Jianlin Feng and Yucai Feng [1998] Mining association rules is an important data mining problem. A fast binary partition-based algorithm (BPA) for mining association rules in large databases is presented in this paper. Basically, the framework of BPA is similar to that of the algorithm Apriori. In the first pass, all the frequent 1-item sets are divided into two disjoint parts. Accordingly, in each subsequent pass k, we partition the set of all the frequent k-item sets into three subsets.

O. R. Zaiane et al. [2000] Despite the overwhelming amounts of multimedia data recently generated and the significance of such data, very few people have systematically investigated multimedia data mining. With our previous studies on content-based retrieval of visual artifacts, we study in this paper the methods for mining content-based associations with recurrent items and with spatial relationships from large visual data repositories. A progressive resolution refinement approach is proposed in which frequent item-sets at rough resolution levels are mined, and progressively, finer resolutions are mined only on the candidate frequent items-sets derived from mining rough resolution levels.

G. Grahne et al. [2001] Previous work on frequent item set mining has focused on finding all itemsets that are frequent in a specified part of a database. We motivate the dual question of finding under what circumstances a given item set satisfies a pattern of interest (e.g., frequency) in a database. Circumstances form a lattice that generalizes the instance lattice associated with datacube. Exploiting this, we adapt known cube algorithms and propose our own, minCirc, for mining the strongest (e.g., minimal) circumstances under which an itemset satisfies a pattern. Our experiments show that minCirc is competitive with the adapted algorithms.

Chiech-Ming Wu and Yin-Fu Huang [2002] In recent ten years, with the developments of information and advances of Internet, vast data have been propagated and recorded in the databases of different applications. Currently data mining is arising to understand, analyze, and use these data. Data mining is designed for finding interesting samples in a large database, and thus it is also the knowledge exploration center part. One of the major tasks of data mining is to find association rules for helping retail industries to understand the consumers' behaviors.
Ren Jiang-tao et al. [2003] In the past years the number of algorithms and techniques for data mining has grown tremendously, much useful and valuable information can be obtained through data mining. With the development of the Intelligent Transportation Systems (ITS), more and more data from loops or cameras can be collected and used, which is from a good basis for the use of data mining approaches in ITS. In this paper, the frequent item-set mining algorithm is used to discover simultaneously congested link-sets in a road network. Because the amount of discovered item-sets is great, a ranking mechanism is adopted to improve the efficiency of picking the most interesting link-sets. The experiment results show that the approaches are useful for the discovering such link sets automatically and quickly.

Xiang-Ping Meng et al.[2003] Association rule mining is one of the important data mining tasks. However, the previously proposed methods still encounter some problems, such as complex data structure, candidate set generation, and so on. To improve efficiency, association rules can be mined in parallel. In this paper, we use a simpler data structure called bit string array and propose a new approach to apply parallel projection and compress technique in parallel mining association rules. It conducts various operations on bit string array according to the frequency of frequent items. For frequent item with less frequency, we conduct set operation on them; for frequent item with more frequency, we adopt compress technique and conduct bit AND operator on them. Moreover, it will reduce the communication cost and also response time. This method can be scaled up to very large databases by parallel projection and compress technique.

Raymond Chan et al. [2003] Traditional association rule mining algorithms only generate a large number of highly frequent rules, but these rules do not provide useful answers for what the high utility rules are. We develop a novel idea of top-K objective-directed data mining, which focuses on mining the top-K high utility closed patterns that directly support a given business objective. To association mining, we add the concept of utility to capture highly desirable statistical patterns and present a level-wise item-set mining algorithm. With both positive and negative utilities, the antimonotone pruning strategy in Apriori algorithm no longer holds. In response, we develop a new pruning strategy based on utilities that allow pruning of low utility itemsets to be done by means of a weaker but antimonotonic condition.

S. Mehta et al. [2004] Discretization is a crucial preprocessing primitive for a variety of data warehousing and mining tasks. In this article we present a novel PCA-based unsupervised
algorithm for the discretization of continuous attributes in multivariate datasets. The algorithm leverages the underlying correlation structure in the dataset to obtain the discrete intervals, and ensures that the inherent correlations are preserved. The approach also extends easily to datasets containing missing values. We demonstrate the efficacy of the approach on real datasets and as a preprocessing step for both classification and frequent item set mining tasks. We also show that the intervals are meaningful and can uncover hidden patterns in data.

Yin-Ling Cheung and Ada Wai-Chee Fu [2004] In classical association rules mining, a minimum support threshold is assumed to be available for mining frequent itemsets. However, setting such a threshold is typically hard. We handle a more practical problem; roughly speaking, it is to mine N k-itemsets with the highest supports for k up to a certain k/sub max/ value. We call the results the N-most interesting itemsets. Generally, it is more straightforward for users to determine N and k/sub max/.

Yo-Ping Huang and Li-Jen Kao [2004] Apriori algorithm, the widely adopted approach, exploits the following property to derive the frequent itemset: if an itemset is frequent, so are all its subsets. That is, Apriori algorithm generates itemsets in a level-wise manner where each candidate in the jth iteration is generated from previous frequent (j-1)-itemsets. A generated candidate can be further pruned if any subset of size j-1 is not a frequent itemset. Apriori algorithm relies on the essential assumption that all itemsets have a uniform minimum support value, i.e., we assume that all items in the dataset have the same nature, e.g., all the items have the same sale price or the same salability condition in different time intervals or locations. However, the assumption may not comply with the rules embedded in the large databases.

Chin-Feng Lee and Tsung-Hsien Shen [2005] Recently, most of the studies on association rules mining focused on improving the efficiency of frequent itemsets generation. To our best knowledge, the FP-growth algorithm, which is based on the FP-tree to generate frequent itemsets is time-efficient. Currently, relevant studies are introduced to improve the FP-growth algorithm. However, they ignore the fact that the FP-tree construction may spend much time. Therefore, the goal of our research is to propose a fast algorithm called frequent pattern split, simply FP-split, for improving the process of the FP-tree construction. The proposed FP-split algorithm contains two main steps. The first step is to scan a transaction database only once for generating
equivalence classes of frequent items. The second step is to sort these equivalence classes of frequent items in descending order so as to construct the FP-split tree.

**T. Washio et al. [2005]** A novel approach to subspace clustering is proposed to exhaustively and efficiently mine quantitative frequent item-sets (QFIs) from massive transaction data. For the computational tractability, our approach introduces adaptive density-based and Apriori-like algorithm. Its outstanding performance is shown through numerical experiments.

**Ruoming Jin and G. Agrawal [2005]** Frequent item set mining is a core data mining operation and has been extensively studied over the last decade. This paper takes a new approach for this problem and makes two major contributions. First, we present a one pass algorithm for frequent item set mining, which has deterministic bounds on the accuracy, and does not require any out-of-core summary structure. Second, because our one pass algorithm does not produce any false negatives, it can be easily extended to a two pass accurate algorithm.

**J. Wang et al. [2006]** Mining traffic to identify the dominant flows sent over a given link, over a specified time interval, is a valuable capability with applications to traffic auditing, simulation, visualization, as well as anomaly detection. Recently, Estan advanced a comprehensive data mining structure tailored for networking data—a parsimonious, multidimensional flow hierarchy, along with an algorithm for its construction. While they primarily targeted offline auditing, use in interactive traffic visualization and anomaly/attack detection will require real-time data mining. We suggest several improvements to Estan's algorithm that substantially reduce the computational complexity of multidimensional flow mining.

**Ke Wang et al. [2006]** The issue of mining association rules with composite items was proposed several years ago. Algorithms with composite items have the potential to discover rules which cannot be found out by other algorithms without composite items. However, much redundant rules which are of trivial significance or even incorrect will be also discovered by these algorithms in certain cases. In this paper, the authors design a novel frequent-pattern tree for finding large composite items first. And then how to measure the reliability of these discovered rules with composite items in order to find out the most reliable association rules is discussed.

**Jianying Hu and A. Mojsilovic [2006]** Despite the wide use of data mining techniques in client segmentation and market analysis applications, so far there have been no algorithms that allow
for the discovery of strategically important combinations of products (or offerings) - the ones that have the highest impact on the performance of the company. We present a novel algorithm for analyzing a multi-product environment and identifying strategically important combinations of offerings with respect to a predefined criterion, such as revenue impact, profit impact, inventory turnover etc. In contrast to the traditional association rule and frequent item mining techniques, the goal of the new algorithm is to find segments of data, defined through combinations of products (rules), which satisfy certain conditions as a group. We present a novel algorithm to derive specialized partition threes, called high yield partition trees, which lead to such segments, and investigate different splitting strategies.

**YhiShiau et al.[2007]** Local region conservation has been studied for many years because biologists believe that local conservation could be highly related to protein functions. The concept of local region conservation comes from a motif, a fragment with biological or functional meaning. Besides, structure-based identification of homologues often succeeds where sequence-alone-based methods fail, because in many cases evolution retains the folding pattern long after sequence similarity becomes undetectable.

**ShuYun Wang et al.[2007]** This paper introduce the algorithm MIBFD (mining frequent items using bloom filter based on damped model) for mining recent frequent items in data streams. Based on an efficient data structure named extensible and scalable bloom filter(ESBF), MIBFD is able to adjust the size of memory used dynamically. Theoretical analysis and experiments show that MIBFD is efficient both in processing time and in memory usage.

**Wassim Ayadi and Khedija Arour[2007]** Discovering association rules that identify relationships among sets of items is an important problem in data mining. Finding frequent itemsets is computationally the most expensive step in association rule discovery and therefore it has grasped significant research focus. Discovery of frequently occurring subsets of items, called itemsets, is the core of many data mining methods. Most of the previous studies adopt Apriori-like algorithms, whom iteratively generate candidate itemsets and check their occurrence frequencies in the database. These approaches suffer from serious costs of repeated passes over the analyzed database. In this paper, we propose a new BDD-based (Binary Decision Diagram) data structure called TreeSupBDD. The TREESUPBDD extends the idea claimed by the authors of FP-TREE and ITL-Tree structures, aiming to improve storage compression and to
allow frequent pattern mining without an "explicit" candidate itemset generation step. To address this problem, we propose a novel method, called TreeSupBDD-MINE, for reducing database activity of frequent itemset discovery algorithms.

**Zhang Hong et al. [2007]** This paper defined a kind of multi-dimension data cube model, and presented a new formalization of generalized association rule based on data cube model. After comprehending the weaknesses of the current generalized association rule mining algorithms based on data cube, we proposed a new algorithm GenHibFreq which was suitable for mining multi-level frequent item set based on data cube. By taking advantage of the item taxonomy, algorithm GenHibFreq reduced the number of candidate itemsets counted, and had better efficiency. We also designed an algorithm GenerateLHSs-Rule for generating generalized association rule from multi-level frequent item set. Demonstrated through examples, algorithms proposed in this paper had better efficiency and less generated redundant rules than several existing mining algorithms, such as Cumulate, Stratify and ML_T2L1, and had good performance inflexibility, scalability and complexity and had new ideas on conducting the generalized association rule mining algorithms in multi-dimension environment and it also has great theoretical meaning and practical value.

**Zhang Zheng and Wang Hui-wen [2007]** In this paper, aim at the inefficient problem of the a priori algorithms, we design a new matrix data structure, called cooccurrence matrix, in short COM, to store the data information instead of directly using the transactional database. In COM, any item sets can be randomly accessed and counted without many times full scan of the original transactional database. Based on COM, we first divide association rule into two kinds of rule and then we present an efficient algorithms (COM_mining) to find the valid association rules among the frequent items. Finally we apply COM_mining algorithm and a priori algorithm simultaneously to analyze up-down association relationship between various industry stock blocks of China A stock market. From analytical result we can find that in China A stock market, there are indeed up-down association relationship between various industry stock blocks. At the same time, through comparing COM_mining algorithm and a priori algorithm in this application, we can see, COM_mining is more efficient than a priori.

**Tsai-Pin Chu et al. [2008]** A data stream is an unbounded sequence of data arriving at high speed and changing unceasingly along with the time. Frequent patterns mining in data streams has been
studied extensively in recent years, with many algorithms proposed and implemented successfully. However, frequent patterns mining often generates a huge number of frequent itemsets and their corresponding association rules, including many redundant, useless ones which are difficult to comprehend and manipulate.

Amiya Nayak and Ivan Stojmenovic [2008] The identification of frequent item sets and of association rules have received a lot of attention in data mining due to their many applications in marketing, advertising, inventory control, and many other areas. First the notion of frequent item set is introduced and we study in detail the most popular algorithm for item set identification: the Apriori algorithm. Next we present the role of frequent item sets in the identification of association rules and examine the levelwise algorithms, an important generalization of the Apriori algorithm.

Yun Yu et al. [2008] This paper has analyzed the Apriori algorithm performance, and has pointed out performance bottleneck question of the Apriori algorithm. Currently those algorithms to mine association rules only pay attention to one aspect of efficiency or accuracy respectively. There is a paradox between efficiency and accuracy. In order to resolve to this conflict, a novel algorithm based on Probability estimate and least square estimate is proposed to mine the association rules from database with the high correlativity and the high confidence. Probability estimate reduce the times of database scanning so as to increase efficiency; least square estimate is based on rigorous and classical mathematical model so as to enhance accuracy.

Zhang Hui et al. [2009] Mining of association rules is an important problem in data mining, given a large set of data, extracting frequent item sets in this set is a challenging job in data mining. Item sets matching is the chief problem in extracting frequent item sets. And item set matching is the bottleneck of the mining process. It also has been proved that extracting frequent free item sets is a useful method. Many efficient algorithms have been proposed in the literature. The idea presented in this paper is to divide the database into multiple partitions and then find frequent free item sets in each partition, then merge the several partitions to generate other frequent free item sets and count the support. The algorithm costs little memory to save additional support numbers of item sets in each partition but greatly reduces the time of item set matching which is the bottleneck of the mining process. The experiments on real datasets have showed its good performance.
Sang Lin et al. [2009] Frequent item mining has been extensively used in association rules mining. The goal of frequent itemset mining is to discover all the itemsets whose supports in the database exceed a user-specified threshold. However, it often generates a large number of candidate itemset, which reduce the effectiveness of the mining algorithms. Constraint-based mining enables users to provide restraints on mining their interested association rules and can greatly improve the efficiency of mining tasks. In this paper, we propose a fast constraint-based algorithm for mining maximal frequent itemsets.

Ding Zhenguo et al. [2009] FP-growth algorithm is one of the most efficient approaches for frequent item set mining. In this article, an improved FP-growth algorithm based on Compound Single Linked List is proposed. There are two contributions in the new algorithm. One is to use the sequencing table and single linked list as the main data structure, the other is that it does not need to generate conditional FP-tree. An experiment is conducted to compare the efficiency between the original one and the improved one. It shows that the new one improves the algorithm both in runtime and the main memory consumption.

Huizhen Liu et al. [2009] How to improve the efficiency of discovering the frequent item sets is a major problem in mining association rules. This paper analysed the idea and performance of the general quantitative association rules algorithm ,and put forward a quantitative association rules mining algorithm based on matrix, the new algorithm firstly transformed quantitative database into Boolean matrix ,then used boolean "and" operation to generate frequent item sets on matrix vector .It effectively solved the bottleneck of Apriori algorithm which iteratively produced frequent item sets in the general quantitative association rules algorithm . The results of experiments and analysis showed that the new algorithm effectively improved the efficiency of mining quantitative association rules.

XuePing Zhang et al.[2010] Based on FP-tree algorithm, this article introduced the method of multi-thread processing and a Multi-Threaded Paralleled frequent item-set mining Algorithm—MTPA was proposed. It has been applied to an enterprise human resources management system. Through the experiments of paralleled mining by using increasing multi-thread processing, it has been proved that MTPA which on the condition of multi-core processors can improve the efficiency of frequent item-set mining effectively.
Hai-yan Zhou and Qi Hui[2010] After analyzing many typical association rule mining algorithms, a new algorithm, named as BOFP-V, is proposed for frequent item set mining. FP-V vectors are introduced in order to convert that of frequent item set mining to the course of the vectors operating. The existing Apriori algorithm produces a lot of candidacy sets and needs scanning database many times, and BOM algorithm entails and operation of k vectors with times. Overcoming these drawbacks, BOFP-V algorithm needs scanning database only once. Therefore, the proposed algorithm is obviously superior to Apriori and BOM algorithm in efficiency.

Bac Le et al.[2010] In this paper, we propose algorithms for mining Frequent Weighted Itemsets (FWIs) from weighted items transaction databases. Firstly, we introduce the WIT-tree data structure for mining high utility itemsets in the work of Le et al. (2009) and modify it for mining FWIs. Next, some theorems are proposed. Based on these theorems and the WIT-tree, we propose an algorithm for mining FWIs. Finally, Diffset for fast computing the weighted support of itemsets and saving memory are also discussed. We test the proposed algorithms in many databases and experimental results show that they are very efficient in comparison with Apriori-based approach.

Hai-yan Zhou and Qi Hui[2010] After analyzing many typical association rule mining algorithms, a new algorithm, named as BOFP-V, is proposed for frequent item set mining. FP-V vectors are introduced in order to convert that of frequent item set mining to the course of the vectors operating. The existing Apriori algorithm produces a lot of candidacy sets and needs scanning database many times, and BOM algorithm entails and operation of k vectors with times. Overcoming these drawbacks, BOFP-V algorithm needs scanning database only once. Therefore, the proposed algorithm is obviously superior to Apriori and BOM algorithm in efficiency.

C S Ashwin et al.[2011] Though we have seen many data-mining methods there are always various discrepancies and we have proposed a frequent pattern mining is an important data-mining for finding the correlations among items. Since the frequencies for various items are always varied, specifying a single minimum support cannot exactly discover interesting patterns. In order to overcome these discrepancies we propose an Apriori-based method to include the concept of multiple minimum supports (MMS in short) on association rule mining. It allows user to specify MMS to reflect the different natures of items. Since the mining of sequential pattern may face the same problem, we extend the traditional definition of sequential patterns to include
the concept of MMS in this study. For efficiently discovering sequential patterns with MMS, we develop a data structure, named PLMS-tree, to store all necessary information from database. After that, a pattern growth method, named MSCP-growth, is developed to discover all sequential patterns with MMS from PLMS-tree. This paper can be used in evolving scenarios and sure to create a revolution in its own field.

**Xiao-hong Huang and Xiu-feng Zhang [2010]** The traditional sequential pattern algorithms often have some limitations in solving the applications. First, ignore the time-interval trait of the event sequences, which is more important to the mining result. Secondly, the sequential item is also without attribute constraint. As a result, in this paper an idea is presented to mine event sequential pattern with multi-attribute constraint. Based on the algorithm of AprioriAll and Apriori, the transition time is taken into account between events. According to the layer idea, the key task is to mine the frequent sequential pattern first, then to find out the association rules in the attribute constraint items. In the end provides a way to mine sequential pattern with multi-attribute constraint by example analysis.

**Zhuang Chen et al. [2011]** The paper analyzes the basic ideas and the shortcomings of Apriori algorithm, studies the current major improvement strategies of it. In order to solve the low performance and efficiency of the algorithm caused by its generating lots of candidate sets and scanning the transaction database repeatedly, it studies the pruning optimization and transaction reduction strategies, and on this basis, the improved Apriori algorithm based on pruning optimization and transaction reduction is put forward.

**Ning Li et al.[2012]** Searching frequent patterns in transactional databases is considered as one of the most important data mining problems and Apriori is one of the typical algorithms for this task. Developing fast and efficient algorithms that can handle large volumes of data becomes a challenging task due to the large databases. In this paper, we implement a parallel Apriori algorithm based on MapReduce, which is a framework for processing huge datasets on certain kinds of distributable problems using a large number of computers (nodes).

**ZhuoboRong et al.[2013]** In the single machine environment, the problems of Apriori and FP-Growth algorithm in large-scale data association rules mining are high memory consumption, low computing performance, poor scalability and reliability and so on. Therefore, we put forward
a new implementation method which is based on MapReduce parallel environment for mining frequent itemsets to generate association rules and is verified by using different sizes of real datasets with different nodes in the cluster, meanwhile, selecting “speedup, scalability and reliability” as an indicator.

Qibing Lu and Buyun Sheng [2013] Along with production process automation and development of new products, manufacturing information in large quantity, contains more dimensions, in order to mine useful information from the manufacturing database, monitor and control manufacturing process effectively. A weighted association rules mining algorithm with fuzzy quantitative constraints (FQC-wed Apriori algorithm) is proposed in this paper. First, find association rules after database mining. Then, mine fuzzy association rules with fuzzy query. Last, find frequent item sets with the improved weighted association rules algorithm. Manufacturing process information can be mined and effectiveness of the mining algorithm can be evaluated. The algorithm is applied to manufacturing process information mining in discrete manufacturing industry.

Ke Zhang et al. [2014] In order to avoid scanning the database multiple times, the database mapping method is changed in this research. Meanwhile, after the support of candidate item sets is get, each candidate item set should be determined whether it is a frequent item set or not based on the prior knowledge of Apriori algorithm. If the candidate item sets generated by the element of the existing frequent item sets are certainly not frequent item sets, the element is not necessary to connect with others, which leads to an optimized connecting step. Lastly, for Apriori algorithm, the intersection operation is introduced to address the disadvantages that it takes many time costs to match with candidate item sets and transaction pattern.

Paresh Tanna and Yogesh Ghodasara [2015] The frequent pattern mining algorithms determine the frequent patterns from a transaction database. When the database is updated, the frequent patterns should be updated as well. However, running the frequent pattern mining algorithms with every update is not adequate. This is called the imperative update problem of frequent patterns and the solution is to formulate an algorithm that can with vitality mine the frequent patterns. In this study, an imperative frequent pattern mining algorithm, which is called Imperative Tabularized Apriori Algorithm (ITAA), is proposed and explained. Performance evaluation is given to prove the proposed work.
Frequent itemset mining is an important step of association rules mining. Traditional frequent itemset mining algorithms have certain limitations. For example, the Apriori algorithm has to scan the input data repeatedly, which leads to high I/O load and low performance, and the FP-Growth algorithm is limited by the capacity of computer's inner stores because it needs to build a FP-tree and mine frequent itemset on the basis of the FP-tree in memory. With the coming of the Big Data era, these limitations are becoming more prominent when confronted with mining large-scale data.

She Xiangyang and Zhang Ling [2016] Under the environment of big data, efficiency is low and there are many candidates when the traditional serial Apriori algorithm in dealing with massive data. This paper proposes a parallel better algorithm based on MapReduce distributed architecture. Based on the basic Apriori algorithm on MapReduce, this paper makes a reconstruction of the original transaction database, and implements parallel in data set fragmentation. The algorithm optimizes the transaction database, candidate item sets counting and pruning strategy. The experimental results show that the improved algorithm proposed in this paper can reduce the candidate items and improve the efficiency.

Syed Zishan Ali et al. [2016] In today's world there is rapid development in every field which contains bulk of data and different types of data. In order to distinguish sample data from the other data mining techniques are used in combination with many useful algorithms. Android development is one the major area where there is an immense needs to implement these algorithms. Combining frequent patterns algorithms with clustering is very effective for android. In this paper the work is done in two levels, initial stage focuses on implementing the k-means clustering algorithm and final stage calculate the frequent pattern in the clusters.

Ju Xiao et al. [2016] After studying existing improved Apriori algorithms, this paper rebuilds the transactional database. Then, with providing the definition of transactional granule, this paper sets forth the relationship between transactions and items from the view of granule. By giving the rules of synthesizing these transactional granules, this paper designs an improved Apriorialgorithm based on transactional granule, thus effectively applying the theory of granular computing to the mining of association rules.
XMART is a retail company that has sold more than 5,500 products. The company intends to increase sales of products with a promotion. XMART has a history of sales up to more than 10,000,000 records. The data can then be processed with the technique used association rules by applying a priori algorithm to obtain the value of the support and confidence of every association rules. The association rules adapted to the sales potential for each store location store has sales characteristics are manifold.