Journal of Critical Reviews

ISSN- 2394-5125 Vol 6, Issue 6, 2019

Review Article

ADAPTIVE MINING OF ASSOCIATION RULES OF INTER-TRANSACTIONAL DOMAINS

Dr.S. NANDAGOPAL[1]

[1]Professor, Department of CSE, Nandha College of Technology, Erode, Tamilnadu, India. asnandu@gmail.com

Received: 05.11.2019 Revised: 12.12.2019 Accepted: 13.12.2019

ABSTRACT

Generalization of itemsets of an organization with support and confidence metrics have helped the association rule mining by describing a preorder of transactions and their origin from the ancestral transactions. Intertransactions of different backgrounds stipulate far beyond metrics to be considered for mining process. Bounding the factors which determine the frequency of transactions, in turn simplifying the generations of candidate sets is the approach CBIT [Categorized and Bounded Inter-Transaction mining algorithm] to be discussed in this paper. Differentiation of intra and intertransactions lies in the limits applied to the customer, date and time or even a maintenance of records for a specific period of time. Intertransactions are limited by a very few of its ancestral constraints as they are intended to mine association rules from transactions and incitation of successive transactions of other domains. Hence the final methodology is adept in deriving the associative level of transactions in a bounded yet large domain of itemsets. Comparisons with previous renowned strategies would best describe the efficiency of this proposed technique.

Keywords: Taxonomies, Generalization, Inter-transactions, Bounded.

DOI: http://dx.doi.org/10.22159/jcr.06.06.54

INTRODUCTION

The events of an organization are maintained periodically in the form of databases for various rationales. These events trigger the futuristic promotions that will not affect the growth and health of the organization by any means. Different methodologies have been proposed for efficient storage and manipulation of data transactions. Efficiency of an organization depends completely on the effective decisions made at the right time and locations. Every domain needs to update its activities with respect to the response of its participating customers. Many successive organizations evolve from a scratch by appropriate decisions initiated at the right instance. A sufficient example of an organization that is based completely on the priority and likability of the participating customers is a supermarket. The supermarket needs to keep track on the products most preferred over the others of same criterion and the associated purchase of products which accompanies the first product. This facilitates the supermarket to perform the services intended to enhance the profit and sustain its ranking in the society. Data mining concepts support the organization to thus retrieve patterns of likability of purchases by the customers. Many standards of data mining abridge the analysis process of transactions and produce a pattern as the outcome. This shortens human efforts in maintenance and sustainability. Explaining the supermarket context, every purchase made by a customer is defined as a transaction. The transaction includes a list of products and provisions. A supermarket cannot subject itself to specific products but promises good quality with differentiations on cost and associative compositions. There exist certain patterns among the transactions of a particular customer with time and other attributes. These patterns are described by the support and confidence values defined for each product. Support is the probability of how often the same combination of products is purchased together and Confidence is the measure of initiation of the second product's purchase by the first. The market basket analysis was framed with this context, and was applied in other domains with adequate measure of feasibility. Offers for promotion, arrangement of products on shelves, organizing the products in a right flow of preferences, variations among the same kind yet explainable and analysis of customer behavior are determined by this market analysis.
An association rule follows the representation as

$$\text{BREAD} \Longrightarrow \text{BUTTER (sup=40\%, conf=60\%)},$$

where bread and butter are individual products and the combination of both purchases has a support of 30%, 60% percent of the customers who purchase bread would go for butter as the next option. There are another kind of customers who prefer jam and bread, some with eggs and bread. The frequency of repetitions of typical purchases would motivate the organization to introduce offers, rearrange the layout of storage, maintenance of stock held and many other reasons. The association rules mining involve two major processes. Deriving a table of transactions with metrics of support and confidence of all combinations made by a single customer on the same day or period of weeks or months is the first process. The next step is to derive the association rules between different transactions. Association rules should satisfy the minimum support and minimum confidence levels in order to be a strong outcome for future analysis. These computations need level wise analysis over the recorded transactions and for deriving the support and confidence measures. Huge amount of input and output conditions have to be analyzed and stored in memory. The outcomes have to be tested for their efficiency to prove their impact over the decisions. Several taxonomies have been proposed to simplify the process of this storage and retrieval from memory, computations over the transactions and final representation as a tree after elimination of unnecessary representations (pruning). The final outcomes have to be fruitful after spending a considerable expense of resource over association rule mining. The association rule mining may tend to eliminate less frequent but significant transactions of a customer and they incur a huge force over being avoided. Determination of minimal support and confidence level terminates the difficulties of traditional approaches.

PROBLEM DESCRIPTION

The environments of the domains are not as simple to merely implement the data mining aspects to retrieve patterns. They have to be preprocessed into certain procedural representations to achieve the outcome. Moreover the domains of the same organizations were implemented with data mining aspects. There are limited factors which makes the association rule mining easier (Agrawal R, Srikant R, 1994). Notion of the transaction is in and around the activities of the unique organization and their motives. Studying and understanding the concept of functionality, boundaries are the basic terminologies to enhance the ordinary model into an inter-transactional domain. But this simplicity cannot be expected from a wider and universal domain of multidimensional environments. A pertinent model is proposed and described in this chapter.

Factors for Consideration in Inter-transactional Domain

In the previous section, the environment of a single and simpler example was explained. There are domains which are interrelated and dependent on the other domains in this society. A stock exchange domain can be defined as a suitable and well known example for an inter-transaction domain. The dominance of increased and decreased prices of one's share has adverse and affirmative effects over the complete independent domain of another. There are certain attributes which describe the activities of an organization; in order to be compared these attributes are subjected to analysis. Each transaction in traditional approaches is defined by an identity, customer id and thus the list of products purchased. An intra-transactional domain can be identified (Tung A K H et al, 2003) and processed easier since they belong to the same organization. They are well known to the administrator, officials and all other handlers of comparisons. As the detailed classifications never fall beyond the limitations of the preset vision of the organization, enlisting the transactions would not be a tiring task.

Since there are absolute no or known few possession of similarities between different domains, their transactions need to be transformed into a general form for better understanding of features. Unlike Intra-transactional domains, a number of independent domains participate in the algorithms to be mined. They are related to each other if a domain takes steps to an action as a response to the action of its competitor. Inter-transactional domains have independent attributes accordingly and cannot be compared straight away. They need to be transformed into a generalized format where these factors fall under a common category to facilitate the further processes. These attributes defines a format of organization's primary factors, transactions, employees and operations.

The transactions, employees and customers are associated with a cyclic association. The organization has the motive of producing products as desired by the customers assuring quality and cost satisfaction. Employees are indulged in planning the tasks, taking decisions and finally marketing the product to the customer to achieve the objectives of the organization. The customers on the other hand will not possess enough knowledge and resources to manufacture the product and use them. They in turn select the right product in the market among a huge list and get to apply them in real life situations. In case of inter-transactional domains, researchers introduced the concept of multidimensional attributes such as time, location, distance, temperature, latitudes, longitudes etc. Attribute gives the factual portrayal of positions, features and kind of transactions. Attributes are also independent to the typical organization corresponding to different modes but the same idea.

Apart from all these difficulties, many organizations urge the association rule mining for a number of applications. An accident on a highway needs to be taken care by diverting traffic flow in other ways to avoid further mishaps; the offers to a combo pack of products, weather prediction, stock exchange predictions, understanding customer behavior and much more domains are inter-transactional domains which imply the strategies. Classical approaches defined two phases for framing the association rules between transactions of inter domains.

1. Generation of a frequent itemset
2. Derivation of association rule mining

The frequent itemset comprises of transactions that satisfies a condition more than or equal to a minsup value. This generation of an itemset needs the study over all the transactions as inputs. After generation, the association rules can be mined straight away.

The same initial phase cannot be incorporated directly into the case of multidimensional and inter-transactional domains. Apriori algorithm follows a level wise examination of repetitions of transactions to derive association rules. There are a number of expensive processes which can be avoided to enhance the performance of algorithm. Certain calculations in Apriori algorithm need additional space and time thereby reducing the performance. Intermediate candidate sets (itemsets) are produced unnecessarily, wasting the memory resources. Whenever a new item or transaction is introduced into the set, the whole transactions are analyzed again from the first. This increases the whole execution time without any relevance. Resources of memory, input variables are indeed to be conserved for better usage without being overused and assigned for incremental processes. Another drawback of this classical approach is the derivation of misleading association rules which may be frequent for a number of times but do not have an impact on the desires of the organization. The following sections discuss the variations of algorithm for efficient pruning and classification of transactions at earlier stages. The preprocesses of the algorithm are explained as follows.

PROPOSED METHODLOGY

Generalization

Transactions obviously possess a similarities and dissimilarities with other transactions irrespective of the domain. Generalization is a technique of categorizing the transactions based on the origin (Ming - Cheng Tseng, Wen-Yang Lin, 2007), from which they have been stimulated. Each of its own kind belongs to bounded closure (Anthony J et al , 2008), that is, it would be a parent to some specific transactions, would be a leaflet or possess no children at all. Considering the supermarket domain, a list of products belongs to different hierarchies rather than being denoted by the organization. The hierarchy defines the nature and characteristics of the product as generalized products in the top level and to the typical kind in lower levels. Otherwise known to be taxonomy, could lead to better association rule mining by ordering the transactions facilitating an easier association rule mining process. Further uses of defining taxonomy in the organization are, to predefine the existing categories and placing a new transaction into the hierarchy it best fits in. This step has eliminated the need of generating new candidate sets as in Apriori algorithm thus conserving resources of memory and computations.

Ancestors are generalized leaders of a specified group of transactions and Descendents are representations of direct events those actually occurred. Ancestors present the familiar name for combining all the transactions even with a small resemblance. Moving down to the descendents present the transactions along with the identities of customers, products and session of the respective customer in the domain. Descendents circumvent the generality and mentions the attributes directly for differentiation. General characteristics of an ancestor usually the domain knowledge are completely opposite to attribute values of transactions in the descendents. A sample set of transactions which could be displayed as taxonomy is shown fig 3.1.

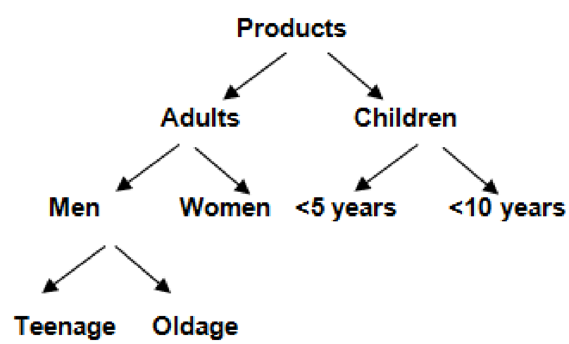

Fig 3.1: Taxonomy of a Supermarket

Let P= {$p_1, p_2, p_3, \dots p_n$} be a list of products defined by individual product ids, CP = {$t_1, t_2, \dots. t_m$} is a set of Customer Purchases with transactions or purchases. Each transaction would comprise of a transaction id (combination of customer id and list of products purchased) denoted by t_i = {t_{id}, P}. It is assumed that all products possess a certain characteristics and they belong to a certain ancestor of taxonomy. The taxonomy defined in this algorithm is a context of a supermarket. The products are categorized as follows based on the type and age group it has been dedicated for. Separation of these products with respect to their age group may or may not be optimistic in all situations. The detailed taxonomy of supermarket algorithm is shown in fig 3.2.

The following tables (3.1, 3.2 and 3.3) display the ids which represents each of the products and transactions. As already defined in the algorithm, no support and confidence metrics are required for construction of candidate sets as in previous strategies. The taxonomy defines the id of each product, which categorizes the transactions of further each customer. In case of an intertransactions domain, the id assigned to each domain is prefixed to the existing ids. The table 3.4 describes the necessity of included ids for an inter-transactional domain. Transactions are the same as intra-transactions varying in context of applicability.

In a multidimensional and interrelated environment, multiple supports for a single transaction exist. This area of extracting different levels of supports for analysis and framing of association rules need absolute care, as a less frequent item may prove watchful and having a serious impact of being avoided. But a candidate set generation only looks for a satisfaction of a minimal support value, which seems to be an uncomplicated process, the pruning of less frequent yet significant transactions prove expensive. Previous methods undergo a strategy which computed the minimal support value min (sup), in the functioning of multidimensional environments, comparing all the values of support of different transactions. It is ensured that minimal support would always fall under the candidate set generation other values found to be lesser than minimal value min (sup) would be pruned from further estimation. Even this step seems to be expensive as subjection of multiple supports of same transaction, as a diversity of min (sup) is present for different combinations

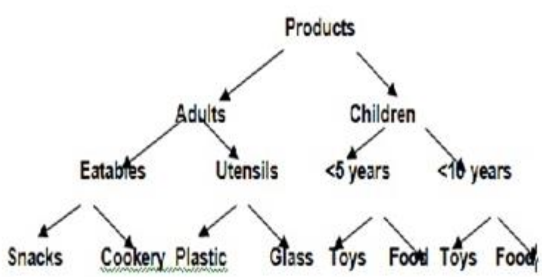

Fig. 3.2: Detailed taxonomy of algorithm

Table 3.1

Product Id	Product Name
A01	Bread
A02	Butter
A03	Jam
A04	Eggs

Table 3.2

Product Id	Product Name
P01	Toys
P02	Milk
P03	Bread
P04	Medicine

Table 3.3

Customer Id	Transactions
C Id	Product List
C Id	Product List
C Id	Product List
C Id	Product List

Table 3.4

Inter Transaction Id	Transactions
D Id, C Id, P Id	Product List
D Id, C Id, P Id	Product List
D Id, C Id, P Id	Product List
D Id, C Id, P Id	Product List

Step 1 Determination of Top Level Taxonomy

Different levels of specification and generalization are analyzed and represented in the preliminary stage of this algorithm. This representation starts from the top levels as the generalized lead of every transaction would comprise of all the characteristics of its descendents[4]. If the same process is reversed, it would drive to a number of additional and time consuming computations. The individual transactions have to be retrieved from storage in order to be compared. Considerable load on input and data transfer would degrade the performance of mining association rules. Thus generalization technique is followed from top level leading down to specifications of transactions. Efficient declarations of ancestors would enhance the futuristic computations of support and confidence values of every transaction.

The top level taxonomy has to be stated clearly to regulate the following steps of the algorithm. But there are some technical difficulties in placing the taxonomical details into association rule mining. The organization needs to define the outcomes and necessities of the same. These outcomes are mentioned irrespective of internal procedures thereby declaring the ancestor's information. Transactions are associated with the auxiliary actions of achieving those objectives or resemble the characteristics of ancestors. Differentiation of primary, supplementary and non user objectives have to be also declared. As all these objectives belong to the organization only the primary is focused with the higher priority. Hence only ancestors which are possessed in a number of transactions more than threshold value are considered to be the primary ancestors. The other ancestors are pruned in the extended set of transactions. Obviously the transactions with items repeated less than the threshold value are pruned. Pruning involves the deletion of the following conditions.

1) Prune ancestors which are not even present in minimal transactional threshold
2) Prune an item from extended set if it is not present in candidate sets
3) Prune non user ancestors even if they exceed the threshold

Products of inter-transactional domains possess attributes of different nature. We consider a more statistical intertransactions domain of stock exchange.

When share prices of Microsoft go up by 15%, the share prices of IBM are likely to go up by 12% in the next two days.

In this stated example, two large organizations are involved and to be accurate all transactions have to be analyzed. In our first step of the algorithm, the top level hierarchy is determined by tracing back the events of both organizations in recent past.

The top level hierarchy in fig 3.3 is based on the type which has to be considered for the typical event that has occurred. Both the organizations record the events and mark them by ids for efficient manipulation from storage. The share prices totally depend on the above stated lower level leaves of taxonomy. When one needs to predict the rise of fall of share prices of associated organizations, they need to refer the ids of equivalent processes in specific.

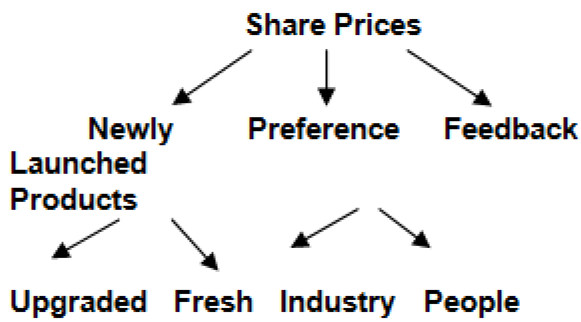

Fig. 3.3: Top Level Hierarchy

Initially many ancestors may be generalized with different characteristics but may not yield optimal performance intensity. The finalist ancestors would encompass all the different transactions within; ancestors with minimal set of descendents can be eliminated (known as pruning) as a performance measure. Pruning at earlier stages of the algorithm contributes a great part in conservation of resources and time spent.

Descendents following the lead of ancestors needs to maintain minimal support and confidence level to stay in bound to the generalized terms. This part of the algorithm requires the introduction of min (sup) and min (conf) that defines the position of leaves of the taxonomy. Descendents which fail to equalize these metrics are pruned from the taxonomy.

Step 2 Bounded Intertransactional Domains

Association mining of interrelated multidimensional environments includes many challenges such as a diversity of attributes (time, location, distance and other vectors) of each transaction in an itemset. Each organization defines its very own attributes for the events and processes[2][3]. When multiple environments needs to be merged and mined for association rules, a common platform has to be established for better understanding of unique terminologies and incorporation of the same. Yet these platforms cannot promise the completed set of association rules and hence the concept of bounding the transactions was introduced to limit the areas of concentration and computations. Bounded intertransactions domains provide only the desired set of association rules rather than a combination unwanted rules from which the user has to evaluate the appropriate.

Definition: A bounded or closed itemset is said so, if no further subsets can be derived from the present set containing the transactions involving the participants or items of the former sets.

When a closed itemset is found then it eliminates the out of bound transactions as no optimal rules could be derived. The database is repeatedly scanned for frequency of each transaction. The list of frequent transactions is reported with its level of support and confidence. Prior to this scanning process, the transactions are preprocessed to remove any empty transactions, noise and redundancy etc. The transactions with very less frequency are omitted as supposed to prevent unwanted computations. The next process is divided into individual conditions those have to be met in order to narrow down to the result as desired. Each condition simplifies the process of succeeding processes. Every condition limits the factors to be checked and the following example best describes this scenario.

Considering the supermarket example, a database maintains the identities of products, employees and customers. The customers enter and make purchases every day, these purchases are updated into the database as transactions. Transactions would carry the customer id, product id and the session of the customer in the supermarket. This database is analyzed for frequency of every purchase made depending on the same customer, same time or same age group (based on some relativity measure). Let us consider an itemset including Product A, Product B, Product C, Product D and Product E. Each product has a predefined id. All purchases from Product A to Product E is also defined with a support and confidence related to their ancestors. After short listing the frequency of products in the whole set of transactions of a completed working day, they are divided into conditions such as

1. Purchases comprising of Product A alone
2. Purchases comprising of Product B alone (without A)
3. Purchases comprising of Product C alone (Without A and B)
4. Purchases comprising of Product D alone (without A, B and C)
5. Purchases comprising of Product E alone (Without A,B,C and D)

The final outcome has constrained the limitations by imposing the boundary that is each itemset cannot be derived into a subset of another since no itemset would consist of transactions. The solution of this closed frequent itemset (Jian pei et al,2000), generation displays the frequency of transactions (purchases) higher than the minimum threshold value min (sup) and omission of less frequent items is achieved. This strategy could be implemented in intertransactional domains and obtain a productive result by enhancing the rate of consolidating the frequent itemsets in two modes. The highly significant mode is to consider the conditions based separations of transactions. The other is to separate the events based on prescribed set size. This is not optimal as there is no surety of framing each set of equal size, and another drawback of leaving out some important transactions during the division process is critical. Defining the taxonomy and boundaries on the transactions divides the area of algorithm's applicability to precise sizes.

Step 3 Defining Length of Interval

The case of inter-transactional domains faces another serious challenge of associating rules which has to be mined at regular or irregular intervals. A customer may purchase a product once in a month and this interval is every thirty days. The drawback of this time span is that rules may not be effective in larger interval transactions. They may not have the affectivity as other association rules found in regular purchases. Number of resources spent on mining those rules also degrades the performance level of the algorithm. Along with the definition of taxonomy, a certain length of interval is described as like in fig 3.4. The interval may be representations in mode of time, space, window size or number of transactions to be compared.

This methodology uses a divide and conquers technique in the number of transactions. Rather than analyzing the entire set of transactions which may be in thousands of numbers, smaller intervals are framed. Each interval occupies only three transactions at a time, finds the association finally deriving the rules of association. It is considered to be an important optimization technique for faster comparisons of innumerable transactions of an inter-transactional domain. The next example considers the transactions to be compared as three frames for simpler calculations and comparisons. Mega sized transactional structures decides the length of interval to be adequate for time conservations. Efficiency can be increased in a certain magnitude than the classical Apriori algorithm and other improvised algorithms. Combinable with taxonomy it adds more flexibility in analyzing the patterns among intertransactions. A notable factor in the assignment of interval is that it is not only applicable to the size of window to enlist the transactions. The interval could also be a specification of number of attributes to be applied for comparisons.

D	E
1	
2	
3	
4	
5	
6	
7	
8	
9	

Fig 3.4: Length of an Interval

Generation of taxonomy categorizes the transactions which already applies a limitation (closed boundary) over the complete itemset. In the already discussed context, purchases made by customers are sectored into Adults, Children and further descendents. From the levels of hierarchy, each horizontal and vertical level eliminates the computations which do not belong to the class. The purchase of baby products is diverted to the closed itemset of Children, showing an opposite deviation from the other class of Adults. It is observed that taxonomy simplifies the process of closed itemset in framing the boundary. In case of inter-transactional domains, a new itemset with new boundaries composing different metrics could be generated as we implement minimal support metrics for comparison. Predominant itemsets are thus derived from the above mentioned techniques.
Length of interval imposes a condition to perimeter the number of transactions for analysis. In a transactional database of supermarket domain, a specific day would record a large number of transactions (purchases) made by individual customers. The ultimate question is that will there be a mechanism to subject all the transactions at once by a single step. Obviously no is the answer and even if there was, the solution is certainly inappropriate. The scanning of frequency is shortened to the only present transactions within the interval length. This process eliminates the extended mining beyond the defined length. Association rules mined over a large set of transactions leads to less interesting measures usually eliminated for consideration. This length of interval argues the limited as well as constrained selection of transactions to eliminate the less interested association rules and save the resources. The number of scans over the maximum window extends throughout all the transaction entries of the database but repeated transactions beyond the required set of association rules are omitted. Transactions scanned in the interval length would generate the extended transactions sets. From the extensions, only the probable list of entries exists as the others are absent in further extensions.

Step 4 Depictions of an Itemset

The itemsets generated and frequencies derived can be illustrated as a FP tree. The Frequent Pattern is a condensed format of an original database of huge volume. FP tree maintains only the id of transactions, which directs to the locations of transactions in the storage space. The processes of representing the FP tree are instructed as two main phases.

1) Scanning the transactional database and ids of transactions
2) Finding the frequency of each transaction with respect to correlation of ids.

The input of the FP tree would be the transactional database, with the short and indexed form of representation with their corresponding ids[3]. The root of each intra-transactional domain is marked to be the general ancestor and based on the following nodes or branches the ids of transactions are registered into the tree. Descending from the root ancestral node, every other node possesses some specificity to the bottom level of the tree. The second phase of the FP tree construction completes the process of mining the transformed transactional database by detecting the frequency of each transaction and determining the association rules between them. The FP tree represents the frequency of respective transactions in form of a prefix. This prefix points the ancestor of a particular transaction from the taxonomy defined in previous phases. Compression of a database into smaller units conserves the on demand resources of memory and execution. With the view of conserving the storage space and computational cost, there should be no mishaps in the reliable content at the end of tree formation. All the transactions have to be included with some degree of memory space lesser than previous methodologies. The intended condensed format ensures that no valuable transactions are left out during the compression stages.
There is another strict condition framed for reducing the size of memory used for the case of longer transactions or patterns. As the long patterns would incur longer search executions and they are reduced to smaller recursive patterns with a suffix link provided to next occurrence of longer sequences. Adding to the condensed structure, this proposed methodology increases the efficiency by trimming down the computations. The same could be used to represent conditional databases, with condition based prefixes. Any number of transactions obeying a condition can be denoted by a single condition facilitating multiple accesses. Conditional databases are even easier to be combined before the determination of support, confidence or patterns of frequency. A condition is abided by transactions or they disobey and grouping them is based on a binary output of yes or no to admittance of the condition. A conditional database originates with predefined constraints and threshold metrics of the organization. Considering a conditional database of a supermarket would be the list of products prohibited for sale to minors, in the context of stock exchange the certain condition to satisfy in order to predict the winners and losers of the working day. Each condition is checked with the transactions registered in the database producing the extended set of itemsets from a large set to categorize and order sets of transactions. The proposed algorithm is as follows:

```
retrieve()
for each domain in intertransactional set
check id(D Id, C Id, T Id)
condition
(check repetition <= LengthofInterval )
if (T Id > threshold, C Id > threshold, D Id >
threshold)
Generate candidateset      ()
Table( Frequent Itemsets)
Table( Frequent Itemsets)
::
Table( Frequent Itemsets)
discard( < threshold)
for each Frequent Itemset()
condition
(association between Item[i++] and Item[j++])
for each item (condition presence > minimal)
derive(rules)
depict()
for each domain in intertransactional set
generate(Frequent Pattern Tree)
End()
```

RESULTS AND DISCUSSIONS

The algorithm has introduced a number of optimality measures by designing a taxonomy, closed itemset frequency mining in inter-transactional domains and finally representations in a simpler way with the same functionality. The previous strategies handled similar methodologies but carried along some notable drawbacks.

The Apriori algorithm laid the foundation of all these derivate algorithms. The importance of data mining and mining association rules in multi domains has argued the intricate necessity of these actions. Stated algorithm has planned different stages of optimal performance and overcame many of demerits and limitations of existing methodologies in turn leading to the same set of results at a faster rate nut limited computations. From the obtained set of results, the number of transactions provided as the input of algorithm is the same. The rules framed from the same set have been improved with the rate of significance. Only strong association rules will be produced as the outcome based on the accuracy level of the initial definition of taxonomy and boundaries on number of contexts and transactions would divide the search space into smaller segments. Faster access to the functional area is facilitated by the first three steps of the proposed algorithm.The first three phases discusses the easier means of formation of taxonomy and closed itemsets for simplifications of intended associations rules to be mined. Assigning a maximum window size for removing extended set of transactions also lends its support to enhance the performance. Depiction of resultant sets of patterns in FP trees gives an understanding better than other modes of representations. On the whole, taxonomy describes the structure and organization of transactions and their classifications. Speaking about today's advancements and the increasing persistence on pattern finding in multidimensional, intertransactional domains requires quite a model for satisfying the requisites.

The supermarket dataset with 24 distinct items are used for the experimental evaluation of CBIT approach. The closed itemsets are identified with total size of 2 GB datasets with certain confidence and support level. The execution time is improved for CBIT with reduction of 2 times as compared to FITI and the memory utilization reduced nearly 2 to 3 times for the data size of 2 GB supermarket data set. The following table 1 shows the performance comparison of the proposed CBIT with the existing FITI approach in terms of execution time and memory usage.

Table 1: Performance Comparison of CBIT with FITI

No of Transactions	FITI Approach		CBIT Approach	
	Execution time (sec)	Memory utilization (bytes)	Execution time (sec)	Memory utilization (bytes)
250	40.4	8400	19.8	3800
300	42.1	9360	20.4	4320
350	43.3	10420	21.2	4980
400	44.2	11530	22.0	5315

CONCLUSION

The mining of association rules in large environments urges sophisticated strategies to be efficient, reasonable and reliable. The algorithm defined in this paper contemplates association mining by various prospects and conditions. Generation of intermediate candidate sets, representation of large databases, trees, and pruning techniques advocates the performance of association rules mining. Enhancements and alterations in these individual stages combine to form a predominant result sets. Ultimately the scope of data mining has been evolving everyday to achieve a scalable result better than previous measures.

REFERENCES

1. Agrawal R, Srikant R, Fast algorithms for mining association rules", Proceedings of International Conference on Very Large Data Bases, Santiago, Chile,pp. 487–499. 1994.
2. Anthony J, Lee T, Chun-Sheng Wang, Wan-Yu Weng, Yi-An Chen, Hue i-Wen Wu, An efficient Algorithm for mining closed inter-transaction itemsets, Elsevier Science Publishers ,Data And Knowledge Engineering , Volume 66, Issue 1, pp 68-91, 2008
3. Jian Pei, Jiawei Han, Runying Mao, CLOSET: An efficient algorithm for Mining Frequent Closed Itemsets, Proceedings of the 5th ACM-SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, pp. 11-20, 2000
4. Ming-Cheng Tseng, Wen-Yang Lin, Efficient Mining of Generalized Association Rules with Non-uniform Minimum Support, Data and Knowledge Engineering, Volume 62, Issue 1. pp 41 -64,2007
5. Tung A K H, Lu H, Han J, Feng L, Efficient mining of intertransaction association rules, IEEE Transactions on Knowledge and Data Engineering 15 (1), pp :43–56, 2003
6. Prakash S, Vijayakumar M, Parvathi RMS, A novel method of mining association rule with multilevel concept hierarchy, Int. J. Comput. Appl.(IJCA), pp.26-29,2011.
7. Saveetha P, Arumugam S and Kiruthikadevi K, Cryptography and the Optimization Heuristics Techniques, Int. Journal of Advanced Research in Computer Science and Software Engg, volume. 4, Issue.10, ISSN: 2277 128X, October 2014.
8. Gokulraj P and Kiruthikadevi K, Revocation and security based ownership deduplication of convergent key creating in cloud, International Journal of Innovative Research in Science, Engineering and technology. Vol. 3, Issue 10, October 2014.
9. Prakash, S. and Vijayakumar, M., An effective network traffic data control using improved Apriori rule mining, Circuits and Systems, Issue 10, Vol. 07, pp. 3162-3173, June 2016.
10. Saranya M and Nithya K, Campus Navigation and Identifying Current Location through Android Device to Guide Blind People, International Research Journal of Engineering and Technology (IRJET), Vol.02,Issue : 08,Nov 2015.
11. Sureshkumar V S, Chandrasekar A, Fuzzy-GA Optimized Multi-Cloud Multi-Task Scheduler For Cloud Storage And Service Applications, International Journal of Scientific & Engineering Research, Vol.04, Issue.3, pp-1-7, 2013.
12. Preethi, B.C. and Vijayakumar, M. A novel Cloud Integration Algorithm (CIA) for Energy Efficient High Performance Computing Applications in Big Data Multimedia Applications, Romanian Journal of Information Science and Technology, vol. 2, no.1, pp. 1-11, March 2018.
13. Vijayakumar M, Prakash s, An Improved Sensitive Association Rule Mining using Fuzzy Partition Algorithm, Asian Journal of Research in Social Sciences and Humanities, Vol.6,Issue.6, pp.969-981, 2016.
14. Prakash S, Vijayakumar M, Risk assessment in cancer treatment using association rule mining techniques, Asian Journal of Research in Social Sciences and Humanities,Vol.6,Issue.10, pp.1031-1037, 2016.
15. Prabhakar E, Enhanced adaboost algorithm with modified weighting scheme for imbalanced problems, The SIJ transaction on Computer science & its application, Vol.6, Issue.4, pp.22-26, 2018.
16. Suresh Kumar VS, Thiruvankatasamy S, Sudhakar R, Optimized Multicloud Multitask Scheduler For Cloud Storage And Service By Genetic Algorithm And Rank Selection Method, Vol.3, Issue.2, pp.1-6, 2014.
17. Prabhakar E, Santhosh M, Hari Krishnan A, Kumar T, Sudhakar R, Sentiment Analysis of US Airline Twitter Data using New Adaboost Approach, International Journal of Engineering Research & Technology (IJERT), Vol.7,Issue.1, pp.1-6, 2019.
18. Dhivyaa CR, Vijayakumar M, An effective detection mechanism for localizing macular region and grading maculopathy, Journal of medical systems, Vol.43, Issue. 3, pp.53-, 2019.
19. Ragunath V, Dhivyaa CR, Privacy Preserved Association Rule Mining for Attack Detection and Prevention, published in International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, Issue 1, March 2014
20. Nithya K, Krishnamoorthi M, KalamaniM, Tweets: Review of Micro-Blog Based Recommendation Systems (RS) for News Recommendation (NR), in International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-7 Issue-4S, November 2018. pp. 444-448

21. K Nithya, M Saranya, CR Dhivyaa, Concept Based Labeling of Text Documents Using Support Vector Machine, International Journal on Recent and Innovation Trends in Computing and Communication, vol. 2, no. 3, pp. 541-544, (2014).
22. Suresh kumar V S , Booma K, Suma K IOT Smart Classrooms, International Journal Of Science Technology and Humanities, Vol 6, pp: 1-6,Issue 1,2019
23. V.P., Karthik S. Journal of Applied Sciences Research VOL, 8, Issue 7, 2012.
24. Kannan R., Selvambikai M., Jeena Rajathy I., Ananthi S. Rasayan, A study on structural analysis of electroplated Nano crystalline nickel based thin films, Journal of Chemistry, Vol 10, issue 4, 2017.
25. Arunvivek G.K., Maheswaran G., Senthil Kumar S., Senthilkumar M., Bragadeeswaran T, Experimental study on influence of recycled fresh concrete waste coarse aggregate on properties of concrete International Journal of Applied Engineering Research" Vol 10, issue 11, 2015.
26. Krishna S.K., Sathya M. Usage of nanoparticle as adsorbent in adsorption process - A review International Journal of Applied Chemistry, vol 11, Issue 2, 2015.
27. Sudha S., Manimegalai B., Thirumoorthy P. A study on routing approach for in-network aggregation in wireless sensor networks, International Conference on Computer Communication and Informatics: Ushering in Technologies of Tomorrow, Today, ICCCI 2014".
28. Satheesh A., Jeyageetha V. Improving power system stability with facts controller using certain intelligent techniques, International Journal of Applied Engineering Research, Vol 9, no 23, 2014.
29. Sathya M., Kumar P.E., Santhi M., Muralidharan B. Removal of acid green 25 dye by using activated carbon prepared from passiflorafoetida[pac-mno2nc]nanocompositein batch adsorption-kinetic study, Rasayan Journal of Chemistry 2018 114
30. L. Santhi M., Kumar P.E., Sathya M. Rasayan, Batch adsorption and isotherm studies for the removal of malachite green and reactive red 4 dyes by using AC-MnO NC prepared from typha angustat Journal of Chemistry, 2018 114
31. Ashok V., Kumar N. Iranian, Determination of blood glucose concentration by using wavelet transform and neural networks, Journal of Medical Sciences, Vol 38, Issue 1, 2013.
32. Senniangiri N., Bensam Raj J., Sunil J."Effects of Temperature and Particles Concentration on the Dynamic Viscosity of Graphene-NiO/Coconut Oil Hybrid Nanofluid: Experimental Study" International Journal of Nanoscience 2019.
33. Dinesh Kumar T., Prabhakar E., Nandhagopal K. An enhanced ensemble classification algorithm (EECA) for airline services big data sentiment analysis, Journal of Advanced Research in Dynamical and Control Systems 2019, 11(7)
34. Rajeswari M., Sivamani C., Saveetha P., Thomas A., Krishna S. Ekoloji, Dispersal of traffic alert messages without missing the reliability for environmental protection, 2019 28(107)
35. Dhivyaa C.R., Vijayakumar M. An Effective Detection Mechanism for Localizing Macular Region and Grading Maculopathy, Journal of Medical Systems 2019 43 (3)
36. Nithya K., Krishnamoorthi M., Kalamani M. Tweets: Review of micro-blog based recommendation systems (RS) for news recommendation (NR), International Journal of Recent Technology and Engineering, 2019, 7(4).
37. P.B. Narandiran, T. Bragadeeswaran,, M. Kamalakannan, V. Aravind, Manufacture of Flyash Brick Using Steel Slag and Tapioca Powder, Jour of Adv Research in Dynamical & Control Systems, Vol. 10, No. 12, 2018, 527-532
38. R. Girimurugan*, N. Senniangiri,, K. Adithya, B. Velliyangiri, Mechanical Behaviour of Coconut Shell Powder Granule Reinforced Epoxy Resin Matrix Bio Composites, Jour of Adv Research in Dynamical & Control Systems, Vol. 10, No. 12, 2018, 533-541.