Association Rule Mining using Improved Apriori Algorithm

Minal G. Ingle
Student
Computer Department,
SSBT's College of Engg. and Tech., Jalgaon,
Maharashtra, India-425001.

N. Y. Suryavanshi
Assistant Professor,
Computer Department,
SSBT's College of Engg. and Tech., Jalgaon,
Maharashtra, India-425001.

ABSTRACT
The past few years have seen a marvellous curiosity in area of data mining. Data mining is usually thought of as the process of finding hidden, non trivial and formerly unknown information in large collection of data. Association rule mining is an significant component of data mining. Association rule are an important class of methods of finding regularities or patterns in data. Association rule mining has been used in several application domains. The ideal application of Association Rule Mining is market basket analysis. Apriori algorithm generates interesting frequent or infrequent candidate item sets with respect to support count. Apriori algorithm can require to produce vast number of candidate sets. To generate the candidate sets, it needs several scans over the database. Apriori acquires more memory space for candidate generation process. While it takes multiple scans, it must require a lot of I/O load. The approach to overcome the difficulties is to get better Apriori algorithm by making some improvements in it. Also will develop pruning strategy as it will decrease the scans required to generate candidate item sets and accordingly find a valence or weightage to strong association rule. So that, memory and time needed to generate candidate item sets in Apriori will reduce. And the Apriori algorithm will get more effective and efficient.

General Terms
Mining, Frequent Itemset Generation

Keywords Data Mining, Association Rule Mining, Apriori algorithm, Frequent Itemsets, Association Rules

1. INTRODUCTION
In today's competitive world, every business has to fight huge competition to achieve success. So it is necessary for every business organization to collect large amount of information. Data mining is the process of discovering actionable information from large sets of data. Data mining uses mathematical analysis to derive patterns and trends that exist in data. Typically, these patterns cannot be discovered by traditional data exploration because the relationships are too complex or because there is too much data [1].

1.1 Data Mining
Data Mining is defined as extracting the information from the huge set of data. Data mining mines the data, information, and knowledge. Data are any facts, numbers, or text that can be processed by a computer. Today, organizations are accumulating vast and growing amounts of data in different formats and different databases. Information is the patterns, associations, or relationships among all this data can provide information. For example, analysis of retail point of sale transaction data can yield information on which products are selling and when. Knowledge is the Information can be converted into knowledge about historical patterns and future trends. For example, summary information on retail supermarket sales can be analyzed in light of promotional efforts to provide knowledge of consumer buying behaviour. Therefore, a manufacturer or retailer could determine which items are most susceptible to promotional efforts[2].

The actually data mining processes work as in three steps. Initially, first we need to clean and integrate the database. While the data source may come from different databases, which may have some inconsistencies and duplications, so it must clean the data source by removing those obstructions or make some negotiations. The second process is to select task related data from the integrated resources and transform them into a format that is ready to be mined. After selection of right data various data mining techniques are applied to the data source, different knowledge comes out as the data mining result. After we get the knowledge, the final process is to visualize the results. This process is try to make the data mining results easier to understand Data mining have several techniques, that have been developing and using in data mining projects.

Data mining have several techniques that have been developing and using in data mining projects. There are different schemes can be used to classify data mining techniques. These schemes are kinds of database to work on, what type of knowledge to be mined, what variety of techniques to be utilized. Based on the next kind of database, data mining techniques are categorized. The types of databases are Relational database, Transactional database, Spatial database, Temporal and time series database, World Wide Web data[3,4]. Data mining is categorized in techniques as association rule mining, classification, clustering. Association rule mining is one of the most important and well researched techniques of data mining. We will discussed it is in next part.

1.2 Association Rule Mining(ARM)
ARM is one of the best milestone of data mining. ARM was first introduced by Agrawal et al. 1993[5]. ARM aims to extract interesting correlations, frequent patterns, association or informal structures between set of items or database or other data repositories. The association rule mining technique is used in market basket analysis to identify a set of products that customers frequently purchase together. Retailers are using association mining technique to investigate customers buying habits. Based on past sale data, retailers might find out that customers always buy crisps when they buy beers, and therefore they can put beers and crisps next to each other to save time for customer and raise the sales[4].

Association rules are extensively used in a variety of areas such as telecommunication networks, market and risk management, inventory control etc. The conceptual model of Association Rule Mining is in Fig. 1.

In ARM model, ARM consist of processes sequentially as selection of required data from database, processing on selected data as per the mining requirement and before it is going to use for mining, then ARM is done on processed data and frequent patterns are generated, by means of the end user's requirement of output knowledge the post-processing is done on frequent patterns.

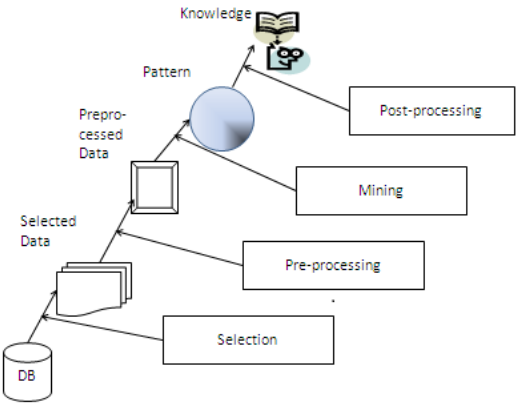

Fig. 1 : Conceptual Model of Association Rule Mining

In ARM model, ARM consist of processes consecutively as selection of required data from database, processing on selected data as per the mining requirement and before it is going to use for mining, then ARM is done on processed data and frequent patterns are generated, by means of the end user's requirement of output knowledge the post-processing is done on frequent patterns.

Association Rules are if/then statements that helps to discover relationships among unrelated data in a data repository. Association rule uses two criteria support and confidence to identify the relationships and rules are generated by analyzing data for frequent if/then pattern. Association rules are generally needs to convince a user specified minimum support and a user specified minimum confidence at the same time.

Support(S) of an association rule is described as the percentage of records that holds union of X and Y to the total number of records in the database.

Confidence(C) of an association rule is defined as the percentage of the number of transactions that contain union of X and Y to the total number of records that include X. Confidence is a measure of strength of the association rule. The Association rule in terms of Support and Confidence is as shown in Fig. 2.

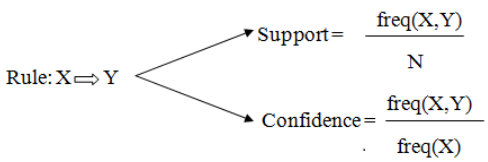

Fig. 2 : Association Rule

Association rule generation is usually split up into two separate steps :

- First, minimum support is applied to discover all frequent item sets in a database.
- Second, these frequent item sets and the minimum confidence constraint are used to form rules . While the second step is straightforward, the first step needs more concentration

The unique motivation for searching association rules came from the need to analyze so called supermarket transaction data, that is, to examine customer behaviour in terms of the purchased products. Association rules describe how often items are purchased together. For example, an association rule "sugar, tea(80%)" states that four out of five customers that bought sugar also bought tea. Such rules can be useful for decisions concerning product pricing, promotions, store layout and many others[2].

ARM have several algorithms such as Apriori algorithm, FP-Growth algorithm, ECLAT algorithm. One of the best ARM algorithm is Apriori algorithm. We will discussed it in next part.

1.3 Apriori Algorithm

Apriori is a great improvement in a history of association rule mining. Apriori is efficient during the candidate generation process. Apriori algorithm is based on the Apriori principle which says that if an item set is frequent, then all of its subsets must be frequent. It Means that the item set 'X' containing item sets X is never large if item set X is not large. Apriori is designed to operate on databases containing transactions. The Apriori principle gives the advantage as it reduce the number of items being considered by only exploring the item sets whose support count is greater than the minimum support count. But still it has the drawback as it takes more time, space and memory for item generation process. And for generating the item sets it requires multiple scans over the database. So that, it needs some improvements in it, in such a way, enhancement will overcome its drawback.

2. LITERATURE SURVEY

To understand the tradeoffs in today's ARM algorithms that is Apriori algorithm , it is helpful to study their history briefly. Apriori is best enhancement in the history of association rule mining. To recognize the Apriori algorithm, it must needed to know about their variations.

Apriori algorithm: Agrawal et al. proposed Apriori algorithm in[6]. Apriori is more efficient during the candidate generation process. It uses a breadth first search scheme to count the support of item sets and uses a candidate generation function which uses the downward closure property. Apriori uses pruning techniques to stay away from measuring certain item sets, while guaranteeing completeness. Apriori takes the merit as any subset of a frequent item sets is also a frequent item set. It takes more time and more memory for candidate generation process. To generate the candidate set it requires multiple scans over the database.

Apriori-TFP algorithm: Yang et al. proposed Apriori-TFP algorithm in[7]. In an association rule mining process, raw data are primarily pre-processed and stored in a partial support tree (P-tree). Then, Association Rule Generation is done. ARM process time is reduced, due to efficient preprocessing. It takes number of scans for dense data.

3D-Apriori algorithm: SHAO et al. proposed 3D-Apriori algorithm in[8]. 3D-Apriori is to understand the logging data. There are mainly two phases, in the first step, attribute data discretion and spatial predicate extraction. Spatial association

rules is calculated between two spatial point objects with just one predicate without considering the spatial distance. Such computation continues until it gets the maximal frequent itemsets. In the second phase, calculate all of the rules that is greater than the minimum support threshold, which is based on the candidate item set. Spatial association rule is mined in a certain space breadth. It increases the efficiency of ARM for the logging data transformation. Time complexity of the 3D-Apriori algorithm is mainly caused by the space predicates.

GP-Apriori algorithm: Zhang et al. proposed GP-Apriori Algorithm in[9]. GP-Apriori performs a parallelized version of the support counting step on the Graphical Processing Unit (GPU)Graphical Processing Unit. Accelerating Apriori with a GPU involves careful consideration of the vertical transaction list representation. Tidsets are stored as linear ordered arrays and when traversing them during the support counting operation, the resulting memory access pattern and instruction stream branching behaviour is unpredictable. Support counting procedure is based on complete intersection. In complete intersection, candidates are copied from main memory to graphic memory. GPU calculates their support ratio value by executing bitwise intersections on their vertical transactions lists and support value results are copied back to main memory. It gives speed up on modern GPU compared with CPU Based Apriori implementations. Complexity is increased in terms of ARM due to GPU.

Apriori Mend algorithm: Delighta et al. proposed Apriori Mend algorithm in[10]. It generates item sets using a Hash function. Initially it generates frequent Itemset using perfect Hash function in the database. The user has to specify the minimum support to prune the database Itemset and deletes the unwanted Itemset. Then pruned database itemsets are grouped according to the transaction length. Apriori Mend algorithm is found to be more admirable than the traditional method Apriori algorithm in terms of efficiency. Execution time is increased.

Parallel Apriori algorithm: Ning et al. uses Parallel Apriori Algorithm Based on MapReduce in[11]. Here, Parallel Apriori algorithm is implemented based on MapReduce, which creates it applicable to mine association rules from large databases of transactions. Map function and Reduce function are used to generate the association rules. Map function performs the procedure of counting each occurrence of candidates of size k and thus the map stage realizes the occurrences counting for all the potential candidates in a parallel way. Then, the Reduce function performs the procedure of summing the occurrences counts. For each iteration, such process is carried out to execute the occurrences computing for potential candidates of size k. P-Apriori can scale fine and efficiently processed large datasets on commodity hardware. It requires more computation power and memory to find association rules.

Apriori Algorithm for Multidimensional Data : Sulianta et al. uses Apriori Algorithm for Multidimensional Data in[12]. In this Apriori algorithm, Multidimensional Data Reduction process is a pre-processing. The problem in this case focused on the process of preparation of data. Cleaning and integration are done after collect the data from the system. Selection and Transformation is used which includes procedures to handle the data which is transformed to standardize data for mining process. Validation levels are implemented to verify the reliability of the association rules - Data training after reduction, Data training without reduction, Data testing. It explores multidimensional data handling methods to build association rules more specific to product effectively. It must require the data reduction.

Apriori algorithm in minimizing candidate generation: Sheila proposed this improved apriori algorithm in[13]. The improved apriori algorithm introduces factors such as set size and set size frequency which are being used in elimination of non significant candidate keys. Set size which is the number of items per transaction and set size frequency which is the number of transactions that have at least set size items.

Average case Performance of the Apriori algorithm: Purdom et. al. studies in[14] that the failure of the apriori algorithm using both analytical and experimental approaches. The analytical work is done on a simple model where each shopper buys at random. The experimental work is three data sets that are more like the usual applications of the algorithm, the buying pattern of the various shoppers are highly correlated . In the best case, the candidacy test always correctly predicts the result of the frequency test and the amount of time spend by the apriori algorithm is essentially the amount of time spent verifying that all the output sets should be output. The worst case time needed by the apriori algorithm is polynomial in the sum of the size of the input plus output. Since time needed by the apriori algorithm is very large.

Apriori algorithm has various variations. Apriori algorithm inherits the drawback of number of scans over the database and more memory space required to candidate generation process in each variance. It highly needs to reduce the scans over the database and reduce memory size, in such a that Apriori will work efficiently and accurately.

3. PROPOSED SYSTEM

With reference to literature work, the initial problems for Apriori algorithm have been noticed. The proposed solution introduces a better way for Apriori algorithm by doing enhancement in existing algorithm.

Improved Apriori algorithm tries to reduce the drawbacks of traditional Apriori algorithm. Improved Apriori algorithm is designed in such a way, it executes effectively and efficiently. It should take less time and minimum number of scans to generate frequent itemsets. The algorithm for Improved Apriori algorithm is as follows:

Improved Apriori algorithm :

1. Start
2. Define the min_sup //Give minimum support count
3. Enter X{} //Generate the itemsets
4. Find λ X } // Number of transactions contains X {}
5. For all transactions T from Database D, Get the Target transactions(Tgt[m, n]) that contains X{}
6. Tgt[m, n] = Get Transactions Ids that contains λ = X {}
7. For each transaction in Tgt[m, n] Do
8. Find the items that are frequently occurred with X{} from Tgt[m, n]
9. Frequent Itemset = [items in transactions in Tgt that are frequently occurred with given itemset] ∪ [Given itemset(X{})].
10. Generate the Association Rules <= Frequent Itemsets
11. Mine the Association Rule(z) = Frequent itemsets in transactions in Tgt ≥ min_sup. //Rules according to support count
12. End

The Improved Apriori algorithm, it first generates the minimum support count. Then according to the given item or itemsets, it locates the transactions that contains the given item or itemsets. After that, algorithm finds the items that are frequently occurred with the given itemsets. Later that generate the association rules. According to given support count it mines the association rules that is the frequent Itemset's occurrence in transaction is greater than or equal to the specified support count.

In order to understand the working of architecture of improved apriori algorithm, it is essential to study the required blocks for architecture. The architecture of improved Apriori algorithm includes the textile database, items selection and quantity selection, two apriori algorithms such as Apriori algorithm and Improved apriori algorithm, Association rule mining(Frequent Itemset mining) and Result Analysis. At the end, the results of Improved Apriori algorithm are evaluated in terms of number of scans taken by the ARM, time required to mined the association rules.

The architecture of a system all the time defines its broad outlines, and may describe specific mechanisms as well. Architecture is a passion, a vocation, a calling in addition to a science and a business. It has been described as a social art and also an artful science. Architecture must be of the uppermost quality of design. The architecture of improved apriori algorithm is shown in Figure 3 and the block description is as follows.

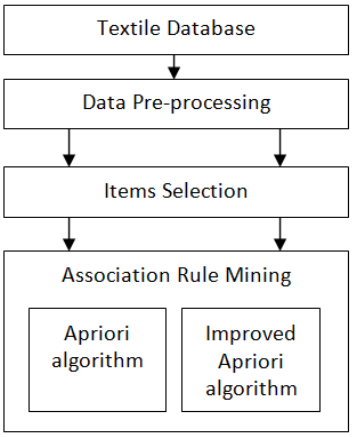

Fig. 3 : Architecture of Improved Apriori Algorithm

Textile Database: The Textile database is a transactional database. Textile database consists of Blanket Item data such as Transaction-id, Date, Customer name, Item-code, items, price, quantity, total bill.

Data Pre-processing: Textile database contains the various parameters such as Transaction-id, Date, Customer name, Item-code, items, price, quantity, total bill. But here, Improved Apriori only needs the parameters as Transaction-id and Items. Hence, only these two parameters are extracted into another file from the original database. After that, conversion of presence of items in transaction is converted into 0 and 1 that is 0 for absent and 1 for present.

Items Selection: Here, it selects the items from the database for finding the frequent itemsets with the given items.

Association Rule Mining: An association rules mining contains the following steps :

- The set of candidate k-item sets is generated by 1-extensions of the large (k-1) item sets generated in the previous iteration.
- Supports for the candidate k-items sets are generated by a scan over the database.
- Item sets that do not have the minimum support are discarded and the remaining itemsets are called large k-item sets.

Association rule uses two criteria support and confidence to identify the relationships and rules are generated by analyzing data for frequent if/then pattern. Association rules are normally needs to satisfy a user specified minimum support and a user specified minimum confidence at the same time.

Let X or Y = Itemsets

- Support count(σ) = Number of Transactions Containing an Itemset
- Support(S) = Percentage of transactions containing an Itemset(both X and Y)
- Frequent Itemset = An itemset whose support is greater than or equal to a min_sup threshold
- Confidence(C) = Ratio between the number of transactions containing both X and Y and the number of transactions containing X.

The equations (1) and (2) are Support and Confidence equations, they are given below.

$$S = \frac{\sigma(X,Y)}{| T |} \qquad (1)$$

$$C = \frac{\sigma(X,Y)}{\sigma(X)} \qquad (2)$$

According to specified Support count confidence Association rules are mined from the database. ARM first generate the rules then mine the rules according to support count.

Apriori algorithm : The Apriori algorithm, it generates the itemsets using only the large itemsets of previous scan without considering the transactions in the database. It involves two processes to find out all the large itemsets from the database in Apriori algorithm. First the itemsets are generated that are frequently occurred with each other, then the database is scanned to check the actual support count of the corresponding itemsets.

Improved Apriori algorithm: The Improved Apriori algorithm, it initially produces the minimum support count. After that according to the given item or itemsets, it traces the transactions that includes the given item or itemsets. Then, algorithm finds the items that are frequently occurred with the given itemsets. Afterwards, generate the association rules. According to given support count it mines the association rules that is the frequent Itemset's presence in transactions is greater than or equal to the given support count.

The Improved apriori algorithm is for finding all the frequent itemsets. Input for the improved apriori algorithm requires as item or itemset from the database. And output comes will frequent itemsets from the database.

According to Support equation, Improved apriori algorithm defines the Support Count that is min sup,

Let X or Y = Itemset

- Support count (σ) = Number of Transactions Containing an Itemset
- Support(S) = Percentage of transactions containing an Itemset(both X and Y)
- Frequent Itemset = An itemset whose support is greater than or equal to a min-sup threshold
- Confidence(C) = Ratio between the number of transactions containing both X and Y and the number of transactions containing X.

After getting the support count, generate the item or itemsets. Let X{} = Item or Itemset. That is finding = λ X{}

For all the transaction T from database D, extract the target transaction(Tgt[m,n]) in which given itemset (X) present in target transactions,

Where m = Tgt_Id, n = Tgt_Items.

$$Tgt[m,n] <= \forall T(D)$$

For all the transactions Tgt from target transactions, find the items that are occurred with X{} from Tgt[m,n],

$$(X'\{\}) <= \forall Tgt[m,n]$$

Where (X'{}) = Frequent Itemsets that are frequently occurred with X{}

Generate the Association Rule,

$$Z <= (X\{\}) \cup (X'\{\})$$

Where Z = Association Rule

Mine the Association Rules as per the Support Count,

$$Z \geq \text{min sup}$$

Improved Apriori algorithm is proposed in such a way it overcomes the drawback of Apriori algorithm. It will prove in next section according to the results.

4. RESULTS AND DISCUSSION

The ARM algorithm such as proposed improved apriori algorithm is evaluated by using parameters such as time and no. of scans it takes to mine the association rules. This algorithm are evaluated on textile dataset. Textile database contains the attributes like item and itemsets.

The Improved Apriori algorithm is appraised by performance metrics as time and number of scans. A performance metrics determines a behaviour and performance of output values. It includes Time which is required to find out the frequent itemsets and No. of Scans which are required to check the transactions for Association Rules occurrences.

Table I shows the four cases. Cases that contains the item and itemsets that are entered during the execution of the Apriori and improved Apriori algorithm. It shows recommended itemsets that is frequent itemsets according to support count. For examples, in case 1, {Alps} item is entered to find the frequent Itemset. It gives the results as frequent Itemset {Alps}Union{Karishma shawl, Glory, Imperial} for Apriori and {Alps}Union{Maria} for Improved Apriori algorithm.

After that, in table I, It specifies the time and no. of scans taken by the Apriori and Improved Apriori algorithm.

Table 1. Results for Apriori and Improved Apriori With Frequent Itemsets, Time and Number of scans

Cases	Itemsets X{}	Recommended Itemsets X'{}		Time(t)		Number of scans	
		Apriori	Improved Apriori	Apriori	Improved Apriori	Apriori	Impro-ved Apriori
Case1	{Alps}	Karishma shawl, Glory, Imperial	Maria,	0.3086	0.0678	272	9
Case2	{Azure, Corolina}	Delight, Chancellor, Vervain, Elixir	Maria, Prince 2SS	0.1588	0.0593	272	8
Case3	{Alps, Amabel, Azure}	Peninsula, Nova, Polar Bear Rug, Arena	Viola, Bhagirath, Alaknanda	0.1582	0.0601	272	9
Case4	{Alps, Amabel, Azure, Corolina}	Command 4SS, Delight, Sanford	Maria, Plateau, Fascino	0.1681	0.036585	272	9

Figure 3 shows the graph based on the time taken by both Apriori and Improved Apriori algorithm. It shows the huge difference between the time required by both algorithms.

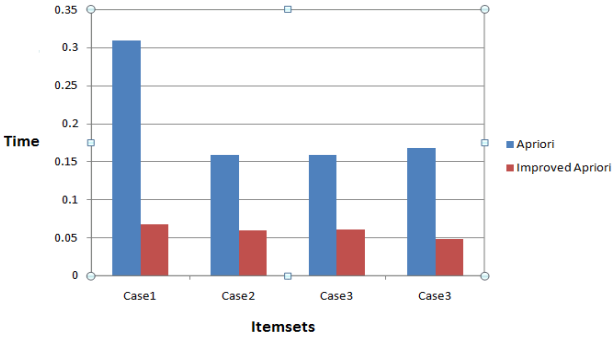

Fig.4 : Apriori and Improved apriori algorithm output for time required

Figure 4 shows the graph based on the no. of scans taken by both Apriori and Improved Apriori algorithm. It shows the vast divergence among the no. of scans needed by both algorithms. Improved Apriori algorithm takes only those transaction in which given item or itemsets are available. So that, algorithm goes through only those transaction to find the frequent itemsets not use the all transactions. Hence, it takes minimum no. of scans. And because of this it requires less time.

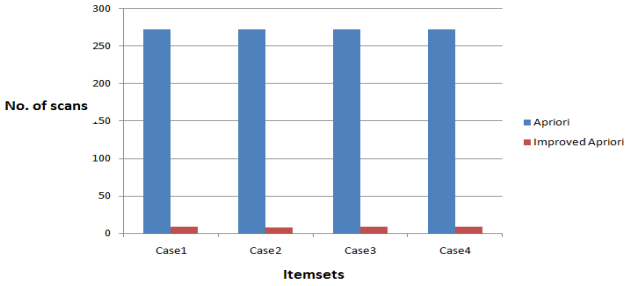

Fig.5 : Apriori and Improved apriori algorithm output for number of scans required

Improved Apriori algorithm performs superior as compared to Apriori algorithm in terms of time required to produce and mine the association rules & number of scans taken for the itemsets generation. Improved Apriori algorithm proved that it is better than the Apriori algorithm as both algorithms are evaluated on the basis of same parameters.

5. CONCLUSION

Association rule mining is a data mining algorithm and plays major role for extracting knowledge and updating of information. ARM algorithm applied on textile dataset has resulted in novel approach which have significance success in mining the association rules from textile database. Improved Apriori algorithm is applied on textile database to find out frequent itemsets. Yet, the main drawback of apriori algorithms like requires more time and takes large no. of scans which are required to mine the frequent itemsets are pointed out. The drawbacks are overcome by proposing improved apriori algorithm in such a way it takes less time and less no. of scans than the apriori algorithm. The evaluation shows the peak improvement in the mining results. In future work, classification can be used for finding frequent itemsets. In classification, Decision tree can be used for the multilevel association rule mining.

6. ACKNOWLEDGMENT

Authors would like to express a deep sense of gratitude to Prof. Dr. Girish K. Patnaik, Head of the Computer Science and Engineering Department for his guidance and support in this work. The authors are also thankful to the Principal, SSBT"s College of Engineering and Technology, Jalgaon for being a constant source of inspiration.

7. REFERENCES

[1] J. Han and M. Kamber, "Data mining: concepts and techniques." Morgan Kaufmann, 2006. [Online]. Available:http://scholar.google.de/scholar.bib?q=info:kYdwviD3IR4J:scholar.google.com/&output=citation&hl=de&assdt=0&scfhb=1&ct=citation&cd=0

[2] S. Brin, R. Motwani, and C. Silverstein, "Beyond market baskets: Generalizing association rules to correlations." in SIGMOD Conference, J. Peckham, Ed. ACM Press, 1997, pp. 265-276. [Online]. Available: http://dblp.uni-trier.de/db/conf/sigmod/sigmod97.html#BrinMS97

[3] M. syan Chen, J. Hun, P. S. Yu, I. T. J, and W. R. Ctr, "Data mining: An overview from database perspective," IEEE Transactions on Knowledge and Data Engineering, vol. 8, pp. 866-883, 1996.

[4] Sourav. S. Bhowmick, Qiankun Zhao, "Association rule mining: A survey," CAIS, Nanyang Techno- logical University, Singapore, Technical Report 2003116, 2003.

[5] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," in ACM SIGMOD Record, vol. 22, no. 2. ACM, 1993, pp. 207-216.

[6] R. Agrawal, R. Srikant et al., "Fast algorithms for mining association rules," Proc. 20th Int. Conf. Very Large Data Bases, VLDB, vol. 1215, pp. 487- 499, September 1994. [Online].Available:http://rakesh.agrawalfamily.com/papers/vldb94apriori.pdf

[7] Z. Yang, W. Tang, A. Shintemirov, and Q.Wu, "Association rule mining-based dissolved gas analysis for fault diagnosis of power transformers," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 39, no. 6, pp. 597-610, 2009.

[8] X. dong Shao, "The application of improved 3dapriori three dimensional association rules algorithm in reservoir data mining," in CIS (1). IEEE Computer Society, 2009, pp. 64-68. [Online]. Available: http://dblp.uni-trier.de/db/conf/cis/cis2009-1.html#Shao09

[9] F. Zhang, Y. Zhang, and J. D. Bakos, "Gpapriori: Gpu-accelerated frequent itemset mining." in CLUSTER. IEEE, 2011, pp. 590-594. [Online]. Available:http://dblp.unitrier.de/db/conf/cluster/cluster2011.html#ZhangZB11a

[10] I. S. P. J. D. Magdalene Delighta Angeline, "Association rule generation using Apriori mend algorithm for student's placement", vol. 2, no. 1, 2012, pp. 78-86.

[11] N. Li, L. Zeng, Q. He, and Z. Shi, "Parallel implementation of apriori algorithm based on mapreduce," in Software Engineering, Artificial Intelligence, Networking and Parallel Distributed Computing (SNPD), 2012 13th ACIS International Conference on, 2012, pp. 236-241.

[12] F. Sulianta, T. H. Liong, and I. Atastina, "Mining food industry's multidimensional data to produce association rules using apriori algorithm as a basis of business strategy," in Information and Communication Technology (ICoICT), 2013 International Conference of, 2013, pp. 176-181.

[13] S. A. Abaya, "Association rule mining based on apriori algorithm in minimizing candidate generation," International Journal of Scientific and Engineering Research, vol. 3, no. 7, pp. 1-4, July 2012.

[14] P. W. Purdom, D. V. Gucht, and D. P. Groth, "Average-case performance of the Apriori algorithm."SIAM J. Comput., vol. 33, no. 5, pp. 1223-1260, 2004. [Online]. Available:http://dblp.unitrier.de/db/journals/siamcomp/siamcomp33.html#PurdomGG04.