Scientific Journal Impact Factor (SJIF): 1.711

e-ISSN: 2349-9745
p-ISSN: 2393-8161

International Journal of Modern Trends in Engineering and Research
www.ijmter.com

REVIEW: Frequent Pattern Mining Techniques

Parag Moteria[1], Dr. Y. R. Ghodasara[2]
[1]*PhD scholar, School of Computer Science, RK University,Rajkot and Assistant Professor, ISTAR, MCA Department, Vallabh Vidyanagar, Gujarat*
[2]*Agricultural Information Technology, Anand Agricultural University, Anand, Gujarat*

Abstract – Frequent pattern mining techniques helpful to find interesting trends or patterns in massive data. Prior domain knowledge leads to decide appropriate minimum support threshold. This review article show different frequent pattern mining techniques based on apriori or FP-tree or user define techniques under different computing environments like parallel, distributed or available data mining tools, those helpful to determine interesting frequent patterns/itemsets with or without prior domain knowledge. Proposed review article helps to develop efficient and scalable frequent pattern mining techniques.

Keywords – Apriori algorithm, FP-tree Algorithm, Frequent itemsets, Hadoop MapReduce framework, Heterogeneous platforms, Parallel and Distributed mining algorithm, Without support threshold

I. INTRODUCTION

Data mining is a very basic operational technique in knowledge discovery and decision making processes. Data mining is the process of finding interesting trends or patterns in large datasets to steer decision about future activities. Knowledge discovery in databases and data mining helps to extract useful information from raw data. Frequent itemsets play an essential role in many data mining tasks that try to find interesting patterns from databases or transactional dataset, such as association rules, correlations, sequences, episodes, classifiers, clusters. Frequent pattern mining is one of the most important and well researched techniques of data mining. Frequent pattern mining techniques have become necessary for massive amount datasets in different computing environments. This review article presents different frequent pattern mining techniques those helpful to make this process more efficient and scalable with different approaches. Using these different techniques, we analyze reduction in scanning of transactional datasets or reduce candidate generation overhead using different computing environments.

II. LITERATURE REVIEW

A. History

[1][2]Frequent pattern mining was first proposed by Agrawal et al. (1993) for market basket analysis in the form of association rule mining. It analyses customer buying habits by finding associations between the different items that customers place in their "shopping baskets". For instance, if customers are buying milk, how likely are they going to also buy cereal (and what kind of cereal) on the same trip to the supermarket? Such information can lead to increased sales by helping retailers do selective marketing and arrange their shelf space.

B. Literature Review

- [3]Mining Frequent Itemsets without Support Threshold: With and without Item Constraints (June, 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. In this paper, author handled a more practical problem; roughly speaking, it was to

mine N k-itemsets with the highest supports for k up to a certain k_{max} value. Author called the results the N-most interesting itemsets. Generally, it was more straightforward for users to determine N and k_{max}. This paper proposed two new algorithms, LOOPBACK and BOMO. Experiments explained that proposed methods outperform the previously proposed Itemset-Loop algorithm, and constraint-based itemsets mining with BOMO performed better than the original FP-tree algorithm, even with the assumption of an optimally chosen support threshold. Author also proposed the mining of "N-most interesting k-itemsets with item constraints." This allowed user to specify different degrees of interestingness for different itemsets. Experiments explained that proposed Double FP-trees algorithm, which was based on BOMO, is highly efficient in solving this problem.

Several sets of synthetic data were generated from the synthetic data generator by the author and mentioned as (Table 1 and Table 2).

Table 1. Parameter setting

Parameter	Description	Value
|D|	Number of transactions	100K, 1000K
|T|	Average size of the transactions	5,10, 20
|I|	Average size of the maximal potentially large itemsets	2, 4, 6, 8, 10
|L|	Number of maximal potentially large itemsets	2000, 10000
M	Number of items	1k, 50k
C	Correlation between patterns	0.25

Table 2. Synthetic Data Description

Dataset	|T|	|I|	|D|	|M|
T5.I2.D100K	5	2	100K	1K
T20.I6.D100K	20	6	100K	1K
T20.I8.D100K	20	8	100K	1K
T20.I10.D100K	20	10	100K	1K
T10.I4.D1M	10	4	1000K	50K

- [4]Frequent Pattern Mining in Web Log Data (2006)

  Frequent pattern mining is a heavily researched area in the field of data mining with wide range of applications. One of them is to use frequent pattern discovery methods in Web log data. Discovering hidden information from Web log data is called Web usage mining. Patterns in Web usage mining like Page sets, page sequences and page graphs helped to identify frequent pattern navigational behavior of the web users from large amount of data collected by Web servers and generated information for advertising purposes, for creating dynamic user profiles and many more.

  Three types of log files used for Web usage mining. Log files are stored on the server side, on the client side and on the proxy servers. Web usage mining system was able to use all three frequent pattern discoveries task mentioned as (Fig. 1).

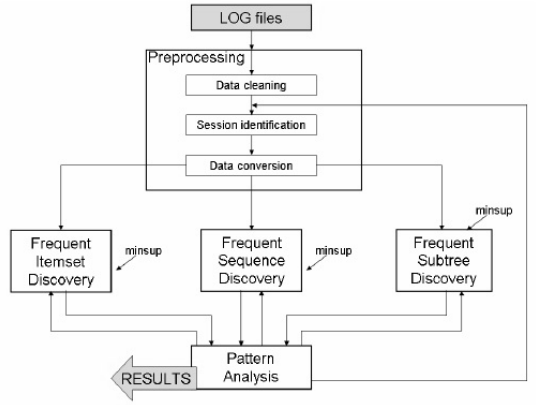

Figure 1. Process of web usage mining

- [5]Distributed Frequent Itemsets Mining in Heterogeneous Platforms (2007)

  Huge amounts of datasets with different sizes are naturally distributed over the network. Author proposed a distributed algorithm for frequent itemsets generation on heterogeneous clusters and grid environments. In addition to the disparity in the performance and the workload capacity in these environments, other constraints were related to the datasets distribution and their nature, and the middleware structure and overheads. The proposed approach uses a dynamic workload management through a block-based partitioning, and takes into account inherent characteristics of the Apriori algorithm related to the candidate sets generation. This approach was evaluated on large scale datasets distributed over a heterogeneous cluster. The block-based approach focused on memory constraints since the basic task may need very large memory space depending on several parameters including the support threshold, and information about the dataset. Author developed an inherent property of the itemsets generation task that explained intermediate communication steps, in classical implementations such as the FDM approach, were performance constraining. Indeed, global pruning strategies did not bring enough useful information in comparison to the generated synchronization and I/O overheads. Furthermore, the workload management strategy attempted the imbalanced workloads of the platform heterogeneity or uneven dataset distribution. Experiments had been conducted on heterogeneous platforms and explained that the proposed algorithm achieved very good performance and high scalability compared to a classical Apriori-based implementation.

  Author implemented experimental on Condor and DAGMan systems. The Condor system provided job management capabilities for the grid through Condor-G (using Globus Toolkit). DAGMan used for directed acyclic graph representation manager, which allows the user to express dependencies between Condor jobs. Author clustered of heterogeneous workstations connected by a Fast Ethernet network to perform experiments. A synthetic dataset and a census dataset (the PUMS dataset available from the UC Irvine KDD Archive) were used. The datasets size was 0.5×10^6 transactions, with average transaction size of 10 for the synthetic dataset, and 30 for the census dataset.

- [6]Parallel and Distributed Mining of Association Rule on Knowledge Grid (June, 2008)

  In Virtual organization, Knowledge Discovery (KD) service contains distributed data resources and computing grid nodes. Computational grid was integrated with data grid to form Knowledge Grid, which implemented Apriori algorithm for mining association rule on grid network. Author described development of parallel and distributed version of Apriori algorithm on Globus Toolkit using Message Passing Interface extended with Grid Services (MPICH-G2). The creation of Knowledge Grid on top of data and computational grid was support for decision making in real time applications. In this paper, the case

study described design and implementation of local and global mining of frequent item sets. The experiments were conducted on different configurations of grid network and computation time was recorded for each operation. Both grid technology and parallel mining algorithm reduced the computational time and increase the speed of the application. In all iteration, the Message Passing Interface extended with Grid services (MPICH-G2) supported communication of frequent mobility patterns between the grid nodes in different clusters. The experimental result described that parallel and distributed version of Apriori algorithm is optimal than distributed algorithm. It performed scalable in terms of the database size and the number of nodes. Grid technology was used as a platform for implementing and deploying geographically distributed knowledge and knowledge management services and applications. The discovered knowledge can be used by the experts to provide various services to the mobile user in web environment. The Knowledge Grid (KG) was integrated with grid services system to support distributed data analysis, knowledge discovery and knowledge management services. Author analyzed result with various grid configurations and it derived speedup of almost super linear computation time.

Author designed grid with different configurations to measure the efficiency of parallel apriori algorithm. Test were conducted on standalone PC and two clusters of three nodes each and three clusters of three nodes each. Nodes within the cluster were connected by LAN link and clusters are connected by WAN link. Each node was installed with the Globus 3 toolkit and deployed with the apriori grid service. Mobile users logs were stored types of data such as video, voice and image in three different database systems: Oracle 10g, PostGreSQL and MySQL.

- [7]Parallel and Distributed Frequent Pattern Mining in Large Databases (June, 2009)

  Recently, a significant number of parallel and distributed algorithms have been proposed to mine frequent patterns (FP) from large and/or distributed databases. Among them parallelization of the FP-growth algorithms using the FP-tree has been proved to be highly efficient. However, the FP-tree-based techniques suffer from two major limitations such as multiple database scans requirement (i.e., high I/O cost) and high interprocessor communications cost (during the mining phase). Therefore, author proposed a novel tree structure, called PP-tree (Parallel Pattern tree) that significantly reduced the I/O cost by capturing the database contents with a single scan and facilitates the efficient FP-growth mining on it with reduced inter-processor communication overhead. Proposed parallel algorithm worked independently at each local site and locally generates global frequent patterns which merged at the final stage. The experimental results reflect that parallel and distributed FP mining with PP-tree outperforms other state-of-the-art algorithms.

  T10I4D100K was synthetic dataset, developed by the IBM Almaden Quest research group and obtained from http://cvs.buu.ac.th/mining/Datasets/synthesis_data/. The other datasets were real and have been obtained from the UCI Machine Learning Repository (University of California – Irvine, CA). Among all the datasets, connect was dense and others were sparse. Datasets characteristics mentioned as (Table 3).

Table 3. Dataset characteristic

	Datasets		
	T10I4D100K	connect	kosarak
Transactions	100000	67557	990002
Items	870	129	41270
Max Transaction Length	29	43	2498
Average Transaction Length	10.10	43.00	8.10
Type	Sparse	Dense	Sparse

- [8]Robust and Distributed Top-N Frequent-Pattern Mining With SAP BW Accelerator (August, 2009)

  Mining for association rules and frequent patterns is a central activity in data mining. Real-world datasets are distributed and modern database architectures are switching from expensive SMPs to cheaper shared-nothing blade servers. Thus, most mining queries require distribution handling. Since partitioning can be forced by user defined semantics, it is often forbidden to transform the data. Most strategies used parameters like minimum support, for which it could be very difficult to define a suitable value for unknown datasets. Since most untrained users unable to set such technical parameters, Author addressed the problem of replacing the minimum support parameter with top-n strategies. Author implemented ECLAT algorithm to improve its performance by using heuristic search strategy for top-n strategies. Author developed an adaptive top-n frequent-pattern mining algorithm that simplified the mining process on real distributions by relaxing some requirements on the results. In the first step, this proposed work combined the PARTITION and the TPUT algorithms to handle distributed top-n frequent-pattern mining. Then, extend this proposed algorithm for distributions with real-world data characteristics. Author divided real world data into equal distribution for frequent pattern mining algorithms because of each tiny partition caused performance bottlenecks. Minimum absolute support threshold defined by MAST approach. MAST approach pruned patterns with low chances of reaching the global top-n result set with high computing costs. In this manner, author simplified the process of frequent-pattern mining for real customer scenarios and data sets. This method made frequent pattern mining accessible for every new user groups. Author presented results of new proposed algorithm implemented on the SAP Net Weaver BW Accelerator with standard and real business datasets and drawn evaluation of top-n mining impacted by numbers of partitions.

  Flow of TPARTITION algorithm using STH algorithm mentioned as (Fig. 2). Note that STH is only responsible for half of the runtime.

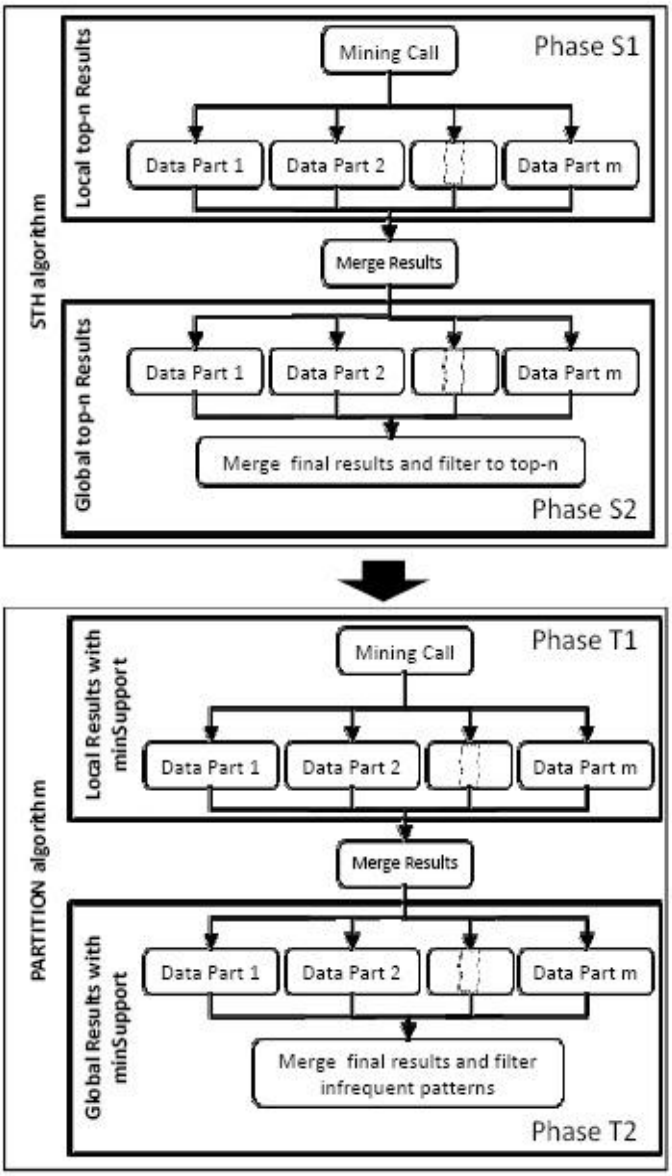

Figure 2. Distributed top-n frequent pattern mining with TPARTITION

Author used well known artificial and real world dataset for experiments mentioned as (Table 4).

Table 4. Dataset characteristics

	Name of Datasets				
	SynthA	SynthB	Retail	CustomerA	CustomerB
Type	Artificial/open	Artificial/open	Real/open	Real/closed	Real/closed
Partitions	1, 2, 4, 6, 8, 10,14	1	1, 2	1, 2, 4, 10, 22	40
Transactions	800000	980000	85146	134167	33542000
Average Transaction Length	19.9	10.2	9.6	3.0	3.3
Distinct Items	772	24000	16398	72252	72025

- [9]Using Distributed Apriori Association Rule And Classical Apriori Mining Algorithms For Grid Based Knowledge Discovery (July, 2010)

  The aim of this paper was to extract knowledge using predictive apriori and distributed grid based apriori algorithms for association rule mining. The paper presented the implementation of an association rules discovery data mining task using Grid technologies. A result of implementation with a comparison of classic apriori and distributed apriori was also discussed. Distributed data mining systems provide an efficient use of multiple processors and databases to speed up the execution of data mining and enable data distribution. Efficiency of the proposed system evaluated on weka tool and performance analysis with apriori and predictive apriori algorithms on a centralized database. The main aim of grid computing was to give organizations and application developers the ability to create distributed computing environments that utilized computing resources on demand. Therefore, it helped increase efficiencies and reduce the cost of computing networks by decreasing data processing time and optimizing resources and distributing workloads, thereby allowing users to achieve much faster results on large operations and at lower costs. Author discussed distributed apriori association rule on grid based environment is mined and the knowledge obtained is interpreted.

  Author implemented grid architecture to achieve distributed data mining and mentioned as (Fig. 3).

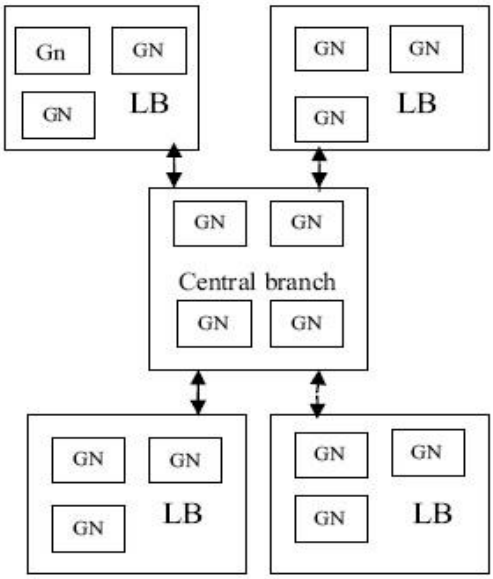

Figure 3. Virtual organization infrastructure using grid technologies

  Author used dataset with form ‘‘TxxIyyDzzzK’’, where ‘‘xx’’ denotes the average number of items present per transaction, ‘‘yy’’ denotes the average support of each item in the dataset and ‘‘zzzK’’ denotes the total number of transactions in ‘‘K’’ (1000 s). The experiments were performed for 4 database sizes (3000, 7000, 10,000 and 50,000 transactions) and the resulting rules resulted had 30%, 25% and 20% support factors.

- [10]An Algorithm for Frequent Pattern Mining Based on Apriori (2010)

  Mining frequent patterns from large scale databases has emerged as an important problem in data mining and knowledge discovery community. A number of algorithms have been proposed to determine frequent pattern. Apriori algorithm is the first algorithm proposed in this field. Three different frequent pattern mining approaches, named, Record filter, Intersection and Proposed Algorithm were discussed based on classical Apriori algorithm. Author declared that, Record filter approach proved better than classical Apriori

Algorithm, Intersection approach proved better than Record filter approach and finally proposed algorithm proved much better than Intersection. Author tested dataset of two thousands transactions of fifty items for comparative study. Author proved that proposed algorithms took less time than that of classical apriori algorithms.

- [11]Frequent Pattern Mining Using Record Filter Approach (July, 2010)

  In today's emerging world, the role of data mining is increasing day by day with the new aspect of business. Data mining has been proved as a very basic tool in knowledge discovery and decision making process. Data mining technologies are very frequently used in a variety of applications. Frequent patterns were the itemsets those frequently visited in database transactions at least for the user defined number of times which known as support threshold. Presently a number of algorithms had been proposed in literature to enhance the performance of Apriori Algorithm, for the purpose of determining the frequent pattern. The main issue for any algorithm was to reduce the processing time. Author proposed a new record filter based algorithm which was a variation of the Apriori algorithm and performed fewer database scans than Apriori and utilizes only transaction of specific sizes for the generation of frequent itemsets. As observed by many researchers counting the occurrences of itemsets is a time consuming activity, this paper introduced a new strategy of considering only those transactions whose length was greater than or equal to the length of candidate set was checked, because candidate set of length k , cannot exist in the transaction record of length k-1 , it might exist only in the transaction of length greater than or equal to k. Due to this, proposed approach took very less time for performing computations during mining process. Experiments performed on synthetic datasets. The results explained that proposed approach performd well in terms of execution time and ultimately enhance efficiency as compared to traditional Apriori approach.

  For the comparative study of classical Apriori and proposed approach, author considered a database of 5000 transactions containing 50 unique items. During this analytical process author considered 1000 transactions to generate the frequent pattern with the support count of 10% and the process was repeated by increasing the transaction gradually.

- [12]A Parallel, Distributed Algorithm for Relational Frequent Pattern Discovery from Very Large Data Sets (January, 2011)

  Heterogeneity and strong interdependence, which characterize ubiquitous data, required a multi relational approach to be analyzed with WARMR and SPADA. However, relational data mining algorithms did not scale well. Author proposed an extension of a relational algorithm for multilevel frequent pattern discovery, which resorted to data sampling and distributed computation in Grid environments, in order to overcome the computational limits of the original serial algorithm. The set of patterns discovered by the proposed algorithm approximates the set of exact solutions found by the serial algorithm. The quality of approximation depended on three parameters: the proportion of data in each sample, the minimum support thresholds and the number of samples in which a pattern had to be frequent in order to be considered globally frequent. Author investigated on the third one.

  Experiments performed by processing both an event log publicly available on ProM web site http://is.tm.tue.nl/_cgunther/dev/prom/ and an event log provided by THINK3 Inc http://www.think3.com/en/default.aspx.

- [13]A Frame Work for Frequent Pattern Mining Using Dynamic Function (May, 2011)

  Discovering frequent objects (item sets, sequential patterns) is one of the most vital fields in data mining. Apriori algorithm is a standard algorithm of association rules mining. We presented a new research trend on frequent pattern mining in which generate Transaction pair, which provided scalability to massive data sets and improving response time. This

framework made pair of transaction instead of item id, so result show more scalable. Author suggested a novel dynamic algorithm for transposed database, mined in transaction pair and found longest common subsequence using dynamic function. Artificial and real-life data sets were tested and result described that proposed FPMDF algorithm was more scalable than Apriori and FP Growth algorithm.

Author performed experiment on T40I4D100K dataset, provided by the QUEST generator of data generated from IBM's Almaden lab.

- [14]Comparative Analysis of Various Approaches Used in Frequent Pattern Mining (August, 2011)

  Frequent pattern mining searched for recurring relationship in a given data set with association rules for interesting k itmesets. Various techniques found to mine frequent patterns with its own pros and cons. Performance of particular technique depended on input data and available resources in different domains like market basket analysis, including applications in marketing, customer segmentation, medicine, e-commerce, classification, clustering, web mining, bioinformatics and finance. This paper presented review of different frequent mining techniques including apriori based algorithms, partition based algorithms, DFS and hybrid algorithms, pattern based algorithms, SQL based algorithms and Incremental apriori based algorithms. Among all of the techniques discussed above, FP- Tree based approach achieved better performed and reduced the computational time. It took less memory by representing large database in compact tree-structure. But a word of caution here that association rules should not be used directly for prediction without further analysis or domain knowledge.

  Author used following real life dataset:

  Kosarak: The kosarak dataset comes from the click-stream data of a Hungarian online news portal, Number of Instances =990,002, Number of Attributes= 41,270.

  Mushroom: This data set included descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms. Each species was identified as definitely edible, definitely poisonous, or of unknown edibility and not recommended. This latter class was combined with the poisonous one. The Guide clearly stated that there was no simple rule for determining the edibility of a mushroom. Number of Instances = 8124, Number of Attributes = 22.

  Chess: A game datasets.

  Attribute Information: Classes (2)**:** White-can-win ("won") and White-cannot-win ("nowin"). Number of Instances= 3196, Number of Attributes=36.

- [15]Performance Analysis of Distributed Association Rule Mining with Apriori Algorithm (August, 2011)

  One of the most crucial problems in data mining is association rule mining. It required large computation and I/O traffic capacity. Author considered grid approach to resolve this problem. It offered an effective way to mine for large data sets. Therefore, author implemented distributed data mining with Apriori algorithm in grid environment. However, usage of grid environment raised some issues about the optimization of the Apriori algorithm, especially the cost of the node to node communication and data distribution. In this paper, an Optimized Distributed Association rule mining approach for geographically distributed data was introduced in parallel and distributed environment and analyzed that this proposed method reduced communication costs. Author implemented experiments on datasets having minimum one million transactions to maximum five million transactions.

- [16]Parallel and Distributed Closed Regular Pattern Mining in Large Databases (March, 2013)

Due to huge increase in the records and dimensions of available databases pattern mining in large databases is a challenging problem. Numbers of parallel and distributed FP mining algorithms have been proposed for large and distributed databases based on frequency of item set. Author introduced a novel method called PDCRP-method (Parallel and Distributed closed regular pattern) to discover closed regular patterns using vertical data format on large databases. Conversion of horizontal database to vertical database format needed one database scan. PDCRP method applied in parallel and distributed environment to mine complete set of closed regular patterns based on user given global regularity and support values which minimize I/O cost and worked at each local processor which reduces inter processor communication overhead and getting high degree of parallelism generates complete set of closed regular patterns. Author derived results from experiments, which described PDCRP method is highly efficient in large databases.

Author implemented PDCRP method from real (Kosarak) and synthetic (T1014D100K) datasets, available from http://cvs.buu.ac.th/mining/Datasets/synthesis_data/ and UCI Machine Learning Repository (University of California – Irvine, CA), these are used by Almanden Quest research group to develop frequent patterns in mining process.

- [17]Mining Efficient Association Rules Through Apriori Algorithm Using Attributes and Comparative Analysis of Various Association Rule Algorithm (June, 2013)

  Apriori is the classical and most famous algorithm. Author considered data (bank data) and tried to obtain the result using Weka a data mining tool. Three algorithms tested and got elapsed time by author, named, Apriori Association Rule, PredictiveApriori Association Rule and Tertius Association Rule. According to the result obtained using data mining tool, author declared that Apriori Association algorithm performs better than the PredictiveApriori Association Rule and Tertius Association Rule algorithms.

  Author implemented experiment on dataset containing six hundred records and eleven attributes.

- [18]Distributed Algorithm for Frequent Pattern Mining using Hadoop MapReduce Framework (2013)

  With the rapid growth of information technology and in many business applications, mining frequent patterns and finding associations among them requires handling large and distributed databases. As FP-tree considered being the best compact data structure to hold the data patterns in memory there has been efforts to make it parallel and distributed to handle large databases. However, it incurs lot of communication over head during the mining. Author proposed parallel and distributed frequent pattern mining algorithm using Hadoop Map Reduce framework, which helped to derive best performance results for large databases. Proposed algorithm partitioned the database in such a way that, it worked independently at each local node and locally generates the frequent patterns by sharing the global frequent pattern header table. These local frequent patterns merged at final stage. This reduced the complete communication overhead during structure construction as well as during pattern mining. The item set count was also taken into consideration reducing processor idle time. Author used Hadoop Map Reduce framework effectively in all the steps of the algorithm. Experiments were carried out on a PC cluster with five computing nodes which shows execution time efficiency as compared to other algorithms. The experimental result described that proposed algorithm efficiently handles the scalability for very large databases.

  Architecture diagram of DPFPM algorithm using MapReduce framework mentioned as (Fig. 4).

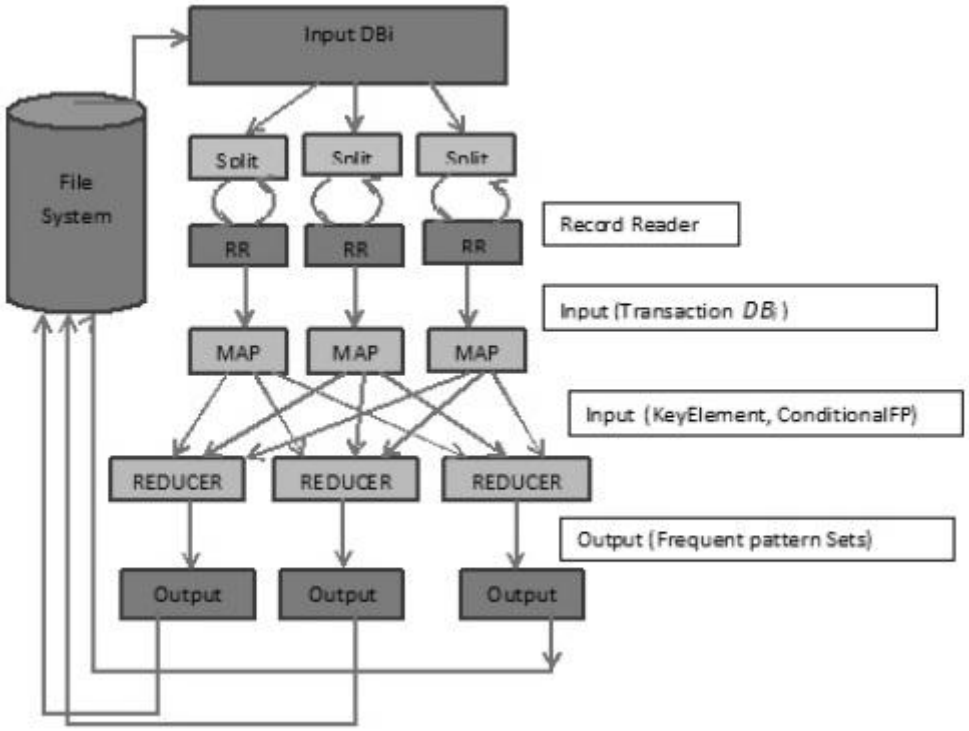

Figure 4. Architecture diagram of DPFPM algorithm using MapReduce framework

Author used Kosarak1G.dat dataset which contained 3,16,80,064 number of transactions and DPFPM algorithm performed on Hadoop Cluster of 5 Nodes (1 master, 4 slaves).

- [19]A complete Survey on Application of Frequent Pattern Mining and Association Rule Mining on Crime Pattern Mining (April, 2014)

  Author presented reviewed on Apriori, FP-Growth and ECLAT algorithms, pertaining to applications of frequent patterns mining and association rule mining in the field of crime pattern detection. It helped to understand about various frequent pattern mining algorithm and its extensions. Author covered different application areas other than legal field like Network Forensic Analysis, Network Cyber Attack, Animal Behavior Analysis, Educational Data, Digital Forensic, Socio-Economic Impact and Banking Sector, where various frequent pattern mining can be found to extract knowledge.

III. CONCLUSION

Each frequent pattern mining techniques have their own characteristics to enhance scalability and efficiency using different methodology like, without support threshold: with and without item constraints, heterogeneous platforms, parallel and distributed mining using grid technologies, record filter approach, dynamic function approach, and Hadoop MapReduce framework for different kinds of dataset like, web log data, large databases includes video, voice and image, and crime datasets. Frequent pattern mining techniques used either synthetic or real datasets for experiments.

REFERENCES

[1] Jiawei Han, Hong Cheng, Dong Xin, *et al*. Frequent pattern mining: current status and future directions. Data Mining and Knowledge Discovery. 2007; 15(1): 55–86. Available from: www.jaist.ac.jp/~bao/VIASM-SML/SMLreading/DirectionsofAssociationMining.pdf

[2] Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM-SIGMOD international conference on management of data (SIGMOD'93). 1993: 207–216. Available from: rakesh.agrawal-family.com/papers/sigmod93assoc.pdf

[3] Yin-Ling Cheung, Ada Wai-Chee Fu. Mining Frequent Itemsets without Support Threshold: With and without Item Constraints. IEEE Transactions on Knowledge and Data Engineering. 2004; 16(6): 1–18. Available from: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1316834&isnumber=29187

[4] Renata Ivancsy, Istvan Vajk. Frequent Pattern Mining in Web Log Data. Acta Polytechnica Hungarica. 77–90. Available from: http://www.uni-obuda.hu/journal/Ivancsy_Vajk_5.pdf

[5] Lamine M. Aouad, Nhien-An Le-Khac, Tahar M. Kechadi. Distributed Frequent Itemsets Mining in Heterogeneous Platforms. Journal of Engineering, Computing and Architecture. 2007; 1(2): 1–12. Available from: www.scientificjournals.org/journals2007/articles/1239.pdf

[6] U. Sakthi, R. Hemalatha, R.S.Bhuvaneswaran. Parallel and Distributed Mining of Association Rule on Knowledge Grid. International Scholary and Scientific Research & Innovation. 2008; 2(6): 292–296. Available from: waste.org/publications/11084/parallel-and-distributed-mining-of- association-rule-on-knowledge-grid

[7] Tanbeer, S.K., Ahmed, C.F., Byeong-Soo Jeong. Parallel and Distributed Frequent Pattern Mining in Large Databases. High Performance Computing and Communications, 2009. HPCC '09. 11th IEEE International Conference. 2009; 407(414): 25–27. Available from: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5167021&isnumber=5166954

[8] Thomas Legler, Wolfgang Lehner, Jan Schaffner. Robust and Distributed Top-N Frequent-Pattern Mining With SAP BW Accelerator. Proceeding of the VLDB Endowment. 2009; 2(2): 1438–1449. Available from: http://www.vldb.org/pvldb/2/vldb09-970.pdf

[9] Sumithra, R., Paul, S. Using Distributed Apriori Association Rule And Classical Apriori Mining Algorithms For Grid Based Knowledge Discovery. Computing Communication and Networking Technologies (ICCCNT) International Conference. 2010; 1(5): 29–31. Available from: http://ieeexplore.ieee.org/stamp/ stamp.jsp?tp=&arnumber=5591577&isnumber=5591555

[10] Goswami D.N., Chaturvedi Anshu, Raghuvanshi C.S.. An Algorithm for Frequent Pattern Mining Based on Apriori. International Journal on Computer Science and Engineering. 2010; 4(2): 942–947. Available from: http://www.enggjournals.com/ijcse/doc/IJCSE10-02-04-16.pdf

[11] D.N. Goswami, Anshu Chaturvedi, C.S. Raghuvanshi. Frequent Pattern Mining Using Record Filter Approach. International Journal of Computer Science Issues. 2010; 7(4): 38–43. Available from: ijcsi.org/papers/7-4-7-38-43.pdf

[12] Annalisa Appice, Michelangelo Ceci, Antonio Turi, et al. A Parallel, Distributed Algorithm for Relational Frequent Pattern Discovery from Very Large Data Sets. Intelligent Data Analysis – Ubiquitous Knowledge Discovery. 2011; 15(1): 69–88. Available from: http://www.di.uniba.it/~ceci/micFiles/papers/IDA.pdf

[13] Sunil Joshi, R S Jadon, R C Jain. A Frame Work for Frequent Pattern Mining Using Dynamic Function. International Journal of Computer Science. 2011; 8(3): 141–147. Available from: www.IJCSI.org

[14] Deepak Garg, Hemant Sharma. Comparative Analysis of Various Approaches Used in Frequent Pattern Mining. International Journal of Advanced Computer Science and Applications, Special Issue on Artificial Intelligence. 141–147. Available from: http://thesai.org/Downloads/SpecialIssueNo3/Paper%2023-Comparative%20Analysis%20of%20Various%20Approaches%20Used%20in%20Frequent%20Pattern%20Mining.pdf

[15] M.A.Mottalib, Kazi Shamsul, Mohmmad Majharul Islam, et al. Performance Analysis of Distributed Association Rule Mining with Apriori Algorithm. International Journal of Computer Theory and Engineering. 2011; 3(4): 484–488. Available from: www.ijcte.org/papers/354-G475.pdf

[16] M. Sreedevi, L.S.S.Reddy. Parallel and Distributed Closed Regular Pattern Mining in Large Databases. IJCSI International Journal of Computer Science Issues. 2013; 10(2): 264–269. Available from: http://ijcsi.org/papers/IJCSI-10-2-2-264-269.pdf

[17] Ms Sweta et al. Mining Efficient Association Rules Through Apriori Algorithm Using Attributes and Comparative Analysis of Various Association Rule Algorithm. International Journal of Advanced Research in Computer Science and Software Engineering. 2013; 3(6): 306–312. Available from: http://www.ijarcsse.com

[18] Suhasini A. Itkar, Uday Kulkarni. Distributed Algorithm for Frequent Pattern Mining using HadoopMap Reduce Framework. Association of Computer Electronics and Electrical Engineers. 2013; 15–24. Available from: searchdl.org/index.php/conference/view/742

[19] D. Usah, Dr. K. Rameshkumar. A complete Survey on Application of Frequent Pattern Mining and Association Rule Mining on Crime Pattern Mining. International Journal of Advances Computer Science and Technology. 2014; 3(2): 264–275. Available from: http://warse.org/pdfs/2014/ijacst05342014.pdf

