TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 13062 - 13068

Effectively Mining on Utility Itemset by using Conventional Method

Suresh .K
Assistant Professor, Department Of Computer Science, Amrita School Of Arts And Sciences Mysore,Amrita Vishwa Vidyapeetham, India.
suresh.kalaimani@gmail.com

Kashyap C
Department Of Computer Science, Amrita School Of Arts And Sciences Mysore, India, Amrita Vishwa Vidyapeetham, India.
Kashyapcherukatte@gmail.com

Article Info
Volume 82
Page Number: 13062 - 13068
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 24 February 2020*

Abstract
Data mining is a method of examining huge data sets to obtain new patterns. Data mining is an interdisciplinary subfield of computing and statistics with an overall goal to extract information from a knowledge set and transform the knowledge into a clear structure for further use. The main task during data mining is high utility itemset mining. High utility itemset mining technique handles a great deal of benefit for the retail industry. Numerous algorithms has been projected to determine HUIs utmost of the algorithm works with positive utility itemsets, but items with positive and negative values are found in real life.To solve this problem we proposed a new algorithm which works for both positive and negative itemset and it is more efficient compared to others in terms of running time and efficiency

Keywords: *Datamining, High utility, Negative utility, Frequent pattern, Itemset mining, Frequent Itemset mining.*

I. INTRODUCTION

Data mining is a method employed by enterprises to convert data into useful information. It is used to find the hidden information from a big database. Interesting patterns and valuable information can be produced from large data sets. The main technique during datamining is high utility mining. High utility itemsets produce greater profit. High utility itemsets contain positive value and negative values. Several methods has been projected to determine HUIs but these methods do not consider negative values. High utility itemset mining is applicable in various domains to find out the best pattern. One such place where high utility mining is used to get list of things customer brought together which gives a great profit. The data mining method helps in the growth of organization

APPLICATION OF DATA MINING

- Data Mining in Farming
- Crime Investigation / Mass surveillance
- Customer analysis
- Educational

The data mining procedure urges associations to develop models reliant on obvious data to envision who will respond to new promoting exertion. For instance, post office based on the web, electronic advancing exertion, etc. Through this desire, promoters can have fitting approach to manage offer beneficial things to centered customers with high satisfaction. Data mining conveys a lot of bit of leeway to retail association comparably as advancing. Through market holder assessment, the store can have a legitimate age strategy in the way that customers can buy visit acquiring things together with enchanting. Likewise, it furthermore help the retail association offers a particular

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 13062 - 13068

discount for explicit things what will pull in customers.

II. LITERATURE SURVEY

Jerry Chun-Wei Lin et al.[2016] described about effectual mining of high utility itemset having negative unit profits. One of the developing data mining tasks, which consist of determining extremely gainful itemset in a very large transactional database is high utility itemset mining. Great commercial itemsets is mostly discovered from large transactional database. Transactional database contains positive and negative unit profits. Most of the procedure assumes that high utility itemset have only positive unit profits. Mining high utility itemset having together positive and negative profit is expensive task. To solve these issues efficient algorithm called Faster High-Utility itemset miner having Negative unit profits (FHN) is proposed. The novel PNU list structure and pruning strategies are used. Pruning strategies are used to improve the performance and reduce candidate generation The results shows that faster high utility itemset miner algorithm is faster and takes less memory

Srikumar Krishnamoorthy [2017] proposed an algorithm named GHUM. It deals with effective mining of high utility itemset having negative unit profits. The important problems in utility mining that consider utilities of item to realize exciting pattern from transactional databank is high utility itemset. The intension of finding negative unit profit is that it helps the decision maker to easily find profitable itemsets. For effective mining of high utility itemset pruning strategies, data structures and algorithm has been proposed. The goal of the paper is to present a general method known as GHUM generalized high utility mining. The projected algorithm consider both the positive unit profit and negative unit profit. This technique uses basic utility list data structure, novel utility based antimonotone property and pruning strategies. This paper also presents another pruning systems to improve execution of mining. In high utility itemset mining utilities are used from transactional database to obtain profitable itemsets. HUI mining problem has gained greater attention consideration because of its potential pertinence in various business and logical applications

Mohsen Marjani et al.[2017] proposed framework for big iot data analytics and its opportunities. The iot device produce large data, big data iot and analytics results are used to get information from these large data. These data are not usefull until proper analytics is done. This study investigates state-of-the art research efforts directed towards Big IOT data analytics. As there is a quick increase in growth of big data and internet of things it thus touching all ranges of technologies and industries by expanding the advantages for associations and people. Big data analytics has become challenging testing in view of the preparing and assortment of information through various sensors and also due to wide spread of IoT devices. IoT big data analytics are used on large IoT data to get fashions, hidden patterns, hidden co-relations and innovative information. These techniques are mainly used to help business associations and organisations to get clear idea of data, which help them to take proper decisions. Big data analytics aim to obtain information using data analytics very quickly and fast. These technique can be used for both generation of data analytics and problem specific method

Mehdi Mohammadi et al.[2018] This paper deals with IoT and data analytics. Analytics is applied in data streams, which discovers new ideas. By using advanced technology IoT transforms traditional object to smart. IoT applications are spread across various field such as health, transportation, smart home, education. IoT is considered as a commendable worldview for organizations and personal satisfaction improving innovation as it helps in applying examination over huge information stream to find new data, foresee future knowledge and settle on control choice. The

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 13062 - 13068

fundamental component of the vast majority of the IoT applications is a keen learning system for forecast which incorporates relapse, classification, and grouping. Out of the many AI draws near, Deep Learning has been effectively used in numerous IoT applications as of late. Aside from the large data investigation, IoT data needs additional new class of examination, rapid and falling data examination, to help applications with fast data streams and necessitating time-touchy activities.

Zhenyu Shou et al. [2018] proposed frequent sequential activity pattern mining. In this paper they make a methodological system to investigate likeness of action designusing frequent sequential pattern mining.The current likeness measures have a few issues. This paper additionally built up an inventive likeness system utilizing frequent sequential pattern mining by attachment of multi-person multi-day GPS trajectories. To extricate frequent designs for every individual prefixspan procedures is utilized prefix span has the advantage that it only explores pattern appearing in the database. For each individual, frequent pattern mining produces numerous frequent patterns

Loan T.T. Nguyen et al. [2018] proposed technique named MEFIM, about the High utility which is an essential process to yield high profit in data mining. HUI is mainly applicable on real time applications which works up on the problem of HUI. MEFIM modified efficient high utility itemset mining algorithm is used which contain p-set to decrease speed of the transmission. The main aims of the frequent itemset mining is to determine the new itemsets. The disadvantages of HUIM algorithm are static that do not overcome the problems in real time. MEFIM reduces the cost and also increases the performance in the p-set structure. High utility is one of the significant roles in datamining. Algorithm used is MEFIM (modified efficient high utility itemset mining) which is used to reduce transactional speed of this mining process. MEFIM is one of the faster algorithm. Only some seconds are required for the working. The disadvantages of HUIM (High utility itemset mining) algorithm are static that do not overcome the problems. MEFIM reduces the cost and also increases the performance in the p-set structure.

Unil Yun et al. [2018] introduced HAI algorithm, which it can mine the normal utility on itemsets. The calculation has high normal utility itemsets by profundity first search based mining procedure to evade the extension of itemsets pruning strategy is utilized. Mostly utilized continuous itemsets are apriori and FP development with create and test approach. To defeat impediment of apriori calculation FP tree structure is utilized and having better execution contrasted with apriori. This paper mainly used high average utility itemsets. It generates depth first search pruning process. There is a pruning technique for finding the upper bounds of the frequent itemsets. Algorithm uses list structures called HAI list for getting the information about mining the itemsets. It is not as that expensive designing to be less cost. To check the performance they have conducted many experiments with their datasets. It shows better result with the algorithm.

Kuldeep Singh et al. [2019] proposed novel method in efficient mining on high utility itemset with negative utility value and length constraints. One of the emerging topic in the frequent itemset mining is high utility itemset. In data mining high utility itemset mining is fundamental research topic and is the subfield of frequent itemset mining. High utility itemset falls under frequent itemset mining which is a significant research topic in data mining. The frequent itemset mining is used to determine frequent pattern from transactional database which has minimum support count. The main aim is to find important itemset rather than frequent itemset related to frequent itemset mining, high utility itemset can get the exact data very easily. High utility itemset

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 13062 - 13068

solves many issues like handling large data, time to scan and so on. The proposed algorithm solves the problematic of producing a enormous quantity of candidates itemsets and itemsets are small in size which damage mining presentation

Loan T.T. Nguyen et al. [2019] proposed algorithm named EFIM about effective way for mining high utility closed itemset. When mining closed high utility itemset includes discovering a delegate set of HUIs. One of the advantages of closed high utility itemset is that it can produce full HUIs without loss of data. There are many methods to mine closed high utility itemset in which most effective are closed high utility itemset miner and EFIM closed. To diminish the search space and save time during rebuilding recessive inspection is performed. To diminish storage space closed transactions are merged.

Vaithinathan K et al.[2019] proposed algorithm named HUIM, which is the addition of frequent item-set mining (FIM).It reflects the yield and quality of item-sets to realize high-utility item-sets (HUTs) and this methodology is be that as it may ,improper in true applications since the utility of the thing sets expands .The quantity of things inside it high normal utility thing set mining(HAUIM) Was intended to quantify the utility size of the thing sets into account. Present calculation just deal with, the fixed databank and the size of information is as often as possible changed, all things considered, circumstances

III. PROPOSED METHODOLGY

This research paper proposed a technique for the operative mining of HUIs in both condensed and light datasets. One of the significant study topics in data mining is high utility itemset mining. There is an assumption that HUIs has only positive unit profit. HUIs with negative unit profit are recently used in market and business organizations. Mining high utility itemset with negative unit profit is a difficult task. Many algorithm are presently existing for minining high utility itemset with negative unit profits such as

- Apriori based
- Tree based
- Utility list based

These procedures are facing many problems while mining itemset. To report this matter we projected a new procedure which works for both positive and negative itemset. From the investigational outcomes it's clear that proposed procedure is more efficient compared to others in terms of running time and efficiency. The new projected procedure can also reduce the scanning time of dataset and also take less memory space to do mining.

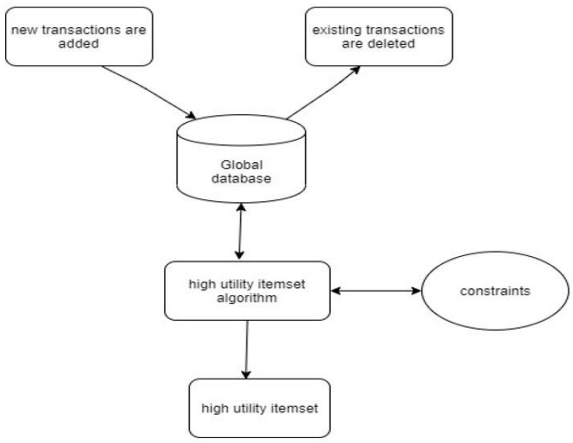

Figure: 1 Architecture for Utility mining

In fig-1, high utility mining method to obtain high utility itemset from adding new transaction to the global database and remove the old transaction. Global Database is a business intelligence provider, created to help businesses and organizations. Now we are applying the algorithm to the global database with some specific constraints. By using this high utility algorithm, we get high utility itemsets.

Over late years information mining has been building up itself as one of the significant factors in industry with developing modern effect. Without a doubt, explore in information mining will proceed and even increment over upcoming periods include mining compound items of

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 13062 - 13068

discretionary sort, quick, straightforward and organized information preprocessing, increasing ease of use. All target understanding shopper conduct, evaluating item request, management and fabricating the brand, following execution of clients or items in the marketplace and dynamic steady income from changing information into data and data into information.

In spite of the fact that information mining is still in its early stages, organizations in a wide scope of ventures are including retail, account, medicinal services, producing transportation, and aviation are now utilizing information mining devices and strategies to exploit authentic information. By utilizing design acknowledgment innovations and measurable and scientific procedures to filter through warehoused data, information mining assists investigators with perceiving huge realities, connections, patterns, examples, exemptions and inconsistencies that may somehow or another go unnoticed.

Algorithm

Input: D, the database which includes transactions, Minutil, the least value constraints M,
Output: HUI, the established elevated value from itemsets

1. Scan D and associate PTWU for all 1-itemsets
2. For each Ti be appropriate D do
3. Sort xi in Tj as per the ordering heuristic
4. UM ← PTaUt + TU(x, Tj)
5. If
6. For both elements ai goes to Tj do
7. If TUI
8. Endif
9. End

IV. EXPERIMENTAL ANALYSIS AND RESULTS

We analyse the viability HUM against a best in class Enhanced method calculation. Figure 1 shows the effects of our analyses on standard datasets. It is obvious from the graphs that projected strategy performs altogether superior to HUM. It is presentation enhancement for the chess dataset is multiple sets of greatness. For the retail dataset, the HUM implements hardly well at advanced least value qualities. Not with standing, at Minor least value qualities enhanced takes altogether lesser performance period than FHN and shows a direct increment in performance period.

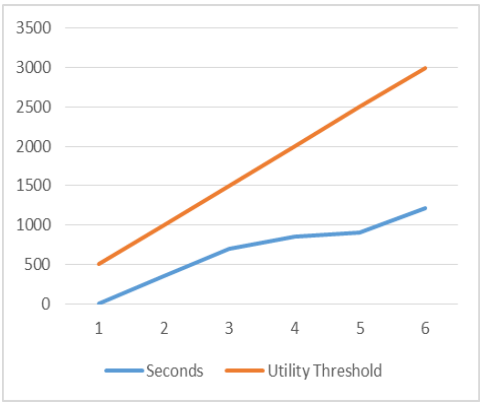

Figure 1. Minimum Utility threshold

The memory utilization for both the calculations is exhibited in Fig 2. The memory utilization execution is additionally seen as better for enhanced on the entirety of datasets aside from mushroom dataset. X-axis represented as seconds and Y-axis represented as number of identified patterns. For the retailer dataset, the development was undoubtedly brought down the base value edge esteems.

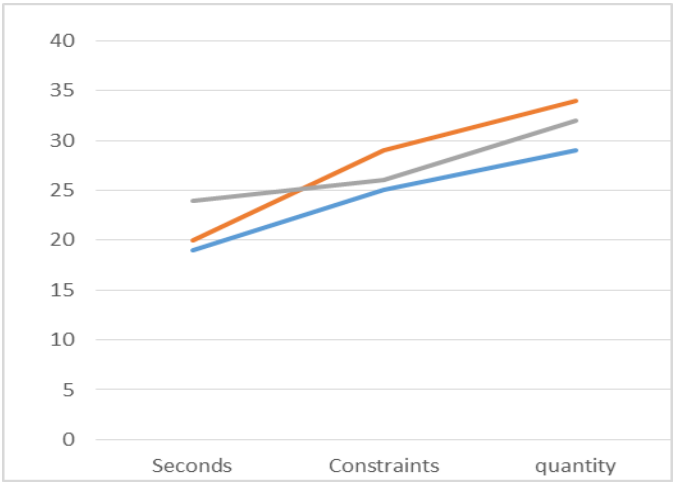

Figure 2. Runtime evaluation for Minimum Utility threshold

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 13062 - 13068

V. CONCLUSION

In this research paper, we projected a procedure to mine HUIs with negative utility. The majority of conventional HUIs mining calculations mines the principles from datasets having just positive utility value. But in actuality, negative utility is significant. In writing, just HUINIV-Mine and FHN procedures are projected to settle the negative utility itemsets mining issue. Be that as it may, the standards mined by the two procedure incorporate extremely enormous number of small itemsets and takes lot of time with repeated scanning of data. We overcome this problem by introducing new algorithm in negative HUIs mining. High Utility itemsets yields higher profit value. It is mainly applicable in online shopping like flipcart and amazon. The main aim is to help the business organisation to yield high profit and also attracting the customers and giving them more satisfaction.

References

1. Lin, J. C. W., Fournier-Viger, P., & Gan, W. (2016). FHN: an efficient algorithm for mining high-utility itemsets with negative unit profits. Knowledge-Based Systems, 111, 283-298.
2. Krishnamoorthy, S. (2018). Efficiently mining high utility itemsets with negative unit profits. Knowledge-Based Systems, 145, 1-14.
3. Suresh. K and Pattabiraman. V, "An improved utility itemsets mining with respect to positive and negative values using mathematical model", International Journal of Pure and Applied Mathematics, Volume-101 No-5, Page No-763-772, 2015.
4. Mohammadi, M., Al-Fuqaha, A., Sorour, S., & Guizani, M. (2018). Deep learning for IoT big data and streaming analytics: A survey. IEEE Communications Surveys & Tutorials, 20(4), 2923-2960.
5. Suresh.K and N. Nalini, "A Fuzzy Inventory Model for solid profits", International Journal of Pure and applied mathematics. International Journal of Pure and Applied Mathematics, Volume-119, No-13, Page No-117-122, 2018
6. Nguyen, L. T., Nguyen, P., Nguyen, T. D., Vo, B., Fournier-Viger, P., & Tseng, V. S. (2019). Mining high-utility itemsets in dynamic profit databases. Knowledge-Based Systems, 175, 130-144.
7. Suresh.K, and Pattabiraman. V, Reduction of large database and identifying frequent patterns using enhanced high utility mining", International Journal of Pure and Applied Mathematics, Volume-109, No-5, Page No-161-169, 2016
8. Singh, K., Kumar, A., Singh, S. S., Shakya, H. K., & Biswas, B. (2019). EHNL: An efficient algorithm for mining high utility itemsets with negative utility value and length constraints. Information Sciences, 484, 44-70.
9. [9] Nguyen, L. T., Vu, V. V., Lam, M. T., Duong, T. T., Manh, L. T., Nguyen, T. T., ... & Fujita, H. (2019). An efficient method for mining high utility closed itemsets. Information Sciences, 495, 78-99.
10. Pernabas, J. B., Fidele, S. F., & Vaithinathan, K. K. (2019). Enhancing Greedy Web Proxy caching using Weighted Random Indexing based Data Mining Classifier. Egyptian Informatics Journal.
11. Han, X., Liu, X., Chen, J., Lai, G., Gao, H., & Li, J. (2019). Efficiently Mining Frequent Itemsets on Massive Data. IEEE Access, 7, 31409-31421
12. Chen, Y., Zhao, X., Lin, X., Wang, Y., & Guo, D. (2018). Efficient Mining of Frequent Patterns on Uncertain Graphs. IEEE Transactions on Knowledge and Data Engineering, 31(2), 287-300.
13. Nguyen, L. T., Nguyen, P., Nguyen, T. D., Vo, B., Fournier-Viger, P., & Tseng, V. S. (2019). Mining high-utility itemsets in dynamic profit databases. Knowledge-Based Systems, 175, 130-144.
14. Yun, U., & Kim, D. (2017). Mining of high average-utility itemsets using novel list structure and pruning strategy. Future Generation Computer Systems, 68, 346-360.
15. Djenouri, Y., Djenouri, D., Belhadi, A., & Cano, A. (2019). Exploiting GPU and cluster parallelism in single scan frequent itemset mining. Information Sciences, 496, 363-377.
16. Aryabarzan, N., Minaei-Bidgoli, B., & Teshnehlab, M. (2018). negFIN: An efficient algorithm for fast mining frequent itemsets. Expert Systems with Applications, 105, 129-143.
17. Truong, T., Duong, H., Le, B., Fournier-Viger, P., & Yun, U. (2019). Efficient high average-utility itemset mining using novel vertical weak upper-bounds. Knowledge-Based Systems.
18. Lin, J. C. W., Ren, S., Fournier-Viger, P., Pan, J. S., & Hong, T. P. (2018). Efficiently updating the discovered high average-utility

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 13062 - 13068

itemsets with transaction insertion. Engineering Applications of Artificial Intelligence, 72, 136-149.

19. Nguyen, L. T., Vu, V. V., Lam, M. T., Duong, T. T., Manh, L. T., Nguyen, T. T., ... & Fujita, H. (2019). An efficient method for mining high utility closed itemsets. Information Sciences, 495, 78-99.
20. Pernabas, J. B., Fidele, S. F., & Vaithinathan, K. K. (2019). Enhancing Greedy Web Proxy caching using Weighted
21. Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I. A. T., Siddiqa, A., & Yaqoob, I. (2017). Big IoT data analytics: architecture, opportunities, and open research challenges. IEEE Access, 5, 5247-5261.
22. Shou, Z., & Di, X. (2018). Similarity analysis of frequent sequential activity pattern mining. Transportation Research Part C: Emerging Technologies, 96, 122-143.
23. Yun, U., & Kim, D. (2017). Mining of high average-utility itemsets using novel list structure and pruning strategy. Future Generation Computer Systems, 68, 346-360.

Published by: The Mattingley Publishing Co., Inc.