ASSOCIATION RULES WITHOUT REDUNDANCY

Amit Kumar Chandanan[1]* , Suman Singh[2]
[1] Department of Computer Science and Information Technology, Guru Ghasidas Vishwavidyalaya, Bilaspur, India
[2] University Institute of Computer Science and Applications (UICSA),RDVV, Jabalpur, India
Corresponding Author: chandanan.amit@ggu.ac.in

Abstract:
Association rules are one of the most researched areas of data mining. This is useful in the marketing and retailing strategies. Association mining is to retrieval of a set of attributes shared with a large number of objects in a given database. There are many potential application areas for association rule approach which include design, layout, and customer segregation and so on. The redundancy in association rules affects the quality of the information presented. The goal of redundancy elimination is to improve the quality and usefulness of the rules. The aim of this work is to remove hierarchical duplicity in multi-level, thus reducing the size of the rules set to improve the quality and usefulness without any loss. Redundancy in association rules mining decreases the speed for rules generation. It causes so many rules to be generated for same set of attribute. This research proposes a hybrid method that is combination of binary particle swarm optimization and mutation i.e. HBPSO.
Key Words: Association Rule, HBPSO, Mutation, non-redundant rule

Introduction:

Association rule discovery (Ceglar A. and J F Roddick J.F. 2006, Agrawal R. and Srikant R. 1995), a successful and important mining task, aim at uncovering all frequent patterns among transactions composed of data attributes or items. Results are presented in the form of rules between different sets of items, along with metrics i.e. joint and conditional probabilities of the pre-existing and subsequent, to judge a rule's importance. A closed set contains its own boundary. In other words, if you are "outside" a closed set, you may move a small amount in any direction and still stay outside the set. Note that this is also true if the boundary is the empty set, e.g. in the metric space of rational numbers, for the set of numbers of which the square is less than 2.

Association rule mining plays key role in information and knowledge discovery, to find association rule there are many approaches available with short comings to discovery quality rules.

Association rule mining has become an important and most used technique in data mining technique. This technique is mainly used to find frequent patterns, interesting co-occurrences and associations from sets of items in large transactional databases. With Traditional approach

there are two steps in finding association rules: firstly, finding the frequent patterns or item sets using the constraint of minimum support and secondly generating the rules from these frequent patterns/item sets using the constraint of minimum confidence (Tanna P. and Y Ghodasara Y. ,2013). With this approach, the basis of an interesting or useful rule is that its confidence exceeds a user defined threshold. This approach is widely known as the frequent item set approach. Much work has been done in developing more and more efficient algorithms or data structures to make computing these rules quicker. Much effort has been focused on improving the determination of the frequent item sets.

Another technique that has developed from the traditional frequent item set approach is the use of frequent closed item sets (Tanna P. and Y Ghodasara Y., 2013, Chandanan A. K. and Shukla M. K., 2013), which has originated from the mathematical theory of Formal Concept Analysis. It was shown to be a powerful technique for data analysis. Its major advantage is its ability to reduce the number of rules as well as provide a more concise representation which is lossless. Usually too many association rules containing redundancies are discovered; often too many to comprehend. Using frequent closed item sets the issue of redundancy can be dealt with by deriving non-redundant association rules.

However, this work has only dealt with redundancy in single level datasets. Multi-level datasets (in which the items are not all at the same concept level) contain information at different levels. The approaches used to find frequent item sets in single level datasets miss information, as they only look at one level in the dataset. Thus techniques that consider all the levels are needed. However, rules derived from multi-level datasets can have the same issues with redundancy as those from a single level dataset (Han J. and Fu Y., 2000). While approaches used to remove redundancy in single level datasets can be adapted for use in multilevel datasets, they still fail to remove all of the redundancies, namely the redundancy of hierarchy, where one rule at a given level gives the same information as another rule at a different level.

Data mining and the use of the discovered data and knowledge is a major field of research. Research in this field is also often taken and applied to real world scenarios. In order to improve the usage of data and the knowledge it contains, it is necessary to develop better techniques. Improvements to rule mining, being one of the major data mining technologies, would result in benefits to many applications. Hence the development of new novel techniques to discover and mine high quality association rules from multi-level datasets and effectively use them is important. Furthermore, to ensure that high quality rules can be identified, it is equally as important to have an assessment for measuring a rule's usefulness or interestingness. By achieving this, then the applications that utilize association rules can be improved.

This paper looks into hierarchical redundancy and with a proposal to derive association rule without redundant rules. The definition for non-redundant rules used here defined by (Shrivastava N. and Singh S. L. (2012), in which minimum antecedent and maximum consequents are expected. With this definition we add some requirement which consider the hierarchy level of data in the rule during discovering redundancy. By this approach more repeated association rules may be removed. It is possible to derive association rules from

this non redundant set of basis rules and finally there no information loss by this generated rules.

Result and Discussions

The results are shown in the form of a table where different items are taken and are numbered in the form of rule number and these items are divided into antecedents and the consequent which will be described below with its full meanings. Here percentage is taken from 100 and are named under support and confidence.

Here the BPSO and HBPSO are run for grocery dataset, the experimental results are show in form of table and graph. The run time of both algorithm has been compare and rule generated by both algorithms.

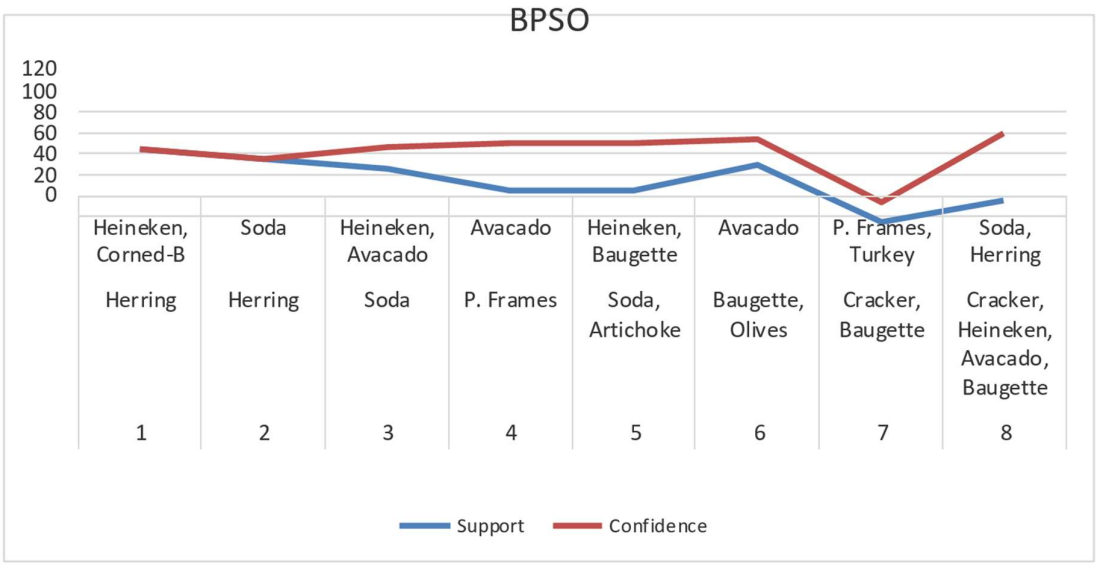

Figure 1: Rules for BPSO with support and confidence

The rules shown in figure 1 are generated by BPSO, eight rules for dataset with their confidence and support, the same dataset was tested with HPBSO and it has been found that only four rules are generated. The run time of HPBSO is lesser than BPSO.

Figure 2: Rules for HBPSO with support and confidence

When the HBPSO is run with 500 records of food data set rules are shown in table 2 with support and confidence

Table 1: Rules generated by BPSO

Rule NumbeAr	tecedent	Consequent	Support	Confidence
1.	Beer	Canned veg	65.00	76.47
2.	Wine	Fruit veg	25.00	45.45
3.	Beer	Canned veg, rozen meal	40.00	47.06
4.	Canned veg	Beer, rozen meal	40.00	50.00
5.	Wine	Beer, rozen meal	30.00	54.55
6.	Wine	Frozen meal, nfectionery	30.00	54.55
7.	Beer	Canned veg, Turkey	40.00	47.06

Table 2: Rules generated by HBPSO

		onsequent	Support	Confidence
1.	Fish	Beer, Canned Veg.	65.00	86.67
2.	Fruit Veg.	Dairy, Turkey	20.00	36.36
3.	Fruit Veg.	Dairy, Frozen Meal	20.00	36.36
4.	Wine	Fish	50.00	90.91

Sarath K.N.V.D. and Vadlamani R (2013) was generate association rules using binary particle swarm optimization algorithm. These result has been compared with HBPSO for 500 food transaction dataset. From figure 3 the run time for binary particle swarm optimization algorithm and HBPSO for 500 transaction dataset, in above comparison shows that performance of this research work in better than binary particle swarm optimization algorithm by 34%

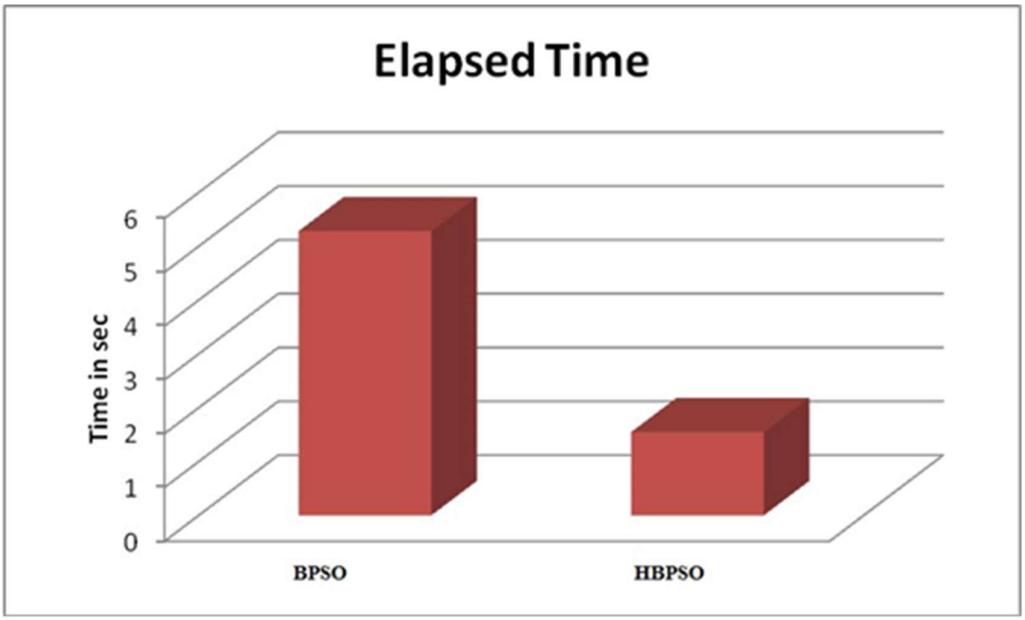

Figure 3 Comparison on Elapsed time (in seconds) for BPSO and HBPSO

Conclusion

Redundancy in association rules mining decreases the speed for rules generation. It causes so many rules to be generated for same set of attribute. The goal of this work was to remove redundancy in the item set it reduce unnecessary time utilization of algorithm. With use of above designed algorithm redundant rules for the large /multilevel dataset could be removed. By reducing the set of rule improve the quality and efficiency of mining without any loss of

information. This proposed approach remove redundancy using upper level closed frequent item set.

References:

Agrawal R. and Srikant R. (1995). Mining sequential patterns. Proceedings of the Eleventh International Conference on Data Engineering.

Ceglar A. and J F Roddick J.F. (2006). Association Mining. ACM Computing Surveys (C SUR). Vol 38(2): p. 13.

Tanna P. and Y Ghodasara Y. (2013). Foundation for Frequent Pattern Mining Algorithms Implementation. International Journal of Computer Trends and Technology (IJCTT). Vol 4(7).

Han J, Y Fu (2000). Mining Multiple Level Association Rules in Large Databases. IEEE Transactions on Knowledge and Data Engineering Vol 11(5): p.798-804 11

Zaki M. J. (2004). Mining Non-Redundant Association Rules. Data Mining and Knowledge. Vol 9: p. 223-248.

Chandanan A. K. and Shukla M. K. (2013). Review on Redundancy Free Association Rule mining. IJCA Proceedings on International Conference on Communication Technology ICCT (5):13-16.

Shrivastava N. and Singh S. L. (2012). Overview of Non-Redundant Association Rule Mining. Research Journal of Recent Sciences. Vol. 1(2): ISSN 2277-2502: p.108-112.