INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 1)
Available online at www.ijariit.com

Sliding Window Control Based High Utility Pattern Mining for Industrial Use

Swapnali Londhe
DR. D. Y. Patil Institute of Technology, Pimpri, Pune, Maharashtra
tejswapn@gmail.com

Rupesh Mahajan
DR. D. Y. Patil Institute of Technology, Pimpri, Pune, Maharashtra
mhjn.rpsh@gmail.com

Abstract: Value of the item set is considered as a utility of that item set and find out the high utility of the item set is the aim of utility mining.Some time database parameters are considered to find out high utility pattern eg. Profit, cost etc. In day to day life, high utility pattern mining plays important role in applications. Its current hot topic in today's research area. Different existing algorithms are present in this area.It, first of all, identifies the candidate item sets by using their utilities and simultaneously identify the exact utility of that candidate pattern. The problem of using this algorithm is a large number of candidate item sets are generated. But after computing exact utility it's clear that most of the candidate having no high utility.

For generating profitable product manufacturing plan it's very important to understanding the customer preferences in an industrial area. One of the approaches in pattern mining, by considering the quantity, quality and cost of each product for generating high profitable product set, which employed to find out high utility pattern. For establishing highly profitable.

Manufacturing plan which allows the corporation to maximize its revenue, high utility pattern mining is an important aspect. A large amount of stream data related to customer purchase behavior used for establishing manufacturing plan. Recent preference of the customers also helps in generating manufacturing plans. This survey work contains a list structure and a novel algorithm for generating high utility pattern over large data, on the basis of Sliding Window Control Mode. This approach avoids the generation of candidate pattern. Due to that algorithm not required a large amount of memory space as well as computational resources for verifying candidate patterns. Due to this, it's very efficient approach.

Keywords: Sliding Window Based Utility Pattern Mining; Manufacturing Plan; Industrial Systems; Frequent Pattern Mining; High Utility Patter Mining.

I. INTRODUCTION

Knowledge discovery within the database is the main aim of data mining. KDD works for finding interesting information from a large amount of massive data in an automatic way. To analyze complex industrial data and find out solutions for knowledge related problems. Classification and pattern mining like approaches has been developed for industrial area databases. To analyze complex data and find out valuable pattern data, pattern mining has been mainly used. Simple pattern mining, frequent pattern mining approaches are used when the availability of meaningful information related to frequency and minimum support present. If consider pattern is an important information than their frequency is a support and it's greater than or equal to the given minimum support. Within FPM, product parameters information does not indicate individually because it's assumed that all items have binary information. Therefore it's very complex to analyze industrial data with product parameter like profit, price, and quantity purchased by the customer.

Discovery of itemset with a high utility like profits of the transactional database required for mining high utility item sets. A large number of algorithms are present which used for generating high utility pattern but it indirectly generate a large number of candidate pattern also due to that mining performance of the system is a decrease in term of execution time and space requirement. When the database contains a large number of transaction and long high utility itemset than performance parameter become worse.

For increasing revenue of the plant, it's important for a manager to recognizing manufacturing plan according to customer need. For that purpose, the manager needs to know which product generate high profit within the plant. To achieving this aim, the manager needs to use high utility pattern mining approach. Pattern information related to the set of items which are profitable, present towards manager also helps in generating more profitable plans. Due to that production of these product increase and revenue of the plant can be maximized.

There are three main methods present in the pattern mining area, sliding window, damped window, landmark window. Within sliding window method stream data divides in to multiple chunks called batches. Recent batches are only used for pattern mining approach. In the damped window, new incoming data treated as more important than the previous one. Due to that importance of data decays as time passes. Within landmark window method, its only use the data in-between specific time period like time period between current time point and specific time point.

Within this approach novel, data structure named SHUP-List has proposed mainly used for storing recent batches information. Also, it's containing a proposed algorithm for efficiently updating remaining utilities. For reducing search space it's introduce pruning techniques. Also introducing proposed algorithm named as SHUPM (Sliding Window based High Utility Pattern Mining).

II. MOTIVATION

This work is motivated by the following problem in industrial areas. In a manufacturing plant, a manager needs to reorganize a manufacturing plan so that the revenue of the plant can be maximized. Therefore, the manager wants to know what products create high profits for the plant. In order to achieve this purpose, HUPM is a proper solution. If the manager obtains pattern information indicating that a set of certain products creates high profit, the manager can establish more profitable manufacturing plans, which increase the production of these products so that the revenue of the plant can be maximized.

III. METHODOLOGY

A sliding window method divides stream data into multiple chunks called batches and uses only recent batches for mining patterns. The main methodology of given Proposal contain following Steps:

Step-1

Proposing a novel list structure named SHUP-List (Sliding window based High Utility Pattern list) that can maintain the information of recent batches and handle them batch by batch keeping reliability.

- Capturing information of recent batches in list structures.
- A list of each pattern is composed of one or more entries representing transactions, which contain the Pattern as its sub-pattern.

Step-2

- Suggesting a novel strategy that allows the proposed algorithm to efficiently update remaining utilities in SHUP-Lists according to a TWU ascending order.

Step-3

High utility pattern mining with sliding window technique.

- Introducing a new pruning technique by combining two lists in order to reduce the search space of HUPM based on the sliding window model

Step-4

Analysis of the given pattern mining algorithm.

Step-5

Performance evaluations with respect to previous techniques such as SHU-Growth etc.

- Run Time Test
- Memory Usage Test
- Scalability Test

With the help of this tests, performance efficiency of given approach is find out.

IV. REVIEW OF LITRATURE SURVEY

In this section, we present the different approach and techniques given by different authors regarding sliding windows control based pattern mining is as follow.

[1] Propose a list structure and novel algorithm for finding high utility pattern over data stream on the basis of sliding window mode. In this approach, candidate patterns are not generated. Due to that large memory space and more computational resources not required for candidate pattern verification. The algorithm used within this approach work efficiently. In this, various real time databases are used for the experimental purpose for generating a result in terms of runtime, memory usability, and scalability.

[2] Specifies two algorithms for generating a pattern from the database. Its shows how the best features of the two proposed algorithm can be combined into one hybrid algorithm called Apriori Hybrid. For a large number of transactions, it's very important to approach. Apriori Hybrid algorithm having excellent scale-up property.

The novel tree structure for efficiently performs incremental and interactive HUP mining. Incremental HUP lexicographic tree is the first tree structure arranged according to an item's lexicographic order. Without any reconstructing operation, it can capture incremental data. IHUP transaction frequency tree is the second tree structure. By arranging items according to their transaction frequency, it's obtaining a compact size. It's also reduces mining time. IHUP- Transaction weighted utilization tree is based on the TWU values of items in descending order. It's very efficient structure for scalable, incremental mining [3].

A novel method Temporal High Utility itemset mine propose for mining temporal high utility itemset from large dataset [4]. The main contribution of this approach is to reduce the generation of candidate itemset and effectively identify the temporal high utility itemsets. Due to that execution time can be reduced in mining. With this approach, item sets can be achieved effectively with less memory space and execution time.

An approach based on the analysis of item co-occurrences to reduce the number of join operations that need to be performed [5]. This approach is called as Fast High Utility Miner (FHM). It's six times faster than the state-of-the-art algorithm as well as it reduces the number of join operations.

[6] Propose an algorithm called High Utility Itemset-Miner. It's used for High Utility Itemset Mining. It's having a novel structure called utility list, used for storing utility information about an itemset and the heuristic information for pruning the search space of HUI-Miner. Its efficiently mine high utility itemset from the utility list generated from the dataset. It's avoiding costly generation of utility computation. Paper also contains a comparison of HUI-Miner with Stat-of-the-Art algorithms on various databases. The result of this algorithm in terms of running time and memory consumption are explained.

In these paper [7] author Present Two-Phase algorithms for efficiently prune down the number of candidates and precisely obtain the complete set of high utility itemset. It's very efficient in terms of speed and memory cost on a real database. It's also efficient for large databases.

A novel algorithm called FP-CDS that can capture all frequent closed itemsets and a new storage structure called FP-CDS tree [8]. It's used for adjusting evolution of itemsets frequencies over time. For updating units, several basic windows of the landmark window used. Itemsets in each basic window are mined and stored in FP-CDS tree based strategies.

An Explained two algorithms, utility pattern growth (UP-Growth) and UP-Growth+, for mining high utility itemset with effective strategies [9]. It's having a data structure called UP-Tree used for mining high utility itemsets information. Due to that candidate itemsets can be generated efficiently with two scan on database. In this paper comparison between UP-Growth and UP-Growth+ presented. This approach not only works on candidate pattern reduction but also its increase the overall performance in term of runtime when database contains lots of long transactions.

10] Explain method for mining association rules that reflect the behavior of past users was proposed for an adaptive search engine. RDF model used for retrieving users behaviors. Using association rules user behavior extracted successfully. Generation low with RDF schema was also proposed for absorbing a variety of user's behaviors.

V. CONCLUSION

Here in this paper, we survey a new approach for pattern mining. Sliding Window Control Based High Utility Pattern Mining approach contains a novel list structure named SHUP-List for storing mined data. Also, it's efficiently updating remaining utilities. The approach promises to, understand the recent purchase preferences of customers. Therefore, it is suitable for establishing profitable manufacturing plans in various industrial areas as well as Algorithm does not consume huge computational resources for verifying candidate patterns because it can avoid the generation of candidate patterns**.** In this system, proposed for extracting and understanding the user patterns from the huge data that generates the useful patterns using Sliding Window methodologies and the frequent items algorithm for effective pattern recognition, to find up-to-date information.

VI. ACKNOWLADGEMENT

I would like to take this opportunity to express my thanks to my guide Prof. Rupesh Mahajan for his esteemed guidance and encouragement. His guidance always helps me to succeed in this work. I am also very grateful for her guidance and comments while designing part of my research paper and learnt many things under his leadership.

REFERENCES

1. Yun, Unil, Gangin Lee, and Eunchul Yoon. "Efficient High Utility Pattern Mining for Establishing Manufacturing Plans with Sliding Window Control." IEEE Transactions on Industrial Electronics, 2017.
2. R. Agrawal and R. Srikant. "Fast Algorithms for Mining Association Rules." Proceedings of the 20th International Conference on Very Large Data Bases. [Online], 1994.
3. C. Ahmed, S. Tanbeer, B.-S. Jeong and Y.-K. Lee. "Efficient Tree Structures for High Utility Pattern Mining in Incremental Databases." IEEE Transactions on Knowledge and Data Engineering. [Online], 2009.
4. C.-J. Chu, V. Tseng, and T. Liang. "An efficient algorithm for mining temporal high utility itemsets from data streams." Journal of Systems and Software. [Online], 2008.
5. P. Fournier-Viger, C.-W. Wu, S. Zida, and V. Tseng. "FHM: Faster High-Utility Item set Mining Using Estimated Utility Co-occurrence Pruning." Proceedings of 21st International Symposium ISMIS. [Online], 2014.
6. M. Liu and J. Qu. "Mining high utility item sets without candidate generation." Proceedings of the 21st ACM international conference on Information and Knowledge management. [Online], 2012.
7. Y. Liu, W. Liao, and A. Choudhary. "A Two-Phase Algorithm for Fast Discovery of High Utility Item sets." Proceedings of 9th Pacific-Asia Conference PAKDD. [Online], 2005.
8. X. Liu, J. Guan, and P. Hu. "Mining frequently closed item sets from a landmark window over online data streams." Computer & Mathematics with Applications. [Online], 2009.
9. V. Tseng, B.-E. Shie, C.-W. Wu, and P. Yu. "Efficient Algorithms for Mining High Utility Itemsets from Transactional Databases." IEEE Transactions on Knowledge and Data Engineering. [Online], 2013.
10. Y. Takama and S. Hattori. "Mining Association Rules for Adaptive Search Engine Based on RDF Technology." IEEE Transactions on Industrial Electronics. [Online], 2007.
11. C. Ahmed, S. Tanbeer, B.-S. Jeong and Y.-K. Lee. "Efficient Tree Structures for High Utility Pattern Mining in Incremental Databases" IEEE Transactions on Knowledge and Data Engineering. [Online], 2009.
12. C. Ahmed, S. Tanbeer, B.-S. Jeong, H.-J. Choi. "Interactive mining of high utility patterns over data streams. Expert Systems with Application [Online]. 2012.

