IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 10, Issue 1, January 2020, ||Series -III|| PP 19-24

Comparative Study of Different High Utility Pattern Mining Techniques

1[st]Ms. S. S. Naghate, 2[nd] Dr. Mrs. S. S. Sherekar, 3[rd]Dr. V. M. Thakare
dept. Computer Science & Engineering SGBAU, Amravati Amravati, India
dept. Computer Science & Engineering SGBAU, Amravati Amravati, India
dept. Computer Science & Engineering SGBAU, Amravati Amravati, India
Received 01January 2020; Accepted 15 January 2020

Abstract: mining high utility pattern is an important research area in data mining. Several business applications have been found to benefit from the discovery of high utility itemsets and association rules from transaction databases. A comprehensive survey and study of various methods in existence for high utility itemset mining, association rule mining with utility considerations have been presented in this paper. In these, the various mining techniques are used such as Incremental Mining of High Utility Patterns, High Average-Utility Patterns with Multiple Minimum Average-Utility Thresholds, Using Bio-Inspired Algorithms, Algorithm for Incremental and Interactive High Utility Itemset Mining, using Temporal-Based Fuzzy Utility Mining. However there are some issues that need to resolve these are discussed in this paper. Various algorithms are studied for the comparative analysis of high utility pattern mining algorithms and new method is proposed.
Keywords: High utility pattern mining, frequent itemset mining, fuzzy utility mining, utility threshold.

I. INTRODUCTION

Data mining is the process of retrieving useful patterns from a dataset. The purpose of this system is to find out High Utility Itemsets with high utility pattern mining. The term utility considers importance, interestingness or profitability of the items. High utility itemsets are those itemsets which have high utility greater or equal to a threshold value. High utility pattern mining is formulated as discovering the patterns whose utilities are no less than a minimum utility threshold [1]. Various types of data mining tasks can be performed on a database to extract useful patterns or models such as association rule mining (ARM) sequential pattern mining (SPM), high-utility itemset mining (HUIM), classification and clustering [2]. Generally the first phase of association rule mining (ARM), frequent itemset mining (FIM) has received considerable attention and been applied in various domains such as web-click stream analysis and business promotion in chain hypermarkets [3] [4]. As a result of the foregoing, the paper presents a research issue in the work to consider both the benefits from the sale of the items and amounts of the items as well as the transaction period of transactions to get temporal high utility itemsets in product databases [5].

In this paper, various utility mining techniques performs a comparative analysis including the data structures; threshold raising strategies with the HUI mining are also presented.

II. BACKGROUND

Many studies on high utility pattern mining itemsets have been done to develop the utility mining algorithms in past years. Such algorithms are:

Incremental Mining of High Utility Patterns presents a naive approach for mining a dynamic database with transaction insertion is to mine each newly updated database instance in its entirety, the algorithms improve the efficiency by avoiding repeatedly mining the legacy transactions [1].

HAUIM more fairly measures the utility of itemsets by considering their lengths (number of items) in which an itemset is a high-utility itemset if its utility is no less than a user-defined minimum high-utility threshold (count) [2].

In bio-inspired algorithm strategy of selecting discovered HUIs probabilistically, instead of maintaining the best values from population to population, improves the diversity of solutions within a limited number of iterations [3].

When the database is updated, an efficient tree structure IHUP-Tree is proposed for incremental and interactive HUIM. IHUP-Tree can be adjusted efficiently when a transaction is added into or deleted from a database, and incremental HUIM can be efficiently performed based on IHUP Tree [4].

An effective framework called temporal-based fuzzy utility mining is to give more attention to the transaction period of given items according to the concept of fuzzy utility mining [5].

The paper is organized as follows:

Section I Introduction. **Section II** discusses Background. **Section III** discusses previous work. **Section IV** discusses existing methodologies. **Section V** discusses attributes and parameters and how these are affected on mining techniques. **Section VI** proposed method and outcome of result. Finally **Section VII** Conclude this analytical paper.

III. PREVIOUS WORK DONE

In research literature, many utility mining technology have been studied to provide various high utility pattern mining algorithms and improve will the high utility itemsets with the average of high utility threshold.

JUNQIANG LIU et. al., (2019) [1] proposed a new algorithm that introduces several novel ideas to efficiently mine high utility patterns in the dynamic databases. The algorithm also improves relevance-based pruning and upper-bound-based pruning, and introduces quick merge of identical transactions. The other pruning is to efficiently determine whether a pattern is a high utility pattern in the original database, which targets a scalability issue with the prior algorithms.

JERRY CHUN-WEI LIN et. al., (2018) [2] proposed to discover all HAUIs using multiple minimum high-average utility thresholds based on the average-utility (AU)-list framework. Each item can be associated with a user-defined minimum high average-utility threshold, which is more realistic than the original HAUIM task in terms of runtime, memory usage, number of candidates and scalability.

WEI SONG et. al., (2018) [3] proposed bio-inspired algorithms, the strategy of selecting discovered HUIs probabilistically, instead of maintaining the best values from population to population, improves the diversity of solutions within a limited number of iterations.

Shiming Guo et. al., (2017) [4] proposed incremental and interactive HUIM, IHUP-Tree can be adjusted efficiently when a transaction is added into or deleted from a database, and incremental HUIM can be efficiently performed based on IHUP Tree.

WEI-MING HUANG et. al., (2017) [5] addresses these issues by presenting an effective framework called temporal-based fuzzy utility mining to give more attention to the transaction period of given items according to the concept of fuzzy utility mining.

IV. EXISTING METHODOLOGY

Many techniques and algorithms have been implemented over the last several decades. There are various techniques that are implemented i. e. on analysis of various high utility pattern mining algorithm, such as Incremental Mining of High Utility Patterns, High Average-Utility Patterns with Multiple Minimum Average-Utility Thresholds, Using Bio-Inspired Algorithms, Algorithm for Incremental and Interactive High Utility Itemset Mining, using Temporal-Based Fuzzy Utility Mining.

A. Incremental Direct Discovery of High Utility Pattern Algorithm:

Id2HUPC, which enumerates patterns by prefix extensions in a depth-first manner, reduces the search space by the relevance-based pruning, the upper-bound-based pruning, and the absence-based pruning, and quickly identifies patterns not in IHUPset by the legacy-based pruning. The utility of X in a database D, denoted by u(X; D), is the sum of the utility of X in every transaction containing X, i.e.

$$u(X, D) = \sum_{t \in TS(X,D)} u(X, t) = \sum_{t \in TS(X,D)} \sum_{i \in X} u(i, t).$$

Algorithm 1 Id^{2}HUP+ $(TS_{niCAUL}(\{\}, D_1), N, \mu, \theta)$

Input: $TS_{niCAUL}(\{\}, D_1)$ representing the original database D_1, the set N of new transactions with $D_2 = D_1 \cup N$, the minimum utility threshold μ, the materialization factor θ;

Output: $IHUP_{set}(\mu, D_2, D_1)$ high utility patterns in D_2 not in D_1;

construct $TS_{niCAUL}(\{\}, D_2)$ based on $TS_{niCAUL}(\{\}, D_1)$;

GrowPatterns$(\{\}, TS_{niCAUL}(\{\}, D_2), \mu, \theta)$;

$TS_{niCAUL}(\{\}, D_1) \leftarrow$ RefreshNICAUL$(TS_{niCAUL}(\{\}, D_2))$;

Algorithm1: Incremental Direct Discovery of High Utility Patterns

B. High Average-Utility Patterns with Multiple Minimum Average-Utility Algorithm:

This algorithm was first proposed to discover all HAUIs with multiple minimum high average-utility counts. Thus it suffers from the same drawbacks as the Apriori algorithm. It efficiently discovers HAUIs without candidate generation and without performing multiple database scans. MEMU Algorithm follows as:

1. calculates *auub* of each item;
2. *LMAU* *min*{*mau*(*i*1), : : : , *mau*(*ir*)};
3. if *auub*(*ij*) < *LMAU* then
4. remove *ij* from *D*;
5. recalculate the *auub* of each remaining item in *D*;
6. sort items in each transaction in *mau*-ascending order;
7. Searches (*MEMU*);
8. return HAUIs;

Algorithm2: MEMU Algorithm

C. Bio-inspired algorithm:

The standard of bio-inspired algorithms, the optimal values of the current population are not definitely retained in the next population all discovered HUIs are subjected to roulette wheel selection to determine the target of the next population. This improves in average the diversity within one population, and also enhances the efficiency and quality of mining. The utility of item ip in transaction Td is defined as

u (ip, Td)= p(ip) X q(ip, Td)

Where, the external utility p(ip) is the unit profit value of item ip and transaction Td.

D. Algorithm for Incremental and Interactive High Utility Itemset Mining:

Novel tree structures IHUIL- Tree and a utility database are represent the information of an original database. The length is the size of the longest transaction and the width is the number of transactions in a database adopts pattern-growth methodology and the bottom-up order; conditional database then a conditional pattern tree and a conditional utility database are constructed. Algorithm for the Construction of IHUIL-Tree and utility database follows as:

1. Create the root R of Tree, and label it as null
2. Initialize the header table Header of Tree
3. Set the prefix utilities of all the items in Header as 0
4. For each transaction Trans in TDB
5. Sort the items in Trans according to lexicographic order
6. Allocate a record from UDB for Trans
7. Call Insert Trans (Trans, R, Record, Header)

Algorithm3: IHUIL-Tree and utility database Algorithm

E. Temporal based fuzzy itemset mining

The proposed TP-TFU algorithm can be efficiently designed. It includes two main phases such as firstly, deriving candidate if their high temporal fuzzy utility ratios, it satisfy the threshold and scanning of the temporal quantitative database to get the fuzzy utility itemsets. Given the membership functions of item iz, the fuzzy set fyz of the value vyz of iz in the temporal quantitative transaction Transy is represented as:

$$f_{yz} = \left(\frac{f_{yz1}}{R_{z1}} + \frac{f_{yz2}}{R_{z2}} + \ldots + \frac{f_{yzl}}{R_{zl}} + \ldots + \frac{f_{yzh}}{R_{zh}} \right)$$

V. ANALYSIS AND DISCUSSION

Id2HUPC invokes the Grow Patterns procedure to recursively mine high utility patterns as the prefix extensions and merges identical transactions and thus is more compact than the vertical data structure. These algorithm uses pseudo projection to compute the transaction set for every enumerated pattern, which greatly saves memory usage [1].

The developed HUINIV-Mine algorithm extends the TWU model to cope with negative unit profit values to mine HUIs efficiently and an effective itemset generation method was developed to avoid generating a large number of redundant candidates and to effectively reduce the number of data scans to find itemsets [2].

Bio- HUIF-PSO and Bio-HUIF-BA are more efficient than all the other algorithms, including the two exact HUIM algorithms, IHUP and UP-Growth shows that bio-inspired algorithms do not need to scan the database many times or construct trees or other structures to transform the original database, and can solve the HUIM problem within a reasonable time [3].

When an incremental database is inserted into an original database, it can use Algorithm to insert the transactions of the incremental database into the global lHUIL -Tree and global utility database constructed from the original database and also maintains an array in a global lHUh-Tree to link to the last node of each transaction [4].

Experiments were first executed to make a comparison of the proposed TP-TFU and the non-temporal TP-HFU in which no time periods were divided. Two parameters were tested such as Experiments were initially conducted for different threshold values and it goes towards the execution time [5].

TABLE 1: COMPARISONS BETWEEN VARIOUS HIGH UTILITY PATTERN MINING TECHNIQUES

Proposed Algorithm Techniques	Advantages	Disadvantages
Incremental Direct Discovery of High Utility Pattern Algorithm	Comparing various utilities instead of searching saved patterns improves the scalability as well as efficiency.	Decrease of the minimum utility threshold.
High Average-Utility Patterns with Multiple Minimum Average-Utility Algorithm	The TWU algorithm reduces the search space also it Improves the efficiency using a single minimum high utility threshold.	It suffer the problem of high memory complexity
Bio-inspired algorithm	Using these, diversity would be improved.	The bio-inspired algorithms do not need to scan the database many times.
Algorithm for Incremental and Interactive High Utility Itemset Mining:	The algorithm shows more than one orders of magnitude faster than the state-of-the-art algorithm.	The number of candidates is usually very huge.
Temporal based fuzzy itemset mining	The results demonstrate that the number of high temporal fuzzy utility itemsets is larger than that of high fuzzy utility itemsets.	The proposed approach needed more computation time.

VI. PROPOSED METHODOLOGY

Various utility pattern mining is a forcing to make lot of efforts as it deals with recent transaction database has been proposed in the following algorithm. It finds the average high utility itemsets (AHUIs) from the transaction database D, the minimum utility value minutil, minimum utility threshold t as input. The representative utility mining techniques, calculates AHUIs with the help of minimum utility threshold by comprehensiveness of all the steps of following algorithm.

Steps of Algorithm:
Algorithm1: **Utility_mining (D, t, B, utilcal, minutil)**
Input: A sequence of transaction database D, a minimum utility threshold t and a minimum utility itemsets minutil.
Output: The average of high utility itemsets with minimum utility threshold

1. Begin
2. Initialize D
3. for each transaction database D
4. set times := 1;
5. if times ≤ minutil
6. then
7. sort minimum itemsets in D according to **utility_trans(D, minutil, u)**
8. end if
9. times := times+1;
10. return minutil;
11. end for
12. end

Algorithm2: utility_trans(D, minutil, u)
Input: A sequence of transaction database D, a minimum utility threshold t, a set of data called batches Band a minimum utility itemsets minutil.
1. Begin
2. Set the utility value of all the items in D
3. If minutil ≤ t
4. Then
5. Minutil := minutil+1;
6. Return the average high utility itemsets
7. End if
8. End

Diagrammatic representation of proposed method is shown as follows:

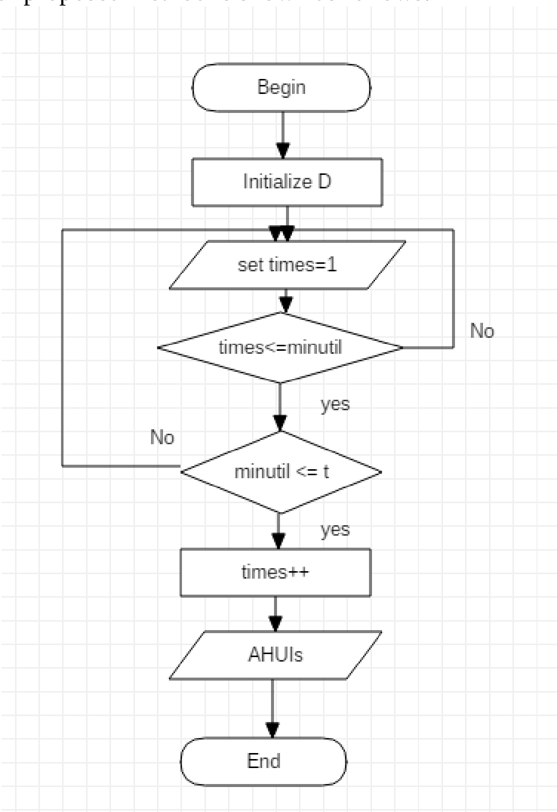

FIGURE 1: Flowchart of Proposed Methodology

VII. OUTCOMES AND POSSIBLE RESULTS

In this paper, the proposed method performs for the various high utility itemsets using pattern mining techniques when the data is in an information form. With the help of these algorithms, the proposed method calculates the output Average High Utility Itemsets AHUIs from a sequence of database with the help of minimum utility threshold. It also improves the scalability as well as efficiency including the execution speed.

VIII. CONCLUSION

The experimental result of proposed algorithm show that orders of magnitude has faster than the state-of-the-art algorithms, and is the most scalable one. It has been developed to mine AHAUIs while reducing the search space. The algorithms in terms of efficiency, the number of discovered HUIs, and convergence speed. A utility database used to maintain the information of itemsets in the original database and the incremental database. Although the proposed approach needed more computation time, it could get more useful itemsets which can be helpful in decision making.

IX. FUTURE SCOPE

In the future, especially for the analysis of customer transactions, items sometimes have negative weights or unit profits. The reviewed algorithms effectively mining high utility itemsets based on the various data structure and constraint techniques. Furthermore, the AHUIs algorithm will be examined using MapReduce or Spark. Additionally, the lifetime of each item such as the shelf-life of each product or specific time of each product in a particular holiday will be considered.

ACKNOWLEDGMENT

It gives me immense pleasure to express my gratitude toDr. S. S. Sherekar and Dr. V. M. Thakare, my Supervisors who provided me constructive criticism and a positive feedback during the preparation of this Research Paper.I am indebted to Dr. V. M. Thakare, Head of Department of Computer Science & Engineering at SGBAU and other teaching and non-teaching staff who were always there whenever I needed any help. Without them and their co-operation, completion of this research work would have been inevitable and their presence behind me is totally indispensable.

REFERENCES

[1]. JUNQIANG LIU, (Member, IEEE), XINYI JU, XINGXING ZHANG, BENJAMIN C. M. FUNG, (Senior Member, IEEE), XIANGCAI YANG, AND CHANGHONG YU, "Incremental Mining of High Utility Patterns in One Phase by Absence and Legacy-Based Pruning", IEEE, Vol.:7, 74168-74180, June 2019

[2]. JERRY CHUN-WEI LIN, (Member, IEEE), SHIFENG REN, AND PHILIPPE FOURNIER-VIGER, "MEMU: More Efficient Algorithm to Mine High Average-Utility Patterns with Multiple Minimum Average-Utility Thresholds", IEEE, Vol.:6, 7593-7609,March, 2018

[3]. WEI SONG AND CHAOMIN HUANG, "Mining High Utility Itemsets Using Bio-Inspired Algorithms: A Diverse Optimal Value Framework", IEEE, Vol.:6, 19568-19582,April 2018

[4]. Shiming Guo, Hong Gao**, "**An Efficient Algorithm for Incremental and Interactive High Utility Itemset Mining", 2017 2nd International Conference on Image, Vision and Computing,996-1001,2017

[5]. WEI-MING HUANG, TZUNG-PEI HONG, GUO-CHENG LAN, MING-CHAO CHIANG AND JERRY CHUN-WEI LIN, "Temporal-Based Fuzzy Utility Mining", IEEE, Vol.:5, 26639-26652, November, 2017

	**Ms. Snehal Naghate** has completed B.E. Degree in Computer Science & Engineering from Sant Gadge Baba Amravati University, Amravati, and Maharashtra. She is pursuing Master's Degree in Computer Science and Information Technology from P.G. Department of Computer Science and Engineering, S.G.B.A.U. Amravati. (e-mail id: naghatesnehal21@gmail.com)
	**Dr. Vilas M. Thakare** is Professor and Head in Post Graduate department of Computer Science and Engg, Faculty of Engineering & Technology, SGB Amravati university, Amravati. He is also working as a coordinator on UGC sponsored scheme of e-learning and m-learning specially designed for teaching and research. He is Ph.D. in Computer Science/Engg and completed M.E. in year 1989. He has done his PhD in area of robotics, AI and computer architecture. His area of research is Computer Architectures, AI and IT. He has published more than 150 papers in International & National level Journals and also International Conferences and National level Conferences. (e-mail id: vilthakare@yahoo.co.in)
	**Dr. Mrs. S. S. Sherekar** working as a professor in PG department of Computer Science and engineering at SGBAU from 1994 and having 24 years of teaching and research experience. Her area of research is Network security, Image Processing and working as supervisor for Ph.D. guidance. Completed one MRP of UGC one MRP of AICTE is going on. She has published more than 68 research paper in International & National level Journals. (e-mail id: ss_sherekar@rediffmail.com)

Ms. S. S. Naghate,et.al. "Comparative Study of Different High Utility Pattern Mining Techniques". *IOSR Journal of Engineering (IOSRJEN),* 10(1), 2020, pp. 19-24.