ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

Survey for Rule Pruning in Association Rule Mining for Removing Redundancy

Ashwini Batbarai[1], Devishree Naidu[2]

P.G. Student, Computer science and engineering, Ramdeobaba College of engineering and management, Nagpur, India[1].

Assistant Professor, Computer science and engineering, Ramdeobaba College of engineering and management, Nagpur, India[2].

Abstract: association rule mining is an important component of datamining. Association rule mining is used in many applications for decision making. It is used in many applications like analysis of market data, purchase histories, web log. These applications having large dataset which give large no of association rule which used for decision making. Many of which are redundant rule. In this paper we have survey many methods for removing redundancy.

Keywords: Association rule, Redundant rule, Itemset mining, Non-redundant rule.

I. INTRODUCTION

Association rule mining is a popular method for finding the interesting relation between variables in large databases. It is intended to identify strong rules discovered in database using different interestingness measure of association rule mining. There are many interesting measures present for finding the association rule. It can use depending on the different application. Based on concept of strong rules, association rules are introduce in supermarket for discovering the regularities between the product in large-scale transaction data which recorded by point of scale system (POS). . For example, the rule {onions, potatoes}→{burger} found in the sales of supermarket would indicate that if customer buys onions and potatoes together they likely to also buy burger. Such information can be used as the basis for decision about marketing activities such as, e.g. promotional pricing or product placements. In addition to market basket analysis association rules are employed today in many application areas like web usage mining, intrusion detection, continuous production, and bioinformatics. As opposed to sequence mining, association rule learning does not consider order of items either within a transactional or across transaction.

Association rule mining is defined as: Let $I=\{i_1,i_2.....i_n\}$ be a set of n binary attributes called items. Let $D=\{t_1,t_2, \ldots,t_n\}$ be a set of transactions called the database. Each transaction in D has a unique transaction ID and contains a subset of the items in I. A rule is defined as an implication of the form X ➔ Y where $X,Y \subseteq I$ and $X \cap Y= \varnothing$. The sets of items X and Y are called antecedent (left-hand-side or LHS) and consequent (right-hand-side or RHS) of the rule respectively.[11]

Association rule mining has two steps for generating rule. First all the frequent itemset are mined from the dataset using the minimum support threshold level. Then rules are generated by combining subsets of every frequent itemset in all possible way. The rules which satisfy the minimum support and confidence threshold level are generated. These rules are the valid association rule.

TID	ITEMS
1	{bread**,** milk**}**
2	{bread, diaper, cola}
3	{bread, milk, diaper}
4	**{**bread**,** milk , diaper, cola}

Bred→ milk sup=0.75% and conf=0.75%
Bred→ diaper sup=0.75% and conf=0.75%

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

Bread →milk diaper sup=0.50% and conf=0.50%
Milk →bread sup=0.75% and conf=0.75%

These are the some rule generated from given transaction dataset base on minimum support =50% and confidence=50% threshold. In this way association rule predict the occurrence of one item based on occurrence of another item.

As the rules which satisfy minimum threshold level are valid rule but some of these rules give the same information as other rule is giving. These rules called the redundant rules.

Redundant Association Rules: Definition: Let X →Y and X'→ Y' be two rules with confidence cf and cf', respectively. X→ Y is said a redundant rule to X' →Y' if X belong to X'; Y' belong to Y, and cf ≤ cf' [4].

From the above example third rule is redundant rule which is combination of Bred→ milk sup=0.75% and conf=0.75% and Bred→ diaper sup=0.75% and conf=0.75% and giving same information as these two rules are giving. Many types of redundancy come. Rule which is the combination of valid rule are the redundant rule. Rule which form by only changing the antecedents and consequents are redundant rule. In the above example rule Milk → bread sup=0.75% and conf=0.75% is formed by only changing the antecedents and consequents is redundant rule. Due to this type of redundancy accuracy of decision making will affect so there is need to reduce the redundancy. In this paper we have survey many approaches for removing redundancy.

II. RELATED WORK

Bastide et al. [1] this approach is based on frequent close itemset mining which using the semantic for extraction of association rule based on closure of Galois connection, the generic basis for exact association rule and informative basis for approximate association rule. It constructed using frequent closed itemset and generators .They is using A **Close** algorithm for frequent closed itemset mining and their generators. It generate the rule which having minimal antecedents and maximal consequents. It gives the user the set of rules covering all the attributes of dataset i.e. containing the rule where union of antecedent (resp. consequents) is equal to the union of antecedent (resp. consequents) of all association rules valid in context. Generating the rules from frequent itemset gives large number of rules. In this frequent closed itemset mining is used which generate the valid minimum rule and efficient than the frequent itemset. In the above example { bred ,milk ,diaper } is frequent close itemset having sup=50%. From this we generate rules by combining subsets of every frequent closed itemset in all possible way. We get 2^n-2 rules where n is number of items in frequent closed itemset. We get 2^3-2 rules six rues. some of which are {bread → milk, diaper} sup=0.50% and conf=0.50% ,{bread, milk → diaper} sup=0.50% and conf=0.75%, {milk → diaper, bred} sup=0.50% and conf=0.75% ,{milk, diaper → bread} sup=0.50% and conf=100%, { diaper → bread, milk } sup=0.50% and conf=0.75%, {diaper, bread → milk } sup=0.50% and conf=100%.Here all the rules are valid rules then after using minimal antecedents and maximal consequents we get rule {milk → diaper, bread} sup=0.50% and conf=0.75%,{ diaper → bread, milk } sup=0.50% and conf=0.75% as non-redundant rule.

Ashrafi et al [2] the proposed methods not only remove redundant rules generated from frequent itemset but also remove redundant rules generated from the frequent closed itemset**.** The proposed methods are not based on any bias assumptions. It verifies all rules that have one or more items in the consequence. Therefore, it has the ability to eliminate redundant rules that contain single or multiple items in the consequence. Example: by using this approach we can reduce following type of redundancy which having fixed antecedent so third rule is redundant rule which having fixed antecedent as first two rule.

Bred→ milk sup=0.75% and conf=0.75%, Bred→ diaper sup=0.75% and conf=0.75%, Bread →milk diaper sup=0.50% and conf=0.50%.

David Lo et al [3] in proposed method several rule sets based on composition of various types of pattern sets namely generators, projected-database generators, closed patterns and projected-database closed patterns. This set evaluated based on the two criteria of completeness and tightness and used as a composite filter , replacing a full set of rules with non redundant subset of rules dose not impact the accuracy of filter. The rule set which contain all rules and having the rule which infer the other rule called complete rule set. The rule set which not contain any redundant rule called tight rule set.

Zaki et al [4] present the framework for generating non-redundant association rule. They are using frequent closed itemset mining using the charm algorithm and generating the rule; then reducing the redundancy using transitivity rule. E.g. {TW→A, TW→AC, CTW→A} in this set all rule having same confidence as 1then put rule TW→A as interesting rule and remove other as redundant rule, because its only adding the items in the antecedent and consequent side items.{A→W,A→CW,AC→W} in this set all rule having same confidence hence put A→W and remove other as

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2014

redundant rule, but rule A→W is not fully characterized the knowledge of rule AC→W so in this we are removing the interesting rule so we are not getting accurate rule set. .
Philippe Fournier-Viger et al [5] this gives the TNR algorithm for removing redundancy. In real time selecting the parameters to generate a desired amount of rules is usually difficult and time-consuming and generate large amount of redundancy in the results. These problems are addressed using TNR algorithm but it costly. Redundancy can be removed by interesting measures. It measures how strongly one attribute implies other based on available data. Many objective measures are present like lift, convention which can use for getting interesting rule depending on application [6],[7],[8].

III. CONCLUSION

In this paper we have study how redundancy come and survey many approaches for removing redundancy. As association rule mining used in many application area like web usage mining, supermarket analysis which having large dataset. For these applications we get large amount of rule some of which will be redundant and give same information which can affect on the accuracy of decision making. As we survey all the paper in that some type of redundancy removed but not getting the interesting rule with non-redundant set. So there is need to generate the interesting rule set which can be understand by end user and which is interesting rule from which can make good decision.

REFERENCES

[1] Bastide Y., Pasquier N., Taouil R., Stumme G., Lakhal L., "Mining Minimal Non-Redundant Association Rules using Frequent Closed Itemsets", 1st International Conference on Computational Logic, pp. 972 – 986,2000.
[2] Mafruz Zaman Ashrafi, David Taniar, Kate Smith "New Approach of Eliminating Redundant Association Rules" 15th International Conference, DEXA2004, Zaragoza, Spain, Proceedings pp 465-474,2004,August 30-September 3, 2004.
[3] David Lo , Siau-Cheng Khoo , Limsoon Wong , "Non-redundant Sequential Rule Theory and Algorithm '' School of Information System, Singapore Management University ,2008.
[4] Mohammed J. Zaki, "Generating Non-Redundant association rule KDD'00 Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. Pages 34-43 ,ACM New York, NY, USA ©2000.
[5] Philippe Fournier-Viger, Vincent S. Tseng "Mining Top-K Non-Redundant Association Rules" 20th International Symposium, ISMIS 2012, Macau, China, December 4-7, 2012.
[6] Pang -Ning Tan, Vipin Kumar, Jaideep Srivastava "Selecting the Right Interestingness Measure for Association Patterns"KDD'02 Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining Pages 32-41 ACM New York, NY, USA ©2002 .
[7] P. D. McNicholas, T. B. Murphy, M. O'Regan "standardising the lift of an association Rule" computational statistics and Data Analysis, volume 52,Issue 10, pages 4712-4721,15 June 2008.
[8] Paulo J. Azevedo, and Alıpio M. Jorge2 "Comparing Rule Measures for Predictive Association Rules" 18th European Conference on Machine Learning, Warsaw, Poland, pp 510-517 Proceedings 2007, September 17-21, 2007..
[9] Jose L. Balcazar "Redundancy, Deduction Schemes, and Minimum size bases for Association Rules" Logical Methods in Computer Science, Vol. 6 (2:3) 2010, pp. 1–33 Jun. 22, 2010.
[10] Bay Vo and Bac Le " A Frequent Closed Itemsets Lattice-based Approach for Mining Minimal Non-Redundant Association Rules" International Journal of Database Theory and Application Vol. 4, No. 2, June, 2011.
Book reference:
[11]Book (Second Edition) Data Mining, Concept and techniques by Jiawei Han and Micheline Kamber.

BIOGRAPHY

Ashwini Batbarai has received B.E. degree in information technology from K.I.T.S. Ramtek Nagpur University in 2011. She is pursuing M.Tech. degree in computer science and engineering from Ramdeobaba college of engineering and management (Autonomous) Nagpur her research interest include association rule mining, Datamining.

Devishree Naidu,has received B.E degree in Computer science and M.E. Specialization in Wireless Communication and Computing from G.H.Raisoni College of Engineering, Rashtrasant Tukdoji Maharaj University (University of Nagpur) INDIA in year2009. She has around Seven years of Teaching Experience, Currently working as a Assistant Professor at Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, INDIA. She has guided undergraduate and post graduate student for research project work in the field of mobile computing, and wireless sensor network. She is currently carrying her research work in sensor networks. Her previous research work was in Header compression Related to TCP/IP Protocol. Also she has interest in association rule mining, datamining