

Designing a Data Structure Utility List and High Utility Sequential Pattern for One phase in Data Mining

Kaushik Dattatraya Kulkarni¹, Rahul P. Mirajkar²

¹ Student, ME (CSE), Bharati Vidyapeeth's College of Engineering, Kolhapur, Maharashtra, India

² Asstt Professor, Bharati Vidyapeeth's College of Engineering, Kolhapur. Maharashtra, India

ABSTRACT

High utility item set mining finds item set from the database which have their utility no less than minimum threshold, the most significant task in data mining is the process to discovering the different type of pattern algorithm that generate the mining pattern. Sequence of database rather than strings and it can capture the set of sequential pattern. Data mining consist extracting information from data stored in databases to understand the data. Pattern mining consists of discovering interesting, useful, and unexpected pattern in databases.

Keywords : Data mining, Database, sequential pattern mining, High utility, item set

I. INTRODUCTION

In sequential pattern mining that pattern can maintain their sequential. Item set may generating in sequential manner without any duplication. Data mining consist extracting information from data stored in databases to understand the data. Pattern mining consists of discovering interesting, useful, and unexpected pattern in databases. Sequential pattern mining is a data mining is a data mining task specialized for analyzing sequential data to discover sequential pattern. Efficient Mining of High Utility Item set from large data set these algorithm search large transactional weighted utilization item in transaction database. It is used to mine the complete set of high utility item set. Implied a structure named High Utility of Pattern tree for maintaining essential information about utility mining.

Each node in enumeration tree will be contain generating different pattern that will be useful for utility of sequential pattern mining to maintain this pattern in linear data structure will be developed. It will be contain information about each item relevant of pattern. . High utility pattern that can be finding

the pattern from database that have a utility value. The utility of pattern defines defines the its importance and makes mined pattern.

1.1 MOTIVATION:

Efficient Mining of High Utility Item set from large data set these algorithm search large transactional weighted utilization item in transaction database. It is used to mine the complete set of high utility item set. Implied a structure named High Utility of Pattern tree for maintaining essential information about utility mining. It avoids scanning of multiple times generating pattern during mining process. Identifying better estimate of the utility value of pattern and systematic search of space for pattern using the estimate. Data structure which helps into computation of better estimate will improve the performance of mining algorithms by effectively search space.

II. OBJECTIVES

- Generating the High Utility Sequential Pattern in one phase

- Minimum Time required for High Utility Sequential Pattern

III. PROPOSED WORK

3.1 Scope

The Scope of the system that high utility pattern that can be finding the pattern from database that have a utility value. The utility of pattern defines the it is importance and makes mined pattern. Utility pattern that can be allows to associates relative importance to different item, and account for multiplicity of items.Each item has different profit values and a transaction can have multiple copies of an item that can be direct role of high utility pattern mining. Generating a large number of candidate patterns in the first phase which are then verified for their high utility property in the second phase. Identifying better estimate of the utility value of pattern and systematic search of space for pattern using the estimate. Data structure which helps into computation of better estimate will improve the performance of mining algorithms by effectively search space.

IV. IMPLEMENTATION MODULE

- ✓ Designing a Data Structure Utility List To Maintain Pattern
- ✓ Sequential Pattern Mining Without Candidate Generation in One Phase

3.2 Designing a Data Structure Utility List To Maintain Pattern

Each node in enumeration tree will be contain generating different pattern that will be useful for utility of sequential pattern mining to maintain this pattern in linear data structure will be developed. It will be contain information about each item relevant of pattern N. also N include summary information of that item in multiple transaction. Utility List will be

hold the utility of item that growing of pattern N that will be storing of utility list will be maintain. High utility pattern that can be finding the pattern from database that have a utility value. The utility of pattern defines the its importance and makes mined pattern. Utility pattern that can be allows to associates relative importance to different item, and account for multiplicity of items..Each item has different profit values and a transaction can have multiple copies of an item that can be direct role of high utility pattern mining. Generating a large number of candidate patterns in the first phase which are then verified for their high utility property in the second phase. Identifying better estimate of the utility value of pattern and systematic search of space for pattern using the estimate. Data structure which helps into computation of better estimate will improve the performance of mining algorithms by effectively search space.

Item	S	u	UItem	USpa
21731	1	25.5	139.12	25.5
22932	1	11.10000000000001	22.20000000000003	11.10000000000001
22933	1	11.10000000000001	22.20000000000003	22.20000000000003
22932	1	15.3	139.12	40.8
71053	1	20.34	139.12	81.14
84402E	1	20.34	139.12	81.48
84429F	1	20.34	139.12	101.82000000000001
84490B	1	22.0	139.12	123.82000000000001
84497A	1	54.08	54.08	54.08
85123A	1	12.299999999999999	139.12	139.12

Figure 1: Utility List To Maintain pattern

Figure 1 show that pattern that is useful for performing the utility list for generating sequential pattern mining for maintain the pattern in linear data structure that will created. In data structure contain list of structure that show that each information about pattern that include the summary of information of item that will done during the multiple transaction. Utility of item that growing of pattern that will be storing of utility list that will be maintain. During the transaction for pattern the list

of item that storing in database. Figure show there is item, s is stock, u is utility, UBitem is upper bound item, UBfpe is upper bound frequent element. Item can show their list of item number that can enter in transaction. Stock that define number of stock of item. Utility show the value in item upper bound item and frequent pattern that show the item that generated during transaction pattern displayed in frequently.

3.3 Sequential Pattern Mining Without Candidate Generation in One Phase

In this module algorithm will be developed to found utility pattern from sequential transaction database without candidate generation. Sequential pattern will be set of item sets structured in sequence database which occurs sequentially with a specific order. Sequential mining is process that will be extracting the pattern in database. The input data is set of sequence that will be performing transaction on sequence list. Transaction will be applied to set of item that will be generating the high utility pattern. High utility item set mining finds item set from the database which have their utility no less than minimum threshold, the most significant task in data mining is the process to discovering the different type of pattern algorithm that generate the mining pattern. Generating a large number of candidate pattern that do not appear in the input database and require to scan the original database many times. Maintain a large amount of intermediate candidates in main memory during the mining process, they are not memory efficient since they need to maintain large amount of intermediate candidates in main memory during the mining process. Sequence of database rather than strings and it can capture the set of sequential pattern. Chain of Accurate Utility List keeps the original utility information for each transaction, while the latter keep the utility estimate. This is the root because we are able to mine high utility patterns without generating candidates.

Algorithm for One Phase High Utility Pattern

Input: Database Transaction

Output: Pattern (sequential)

Steps: 1 Scan database transaction for item i

2 Compute $s[i]$, $u[i]$

3 Build transaction with descending order

4 Apply DFS

5 if pattern $[N] >$ threshold

Display pattern

if clouser property is true

output of every prefix extension of pattern

$[N]$ with relevant item

if singleton property is true

Output of relevant item and pattern

else each item i element of W

$C \leftarrow$ child node i

$TS \leftarrow$ search $[TS, i]$

Apply DFS

End if

End

Term Description: N denoted as total number of transaction

C denoted as child node

TS denoted as transaction set

W denoted as item set



Figure 2: Sequential Pattern Mining Without Candidate Generation in One Phase

Figure 2 show that generating a large number of candidate patterns in the first phase which are then verified for their high utility property in the second phase. Identifying better estimate of the utility value of pattern and systematic search of space for pattern using the estimate. Data structure which helps into computation of better estimate will improve the performance of mining algorithms by effectively search space. In utility mining item set share the frame work adopt a two phase candidate generation approach that is first find candidates of high utility pattern in the first phase and scan the raw data one more time to identify high utility patterns from the candidates in the second phase. The number of candidates can be huge which poses a challenge which is in the scalability and efficiency. Utility pattern from sequential transaction database without candidate generation Sequential pattern will be set of item sets structured in sequence database which occurs sequentially with a specific order. Sequential mining is process that will be extracting the pattern in database. The input data is set of sequence that will be performing transaction on sequence list. Transaction will be applied to set of item that will be generating the high utility pattern.

V. RESULT ANALYSES

One phase is process that transaction will be performing in set of item that generating the high utility pattern. For testing the performance of the proposed work dataset used is of Online Retail Transaction. For the proposed system performance evaluation is done on two parameters.

1) min U vs. Running Time

Here analysis is done by varying the value of min U and calculate the time required for each one. The table and its associated graph prove that as the value of min U is increasing it reduce the running time. Table 1 shows that min U value with their time.

min	Time(s)
	27
0.005	23
0.01	19
0.05	14
0.1	9

Table 1: Time are required to define the min U

With technique will be processed within the scanning that will be improved their performance and maintain their scalability and efficiency of transaction. High utility pattern will be generating in transaction process.

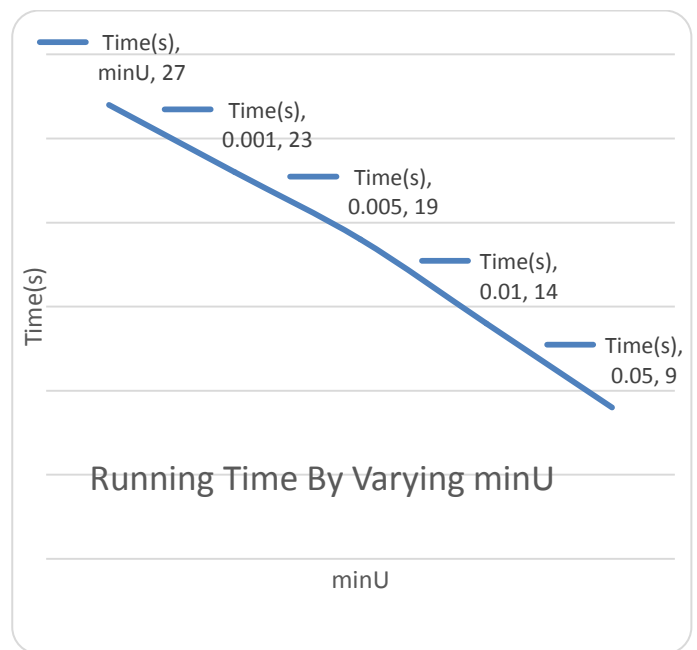


Figure 3: Running Time By Varying min U

Figure 3 shows that their min U that shows the minimum utility with Time takes to generating the sequential pattern. Their will minimum utility min U that defines as minimum time are required to generate pattern. Graph show that the time is reduced the increase the running time.

2) Number of Transactions vs. Running Time

In the proposed system database used is Online Retail transactions that is available from machine learning

databases. This dataset includes 45,222 Invoices that is number of transactions. By taking different sets of transactions, I have calculated running time required to process the data. It proves that one phase high utility pattern system takes slight increase in time as dataset goes an increase. Table 2 shows that Time and Number of Transaction value that taking to complete the process.

Number of	Time(s)
10000	180
15000	
20000	
25000	
30000	

Table 2: Time and No of Transaction are define to generate pattern

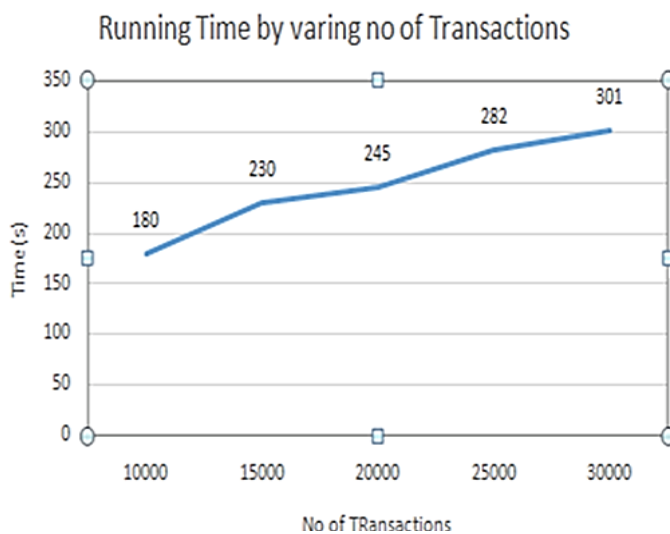


Figure 4: Number of Transactions vs. Running Time

Figure 4 shows that number of transaction with taking time for complete the process to generating the sequential pattern. Some of time takes the progress of graph increase step by step. Number of transaction increases the time is slightly increases.

VI. CONCLUSION

This shows that Mining may be useful for generating the patterns. Frequent pattern is important for data

mining that all the item in transaction that may represent in values. Finding the item set in data mining it is important. In high utilities patterns of item set may be consist of the group of item in transaction it is called as item set. High utility refers to the set of item that may refer to the database system. High utility item set may mined to the list may generate from mined database. High utility of pattern may identify the item set that utility may satisfy. Mining of high utility of item set is efficiently most of challenging task may consist of cost, quantity, profit that measure the utility.

VII. REFERENCES

1. Alva Erwin, Raj P. Gopalan, N.R.Achuthan," Efficient Mining of High Utility Item set From Large Data set" Nov 2015
2. CF.Ahmed, S Khairuzzaman Tanbeer, Byeong-Soo Jeong, Young-Koo Lee," Efficient Tree Structure for High Utility Pattern Mining in Incremental Databases"Nov 2015
3. Menghchi Liu, Junfeng Qu," Mining High Utility Item set Without Candidate Generation" Nov 2015.
4. Vincent S Tseng, Bai-En Shie, Cheng Wu, Philip S Yu," Efficient Algorithm For Mining High Utility Item set from Transactional Databases" Nov 2015
5. Yao H and Hamilton Hj "Mining item set Utility from Transaction database" Nov 2015
6. Yung Liu, Woe heng Liao and Alok Choudhary," A Two Phase algorithm for Fast Discovery of High Utility Item set" Nov 2015
7. HF.Li, H.Y.Hung, Y.C.Chen, Y.J.Liu and S.Y. Lee" Fast and Memory Efficient Mining Of High Utility Item set in Data Stream" Aug 2015
8. Mahdi Esmaeili and Fazekas Gabor," Finding Sequential Pattern from Large Sequence" May 2015