Finding of K-Most Demanding Products using Greedy Algorithm

V.Jayanthi[1] K.K.Kavitha[2]
[1]M.Phil Student [2]H.O.D
[1,2]Department of Computer Science and Engineering
[1,2]Selvamm Arts and Science College, Namakkal, TamilNadu, India

Abstract— Data mining offers a facility to find most demanding products from the number of products based on the customer satisfaction in a particular organization or company to improve production plans [1]. The main objective of this research is to improve the execution time from the previous method by eliminating the products which are not possible to enter top-k list using minimum support threshold concept in greedy algorithm.

Keywords: Products, Customer satisfaction bit, Minimum support, Performance evaluation and Top k

I. INTRODUCTION

The biggest challenges now-a-days for a company to take decisions to produce a most demanding product by a customer. Customer's preference is an important factor to make a decision for production plan. The production plans have a high utility. The utility of a production plan is calculated based on the expected number of the whole customers for the particular product [1]. Manual data analysis has been around for sometimes now, but it creates bottleneck for large data analysis. Data mining allows user to analyze data from different dimensions categorized it and summarize the relationship, identified during mining process.

The main task of data mining is applying various kinds of methods and algorithms to discover and extract patterns of stored data [4].

Proposed research presents data mining process for evaluating the expected number of customers for a particular product by greedy algorithm with minimum support count. The minimum support count is there for filtering the product that is not possible to enter top-k list.

II. EXISTING SYSTEM

In the existing system all the candidate product (CP) details, customer requirements, expected number of total customers for existing product details are uploaded as input and stored into the database. The customer satisfaction bit string of candidate products and the expected number of total customers for each product are calculated. Finally Ranking function is used to find the top k-MDP (Most Demanding Product) using expected number of customer's value.

A. Drawbacks of Existing System:

All the candidate product details are stored into the database. Stores all the candidate product details make memory wastage, because all the candidate products are not possible to enter the top k list.

Calculating the expected number of customers for each candidate product make time consuming. It will affect the execution time.

III. PROPOSED SYSTEM

To speed up the execution process partition concept is used to split the candidate product dataset into number of datasets. Another proposed system is fixing the minimum support threshold value, it is the value given as an input by the user.

A. Advantages of proposed system:

Eliminate the candidate products, which are not possible to enter the top k list using minimum support threshold. It leads to speed up the execution process and reduce the execution time.

IV. DESIGN AND WORKING

The methodology will be followed for the current system is explained in this section. The flow of the methodology is given diagrammatically as follows in Figure 1.

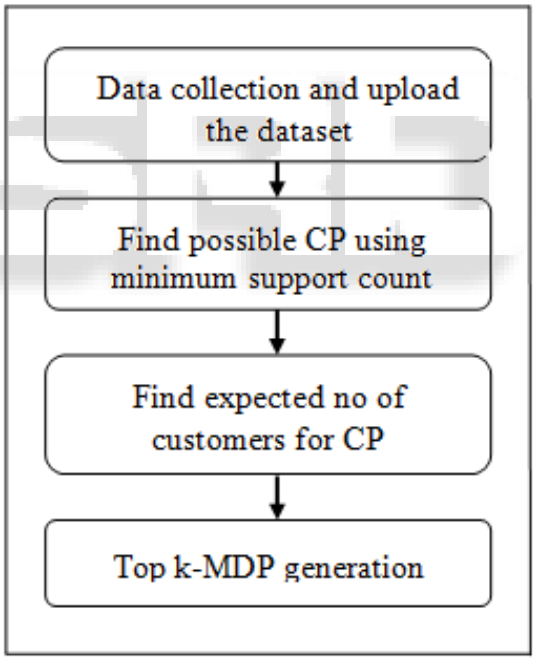

Fig. 1: Flow Diagram For Proposed Research

A. Data Collection And Upload The Dataset:

The first step of the proposed research is to design sample and gather product data. It is used to construct the dataset needed for data mining process. The dataset consists of 300 records by collecting people from various places for a period of 1 month. Then the data's are transformed into processing data and uploaded into database.

B. Find Possible Products To Enter Top K List:

In the second step compute the satisfaction bit strings of candidate product for all the customers using customer requirements and candidate product (CP) data. Then calculate the bit count for each CP.

Then minimum support (min_sup) count is applied for this count, it eliminate the CP, which are not satisfy the min_sup count. The other products are move on to the next stage.

C. Find expected number of customers for CP:

In the third step find the maximum expected number of customers for the selected candidate product with the help of existing product and customer satisfaction bit.

D. Top k-MDP generation:

In the final step get the value of k from the user and then perform the comparison between the expected numbers of values for the CP and produce the result and execution time.

V. IMPLEMENTATION

```
1.  i=0
2.  For each CP compute the satisfaction bit
    and their count
3.  Compute the bit count.
4.  Get the min_sup value
5.  Eliminate the product that does not
    satisfy the min_sup value
6.  Find the expected number of customers
    E(S, C) for selected products.
7.  i=i+1
8.  Get the top k value.
9.  If i<=k
    {
        Insert E(S, C) into the top k list.
    }
10. Else if E(S,C)>smallest value in the top
    k list
    {
        Replace the smallest value with
    E(S, C).
    }
11. Put the corresponding CP in the top k
    list.
12. Return the top k list.
```

Fig. 2: Pseudocode for top k-MDP generation

VI. PERFORMANCE EVALUATION

In this study there are two different databases are used. The database names are sample1 and sample2.Sample1 database contain fifty customers detail and five product details. Sample2 database contain fifty customers detail and ten product details.

Algorithm	Sample 1	Sample 2
Existing algorithm	0.35	0.39
Proposed algorithm	0.27	0.32

Table 1: Result of time efficiency in seconds

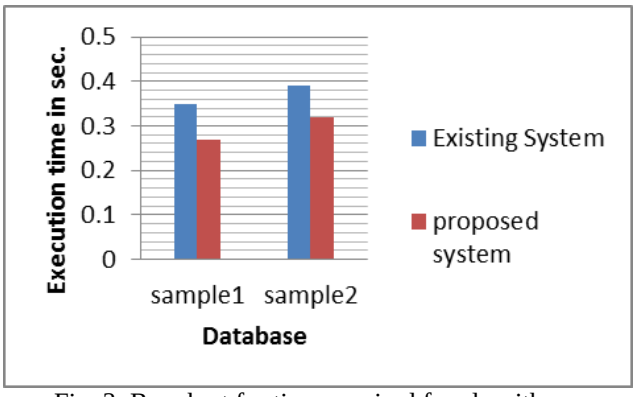

Fig. 3: Bar chart for time required for algorithms.

VII. CONCLUSION

In the proposed research the top k-MDP generation is determined with the minimum support concept and compared with the existing system in the terms of execution time. Two applications were developed, one is for existing algorithm, and another one is for proposed algorithm. Finally the whole execution time has been calculated for both application and compared. Existing algorithm takes 0.35 seconds for sample1 database and 0.39 seconds for sample2 database. Proposed algorithm takes 0.27 seconds for sample1 database and 0.32 seconds for sample2 database. The comparison demostrate the proposed system reduce the execution time.

REFERENCES

[1] Chen-Yi Lin, Jia-Ling Koh, and Arbee L.PChen,"Determining k-Most demanding prooducts with Maximum Expected Number of Total Customers",IEEE Transactions on KnowledgeandData ngineering,Vol.25,No. 8,August 2013.

[2] Kleinberg.J, C.Papadimitriou,and P.Raghavan, "A Microeconomic View of Data Mining", Data Mining and Knowledge Discovery,vol.2 no. 4,pp.311-322,1998.

[3] Mankiw.N.G, Principles of Economics, fifth ed. South-Western College Pub, 2008.

[4] Priyanka Rahul Patil ,Dr. R.S.Kamath "Assessment of Graduate Students' Performance using Data Mining: Proposed Research", IJSRD, vol. 4,Issue.05,2016.

[5] Tanz.K.L, P.-K.Eng, and B.C.Ooi, "Efficient Progressive Skyline Computation ", Proc. 27th Int'l Conf. Very Large Data Bases, pp. 301-310,2001.

[6] Zhang.Z,L.V.S.Lakshmanan, and A.K.H. Tung, "On Domination Game Analysis for Microeconomic Data Mining", ACM Trans. Knowledge Discovery from Data, vol. 2, no. 4,pp. 18-14,2009.

