A Business Prediction system based on Granule Association Rule mining

R.V.Argiddi
Professor in WIT, Solapur Walchand Institute of Technology, AkklkotRoad ,Solapur.413003

H. M. Gaydhankar
Student in WIT, Solapur Walchand Institute of Technology, Akklkot Road ,Solapur.413003

A.S.Pawade
Student in Amina College of Technology JNTU University ,Hydrabad.

ABSTRACT
It has been widely accepted that association rules mining, the task of searching for correlations between items in a database, can discover useful rules in stock analysis. Among many techniques of data mining, association rule mining is widely accepted for finding co-relation between items in databases, which can be effectively used for forecasting the trends in stocks, future market analysis and setting strategies for super markets etc. Previous studies on association rule mining mainly emphasize on mining intratransaction associations.

Intertransaction association rule mining is used to discover patterns between different transactions. It breaks the scope of association rule mining on the same transaction. Currently the FITI algorithm is the state of the art in intertransaction association rule mining. However, the FTTI introduces many unneeded combinations of items because the set of extended items is much larger than the set of items. Thus, we propose an alternative approach of granule based intertransaction association rule mining, where a granule is a group of transactions that meet a certain constraint. The experimental results show that this approach is promising in real-world industry.

INTRODUCTION
The rapid development of knowledge discovery means that data mining plays an important role in terms of data preparation, data processing and knowledge representation [2]. Data mining stands in the central position, where data from a data warehouse can be processed into identified patterns for knowledge representation [3, 4, 5]. Data mining has four major tasks: summarization, classification, clustering and association mining [6]. Association rule mining is one essential aspect in data mining that remains relevant with the phenomenal growth of electronic data and information. However, it is difficult to get enough useful knowledge to support intelligent business through the current knowledge discovery process [7]. The quality of knowledge discovery seems to be the problem for this inefficiency and ineffectiveness [8, 9].

Association mining, which is widely used for finding association rules in single and multidimensional databases, can be classified into intra and inter transaction association mining. Intratransaction association refers to association in the same transaction; intertransaction association indicates association among different transactions [11]. Intertransaction association mining was proposed in 2000 [11] and has a broad range of applications, though its basic idea extends from intratransaction association mining.

Both intra and inter transaction association mining are expensive in terms of time cost in both phases of association mining. Though various algorithms have attempted to improve efficiency in finding frequent patterns and rule generation in multidimensional associations, there remains a gap between user requirements and achievements maintained. Extracting the knowledge for the intratransaction and intertransaction association mining is computationally expensive. Analysis of current bottleneck problems concentrates on finding how to discover the knowledge in databases efficiently to meet what users want.

Granule mining, proposed in intratransaction association mining [16], extends condition attributes and decision attributes in decision tables into condition granules and decision granules. It is difficult for decision tables to deal with the many attributes. Granule mining adopts multi-tier structures and association mappings to process the associations among condition granules and decision granules. It enables association mining to process many attributes in an information table. The multi-tier structure divides the condition granules or decision granules into several levels; so the large granule can be viewed as a group of small granules in different levels. Association mappings link granules in different levels into a network. The multi-tier structure can be described as a graph, where the nodes at each tier are granules and the links are the association mappings. To evaluate the association mining, the concept of general rules is presented, in which a general rule is an association rule with a short premise. This concept can be used to remove meaningless rules. Rule redundancy can be obviously reduced in the experiments based on this approach; efficiency and effectiveness can also be improved in association mining.

Literature Review / Survey
Intra association mining developed very well and has obtained many significant solutions, e.g. Apriori based approaches and FTP based approaches [18] [19].The inter association mining tries to find associations between items in different transactions.

Lu et al. [20] first proposed the concept of inter association mining and contributed E-Apriori and EH-Apriori algorithms to this area. To improve the performance, Feng et al. [21] presented a template model for this problem.

Moreover, Tung et al. [22] recently proposed the FITI (First Intratransaction Then Intertransaction) algorithm. In FITI, if the average size of the transactions is very large, the extended transactions should be very long. It generates many extra combinations of items because the set of extended items is much larger than the set of items. Thus, this method is very slow if the average size of the transactions is large.

To reduce the size of discovered knowledge in large databases, Pawlak [23] and Li and Zhong [24] used decision tables for association mining. Li et al. [25] also proposed the concept of granule mining for intra association mining recently. Granule mining is a new initiative that tries to improve the quality of discovered knowledge in databases.

Problem Statement

Design a Granule based association mining technique for business prediction system (for e.g.: stock market, future market etc) with the additional consideration of initial data filtering to achieve better performance in terms of time, cost and accuracy.

Proposed Work

Time cost and rule redundancy are the bottlenecks stopping the discovery of the interesting patterns in industrial applications. The previous methods, frequent pattern mining and rule generation, could not show any further potential in efficiency and effectiveness.

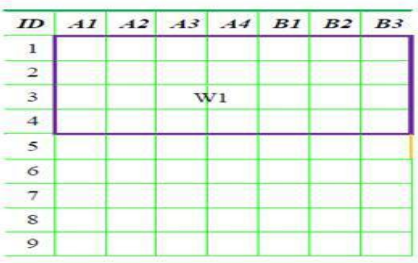
The sliding window approach is the current competitive tool in intertransaction association mining, but the width and length of a sliding window restrict an information table with many attributes and long intervals in efficiency and effectiveness. Previous frequent pattern mining in the sliding window approach was extended from the Apriori algorithm. The candidate set can generate a very large number of itemsets with itemset joins. It is hard to filter out the noise data and useless itemsets. Additionally, intertransaction associations are often related to the prediction in industry.

In order to use the sliding window approach to illustrate intertransaction associations directly and clearly, time and cost, rule filtering and redundancy must all be considered. In particular the approach should match the requirement of the prediction with the user requirements. Intertransaction associations are different from intratransaction associations.

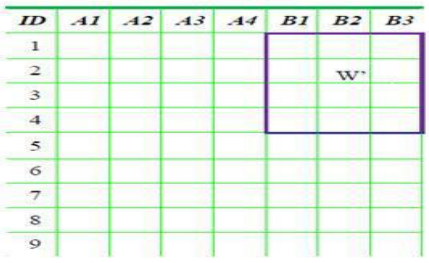
Instead of single item based frequent itemsets, granule mining finds interesting associations between granules in databases [13, 14], where a granule is a predicate that describes common features of a set of objects (e.g., records, or transactions) for a selected set of attributes (or items). For example, a granule refers to a group of transactions that have the same attribute values. Granule mining extends the idea of decision tables in rough set theory [15] into association mining. The attributes in an information table consist of condition attributes and decision attributes, with users' requirements. Condition attributes are associated to decision attributes: this can generate the decision rule directly. Granule mining applies this advantage in association mining, which attempts to solve the problem at the point of knowledge representation. The granule-based structure impacts the innovation in association mining at both its phases: frequent pattern mining and rule generation.

Granule-based intertransaction association mining attempts to solve the complex issue in intertransaction associations [17]. The intertransaction associations have broad ranges of applications but are very complex. The sliding window based approach is the most competitive method in the intertransaction association mining. Each sliding window contains all items among intertransaction within fixed intervals. This approach generates the frequent patterns from all sliding windows using traditional intratransaction association mining methods based on a given minimum support. The major advantage of the sliding window based approach is that intertransaction associations can be interpreted directly and clearly. The FITI algorithm has been the state of the art method for the sliding window based approach. However, it was extended from the Apriori algorithm based on the item joins: this causes very expensive time cost and generates many extra or meaningless rules if there are many attributes in an information table. Additionally, the intertransaction associations are often related to the issue of prediction in industry being able to meet users' needs.

In order to solve this problem in intertransaction association mining, this thesis presents a granule-based approach that reduces the width of the sliding window: for example, it may cover only decision attributes, where the width and the length of the sliding window refer to the number of attributes and the intervals of transactions, respectively. The granule- based approach is different from the FITI algorithm: it uses the sliding windows to separate attributes into tiers, and uses small sliding windows for certain tiers, based on users' constraints. In Figure, there are seven The experimental data source is divided into training sets and testing sets. Measures for evaluating the quality of intertransaction association rules are also proposed in this thesis. The idea is to test the attributes in an information table. The sliding window in the FITI algorithm includes all attributes; the number of intervals is four (in Figure (a)). The granule-based approach divides the attributes into two tiers: condition attributes (A1, A2, A3, A4) and decision attributes (B1, B2, B3). The sliding window is set up only on decision attributes and the number of intervals is four (in Figure (b)). This innovation can largely reduce the number of extended itemsets; therefore, people can use long intervals for intertransaction association mining in real applications.



(a) FITI algorithm



(b) Granule-based approach

The experimental data source is divided into training sets and testing sets. Measures for evaluating the quality of intertransaction association rules are also proposed in this thesis. The idea is to test the percentage of rules selected from the training set that occur in the testing set. This is the first application evaluation measure in intertransaction association mining in order to evaluate the effectiveness of intertransaction association mining. To analyze impacts of intratransaction

associations on intertransaction associations, experiments of correlation analysis on the dataset are also implemented. The experimental results clearly demonstrate that there is no relation between the data correlation of intra and intertransaction associations.

Objectives and Scope

Association mining consists of two phases [10].

- The first phase is finding frequent itemsets or frequent patterns among all transactions in databases, where an itemset is a subset of items. This phase is the core of association mining, but also creates the bottleneck for applications.
- The second phase is rule generation. Most algorithms focus on the first phase.

As the application is of granule mining in the intertransaction associations, the sliding window approach can be enhanced by reducing the width of the sliding windows.

The new approach consists of two main steps.

- The first step is to divide attributes into tiers.
- The second step is to define sliding windows in tiers.

All the objectives must satisfy all the following features of Granule based association rule approach.

- The granule-based approach reduces the time and cost of using sliding windows. It can make processing many attributes and long intervals for the intertransaction associations feasible.
- The measure, precision, is proposed to remove the meaningless or useless rules for the purpose of prediction in intertransaction associations.

Conclusion

Thus the proposed system which is to be built would be very much useful software that can make the all sort of forecasting's and analyzing of most of the business (stock exchange, future market etc) related task very much simpler, thereby making a lot of profit into the business.

References

[1]. Granule Based Intertransaction Association Rule Mining by Wanzhong Yang, Yuefeng Li, Yue Xu 19th IEEE International Conference on Tools with Artificial Intelligence 2007.

[2]. U. Fayyad, G. Piatetsky-shapiro, and P. Smyth (1996). "From data mining to knowledge discovery in databases‖ AI Magazine 17: 37—54

[3]. A. BONIFATI, F. CATTANEO, S. CERI, A. FUGGETTA, and S. PARABOSCHI (2001). "Designing data marts for data warehouses." ACM Trans. Softw. Eng. Methodol. 10(4): 452-483.

[4]. C. A. Hurtado, C. Gutierrez, and A. O. Mendelzon (2005). "Capturing summarizability with integrity constraints in OLAP." ACM Trans. Database Syst. 30(3): 854-886.

[5]. S. Parthasarathy (2007). Data mining at the crossroads: successes, failures and learning from them. Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. San Jose, California, USA, ACM.

[6]. Y. Fu (1997). "Data Mining Tasks, techniques, and applications." IEEE Potentials 16(4): 18- 20.

[7]. N. Zhong, J. Liu, Y.Y. Yao, and S. Ohsuga (2000). Web Intelligence (WI). The 24th IEEE Computer Society International Computer Software and Applications Conference

[8]. K. Wahlstrom and J. F. Roddick (2000). On the Impact of Knowledge Discovery and Data Mining The 2nd Australian Institute of Computer Ethics Conference Australia.

[9]. Y. Xu and Y. Li (2007). Generating concise association rules. Proceedings of the sixteenth ACM conference on Conference on information and knowledge management. Lisbon, Portugal, ACM.

[10]. J. Han and M. Kamber (2001). Data Mining Concepts and Techniques, Morgan Kaufmann Publishers.

[11]. H. Lu, J. Han, and L. Feng (2000). "Beyond intratransaction association analysis: miningmultidimensional intertransaction association rules." ACM Transactions on Information Systems 18(4): 423-454.

[12]. A.K.H. Tung, H. L., J. Han, and L. Feng (2003). "Efficient mining of intertransaction association rules." IEEE Transactions on Knowledge and Data Engineering 15(1): 43-56

[13]. Y. Li, W. Yang, and Y. Xu (2006). Multi-Tier Granule Mining for Representations of Multidimensional Association Rules. ICDM, HongKong.

[14]. W. Yang, Y. L., J. Wu, and Y. Xu (2008). "Granule Mining Oriented Data Warehousing Model for representations of Multidimensional Association Rules." International Journal of Intelligent Information and Database Systems 2(1): 125-145.

[15]. Z. Pawlak (1982). "Rough Sets." International Journal of Computer and Information Science 11(5): 341-356.

[16]. Y. Li, W. Yang, and Y. Xu (2006). Multi-Tier Granule Mining for Representations of Multidimensional Association Rules. ICDM, HongKong.

[17]. W. Yang, Y. Li, and Y. Xu (2007). Granule based Intertransaction Association Rule Mining. 19th IEEE International Conference on Tools with Artificial Intelligence Greece.

[18]. Agraw, R., Imielinski, T., Swami, A., ―Mining association rules between sets of items in large database‖, Proceedings of ACM-SIGMOD, Montreal, Canada, 1993, pp. 207-216.

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

[20]. Lu, H., Han, J. and Feng, ., ―Beyond intratransaction association analysis: mining multidimensional intertransaction association rules‖, *ACM Transactions onInformation Systems*, 18(4), pp.423 - 454, 2000.

[21]. Feng, L., Yu, J. X., Lu, H. and Han, J., ―A template model for multidimensional inter-transactional association rules‖, *The International Journal on Very Large Data Bases*, 11(2), pp. 153 -175, 2002.

[22]. Tung, A.K.H., Lu, H., Han, J. and Feng, L., ―Efficient mining of intertransaction association rules‖, IEEE Transactions on Knowledge and Data Engineering, 15(1), pp.43 – 56, 2003.

[23]. Pawlak, Z., ―In pursuit of patterns in data reasoning from data, the rough set way‖, *3rdInternational Conference on Rough Sets and Current Trends in Computing*, USA, 2002, pp. 1-12.

[24]. Lu, H., Han, J. and Feng, ., ―Beyond intratransaction association analysis: mining multidimensional intertransaction association rules‖, *ACM Transactions on Information Systems*, 18(4), pp.423 - 454, 2000.

[25]. Li, Y., Yang, W. and Xu, Y., ―Multi-Tier Granule Mining for Representations of Multidimensional Association Rules‖, *6th IEEE International Conference on*