isrjournals

International Journal of Advanced Research in Computer Science Engineering and Information Technology

Volume: 4, Issue: 3,Special Issue: 2 ,Apr,2016 ,ISSN_NO: 2321-3337

E-Commerce Website Without Review Using Hui Algorithm

Poornachander Karunagaran [1], Vignesh Ravikumar [2], Mr.A.Thiyagarajan [3]

Student, Department of Computer Science and Engineering, Agni College of Technology,India[1,2] .

Assistant Professor, Department of Computer Science and Engineering, Agni College of Technology, India [3].

ABSTRACT

Mining high utility item sets (HUIs) from databases is an important data mining task, which refers to the discovery of item sets with high utilities (e.g. high profits). However, it may present too many HUIs to users, which also degrades the efficiency of the mining process. To achieve high efficiency for the mining task and provide a concise mining result to users, we propose a novel framework in this paper for mining closed high utility item sets (CHUIs), which serves as a compact and lossless representation of HUIs. We propose three efficient algorithms named AprioriCH (Apriori-based algorithm for mining High utility Closed þ item sets), AprioriHC-D (AprioriHC algorithm with Discarding unpromising and isolated items) and CHUD (Closed þ High Utility Item set Discovery) to find this representation. Further, a method called DAHU (Derive All High Utility Item sets) is proposed to recover all HUIs from the set of CHUIs without accessing the original database. Results on real and synthetic datasets show that the proposed algorithms are very efficient and that our approaches achieve a massive reduction in the number of HUIs. In addition, when all HUIs can be recovered by DAHU, the combination of CHUD and DAHU outperforms the state-of-the-art algorithms for mining HUIs.

Keywords---Frequent item set, closed high utility item set, lossless and concise representation, utility mining, and data mining

Volume: 4, Issue: 3,Special Issue: 2 ,Apr,2016 ,ISSN_NO: 2321-3337

1. INTRODUCTION

FREQUENT item set mining (FIM) is a fundamental research topic in data mining. One of its popular applications is market basket analysis, which refers to the discovery of sets of items (item sets) that are frequently purchased together by customers. However, in this application, the traditional model of FIM may discover a large amount of frequent but low revenue item sets and lose the information on valuable item sets having low selling frequencies. These problems are caused by the facts that FIM treats all items as having the same importance/unit profit/weight and it assumes that every item in a transaction appears in a binary form, i.e., an item can be either present or absent in a transaction, which does not indicate its purchase quantity in the transaction. Hence, FIM cannot satisfy the requirement of users who desire to discover item sets with high utilities such as high profits. To address these issues, utility mining emerges as an important topic in data mining. In utility mining, each item has a weight (e.g. unit profit) and can appear more than once in each transaction (e.g. purchase quantity). The utility of an item set represents its importance, which can be measured in terms of weight, profit, cost, quantity or other information depending on the user preference. An item set is called a high utility item set (HUI) if its utility is no less than a user-specified minimum utility threshold; otherwise, it is called a low utility item set. Utility mining is an important task and has a wide range of applications such as website click stream analysis, cross marketing in retail stores, mobile commerce environment and biomedical applications [6].However, HUI mining is not an easy task since the downward closure property in FIM does not hold in utility mining. In other words, the search space for mining HUIs cannot be directly reduced as it is done in FIM because a superset of a low utility item set can be a high utility item set. Many studies were proposed for mining HUIs, but they often present a large number of high utility item sets to users.

A very large number of high utility item sets makes it difficult for the users to comprehend the results. It may also cause the algorithms to become inefficient in terms of time and memory requirement, or even run out of memory. It is widely recognized that the higher utility item sets the algorithms generate, the more processing they consume. The performance of the mining task decreases greatly for low minimum utility thresholds or when dealing with dense databases.

2 . EXISTING SYSTEM

- Existing techniques for high utility item sets is Integrating concepts of concise representations from FIM into HUI mining may produce a loss representation of all HUIs or a representation that is not meaningful to the users.
- The representation may not achieve a significant reduction in the number of extracted patterns to justify using the representation.
- The Unwanted representation of data leads to misunderstanding.

3. PROPOSED SYSTEM

- We propose a top-down method named DAHU (Derive All High Utility itemsets) for efficiently recovering all HUIs from the set of CHUIs. The proposed representation is lossless due to a new structure named utility unit array that allows recovering all HUIs and their utilities efficiently.
- We propose three efficient algorithms named AprioriHC (Apriori-based algorithm for mining High utility Closedþ item set), AprioriHC-D (AprioriHC algorithm with Discarding unpromising and isolated items) and CHUD (Closedþ High Utility item set Discovery) to find this representation.
- Using algorithm, HUI can be found and unwanted item sets are removed. Users can view needed information alone.

4. SYSTEM IMPLEMENTATION

4.1 Product insertion

Admin has the authority to maintain the website. Admin add products into database with product details which include product id, brand name, and product description with rate to main database with secure operations. Dynamically it is displayed in website. It can be retrieved by user in website with product name.

4.2 View transaction log

Admin can do operations in the transaction log. Admin views table of transaction in which products transaction details displayed. Updating in transaction table can be accessed by admin. In the table using the algorithm the no. of view count in a product and no. of sold count in a product has been displayed which is generated by users.

4.2 User View

User can register in the online shopping site to login in the website. User can view all products which are displayed in the website. When user hits the image of a product to view the whole details of product user can view the product with information of product sold and view count. After ordering the product user gets notification mail about ordered product. User can view the order list in the account.

4.4 High utility item sets

After User operation in the website the product values has been changed. Admin checks the table details of product log, the high utility item sets (ratio of product count and sold count is partially equal) is backed up and less utility (ratio of product count and sold count is not equal) is deleted by admin with view and sold count. The product which has high view count without sold count or only view count is less utility item set which will be deleted by admin.

5. ARCHITECTURE DIAGRAM

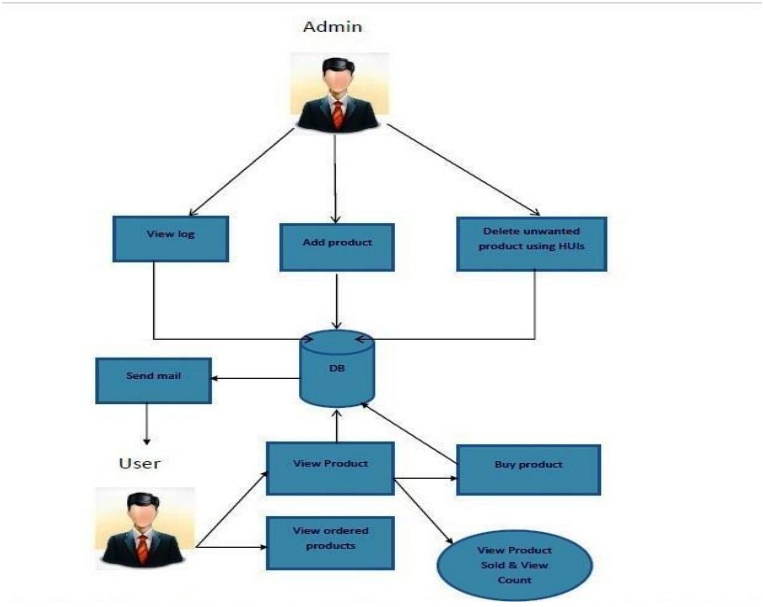

Figure 1: Architecture Diagram.

6. OUR CONTRIBUTION

We propose a top-down method named DAHU (Derive All High Utility item sets) for efficiently recovering all HUIs from the set of CHUIs. The proposed representation is lossless

due to a new structure named utility unit array that allows recovering all HUIs and their utilities efficiently. We propose three efficient algorithms named AprioriHC (Apriori-based algorithm for mining High utility Closed item set),AprioriHC-D (AprioriHC algorithm with Discarding unpromising and isolated items) and CHUD (Closed High Utility item set Discovery) to find this representation. Using algorithm, HUI can be found and unwanted item sets are removed. Users can view needed information alone.

7. CONCLUSION

In this paper, we addressed the problem of redundancy in high utility item set mining by proposing a lossless and compact representation named closed high utility item sets, which has not been explored so far. To mine this representation, we proposed three efficient algorithms named AprioriHC (Apriori-based approach for mining High utility closed item set), AprioriHC-D (AprioriHC algorithm with Discarding unpromising and isolated items) and CHUID (Closed High Utility item set Discovery). AprioriHC-D is an improved version of AprioriHC, which incorporates strategies DGU [24] and IIDS [19] for pruning candidates. AprioriHC and AprioriHCD perform a breadth-first search for mining closed high utility item sets from horizontal database, while CHUD performs a depth-first search for mining closed high utility item sets from vertical database. The strategies incorporated in CHUD are efficient and novel. They have never been used for vertical mining of high utility item sets and closed high utility item sets. To efficiently recover all high utility item sets from closed high utility item sets, we proposed an efficient method named DAHU (Derive All High Utility item sets). Results on both real and synthetic datasets show that the proposed representation achieves a massive reduction in the number of high utility item sets on all real datasets (e.g. a reduction of up to 800 times for Mushroom and 32 times for Food mart). Besides, CHUD outperforms UP-Growth, one of the currently best algorithms by several orders of magnitude (e.g. CHUD terminates in 80 seconds on BMSWebView1 for min utility 1⁄4 2%, while UP-Growth cannot terminate within 24 hours). The combination of CHUD and DAHU is also faster than UP-Growth when DAHU could be applied.

8. REFERENCE

[1] Vincent S. Tseng, Cheng-Wei Wu, Philippe Fournier-Viger, and Philip S. Yu, Fellow Efficient Algorithms for Mining the Concise and Lossless Representation of High Utility Item sets, IEEE,Vol .27,no.03 2015,pp. 487–499.

[2] C. F. Ahmed, S. K. Tanbeer, B.-S. Jeong, and Y.-K. Lee, "Efficient tree structures for high utility pattern mining in incremental databases," IEEE Trans. Knowl. Data Eng., vol. 21, no. 12, pp. 1708–1721, Dec. 2009.

[3] J.-F. Boulicaut, A. Bykowski, and C. Rigotti, "Free-sets: A condensed representation of Boolean data for the approximation of frequency queries," Data Mining Knowl. Discovery, vol. 7, no. 1,pp. 5–22, 2003.

[4] T. Calders and B. Goethals, "Mining all non-derivable frequent item sets," in Proc. Int. Conf. Eur. Conf. Principles Data Mining Knowl. Discovery, 2002, pp. 74–85.

[5] K. Chuang, J. Huang, and M. Chen, "Mining top-k frequent patterns in the presence of the memory constraint," VLDB J., vol. 17, pp. 1321–1344, 2008.

[6] R. Chan, Q. Yang, and Y. Shen, "Mining high utility item sets," in Proc. IEEE Int. Conf. Data Min., 2003, pp. 19–26.

[7] A. Erwin, R. P. Gopalan, and N. R. Achuthan, "Efficient mining of high utility item sets from large datasets," in Proc. Int. Conf. PacificAsia Conf. Knowl. Discovery Data Mining, 2008, pp. 554–561.

[8] K. Gouda and M. J. Zaki, "Efficiently mining maximal frequent item sets," in Proc. IEEE Int. Conf. Data Mining, 2001, pp. 163–170.

[9] T. Hamrouni, "Key roles of closed sets and minimal generators in concise representations of frequent patterns," Intell. Data Anal vol. 16, no. 4, pp. 581–631, 2012.

[10] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2000, pp. 1–12.