K-Most Demanding Products Discovery with Maximum Expected Customers

Sofiya S. Mujawar
Department of Computer Engineering,
Dr. D. Y. Patil College of Engineering,
Ambi, Talegaon.

Dhanashree Kulkarni
Department of Computer Engineering,
Dr. D. Y. Patil College of Engineering,
Ambi, Talegaon.

ABSTRACT
This paper formulates a problem for production plan as k-most demanding products (k-MDP). Given a set of customers demanding a certain type of products with multiple features, a set of current products of the category, a set of candidate products that company is capable to offer, and a positive integer k, it helps the company to select k products from the candidate products such that the predicted number of the total customers for the k products is maximized. One greedy algorithm is implement to search inexact solution for the issue presented in this paper is NP-hard when the number of standards explains or features is 3 or more than 3. To find imprecise solution for this issue, Apriori-Based (APR) Algorithm and Upper Bound Pruning (UBP) Algorithm are proposed. Upper bound of expected figures of total customers is also implemented to find optimal solution of the problem. In addition to that, for computing least demanding products, an algorithm is proposed to search the k-least demanding products. This can also be beneficial to production plans.

Keywords
K-MDP, Decision support, Production plan.

1. INTRODUCTION
For that manufacturer must have products which fulfill the desires of customers. Because of the vicious rivalry in the market, everyone is engage with getting the greater focus of people. Large scale study is going in this field. In such cases, customer needs are very essential. The figure of production strategies can be modeled as a function that impacts the communication of the company with different procurer, for example, customers and competitors. The issue focused in this paper is to identify the production strategies with the maximum utility for a company, where the utility of production strategies is evaluated by predicted number of the customers for the preferred products in the plan.

Let EP and CP specification the set of current products and the set of candidate products, separately. Moreover, kCP means the set of k products selected from CP, cp indicates a candidate product from kCP, and c signifies a customer whose demands are satisfied by cp. The possibility for c picking cp is converse proportionate to the total number of products, including EP and kCP, which fulfill c. Thusly; the anticipate number of the customers for cp is affected not just by the number of customers fulfilled by cp more the total number of different products that satisfy the same set of customers. Identified that it is possible that the commodities in kCP will contend with one another if they satisfies the same set of customers. Consequently, no simple methodology can be fastening to discover the set of k candidate commodities with the biggest anticipated number of the customers. Instructions to give a proficient and effective algorithm for fathom the k-MDP finding problems are the goals of this paper. The major

works of this paper are: the issue of the k-MDP finding to be an escalation problem of a target function is computed; the k-MDP searching problems is NP hard when the properties for a product is 3 or more. Greedy algorithms are proposed to discover inexact answers for the k-MDP searching problem; an attempt to discover the optimal solution of the problem by assessing the upper and lower leap of the anticipated number of the customers for a set of k candidate products for diminishing the search space of the optimal solution. Two algorithms are then proposed to discover the optimal solution of the issue by utilizing the thin methods. One algorithm is proposed for finding the k-least products which is also essential for production plan.

Let us assume the condition of the rentable house market at a city, where the distance to a hospital and to a market are primary prerequisites of the customers demanding a rentable property. To construct a promoting decision, a rental company has collected the prerequisites of the distance to a market and to a hospital from the customers. Now assume, the rental company claims a group of attributes. The manager of the rental company required to opt to k attributes to contend with the current rentable houses for rental. For getting most profit, a technique is to get more expected number of the customer for the k chosen attributes. It is postulate that every customer will pick one of the rentable houses satisfying his/her requisites. At the point when more than one rentable house satisfies the necessities of a customer, the client will pick one of the attributes as per his/her implicit preference. For the cause of simplicity, it is obtaining that a customer will pick any capable rentable house with equal probability [2].

2. LITERATURE SURVEY
Many research has dealt the strong customers finding, for example, reverse k-nearest neighbor query [3], [4] the inverse skyline query [5], and the inverse top-k query [6]. The plan of these tasks is corresponding. Given a set of customer preferences and a particular product, the queries assumed in [3] furnish customers whose favorite products consist of the specific product as shown by their customer choices. The display of the product is, in this way, compel to the strong customers.

First of all, A. Vlachou et al [6] proposed a paper on converse top-k queries. The monochromatic and bichromatic these two categories of reverse top-k queries are represented. At that point, an algorithm for evaluating monochromatic reverse top k queries is initiated, based on the geometrical attributes of the result set. After that, they initiated an effective threshold-based algorithm (RTA) for computing bichromatic inverse top k queries, which eagerly discards candidate user choices, without required to evaluate the relevant top-k query. Besides, they initiated an indexing patterns depends on space partitioning, which emerges reverse top-k perspectives, to

advance reverse top-k query operation importantly further. They direct exhaustive test evaluations that unveil the expertise of their algorithms. RTA reliably improves 1 to orders of magnitude the naive technique. There are number of fascinating measure for future work. It is wide to study in more detail the monochromatic reverse top k query, specifically for higher admeasurements, since the geometrical attributes of the output set are important for operating the bichromatic reverse top-k query expertise.

A Reverse k-Nearest-Neighbor query searches the goals that are impacted by the querying object. It can be implemented in Location-Based Services to answer fascinating sites relevant queries. W. Wu et al [7] have founded their answers for evaluating RkNN queries on site data. They characterize RkNN query's fetch region and proposed an algorithm called FINCH to compute it concentrated on the queries and a set of data objects. FINCH is then consumed as a part of their RkNN solutions for filter and encloses the search space for output candidates. They moreover represented a method for implementing (monochromatic) RkNN algorithms to evaluate bichromatic RkNN queries. Investigating outputs represent that the fetch region computed by FINCH has a solid thin force and it speedup the filter operation. These elements make RkNN solutions a great deal more productive than the current RkNN algorithms.

M. Miah et al [8] proposed the problem of selecting the best properties of new tuple, such that this tuple will be ranked highly, given a dataset, a query log, or both, i.e., the tuple "emerges in the crowd". They initiated difference of the issue for Boolean, classifying, text and numeric data, and represented that notwithstanding the fact that the issue is NP-complete; optimal algorithms are obtainable for small inputs. Besides, they exhibited greedy algorithms, which are presented to fabricate good inexact ratios. While the problems assumed in this paper are novel and vital to the range of ad-hoc data investigation and access, they investigate that their specific issue definition has incarceration. After that, a query log is just an approximated substitute of genuine user choices, and moreover, in a few applications neither the database, nor the query log may be attainable for analysis; in this way they required to make presumptions about the manner of the rivalry and about the user choices. At last, in this issue they were concentrated on preferring what subset of properties to hold off a product.

The existing algorithms for RNN query are apposite in cramped condition. Y. Tao et al [9] proposed the first normal approach for accessing of an arbitrary number of reverse nearest neighbors in several dimensions. Despite its applicability and compliancy, their solution is superior to the past techniques moreover concerning expertise and versatility. Furthermore, presently there exists no any cost model for evaluating the operation time of RNN techniques. A fascinating heading for future work is to manage the presented approach to several diversity of RNN measures. The enhancement of such a model will not motivate query escalation, yet might likewise uncover new problems characteristics that could prompt importantly quicker algorithms.

S. Borzsonyi et al [10] presented how a database system can be outstretching so as to compute the skyline of a set of concentrations. The Skyline operation is helpful for several database applications, combining decision support and visualization. Their investigate outputs represented that a database system ought to build a block-nested-loops algorithm for good cases and a divide-and conquer algorithm for major

cases. They presented the SKYLINE OF condition as a basic extension to SQL's SELECT statement, unveil and exploratory evaluate optional algorithms to compute the Skyline, observe how inventory can be utilized to help the Skyline performance, and shown how the Skyline operation cooperates with other query operators that is join and Top N. All the more specifically, they presented to construct a block-nested loops algorithm with a window that is collected as a self-organizing list and a divide-and-conquer algorithm that does m-way separating and "Early Skyline" operations.

X. Lin et al [11] explore the problem of calculating the top-k demonstrative skyline points. This is among the first practices to generate expertise and adaptable algorithms to manage the problem. After proposing the novel skyline operator: top-k demonstrative skyline points, they exhibit an effective dynamic programming based algorithm for a 2D-space in which a specific solution can be accomplished. This measure is NP hard for space with dimensionality $d \geq 3$ and the greedy heuristic for set cover problem can be promptly linked to give the rough inexact proportion 1 - 1/e. they then developed an effective, versatile randomized algorithm with a hypothetical precision assure.

E. Dellis et al [12] proposed the idea of Reverse Skyline Queries (RSQ). With a particular end objective to operate the reverse skyline of an capricious query point, they startlingly proposed a Branch and Bound algorithm (called BBRS), which is an improved customization of the BBS algorithm. Provided a set of information points P and a query point q, a RSQ provides the data goals that have the query object in the set of their random skyline. It is the flattering problems to that of discovering the dynamic skyline of a query object. Such sort of dynamic skylines differentiate to the skyline of a change information space where point q turns into the origin and all points are proposed by their distance to q. Besides, they separate a super set of the reverse skyline that allows us to bound the space searched amid the reverse skyline operation. These inexactness are utilized to separate whether a point fits in with the reverse skyline or not. To further diminishes the computational cost of figuring out whether a point fits in with the reverse skyline, they presented an upgraded algorithm (called RSSA), that is depends on precise pre-computed inexactness of the skylines. For two-dimensional information, they initiated an ideal algorithm, while for higher dimensions a greedy algorithm is proposed.

3. PROPOSED SYSTEM

A methodology is presented to search k most preferred products. In this methodology, Bitmap Index Structure that is BMI index structure is computed to compute number of present product satisfying customer c. Bitmap Index structure is also implemented to keep fulfilling information of properties of product. Two greedy algorithms to search out the inexact solution are implementing in this methodology. These two algorithms are namely Single- Product-Based and Incremental-Based Greedy Algorithm. To search out the optimal solution, Apriori based (APR) and Upper Bound Pruning (UBP) algorithm is presented. These two algorithms implement upper bound and lower bound to thin the products which cannot become optimal solution. To search out products which are not essential to producer anymore and also will be useful to production strategy can be implemented, additionally this methodology compute the k-least demanding products. For this algorithm is proposed which is given below.
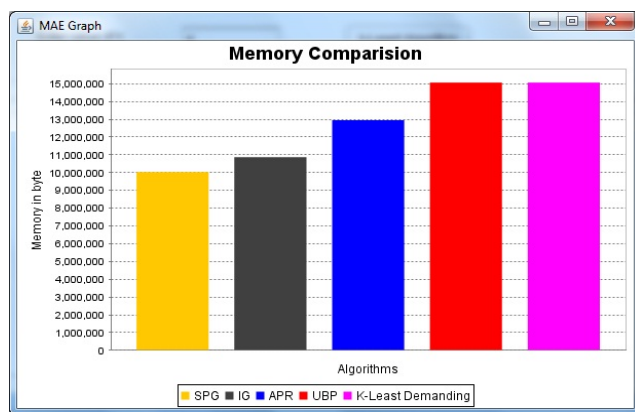
3.1 Algorithm

Input: N_vector(EP, C), the set C of customer requirements, the set CP of candidate products, and the value of k.

Output: a set of k candidate products.

1: For each candidate product cp in CP

2: {Compute the satisfaction bit string of cp; S = {cp};

3: Compute E(S, C); }

4: SL = <cp1', cp2',...,cp|CP|' >; //according to the increasing order of the values of

E(S, C);

5: kCPb = {cp1', cp2', ..., cpk'}; base = E(kCPb, C);

6: kCP ={cp1', cp2', ..., cpk-1', cpk+1'};

7: prune = {cp|CP|-k+1', cp|CP|-k+2', ..., cp|CP|'};

8: While (true)

9: { compute LB_E2(kCP, C); If LB_E2(kCP, C) < base

10: {compute E(kCP, C); If E(kCP, C) < base

11: {base = E(kCP, C); kCPb = kCP; }}

12: Else prune = kCP;

13: kCP = NextCandidateGen(SL, prune, k); If kCP == Φ

Break;}

14: Return kCPb;

15: Function NextCandidateGen(SL, prune, k) While (true)

16: { kCP = the next set of k candidate products according to <t;

17: If prune < r kCP Continue; Else Break;}

18: Return kCP;

 = N vector(EP, C)

4. RESULTS AND DISCUSSION

In the following first graph the Memory Comparison is depicted. On the X-axis the different algorithms are given which are SPG, IG, APR, UBP, K-Least Demanding these all are shown. Which are gradually increases in its memory bytes. The memory comparison of the UBP and K-Least demanding both are at the top among all the algorithms which are implemented. On the Y-axis the memory in bytes is computed which is started from 0 to 15,000,000.

In the following second graph the Time Comparison is depicted. On X-axis the similar algorithms are used which are used in first graph. Among all the graphs APR takes more time to execute where SPG and IG both take lesser time to execute or to perform the operations. On the Y-axis the time

in milliseconds is given to measure the execution time of the algorithms. The time is measured from 0 to 80ms.
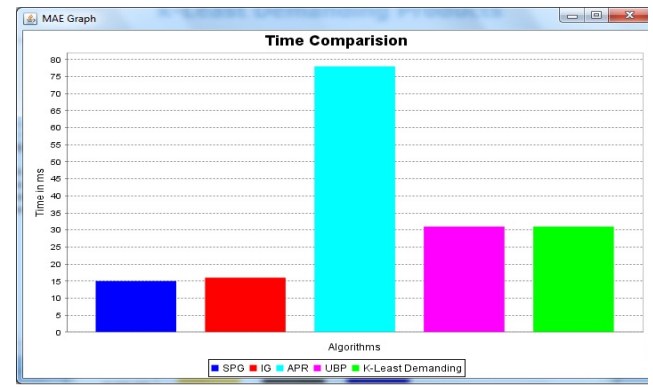
5. CONCLUSION

The k-MDP finding issues for searching k most required products with the most maximum anticipated number of the implements is formulated in this paper. The issue is NP hard when number of standard describers is 3 or more than 3. First we computed Bitmap Index pattern, Fulfillment Bit String and N vector (EP, C), which are given as input to algorithms. Likewise, two algorithms are proposed, which are the SPG and the IG algorithm, for finding the inexact solution. In this paper k-most preferred product system is used. It consist of four stages in first stage the data is gathered from the registration of the customers, in second stage preprocessing is done on this data to created bitmap index, In next stage 4 different algorithms are used to find the top preferred products. At the end the algorithms are studies to search the better algorithm. Here generated an online electronic shopping site for synthetic dataset. In this paper the propose system have some limitation such as it is used for only construction field. In future it can be extend for mobile industries or automobile industries.

6. REFERENCES

[1] 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.

[2] 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-EEEhowto:kopka 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 Intl Conf. Management of Data, pp. 515-526, 2006.

[3] F. Korn and S. Muthukrishnan, "Influence Sets Based on Reverse Nearest Neighbor Queries," Proc. 19th ACM SIGMOD Intl Conf. Management of Data, pp. 201-212, 2000.

[4] X. Lian and L. Chen, "Monochromatic and Bichromatic Reverse Skyline Search over Uncertain Databases," Proc. 27th ACM SIGMOD Intl Conf. Management of Data, pp. 213-226, 2008.

[5] A. Vlachou, C. Doulkeridis, Y. Kotidis, and K. Norvag, "Reverse Top-k Queries," Proc. 26th Intl Conf. Data Eng., pp. 365-376, 2010.

[6] W. Wu, F. Yang, C.Y. Chan, and K.L. Tan, "FINCH: Evaluating Reverse k-Nearest-Neighbor Queries on Location Data," Proc. 34th Intl Conf. Very Large Data Bases, pp. 1056-1067, 2008.

[7] M. Miah, G. Das, V. Hristidis, and H. Mannila, "Standing Out in a Crowd: Selecting Attributes for Maximum Visibility," Proc. 24th Intl Conf. Data Eng., pp. 356-365, 2008.

[8] Y. Tao, D. Papadias, and X. Lian, "Reverse kNN Search in Arbitrary Dimensionality," Proc. 30th Intl Conf. Very Large Data Bases, pp. 744-755, 2004.

[9] S. Borzsonyi, D. Kossmann, and K. Stocker, "The Skyline Operator," Proc. 17th Intl Conf. Data Eng., pp. 421-430, 2001.

[10] X. Lin, Y. Yuan, Q. Zhang, and Y. Zhang, "Selecting Stars: The k Most Representative Skyline Operator," Proc. 23rd Int'l Conf. Data Eng., pp.86-95, 2007.

[11] E. Dellis and B. Seeger, "Efficient Computation of Reverse Skyline Queries," Proc. 33rd Int'l Conf. Very Large Data Bases, pp. 291-302, 2007.

[12] J. Kleinberg, C. Papadimitriou, and P. Raghavan, "A Microeconomic View of Data Mining," Data Mining 311-322, 1998.