ISSN 2319-2720

Volume 1, No.1, July – August 2012
International Journal of Computing, Communications and Networking
Available Online at http://warse.org/pdfs/ijccn03112012.pdf

A Competent Technique on Cluster Based Master Slave Structural Design

Ms. Neha Saxena[1], Mr. Niket Bhargava[2,] Ms. Urmila Mahor[3]
[1] M.Tech(CSE), B.I.S.T., Bhopal, India, saxenaneha07@gmail.com
[2] Asst. Professor, CSE Dept.,B.I.S.T., Bhopal, India, niket.bhargava@rediffmail.com

ABSTRACT

We are in an age repeatedly referred to the information age. Information is gathered almost everywhere in our everyday lives. The amount of information stored in modern databases makes manual analysis intractable. Data mining provides tools to reveal previously unknown information in large databases. A number of algorithms have been proposed to determine frequent pattern. Apriori algorithm is the first algorithm proposed in this field. With the time a number of changes proposed in Apriori to enhance the performance in term of time and number of database passes. A well-known data mining technique is association rule mining. Association rule mining is a well-known technique in data mining. It is able to reveal all interesting relationships, called associations, in a potentially large database. However, how interesting a rule is depends on the problem a user wants to solve.This paper describes new technique cluster based Master-Slave structural. It uses Hybrid technique, the grouping of bottom up and top down technique for search repetitive item sets. It reduces the time in use to get out the support count of the item sets. The Prime number show present offers the give for verify rules and provides decrease in the data complexity.

Keywords: Apriori Algorithms, Association Rules Mining, Master Slave Structural, Prime Number

1. INTRODUCTION

Data mining refers to extract or "mining", "knowledge from large amounts of data". Data mining should have been more suitably named knowledge mining from data. [1]There are many other terms carrying a similar or slightly different meaning to data mining, such as knowledge mining from databases, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. Data mining is a widely-used technique for discovering information from large databases, within which 'association rules discovery', is one of the most popular technologies. It was first formulated in 1993 by Agrawal et al. [1]. A year later, Apriori, one of the most noticeable algorithms, was proposed by the same authors.

Over the next few years, studies on improvements or extensions of Apriori have been extensive. Many modifications to Apriori have been done. Cluster-based technique creates cluster tables by scanning the database once in a time. [2].The algorithm's support count is performed on the cluster table and it need not scan all transactions stored in the cluster table.

2. PROBLEM DEFINITION AND RELATED WORK

Data mining normally involves four classes of task classification, clustering, regression, and association rule learning. Data mining is use to determine knowledge in large quantity of information. Now a days, Association rule serve a very important role. Data mining association rule is most of the current data mining researches. Association rules are used to prove the connection between data items. Association rules are repeatedly used in marketing, publicity and inventory control .Association rules identify regular usage of records.[2].This problem is motivated by applications known as market basket analysis to find interaction between items purchase by customers [4], that is, what kinds of products tend to be purchased jointly.

This algorithm finds the frequent itemsets or patterns by partitioning the database transactions into clusters. Clusters are designed based on the comparison measurement among the communication. Then after find out the frequent itemsets. With the transactions in the clusters and items openly via improved Apriori algorithm which supplementary reduces the number of amount scans in the database and that's why improve the proficiency. When buy the product when another product is buy presents to the association rules. [7] The Apriori algorithm perform only find out the frequent itemsets. The Apriori algorithm is the base for association rules.

These technique have been planned to consider and recognize the executive implications of a spectrum of trade-offs between calculation, conversation, usage of memory,

synchronizing, and the use of particular problem-specific information data mining in parallel [3].

The problem of mining association rules could be decomposed into two sub problems:

- Find out all large itemsets or patterns and their support counts. A large itemset or patterns is a set of items which are contained in a sufficiently large number of transactions, with respect to a support threshold minimum support.
- The set of greatest itemsets or patterns catch, find it all there association rules that have a confidence value beyond a confidence threshold value lowest.

Since the solution of the second sub problem is basic, here we are concentrating only on the first sub problem.

Clustering is a data mining technique used to position data elements into related groups without advance knowledge of the group definitions. Clustering can be consider the most essential unsupervised learning problem; so, as every further problem of this kind, it deals with finding a structure in a collection of unlabeled data. A cluster is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters.

This example demonstrates the clustering of groups of small balls,

We are interested in grouping the balls into clusters by different values as shown in Figure 1

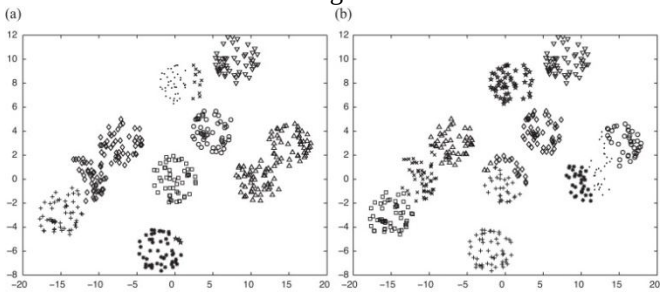

Figure 1: Clustering in groups

The partition algorithm is based in the study that the frequent sets are normally very few in number compared to the set of all item sets. By using partitioning, cluster based and/or distributed technique can be easily created, where each partitioning might be handle by a separate machine In order to find the Largest item set it is enough to go through the transactions with in the clusters.[6] The cluster that has the full amount of transactions smaller than some threshold cost will be deleted. For finding the greatest item sets it is sufficient to go through the transactions with the clusters. There is no need to go through the entire database again. Hence it compresses the unnecessary database scan and better the efficiency. If we apply the master slave structural after portioning it will further reduces the number of scans and improves efficiency [7].

3. MASTER SLAVE STRUCTURAL DESIGN

This technique is based on Cluster based Master-Slave structural design using candidate division technique. Candidate division technique reduces the communication overhead between master and slave nodes. Candidate division technique assigns the candidate item sets generated from different parts of database to different processors and each processor is assigned disjoint candidates, independent of further processors.[8] The Master node prunes the transactions by removing 1-infrequent itemsets and stores the Prime number multiple for each transaction in mutual memory. It finds the maximal length transaction size Maxlength and puts in mutual memory. It divides the transactions equally to each node for candidate generation. Though horizontal partition, Vertical partition techniques can be used to divide a division the transactions, horizontal partitioning technique is adopted, as it demands minimum communication. [4]If there are S number of slaves and T number of transactions, then T/S number of transactions are assigned to each slave if T is a integral multiple of S. Otherwise, S-1 slaves will be assigned T/S transactions and S^{th} slave will be assigned (T/S + mod (T/S)) transactions. Master connects to each slave node and initiates the process of finding the frequent itemset. Finally, the master node shows the global frequent itemsets after gathering the local frequent itemsets. After the Master node initiates the slave node, it reads the allotted number of transactions and Maximal length transaction size Maxlength[9]

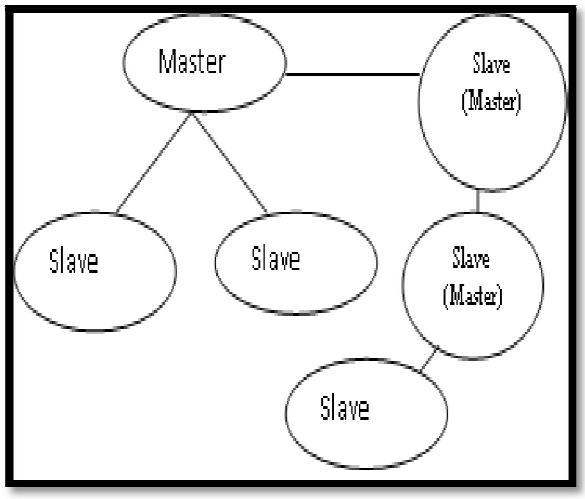

Figure 2: Master Slave structural design

4. PROPOSED TECHNIQUE

A new approach for Frequent Item sets using cluster based master slave structural design efficient approach is being proposed. This technique integrates both the bottom up explore as well as the top-down explore. The top down technique used to find the frequent subsets of itemsets and this bottom up technique to find the supersets of the frequent itemsets. Most of the technique for mining the frequent items is based on bottom-up search technique. In this approach, the technique starts from 1 itemsets and extends one stage in each pass until all maximum frequent item sets are found. These techniques perform fit if the length of the maximum itemset is short. If the maximum itemset is longer, top down technique is suitable for a transaction with a normal sized assigned a exclusive prime number as shown in table 1 maximum frequent set; a combination of both these techniques performs fit. This technique adopts Candidate partition technique to allocate the candidates among all nodes. The Prime number illustration reduces the memory needed for store the items of the transactions by conveying an exclusive prime number for each item. This technique uses prime numbers correspond to the items in this transaction. Each item is specify a exclusive prime number. Each Transaction of data base is the product of the corresponding prime number of specific items in the transaction, while the product of the prime number is exclusive, modulo division operation of a transaction's prime product. The prime product result of the item set can assure the occurrence or nonexistence of the item set in any transaction.

- If the residue is zero, then the item set is event in the transaction.
- If the residue is Non-zero, then the item set is nonexistence in the transaction

S.No	Items	Equivalent Prime
1	A	2
2	B	3
3	C	5
4	D	7
5	E	11
6	F	13
7	G	17
8	H	19
9	I	23
1`0	J	29
11	K	31
12	L	37
13	M	41

Table 1: Database Example Equivalent Prime Number

Step 1: Consider a example of database and equivalent prime number as shown in Table 1.

Step 2: In this technique each item is assigned a unique prime number and product of the corresponding prime numbers of the item set shown in table 2.

Table 2: Assign Prime Number and Its Multiplication

S. No	Transaction id	Transaction	Transaction Multiple	Result
1	I1	ABGH	2*3*17*19	1938
2	I2	ABC	2*3*5	30
3	I3	EFGHKL	11*13*17*19*31*37	1642342273
4	I4	ACDE	2*5*7*11	770
5	I5	BHIM	3*19*23*41	53751
6	I6	AJKL	2*29*31*37	66526
7	I7	ADEG	2*7*11*17	2618
8	I8	ADEGH	2*7*11*17*19	49742
9	I9	CEK	5*11*31	1705
10	I10	ADE	2*7*11	154
11	I11	AB	2*3	6
12	I12	ADEFG	2*7*11*13*17	34034

Step 3: To find the support count in given data base using modulo division operation, support count of item set {D(7), E(11)} is that 77 mode,

Table 3: Find Out Support Count

S.NO	T id	Modulo Division	Remainder	Item's Presence
1	I1	1938 mod 77	Non-zero	No
2	I2	30 mod 77	Non-zero	No
3	I3	1642342273	Non-zero	No
4	I4	770 mod 77	zero	Yes
5	I5	53751 mod	Non-zero	No
6	I6	66526 mod	Non-zero	No
7	I7	2618 mod 77	zero	Yes
8	I8	49742 mod	zero	Yes
9	I9	1705 mod 77	Non-zero	No
10	I10	154 mod 77	zero	Yes
11	I11	6 mod 77	Non-zero	No
12	I12	34034 mod	zero	Yes

Table 4: Experimental Result With Time

S. No	Transaction	Time in second		
		APriori	Clustering Technique	Master Slave
1	200	26	20	17
2	400	46	41	25
3	600	64	55	33
4	800	92	82	52
5	1000	100	87	70
7	1200	123	109	80
8	1400	143	127	109
9	1600	162	144	116
10	1800	182	162	156
11	2000	201	179	164

5. EXPERIMENTATION AND PERFORMANCE EVOLUTION

We have implemented the proposed approach using cluster based master slave structural design. The results of this technique are given in this style. There are two conditions occurs, are as follows:

If the support count is larger than or equal to the support count(minimum), it is treat as the maximum frequent item set and the process terminated.

If the support count for the itemsets of time slave A is smaller than the support count(minimum), the subsets of time-slave equal to P=A/2 is produce and their support count is recognize .

If the modulo division operation gives a zero residue, it indicates that the items or item pattern is in the transaction. If the residue is non-zero, it indicates that the items or item pattern is not there in the transaction. In this representation, support count of item set {D,E} can be found by exection the modulo division of each transaction's prime number by the product '77' of item 'D's equivalent prime number '7' and item 'E's equivalent prime number '11' as shown in table 1 The support count of item set {D,E} is found to be '5' as the modulo division operation of the four transactions with 77 gives zero residue and the modulo operation with other transactions resulted in a Non-zero residue.

The study of the proposed technique cluster based master slave structural design we have taken a database of 2000 transaction of 13 items. In this systematic process we considered 2000 transactions to generate the frequent itemset pattern with the support count 25% .We have repetitive the similar process by increasing the transaction, after the testing on both technique, so find the finalized this result, we have designed a graph and summarized a result in the following Table 4 .

Show the comparisons graph and result with time & transaction Figure 3.

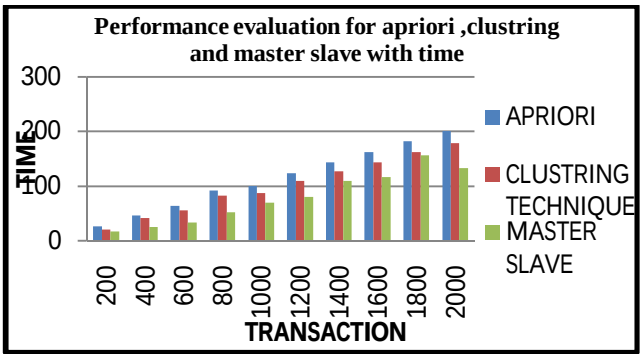

Figure 3: Performance evolution with time & transaction

For the comparative study of clustering and cluster based master slave structure design, we have taken a database of 1000 transaction of 13 items. In this systematic process we considered 1000 transactions to generate the All frequent itemset and pattern with the support count 25%.shown in the Table 5 and Figure 4.

Table 5: Experimental Result with all Frequent Itemset

S.No	Transaction	All Frequent Itemsets	
		Clustering Technique	Master Slave
1	200	26	20
2	400	21	14
3	600	20	18
4	800	22	25
5	1000	25	29

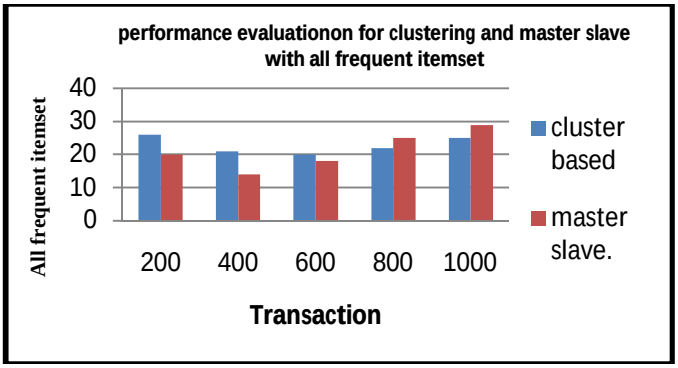

Figure 4: Performance evolution with Transaction & all frequent itemset

6. CONCLUSION

In this paper, a new Approach for cluster based master slave structural design was proposed. The innovative Prime number representation stores only one number for each transaction, it may need less memory. The computational complexity is reduced as the product of their assigned prime numbers represents each candidate item set. The pruning of infrequent items in the first scan reduces the size of the dataset in the main memory. Large transaction database including a great number of item set will generate. The use of traditional technique will be difficult to meet the new demand for data mining, so the new data mining algorithm proposed in this paper is meaningful. This paper increase data mining efficiency significantly. This master - and-slave technique can solve the algorithm space problem,

REFERENCES

1. R. Agrawal and R. Srikant. **Fast Algorithms for Mining Association Rules in Large Database,** Proc. 20th Int'l Conf. Very Large Databases (VLDB 94), Morgan Kaufmann, 1994,pp. 407-419.

2. R. Agrawal and J.C. Shafer. **Parallel Mining of Association Rules,** *IEEE Tran. Knowledge and 16 IEEE* Distributed Systems Online March 2004 Data Eng. , *V*ol. 8, No. 6, 1996, pp. 962-969.

3. Dr.Sujni Paul. A**n optimized Distributed Association Rule Mining algorithms in Parallel and Distributed Data Mining with XML data for improved response time,** International Journal of Computer Science and Information Technology, Volume 2, No. 2, April 2010.

4. Dr.Dev Prakash Agrawal. C**luster based partition algorithms for mining frequent itemset,** IJCSNS International Journal of Computer Science and Network Security, Vol.9 No.6, June 2009.

5. C. Raissi, P. Poncelet and Teisseire. **Towards a new approach for mining frequent itemsets on data stream,** Journ. Intell Inf Syst , 2007, Vol. 28, pp.23–36, 2007.

6. J,.Xu Yu , Z.Chong , H. Lu, Z. Zhang and A. Zhoub. **A false negative approach to mining frequent itemsets from high speed transactional data streams,** Information Sciences 176, 1986–2015,2006.

7. ZENG Wandan, ZHOU Xubo and DAI Bo. **An Association Mining Algorithm Based on Matrix,** Journal of Computer Engineering., Vol.32, No.2, pp.45-47,2006.

8. NI Zhiwei and GAO Yazhuo. **Incremental updating algorithm based on matrix for mining association rules,** Journal of Computer Application.

9. LIU Yian and YANG Bin. **Research of an improved Apriori algorithm in mining association rules**, Journal of Computer Applications, Vol.27, No.2, pp.418-420, 2007.

10. Sheng Chai, Jia Yang and Yang Cheng. **The Research of Improved Apriori Algorithm for Mining Association Rules,** Proceedings of the Service Systems and Service Management, 2007 International Conference, 9-11 June 2007 pp. 1 – 4.

11. Margatet H. Dunham. **Data Mining, Introductory and Advanced Topics**: **Upper Saddle Rive**r, New Jersey: Pearson Education Inc.,2003

12. Jiawei Han. **Data Mining, concepts and Techniques,** San Francisco, CA: Morgan Kaufmann Publishers.,2004.

13. M.J. Zaki and Y. Pin. **Introduction: Recent Developments in Parallel and Distributed Data Mining,** Journ. Distributed and Parallel Databases , Vol. 11, No. 2, 2002, pp. 123-127.[14].J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation, Proc. ACM SIGMOD Int'l. Conf. Management of Data, ACM Press, 2000, pp. 1-12.

14. Jiawei Han. **Data Mining, concepts and Techniques**: SanFrancisco, CA:Morgan Kaufmann Publishers. ,2004

16. A. Savasere , E. Omiecinski, and S.B. Navathe. **An Efficient Algorithm for Mining Association Rules in Large Databases,** Proc. 21st Int'l Conf. Very Large Databases (VLDB 94), Morgan Kaufmann, pp.432-444, 1995

17. Rakesh Agrawal and R. Srikant. **Fast Algorithm for Mining Association Rules in Large Databases,** Proceedings of the 20th International Conference on Very Large Databases, Santigo, Chile, pp. 487-499.