International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 - 5161

Available at http://inpressco.com/category/ijcet

Research Article

A Comprehensive Overview of ARM Algorithms in Real Time Inter Transactions

Kranti M. Jaybhay[Ȧ*] and R.V.Arggiddi[Ȧ]

[Ȧ]Computer Science Department, Walchand Institute of Technology, Solapur, India

Accepted 16 August 2014, Available online 25 Aug2014, **Vol.4, No.4 (Aug 2014)**

Abstract

Association rule discovery from large databases is one of the tedious tasks in data mining. Most of the previous studies on mining association rules are on mining intra transaction associations, i.e., the associations among items within the same transaction where the notion of the transaction could be the items bought by the same customer, the events happened on the same day, etc. Mining intertransaction associations has more challenges on efficient processing than mining intratransaction associations because the number of potential association rules becomes extremely large after the boundary of transactions is broken. In this paper, we reviewed the notion of intertransaction association rule mining given by algorithms such as EH-Apriori and FITI. FTTI generates many unnecessary combinations of items because the set of extended items is much larger than the set of items. Thus, In order to provide efficient approach of g-based intertransaction association rule mining is used, where this group of transactions are following certain constraints. This paper helps to do comparative study of algorithms which will help in analyzing the predictions on stock market data.

Keywords: *ARM Association rule mining, EH-Apriori, FITI, Granule-based Transactions.*

1. Introduction

A various data mining applications are used in predictive learning and discovering knowledge from large databases. Stock market prediction is one of the challenging areas of research nowadays because it involves many transactions within single day. Intratransaction and intertransaction of stock market can give highly important correlations among them which can be helpful information in forecasting nature of stock. Association rules mining are also used in providing the predictive associative rules from historical data of stock market. Thus financial information in the form of stock quotes time series from financial websites is used as datasets. So far lot of work is done on application of this technique in intratransaction or those transactions happening on the same traded day. To do analysis of stock market over two days, four days, weekly, quarterly or yearly we need to work with the Intertransaction. To find out the correlations among transactions happening periodically association rules mining involves great part. Association rule mining, the most important and well researched techniques of data mining. It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases.

1.1 Association rule mining

Association rule mining find out association rules that satisfy the predefined minimum support and confidence from a given database. The problem can be decomposed into two sub problems. First to find those itemsets whose occurrences exceed a predefined threshold in the database; those itemsets are called frequent or large itemsets. The second problem is to generate association rules from those large itemsets with the constraints of minimal confidence.
Let, one of the large itemsets is Tk, Tk = {I1, I2… Ik}, association rules with this itemsets are generated in the following way: the first rule is {I1, I2… Ik-1} ⇒ {Ik}, by checking the confidence this rule can be determined. Then other rules are generated by deleting the last items in the antecedent and inserting it to the consequent, further the confidences of the new rules are checked to determine the interestingness of them. This process is iterated until the antecedent becomes empty.

We will also investigate the application of Intertransaction association rules mining in stock price predication and the possibility of generalizing this method to futures market.

1.2 Intertransaction association rule mining

A two-dimensional intertransaction association rule example is *After McDonald and Burger King open branches, KFC will open a branch two months later and one mile away,* which involves two dimensions: time and space (Hongjun Lu ,Ling Feng and Jiawei Han, 2000). Mining intertransaction associations poses more challenges on efficient processing than mining intratransaction associations. Interestingly, intratransaction association can be treated as a special case of intertransaction association from both a conceptual and algorithmic point of view.

*Corresponding author: **Kranti M. Jaybhay**

Table No. 1

Year	Researchers	Association Rules Extraction (mining) Criteria
1994	Agrawal and Shrikant	Apriori like methods
1994-1997	Klemettinen *et al.*	Generalized, multilevel , or quantitative association rules are extracted
	Savasere	
	Park *et al*.	
	Toivonen	
	Zaki	
1995-1997	Han and Fu	Association rule mining query languages
	Fukuda et	
	Miller and Yang	
	Kamber *et al*	
1997-1998	Brin *et al*. and Silverstein *et al*.	Cyclic and interesting association rule mining
2000	Han, J. and Pei, J.	FP-Tree frequent pattern searching by pattern growth
2002	Wojciechowski, M., Zakrzewicz, M.,	Dataset Filtering Techniques in Constraint-Based Frequent Pattern Mining
2005	Chuang, K., Chen, M., Yang, W.,	FDM algorithm is a parallelization of Apriori to (shared nothing machines, each with its own partition of the database.)
2009	Wanzhong Yang	Granule based knowledge representation for Intra and Inter transactions association mining
2011	Kerana and Rangaswamy	Efficient partition algorithm using clustering to reduce frequent itemsets to improve Apriori
2011	Kumar and Chandrashekharan	Attribute correction using data mining which follows two approaches, clustering techniques for context independent correlation and association rule for context-dependent correction
2012	Argiddi and Apte	Fragment based mining approach which focuses on minimizing the length of the transaction table of the stock market
2013	Kumar and Valli	RFPID-algorithm to mine regular frequent Patterns using vertical data format which performs better with large number of transactions and long Item sets with one database scan

In this study, we have introduced the notion of multidimensional intertransaction association rules, study their measurements—*support* and *confidence*—and develop algorithms for mining intertransaction associations by extension of Apriori.

The problem of intertransaction association rule mining explored by Anthony in 2003.The difference between the traditional association rule for intratransaction and intertransaction association rule can be stated as the following:

R2: ‘‘When the prices of A and B go up, the price of C will increase on the same day with probability of 80%.’’ However, stockjobbers may be more interested in the following rule.

R3: ‘‘If the prices of A and B go up on the first day, the price of C will increase four days later with probability of 80%.’’

Above classical association rules, like R2, discover the relationship among items within the same transactions, called as Intratransaction while R3 expresses association among items of different transactions along certain dimension, called Intertransaction.

The remainder of this paper is organized as follows: In section 2, Literature review is presented. In section 3 we describe the problem of Intertransaction association rule mining in general. In section 4 and 5, we will discuss algorithms like EH-Apriori traditional approach, FITI algorithm and group-based intertransaction association rule mining and their comparative study. And finally section 6 concludes the paper.

2. Related Work

In 1993 Agrawal *et al*. introduced the problem of mining association rules, since large amount of work is done been done in various directions including efficient, Apriori-like mining methods.

In 2006 Sotiris Kotsiantis presented a recent review about work in association rules mining. In addition to the basic Apriori algorithm's mining methods various new criteria for rules extraction are in use.

3. Methodology

3.1 Problem Description

Let B = {t_1, t_2 ... t_n } be a transaction database, and each transaction is a set of items. Tung in 1999 used the sliding window and extended-items to describe the intertransaction (see Definition 1).

Definition 1: A sliding window W in a transaction database B is a block of w continuous intervals along domain D, starting from interval d0 such that B contains a transaction at interval d0. Each interval dj in W is called a sub window of W denoted as W[j], where j= dj -- d_0. We call j the sub window number of dj within W.

Each sliding window W can be viewed as a continuous ω (a fixed interval called maxspan, or sliding_window_length) sub-windows such that each sub-window contains only one transaction. Let $e_i$be an item, its occurrences in different transactions in a sliding window can be extended from $e_i(0)$ to $e_i(\omega)$, where 0, ω are positions of transactions in the window. The transactions in a sliding window W can be merged into a *megatransaction* (or extended transaction) by putting all of W's extended items in a collection. Hence, an inter itemset refers to a set of extended-items, and an inter association rule can be represented as $X \rightarrow Y$, where X and Y are both a set of extended-items and $X \cap Y = ø$. The definition of the support and confidence in inter association mining {b[0], d[0]} ⟹ a[2] (support=0.4, confidence=1) (Anthony,2003) follows up the intra association mining. Let N be the number of megatransactions and, X and Y both be a set of extended-items and $X \cap Y = ø$. Let Txy be the set of megatransactions that contains X and also Y, and Tx be the set of megatransactions that contains X [3].We have,

$Sup(X \rightarrow Y) = |Txy| / N$,
$Conf(X \rightarrow Y) = |Txy| / |Tx|$ [3]

Example let, the five transactions are located at intervals 1, 2, 4, 5, 6. Let w=4, we now have five sliding windows W1, W2, W3, W4 and W5, with addresses of 1, 2, 4, 5, 6, respectively. Each window contains 4 subwindows. For example, W1 has four subwindows W1[0] (with items a, b), W1[1] (with items b, d), W1[2] and W1[3] (with items a, b, c, d). Each sliding window forms a megatransaction, which is the itemset of all the items in one sliding window. In our case, the megatransaction in W1 is {a[0], b[0], b[1], d[1], a[3],b[3],c[3],d[3]}. Thus, we have, Σ ={ a1[0], b1[0], b1[1], d1[1], a1[3], b1[3], c1[3], d1[3], b2[0], d2[0], a2[2], b2[2], c2[2], d2[2], b2[3], c2 [3], a3[0], b3[0], c3[0], d3[0], b3[1], c3[1], a3[2], b4[0], c4[0], a4[1], a5[0]}. Then, after we set the two essential parameters minsup (denotes minimum support level) and minconf (denotes minimum confidence level), we can mine intertransaction association rules from the transaction database. For instance, setting minsup=0.4 and minconf=0.8, we can get one rule mined from Table 1 (see Figure 1),

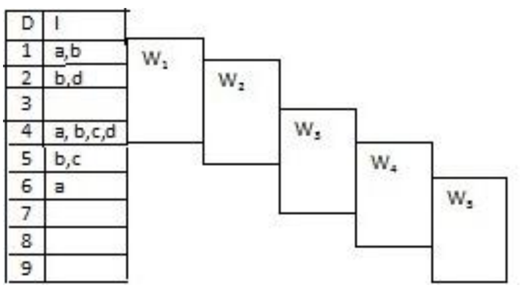

Figure1ARM Intratransaction

4. Association Rules Mining Algorithms

4.1 EH-Apriori

To discover frequent intertransaction itemsets with the existence of a megatransaction within each sliding window leads to use the extended Apriori algorithm. Apriori algorithm assumes the existence of lexicographic order among the extended-items, the extended-items in the megatransaction will be ordered using Definition 1 [1].

Definition 1. Let e_i (d_i) and e_j (d_j) be two extended-items in a megatransaction. We say that $e_i (d_i) < e_j (d_j)$ if either of the following two conditions holds:

1. $d_i < d_j$
2. $d_i = d_j$ and $e_i < e_j$.

Otherwise, we say that $e_j (d_j) < e_i (d_i)$ [1].By using this definition, the Apriori algorithm is used to discover frequent intertransaction itemsets. To enhance the efficiency further, a hashing technique similar to the one in is used.

The general idea of the algorithm is as follows:
When the support of candidate intertransaction one itemsets is counted by scanning the database, information about candidate intertransaction two-itemsets is collected in advance in such a way that all the possible two-itemsets are hashed to a hash table. Each bucket in the hash table consists of a number to represent how many itemsets have been hashed to this bucket so far. The hash table is then used to reduce the number of candidate intertransaction two-itemsets. This is done by removing a candidate two itemsets if its corresponding bucket value in the hash table is less than minsup. We call this algorithm the Extended Hash Apriori or EH-Apriori (Anthony, 2003).

4.2 FITI [First Intra Then Inter] algorithm

Unlike EH-Apriori which is modified from the Apriori algorithm, FITI is an algorithm designed specifically for discovering frequent intertransaction itemsets (Hongjun Lu, Ling Feng and Jiawei Han, 2000).The problem of mining intertransaction association rules can be decomposed as follows:

1. Find all Inter-transaction itemsets with support higher than minsup. They are called as **Frequent Inter-transaction itemsets.**
2. For every frequent Inter-transaction itemset F and for all possible combination of X C F,

output a rule X ⟹ (F - X) if its confidence is higher than minconf (Hongjun Lu ,Ling Feng and Jiawei Han 2000).
FITI makes use of the following property to enhance its efficiency in discovering frequent inter-transaction itemsets (Anthony, 2003).
Property 1. Let F be a frequent intertransaction itemset.
Let, $A_i = \{e_j \,|1 \le j \le u,\, e_j\,(i) \in F\}$
For all i, $0 \le i \le (w-1)$ and Ai must be a frequent intratransaction itemset
Proof. Prove this property by contradiction. Let F be a frequent intertransaction itemset.
Let $A_i = \{e_j \,|1 \le j \le u,\, e_j\,(i) \in F\}$ for all, i, $0 \le i \le (w-1)$
Assume that ∃ Ai, such that Ai is not a frequent intratransaction itemset. Denote the support of support.F. and support of Ai as support (Ai) Since F frequent intertransaction itemset, support(F) ≥ minsup. Also, since Ai is not a frequent intratransaction itemset, then support (Ai) < minsup.
Hence, we have
support(F) > support(Ai)

However, as we know that for any sliding window W that contains F, Ai will occur in W[i] and each W[i] refer to a different transaction for different W. It can thus conclude that support(Ai) >= support[F] giving a contradiction and thus proving that Property 1 holds.

This property provides a different view of mining frequent intertransaction itemsets. Instead of viewing mining as an attempt to identify frequently occurring patterns formed from the extended items, we can view it as an attempt to discover frequently occurring patterns formed from frequent intratransaction itemsets.

As such in FITI, frequent intratransaction itemsets are first discovered and then frequent intertransaction itemsets are formed from them. This gives rise to the name of FITI which stands for **First Intra Then Inter.**

In FITI, in the first phase frequent intratransaction itemsets are discovered and stored in a data structure designed to facilitate the mining of frequent intertransaction itemsets in the later phase.

Each frequent intratransaction itemset is given a unique number called an ID. By using this ID as an index into the data structure, FITI is able to gather information on intratransaction itemsets quickly. To avoid the need to regenerate frequent intratransaction itemsets during the discovery of frequent intertransaction itemsets, the original database is transformed into another database that stores the IDs of frequent intratransaction itemsets presented in each transaction of the original database.

When mining the frequent intertransaction itemsets, each intertransaction itemset is represented as a tuple of w IDs. Using this encoding, they formulate two types of joins to generate candidates' intertransaction. (k+1)-itemsets from two existing frequent intertransaction k-itemsets (Hongjun Lu , Ling Feng and Jiawei Han 2000).

In general, FITI consists of the following three phases.

1. Phase I: Mining and Storing Frequent Intratransaction Itemsets
2. Phase II: Database Transformation
3. Phase III: Mining Frequent Intertransaction Itemsets

Phase I: Mining and Storing Frequent Intratransaction Itemsets

In this phase, frequent intertransaction rules are mined and stored in data structure called as FILT (Frequent – Itemset Linked Table). Many fast algorithms are developed to mine intratransaction rules which can be applied to this step as well. The space needed for FILT data structure is smaller as it only store intratransaction rules rather than Candidate itemsets (Hongjun Lu, Ling Feng and Jiawei Han 2000).

FILT data structures

1. Lookup Links: The data structure consists of an itemset Hash Table, with nodes linked by several kinds of links. Each frequent intratransaction itemset is assigned a unique ID number that corresponds to a row number in the itemset Hash Table. Each itemset is stored in a node pointed to by a lookup link from the corresponding row in the table,

Example. Let {a},{b},{c},{e},{a,b},{b,c},{b,c},{a, b,c} be the frequent intratransaction itemsets derived by Apriori Each is then inserted into FILT by with the node pointed to by corresponding lookup link .Here ID of {a} is 1 and ID of {a, b,c} is 8

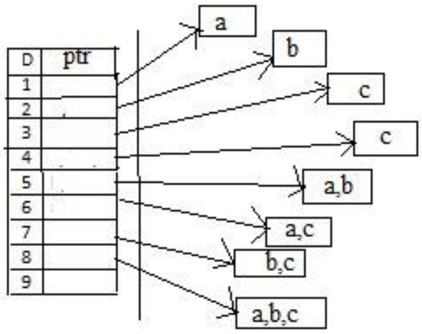

Figure 2 Lookup Links.

2. Generator and Extension Links: Given a node NF that contains an intratransaction k-itemset F, the generator links of NF point to the two (k-1) - itemsets that are combined to form F in the Apriori algorithm..The itemset {a, b, c} has its two generator links pointing to {a, b} and {a, c}.On the other hand {a, b} and {a, c} are combined to {a, b,c} both of them will have extension link pointing to {a, b,c}.Because of the nature of generator and extension links, they are depicted in the same diagram and the generator/extension relationship is represented by a bidirectional arrow. See Figure 3.

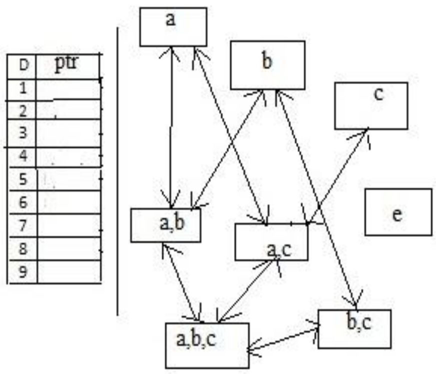

Figure 3 Generator and Extension Links

3. Subset Links. Given a node NF that contains an intratransaction k-itemset F, the subset links of NF point to all subsets of F with size k - 1. For example, the subset links of {a, b, c} point to {a, b}, {a, c} and {b,c} in Figure 4.

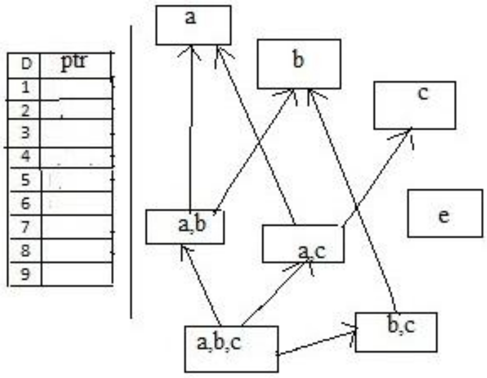

Figure 4 Subset Links

4. Descendant Links: FILT is composed of an array and a hash-tree. Given a node NF that contains an intratransaction k-itemset F, the descendant links of NF point to all of its descendants in the hash-tree. If F = $\{e_1 \ldots e_k\}$ then its descendants will be $\{e_1,..e_k$, ek+1} For example, descendents of {a} will be {a,b} and {a, c}. Unlike the subset links, the descendant links of a node points to other nodes that share a common suffix with it.

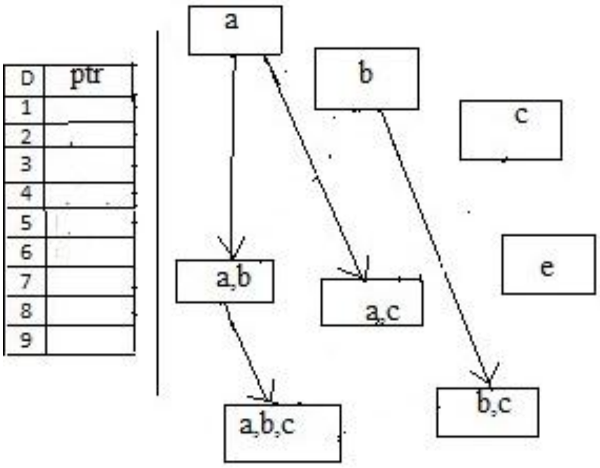

Figure 5 Descendant Links

Phase II: Database Transformation

After forming the data structure FILT, the next step of FITI is to transform the database into a set of encoded Frequent-Itemset Tables, (called FIT tables) We have in total max_k FIT tables, $\{F_1; \ldots ; Fmax_k\}$ where max_k is the maximum size of the intratransaction itemsets discovered in Phase I. Each table F_k will be of the form, $\{d_i, IDset_i\}$ where d_i is the value of the dimensional attribute and $IDset_i$ is the IDs of frequent k-itemsets that are found in the transaction.

Phase III: Mining Frequent Intertransaction Itemsets

After database transformation the next phase is to mine frequent intertransaction itemsets. Intertransaction itemsets are represented by their ID encoding form (Ping Li, Wenjing Xing, 2010).

Algorithm for Mining of frequent intertransaction itemsets

Input: A set of FIT tables: F1… Fmax, and the minimum support threshold: minsup.
Output: The complete set of frequent intertransaction itemsets
Generate frequent intertransaction 2-itemsets, L2; k=3;
While (Lk-1 ≠ Φ)
{
Generate candidate intertransaction k-itemsets, Ck;
Scan transformed database to update the count for Ck;
Let Lk ={c Є C_K support(c) ≥ minsup};
k++;
}

Algorithm for Generation of all the (k-itemset) subsets of an intertransaction (k+1)-itemset.

Let S be the set of k-subsets of I;
S= {};
For (p=0; p < w; p++)
{
if (Ip! = 0)
{
If (Ip is an intratransaction one-itemsets) {
If (p! =0)
Add {I0… Ip-1, 0… Iw-1} to S
Else
Add {I0… Iw-1, 0} to S
} else
{Let Ip be an intertransaction h-itemsets, h > 1
For each (h-1)-subset of Ip
{Let t be the ID of the (h-1)-subset
add {I0,…, Ip-1, t… Iw-1} to S}}}}
Return S;

4.3 Granule-based Intertransaction association rule mining (Wanzhong 2009)

Compared to FITI algorithm, granule based association mining can process many attributes in sliding window and so it avoids the restriction on interval number.

In real time applications long pattern involves so it makes easier intertransaction by avoiding interval numbers. It not only separates attributes into different tiers but also transforms intratransaction associations into intertransaction association by setting sliding window on decision attribute.

Though FITI outperforms than EH-Apriori in efficiency, it still generates many unneeded combinations that should be avoided. Hence we have alternative approach of group based intertransaction association rule mining, where this group is set of transactions that meet a certain constraint.

This method of group-based inter-association mining can reduce the complexity of inter-transaction mining, as it also reduces the width of sliding windows and uses this set to replace extended itemsets, thus there is no need to consider too many combinations of extended items.

5. Comparative studies of EH-Apriori and FITI

- EH-Apriori algorithm is based on Apriori approach to discover frequent intertransaction itemsets. Thus, uses BFS (Breadth First Search) like level by level search so at each level database must be scanned, it generates large number of candidate patterns at each level, and they are prone to memory shortage during mining process.
- The advantage of EH Apriori is it provides output of complete set of itemsets. The weakness about this algorithm is complexity in operations.
- The intertransaction associations are very complex, so in particular industry and when multidimensional parameters involve to apply Apriori is of no use.
- The feature of FITI uses idea of Apriori twice, to generate frequent set for intratransaction and for

Comparative Results

Parameters	EH- Apriori	FITI	Granule based Intertransaction algorithm
Technique	Use Apriori property and join and prune property	Uses Apriori property twice, to generate frequent set for intra and then for intertransactions	Uses Sliding window setting on decision attribute or constraints and uses SUM measure
Memory Utilization	Due to large number of candidate are generated so require large memory space	Due to compact structure and no candidate generation require less memory	Due to sliding window set only for decision attribute less memory requires
No. of scans	Multiple scans for generating candidate sets	Scans the database twice only	It also scans the database twice
Time	Execution time is more as time is wasted in producing candidates every time	Execution time smaller than Apriori	Efficient than FITI as it avoids cross join operations like FITI and takes less execution time
Real time applications	Rarely applicable	Mostly used in real applications	More applicable in real industry

intertransaction. FITI algorithm provides faster performance than EH- Apriori.

- Currently the FITI algorithm is the state of the art in intertransaction association rule mining.
- The advantage of FITI algorithm is it provides complete set of frequent itemsets however, however the FITI introduces many unneeded combinations of items because the set of extended items is much larger than the set of items.
- In particular industry such as Stock Market such itemsets are useless. The present form of FITI algorithm predicts along a single dimension it can be enhanced to n-dimensional intertransaction association-rules.

Conclusion and Future Work

Association rule mining is the most tedious task in the Intertransaction. In this paper we have reviewed EH-Apriori and FITI algorithms. These algorithms are compared according to their algorithmic structures and itemsets used, we found that FITI is much better than EH-Apriori. FITI algorithm produces many extra and meaningless rules and makes the process complex. Thus we have stated another technique called granule -based transactions to have efficient mining process.

In future, this method will be applicable to the Stock market data to predict and analyze the intertransaction rules and for knowledge discovery.

References

Anthony K.H. Tung, Member, IEEE, Hongjun Lu, Member, IEEE, Jiawei Han, Member, IEEE, and Ling Feng, Member, IEEE.(2003):Efficient Mining of Intertransaction Association Rules, *IEEE Transactions On Knowledge And Data Engineering*, VOL. 15, NO. 1.

Ping Li, Wenjing Xing, Huang Guangdong IEEE (2010): Financial Asset Price Forecasting Based on Intertransaction Association Rules Mining, *International Conference on E-Business and E-Government* 978-0-7695-3997-3/10

Wanzhong Yang, Yuefeng Li, Yue Xu (2009): Granule Based Intertransaction Association Rule Mining, Faculty of Information Technology *Queensland University of Technology.*

A. K.H. Tung, H.J. Lu, J.W. Han and L.Feng.(1999): Breaking the barrier of transactions: mining inter-transaction association rules, Proc. *ACM SIGKDD Intl. Conf. Knowledge Discovery and Data Mining.*

Sotiris Kotsiantis, Dimitris Kanellopoulos (2006):Association Rules Mining: A Recent Overview,*GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), 2006, pp. 71-82*

H. Lu, J. Han, L. Feng (1998): Stock movement prediction and n-dimensional inter-transaction association rules, in: *Proceedings of ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge, pp. 1–7.*

Hongjun Lu ,Ling Feng and Jiawei Han (2000): Beyond Intratransaction Association Analysis: Mining Multidimensional Intertransaction Association Rules , Hong Kong University of Science and Technology, Tilburg University, Simon Fraser University *ACM Transactions on Information Systems, Vol. 18, No. 4, October 2000, Pages 423–454.*

Han, J. and Pei, J (2000): Mining frequent patterns by pattern-growth: methodology and implications, *ACM SIGKDD Explorations Newsletter 2000, 2, 14-20.*

M. H. Dunham(2005):Data Mining Introductory and Advanced Topics, *India–Pearson Education.*

J. Han and M. Kamber(2006):Data Mining: Concepts and Techniques , *2nd ed. Morgan Kaufmann.*

D.Kerana Hanirex Dr.M.A.Dorai Rangaswamy(2011): Efficient Algorithm for Mining Frequent itemsets using Clustering techniques, *D.Kerana Hanirex et al. International Journal on Computer Science and Engineering (IJCSE) ISSN : 0975-3397 Vol. 3 No. 3*

R Kavita Kumar and Dr. RM.Chandrasekharan (2011):Attribute Correction-Data Cleaning using Association Rule and Clustering Methods, *International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.1, No.2.*

Rajesh V. Argiddi, Sulabha S. Apte,(2012) : Fragment Based Approach to Forecast Association Rules from Indian IT Stock Transaction Data , *Rajesh V. et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012,3493-3497*

Vijay Kumar G., Valli Kumari V., Incremental Mining for Regular Frequent Patterns in Vertical Format ,Vijay Kumar G. *et al*., *International Journal of Engineering and Technology* (IJET) ISSN : 0975-4024 Vol 5 No 2 Apr-May 2013