

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Improved Probabilistic Frequent Itemset Analysis Strategy of Learning Behaviors Based on Eclat Framework

Xiaona Xia

Abstract

Interactive learning environment is the key support for education decision making, the corresponding analytics and methodology are the important part of educational technology research and development. As an important part and the research challenge, learning behaviors are uncertain and produce complex data relationships, which makes the learning analysis process more difficult. This chapter studies the feasibility of Eclat framework applying in educational decision making and get the corresponding the data analysis results. We take probabilistic frequent itemsets and association rules as research objectives, extract and standardize multiple data subsets; Based on Eclat framework, using data vertical format, we design and improve the models and algorithms in the process of data management and processing. The results show that the improved models and algorithms are effective and feasible. On the premise of ensuring robustness and stability, the mining quality of probabilistic frequent itemsets and association rules is guaranteed, which is conducive to the construction of key execution topology of learning behaviors, and improves the accuracy and reliability of data association analysis and decision prediction. The whole analysis methods and demonstration processes can provide references for the study of interactive learning environment, as well as decision suggestions and predictive feedback.

Keywords: Learning Analytics, Decision Making, Eclat Framework, Probabilistic Frequent Itemsets, Association Rules, Decision Prediction, Interactive Learning Environment

1. Introduction

Content resources, interaction patterns, collaborative models, organizational planning and influencing factors related to learning processes constitute learning behaviors, which are also key elements to describe learning behaviors [1, 2]. The learning processes supported by online technology and data technology ensure the completeness and continuity of learning behavior data. Massive learning behaviors is an important part of education big data, which provides the possibility for the full development of learning analytics [3, 4]. Learning behavior data can be divided into two categories: horizontal format and vertical format from the perspective of data

structure and feature attributes. These two categories are inseparable about the components of learning behaviors, which are the atomic units to describe learning behaviors, such as url, forumng, questionnaire, etc. The horizontal format of learning behaviors is a vector set composed of multi-dimensional attributes, and the vertical format is a vector set composed of multi-level learners. From the perspective of horizontal format, the researches define learning behaviors as the collection of learners, appropriate learning analytics and tools are used to carry out data statistics and rule exploration. However, it is difficult to calculate and compare the influences of components of learning behaviors, which is not conducive to the construction of a new education mode, and it is relatively difficult to implement the calculation and comparison of the influences of learning behaviors more passive.

Learning behaviors represent continuous learning processes, and there are associated needs and execution results [5]. The analysis of learning behavior based on vertical format can provide more intuitive and accurate characteristics for the study of the groupness and individuality of the learning behavior components. However, the analysis process based on the vertical format is a complex problem with multiple factors. It is impossible to find a suitable decision making and prediction framework. Through sampling, the breadth and depth of data processing are limited, and it is difficult to achieve a feasible decision. Due to the shortcomings and gaps in technology and model, learning behaviors constitute data and potential relationships cannot be gotten fully mining and complete analysis. In terms of research methods and application practices of learning behaviors, there are still many problems to solve [6, 7].

In this chapter, vertical data is analyzed for an online learning behavior big data set. The vertical data analysis of learning behaviors is carried out from the data structure and characteristics. Based on Eclat framework, a probabilistic frequent itemset learning algorithm is designed, and its feasibility and reliability are demonstrated and compared. Within the effective performance indicators, the probabilistic frequent itemsets and association rules are calculated and mined from the learning behavior components, and the correlation is demonstrated. Then we explore the rules and characteristics of learning behaviors, and provide decision feedback and suggestions for the design improvement and relationship of learning behaviors.

2. Related work

Mining probabilistic frequent itemsets is a branch of data analytics. There are explicit or implicit association data, which is the key basis for prediction, decision making and recommendation of other learning behavior components. On the current big data platform, the decision algorithm and recommendation algorithm based on frequent itemsets mining are used to track data. However, due to the particularity and complexity of learning behaviors, as well as the autonomy and randomness of learning processes, there is no general technical means to ensure the integrity and sufficiency to implement the analysis and calculation with the goal of decision making and prediction. In this regard, it is necessary to participate in benign learning behavior component construction and recommendation. The research on frequent itemsets has shown an urgent technical demand in the field of education big data. There have been relevant results to demonstrate the urgency and reality of empirical methods and technical means.

The research on probabilistic frequent itemsets of learning behavior components, after combing the relevant theoretical and application results, mainly focuses on the data statistics and association rules of horizontal format, which is reflected in the following aspects:

2.1 Frequent itemset mining based on apriori

Frequent itemset mining based on Apriori takes the construction of itemset association rules as the premise. The mining process is based on the horizontal format and completes the extraction of rules through iterative search strategy. After data connection and pruning, the itemsets satisfying the association rules are formed [8–11]. If one itemset satisfies the minimum support and a certain confidence, it is defined as a frequent itemset. Apriori algorithm is used to analyze the relevance of learning behaviors, the main idea is to select the learning platform, locate the components of learning behaviors, realize the association between learning behaviors and learning effects, define learning behavior as “cause”, and define learning effect as “result”. The traditional Apriori algorithm is improved flexibly. With the help of clustering, weighted balance, decision tree evaluation and other means, the data tracking are realized. The research target is to optimize learning behaviors and improve learning efficiency. However, the frequent itemset mining process of Apriori needs to scan the original data many times. When the original data is large, the number of times of iterative scanning is too much, which seriously affects the efficiency of the algorithms.

2.2 Frequent itemsets mining based on FP-growth

Frequent itemset mining based on FP-growth also uses horizontal data format, but the data structure is essentially different. The process of data analysis is mainly divided into two steps: constructing FP tree and mining frequent itemsets. Through the construction of FP tree, the expression of itemsets associated transaction is realized, that is, one path of FP tree corresponds to a transaction, and the transaction is composed of items. Different transactions may have the same items, which makes the path of FP tree overlap. The more overlapped, the greater the path compression space, the higher the access efficiency of FP [12–15]. FP-growth is used to mine frequent itemsets of learning behaviors. Its main idea is similar to Apriori. According to the research target of learning behaviors, users require to select the data set of learning behaviors, define the itemsets and research target, put forward hypothesis test, explore the rules by means of classification, clustering and decision making, draw a conclusion, and verify the existing education and teaching according to the data analysis results, but there are some problems. Due to the diversity, randomness and complexity of learning behaviors, FP-growth algorithm has obvious limitations in the study of learning behaviors. When the itemsets of learning behavior are too many or the relationships are complex, it will lead to too many sub nodes of FP tree, which will greatly reduce the efficiency of the algorithms, and can not get accurate and complete frequent itemsets. FP-growth algorithm is very difficult to learn.

2.3 Frequent itemsets mining based on Eclat framework

Compared with Apriori and FP(Frequent Pattern)-growth, The fundamental difference of Eclat is that the algorithm analysis of Eclat uses vertical data format, and is essentially a deep optimization search mechanism. The rule search space is effectively divided into subspace sets through concept lattice and equivalence relationships. The support calculation of each itemset does not require repeated retrieval of the entire dataset [16–19]. The main idea of using Eclat framework to study learning behaviors need the support of big data set of learning behaviors, through data transposition and standardization processing, we can get the itemsets and the transaction set. On this basis, the relevant models and algorithms of Eclat framework are improved and redesigned. On the premise of support, confidence and promotion, frequent itemsets

and association rules are mined. Taking the final frequent itemsets and association rules as the references. Vertical data analysis and research based on Eclat framework can improve the speed of data search, association and analysis, and also improve the reliability of data validation results to a certain extent.

However, the Eclat framework is rarely used in the data processing of learning behaviors. Therefore, the improvement of algorithms and models has no effective results, which is directly related to the difficulty of technology caused by the complexity of learning behaviors. If Eclat is used to transpose and intersect all items and transactions, or if the number of items and transactions is too large, the efficiency of the algorithms will be affected. Therefore, the mining of frequent itemsets in Eclat framework should be assisted by other algorithms and tools, which is more practical. This chapter will integrate the advantages and feasible attempts in the application of Eclat framework, such as technical improvement, model design, tool application, etc., so as to provide more effective methods for the follow-up study of big data of learning behaviors and others.

3. Elements of learning behaviors and research problems

We select a big data set of learning behaviors of UK open university in four periods in recent two years, and the data scale reaches hundreds of millions. From the perspective of course category, we realize the tracking and comparison of learning behaviors of the same category and different categories, and make adaptive decision. The courses are divided into two categories: Literature and Technology. For each category, two courses are selected, namely L1 and L2, T1 and T2. Different courses have different periods of learning behaviors, with the help of assessment, the learning effects are achieved. There is correlation between learning behaviors and learning effects, and there is mutual restrictive and driving relationships between the components of learning behaviors. The empirical problems and testing strategies are established between learning behaviors and learning effects, the research conclusions and decision making reflection are the basis for the improvement and optimization of data-driven learning behaviors.

Tables 1–4 show the components and indicators of learning behaviors corresponding to the four courses of L1, L2, T1 and T2. The four tables involve four learning periods: P1, P2, P3 and P4. The data distribution of the tables indicates that not all courses have learning behavior in each period. The indicators involve two statistics: the median and the mode, which are used to investigate the population trend. Different indicators are selected according to different types of components. “assessment” represents the assessment method of courses, that is composed of enumeration components, mainly including CT (Computer Test), TT (Teacher

P2		P4	
forumng	668	content	1080
homepage	369	resource	8
content	147	subpage	16
resource	1	url	0
assessment	TT	assessment	TT
final_result	pass	final_result	pass

Table 1.
Components and indicators of L1.

P3		P4	
forumng	23	forumng	26
homepage	106	homepage	112
collaborate	0	collaborate	0
content	17	content	24
page	0	page	0
quiz	276	quiz	312
resource	32	resource	34
subpage	56	subpage	54.5
url	4	url	4
assessment	TT	assessment	TT
final_result	With-drawn	final_result	With-drawn

Table 2.
Components and indicators of L2.

P2		P3		P4	
dualpane	3	dualpane	0	dualpane	0
forumng	112.5	forumng	73.5	forumng	126
homepage	195	homepage	160.5	homepage	189
collaborate	0	collaborate	1	collaborate	1
content	506	content	447.5	content	466
wiki	117	wiki	86.5	wiki	126
page	0	page	0	page	0
quiz	137	quiz	113	quiz	108
resource	7	resource	11	resource	9
subpage	22	subpage	17	subpage	19
url	27	url	27.5	url	22
assessment	TT	assessment	TT	assessment	TT
final_result	withdrawn	final_result	withdrawn	final_result	pass

Table 3.
Components and indicators of T1.

Test) and exam (computer and Teacher joint test); “final_result” represents the result of the course assessment and is also an enumeration type, including four components: excellent, pass, fail and withdrawn. “assessment” and “final_result” measure the group tendency of courses. Other components are the main parts of the interaction processes. They all describe the interaction frequency, which has the autonomy and randomness of learners. The strength of interaction frequency is assessed by the median to investigate the distribution range.

From **Tables 1–4**, we can see that the concentration of group selection of “assessment” is very obvious. Most of the learners have completed the course assessment by teachers, but the assessment results are quite different, and the assessment results of the same course in different learning periods are also different. About P4 of L2, as same as P2 and P3 of T1, learners tend to give up the assessment.

P1		P2		P3		P4	
dataplus	0	dataplus	0	dataplus	0	dataplus	0
dualpane	2	dualpane	0	dualpane	0	dualpane	0
forumng	229.5	folder	1	folder	1	forumng	150
glossary	0	forumng	183	forumng	143	glossary	0
homepage	282	glossary	0	glossary	0	homepage	229
content	795	homepage	234	homepage	201	htmlactivity	4
illuminate	8	collaborate	1	oucollaborate	1	collaborate	1
wiki	13	content	566	oucontent	482	content	638
page	9	wiki	11	ouwiki	8	wiki	9
questionnaire	3.5	page	7	page	7	page	2
quiz	581.5	questionnaire	3	questionnaire	0	questionnaire	2
resource	32	quiz	543	quiz	521	quiz	557
subpage	219.5	repeatactivity	0	resource	22	repeatactivity	0
url	23	resource	26	subpage	162	resource	25
assessment	TT	subpage	180	url	12	subpage	184
final_result	Pass	url	13	assessment	TT	url	14
		assessment	TT	final_result	pass	assessment	TT
		final_result	pass			final_result	excellent

Table 4.
Components and indicators of T2.

In P4 of T2, most of the learners obtain excellent assessment results, and most of them pass the course. From “assessment” and “final_result”, the group indicators of Literature courses and Technology courses are similar.

As for other components of learning behaviors, it can be found from the data that the category of components and the participation of isomorphic components show strong discrete characteristics. The results show that the types of interaction components in two learning periods of L2 and three learning periods of T1 are consistent, and the median is relatively close, which indicates that the distribution of learners’ participation in these interactive components is basically consistent. The two learning periods of L2 have the same “final_result” mode, and the assessment results of T1 have obvious differences. The comparison of the types or numbers of interaction components related to the same course in different learning periods directly shows the differences. The interactions are significantly different, and there is a gap in the median of the same interaction component, such as “content” of two learning periods of L1. At the same time, the types of interaction components that belong to Literature or Technology courses are subject to the courses. The learners of L1 and L2 have their own interactive components, and T1 and T2 are the same.

Therefore, their interaction components of L1, L2, T1 and T2 reflect the autonomous learning characteristics, and the component constraints of assessment methods and results realize the differentiation of learners. The problems and relationships are shown in **Figure 1**, which is divided into the following four steps:

1. The mining of frequent itemsets will take different interaction components as reference items, and realize the analysis and mining of frequent itemsets based on reference items according to certain probability;

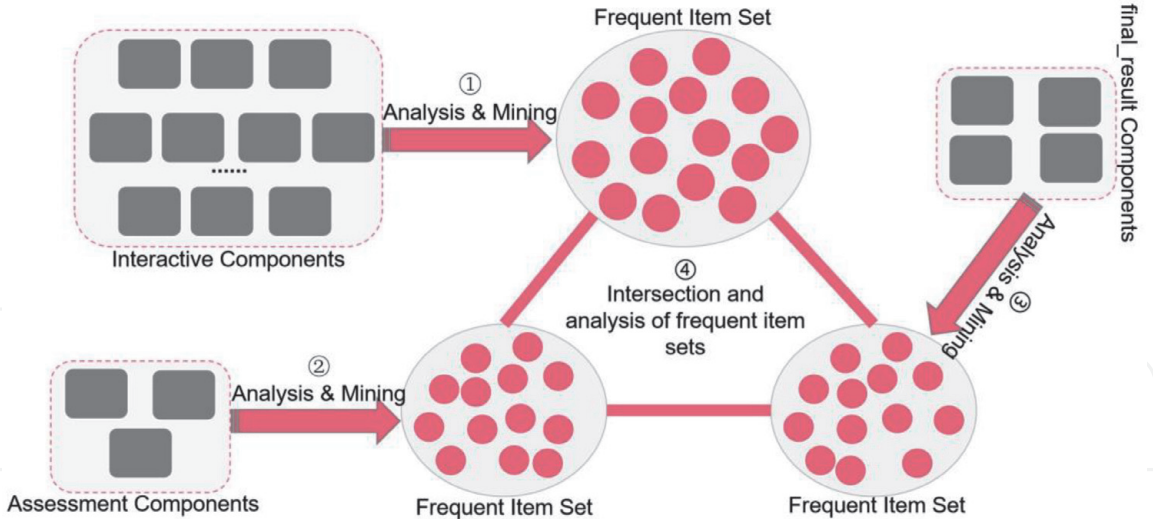


Figure 1.
The research problems and logical relationships.

2. Taking three enumeration methods of “assessment” as component reference items, according to certain probability, the frequent itemsets analysis and mining based on reference items are realized;
3. The four enumeration methods of “final_result” are component reference items. Based on a certain probability, the analysis and mining of frequent itemsets based on reference items are realized;
4. Based on a certain probability, the intersection of the three groups of frequent itemsets obtained from (1), (2) and (3) is solved and analyzed, and the inherent association logic and restrictive conditions are evaluated. On this basis, the rule of data-driven learning behaviors, prediction direction and decision making are explored.

The certain probability in the four steps depends on the selected algorithm requirements and measurement support. Based on the improved Eclat framework, we complete the four steps of the research problems, uses the three indicators “Support”, “Confidence” and “Lift” to realize constraints, analyzes threshold and test criteria, and mines probabilistic frequent itemsets and association rules.

4. Improved Eclat framework

For the four learning behavior datasets corresponding to L1, L2, T1 and T2, the execution results of the reference items can be described in the form of probability, but not the “Support” calculation mode. The expected “Support” of reference items should be used to describe the execution frequency of uncertain components [20], that is a feasible and target analysis strategy, which is the model basis for improving the Eclat framework.

4.1 Related models

The Related Models for the improvement of Eclat framework are as follows.

4.1.1 Expected “Support” of reference items

Given a probabilistic data set with N reference item instances, the expected “Support” of a reference term X is expressed as the cumulative value of the probability in the probabilistic data set. The calculation formula is $expect - sup(X) = \sum_{i=1}^N p_i(X)$.

4.1.2 Frequent itemsets

Based on the expected “Support” of a reference item, a probabilistic data set with N reference item instances is given, if it meets $expect - sup(X) \geq N \times \min_RST$, the reference item X is a frequent item set. \min_RST is the minimum relative “Support” threshold, which is calculated by the ratio of the minimum absolute “Support” threshold to the reference item instance. Generally, this value can be specified according to the data distribution.

4.1.3 Probability frequency

Combined with the conditions of frequent itemsets, given a probabilistic data set with N reference item instances, the probability frequency of the reference term is defined as: $proF(X) = proF\{expect - sup(X) \geq N \times \min_RST\}$.

4.1.4 Probabilistic frequent itemsets

Given a probabilistic data set with N reference item instances, if meeting $proF(X) \geq \min_proF$, the reference item X is a probabilistic frequent itemset, \min_proF is the minimum frequent probability threshold, which can also be specified according to the data distribution.

4.2 Algorithm design

Many algorithms for mining frequent itemsets mostly use horizontal data format with transaction as vector [5, 21]. The uncertainty of learning behavior data makes the analysis of learning behavior need vertical data format. One complete learning behavior of learners constitutes a transaction. Based on Eclat framework, it is suitable to adopt *tidlist* data structure, and add a probability parameter to each item of learning behaviors to indicate the possibility of a specific transaction.

The vertical data format of learning behavior data is a binary tuple $(x, tidlist(x))$, which represents the item set of learning behaviors, and x is the identifier of each item, that is, the number of each learning behavior, $tidlist(x)$ is the list of items of x . If each item contains an identifier i_i and an existence probability $p_X(i_i)$, $tidlist(x)$ is expressed as a tuple $\{(i_1, p_x(i_1)), (i_2, p_x(i_2)), \dots, (i_i, p_x(i_i))\}$. In the algorithm design of vertical data format, it is necessary to complete the calculation of probability frequency. Here we use two-dimensional array $P_x[i, j]$ to represent the probability quality function, which means the X probability of the i occurrence in the previous j reference items. Therefore, the calculation process of probability frequency is described as PFC(Frequent Pattern Calculation) program.

PFC program

Input: Item set $X(i_i, p_X(i_i)) // 1 \leq i \leq |I|$, I represents the maximum number of transactions.

Output: $P_x[i, j]$

Process

1. PFC()
2. For $j=0$ to $|I|$
3. $P_x[0, j]=1$
4. EndFor //Initialize the first row units of $P_x[i, j]$ of a to 1
5. For $j=0$ to $|I|$
6. For $i=0$ to $\min_Value(j, \min_RST)$ // $\min_Value(j, \min_RST)$ is used to compare j and \min_RST , then return the minimum value.
7. If $i > j$ then $P_x[i, j]=0$
8. Else if $i > j$ then $P_x[i, j] = \prod_{i=1}^j p_x(i_i)$
9. Else if $i < j$
10. Then $P_x[i, j] = P_x[i-1, j-1] \cdot p_x(i_j) + \max(P_x[i, j-1], P_x[i-1, j]) \cdot (1 - p_x(i_j))$
- /*This formula is a kind of dynamic decision programming, and the maximum probability frequency is obtained by the adjacent units.*/
11. End If
12. End If
13. End If
14. End For
15. End For
16. Output: $P_x[i, j]$

Based on the calculation results of probability frequency, Eclat algorithm is designed. There are three main steps:

Firstly, according to the vertical data format, the transactions and corresponding items are extracted from the learning behavior data set, with the help of bi-directional sorting strategy, transactions are initialized. The items are stored according to *tidlist*. Then, it analyzes the “Support” of the transactions stored in *tidlist*, and discards the transactions with lower “Support” (support $< \min_RST$).

Secondly, the items of learning behaviors are pruned and optimized, and the k -item set from *tidlist* is extracted by intersection, and the probability frequency of k -item set is realized by multiplication operation.

Thirdly, mining probabilistic frequent itemsets recursively in candidate itemsets. In the mining process, pruning strategy based on *tidlist* is implemented to reduce the search time complexity. Furthermore, based on the projection of k -frequent itemsets, the probability data composed of frequent itemsets are obtained.

These three steps constitute a recursive process, and the whole algorithm process is described as LB(Learning Behavior)-Eclat program.

LB-Eclat Program

Input: T // T is the data set for storing vertical data formats.

Process:

1. LB-Eclat(T)
 2. While all $X_i \in T$
 3. $I_i = \varnothing$
 4. While $X_j \in T$ & $\text{expect} - \text{rup}(X_i) > \text{expect} - \text{rup}(X_j)$
 5. $X_{ij} = X_i \cup X_j$
 6. $\text{tidlist}(X_{ij}) = \text{tidlist}(X_i) \cap \text{tidlist}(X_j)$
 7. Call PFC(X_{ij})//call PFC program
 8. If $P_X[i, j] \geq \text{min_proF}$
 9. Then $T = T \cup \{X_{ij}\}$; $I_i = I_i \cup \{X_{ij}\}$
 10. End While
 11. End While
 12. While $I_i \neq \varnothing$
 13. LB-Eclat(I_i)
 14. End While
 15. Output: all probabilistic frequent itemsets.
-

5. Experiments

The learning behavior components shown in **Tables 1-4** are different in scale and sparsity. Combined with the density of learning behavior components, the specific situation is shown in **Table 5**. In order to realize the comparison and test of the algorithms, the traditional Eclat algorithm and the Eclat algorithm based on descending "Support" (DES Eclat) are selected to carry out the experiments.

L1-P2	sparse density	L1-P4	sparse density	L2-P3	moderate density	L2-P4	dense density
T1-P2	sparse density	T1-P3	sparse density	T1-P4	moderate density		
T2-P1	moderate density	T2-P2	dense density	T2-P3	moderate density	T2-P4	dense density

Table 5.
Density of data sets.

5.1 Performance Indicators

Based on the Eclat framework, the traditional Eclat algorithm, des-Eclat algorithm and LB-Eclat algorithm are written into Python 3.7 and run in the same experimental configuration. In the whole experiment process, we set different min_RST to mine frequent itemsets, and record the indicators generated in the whole processes, which are mainly reflected in the running time of the algorithm, the proportion of memory and the number of probabilistic frequent itemsets.

The test of each indicator is divided into three series according to the sparsity of the data set. The comparative statistical results of corresponding time are shown in **Figures 2–4**. The larger the min_RST, the smaller the time curve distribution of each subgraph. It can be seen from **Figure 2** that the algorithm execution results of sparse density dataset based on the same value show that the traditional Eclat algorithm has advantages. The special sorting of data of DES-Eclat and LB-Eclat increases the time complexity, and the analysis process increases the data time. The execution time of LB-Eclat algorithm is the lowest in **Figures 3 and 4**, which

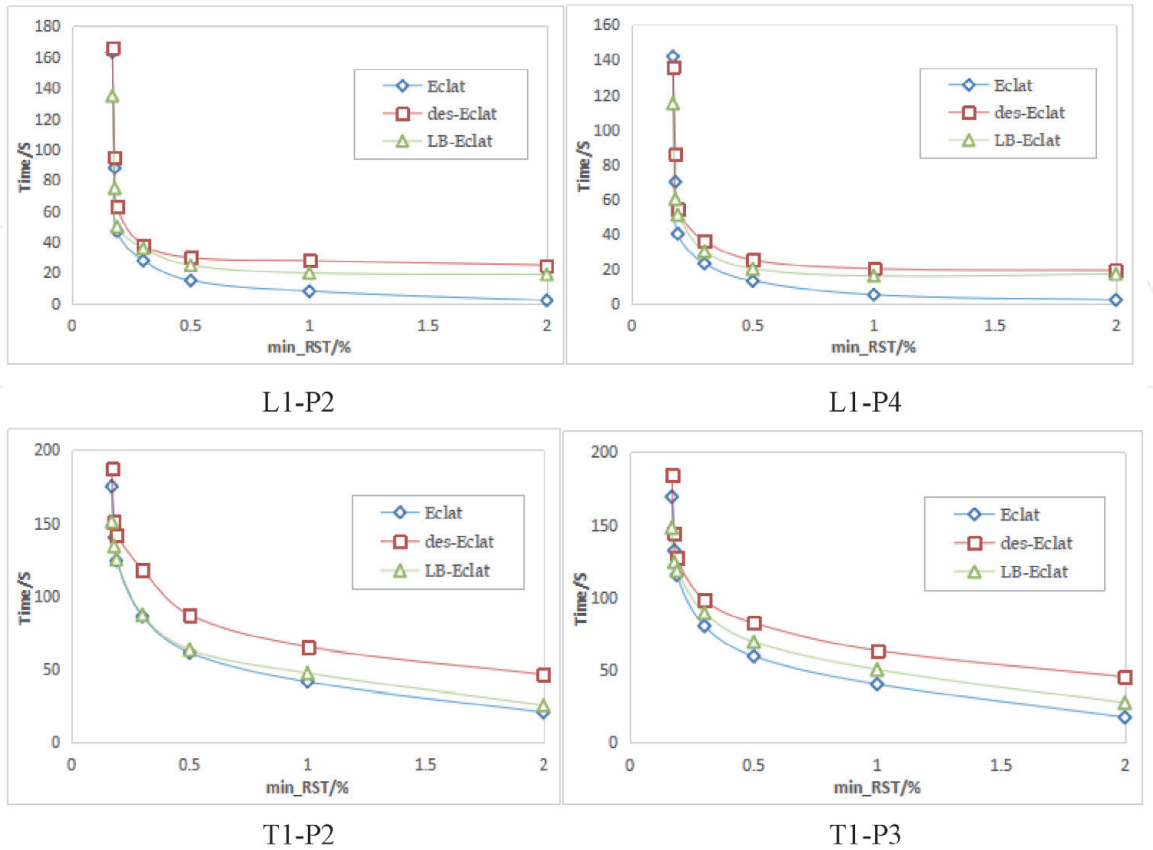


Figure 2.
Comparison of running time of three algorithms on sparse density datasets.

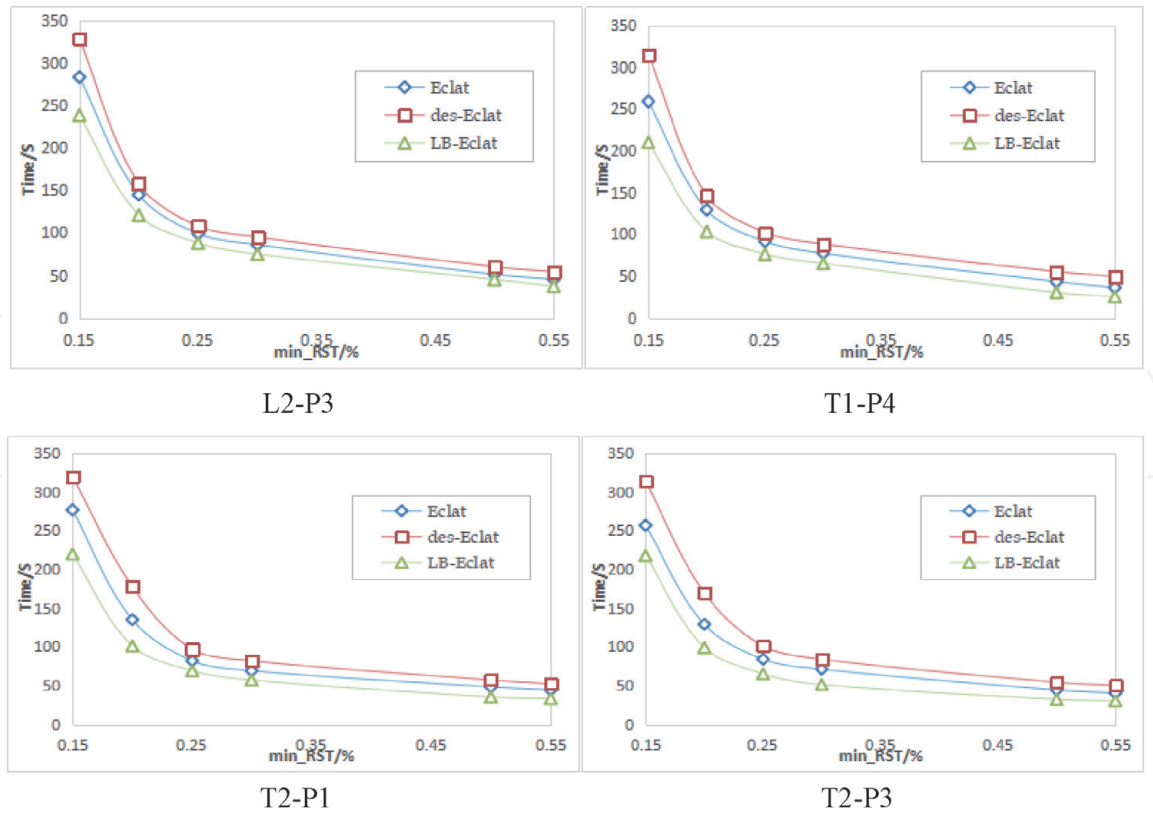


Figure 3.
Comparison of running time of three algorithms on moderate density datasets.

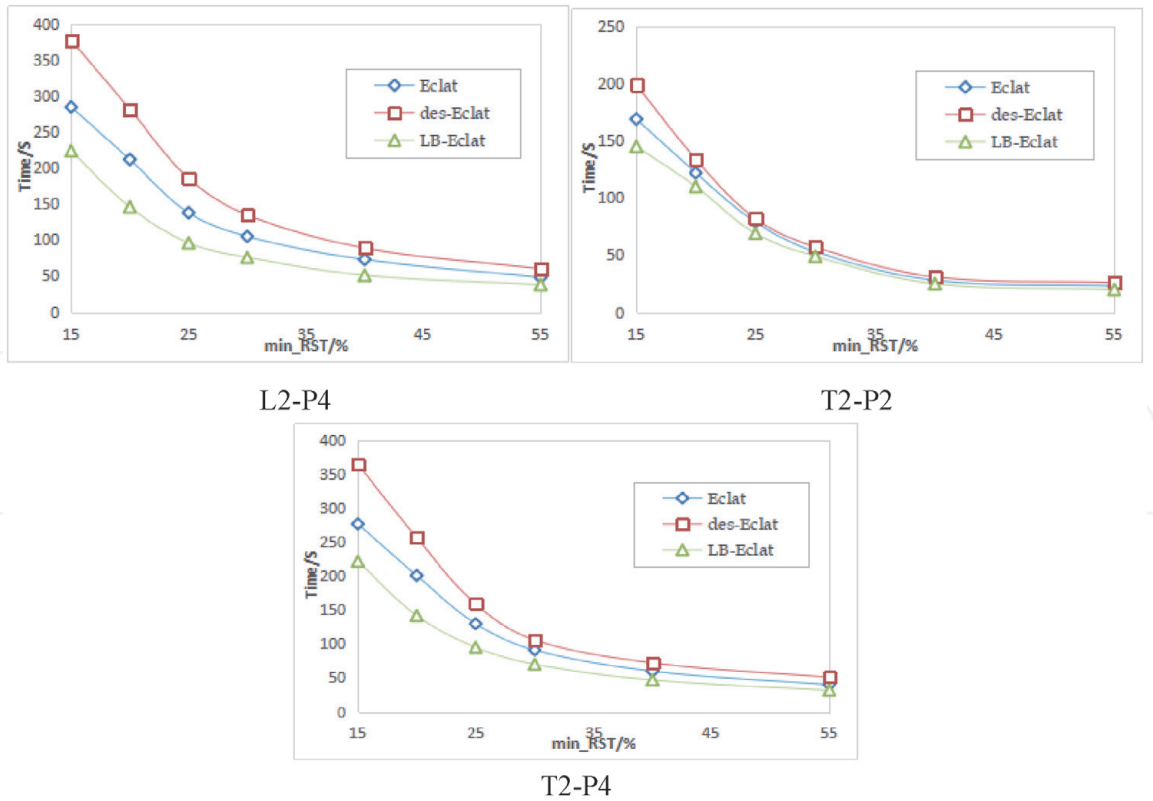


Figure 4.
Comparison of running time of three algorithms on dense density datasets.

indicates that the improvement of the algorithm is more conducive to the analysis of data sets with higher density, and is more effective for mining and processing frequent itemsets of learning behaviors. It can not be found from the time that the DES-Elat algorithm based on the reverse order strategy has a long running time.

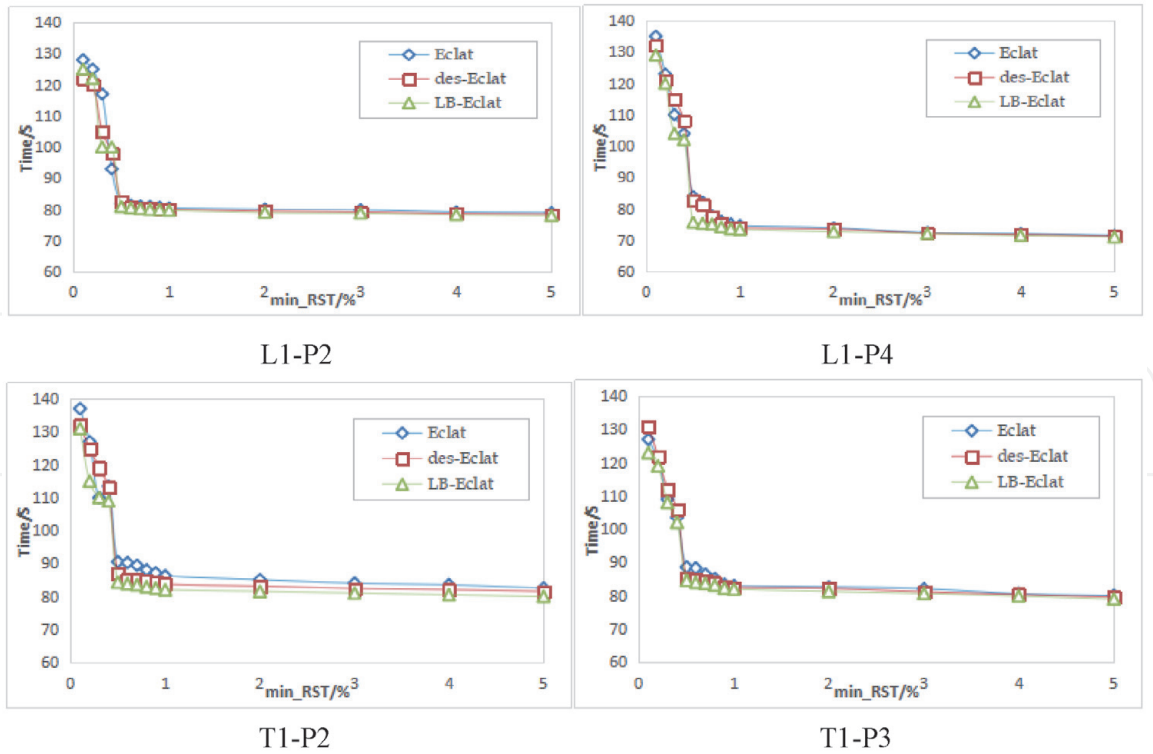


Figure 5.
Comparison of memory space of three algorithms on sparse density datasets.

The comparative results of memory space of the three algorithms are shown in **Figures 5–7**. No matter what the density of data set, the three algorithms occupy the same memory space distribution, the value change trend is the same, LB-Eclat algorithm is slightly smaller than other algorithms, the larger the data set density,

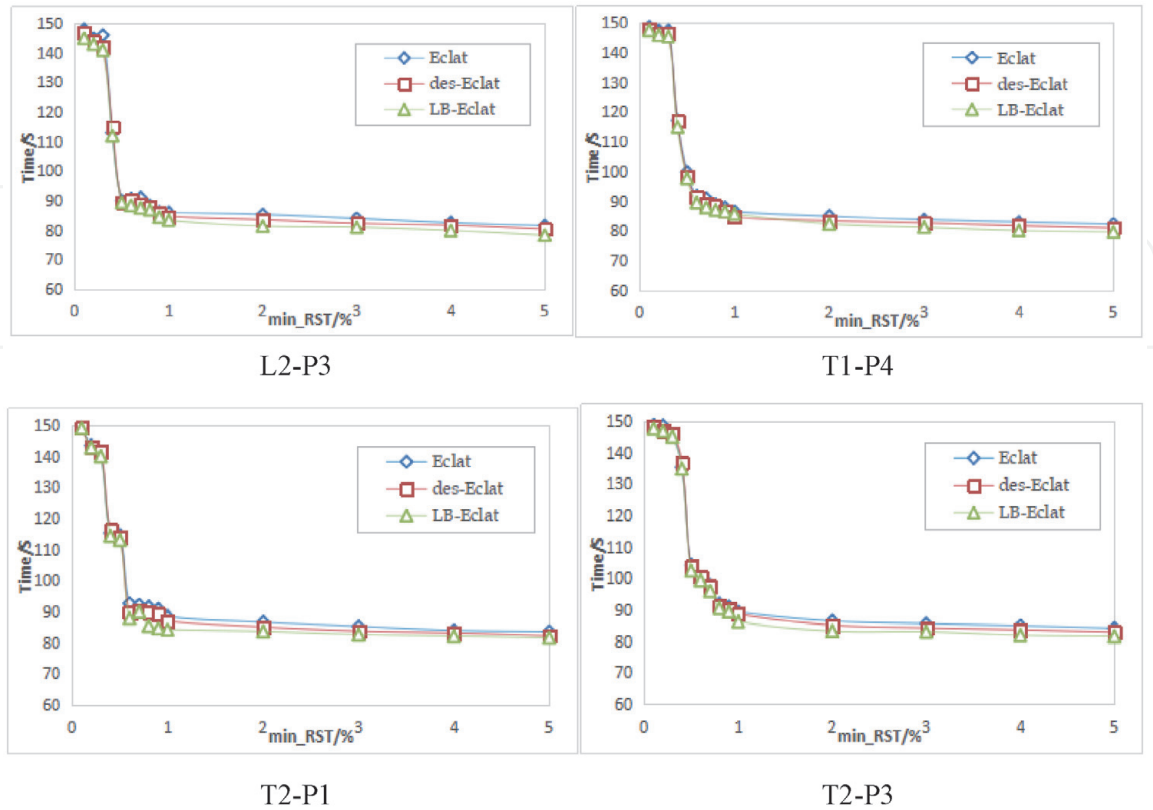


Figure 6.
Comparison of memory space of three algorithms on moderate density datasets.

compared with the traditional Eclat algorithm and des-Eclat algorithm, the smaller the space complexity, that improve the utilization of memory.

The comparison results of probabilistic frequent itemsets mined by the algorithms are shown in **Figures 8–10**. With different `min_RST`, the number of

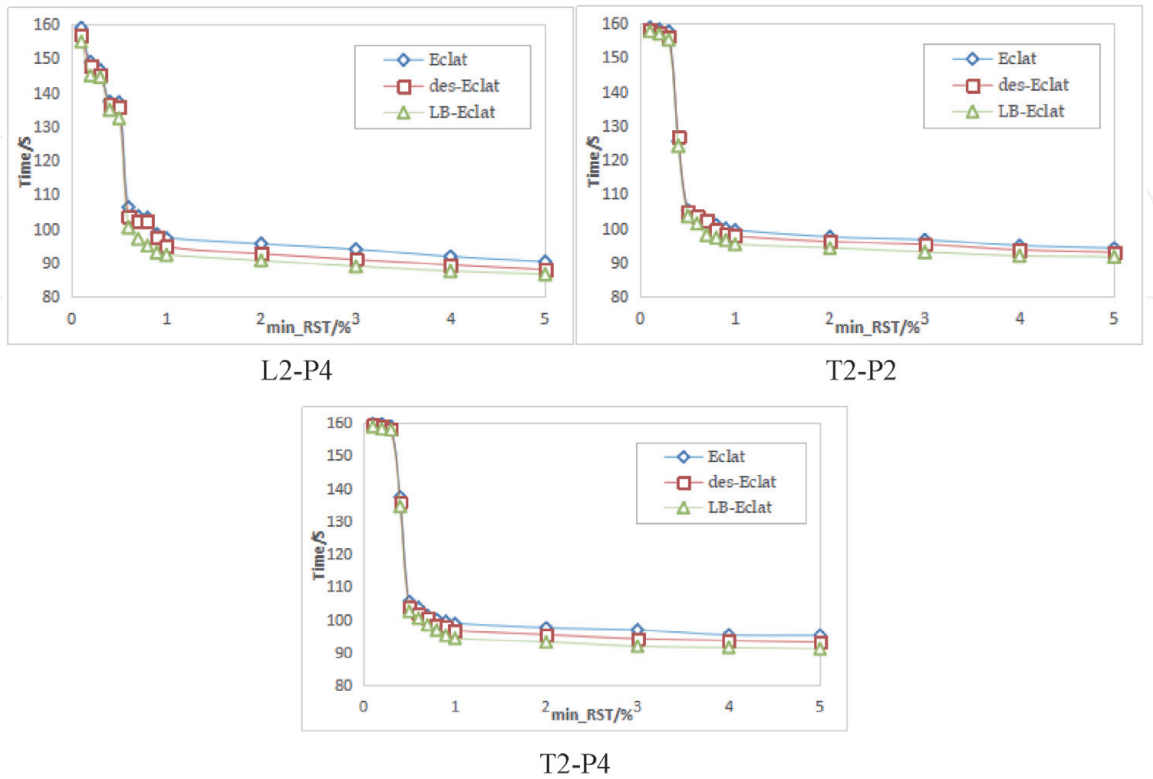


Figure 7.
Comparison of memory space of three algorithms on dense density datasets.

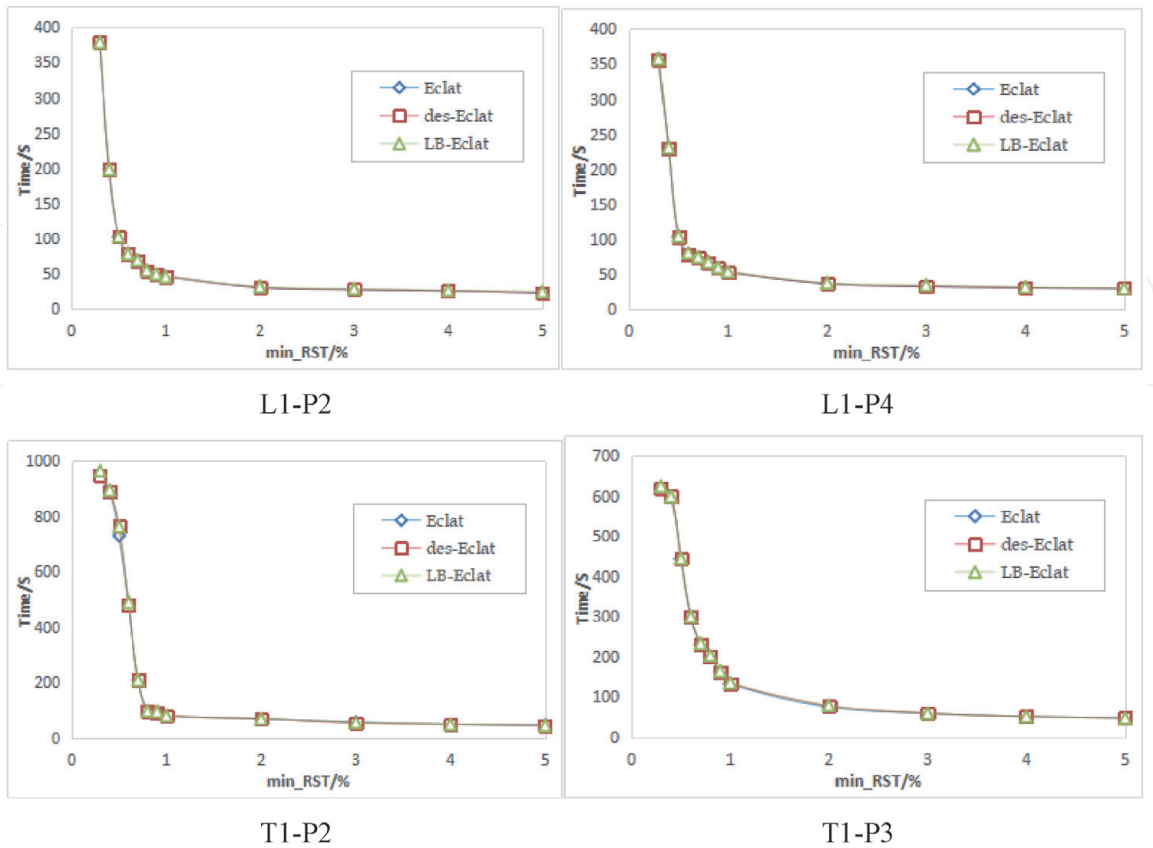


Figure 8.
Comparison of probabilistic frequent Itemsets of three algorithms on sparse density datasets.

probabilistic frequent itemsets depends on the items of learning behaviors and the density of transactions. Although the running time and memory space of the three algorithms are different on the same dataset, the number of probabilistic frequent

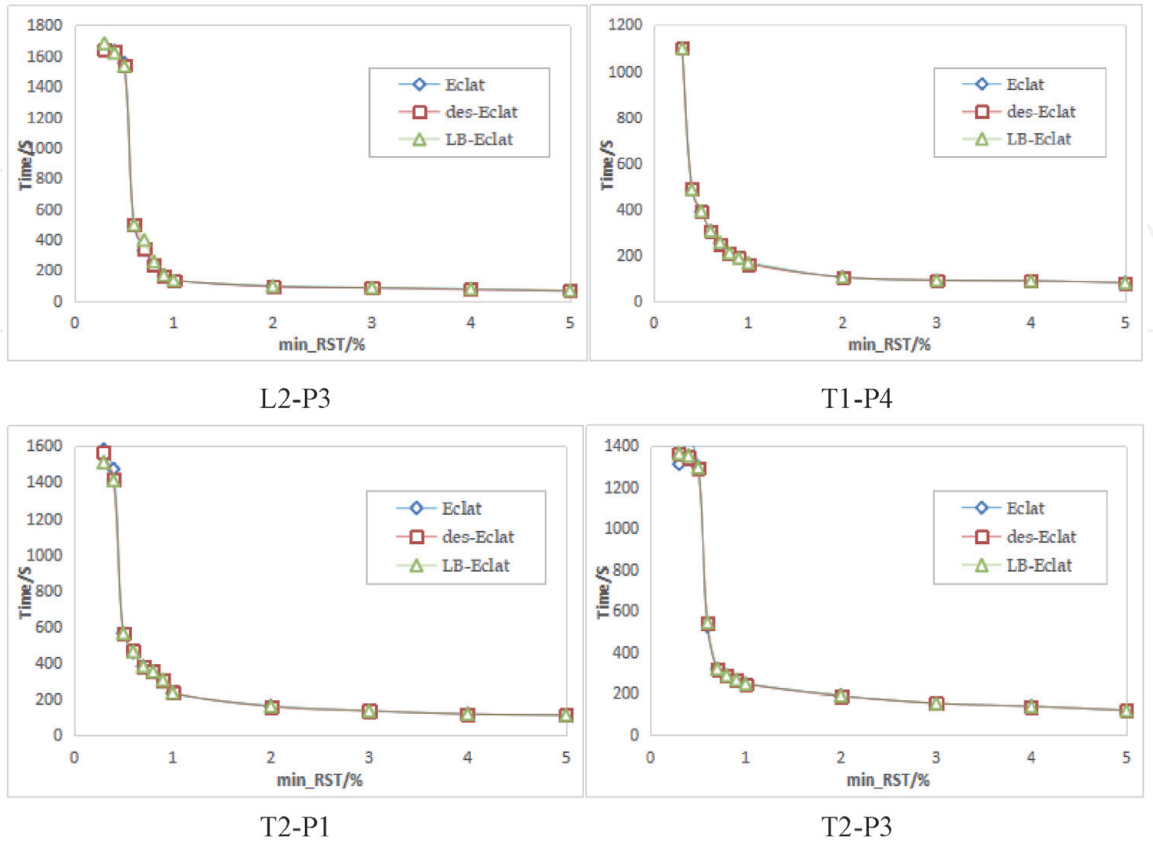


Figure 9.
Comparison of probabilistic frequent Itemsets of three algorithms on moderate density datasets.

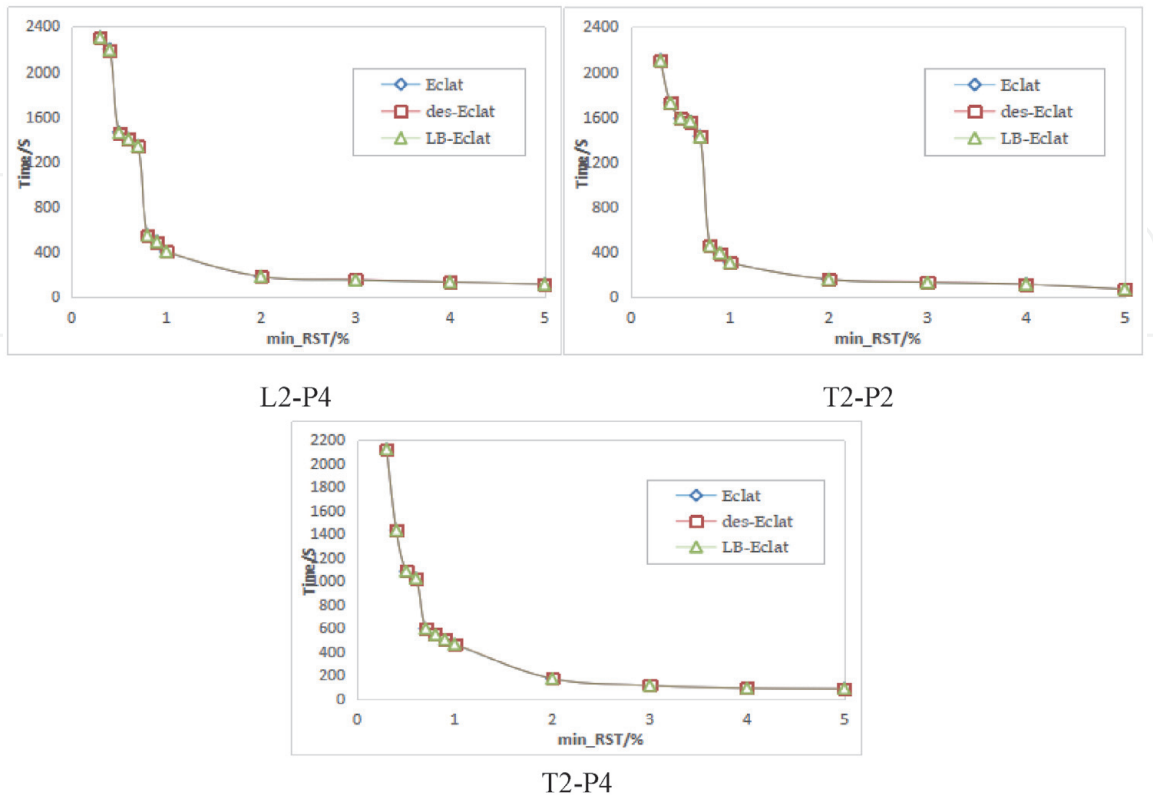


Figure 10.
Comparison of probabilistic frequent Itemsets of three algorithms on dense density datasets.

itemsets obtained is basically the same. With the increase of min_RST , the fewer the number, the smaller the value, The larger the number, the more transactions and items need to be analyzed and calculated, which will inevitably increase the time complexity and space complexity.

The experimental results show that the LB-Eclat algorithm is effective in the study of uncertain learning behavior probabilistic frequent itemsets. About the running time and memory space, LB-Eclat is better than the other two approximate algorithms in mining and analyzing the probabilistic frequent itemsets of sparse density data sets, moderate density data sets and dense density data sets. Since there are 11 learning behavior data sets, the data are all from the real learning processes, and the comparison test process is fully complete. The indicators show that LB-Eclat algorithm are robust and realistic.

6. Probabilistic frequent itemsets analysis of learning behaviors

Based on LB-Eclat algorithm, the probabilistic frequent itemsets of 11 data sets of learning behaviors are mined, and the itemsets with high probability are found. On the basis of “Support” (>0.3) and “Conference” (>0.7), the probability frequent itemsets of each dataset are mined, and then the association degree of rules generated by itemsets is verified by “Lift”. If “Lift” > 1 , the association degree of relevant rules is high. In the mining results of probabilistic frequent itemsets, 2-itemsets are the most, as shown in **Tables 6–8**, the other 3-itemsets and 4-itemsets are mainly based on the intersection and combination of 2-itemsets. The higher the density of data sets, the more frequent itemsets are mined. Based on the constraints of “Support” and “Confidence”, some data sets are limited to 2-itemsets, such as L1-p2 and L1-p4.

From the distribution of frequent 2-itemsets in **Tables 6–8**, they have the following characteristics:

- 1. There is a strong correlation between the components of learning behaviors, and even has a more obvious impact on the components of learning results. In the data set of approximate density, the frequent itemsets of Technology courses are significantly more than that of Literature courses. It shows that the learning behavior components of Technology courses have a strong diversity, and there is a continuous and serial interaction between the components, which makes learners form the approximate frequency participation. Compared with Literature courses, the components of Technology courses are more conducive to the formation of frequent itemsets of learning behaviors.

L1-P2	T1-P2	T1-P3	
forumng, homepage	forumng, homepage	content, url	forumng, homepage
forumng, content	homepage, content	wiki, url	forumng, wiki
content, final_result	homepage, wiki	subpage, url	forumng, url
forumng,final_result	homepage, subpage	homepage, final_result	homepage, content
L1-P4	homepage, url	content, final_result	homepage, wiki
	content, wiki	wiki, final_result	homepage, url
content, subpage	content, quiz	url, final_result	content, wiki
content,final_result			
T1-P3	content, url	homepage, final_result	wiki, final_result
	wiki, url	content, final_result	url, final_result

Table 6.
Probabilistic frequent 2-itemsets of sparse density data sets.

L2-P3	T1-P4	T2-P1	
forumng, homepage	forumng, homepage	dataplus, content	content, url
forumng, subpage	forumng, wiki	dataplus,questionnaire	wiki, subpage
forumng, url	forumng, url	dataplus, url	page, questionnaire
homepage, quiz	homepage, content	dualpane, content	page, subpage
homepage, subpage	homepage, wiki	dualpane,questionnaire	page, url
homepage, url	homepage, url	dualpane, subpage	questionnaire,subpage
quiz, subpage	content, wiki	dualpane, url	questionnaire, url
resource, subpage	wiki, url	forumng, homepage	quiz, subpage
subpage, url	forumng,final_result	homepage, content	resource, subpage
homepage, final_result	homepage,final_result	homepage,	subpage, url
page, final_result	content, final_result	questionnaire	dataplus, final_result
quiz, final_result	wiki, final_result	homepage, subpage	dualpane, final_result
resource, final_result	quiz, final_result	homepage, url	forumng, final_result
subpage, final_result	subpage,final_result	content, page	homepage, final_result
	url, final_result	content, questionnaire	content, final_result
quiz, final_result	subpage, final_result	content, quiz	page,final_result
resource, final_result	url, final_result	content, resource	questionnaire,
		content, subpage	final_result
T2-P3	forumng, homepage	content, subpage	dataplus, final_result
	forumng, subpage	content, url	folder, final_result
dataplus, content	homepage, content	wiki, questionnaire	forumng, final_result
dataplus,	homepage, wiki	wiki, subpage	homepage, final_result
questionnaire	homepage,	wiki, url	content, final_result
dataplus, url	questionnaire	questionnaire, subpage	questionnaire,
dataplus, subpage	homepage, subpage	questionnaire, url	final_result
folder, quiz	homepage, url	quiz, subpage	quiz, final_result
folder, subpage	content, wiki	subpage, url	subpage, final_result
	content, questionnaire		url, final_result

Table 7.
Mining results of probabilistic frequent 2-itemsets of moderate density data sets.

2. For sparse density data sets, “forumng”, “homepage” and “content” are beneficial to form frequent 2-itemsets with other components, which is obviously reflected in different data sets of Literature and Technology courses. “wiki” also has frequent interaction with other components in Technology courses; For moderate density and dense density data sets, frequent 2-itemsets are similar, “forumng”, “homepage”, “content”, “url”, “quiz” and “subpage” all have strong component correlation. For Technology courses, frequent itemsets formed by “dataplus”, “dualpane”, “wiki” and “questionnaire” are used widely and frequently.

For the frequent itemset association rules of learning behavior components, three indicators are used to measure, which are “Support”, “Confidence” and “Lift”. “Support” determines the correlation between the components. “Lift” > 1 indicates that there is association and has positive correlation. The higher “Lift” is, the more valuable the association rules are; if “Lift” < 1 and smaller, there is negative correlation; if Lift = 1, the components are independent and have no correlation. The association rules with “Lift” > 1 and high confidence are listed and shown in **Table 9**, these association rules are the basis for tracking, adjusting and optimizing learning behaviors.

On the whole, the association rules corresponding to the probabilistic frequent itemsets of sparse density data sets are less, and the association rules of Literature courses are less in the same density data sets [22]. For the moderate density and dense density data sets of Technology courses, rules are formed among the

L2-P4	T2-P2		
forumng, homepage	dataplus, questionnaire	content, wiki	subpage, url
homepage, subpage	dataplus, dualpane	content, page	dataplus, final_result
quiz, subpage	dataplus, content	content, questionnaire	dualpane, final_result
forumng, final_result	dataplus, page	content, quiz	folder, final_result
homepage, final_result	dataplus, url	content, subpage	forumng, final_result
page, final_result	dualpane, content	content, url	content, final_result
quiz, final_result	dualpane, page	wiki, questionnaire	homepage, final_result
subpage, final_result	dualpane, questionnaire	wiki, subpage	content, final_result
	dualpane, subpage	wiki, url	wiki, final_result
	dualpane, url	page, questionnaire	page, final_result
	folder, subpage	page, subpage	questionnaire, final_result
	forumng, homepage	page, url	quiz, final_result
	homepage, content	questionnaire, subpage	subpage, final_result
	homepage, wiki	questionnaire,	url, final_result
	homepage, subpage	urlquiz, subpage	
	homepage, url		
T2-P4			
dataplus, dualpane	dualpane, questionnaire	content, subpage	page, subpage
dataplus, content	homepage, content	content, url	page, url
dataplus, wiki	homepage, subpage	wiki, page	questionnaire, subpage
dataplus, page	homepage, url	wiki, questionnaire	questionnaire, url
dataplus, questionnaire	content, wiki	wiki, subpage	quiz, subpage
dataplus, subpage	content, page	wiki, url	resource, subpage
dataplus, url	content, questionnaire	page, questionnaire	subpage, url
dualpane, page			

Table 8.
Probabilistic frequent 2-itemsets of dense density data sets.

	Support	Conference	Lift	Rules
L1-P2	0.2301	0.8127	1.2832	{homepage, content} → {forumng}
L1-P4			None	
T1-P2	0.2152	0.7836	1.5241	{homepage} → {forumng}
T1-P3	0.2701	0.7435	1.7195	{content, wiki, subpage, url} → {homepage}
	0.2558	0.8281	1.6399	{content, wiki, subpage} → {url}
L2-P3	0.3291	0.8536	1.8408	{homepage, subpage, url} → {forumng}
	0.2453	0.7166	1.6807	{quiz, subpage, url} → {homepage}
	0.2132	0.5369	1.3577	{subpage, quiz} → {final_result}
T1-P4	0.1731	0.8467	1.7063	{homepage, wiki, url} → {forumng}
	0.2229	0.8757	1.7530	{content, wiki, url} → {homepage}
	0.3681	0.5355	1.2122	{content, wiki} → {final_result}
T2-P1	0.3522	0.7739	1.4773	{content, questionnaire, url} → {dataplus}
	0.4119	0.7049	1.6980	{content, questionnaire, subpage, url} → {dualpane}
	0.3859	0.7978	1.5795	{homepage} → {forumng}
	0.4619	0.7953	2.0856	{content, questionnaire, subpage, url} → {homepage}
	0.4207	0.8682	1.7532	{page, questionnaire, quiz, resource, subpage, url} → {content}
	0.3361	0.7452	1.6985	{questionnaire, subpage, url} → {page}
	0.4747	0.7210	1.4747	{subpage, url} → {questionnaire}

	Support	Conference	Lift	Rules
T2-P3	0.5151	0.8386	2.0553	{resource, url} → {subpage}
	0.3361	0.5548	1.2858	{content, subpage, quiz} → {final_result}
	0.2023	0.7734	1.6246	{content, questionnaire, url, subpage} → {dataplus}
	0.1285	0.8243	1.6447	{homepage, subpage} → {forumng}
	0.2704	0.8609	1.7790	{content, wiki, questionnaire, subpage, url} → {homepage}
	0.3022	0.8631	1.7404	{wiki, questionnaire, subpage, url} → {content}
	0.3253	0.8098	1.5003	{url} → {subpage}
L2-P4	0.2521	0.5250	1.5236	{folder, content, quiz, subpage} → {final_result}
	0.1781	0.8473	1.4871	{homepage} → {forumng}
T2-P2	0.3934	0.5563	1.0403	{subpage} → {final_result}
	0.1989	0.7732	1.6777	{questionnaire, dualpane, content, page, url} → {dataplus}
	0.2486	0.7531	1.6984	{content, page, questionnaire, subpage, url} → {dualpane}
	0.2019	0.8072	1.5534	{homepage} → {forumng}
	0.3947	0.7515	1.7227	{content, wiki, subpage, url} → {homepage}
	0.3025	0.8761	1.7971	{wiki, page, questionnaire, quiz, subpage, url} → {content}
	0.3342	0.7527	1.6687	{questionnaire, subpage, url} → {page}
	0.2624	0.7085	1.2884	{subpage, url} → {quiz}
	0.3128	0.8552	1.6777	{url} → {subpage}
T2-P4	0.3760	0.5452	1.4449	{folder, content, quiz, subpage} → {final_result}
	0.2763	0.8002	1.7168	{dualpane, content, wiki, page, questionnaire, subpage, url} → {dataplus}
	0.2262	0.7934	1.7061	{content, subpage, url} → {homepage}
	0.2756	0.8858	1.7399	{wiki, page, questionnaire, subpage, url} → {content}
	0.3966	0.7606	1.7510	{questionnaire, subpage, url} → {page}
	0.3149	0.8222	1.9056	{url} → {subpage}

Table 9.
Association rules generated by probabilistic frequent Itemsets.

components of learning behaviors, and some of the components can produce rules with high credibility and strong relevance with the final assessment results.

It can be seen from **Table 9** that there are common association rules of components among different data sets, which indicates that these rules have strong generality; for Literature courses or Technology courses, there are some similarities in association rules, but there are also obvious differences; For the same course, in different periods, the results show that the association rules of probabilistic frequent itemsets have both intersection and differences. About {content, questionnaire, subpage, url} → {homepage}, {resource, url} → {subpage} and {resource, url} → {subpage}, the “Lift” values are higher, indicating that the association degree is very high. From the table, it is easy to form strong association rules around “questionnaire”, “quiz”, “forumng”, “homepage”, “resource”, “subpage”, “url” and so on. “dataplus”, “dualpane”, “folder”, “wiki” and so on have strong relevance in Technology courses. Some of components have an obvious impact on the learning results. The extraction of these association rules can greatly simplify the categories of components in **Tables 1–4**.

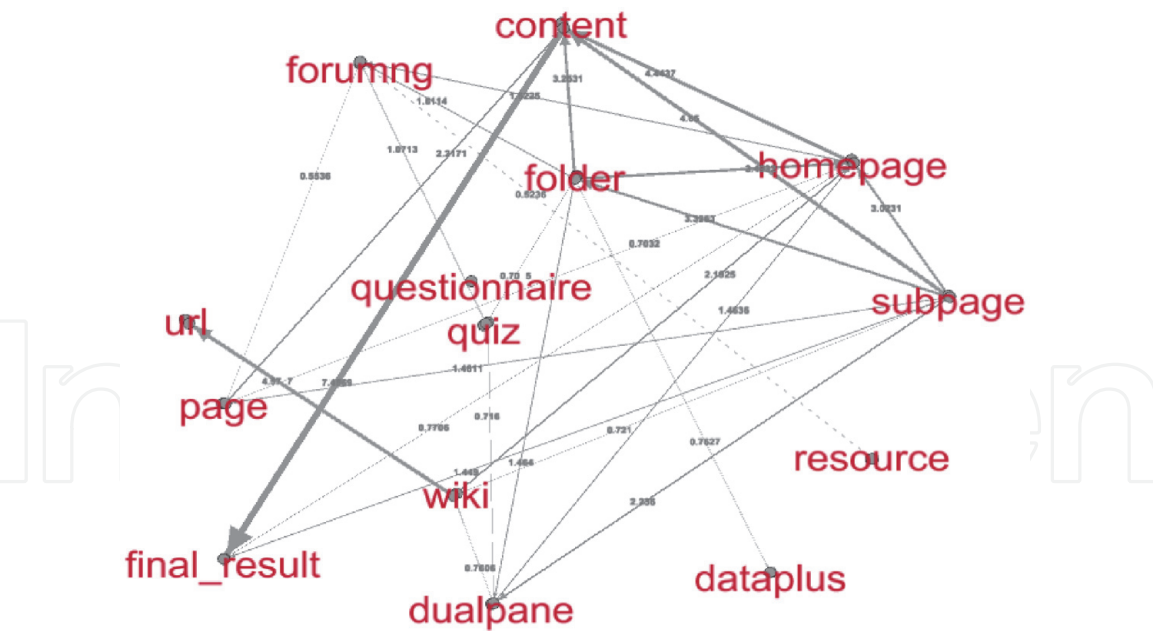


Figure 11.
The key topology of learning behaviors based on probabilistic frequent Itemsets and association rules.

The mining of probabilistic frequent itemsets and the learning of association rules are conducive to the evaluation and recommendation of components in the construction of learning behaviors [22–24]. At the same time, the formation process of learning behaviors can realize the aggregation of effective components according to these association rules. For the components related to association rules, we can build elastic proximity relationships or timely guidance strategies and recommendation mechanisms. This can effectively guide the learning processes, on the other hand, according to the needs of learning objectives, we can design association rules of probabilistic frequent itemsets according to the historical data, which is conducive to analyze and predict feasible participation components.

Based on the data in **Tables 6–9**, the nodes and edges of component interaction processes are constructed, and the key constituent units of learning behavior data sets are generated by Gephi. **Figure 11** shows the topological structure and relationship weight of probabilistic frequent itemsets. There are 14 participation components involved and the weight of each relationship (edge) is calculated automatically. The thickness of the line indicates the strength of the relationship, and the dotted lines represent the potential relationships. The construction and extraction of the key topology of learning behaviors supported by probabilistic frequent itemsets are completed, which is a referential result of data-driven learning behavior prediction and decision making.

7. Decision-making scheme for improving learning behaviors

Studying learning behavior through big data can promote learners to improve their learning processes and learning effects [25]. Aiming at the mining and association analysis of probability frequent itemsets, we realize 11 data subsets of learning behaviors with components as the basic structure characteristics. On the basis of Eclat framework, the vertical data format is adopted to design and improve the data structure and analysis algorithm for learning behavior components. Through the indicator comparison of approximate algorithms, the improved algorithm is effective and feasible for the analysis processes of data subsets, especially in the application of moderate density and dense density data set. Based on the data analysis

results, “Support”, “Confidence” and “Lift” are the measurement indicators, and the corresponding thresholds are set. The probabilistic frequent itemsets and association rules are mined, and the key topology of learning behaviors supported by the probabilistic frequent itemsets are constructed. The whole processes of mining and analyzing probabilistic frequent itemsets are based on the vertical data format, which ensures the depth and breadth of data research results for decision prediction.

The research of learning behaviors is a specific branch of big data. It is different from other types of data characteristics. Because of the periodicity, continuity, collectivity and individuality of learning behaviors, there may be greater instability and discreteness between the generated data and the expected data. It is very difficult in data analysis and decision making, so it is necessary to design appropriate data structures and algorithms [26, 27] to carry out multi-dimensional empirical study on learning behaviors. Through a series of work and research results of probabilistic frequent itemsets analysis, the following decision schemes are obtained.

7.1 Learning content will affect the frequent itemsets of learning behaviors

Learning content determines learners’ tendency. The data of learning behaviors focuses on two Literature courses and two Technology courses, which correspond to multiple learning periods respectively. On the whole, the learning process of Technology courses more complicated, the learning behavior components are more diverse, and the online learning process description is also quite complete and comprehensive, that forms larger scale datasets. Learning content will affect the data density, components and the actual learning processes of learners, which determines the frequent itemsets mining results. For example, from the probabilistic frequent itemsets of the two learning periods of L1 course, the online learning processes corresponding to the learning contents do not have advantages, there is no effective correlation between the components and the learning assessment results, and the advantages of online learning mode are not obvious, which may be more suitable for the teaching mode.

Therefore, the construction of learning behaviors depends on the learning content. According to the mining results of frequent itemsets of historical data and the analysis of association rules, the learning mode of the course is optimized in the new learning period. Based on the learning content, we guide or expand the components of learning behaviors, so as to enhance the learning interest.

7.2 Teaching goals will affect the frequent itemsets of learning behaviors

The same learning content in different learning period, can produce different learning behavior data density, so as to get different frequent itemsets. In different learning periods, the frequent itemsets and association rules obtained by the algorithm are similar, but there are also obvious differences. The components are not the same, and some data sets are quite different. Learners in different periods have different teaching needs, and then correspond to different teaching objectives; On the other hand, the participation and traction in the learning process make the different participation components, and the stickiness of different components are different, which determines the frequent items, and thus produces different association rules, it even affects learners’ assessment methods and learning results.

Therefore, the construction of learning behaviors should consider the learning periods and the actual learners, flexibly construct teaching objectives, and design adaptive learning behavior components. In the learning processes, we should also

timely analyze the learning behaviors, mine the existing problems and learners' preferences, adjust the components in time, and optimize the learning methods appropriately. We should build a real-time and effective data tracking and analysis mechanisms.

7.3 The frequent itemsets of learning behavior have the characteristics of explicit and implicit association

There are differences in interaction mode of learning behavior components in different platforms, but the demands of serving learning behaviors are the same, that is to realize the continuity of learning behaviors and achieve the learning effects through the interaction of components. Through the mining of probabilistic frequent itemsets and the analysis of association rules, the components of frequent itemsets have explicit association features, and different frequent itemsets may also have implicit association features. It has a strong recommendation value for the prediction and feedback of latent learning behaviors. The key topological relationships of learning behaviors are shown in **Figure 11**, that can provide references for the follow-up learning processes of similar or the same courses, and expand learning methods.

Therefore, the construction of learning behaviors should not only consider the learning content and teaching objectives, but also refer to the historical effective learning behaviors, and also need to carry out effective learning process reform and learning strategy change based on data analysis, gradually promote learners to develop effective learning habits and methods, and construct new learning behavior components. According to the learning situation, stage learning feedback, potential behavior recommendation and implicit interest mining are achieved in order to improve the learning quality.

7.4 Learning behavior needs the adaptive support of specific algorithm and data structure

The generation of learning behaviors is a multi-dimensional process. The research strength of these data determines the cognitive strength of learning behaviors. There are different perspectives on the composition of learning behaviors, which determines different research methods. How to carry out relatively sufficient modeling description and business processing of learning behaviors presents challenges to learning analytics. Some existing software tools and analysis methods can not guarantee the appropriate quantification, standardization and initialization, the analysis process and experimental conclusion may not be thorough and comprehensive. Compared with the statistics and test of learning behaviors carried out by sampling, the effective and comprehensive analysis of learning behaviors is more convincing.

Therefore, the empirical analysis of learning behaviors should be the comprehensive application process of data-driven technologies and methods. Combined with the data characteristics, the technical requirements are demonstrated, and the structures and algorithms suitable for data attributes and process characteristics are designed. This aspect has huge research space and prospect in the field of education big data, which poses challenges for researchers. Learning analytics of educational big data is essentially data analysis, and it is a comprehensive application of computer science and technology, statistics, engineering, etc., and the design and development of general tools in this respect still need time [28]. For a specific data set, it is feasible and more realistic to design suitable data structures and algorithms for decision making.

8. Conclusion

The learning analysis of learning behaviors is a complex process. The data structure, attribute characteristics and relationship categories bring more difficulties. Moreover, the data has strong uncertainty and instability, so it is difficult to achieve technical unity and generality [29]. The development of online learning model gives new definitions and norms to learning behaviors, and also requires new data structure, attribute characteristics, relationship categories, etc. many technologies and methods that can be used in the research of learning behaviors may be inefficient for new data, or do nothing for the new research branches. This research is about the design and application of intelligent data mining technology on a big data set of learning behaviors. Based on Eclat framework, the data structure and algorithms are improved. Starting from the vertical data format, mining probabilistic frequent itemsets, analyzing association rules, and realizing data-driven decision making. In the subsequent research of learning behaviors, for uncertain data, we continue to conduct in-depth research and demonstration of methods and technologies, improve the quality of data analysis and relationship perspective, and provide more valuable conclusions for decision making and prediction feedback of learning behaviors.

Compliance with ethical standards

The authors certify that there is no conflict of interest with any individual/organization for the present work.

A list of acronyms

FP	Frequent Pattern
PFC	Frequent Pattern Calculation
LB-Eclat	Learning Behavior Eclat
Descending Eclat	Descending “Support”

IntechOpen

IntechOpen

Author details

Xiaona Xia^{1,2,3}

1 Faculty of Education, Qufu Normal University, Qufu, Shandong, China

2 Chinese Academy of Education Big Data, Qufu Normal University, Qufu, Shandong, China

3 School of Computer Science, Qufu Normal University, Rizhao, Shandong, China

*Address all correspondence to: xiagn@sina.com

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Kusemererwa, C., Munene, J. C., Laura, O. A., & Balunywa, J. W. (2020). Individual learning behavior: do all its dimensions matter for self-employment practice among youths in uganda?. *Journal of Enterprising Communities: People and Places in the Global Economy*, Vol. 14 No. 3, pp. 373-396. DOI: 10.1108/JEC-02-2020-0012
- [2] Wang, X., Guchait, P., & Paamehmetolu, A. (2020). Tolerating errors in hospitality organizations: relationships with learning behavior, error reporting and service recovery performance. *International Journal of Contemporary Hospitality Management*, Vol. 32 No. 8, 2635-2655 DOI: 10.1108/IJCHM-01-2020-0001
- [3] B, Z. Y. A., C, G. Z. A. B., C, D. L. A. B., & B, H. L. A. (2020). Learning crowd behavior from real data: a residual network method for crowd simulation. *Neurocomputing*, 404(3), 173-185, DOI: 10.1016/j.neucom.2020.04.141.
- [4] Schmerse, D. (2020). Preschool quality effects on learning behavior and later achievement in germany: moderation by socioeconomic status. *Child Development*, Volume 00, Number 0, 1-18, DOI: 10.1111/cdev.13357.
- [5] Lai, S., Sun, B., Wu, F., & Xiao, R. (2020). Automatic personality identification using students' online learning behavior. *IEEE Transactions on Learning Technologies*, 13(1), 26-37, DOI:10.1109/TLT.2019.2924223
- [6] Ines Šarić-Grgić, Ani Grubišić, Ljiljana Šerić, & Robinson, T. J. (2020). Student clustering based on learning behavior data in the intelligent tutoring system. *International Journal of Distance Education Technologies*, 18 (2), 73-89, DOI:10.4018/IJDET.2020040105.
- [7] Yokoyama, M., & Miwa, K. (2020). STUDENTS' CONCEPTION OF LEARNING AND LEARNING BEHAVIOR FROM MULTIPLE-GOALS PERSPECTIVE. *7th International Conference on Educational Technologies 2020*, 33-40, DOI: 10.33965/icedutech2020_202002L005.
- [8] Silva, J., Varela, N., Luz Adriana Borrero López, & Rafael Humberto Rojas Millán. (2019). Association rules extraction for customer segmentation in the smes sector using the apriori algorithm. *Procedia Computer ence*, 151, 1207-1212, DOI: 10.1016/j.procs.2019.04.173.
- [9] Hossain, T. M., Watada, J., Jian, Z., Sakai, H., & Aziz, I. A. (2020). Missing well log data handling in complex lithology prediction: an nis apriori algorithm approach. *International journal of innovative computing, information & control: IJICIC*, 16(3), 1077-1091, DOI: 10.24507/ijicic.16.03.1077
- [10] Hong, J., Tamakloe, R., & Park, D. (2020). Discovering insightful rules among truck crash characteristics using apriori algorithm. *Journal of advanced transportation*, 2020(2), 1-16, DOI: 10.1088/1742-6596/1477/2/022032
- [11] Bashkari, S., Sami, A., & Rastegar, M. (2020). Outage cause detection in power distribution systems based on data mining. *IEEE Transactions on Industrial Informatics*, PP(99), 1-1, DOI: 10.1109/TII.2020.2966505
- [12] Abdullah, S. S., Sedig, K., & Rostamzadeh, N. (2020). Multiple regression analysis and frequent itemset mining of electronic medical records: a visual analytics approach using visa_m3r3. *Data*, 5(2), 1-26, DOI: 10.3390/data5020033.
- [13] Chaghari, A., Mohammad-Reza Feizi-Derakhshi, & Mohammad-Ali

- Balafar. (2020). The combination of term relations analysis and weighted frequent itemset model for multidocument summarization. *Computational Intelligence*, 36(2), 783-812, DOI: 10.1111/coin.12270
- [14] Raj, S., Ramesh, D., Sreenu, M., & Sethi, K. K. (2020). Eafim: efficient apriori-based frequent itemset mining algorithm on spark for big transactional data. *Knowledge and Information Systems*, 62(4), 3565–3583, DOI: 10.1007/s10115-020-01464-1
- [15] Rahman, A., Mutiarawan, R. A., Darmawan, A., Rianto, Y., & Syafrullah, M. (2020). Prediction Of Students Academic Success Using Case Based Reasoning. *2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*. IEEE, 171-176, DOI:10.23919/EECSI48112.2019.8977104
- [16] Singh, P., Singh, S., Mishra, P. K., & Garg, R. (2020). RDD-Eclat: Approaches to Parallelize Eclat Algorithm on Spark RDD Framework. *Second International Conference on Computer Networks and Communication Technologies*, 755-768, DOI: 10.1007/978-3-030-37051-0_85
- [17] Zhang, C., Tian, P., Zhang, X., Jiang, Z. L., Yao, L., & Wang, X. (2019). Fast eclat algorithms based on minwise hashing for large scale transactions. *IEEE Internet of Things Journal*, 6(2), 3948-3961, DOI: 10.1109/JIOT.2018.2885851
- [18] Robu, V., & Santos, V. D. D. (2019). Mining Frequent Patterns in Data Using Apriori and Eclat: A Comparison of the Algorithm Performance and Association Rule Generation. *2019 6th International Conference on Systems and Informatics (ICSAI)*, 1478-1481, DOI: 10.1109/ICSAI48974.2019.9010367
- [19] Man, M., Jusoh, J. A., Saany, S. I. A., Bakar, W. A. W. A., & Ibrahim, M. H. (2019). Analysis study on r-eclat algorithm in infrequent itemsets mining. *International Journal of Electrical and Computer Engineering*, 9 (6), 5446-5453, DOI: 10.11591/ijece.v9i6.pp5446-5453
- [20] Atis KAPENIEKS, Iveta DAUGULE, Kristaps KAPENIEKS, Viktors ZAGORSKIS, Janis KAPENIEKS Jr, Zanis TIMSANS, & Ieva VITOLINA. (2020) TELECI Approach for e-Learning User Behavior Data Visualization and Learning Support Algorithm. *Baltic Journal of Modern Computing*, Vol. 8, No. 1, 129-142, DOI: 10.22364/bjmc.2020.8.1.06.
- [21] Lin Tan, Yali Chen, Runhan Yang, Li Lai. (2020) Empirical Research on the Effect of Collaborative Learning in Blended Learning Mode Based on KNN Algorithm. *ICIET 2020: Proceedings of the 2020 8th International Conference on Information and Education Technology*, March 2020, 48-52, DOI: 10.1145/3395245.3395251
- [22] Chebi, H., Tabet-Derraz, H., Sayah, R., Meroufel, A., & Meraihi, Y. (2020). Intelligence and adaptive global algorithm detection of crowd behavior. *International Journal of Computer Vision and Image Processing*, Volume 10, Issue 1, 2020, 24-40, DOI: 10.4018/IJCVIP.2020010102
- [23] Xia, X. (2021). Interaction recognition and intervention based on context feature fusion of learning behaviors in interactive learning environments. *Interactive Learning Environments*, Advance online publication 17 Jan 2021. 1-19. DOI: 10.1080/10494820.2021.1871632
- [24] Yuniarti, T., Widhianningrum, P., & Sulistyowati, N. W. (2020). A study of accounting learning achievements using emotional intelligence and learning behavior, *ASSETS Jurnal Akuntansi dan Pendidikan*, 9(1):52-60, DOI: 10.25273/jap.v9i1.4179

[25] Xia, X. (2020). Random field design and collaborative inference strategies for learning interaction activities. *Interactive Learning Environments*, Advance online publication 30 Dec 2020. 1–25. DOI: 10.1080/10494820.2020.1863236

[26] Smiderle, R., Sandro José Rigo, Marques, L. B., Coelho, J. A. P. D. M., & Jaques, P. A. (2020). The impact of gamification on students' learning, engagement and behavior based on their personality traits. *Smart Learning Environments*, 7(1), 1-11, DOI: 10.1186/s40561-019-0098-x.

[27] Xia, X. (2020). Learning behavior mining and decision recommendation based on association rules in interactive learning environment. *Interactive Learning Environments*. Advance online publication 4 Aug 2020. 1–16. DOI: 10.1080/10494820.2020.1799028

[28] John, C., & Meinel, C. (2020). Learning Behavior of Men and Women in MOOC Discussion Forums – A Case Study. 2020 IEEE Global Engineering Education Conference (EDUCON). IEEE, 300-307, DOI:10.1109/EDUCON45650.2020.9125322

[29] Jiang, L., & Dong, K. (2020). Artificial intelligence-based learning behavior data mining and network teaching quality monitoring mechanism. *Journal of Physics Conference Series*, 1533, 032058, DOI: 10.1088/1742-6596/1533/3/032058