ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

Obtaining K-Most Demanding Product Using Exact Top K Algorithm

Vidya R. Warghade, Prof. S. Pratap Singh

P.G. Student, Dept. of Computer Engineering, SP.'S Institute of Knowledge, College of Engineering, Shirur, Savitribai Phule Pune University, Pune, India

Asst. Prof., Dept. of Computer Engineering, SP.'S Institute of Knowledge, College of Engineering, Shirur, Savitribai Phule Pune University, Pune, India

ABSTRACT: it is very much basic to decide what items to create for accomplishing most extreme number of clients. That is the reason Product planning is a fundamental stage; attributes which are impacting the conclusions are should have been made to the middle to get the amended conceivable outcomes. For the same reason issue of production plans known as k-Mist Demanding Products (k-MDP) is formulated. To choose which items to implement, producers need to investigate the customers' requirements. An set of clients asking for a specific kind of items with various attributes, an arrangement of existing results of the type, a set of candidate products, and a positive number k are given. This is to help the organization to choose k item form the candidate products like total count of overall clients regarding to the k item is increased. At the time when the attribute count for the item is 3 or more then 3, then the problem became NP hard. Algorithms are given for finding proper as well as optimal solution for this issue. Greedy algorithms provide the optimal solution. For getting optimal solution algorithms makes use of upper bound pruning methods. Furthermore, Exact Top K algorithm is developed for finding the k-Most Demanding Products to increase the expected number of customers.

KEYWORDS: k-MDP, decision support, production plan.

I. INTRODUCTION

In the Microeconomics, problems such as how clients and producers settle on their choices and how they impart or communicate in market are analyzed [1]. Requirements of the clients are the vital part in formulation of product planning which transforms into one unimportant thing in field of microeconomics. While creating the production plans, it is essential for organizations to perceive the one with highest values or utilities.The cost or utility of production plan can be considered as a function which demonstrates the communication or interaction amongst organization and clients [2].The problem analyzed in this paper is to perceive the production plan for which will have most astounding cost or utility for an organization. The production plan is evaluated according to the normal number of clients for chosen items.

Assume the circumstance of the real estate property at a city, where the distance to a school and to a metro station are major requirements of the clients requesting a property. To settle on a suitable market choice, a rental organization has accumulated the prerequisites of the distance to a school and to a metro station from the clients. Likewise there are some current rental properties. Presently assume that the rental organization has an arrangement of properties whose locations. The administrator of the rental organization needs to pick k properties to battle with the present properties for rental.

For getting most extreme advantage, a technique is to get greatest expected number of the all clients for the k selected properties. It is normal that each client will choose one of the rentable properties satisfying his/her requirements. Right when more than one investment property satisfies the requirements of a customer, the customer will choose one of the properties according to his/her preference. For the effortlessness, it is normal that a client will choose any qualified rentable property with identical likelihood [3].

According to the above application, the problem regarding to k-Most Demanding Product (k-MDP) searching is characterized. An group of clients asking for a specific kind of items with various attributes, an group of existing

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

results of the kind, and an group of candidate items are given, the expectation is to help the organization to pick k items from these hopeful items to such an extent that the normal number of the overall clients for the k items is supported. Let EP and CP represent the set of existing items and the set of hopeful items. Furthermore, kCP represents to the set of k items chosen from CP, cp represents to a candidate product in kCP, and c represents to a client whose necessities are satisfied by cp.

The possibility for c picking cp is inverse proportional to the total number of items, including EP and kCP, which satisfy c. In this way, the normal number of the clients for cp is influenced not simply by the amount of clients satisfied by cp additionally the total number of various items that satisfy a similar set of clients. See that it is possible that the items in kCP will compete with each other on the off chance that they satisfy a similar set of clients. In this way, no straightforward procedure can be connected to find the set of k candidate items with the most expected number of the aggregate clients. The best strategy to give a beneficial and effective calculation for finding the k-MDP finding issue is the goal of this paper.

II. RELATED WORK

In Microeconomics, how clients and producers settle on choices and how they communicate in organizations is contemplated [1]. Client preference is a noteworthy component in settling on choices of item deals, which in this manner transforms into one most essential worry in microeconomics. Kleinberg et al. [2] ensured that couple of microeconomic problem can be settled by data mining frameworks, which move the analysts in the database gathering to handle the microeconomic problem. Due to various applications, the studies related to the microeconomic issues can be sorted into three sorts including the potential clients finding, the item points of interest finding, and the item positioning. The writing review is partitioned into three classifications which are: reverse k-nearest neighbor query, reverse skyline query, reverse top-k query.

Reverse k-nearest neighbor query:

E. Achtert et al [4] exhibited the MRkNNCoP-Tree the index structure for reverse k-nearest neighbor (RkNN) search in metric spaces where the cost for k is given at query time. Their list is centered around the pruning power of the kNN distances of the database focuses. They introduced to a system to surmise these kNN distances by a function conservatively and progressively to dodge an enormous storage overhead. They showed how these estimate capacities can gainfully be resolved and how any tree-like metric record structure can be amassed with this collected information. They showed that their MRkNN-CoP-Tree proficiently supports the generalized RkNN search in in random metric spaces. In particular, their technique yields a tremendous accelerate over the sequential search and other indexing solutions. Likewise, they demonstrated that their proposed thoughts are moreover applicable for Euclidean RkNN search. They have exhibited that their MRkNNCoP-Tree performs superior to anything existing answer for RkNN search for Euclidean vector data.

Y. Tao [5] built up the primary general method for recovery of a arbitrary number of reverse nearest neighbors in various measurements. Furthermore its adaptability and applicability, their answer is better than the past procedures in like manner to the extent profitability and versatility. A fascinating intriguing for future work is to conform their way to deal with various assortments of RNN issues. Facilitate, at present there does not present any expense model for evaluating the execution time of RNN frameworks. The progression of such a model won't simply ease question enhancement, yet may similarly reveal new issue properties that could provoke fundamentally quicker algorithms.

A Reverse k-Nearest-Neighbor query finds the items that are influenced by the questioning object. It can be utilized as a part of Location-Based Services to settle location related issues. W. Wu et al [6] have presented answers for assessing RkNN questions on location data. They describe RkNN question's inquiry territory and proposed a algorithm known as FINCH to process it concentrated on the inquiry and an set of data objects. FINCH is then used as a piece of their RkNN solutions for channel and limits the search space for result applicants. They also displayed a technique for applying (monochromatic) RkNN algorithms to evaluate bichromatic RkNN inquiries.

Reverse skyline query:

E. Dellis and B. Seeger [7] given Reverse Skyline Queries (RSQ). At the point when an set of information focuses P and an inquiry point q are given, a RSQ gives back information objects which have the question protest in the arrangement of their dynamic horizon. It is the complimentary issue to that of finding the dynamic skyline of a query object. Such kind of dynamic skyline compares at to the skyline of a changed data space where point q transforms into

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

the cause and all focuses are spoken to by their separation to q. To figure the reverse skyline of an inquiry point, they displayed a Branch and Bound algorithm called BBRS, which is an improved customization of the main BBS algorithm. Furthermore, they perceived a super set of the reverse skyline that grants us to tie the space looked in the midst of the reverse skyline calculation.To further decline the computational complexity nature of making sense of whether a point has a place with the turnaround skyline, they proposed enhanced algorithms called RSSA, which depends on exact pre-processed approximations of the skylines. These nearby estimations are used to perceive whether a direct has a place toward the turnaround skylines or not. They proposed perfect algorithms for two-dimensional information, while for higher dimensions a greedy algorithm is exhibited.

Reverse skyline query has various huge applications, which recovers data objects whose dynamic skyline focuses contain a given query point. In view of the characteristic vulnerability in various genuine information, the question preparing strategies that are expected for correct information can't be particularly used to handle indeterminate data. X. Lian and L. Chen [8] concentrated on the reverse skyline query processing over uncertain information, to be specific probabilistic reverse skyline, in both monochromatic and bichromatic cases that is MPRS and BPRS, individually. Specifically, they proposed successful pruning techniques to decrease the search space of MPRS and BPRS queries, and flawlessly incorporate them into effective query methodology. Additionally, the improved query processing techniques are likewise proposed by means of pre-computation methods.

Reverse top-k query:

Rank-aware query computing has gotten to be vital for some applications that return to the user the top-*k* objects focused on the user's preferences. Top-*k* queries have been studied from the perspective of the user, concentrating principally on effective query processing. A. Vlachou [9] studied top-*k* queries from the perspective of the producer. Given a potential product, which are the user preferences for which this product is in the top *k* query result set? They recognized a new query type, in particular reverse top-*k* query that is vital for producers to assess the potential market and effect of their products focused on the competition.

They defined reverse top-*k* queries and presented two forms of the query, to be specific monochromatic and bichromatic. They presented a geometric interpretation of the monochromatic reverse top-*k* query in the solution space that serves to comprehend the reverse top-*k* query theoretically. At that point, they concentrate on in more details the instance of bichromatic reverse top *k* query, which is additionally intriguing for practical applications. Such a query, if processed in a direct way, obliges assessing a top-k query for every user preference in the database, which is restrictively costly actually for moderate datasets. They proposed a proficient threshold-based algorithm that removes candidate user preferences, without processing the individual top-*k* queries. Besides, they presented an indexing structure focused on materialized reverse top-*k* perspectives so as to accelerate the calculation of reverse top-*k* queries. Materialized reverse top-*k* perspectives trade preprocessing cost for query accelerates in a controllable way.

III. PROPOSED WORK

In this paper, a technique is proposed to find out the k-Most Demanding Products (k-MDP) with maximum expected number of customers. First of all, to calculating number of existing products satisfying customer c, i.e. *N(EP,C)*, a bitmap index structure named BMI is used for maintaining the satisfying data among the quality constraints of customers in C and the quality attributes of existing products in EP. Whether an existing product ep satisfies the requirements of a client c can be proficiently checked, by utilizing the BMI index structure. Then two greedy algorithms are presented for finding k-Most Demanding Products. Single-Product-Based and Incremental-Based Greedy Algorithms are greedy algorithms which intend to find the approximate solution for k-most demanding products (k-MDP) discovering.Also the Apriori based (APR) algorithm is presented which uses upper and lower bounds to prune sets of candidate products whose supersets are impossible becoming the optimal solution. The UBP algorithm embraces the result found by the SPG algorithm as the baseline solution kCP_b. The estimation of E(kCP_S, c) is obliged to be processed just when the upper bound of E(kCP_S, C) is bigger than E($kCPS_b$, C). The baseline solution $kCPS_b$ will be updated if a superior arrangement is found. Therefore, the final result of kCPb is the optimal solution. An Exact Top k Algorithm is presented which minimizes time required for finding the optimal solution it has also has higher accuracy.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

A. ALGORITHM DETAILS

Exact Top-k Algorithm
for*all model t in S* **do**
$d_{maxt}[i] \leftarrow d_{i,j}$ *where* $p_j \in FP(c_i \mid P_E)$;
$e_t[i] \leftarrow |FP(c_i \mid PE)|$;
$m_t[i] \leftarrow |FP(c_i|PE) \cap P_M|$;
end
max_increase$\leftarrow$0;
for*all* $p_j \in P_N$ **do**
for*all* $c_i \in PC(p_j)$ *under each model t in S* **do**
$\Delta sales_t (p_j) \leftarrow 0$;
if$d(i, j) > d_{maxt}[i]$**then**
$\Delta sales_t (p_j) \leftarrow \Delta sales_t (p_j) + (1- \frac{m_t[i]}{e_t[i]})$;
else if$d(i, j) = d_{maxt}[i]$**then**
$\Delta sales_t(p_j) \leftarrow \Delta sales_t(p_j) + (\frac{m_t[i]+1}{e_t[i]+1} - \frac{m_t[i]}{e_t[i]})$;
end
$\Delta sales_t (p_j) \leftarrow \Delta sales_t(p_j) + \ldots + \Delta sales_{m'} (p_j)$
if$\Delta sales_t (p_j) >$ *max_increase* **then**
res $\leftarrow p_j$;
max_increase$< \Delta sales(p_j)$;
end
end
return*res*

B. Mathematical Model

Set Theory:
Let S, be a system such that,
S = {I, DP, A, O}

1. Input Review Data and Customer Requirements

I = {D, CR}
Where,
D = TripAdvisor Review Text dataset.
D = {d1,d2,…..dn}
Where,
D is set of input review data and
d1, d2 … dn are number of data in set D
CR = {cr1, cr2,….,crn}
cr1, cr2 … crn are number of Customer Requirements.

2. Data Preprocessing

DP = {BMI, SI, N}
Where,
DP = set of Data Preprocessing
BMI = BMI index Structure
SB = Satisfaction Bit String
N = N_vector(EP, C) the sum of the bit on the satisfaction bit strings of existing products

3. Algorithms for obtaining k-MDP

A = {SPG, IG, APR, UBP, TOP-K}

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

Where,

SPG = Single Product Based Greedy.
IG = Incremental-Based Greedy Algorithm.
APR = Apriori-Based (APR) Algorithm.
UBP = Upper Bound Pruning Algorithm.
TOP-K = Exact Top k.

4. Output

O = k-Most Demanding Products

IV. RESULTS AND DISCUSSION

Figure below show the time comparison graph of different algorithms with the normal execution and same algorithms after removing the K list item removed.

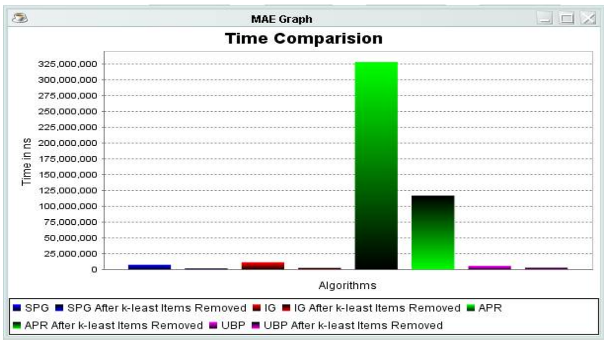

Fig 1. : Time Graph

Figure below show the memory comparison graph of different algorithms with the normal execution and same algorithms after removing the K list item removed.

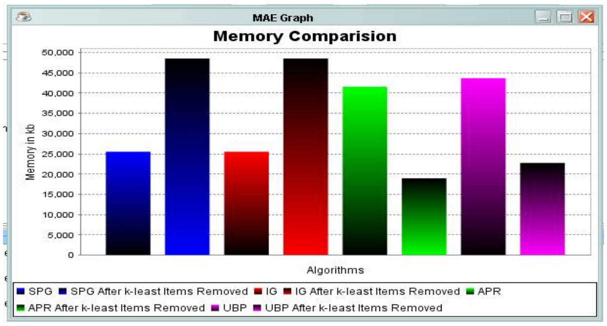

Fig. 2: Memory Graph

V. CONCLUSION

Here, the problem of k-MDP finding for deciding k most demanding products with the maximum expected number of the customers. Firstly, BMI index structure is used in which satisfaction bit string and calculate N_vector(EP, C) which used to store the sum of the satisfaction bit strings of existing products. This is given as input to algorithms for discovering the k-most demanding products. There are two greedy algorithms namely SPG and IG algorithms, which are presented for approximate answer. For finding the optimal solution, APR and UBP are presented which uses pruning techniques. In addition, one algorithm is presented which takes less time for computing and also provides optimal solution, called as Exact Top K algorithm. One real data set is used, that is, the user reviews for hotels on TripAdvisor in experiments.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

There is chance of a product selected by a customer get affected by the values of the attributes of the product. Nominal attributes are used to describe qualities of a product in some perspective. The research can be directed towards solving these issues.

REFERENCES

1. N.G. Mankiw, Principles of Economics, fifth ed. South-Western College Pub, 2008.
2. J. Kleinberg, 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. Z. Zhang, 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. 18-44, 2009.
4. E. Achtert, C. Bohm, P. Kroger, P. Kunath, A. Pryakhin, and M. Renz, “Efficient Reverse k-Nearest Neighbor Search in Arbitrary Metric Spaces,” Proc. 25th ACM SIGMOD Int’l Conf. Management of Data, pp. 515-526, 2006.
5. Y. Tao, D. Papadias, and X. Lian, “Reverse kNN Search in Arbitrary Dimensionality,” Proc. 30th Int’l Conf. Very Large Data Bases, pp. 744-755, 2004.
6. W. Wu, F. Yang, C.Y. Chan, and K.L. Tan, “FINCH: Evaluating Reverse k-Nearest-Neighbor Queries on Location Data,” Proc. 34th Int’l Conf. Very Large Data Bases, pp. 1056-1067, 2008.
7. E. Dellis and B. Seeger, “Efficient Computation of Reverse Skyline Queries,” Proc. 33rd Int’l Conf. Very Large Data Bases, pp. 291-302, 2007.
8. X. Lian and L. Chen, “Monochromatic and Bichromatic Reverse Skyline Search over Uncertain Databases,” Proc. 27th ACM SIGMOD Int’l Conf. Management of Data, pp. 213-226, 2008.
9. A. Vlachou, C. Doulkeridis, Y. Kotidis, and K. Norvag, “Reverse Top-k Queries,” Proc. 26th Int’l Conf. Data Eng., pp. 365-376, 2010.
10. Chen-Yi Lin, Jia-Ling Koh, and Arbee L.P. Chen, “Determining k-Most Demanding Products with Maximum Expected Number of Total Customers”, IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 8, August 2013.