

SEQUENTIAL PATTERN MINING WITHOUT CANDIDATE GENERATION ON TRANSACTION DATABASE

Kaushik Dattatraya Kulkarni¹, Avani A. Karyakarte²,

¹ *Student of ME (Computer Science & Engineering),*

Bharati Vidyapeeth's College of Engineering, Kolhapur ,

² *Assistant Professor, Mechanical Department, GSMCOE, Balewadi, Pune.(India)*

ABSTRACT

Utility of an item set is considered as the value of this itemset, and utility mining aims at identifying the itemsets with high utilities. The temporal high utility itemsets are the itemsets whose support is larger than Temporal high utility item sets are the item sets whose support is larger than a pre-specified threshold in current time window of the data stream. Discovery of temporal high utility itemsets is an important process stream. Discovery of temporal high utility itemsets is an important process for mining interesting patterns like association rules from data streams. Data mining finding frequent item set. Pattern can be generated in sequentially.

Keywords: *Data Mining, Database, Sequential Pattern Mining*

I INTRODUCTION

Data mining is process of analysis that may large number of quantities of data in order to discover meaningful patterns and rules. 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.

The property that provides the pruning technique ones the infrequent item set are identified algorithm that not be identifies the long sets of item set. In market may be consist of item may have price each item in transaction have count that the item have brought. In item set in transaction may consist there is two aspect one may consist of single transaction called as internal utility and second one is different transaction may call external utility. 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.

1.1 Motivation

It may consist of generating candidate in first phase with high utility of pattern to scanning of data in that may identifies the high utility pattern. Two phase may consist of their may have large number of transaction that may reduce performance and scalability of transaction of item. Two phases may be consist of it can pruning by number of candidates it can take long time to generate high utility pattern because their large number of candidates that generate. Utility mining technique phase candidate generation approach is not scalable with large dataset and also suffers for scalability issue due to huge number of candidates generation that overcome the problem by using single phase without generating candidates.

1.2 Dataming System Model

High utility refers to the sets of item that may refer to the database system. High utility item sets may mine to the list may be generated from mined database. Mining of frequent item set that identify the sets of item that may generate in transaction on database. Number of transaction may contain set of item. All the frequent item sets may as frequent. Figure 1 show that it will be consist of Transaction database that performed the transaction operation on item set. Result of transaction operation stored in database. Transaction will be scanning of entire item set. There will be two part that is Reverse set Enumeration Tree. One Phase Sequential mining. Reverse set Enumeration Tree will be generating the tree structure that set of item order in reverse. One phase sequential pattern will be consist pattern may generate in sequence and generating sequence of item in one phase. One Phase Sequential Mining will be consist of two part that is DFS To Extract Pattern And Relevance item clouser property. DFS To Extract Pattern will be consist Extracting pattern by applying the DFS algorithm. Relevance item will be follow clouser property that is pattern of X and set of item of W then. $X \cap W = \phi$ Property will denoted as X, W and minimum utility that is $minU$ is satisfied.

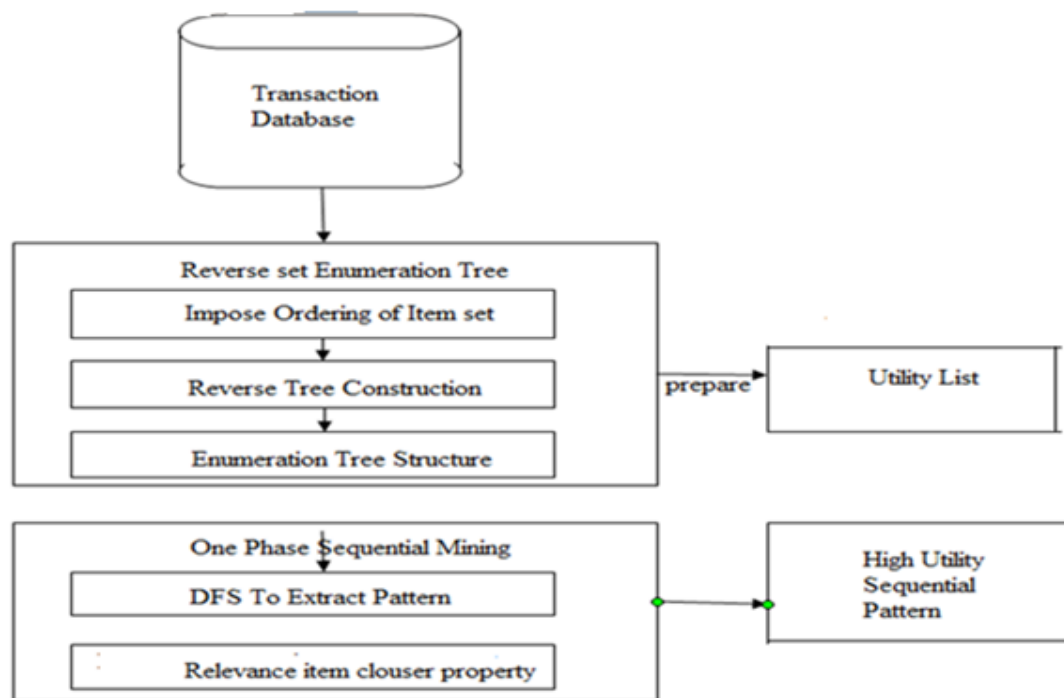


Figure 1: System Architecture

1.3 Implementation- of sequential Pattern Mining Phases

Basic data mining sequential pattern implementation phases are :

- Reverse Set Enumeration Tree
- Developing a DFS technique
- Designing a Data Structure Utility List To Maintain Pattern
- Sequential Pattern Mining Without Candidate Generation
- One Phase High Utility Pattern

II CASE STUDY-NEW SYSTEM: Sequential Pattern Mining

2.1 Reverse Set Enumeration Tree

For given database D the first reverse set enumeration tree will be generated. It is a tree that includes reverse lexicographic order. Here pattern will be search before its super set by using DFS. In this tree root will not be contain any item and each node N will be filled by pattern. The child nodes of N will be labeled by its sequential transaction item for sequential pattern mining.

2.2 Developing a DFS Technique

Reverse set enumeration tree patterns will be enumerated first to get the pattern. DFS technique will be proposed. Depth First Search will be related to exploring the depth of tree. The graph is traversed in top to bottom and left to right. The nodes at higher level in tree will be visited before nodes at lower level during the DFS. In this system root of reverse set enumeration tree will be transaction set of N pattern. Minimum upper bound will be transferred to DFS as input parameter. Depth First Search of a graph differs from exploration of vertex v is suspended as soon as new vertex will be reached. When new vertex will be explored then vertexes of v continue. The search terminates when all reached vertices will be fully explored.

2.2.1 Algorithm for DFS

Input: Node, Transaction set, Item order

Output: High Utility Pattern

Steps: 1 Label N as discovered
2 Find all edges from N in TS as W
3 if W is not labeled as discovered then
4 Recursively call DFS
5 for each unvisited neighbor W of N
6 if (not visited [W]) then
 Visited [W] = true
 End if
End

Term Description: N denoted as total number transaction

TS denoted as transaction set

W denoted as item set

2.3 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.

2.4 Sequential Pattern Mining Without Candidate Generation

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.

2.5 One Phase High Utility Pattern

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. One phase is process that transaction will be perform in set of item that generating the high utility pattern.

2.5.1 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

Term Description: N denoted as total number of transaction

C denoted as child node

TS denoted as transaction set

III. CASE STUDY-OLD SYSTEM

Two Phase System

In sequential pattern consist of generating pattern in sequentially. The old system show that pattern can be generated in two phase system means transaction can take more time that produced them item their scalability and efficiency minimize. Transaction may take lot of time to complete. In two phase scanning of data take lot of time generating the result of transaction. The solution is use one phase transaction that take less transaction time to complete transaction without generating the candidates. . Two phases may be consist of it can pruning by number of candidates it can take long time to generate high utility pattern because their large number of candidates that generate .

IV. COMPARATIVE ANALYSIS

The following table contains comparative features of existing and new solution

Table 1: Comparative features of existing and new solution based on

	Existing system	New system
Memory Space	100 MB	150 MB
Avg. query time	3 sec	< 0.5 sec
User defined queries	NO	YES
Detection of faulty data	NO	YES
Sophisticated analysis	NO	YES
Used for	Data processing and data analysis	Data analysis
Data security	NO	YES
Data granularity	Detailed data	Aggregated data
Dependency analysis	NO	YES
Complex statistical analysis	NO	YES

V. CONCLUSION

This show that provides the pruning technique ones the infrequent item set are identified algorithm that not be identifies the long sets of item set. In market may be consist of item may have price each item in transaction have count that the item have brought. In item set in transaction may consist there is two aspect one may consist of single transaction called as internal utility and second one is different transaction may call external utility. 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.

REFERENCES

- [1] Alva Erwin, Raj P. Gopalan, N.R.Achuthan,“ Efficient Mining of High Utility Item set From Large Data set” Nov 2015
- [2] C.F.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] H.F.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 Esmacili and Fazekas Gabor, ” Finding Sequential Pattern from Large Sequence” May 2015