.739 |
| 2 | {Question 4,3 question 8, 0} | {Question 1,1} | 2.068 | 0.98 | 3.641 |
| 3 | {Question 6,0 question 8, 0} | {Question 1,1} | 2.069 | 0.937 | 3.549 |
| 4 | {Question 5,1 question 8, 0} | {Question 1,1} | 2.095 | 0.944 | 3.232 |
| 5 | {Question 7,0 question 8, 0} | {Question 1,1} | 2.104 | 0.977 | 3.163 |
| 6 | {Question 8,0} | {Question 1,1} | 2.125 | 0.934 | 2.958 |
| 7 | {Question 4,3 question 8, 0} | {Question 7,0} | 2.167 | 0.985 | 2.942 |
| 8 | {Question 1,1 question 8, 0} | {Question 7,0} | 2.169 | 0.958 | 2.701 |
| 9 | {Question 5,1 question 8, 0} | {Question 7,0} | 2.175 | 0.984 | 2.662 |
| 10 | {Question 1,1 question 8, 0} | {Question 6,0} | 2.184 | 0.932 | 2.579 |
| 11 | {Question 8,0} | {Question 7,0} | 2.195 | 0.925 | 2.482 |
| 12 | {Question 6,0 question 8, 0} | {Question 7,0} | 2.229 | 0.983 | 2.461 |
| 13 | {Question 5,1 question 8, 0} | {Question 6,0} | 2.257 | 0.967 | 2.172 |
| 14 | {Question 4,3 question 8, 0} | {Question 6,0} | 2.287 | 0.99 | 2.094 |
| 15 | {Question 1,1 question 7, 0} | {Question 6,0} | 2.297 | 0.987 | 2.067 |

Firstly, all the rules are visualized and analyzed, and a number of valid rules are grouped and presented in the form of bubble charts, Figure 12 shows the rule grouping view. The horizontal coordinate is the left operation of the grouped rules, the vertical coordinate is the right operation of the grouped rules, the circle size is the support of the rules, the larger the circle the higher the support, the color is the enhancement of the rules, the darker the color the higher the enhancement. It is found that the support value of the association rules ranges from 2 to 2.5, while the corresponding enhancement degree is from 2.1 to 3.8, which indicates that the algorithm in this paper can accurately reflect the real development of the dissemination path of ideological and political education in colleges and universities, and it has a guiding and referential value for the development and innovation of ideological and political education in colleges and universities.

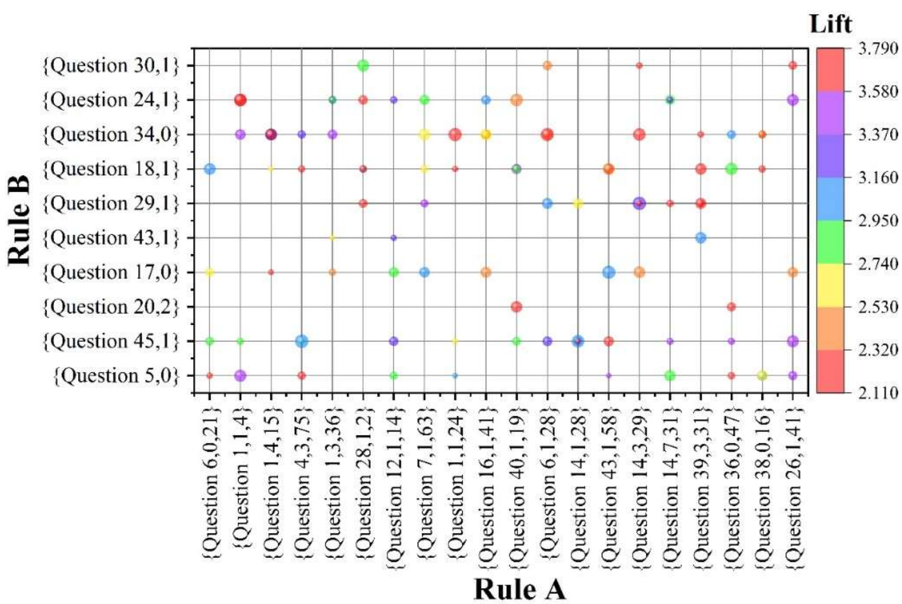



Figure 12: Rule group view

We can analyze the rules for a single outcome, e.g., Question 1 1 represents that the respondent is a student, and we do a special analysis of the rules associated with it. Figure 13 shows a directed graph of the special analysis rules, where the direction of the arrow points to the direction of the derivation of the rule. Figure 14 is a scatter plot of the rule analysis, where the horizontal coordinate is the support, the vertical coordinate is the confidence, and the color of the scatter represents the degree of enhancement, and the darker it is, the higher the degree of enhancement. Based on these rule analyses, we can combine the questionnaire to derive the association relationship between different questions and answers. Based on the association rule mining analysis technology introduced to the ideological and political education effect content dissemination, both the ideological and political dissemination and path optimization are achieved.

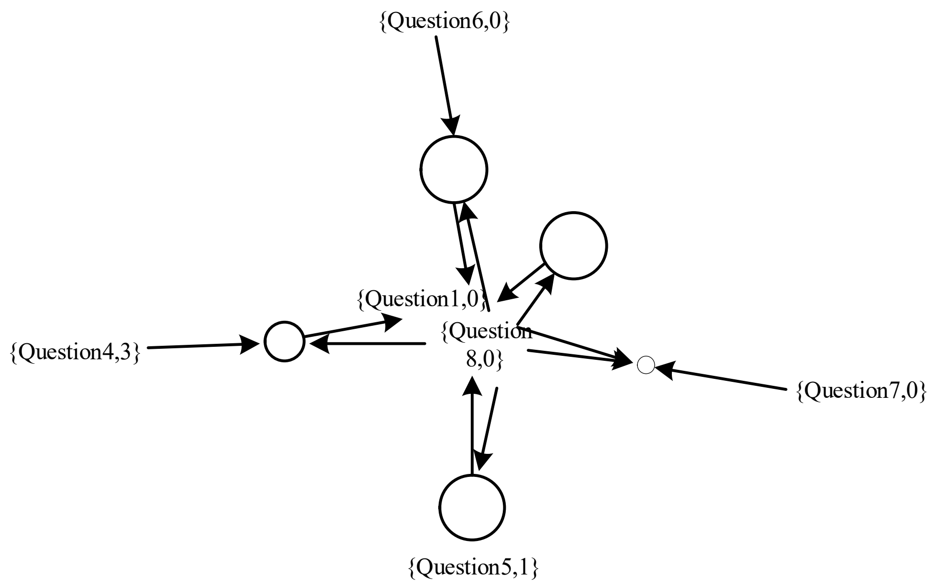



Figure 13: Specific analysis rule directed graph

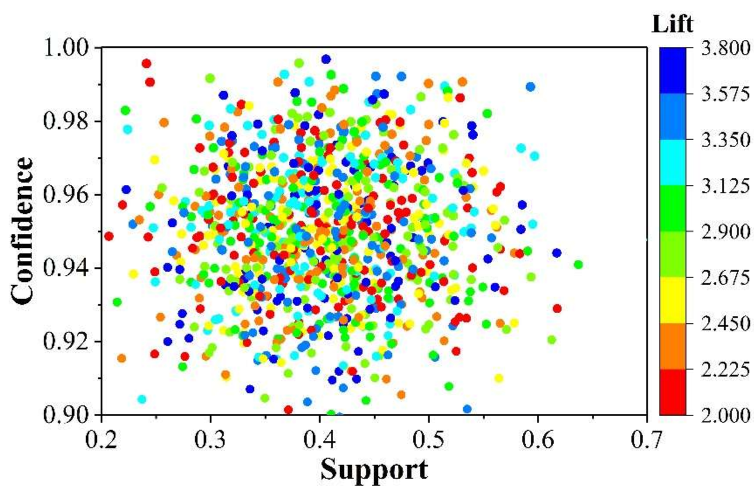



Figure 14: Rule analysis scatter plot

## IV. Conclusion

On the concept of data mining technology, this paper proposes an association rule mining algorithm based on transaction matrix and user's interest degree, in order to achieve the effect of ideological and political education content dissemination path optimization, and use the algorithm to carry out empirical analysis of ideological and political education content dissemination path. When the number of ideological and political education resources is 50, the subjective satisfaction value of this paper's algorithm performs better than the traditional association rule mining algorithm, and there exists a type of law when it is 100 and 200. In addition, it is found that this paper's algorithm can accurately reflect the current ideological education content dissemination path and the current situation, which verifies the application value of this paper's algorithm in the optimization of ideological and political education content dissemination path, and accelerates the development and dissemination of ideological and political education in colleges and universities.

## References

[1] Ampadu, Y. B. (2023). Handling big data in education: a review of educational data mining techniques for specific educational problems. AI, Computer Science and Robotics Technology, (13).
[2] Hooda, M., & Rana, C. (2020). Learning analytics lens: Improving quality of higher education. International journal of emerging trends in engineering research, 8(5).
[3] Li, Y., & Mao, H. (2022). Study on machine learning applications in ideological and political education under the background of big data. Scientific Programming, 2022(1), 3317876.
[4] Zhao, M., Ning, Y., & Yang, Y. (2020). A new thought analysis on the ideological and political education of college students under the socialist core values. In Annual Conference of Education, Teaching and Learning (Vol. 21, p. 432).
[5] Zhang, H., & Yuan, C. (2020). Research on University Students' Ideological and Political Education in the New Era. Open Journal of Social Sciences, 8(6), 217-221.
[6] Gao, H. W. (2023). Innovation and development of ideological and political education in colleges and universities in the network era. International Journal of Electrical Engineering & Education, 60(2_suppl), 489-499.
[7] Zhu, G., & Tao, T. (2021). Strategies for Ideological and Political Education in Colleges and Universities from the Perspective of General Education. Academic Journal of Humanities & Social Sciences, 4(11), 18-23.
[8] 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.
[9] Zhang, Q. (2025). Research Status and Evolutionary Trends of Student Ideological Education: A Knowledge Mapping Analysis Based on CiteSpace. Studies on Religion and Philosophy, 1(1), 109-120.
[10] Ma, Y., Sun, X., & Ma, A. (2025). On the assessment and reliability of political and ideological education in colleges using deep learning methods. Alexandria Engineering Journal, 119, 511-517.
[11] Hong, N., & Liu, X. (2025). Design of an IoT-Based Intelligent Platform for Ideological and Political Education Management. International Journal of High Speed Electronics and Systems, 2540281.
[12] Ma, J. (2021, April). University ideological and political education management based on K-means mean value algorithm. In Journal of Physics: Conference Series (Vol. 1852, No. 4, p. 042023). IOP Publishing.
[13] Shuang Wang,Chenhui Bao & Dongmei Pei. (2024). Application of Data Mining Technology in the Screening for Gallbladder Stones: A Cross-Sectional Retrospective Study of Chinese Adults. Yonsei medical journal,65(4),210-216.
[14] Wu Ling,Tian Yanru,Lu Jinlu & Guo Kun. (2025). AAHIN: attribute-aware heterogeneous information network representation learning for data mining. International Journal of Web Information Systems,21(2),158-179.