ISSN:0975-9646
Ranjeetsingh B.Parihar et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012,3895-3899

Combined Intra-Inter transaction based approach for mining Association among the Sectors in Indian Stock Market

Ranjeetsingh B.Parihar , Rajesh V. Argiddi , Sulabha S.Apte
Computer Science Department, Walchand Institute of Technology, Solapur, India.

***Abstract*— The previous work is carried out on windows width for mining inter-transaction rules. In this paper we present the approach to find out association within inter-transaction with different window length. This approach first mines rules among same transaction and then find out rules among different transactions with different window length is called as Combined Intra-Inter transaction approach This experimental work find out effect of window length and select the minimum window length which will best suited for processing huge amount of data . We conclude that this approach is promising one and will be suitable for predictions and useful in stock trading platforms for proper investment in Real Estate depend on finance sector.**

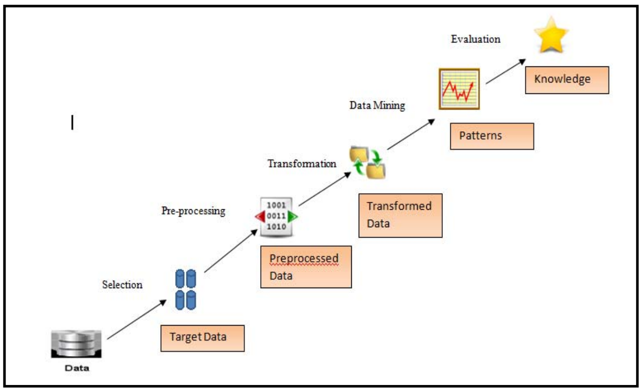
Keywords- Apriori; Combined Intra-Inter transaction mining; Stock Market; BSE.

INTRODUCTION *(HEADING 1)*

An informal definition of Knowledge Discovery in Databases (KDD) is to find useful and interesting patterns in data. Data mining is one of the tasks of KDD and is defined as a method to find a part of data which has interesting common features. Most of data mining methods that have been proposed to achieve the task try to find interesting patterns in database.

Data mining and knowledge discovery in databases (or KDD) is used as synonyms for each other but data mining makes use of algorithm to find out patterns in the knowledge discovery process. The KDD process generally involves a following processing steps, namely, data selection, feature-value selection, transformation, Data mining, Presentation and evaluation

KDD Process.



Classification is used to analyses the relationship between attributes and classes of objects in transaction table. Clustering is used to identify the classes where they also viewed as groups for set of objects whose classes are unknown. Association refers to discovery of associative relationship among objects [2].

Various techniques, such as statistical analysis, machine Learning, information theory and association rule mining have been used for extraction of knowledge in the literature. for our project we will concentrate on association mining.

Traditional association rule mining algorithms focus on association rules among item sets within a transaction. This classical association rule expresses the associations among items within the same transaction, thus we call it intra-transactional association rule. Most of work has been carried out on Intra-transaction association rule mining. Inter-transaction association indicates association among different transactions [3]. Work related to Inter-transaction association mining was proposed in 2000 and has a broad range of applications, though its basic idea extended from intra-transaction association mining. [4]

Stock market is very dynamic and nondeterministic in nature [12]. There are quick changes in market because of the underlying nature of the financial domain and following are the known parameters (Opening Price, Closing Price, P/E Ratio etc) and unknown factors (like Election Results, Rumors, suddenly changes like earthquake etc.) [8] which effects on market price.

In this research we have choose the original data sets of Bombay Stock Exchange (BSE) of different Banks such as Andra Bank, BOI, SBI,YES Bank etc from Yahoo Finance and try to find the association among the shares of Nationalized Bank and Private sector banks.

As we know that there are always some dependencies between different sectors in stock market[6]. Our aim is to find whether shares of nationalized banks affect the private sector banks. We collected the data of last 3 years for this experimental work

Some experimental results shows that there is a strong relation between nationalized and private sector banks and major of the times when the share price of nationalized banks goes up, share price of private sector goes up also it will work vice-versa.

Combined Intra-Inter transaction finds interesting associations between same transaction and afterword finds association in different transaction in the databases. Here we extend the scope of the Combined Intra-Inter transaction with varying the window size and find out the effectiveness in association rule mining. In this method we work out with window size denoted by ώ in the transaction database.

As in this, Combined Intra-Inter transaction we fixed such sliding window size the ώ which reduce the processing time for mining Inter transaction rules on data sets fed to the algorithm. In contrast to FITI, our Combined Intra-Inter transaction based approach, mining the association rules effectively within best time as well as with high confidence.

RELATED WORK

In the previous research, different data warehouse systems presented different techniques to support data mining; Ahmed et al. [9] presented the data warehouse backboned system integrated data mining and OLAP techniques. This system makes use of a router to adopt the previous mining result stored in the data warehouse, accordingly avoiding processing large amounts of the raw data. [8]

Initially rough set theory is used to describe knowledge in information table. Further, rough set theory based decision tables presented by Pawlak[13].

Wanzhong Yang [10] also presented a structure to disconnect the condition granules and decision granules in order to improve the efficiency of generating association rules from decision tables.

The discovery of inter-transaction association rules was first proposed with the E-Apriori and EH-Apriori algorithms. These Apriori inspired approaches make multiple passes over the database to find the set of frequent association rules [4].

The First Intra Then Inter (FITI) algorithm [5] is a more efficient E-Apriori-like algorithm that initially finds the complete set of frequent intra-transaction item sets as a basis for transforming the database into a structure that aids subsequent mining of the inter-transaction item sets

Our aim is to extend the work in this area and provide some improved methodology which should efficiently mine the rules.

BACKGROUND

A. Association Rule Mining

Association rule mining[3] focuses on searching for interesting relationships among items in a given data set. Association rule mining, also viewed as frequent pattern mining, refers to finding all frequent item sets and generating strong association rules from the frequent item sets.

a) Intra-transaction Association Mining

Inter transaction association mining refers to the associations among the same transaction, also called traditional association mining. The traditional association mining is classified into single dimension and multi-dimension.

b) Inter-transaction Association mining

The inter-transaction associations illustrate the associations among different transactions .One well-known example of multidimensional inter-transaction association is that of the fast food restaurants. If McDonald and Burger King open branches, KFC will open a branch two months later and one mile away.

Apriori Algorithm

This algorithm is developed by Agarwal and Srikant in 1994[7], which provide Innovative way to find association rules on large scale, allowing implication outcomes that consist of more than one item, Based on minimum support threshold.

The property of the Apriori [11] algorithm is that all non empty subsets of a frequent item set must be also frequent. It implies that no superset of any infrequent item set could be generated or tested. This property is used widely in data mining techniques. In particular, it is supported by the anti-monotone property. If a set cannot pass a test, all of its supersets fail the same test as well.

The algorithm attempts to find subsets which are common to at least a minimum number C (the cutoff, or confidence threshold) of the item-sets.

Apriori uses a “bottom up” approach, where frequent subsets are extended one item at a time a step known as candidate generation, and groups of candidates are tested against the data.

The algorithm terminates when no further successful extensions are found.

Apriori uses breadth-first search and a hash tree structure to count candidate item sets efficiently.

FITI(First Intra then Inter) Algorithm

In this approach first find out rules among the same transactions and then find out association rules among the different transactions.

The FITI algorithm [5] is based on the following property, a large inter-transaction item-set must be made up of large intra-transaction item-sets, which means that for an item-set to be large in inter-transaction association rule mining, it also has to be large using traditional intra-transaction rule mining methods. By using this property, the complexity of the mining process can be reduced, and mining inter-transaction association rules can be performed in a reasonable amount of time. First FITI introduces a parameter called maxspan (or sliding window size), denoted w. This parameter is used in the mining of association rules, and only rules spanning less than or equal to w transactions will be mined.

Second, every sliding window in the database forms a mega transaction. A mega transaction in a sliding window ώ is defined as the set of items ώ, appended with the sub window number of each item. The items in the mega transactions are called extended items.

T_{xy} is the set of mega transactions that contain the set of extended items X, Y, and T_x is the set of mega transactions that contain X. The support of an inter-transaction association rule X=> Y is then defined as”

Support = $|T_{xy}|$ /S, Confidence = $|T_{xy}|/|T_x|$

METHODOLOGY

There are some weaknesses in the previous FITI approaches such as time and space involved in processing the data is more. In FITI approach it is difficult to process an information table with many attributes and long intervals for inter transaction associations. This results into large amount of time and cost in processing the data.

Combined Intra-inter transaction mining approach work on the length of sliding window size used for finding effective association rules and try to avoid meaningless rules.

Indian stock market for bank Transaction Table

ID	Date	A	B	C	P	Q	R
1	1/1/2008	2613	70	612	1088	264	1739
2	2/1/2008	2649	73	620	1155	282	1825
3	3/1/2008	2625	77	700	1150	264	1725
4	4/1/2008	2590	81	700	1200	285	1733.5
5	7/1/2008	2560	78	680	1139	290	1729
	……	…	…	…..	…	…	….
982	29/12/2012	695	17	400	57	20	365.3
983	30/12/2012	702	16	401	56	18	370.55

Let T= {ID1, ID2, ID3,…, IDn} be a transaction database as shown in the Table 1.There are total 983 rows are consisted in the above transaction table. We have considered total 3 years of stock data for this work.

In this table A, B, C is the shares from Finance sector from BSE that represent Bajaj Holding, TCFC, RELIGARE, respectively.

Here P, Q, R are the shares of Real Estate that represents corresponding HDIL, UNITECH and SUNTECK. Here we are interested to find out to find out how Finance sector affects real estate.

Here share price refers only for the open price at the transaction data.

Here in this combined Intra-inter transaction algorithm we are varying length size of sliding window and finally fixed the window length ώ.

Our main aim is to identify the fixed length of sliding window of the transaction table and increase the performance to find out best association rules.

CONVERTED TRANSACTION TABLE

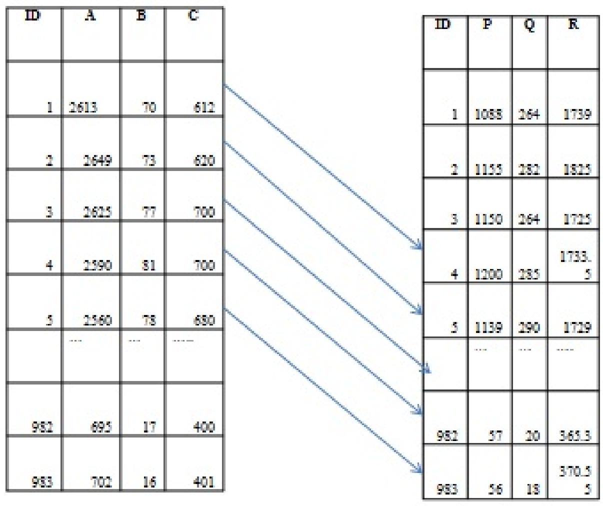
ID	Date	A	B	C	P	Q	R
1	1/1/2009	1	1	1	1	1	1
2	2/1/2009	0	1	1	0	0	0
3	3/1/2009	0	1	0	1	1	1
4	4/1/2009	0	0	0	0	1	0
5	7/1/2009	1	1	1	1	0	0
…	……	…	…	…	…	…	….
982	29/12/2012	0	0	1	0	0	1
983	30/12/2012	1	0	0	0	0	0

In above Table II we represented increase in share price by 1; otherwise decrease in price is represented it by 0. For that purpose we compared opening price of shares for two consecutive days.

A sliding window W for transaction table I is a block of ώ continuous intervals along time dimensions. In table III the transaction table is form of continuous sliding windows and window length is fixed as 4.

So, first rules are generated within same transaction and then within different transactions.

CONVERTED TRANSACTION TABLE WITH ATTRIBUTE



ID	A	B	C
1	2613	70	612
2	2649	73	620
3	2625	77	700
4	2590	81	700
5	2560	78	680
	---	---	----
982	695	17	400
983	702	16	401

ID	P	Q	R
1	1088	264	1739
2	1155	282	1825
3	1150	264	1725
4	1200	285	1733.5
5	1139	290	1729
	---	---	----
982	57	20	365.3
983	56	18	370.55

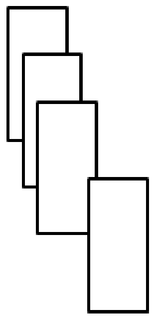
In above Table III we considered window length as four for finding inter-transaction rules. We choose only the price for Open as the value of the share price in this transaction. We ignore the change of each share in the same day. We consider each transaction date as one transaction for all shares. For each share in each transaction there is only one value. We apply support and confidence to control the data processing. In above Table IV, ID1 represents transaction one and ID 2 represent the transaction two.

In above transaction table (table III) we divided the attributes into condition attributes and decision attributes. Let C be the condition attributes where $C = \{a_1, a_2, \ldots, a_m\}$ and let D be the decision attributes where $D = \{a_{m+1}, \ldots, a_n\}$.

In below Table IV we considered window length as four for finding inter-transaction rules. A sliding window W can be viewed as a block of transactions with fixed intervals, which is called maxspan. All items in the sliding window can be viewed as extended items. Hence, an inter-transaction item set refers to a set of extended items.

TRANSACTION TABLE AND SLIDING WINDOWS WITH WINDOW LENGTH IS FOUR

ID	Date	A	B	C	P	Q	R
1	1/1/2009	1	1	1	1	1	1
2	2/1/2009	0	1	1	0	0	0
3	3/1/2009	0	1	0	1	1	1
4	4/1/2009	0	0	0	0	1	0
5	7/1/2009	1	1	1	1	0	0
...	......	...	...	...	...	...	....
982	29/12/2012	0	0	1	0	0	1
983	30/12/2012	1	0	0	0	0	0



EXPERIMENTS AND RESULTS

Combined Intra-Inter transaction Algorithm with sliding window length as four

In this method we have collected last 3 years data of Indian Finance sector and Real Estate from Yahoo Finance and converted that into a tabular format and applied Combined Intra-Inter transaction algorithm on that data set.

Input Data:

ID	Date	BAJAJ	TCFC	RELIGARE	HDIL	UNITECH	SUNTECH
1	1/1/2008	1	1	1	0	1	0
2	2/1/2008	0	1	1	1	0	0
3	3/1/2008	0	1	0	1	0	0
4	4/1/2008	0	0	0	1	0	1
5	7/1/2008	1	1	1	0	1	0
. . .							
981	28/1/2012	0	1	0	0	0	1
982	29/1/2012	0	0	1	0	0	1
983	30/1/2012	1	1	0	0	0	0

Output Association Rules Mining:

1. SUNTECK=1 491 ==> TCFC=0 276 conf:(0.56)
2. Religare=0 527 ==> TCFC=0 293 conf:(0.56)
3. Bajaj=0 520 ==> Religare=0 288 conf:(0.55)
4. UNITECH=0 524 ==> TCFC=0 290 conf:(0.55)
5. TCFC=0 530 ==> Religare=0 293 conf:(0.55)
6. UNITECH=0 524 ==> Religare=0 289 conf:(0.55)
7. SUNTECK=1 491 ==> Religare=0 270 conf:(0.55)
8. SUNTECK=0 488 ==> UNITECH=0 268 conf:(0.55)
9. HDIL=0 505 ==> UNITECH=0 277 conf:(0.55)
10. Religare=0 527 ==> UNITECH=0 289 conf:(0.55)

The first association rule shows that SUNTECK and TCFC has .56 confidence, that if share price of SUNTECK goes high (↑) then TCFC will also go low (↓).

And the 8th association rule shows that SUNTECK and UNITECH has .55 confidence, that if share price of SUNTECK goes low (↓) then UNITECH will also goes low (↓).The above kind of 488 transactions are consisted in the transaction table.

Combined Intra-Inter transaction Algorithm with sliding window length as six

After applying the Combined Intra-Inter transaction approach with window length six, now we apply the Apriori on this processed data and find the association rules among the attributes.

Input Data:

ID	Date	BAJAJ	TCFC	RELIGARE	HDIL	UNITECH	SUNTECH
1	1/1/2008	1	1	1	1	0	0
2	2/1/2008	0	1	1	1	0	1
3	3/1/2008	0	1	0	0	1	0
4	4/1/2008	0	0	0	1	0	1
5	7/1/2008	1	1	1	0	1	0
. . .							
981	28/1/2012	1	1	1	0	1	0
982	29/1/2012	0	1	0	1	1	0
983	30/1/2012	1	1	1	0	0	1

Output Association Rules Mining:

1. HDIL=0 501 ==> TCFC=0 292 conf:(0.58)
2. HDIL=1 472 ==> RELIGARE=0 268 conf:(0.57)
3. RELIGARE=0 524 ==> TCFC=0 292 conf:(0.56)
4. BAJAJ=1 456 ==> UNITECH=0 254 conf:(0.56)
5. UNITECH=1 454 ==> BAJAJ=0 252 conf:(0.56)
6. UNITECH=1 454 ==> TCFC=0 252 conf:(0.56)
7. BAJAJ=0 517 ==> RELIGARE=0 286 conf:(0.55)
8. TCFC=0 528 ==> RELIGARE=0 292 conf:(0.55)
9. TCFC=0 528 ==> HDIL=0 292 conf:(0.55)
10. UNITECH=1 454 ==> RELIGARE=0 250 conf:(0.55)

The first association rule shows that HDIL and TCFC have .58 confidences, that if share price of HDIL goes low (↓) then TCFC will also go low (↓).

And the 5th association rule shows that UNITECH and TCFC has .68 confidence, that if share price of UNITECH goes high (↑) then TCFC will also goes low (↓).

Combined Intra-Inter transaction Algorithm with sliding window length as eight

After applying the same approach on transaction data with window size eight we found different results.

Input Data:

ID	Date	BAJAJ	TCFC	RELIGARE	HDIL	UNITECH	SUNTECH
1	1/1/2008	1	1	1	0	1	0
2	2/1/2008	0	1	1	0	0	1
3	3/1/2008	0	1	0	1	1	0
4	4/1/2008	0	0	0	0	0	0
5	7/1/2008	1	1	1	1	0	0
. . .							
981	28/1/2012	1	1	1	0	0	1
982	29/1/2012	1	0	1	0	0	1
983	30/1/2012	0	1	1	0	0	0

Output Association Rules Mining:

1. SUNTECK=0 485 ==> RELIGARE=0 276 conf:(0.57)
2. HDIL=0 504 ==> TCFC=0 286 conf:(0.57)
3. UNITECH=0 521 ==> RELIGARE=0 293 conf:(0.56)
4. RELIGARE=0 525 ==> TCFC=0 293 conf:(0.56)
5. RELIGARE=0 525 ==> UNITECH=0 293 conf:(0.56)
6. SUNTECK=0 485 ==> BAJAJ=0 269 conf:(0.55)
7. BAJAJ=0 518 ==> RELIGARE=0 287 conf:(0.55)
8. TCFC=0 529 ==> RELIGARE=0 293 conf:(0.55)
9. HDIL=0 504 ==> RELIGARE=0 278 conf:(0.55)
10. SUNTECK=0 485 ==> TCFC=0 267 conf:(0.55)

Here we found that sliding window length means gaps between the transactions to find out inter-transaction rules. After careful observation we noticed it is better to consider latest transaction to mine the rules because the confidence is approximately equal while considering the window length. So, first approach with sliding window length four is best one among all.

CONCLUSION

The Experimental work shows after varying the sliding window length; that Combined Intra-Inter transaction approach is promising one for extracting some association rules in Inter-transaction association. It considers latest transactions to mine the rules, so suitable to choose minimum window length. It is efficient when there is huge amount of data also reduce processing time with best confidence. It will be suitable for predictions and useful in stock trading platforms for proper investments in different sectors and relation among different sectors.

REFERENCES

[1] Osmar R.Zaiane, "Principles of Knowledge Discovery in Databases",1999

[2] Han, J. and Kamber, M. Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2006.

[3] Agraw, R., Imielinski, T., Swami, A., "Mining association rules between sets of items in large database", Proceedings of ACM-SIGMOD, Montreal, Canada, 1993.

[4] Tung, A.K.H., Lu, H., Han, J. and Feng, L., "Efficient mining of intertransaction association rules" IEEE Transactions on Knowledge and Data Engineering, 2003.

[5] H. Lu, L. Feng, J. Han, Beyond intra-transaction association analysis: mining multidimensional inter-transaction association rules, ACM Transactions on Information Systems 18 (4)

[6] Dattatray P.Gandhmal, Ranjeetsingh Parihar,and Rajesh Argiddi "An Optimized approach to analyze stock market using data mining technique", IJCA, ICETT 2011

[7] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Proceedings of International Conference on Very Large Data Bases, Santiago, Chile, 1994

[8] Reuters, An Introduction to Technical Analysis. John Wiley & Sons (Asia) Pte Ltd., 1999.

[9] K. M. Ahmed, N. M. El-Makky, and Y. Taha (1998). Effective data mining: a data warehouse-backboned architecture. The 1998 conference of the Centre for Advanced Studies on Collaborative research, Toronto.

[10] Wanzhong Yang, "Granule Based Knowledge Representation for Intra and Inter Transaction Association Mining", Queensland University of Technology, July 2009.

[11] Professor Lee "Apriori Algorithm Review for Finals" Spring 2007.

[12] Eugene F. Fama "The Behavior of Stock Market Prices", The Journal of Business, Jan 1965

[13] Z. Pawlak . In Pursuit of Patterns in Data Reasoning from Data - The Rough Set Way. Rough Sets and Current Trends in Computing. J. F. P. James J. Alpigini, Andrzej Skowron, Ning Zhong. Berlin Heidelberg, Springer-Verlag ,2002