PAPER • OPEN ACCESS

Summary of Association Rules

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Summary of Association Rules

Foxiao Zhan[1, a], Xiaolan Zhu[1, b], Lei Zhang[1, 2, *], Xuexi Wang[1, c], Lu Wang[1, d],

Chaoyi Liu[1, e]

[1]Department of Computer Technology and application Qinghai University Xining, China
[2]College of Computer Science, Information Management Center Sichuan University China

*Corresponding author e-mail: zhanglei@scu.edu.cn, [a]1376806045@qq.com,
[b] zxlanscu@126.com, [c]774793339@qq.com, [d]wlqgl@126.com,
[e]3305306186@qq.com

Abstract. In recent years, scholars at home and abroad have conducted a large number of researches on association rules, in order to deeply understand the mining technology of association rules, and master its research status and current situation, The development trend, first of all, related definitions and classification methods of association rules are introduced. Secondly, the general methods of association rule mining are summarized from serial and parallel perspectives, and some typical association rule mining modes are summarized and analysed. Finally, the quality improvement of association rule mining and its field application are discussed.

1. Introduction

With the rapid development of society and the continuous progress of science and technology, the speed of information circulation is getting faster and faster. A lot of data has been produced in all works of life. How to excavate potential and valuable information from a large amount of data has become a concern of all industries. Association rule mining is one of the important research methods in the field of data mining, which is widely used in finance, medicine, Internet and other fields. The original association rules mining is put forward for basket analysis problem, its aim is to find the relationship between different commodity transaction database, to obtain the general rules of customer buying patterns. These rules can guide merchants to reasonably arrange the purchase, inventory and shelf design. Agrawal et al. [1] proposed the earliest Apriori algorithm based on frequent itemsets. After that, researchers at home and abroad have deeply studied the mining of association rules. Related work includes optimization of Apriori algorithm, parallel association rule mining, quantitative association rule mining and association rule mining theory.

2. Basic concepts of association rules

2.1. Definition of Association Rules

First introduce the traditional association rules algorithm, association rule is the containing type shape such as X and Y, among them, the X and Y, respectively, known as the forerunner of association rules

(antecedent or left - hand side, LHS) and subsequent (consequent or right - hand - side, RHS). Where, the association rule XY has support and trust.

Definition: suppose I {I1, I2, I3...IM} is the set of terms. Given a Transaction database D, where each Transaction (Transaction)t is a non-empty subset of I, that is, each Transaction corresponds to a unique identifier TID(Transaction ID).The support of association rules in D is the percentage of transactions in D containing both X and Y, that is, the probability. Confidence is the percentage of Y in the case that the transaction in D already contains X, that is, the conditional probability. If the minimum support threshold and minimum confidence threshold are met, the association rule is considered interesting. These thresholds are set manually for mining purposes.

2.2. Mining Process

Association rule mining process consists of two phases: the first stage must first find out all of the high frequency from data collection team (the Frequent Itemsets, i.e., calculate meet support team (sets), the second stage again by the high-frequency team in Association Rules (Association Rules), namely: calculate again meet the confidence of the team.

The first stage of association rule mining must identify all Large frequency sets items from the original data set. High frequency means that the frequency of a project's presence must reach a certain level relative to all records. The frequency of the occurrence is called A project team Support (Support), with A 2 - with A and B two items itemset, for example, we can through the formula contains {A, B} (1) obtained Support the project team, if the Support is greater than or equal to the set of Minimum Support (Minimum Support) threshold value, the {A, B} is called high frequency project team. A k-itemset that satisfies the minimum support is called Frequent k-itemset, which is usually expressed as Large k or Frequent k. The algorithm generates Large k+1 from the Large k project group until it can no longer find a longer high frequency project.

The second stage of Association Rules mining is to generate Association Rules. Association rules from high frequency project team, is to use the high frequency of the previous step k - program to generate the rules, the Minimum reliability (Minimum Confidence) under the condition of threshold, if the rule is obtained by reliability to meet the Minimum reliability, according to the rules for the association rules. For example, the reliability of rule AB generated by high frequency k-project group {A,B} can be obtained by formula (2). If the reliability is greater than or equal to the minimum trust, then AB is called the association rule.

Association rule mining is generally applicable to the case where the index in the record is taken as a discrete value. If the original indexes is continuous data in the database, in association rules mining should be performed before the appropriate data discretization (in fact, is to a value corresponding to a certain value range), data discretization is an important part of the former data mining, the process of discretization is reasonable will directly affect the results of association rule mining.

2.3. Classification of Association Rules

2.3.1. Based on the type of values processed in a rule or pattern. Variables processed by association rules can be classified into Boolean and numeric types. Boolean association rules deal with values that are discrete and categorized, which shows the relationship between these variables. And numeric association rules can be combined and multidimensional association or multi-level association rules, to deal with numeric fields, the dynamic segmentation, or directly to the original data, of course can also contain numeric association rules are variable. For example, gender = "male" => occupation = "lawyer", which is the Boolean association rule. Gender = "female" =>avg (income) =2300, the income involved is a numerical type, so it is a numerical association rule.

2.3.2. Mining based on infrequent mode and negative correlation mode. What we often focus on in life may be the mining of frequent itemsets, but sometimes it's interesting to find rare or negative patterns. Sales of diamond watches are infrequent in jewelry sales data, but transactions involving

diamond watch sales can be interesting. There is data in the supermarket, if we find customers frequent for classic coke and diet coke, but can't be bought, two are together buy coke and diet coke classic is considered to be a negative correlation model.

2.3.3. Mining based on multi-layer and multi-dimensional space. Data in association rules can be divided into single-dimensional and multi-dimensional data and single-layer and multi-layer data. In the one-dimensional association rules, we only involve one dimension of data, such as items purchased by users. In a multidimensional association rule, the data to be processed will involve multiple dimensions. In other words, a single-dimensional association rule deals with relationships within a single attribute; Multidimensional association rules are rules that deal with certain relationships between attributes. In multi-level association rules and association rules between a data item value often appear in the concept of the higher layer, therefore, mining association rules in multi-level association rules than single layer can get a more in-depth knowledge. According to the granularity level of corresponding items in rules, multi-layer association rules can be divided into the same layer and inter-layer association rules. Two strategies for setting support for multi-layer association rule mining set different minimum support for unified minimum support and different levels. The former is relatively easy to generate rules, but does not take into account the precision of each level, which is easy to cause information loss and information redundancy. The latter increases the flexibility of mining.

2.3.4. Mining association rules based on constraints. Thousands of rules can be discovered from a given data set during data mining, most of which are not relevant or interesting to the user. Often users have good judgment about which direction to dig for rules that might generate interest. Therefore, a good heuristic method is to let users explain their intuitions or expectations as constraints to the search space. This strategy is called constraint based association rule mining

3. Association rule algorithm

3.1. Classical Association Rule Mining Algorithm

Apriori classical association rule mining algorithm is based on two core theories: the subset of frequent item set is frequent item set; The superset of an infrequent itemset is an infrequent itemset. Its main ideas are:

(1) $L1 = \{large\ 1\text{-}itemsets\}$;

(2) for $(k = 2; Lk - 1 \neq \emptyset\ ; K + +)$ do begin

(3) $Ck = appriori\text{-}gen\ (Lk - 1)$;

(4) for all transactions $t \in D$ do begin

(5) $Ct = subset\ (Ck, t)$;

(6) for all candidates $c \in Ct$ do c. count $+ +$;

(7) end

(8) $Lk = \{\ c \in Ck \mid c.\ count \geq minsup\_\ count\}$

(9) end

(10) $L = \cup Lk$;

Through single-pass scan data sets get frequent 1 - itemsets, layered iteration after discovered by the previous iteration (k - 1) - frequent itemsets generate candidate k - itemsets, when there is no new frequent itemsets over. Apriori algorithm generates candidate itemsets based on Apriori property, which greatly reduces the size of frequent sets and shows good performance. But as the database grows, there are also two fatal performance bottlenecks: a large number of candidate sets are generated; Scanning the database multiple times requires a considerable I /O load. In view of the disadvantages of Apriori algorithm, subsequent researchers proposed improved algorithm, such as serial algorithm

based on hash technology, partition based, sampling based, FP growth and parallel distributed algorithm.

3.1.1. Serial Algorithm. Park et al. [2] is proposed based on a hash algorithm of frequent itemsets (hash) technology, it USES hash function to the candidate project distribution in the column to a different hash bucket to count, and the project in each bucket subset will be greater than or equal to the minimum support count project subset called frequent itemsets. This method filters the candidate set generated by k-item set connection through hash table, and reduces the data set effectively. DHP were studied. the optimum processing algorithm is to generate the hash tables and stored data set for overhead storage space for the improvement of the performance of the algorithm, when large itemsets in a database, and generated in the process of the hash table efficiency due to the large amount of calculation will be reduced.

Algorithm based on partition (partition) [3] the main idea is to database is divided into several mutually disjoint, fully stored in the memory in the transaction database, using the algorithm of mining the frequent itemsets of the parts, then the set of all frequent items and generate candidate set of transaction databases, and support by scanning calculation and the frequent itemsets of transaction databases. The algorithm only needs to scan the database twice in the process of execution, which overcomes the bottleneck of high cost Apriori algorithm I /O.

For number according to the library number according to the large amount, brush stroke number according to the library once more, Toivonen [4] is proposed based on Sampling from association rules algorithm thought, first extract sample data from a database D D 'get some strong association rules, with the rest of the database D - D' to verify if it is correct. Sampling technology can quickly realize frequent itemsets mining, but as a result from algorithm using random Sampling method to sample and easy to cause data distortion (data skew) problems, lead to the sample database mining results with the database D mining results error increase. In view of the data skew problem, Lin et al. [5] proposed the anti-skew algorithm. Han et al. [6] proposed a does not produce the discovery of the candidate set of frequent itemsets mining methods FP - growth, it USES a form of FP - tree data structure to compress the data, in the process of generating frequent itemsets partition strategy bottom-up generated suffix item set and FP - tree structure condition, and then explore a particular item at the end of the frequent itemsets. Experiments show that the generated fp-tree is small enough or the path overlap is large enough, and the operation speed of fp-growth algorithm is several orders of magnitude faster than that of Apriori algorithm.

3.1.2. Parallel distributed algorithm. Practical applications, the association rules to deal with the amount of data show exponential growth, which makes the problems focused on the algorithm of association rules mining efficiency and I/O load, even if used on a single processor optimized serial algorithm cannot meet the needs of mining properties, and the use of a multiprocessor system for parallel computing can improve the mining efficiency. Agrawal et al. [7] proposed three parallel algorithms, CD, DD and CaD.CD algorithm to generate candidate set all stored in each processor, using the Apriori algorithm to calculate the candidate set on a local database support count, then exchange the processor's local candidate set support count, makes each processor global support count to find the global frequent itemsets. Each processor synchronizes at the end of each cycle.CD algorithm achieves parallelization by dividing database, but it does not realize parallelization when generating candidate sets. When candidate sets are large, there is insufficient memory.DD algorithm divides candidate sets into different processors, which overcomes the problem of low memory utilization of CD algorithm, but this algorithm has a large traffic volume and must be completed on the processor with high communication speed. The CaD algorithm integrates CD and DD algorithms, and redistributes the database while allocating frequent 1-item sets, enabling each processor to independently complete the work of generating candidate sets. Product data management (PDM) algorithm [2] is by Park et al in DHP were studied. the optimum processing algorithm based on the improved parallel algorithm, the algorithm is similar to the CD algorithm, the candidate set generated

by parallel and parallel to determine the composition of frequent itemsets, when o 2 - frequent itemsets, only exchange hash table to satisfy minimum support count itemsets. PDM algorithm not only inherits the advantages of DHP in reducing the number of candidate sets and transaction databases, but also realizes the parallelization of hash table composition, and the input robustness is quite good. Based on DIC minds, Cheung and others brought by [8] APM line and calculate method, APM using global pruning technology about reducing candidate 2 - itemsets, which is effective when the data distribution, and DIC algorithms require all the data in the database. For this reason, the APM algorithm needs to cluster the database into a homogeneous sub-database according to the number of processors, and then implement DIC algorithm on the sub-database by each processor until no new candidate set is generated. For the low efficiency of DD algorithm, IDD and HD algorithms are introduced [9]. In the IDD algorithm, the local database communicates with other operators through a comprehensive looped broadcast mode. Each processor has a sending buffer (sbuf) and a receiving buffer (rbuf) for asynchronous sending and receiving operations between adjacent nodes of each processor. Compared with DD algorithm, IDD algorithm only performs point-to-point communication once between adjacent processors, reducing communication times and eliminating network competition. To reduce the redundancy of candidate set generation, the IDD algorithm filters out those prefixes by detecting the prefixes of the item set. HD algorithm integrated the CD and IDD algorithm, the algorithm could be divided into equal to the size of the processor, using CD algorithm between groups, groups within the IDD algorithm, HD algorithm has some advantages in dealing with large-scale database, and load balancing.

3.2. Mining Association Rules Based on Data Flow
A new data mining model based on the data flow in the financial management, the stock trend analysis, wireless sensor network, and other fields are widely used, it is different from the traditional static database model, data flow is generated by an online online, its characteristics, such as continuous and unbounded 2 cui yan, such as: the characteristics of the review, 133, sequence association rules mining. Therefore, association rule mining based on data flow should not adopt the previous form of scanning data multiple times, but should carry out single scanning with data update. In order to adapt to the characteristics of data flow and solve the problem of insufficient storage space, the window query of data flow is generally carried out by means of sliding window technology. Giannella et al. [10] proposed FP - stream algorithm for mining frequent itemsets is a classic algorithm of data stream mining, to embed the tilt time window table based on memory frequently and the frequent sequence information in the frequent pattern tree (pattern - tree), dig up with approximate support time sensitive frequent pattern, and realize the time granularity

Pattern incremental maintenance. According to the concept of "provide a frequent closed itemsets without losing support information of minimum said", for mining frequent closed itemsets, on this in time and space complexity of improve the mining efficiency. Chi et al. [11] the Moment algorithm with sliding window technique based on data stream mining frequent itemsets, the use of addition and deletion two operations update sequence, improve the internal storage update speed system. In the literature [12] DS - CFI algorithm sliding window is first divided into multiple Windows, the calculation of each window frequent item sets in the DSCFI - tree, closed recycling DSCFI - tree incremental updating dynamic mining frequent closed itemsets in the sliding window. Literature [13] to study the time sensitive data stream mining algorithm of frequent itemsets FIMoTS, the algorithm use type change line concept to dynamic classification of itemsets, according to the size of the sliding window change only to deal with beyond the boundaries of itemsets in order to improve the computational efficiency. Literature [14] proposed a fiut-stream algorithm, which took bittable as the basic structure, compressed the data into the table, updated the list nodes through the sliding window, and used FIUT algorithm to mine frequent itemsets. Manku et al. [15] proposed data stream mining algorithm Sampling and Lossy counting based on the idea of data segmentation. Literature [16] uncertain data stream is studied under the environment of top - k frequent query algorithm, incremental updating algorithm USES the existing results, and by using poisson distribution structure

elements become frequent items on the probability of lower bound to the data filtering, greatly reduce the query time.

3.3. Association Rule Mining Based on Graph

Graph mining refers to the application of association analysis to graph based data, and the discovery of a set of common substructures in the collection of graphs, that is, frequent subgraph mining. According to the mining search path frequent subgraph mining algorithm is divided into BFS broad first search and DFS deep first search. Breadth first search algorithms include AGM [17] and FSG [18].AGM algorithm takes frequent vertices as the initial set based on Apriori, and USES recursive method to gradually add nodes to excavate all frequent subgraphs. On the basis of AGM algorithm, FSG has made improvements. Instead of increasing the vertex generation candidate subgraph pattern, it has optimized the pruning strategy of candidate subgraph and improved the speed of computing support.In view of the disadvantages of Apriori algorithm producing a large number of candidate subgraphs, fp-growth based fp-growth algorithm gSpan, FFSM, close Graph are presented. Yan et al. [19] proposed the gSpan algorithm based on the depth-first search algorithm for the first time. In the algorithm, the minimum DFS encoding is marked and the right-most extension is carried out to avoid the generation of duplicating graphs to reduce redundancy. In view of the existing algorithm for mining frequent subgraph output subgraph too many questions, literature [20] put forward a new measurement between figure and hypergraph model: Δ - jump model, converts mining frequent subgraph patterns in the database to jump mining frequent patterns, and designed based on extended outside cutting and cutting technology in efficient mining algorithm GraphJP, through a large number of data validation, GraphJP algorithm can efficiently achieve the frequent leap pattern mining.[21] studied a kind of uncertain figure of frequent subgraph mining method, to graph S in uncertain database on the determination of D is included in the database support the probability distribution of expectations subgraph importance as a threshold value evaluation, using the search space pruning technology at the same time, this model reduces the number of subgraph isomorphism testing, improve the efficiency of the search space.[22] introduced a search strategy based on partitioning mixing depth and width of uncertain EDFS frequent subgraph mining algorithm, namely first carries on the depth first search, in its search results after using breadth-first search algorithm.

3.4. Sequential Association Rule Mining

Agrawal and Strikant first proposed the concept of sequential pattern mining, which is the process of mining frequent subsequences from a sequence database with minimal support. Sequential pattern mining is different from association rule mining, which mainly studies the relation between item sets. Agrawal et al., put forward three candidate code generation based on Apriori nature - test mining algorithm: AprioriAll, AprioriSome, DynamicSome [23], and then puts forward the classic Apriori algorithm GSP [24], the above algorithms are based on the level of format of the algorithm. Zaki [25] sequence pattern algorithm was proposed based on the vertical format SPADE, converts the sequence database to record a set position of the vertical format database, and then dynamically linked mining frequent sequential patterns, three times algorithm by scanning database, reduced the I/O overhead.[26] proposed a model based on projection PrefixSpan growth algorithm, it USES frequent sequence prefix projected recursive producing frequent sequential patterns, algorithm avoids the overhead of the problems of the candidate set generated. Lin et al. [27] was proposed based on the sequence of memory index MEMISP model method, algorithm for frequent sequences using the index technology, scan the database once and do not produce the candidate sequences and projection database, PrefixSpan is obviously better than the memory utilization. Maximum sequence pattern mining and closed sequence pattern mining can effectively compress the number of result sets of traditional sequence pattern mining, for which, tong yongxin et al. [28] applied these two technologies

Combined with the advantages of the proposed compressed sequence pattern mining algorithm CFSP, a small number of representative frequent sequence combinations to represent all sequence pattern information. The two-step method is adopted in the algorithm. First, most of the representative

sequence patterns are mined according to the interference dominant sequence pattern detection mechanism. Secondly, the remaining representative sequences are mined with a small amount of time. In addition, there are other extended sequential pattern compression algorithms for different requirements.[29] for the first time introduced the ideas of logic in frequent sequential pattern mining, reducing the dependence of the mining results to support threshold, algorithm by logical equivalence rules to filter results, delete do not conform to the rules of logic, greatly improve the intelligibility and practicability of the set of rules.[30] with a wildcard is designed (wildcards) of One - Off mining, sequence pattern mining method, the calculation mode of support at the same time satisfy the conditions of One - Off: model any two time don't share the same position in the sequence of characters. More meaningful patterns can be found in biomedical and Web log fields by introducing wildcard sequence patterns.

4. Developement trend of association rules

Association is following the bayesian classification algorithm (Bayes theorem), support vector machine (SVM), and other machine learning methods after another important classification techniques, are widely used in text categorization, a Web document classification, medical image classification. Classical association classification algorithms are CBA, CMAR and CPAR. [42] will be applied to text classification, associative classification and traditional associative classification algorithm depend on the test set the classification of the support and confidence threshold low effect problem, put forward a kind of based on simulated annealing technology support and confidence associative classification algorithm of intelligent optimization. Literature [43] proposed a method of high precision classification of related text based on category similarity aggregation. Medical image plays an important auxiliary role in medical diagnosis. Medical image data mining is different from general data mining, which requires stability, high efficiency and reliability. [44] put forward a kind of suitable for medical image data mining CARMI associative classification algorithm, by introducing a double degree reduced image data intensive support, support degree classification of interference caused by higher characteristic, realize high-efficient and classification of image data. Network intrusion detection technology as a powerful assistant firewall technology, effectively make up for the firewall policy when dealing with hacker attacks, information theft, become a reliable safeguard network security monitoring system. Traditional intrusion detection system (IDS) USES association rules to generate standard intrusion detection rules library, and generates alarm data flow by judging the matching degree between network data and rule library. [45] proposes a network intrusion detection model based on the context to validate (CVNIDM), will attack external conditions into consideration in order to distinguish of IDS alerts information, so as to improve the accuracy of IDS alerts. In literature [46], a comprehensive correlation confidence degree (h-confidence degree) association rule algorithm for calculating adjacent alarms is proposed to effectively discover the alarm association rules with low support degree and high confidence degree. Privacy data mining refers to the interference in the process of data mining by adopting data and query restrictions to protect the original data, the basic strategy to avoid business sensitive data, personal data privacy leak. Literature [47] proposed a privacy protection association rule algorithm EMASK based on data interference and distribution reconstruction. BEMASK algorithm [48] improved the EMASK algorithm based on the idea of granularity calculation, and transformed relational data tables into machine-oriented relational models by means of granularity

The calculated frequent itemsets are converted to calculate the intersection of basic particles, and the data are represented vertically, which reduces the number of I /O operations compared with EMASK algorithm.[49] put forward a kind of meet the uniform distribution of uncertain data association rules algorithm UFI - DM, effectively solve the traditional data mining technique is applied to uncertain factors in the privacy protection problem when inefficient or infeasible problem. USES association rules for protein structure prediction, according to the characteristics of the protein sequences and handle these sequence data can be quantitatively, and the association rules algorithm used in the data sets for protein sequences in relationship. Literature [50] knowledge discovery based on inner cognitive mechanism (KDTICM) and knowledge discovery in database (KDD) model, put

forward a kind of protein secondary structure prediction method based on associative classification SAC, the algorithm with 85% accuracy the successful prediction of protein sequences. With the increase of the number of satellites such as remote sensing navigation and positioning, geophysics, etc., big data of space earth brings new opportunities for the study of earth science. The traditional data analysis method is mainly based on statistical analysis and nonlinear fitting. The mining model based on association rules can reveal the relationships among the earth data such as ocean, land and atmosphere, thus promoting the development of the field of global change and disaster science. Literature [51] used multiple constraints to excavate the time-series association rules, and analyzed and obtained the relationship between climate index which was highly consistent with the actual situation and abnormal precipitation events in the land area.

5. Conclusion

After a long period of research and development, association rule mining has become increasingly mature in the design and optimization of frequent pattern mining algorithms and is widely used in Internet, finance, biological information and other fields. But the direction of future research in this field is still challenging Work: design more efficient mining algorithm; Realize the interaction between the user and the mining system and develop the easy-to-understand visual interface. To improve the extended mining algorithm in special field, such as periodic pattern mining. Extend the application domain of association rules

Acknowledgments

References

[1]   Agrawal Ｒ, Srikant Ｒ. Fast algorithms for mining association rules [C] / /Proc of International Conference on Very Large Databases. 1994: 487-499.

[2]   Park J S, Chen M S, Yu P S. An effective hash-based algorithm formining association rules [J] . SIGMOD Ｒecord, 1995, 25 (2) : 175-186.

[3]   Savasere A, Omiecinski E, Navathe S. An efficient algorithm for mining association rules in large databases [C] / /Proc of the 21st International Conference on Very Large Databases. 1995.

[4]   Toivonen H. Sampling large databases for association rules [C] / /Proc of the 22nd International Conference on Very Large Databases. 1996.

[5]   Lin Junlin, Dunham, M H. Mining association rules: anti-skew algorithms [C] / /Proc of the 14th International Conference on Data Engineering. 1998: 486-493.

[6]   Han Jiawei, Pei Jian, Yin Yiwen. Mining frequent patterns withoutcandidate generation [C] / /Proc of ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 2000: 1-12.

[7]   Agrawal Ｒ, Shafer J. Parallel mining of association rules [J] . IEEETrans on Knowledge and Data Engeering, 1996, 8(6) : 962-969.

[8]   Cheung D W, Han J, Ng V, et al. A fast distributed algorithm formining association rules [C] / /Proc of International Conference onParallel and Distributed Information Systems. 1996: 31-44.

[9]   Han E S, Karypis G, Kumar V. Scalable parallel data mining for association rules [C] / /Proc of ACM SIGMOD International Conference onManagement of Data. New York: ACM Press, 1997: 277-288.

[10]   Giannella C, Han Jiawei, Pei Jian, et al. Mining frequent patterns indata streams at multiple

time granularities [J] . Next GenerationData Mining, 2006, 35 (1) : 61-84.

[11] Chi Yun, Wang Haixun, Yu P S, et al. Moment: maintaining closedfrequent itemsets over a stream sliding window [C] / /Proc of the 4th International Conference on Data Mining. 2004: 59-66.

[12] Liu xuejun, xu hongbing, dong yisheng, etc.Closed frequent mode of data flow based on sliding window [J].Computer research and development, 2006,43 (10) : 1738-1743.

[13] Li haifeng, zhang ning, zhu jianming, etc.Frequent itemset mining on time-sensitive data streams Law [J].Journal of computer science, 2012,35 (11) : 2283-2293.

[14] Kou xiangxia, ren yonggong, song kuiyong.A data stream frequent item set mining based on sliding window Mining algorithm [J].Computer applications and software, 2013,30 (1) : 143-146.

[15] Manku G S, Motwani Ｒ. Approximate frequency counts over streaming data [C] / /Proc of the 28th International Conference on Very Large Databases. 2002: 346-357.

[16] Wang shuang, wang guoren.Study on top-K probability frequent item query algorithm based on sliding window [J].Computer research and development, 2012,49 (10) : 2189-2197.

[17] Inokuchi A, Washio T, Motoda H. An Apriori-based algorithm formining frequentsubstructures from graph data [C] / /Proc of the European Symposium on the Principle of Data Mining and Knowledge Discovery. 2000: 13-23.

[18] Kuramochi M, Karypis G. Frequent subgraph discovery [C] / /Proc of the 1st IEEE International Conference on Data Mining. 2001: 313-320.

[19] Yan Xifeng, Han Jiawei. gSpan: graph-based substructure pattern mining [C] / /Proc of the 2nd IEEE International Conference on Data Mining. 2002: 721-724.

[20] Liu yong, li jianzhong, gao hong.Mining frequent skip patterns from the graph database [J].software Journal, 2010,21 (10) : 2477-2493.

[21] Zou zhaoxin, li jianzhong, gao hong, etc.Mining frequent subgraph patterns in uncertain graphs [J].Software journal, 2009,20 (11) : 2965-2976.

[22] Hu jian, he linbo, MAO yimin, etc.Research on an improved algorithm for mining uncertain frequent subgraphs Look at [J].Computer engineering and applications, 2015,51 (3) : 112-116.

[23] Agrawal Ｒ, Srikant Ｒ. Mining sequential patterns [C] / /Proc of the 11th International Conference on Data Engineering. 1995: 3-14.

[24] Srikant Ｒ, Agrawal Ｒ. Mining sequential patterns: generalizations and performance improvements [C] / /Proc of the 5th International Conference on Extending Data Base Technology. 1996: 3-17.

[25] Zaki M J. SPADE: an efficient algorithm for mining frequent sequences [J] . Machine Learning, 2001, 42(1) : 31-60.

[26] Pei Jian, Han Jiawei, Mortazavi-Asl B, et al. PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth [C] / /Proc of the 17th International Conference on Data Engineering. 2001: 215-224.

[27] Lin M, Lee S. Fast discovery of sequential patterns by memory indexing [C] / /Proc of the 4th International Conference of Data Warehousing and Knowledge Discovery. 2002: 150-160.

[28] Tong wing - xin, zhang yuanyuan, yuan mei, etc.An efficient algorithm for mining compressed sequence patterns [J].Computer research and development, 2010,47 (1) : 72-80.

[29] Liu duanyang, feng jian, li xiaofan.A frequent sequence pattern mining algorithm based on logic [J].Computer science, 2015,42 (5) : 260-264.

[30] Wu xindong, xie fei, huang yongming, etc.A sequence pattern with wildcards and one-off conditions Dig [J].Software journal, 2013,24 (8) : 1804-1815.

[31] Cui lixin, yuan senmiao and zhao chunxi.Binding associative rule discovery method and algorithm [J].Journal of computer science, 2000,23 (2) : 216-220.

[32] Zhao jialu, Yang jun, han jing, etc.Constrained association rule mining based on a set of transaction ids Mining algorithm [J].Computer engineering and design, 2013,34 (5) : 1663-

        1667.
[33] Wang jiale, gu jin.Constrained association rule mining algorithm based on attribute bit reuse
        [J].Computer engineering and applications, 2011,47 (7) : 131-134.
[34] Dong yanshi, cheng yiyu, pan yunhe.Item constraint association rule discovery based on high
        frequency pattern tree Method [J].Journal of zhejiang university, 2002,36 (4) : 64-69.
[35] Song yuqing, zhu yuquan, sun zhihui, etc.A constrained maximum frequency based on a
        frequent pattern tree Complex item set mining and its updating algorithm [J].Computer
        research and development, 2005, 42 (5) : 777-783.
[36] Hua hong Juan, zhang jian, Chen shaohua.Constraint maximum frequent itemset mining based
        on frequent pattern trees Algorithm [J].Computer engineering, 2011,37 (9) : 78-80.
[37] Fu dongmei, wang zhiqiang.Association rule mining algorithm based on fp-tree and constraint
        concept lattice And applied research [J].Computer application research, 2104,31 (4) : 1103-
        1109.
[38] Srikant Ｒ, Agrawal Ｒ. Mining generalized association rules [C] / /Proc of the 21st
        International Conference on Very Large Databases. 1995:407-419.
[39] Klemettinen M, Mannila H, Toivonen P Ｒ H, et al. Finding interesting rules from large sets of
        diacovered association rules [C] / /Proc of the 3rd International Conference on Information
        and Knowledge Management. 1994: 401-407.
[40] Zhou xin, sha chaofeng, zhu yangyong, etc.Degree of interest - another threshold for association
        rules [J].Computer research and development, 2000,37 (5) : 627-633.
[41] Ma jianqing, zhong yiping, zhang shiyong. Association rule mining algorithm based on interest
        degree [J].Computer engineering, 2006,32 (17) : 121-122.
[42] Wang xiuzhi, an was built. Association classification algorithm based on intelligent
        optimization of support and confidence[J].Computer applications and software, 2013,30
        (11) : 184-186.
[43] Tian feng, GUI xiaolin, Yang pan, etc.An association text classifier that USES category
        similarity aggregation Law [J].Journal of xi 'an jiaotong university, 2012,46 (12) : 6-11.
[44] Wuyuan garden, yu qing song, zhu yuquan. Medical image classification algorithm based on
        association rules[J].Computer engineering and design, 2008,29 (12) : 3234-3236.
[45] Tian zhihong, wang bailing, zhang weizhe, etc.Network intrusion detection module based on
        context validation Type [J].Computer research and development, 2013,50 (3) : 498-508.
[46] Gao huisheng, li yingmin. Mining alarm association rules based on comprehensive correlation
        confidence [J].Telecommunication science, 2014 (10) : 110-115.
[47] Agrawal S, Krishnan V, Haritsa J Ｒ. On addressing efficiency concerns in privacy-preserving
        mining [C] / /Proc of the 9th International Conference on Database Systems for Advanced
        Applications. 2004: 113-124.
[48] Cheng shutong, xu from rich, but hongwei. Research on efficient privacy protection frequent
        pattern mining algorithm [J].Computer science, 2015,42 (4) : 194-198.
[49] Chen aidong, liu guohua, fei fan, etc. Uncertain data association rules that satisfy uniform
        distribution Mining algorithm [J].Computer research and development, 2013,50
        (supplement) : 186-195.
[50] Yang bingru, zhou zun, hou wei.Association multiple classification algorithm based on protein
        secondary sequence [J].Systems engineering and electronics, 2010,32 (6) : 1318-1324.
[51] Shi yan, deng min, liu qiliang, etc.Mining methods for association rules of Marine and
        continental climate events [J].to Ball information science, 2014,16 (2) : 182-189.