IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-8, Aug 2018 E-ISSN: 2347-2693

Discovery of High Utility Patterns from Retail Database with adding constraints on LBHUP algorithm

A. G. Deshpande[1*], R. J. Deshmukh[2]

[1,2]Dep. of Computer Science & Technology, Department Of Technology, Shivaji University, Kolhapur, India

Corresponding Author: anuja.g.deshpande@gmail.com, Tel.: +91-9766075094

Available online at: www.ijcseonline.org

Accepted: 16/Aug/2018, Published: 31/Aug/2018

Abstract— Association-rule mining is commonly used mining technique for finding frequent patterns. In real world applications traditional association-rule mining is not appropriate since the purchased item have different factors, for example, amount and benefit. High utility pattern mining was used for unravelling the limitations of the association-rule mining as far as amount and benefit. There are algorithms for finding high utility patterns from static databases. Some researches worked for handling dynamic dataset but huge computational time and multiple database scan was required. In this paper, a system is proposed for finding high utility patterns which uses list based data structure for dynamic dataset. To improve computational time and memory, constraints like item, date or length are used. A few experiments are led to demonstrate the execution of the proposed system with and without using constraints regarding time and memory.

Keywords—Association rules (ARs); Association rule mining; frequent patterns; high utility patterns; pattern mining

I. INTRODUCTION

The quick change of database frameworks supports the limit and usage of dynamic data from business organizations, governments, and sensible affiliations. Well ordered directions to gain productive information from various databases have become broad thought, which achieves the sharp rising of related research subjects. High utility pattern mining is the most vital issue and it is an augmentation of frequent pattern mining [5]. Association-rule mining [8] from a value-based database is a crucial undertaking for uncovering the connections among things. The Apriori [13] was the primary algorithm to mine the affiliation administers in a level-wise manner. It utilizes produce and test instrument to discover the candidate itemsets and afterward infer the frequent itemsets in light of the minimum limit. The association rules are then uncovered from the found frequent itemsets in light of the minimum limit.

High-utility mining (HUM) [1],[2],[3],[4] was along these lines proposed to mostly unravel the constraints of association-rule mining. It might be thought of as an augmentation of frequent-itemset mining by thinking about the sold amounts and benefits of the things. The utility of an itemset can be estimated as far as amount and benefit, which can be characterized by client inclination. For instance, somebody might be occupied with finding the itemsets with great benefits and another may centre around the itemsets with low contamination while fabricating. At the moment that the utility of an itemset is greater than or equal to the base utility check, an itemset is considered as a high-utility itemset. A couple of algorithms have been proposed to mine high-utility itemsets in a static database. In day to day applications these algorithms are not proficient, for that new algorithms are creating for dealing with dynamic information. In this paper, a system is proposed for finding high utility patterns which uses list based data structure for dynamic dataset. To improve computational time and memory, constraints like item, date or length are used. A few experiments are conducted to demonstrate the performance of proposed system with and without using constraints regarding time and memory.

II. RELATED WORK

Itemset mining is used for finding ARs [8] in light of the fact that they give a subset of examples inside which to seek. HUCI algorithm [8] uses utility confidence framework. This algorithm generates utility with the help of non-redundant association rules. Mainly used algorithm used to find frequent itemset is Apriori, which was proposed by Agrawal and Srikant [13].

High utility itemsets imply the courses of action of things with a high utility like advantage in a database, and powerful mining of high utility itemsets accept a significant part in some certified applications and is a fundamental research issue in data mining zone. They create candidate itemsets for finding high utility patterns. These calculations achieve the issue that a far reaching number of contenders are made, anyway an extensive segment of the candidates are found to be not high utility after their right utilities are handled. HUI-Miner [1] utilizes a novel structure, called utility-list. The list stores the utility data of itemset and some additional information for quick seeking.

III. Proposed Work

In proposed work, to find high utility patterns from transactional retail databases we use an efficient high utility pattern mining algorithm which is list-based. There is no generation of candidates.

It consists of following processes-

1) Data Pre-processing.
2) List-based pattern mining algorithm for high utility items.
3) Formation of List Data Structure.
4) Global List Data Restructure.
5) Constraints.

Data Pre-processing

The database consists of set of transactions of different items. The real life dataset of retail market is available in different format like text file or excel file. In the pre-processing step the available file format is converted into suitable file format required for input of the proposed algorithm.

List-based pattern mining Algorithm for high utility items

The mining process in the proposed system is vital operation. The inputs given to the mining algorithm are utility list, minimum threshold and other constraints.

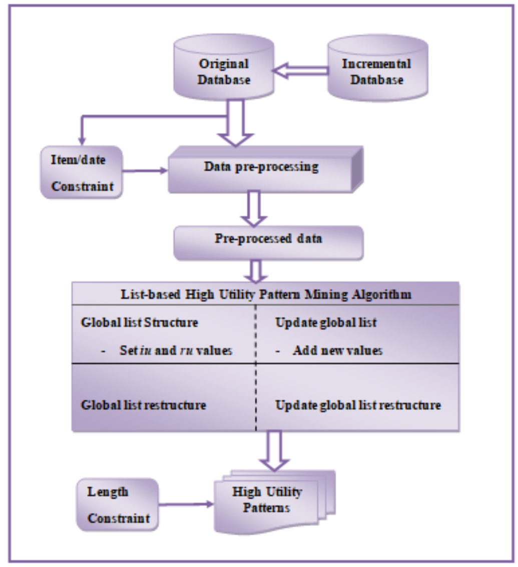

Fig.1 System block diagram

The output of the algorithm is the high utility pattern list which helps the analyst to analyse the data and take precise decision.

Formation of List Data Structure

At first every transaction is read from the source dataset. Sorting of items in the transaction is done according to current sorting order. List of utility items is made called as global list of high utility items. For each item in the utility list iu value refers to utility of item in the transaction and ru value is assigned to zero.

Global List Data Restructure

The restructure procedure is used for newly added transactions. When new transaction is added, the sorting order should be modified. The restructure procedure at first sets all ru values in the list as zero. Then consider the transactional utility weight for each item and according to twu values, the list is arranged in ascending order.

Constraints

- The threshold value $minutil$ is entered by user. The high utility patterns are those patterns whose utilities are greater than or equal to the threshold value.
- For increasing the efficiency of the mining process some constraints like length, attribute etc. are added by user.

It is useful for better and precise decision making process. The use of more constraints increases the efficiency and performance of the mining algorithm.

IV. EXPERIMENTATION

The online retail dataset is a transactional dataset obtained from UCI machine learning repository (https://archive.ics.uci.edu/ml/datasets/online+retail). It contains retail transaction data attributes like Invoice no, Stock code, Description, Quantity, Invoice date, Unit Price, Customer code and Country. Day by day transactional data goes on increasing because of electronic usage. Experimentation uses minutil as threshold utility where utility is the product of unit price and respectively purchased quantity. For the experimentation, transactional data of one year with 25MB size is used.

Experiments are performed on intel core i5 processor with 8GB memory, running on windows10 operating system.

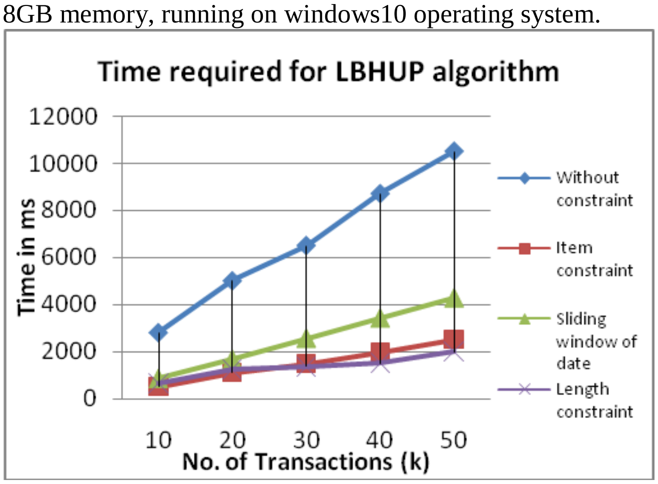

Fig.2 Experimental results for required time for LBHUP algorithm

The first experimentation gives the time requirement of LBHUP algorithm. Figure 2 shows the time required for LBHUP algorithm with varying number of transactions. Figure 2 shows the time required by an algorithm without constraint and with adding different constraints such as item, item length and sliding window for date attribute. Graph in figure 2 shows the time required with adding constraint is less than that of the algorithm without any constraint. Similar to date attribute other user interested attribute can be chosen.

The second experimentation gives memory utilization of LBHUP algorithm. Figure 3 shows the memory utilization of LBHUP algorithm with different threshold values i.e. different $minutil$ values. Figure 3 shows the memory utilization of an algorithm without constraint and with adding different constraints such as item, item length and sliding window for date attribute. Graph in figure 3 shows the memory utilization with adding constraint is less than that of the algorithm without any constraint. The space complexity of algorithm is $O(n)$ where n refers to number of items. This experiment shows that with adding constraints such as date, item and length memory required for algorithm is reduced.

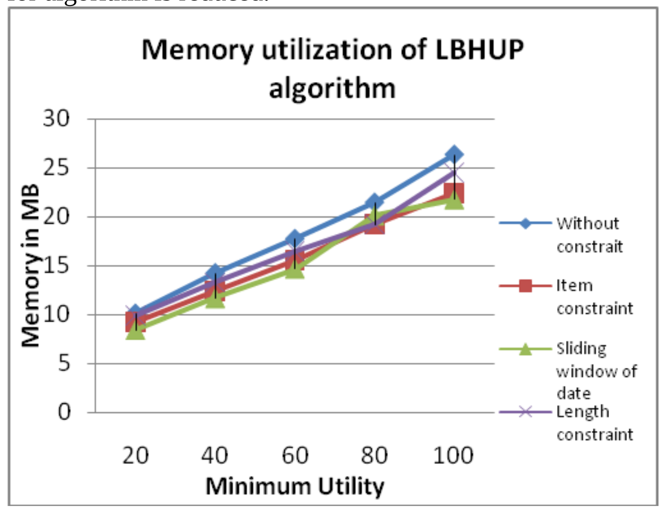

Fig.3 Experimental results for memory utilization of LBHUP algorithm

V. CONCLUSION

The proposed system discovers constraint based high utility patterns. The previous study of high utility mining algorithms such as HUI-Miner, HUI-list-INS which discovers high utility patterns, but it requires an additional database scan, which consumes more data processing time.

The proposed system of List-based High Utility Pattern Mining (LBHUP) algorithm is used to discover high utility patterns with the use of threshold ($minutil$) and some user interested constraints like length, item, date which are used for more fine prediction and analysis. The proposed system is useful to handle dynamic database. Experimental study shows that adding constraints reduces execution time and memory requirement.

ACKNOWLEDGMENT

This paper work was guided and supported by Mrs. R. J. Deshmukh. I would like to thank the guide, anonymous reviewers for their valuable and constructive comments on improving the paper.

REFERENCES

[1] M. Liu , J.-F. Qu , “Mining high utility itemsets without candidate generation”, in: International Conference on Information and Knowledge Management (CIKM 2012), 2012, pp. 55–64.

[2] J.C.-W. Lin , T. Li , P. Fournier-Viger , T.-P. Hong , J. Zhan , M. Voznak , “An efficient algorithm to mine high average-utility itemsets”, Adv. Eng. Inform. 30 (2) (2016) 233–243.

[3] J. Liu , K. Wang , B.C.M. Fung , "Mining high utility patterns in one phase without generating candidates", IEEE Trans. Knowl. Data Eng. 28 (5) (2016) 1245–1257 .

[4] J. Liu , K. Wang , B.C.M. Fung , "Direct discovery of high utility itemsets without candidate generation", in: Proceedings of the 2012 IEEE International Conference on Data Mining (ICDM 2012), 2012, pp. 984–989.

[5] L. Troiano , G. Scibelli , "Mining frequent itemsets in data streams within a time horizon", Data Knowl. Eng. 89 (2014) 21–37.

[6] C.F. Ahmed , S.K. Tanbeer , B.-S. Jeong , H.-J. Choi , "Interactive mining of high utility patterns over data streams", Expert Syst. Appl. 39 (15) (2012) 11979–11991.

[7] V.S. Tseng , C.-W. Wu , P. Fournier-Viger , P.S. Yu , "Efficient algorithms for mining the concise and lossless representation of high utility itemsets", IEEE Trans. Knowl. Data Eng. 27 (3) (2015) 726–739.

[8] J. Sahoo , A.K. Das , A. Goswami , "An efficient approach for mining association rules from high utility itemsets", Expert Syst. Appl. 42 (13) (2015) 5754–5778.

[9] U. Yun , D. Kim , H. Ryang , G. Lee , K.-M. Lee , "Mining recent high average utility patterns based on sliding window from stream data", J. Intell. Fuzzy Syst. 30 (6) (2016) 3605–3617.

[10] C.-W. Lin , G.-C. Lan , T.-P. Hong , "Mining high utility itemsets for transaction deletion in a dynamic database", Intell. Data Anal. 19 (1)(2015) 43–55.

[11] H. Ryang , U. Yun , K. Ryu , "Fast algorithm for high utility pattern mining with the sum of item quantities", Intell. Data Anal. 20 (2) (2016) 395–415.

[12] U. Yun , D. Kim , H. Ryang , G. Lee , K.-M. Lee , "Mining recent high average utility patterns based on sliding window from stream data", J. Intell. Fuzzy Syst. 30 (6) (2016) 3605–3617.

[13] R. Agrawal , R. Srikant , "Fast algorithms for mining association rules", in: Proceedings of the 20th International Conference on Very Large Data Bases (VLDB 1994), 1994, pp. 487–499.

Authors Profile

Miss Anuja G. Deshpande pursed Bachelor of Engineering from Shivaji University, Kolhapur, India in 2012. She is currently pursuing Master of Technology from Shivaji University, Kolhapur, India.

Mrs R. J. Deshmukh pursed Bachelor of Engineering from Shivaji University, Kolhapur, India and Master of Technology from Shivaji University, Kolhapur, India in year 2009. She is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Technology, Department of Computer Science and Technology, Shivaji University, Kolhapur, Maharashtra, India. She has published more than 10 research papers in reputed international journals and conferences including IEEE,Elsvier,Springer and it's also available online. Her main research work focuses on Data Mining.

