Reviving Dominated Points in Skyline Query

Taufik Djatna[†] and Yasuhiko Morimoto[††],

[† &††]Hiroshima University, Japan, [†]Bogor Agriculture University, Indonesia

Summary
We extend the capability of a skyline query to solve the market positioning evaluation of the dominated points of products. This capability can assist both manufacturer and customer to plan product features according to their approximate distances to the preference points. For this purpose we develop a distance measurement on a convex skyline approach. First, we present data sets contain record of multidimensional product, where every dimension represents one attribute of product feature. Then we evaluate the skyline query of a data set and divide the data set into a collection of preferable objects in skyline and another are the dominated points. Here we assume that each dominated point is potentially entering the preferable region by moving their attribute values into customer preference's points. We provide the query to find potential products to enter the skyline with a lower additional distance (as cost). This approach compute minimum additional cost to revive the dominated points based on a user's elicitation of a maximum threshold. Results of our comprehensive experiments show the effectiveness of this approach both in real world and synthetic data sets.

Key words:
skyline query, dominated points, customer preference.

1. Introduction

Skyline query has emerged as an interesting novel technology in database community particularly for *Customer and Supplier Relationship Management* (C/SRM). Recent researches show skyline application for user preferences [1] [2]. The concept of dominance of skyline has also recently attracted researcher's interest in context to answering preferences queries. Regarding of this, we find that skyline queries are potential tools for C/SRM to recognize important role of future preference of product features, like durability, high capacity, lighter weight, smarter response and cheaper price, such as common in electronic products.

We investigate that in conventional skyline concept, this capability is a hidden capacity, in which skyline itself is a *decision boundary* that resulted from both customer and manufacturer or supplier to their own preferences. This boundary separate clearly, which points of product are really in customer and manufacturer preference and, in other side, are the dominated points of product or services. Normally dominated products exert to enter this boundary as the consequent of market mechanism of supply and demand of the product features.

Potential product, in microeconomics, might be a new preferable one in the market if the manufacturer can recognize how to fulfill the customer preference in the future. Additional resource (*cost and patches*) should be allocated to increase preference of the points of object (products). The challenge is how to allocate the scarce resource to different dimensions using the skyline query results. This motivation is the basic customer relationship management motive to win competition in the market. Customer also can take benefit if additional cost to any product which partially satisfies her preference for purchasing is already available. When a manufacturer wants to increase the sale by patching their product dimension into preferable ones, such as higher capacity or lower price, additional costs will incur. This change might move the product into the preference region of a skyline. This additional cost can be either direct or indirect, according to which dimension value is considered.

One important paradigm in computation area dealing this issue is the application of skyline query, which enable us to evaluate characteristic of multi-dimensional value in hyper plane. For this purpose, we construct an enabling approach for companies to evaluate their point of product position in certain threshold and adjust their marketing strategy to dominate other product for eventually win the customer preferences. Each manufacturer would strive to maximize their market share with higher number of customers to acquire their products. To gain this, each manufacturer would allocate resources to find a better positioning in the market, which means manufacturers exert to dominate as many customers as possible within the constraint that called as the *multi dimensional attribute value in hyper plane*.

We propose a method to elaborate the requirement of dominated points a multi dimensional hyper plane to move into preferable region in the skyline. In order to find our goal to utilize skyline result for improvement market positioning, we study following questions (1) How to evaluate the most prospective product in the dominated region for entering a new positioning in skyline, (2) What is the basic formulation to minimize the cost of entering preferable region, (3) What is the boundary for an efficient evaluation among dominated point of the skyline query. These questions are important to support the revival of dominated point of products in a data set.

In this work, we arbitrarily compute additional costs incur

Manuscript received January 5, 2009
Manuscript revised January 20, 2009

to move dominated points bounds to a maximum threshold value in a skyline query. Our approach iteratively examines the initial skyline points and computes the prospective product point in each pair of dimensional plane. We prioritize the lowest cost dimension base on that threshold. Our contributions are as follows:

- Providing an effective formulation for reviving position of prospective points of product to enter the skyline
- Elaborate the user elicitation preference for prospective product using a maximum threshold of additional cost.

Our results show that the lowest prospective products are easily recognized using the elicitation of customer in skyline query. This finding is previously missing in skyline and k-dominance query works. In the following section 2 discusses the preliminaries for our work, then Section 3 provides the proposed method about revival query in skyline to improve to dominate points in skyline. Section 4 provides related works that support this work. Section 5 gives the experimental setup and results, then follow discussion of the improvement and performance finding. We sum up our work in section 6 with conclusions.

2. PRELIMINARIES

A. Skyline Query Definition

Skyline is a set of objects which are not dominated by another object in the data set, which divides object of data set $O=\{o_1,o_2,..o_n\}$ into Skyline set $S=\{s_1,s_2,..,s_m\}$, and dominated points $P=\{p_1,p_2,..,p_l\}$, with $m+l=n$. Skyline is a boundary line that separate preference search region (PSR) and non-preference region (NPR) as we illustrate in Fig.1. *Skyline query* retrieves a set of data points that are not dominated by any other points.
Assume a database having n records. An elicitation query of a user specifies a subspace having k attributes among attributes in the database. In the rest of this paper, we consider the each query consists of single sized, projected database D having k-attributes and n records and we assume lower value in each attribute is preferable without losing generality of the problem. For the reciprocal value of preference, we sign it with a negation.
*Definition 1 (**Dominance**)*: Given o, $o \in O$, o dominates o' iff $\forall\, d_i \in D = \{d_1, d_2, \ldots, d_k\}$, $o.d_i \le o.d_i, \exists\, d_j\, D,\ o.d_j < o'.d_j$.

Formally, using example in Figure 1 a skyline points set S = $\{s_1, s_2, s_3\}$ dominates another object set p= $\{p_1, p_2, \ldots, p_7\}$ in 2-dimensional space if s_1 is no worse that p_1 with respect to every attribute $\forall i : p^i$. That is, in this scenario, where smaller attribute values are desirable, e.g. for attributes such as x and y, s_1 dominates $p1$, if $s^i_1 \le p^i_2$; $\forall i = 1,..,d$ and there exists j, $1 \le j \le d$. such that $p^i_1 < p^j_2$. All of which can be measured by an integer number.
In the microeconomic, according to Figure 1, we can view the dominant products, such as product s_1 dominates p_6 and p_7 in the market. In other word, p_1 does not have a *market* as long as s_1 exists, if s_1 no worse than p_6 and p_7 in all d-attributes. As a consequence, skyline products are view as preferable choices which outperform all the other products in at least one attribute dimension. The rest or non skyline products are assumed as non-*preferable*.
We provide a method to revive each low cost non-skyline product *p which is* manufactured by improving p with minimization cost C (p, p'). We call this query as a *Revival query*. This query computes minimum additional cost for a higher marketability for a non-skyline product *p*. We define the *Revival Query* as follows:

Definition 2: (**Revival Query**) The revival query on a non preferable product p returns a preferable counterpart $p = argmin_p\ C(p, s)$, where $C(p, s) = \sum_{i=1}^{D} c_i(s_i - p_i)$, c_i is the additional cost for improving a unit of i-th attribute dimension d.
With the resource scarcity for manufacturing a point of product, it is very important to consider the profit constraint hyper plane h_i as the non preferable region to produce the product with highest quality just to attract customers.

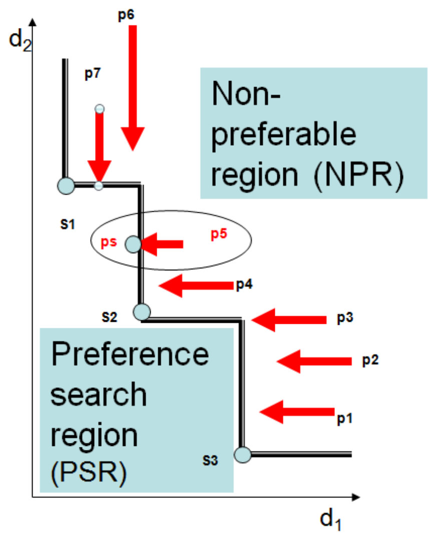

Figure 1. Dominated and preference search region in Skyline query in 2D

Definition 3: (**Maximum threshold**) the maximum threshold Γ is user defined values that limit upper bounds of additional costs occur when move dominated points of object into preferable region in skyline query.

It is possible for any product point in the dominated region to offer their new product features that will change the current market position into the preferable region. This process is in a hyper plane of all attribute dimensions that is considered applied on a product. This hyper plane h_i divides the multidimensional space into two regions, a preferable region which is profitable and a non -profitable one. We provide Figure 1 to illustrate these regions, and how any product point can change position into preferable region.

With the problem formally defined, it is important to know for a non skyline product the way to find the cost optimal to enter the profitable and preferable region. In short we need to formulate the new positioning for a non skyline product so that it transforms into the dominance region. For this purpose, we provide a visual to depict the problem and initial step to solve it in the Fig.1

B. Convex Skyline using Hand-Probing Approach

*Definition 4 (***Convex Skyline***):* A convex skyline is a set of point $s_i \in S$ that lies on the lower side of the convex hull of *S*. The convex hull of *S* is the minimum *k*-dimensional polyhedron that contains all of *S*. Note that a convex skyline must be a skyline since the point is not dominated by any other point in *O*. Figure 2 shows this construction.

Each record in *S* can be represented as a *k*-dimensional point x = $(x_1, x_2, \ldots, x_k)$ where coordinate value of the point are values of *k* attributes of the record. Therefore, the database *S* is a set of |*S*| points in the *k*-dimensional space. Points on the convex hull of the set of |*S*| points can be computed efficiently by, so called, *hand-probing* function. Hand-probing function finds the tangent point of the convex hull of the set of |*S*| points and a hyperplane whose normal vector is Θ.

Figure 2 is an example how to compute convex skyline. Assume there is a line whose normal vector is $\Theta_1 = (-1, 0)$ in the two-dimensional space. In order to find the tangent point with the line and the convex hull without pre-computing all points in *S*, we compute inner products of the normal vector and each atomic points (Θ_1, a).

Note that once we specify a normal vector, we can find the tangent point with convex hull, which is a skyline S, in O(n) by the hand-probing function.

Convex Hull Search First of all, we compute initial k tangent points that can be computed by hand-probing with initial k vectors $\Theta = (\theta_1, \theta_2, \ldots, \theta_k)$ where $\theta_i = -1$ if i = x, otherwise $\theta_i = 0$ for each x = 1,…,k. After computing the initial k tangent points, we have the initial facet (k-dimensional hyperplane). Next, we compute the normal vector of the facet. The new tangent point divides the initial facet into two facets. We recursively compute tangent point for each facet until there is no new point outside every facets.

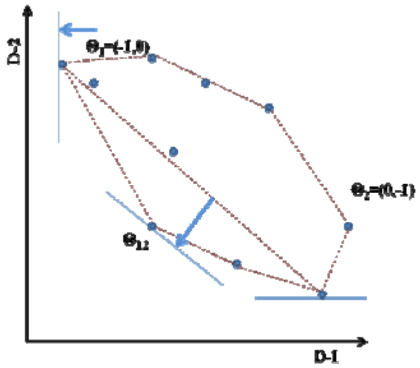

Figure 2. Convex skyline

We can apply this recursive operation for higher k-dimensional space as we performed in [3]. In the k-dimensional case, new tangent point, which is found by the hand-probing, divides the initial facet into k facets.

In higher dimensional case, the normal vector of facet can be computed as follows:

3D case: Assume we have three points P1 = $(p1_1, p1_2; p1_3)$, P2 = $(p2_1, p2_2, p2_3)$ and P3 = $(p3_1, p3_2, p3_3)$. If we imagine three points from three edges in the plane then we can take two of the edges and calculate cross product between them.

Suppose the edges vectors are V1 = $(v1_1, v1_2, v1_3) = (p2_1; p2_2, p2_3) - (p1_1, p1_2, p1_3)$ and V2 = $(v2_1, v2_2, v2_3) = (p3_1, p3_2, p3_3) - (p1_1, p1_2, p1_3)$. Then the normal is defined as the expansion of the following symbolic determinant.

$$V1 \otimes V2 = \begin{vmatrix} e_1 & e_2 & e_3 \\ v1_1 & v1_2 & v1_3 \\ v2_1 & v2_2 & v2_3 \end{vmatrix}$$

Where e_1, e_2 and e_3 are the elementary vectors (1,0,0), (0,1,0) and (0,0,1) respectively.

4D case: Again assume that we have four points P1 = $(p1_1, p1_2, p1_3, p1_4)$; P2 = $(p2_1, p2_2, p2_3, p2_4)$, P3 = $(p3_1, p3_2, p3_3, p3_4)$ and P4 = $(p4_1, p4_2, p4_3, p4_4)$. Using similar concept of 3D case we can compute three vectors V1 = $(v1_1, v1_2, v1_3, v1_4)$ = P2-P1, V2 = $(v2_1, v2_2, v2_3, v2_4)$ = P3-P1 and V3 = $(v3_1, v3_2, v3_3, v3_4)$ = P4-P1. Then the normal vector can be defined as the expansion of the following symbolic determinant:

$$V1 \otimes V2 \otimes V3 = \begin{vmatrix} e_1 & e_2 & e_3 & e_4 \\ v1_1 & v1_2 & v1_3 & v1_4 \\ v2_1 & v2_2 & v2_3 & v2_4 \\ v3_1 & v3_2 & v3_3 & v3_4 \end{vmatrix}$$

where the e_1, e_2, e_3 and e_4 are the elementary vectors

(1,0,0,0), (0,1,0,0), (0,0,1,0) and (0,0,0,1).

kD case: The same concept can be expand for k-dimensional case. Assume we have P1 = ($p1_1$, $p1_2$,…, $p1_k$), P2 = ($p2_1$,$p2_2$,…,$p2_k$), … ,P_k = (pk_1, pk_2,..,pk_k). Now, we can calculate (k -1) vectors like V1, V2,…,V($_{k-1}$). Then, the normal can be defined as the expansion of the following determinant:

$$V1 \otimes \ldots \otimes V(k-1) = \begin{vmatrix} e_1 & \ldots & e_k \\ v1_1 & \ldots & v1_k \\ \ldots & \ldots & \ldots \\ v(k-1)_1 & \ldots & v(k-1)_k \end{vmatrix}$$

3. REVIVAL QUERY IN SKYLINE

There are three possible approaches that enable solution for the problem of reviving dominated skyline points. The first one is to model the problem into a mixed integer programming (MIP). In the MIP approach, first, we have to set the target function, where we want to find the minimized additional cost for moving the dominated points into the preferable region.

These costs might occur for moving the attribute value of a product from less preferred into the preferable one. Then we have to examine which attributes will result the most effective changes for moving. In this case, we have to set up the status of each dimension into binary. The solution is easy for lower dimension as two attributes. But it will risk *curse dimensionality problem* in higher dimension, which make it impossible to solve using only *MIP* approach. It is also hard to maintain the dynamic changing of data set which results on very possible combination of customer elicitation. These combinations are consisted of different dimension patterns.

The second possible approach is to model a numerical analysis using curve fitting such as *b-spline* or higher polynomial computation to cover all dimensions. This approach is risky for higher computational cost. Implicit surface modeling is also another approach in numerical analysis with the same higher computational risk to model in a higher dimensional surface.

Our proposed method utilizes the idea of separation between two parties of skyline in one side and dominated points in other side. This idea is also considered in previous methods such as *b-spline* and *polynomial* as the input for searching the minimum distance. In this case, the main idea is to allocate skyline points set (*S*) in different space from the dominated points (P). We then treat each skyline point as pivot to measure the nearest distance of the dominated points in two ranges, upper and lower rectangular. The upper rectangular is defined as the maximum value of the dimension value. A pair of maximum value of two dimensions will make an upper rectangular. Lower rectangular is constructed from the intersection of two upper rectangular. While most of dominated points are scanned using the upper rectangular, lower rectangular measure the regions that are left from the intersection. Figure 3 gives detail of these two rectangular.

Modeling reviving dominated points in skyline query

We begin by studying the properties of convex skyline query *S* and the rest point's result which are lied in dominated region *P*.

We evaluate the characteristic of multi-dimensional value in hyper plane. The constraint of hyper plane should satisfy the following conditions:

- *Intersection testable* in which a constraint hyperplane h_i and a rectangle with two diagonal corners—lower and upper corners, at *(l[1],...,l[d])* and *(u[1],...,u[d])* are testified for their intersection in a constant time , where
- *Intersection extensible* in which the intersection between a constraint hyperplane h_i and (p_i[1], p_i[2],..., p_i[*d*]) are extensible with another (p'_i[1],..,p'_i [*k*+1],…,p'_i[*d*]) exactly on h_i in constant time.

We assume that each $p_i \in P$ is referred to a normal dynamic market, where maximum profit is the motive to enlarge the volume of customer preference. We propose a heuristic approach for practical solution in the problem of dominated points in skyline query. Using the convex skyline algorithm, we remove skyline set from initial data set and separate from dominated points virtually. First, we construct the upper and lower boundaries for each dimension attribute value to catch the range of numerical information within each dimension. The complexity of this computation give $\Theta(|S|\ \ |P|)$ where |S|+|P|=N is the cardinality of the data set.

Assume that the dominated region from a skyline query is composed of *L* objects $p_1, p_2,..,p_l$. Given a query object p_k which may be contained in the Revival query asks for the Skyline set S where $s_i \neq p_k$ which is closer to s_i than any other object in the database. A more general form of revival query is to ask for the nearest objects instead of just the closest one to p_k. Therefore, the revival query asks for the dominated objects that are less to the range of threshold Γ. The output of a *revival* query is a list of objects sorted in increasing distance order from the query object. As illustrated in Figure 3, we set each skyline point (s_i) that retrieved using convex skyline algorithm as the pivot point. This pivot point will be the base point for scanning each dominated point p_i in its range of {*max*(d_i), *max*(d_j)}, where d_i, $d_j \in D$. The Manhattan Distance (L_1) is deployed with boundary to the threshold value Γ that elicited by user. This measure computes additional cost (AC_i) for moving each object, formulated as follows:

$$AC_i = \sum_{i=i}^{D} | s_i - p_i |$$

where D is number sub-dimension, which is chosen by user's elicitation query. s_i is the pivot of skyline point and p_i is the dominated points.
The distance counts (AC_i) from each pivot is stored in a memory buffer B_i. Eventually the list of $Bi \leq \Gamma$ is the object points that fulfill the requirement to move to preferable region at maximum threshold Γ.

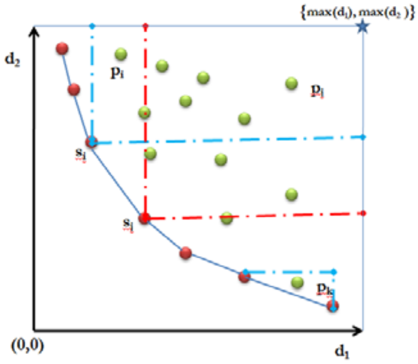

Figure 3. Scan distance on dominated points

4. RELATED WORK

To the best of our knowledge, this is the first study about skyline query analysis utilization for reviving the dominated products and to warrant improvement of their marketing preference. In many previous works, researchers in skyline queries had rather identified the promising objects in terms of the retrieval method characteristic and its efficiency, particularly in high dimensional space with low formulation over head. There is no discussion about the dominated points improvement over the rest of skyline points retrieved. Next, in this section we show related studies that apply skyline ideas, both in microeconomics and how the query result and properties are utilized in different purposes.

A. Skyline Perspective of Microeconomic

First introduced by Kleinberg et al. [4], [5] by formalizing the optimization problem of enterprise based on data. This optimization allows the enterprise to predict the utility of a customer w.r.t. a chosen decision. Various examples are provided in [5] to illustrate utility oriented mining. From these idea, researchers studied the opportunity of profit oriented association discovery [6], customer oriented catalog segmentation [4]. Kleinberg *et al.* [4] show how sensitivity analyses of microeconomic optimization can distinguish interesting from uninteresting changes of the decision in an enterprise. They investigate segmentation problems in more details as an approximate function for catalog segmentation problem. This outlined a sampling-based algorithm by enumerating ad measuring all possible partition of the customers in the sample. Eventually this approves probabilistic bounds for its result quality and runtime. Li *et al.* [7] studied the efficiency issue of extracting interest patterns from raw data, particularly on min/max attributes. Additionally [2] consider not only min/max attribute value in a skyline, but also created a spatial attribute consideration for microeconomic purpose.
They investigate two alternative approaches for efficient query processing, a symmetrical one based on off-shelf index structure and a symmetrical one based on index structure with special purpose extensions. In their model, linear optimization query was resolved, but limited to a lower dimension degree. In this paper we show that linear approach is not cover with higher dimensionality. This approach tends to risk curse of dimensionality problem.

B. Dominance Relationship Query

Skyline queries are a recent well-studied topic due to their roles in multi-criteria decision making and related applications [8]. The skyline operation was first introduced to the database community in [9]. A large number of methods have been proposed for computing conventional relational databases. These methods can be grouped into two general categories depending on whether they use indexes such as Nearest Neighbor [10], Index [11], branch and bound [12] or not, such as sort first skyline [13]. Moreover skyline has been studied in case of mobile devices [14], Distributed system [15], and structure network [16]. Subspace skyline query has been studied extensively in [17], [18] and [19]. Several recent researchers also focus on computation of skyline with data set has specific properties such as [20] that extended *Branch and Bound* skyline for the case where attribute take values from partially-ordered domains. Additionally [21] concentrate on skyline processing for domain with lower cardinality. Other interesting variants of the basic definition have been proposed. *Spatial skyline* [22] returns the set of data point that can be the nearest neighbors of any point given query set.
Our work approaches focus on the utilizing the number of dominated products point as the base for improvement market positioning, which is important for both purchasing planning and customer preference problem. The result supports measuring one market positioning for the manufacturing planning.

5. EXPERIMENTAL EVALUATION

Our experimental tools consisted of a dual-Xeon quad-processor 2.33 GHz machine with 8 GB of RAM running MySQL 5.1 RDBMS engine and MS-Windows server 2008. We develop a Java based prototype to test the method using both synthetic (correlated, anti-correlated and independent) and real world numeric data from CRM cases. The test bed includes two type data sets: (1) Real life data set of 5400 records of CoIL2000 dataset from UCI [23] and e-shopping from kakaku.com that sells electronics appliance in Japan online. (2) Synthetic data sets, consists of three forms, correlated data set, anti-correlated and independent. Each synthetic data set is generated in groups of 1K, 10K, 50K and 100K cardinalities.

Discussion

Here we discuss the performance evaluation of the proposed method. We proceed to revive the dominated points of object. Using our approach to compute minimum additional cost incurred in moving the dominated points, we compute the number of object moving into the preferable region in skyline and the time required. For these purposes, we find the result as follows.

According to Figure 4, we first examine the running time of the proposed program compare to numbers of most economical points to move. Then we compute number of points of object that are retrieved using our method. We then evaluate our method using synthetic data set. Based on these results, we analyze the performance.

1) Real-life data set: We evaluate the effectiveness of our approach on different threshold value using real data set from UCI [23] called CoIL2000. We set the threshold Γ, in five forms {10, 20, 30, 40, 50}. It is worth noticing that our method can handle to *k*-dimensional cases. For simplicity we provide a case in 3 and 5 dimensions.

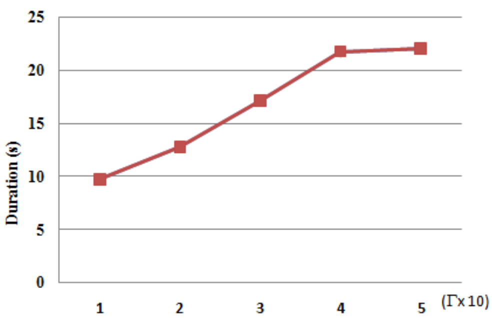

Figure 4. Duration on different threshold

Figure 4 shows the duration required to compute data item, divide the data set into skyline set and dominated set and calculate the distance as the minimum additional cost. Notice that the duration in second increases, as the threshold value rises. It reflects the revival query works on filtering the increasing number of objects selected from the skyline. These incremental numbers of object are not only derived from convex skyline algorithm we deploy, but also from the outer scanning of revival query upon objects between two skyline points. Further, next in Figure 5 provides the number of point of object retrieve using our approach in 3 dimensions. We also find the number of points of object increase linearly with the number of threshold.

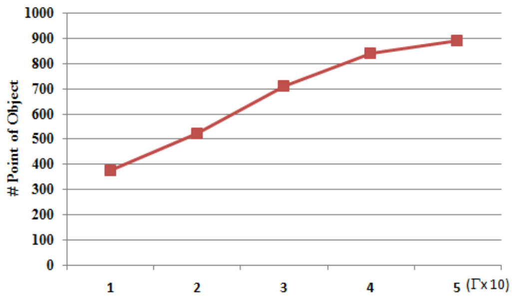

Figure 5. Number of object on different threshold

2) Correlated data distribution: In this case, only a few points are resulted which we use as pivot points to search the minimum additional cost. With majority of dominated data, we found the retrieval time of minimum points significantly increase as the cardinality rise. Figure 6 provides the duration required for correlated synthetic data set results. Note that we provide 5 dimensional cases in this synthetic data set evaluation.

3) Anti - Correlated data distribution: Here, more data points are not dominated, if we compare with previous correlated data set. Higher numbers of pivot points help us to identify more prospective products for moving into preferable region. Figure 7 provides the duration required for anti-correlated synthetic data set results.

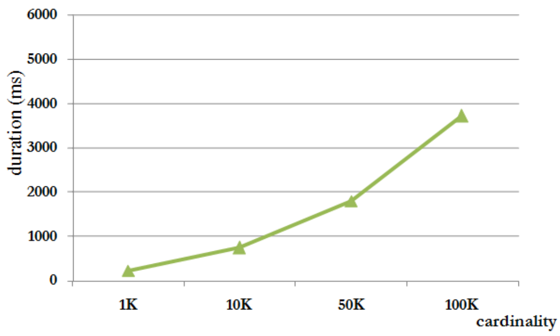

Figure 6. Synthetic Correlated Dataset

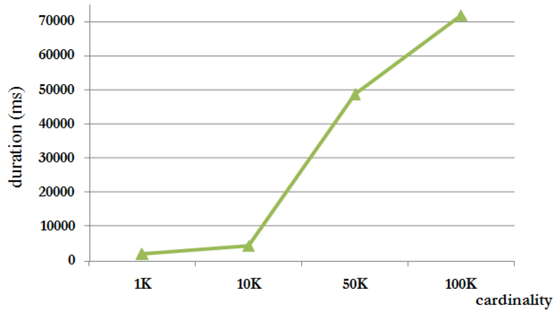

Figure 7. Synthetic Anti-Correlated Dataset

4) Independent data distribution: For this data distribution, m out of n data points is skyline points. As it has been shown in [24], the expected value of $m = \Theta((log(n))^{d-1}/(d-1)!$ that is highly dependent on the data dimensionality (d). Based on this, given n=100K, if d=3, m is 2K (2%). For high dimensionality, the processing time become close to Θ ($n.d.logd(n)$). For low dimensionality, few data points are skyline points and only small numbers of nodes are examined. Figure 8 provides the duration required for independent synthetic data set results. From Figure 7 and 8, we observed that the time spent on both synthetic anti-correlated and independent datasets tend to be polynomial with the expansion of the data set cardinalities. Note that in this synthetic data set evaluation, our duration measurement unit is in mili second (ms).

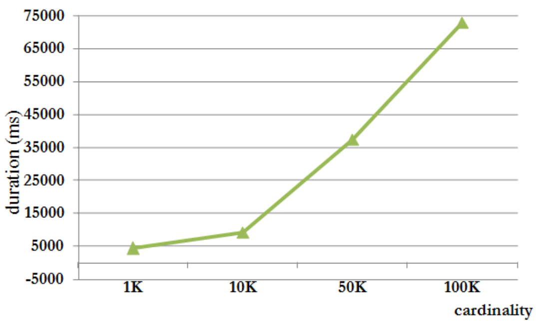

Figure 8. Synthetic Independent Dataset

From mentioned figures, we can conclude the effectiveness of this approach to compute the minimum additional cost to revive the dominated point in the skyline query. As we mentioned in section **IV**, currently, we cannot find any similar previous works to compare the performance of this result. Because of limited space we omit the e-shopping evaluation results.

6. CONCLUSIONS

We proposed the method to revive the dominated points of object that are widely applied in commercial data base application. This work was motivated by the opportunity to utilize the skyline query to support decision process in marketing for competing products by manufacturer, supplier and customer. Our method is based on convex hull skyline method, with target to minimize additional cost incurred for moving the dominated product to be preferable ones in the market. Using both real-life and synthetic data sets, we show that our method effectively evaluate the prospective points of object to move with a certain maximum threshold value. The applicability of this approach will enhance customer preference on a product with high dimensional features.

ACKNOWLEDGEMENT

This work was supported by KAKENHI (#19500123). Taufik Djatna was supported by the scholarship of MEXT Japan.

REFERENCES

[1] Y. Tao, X. Xiao, and J. Pei, “Subsky: Efficient computation of skylines in subspaces,” in *Proc. 22nd International Conference on Data Engineering ICDE ’06*, 03–07 April 2006, pp. 65–65.

[2] C. Li, A. Tung, W. Jin, and M. Ester, “On dominating your neighborhood profitably,” in *Proceedings VLDB 07 Vienna Austria*, 2007.

[3] T. Fukuda, Y. Morimoto, S. Morishita, and T. Tokuyama, “Constructing efficient decision trees by using optimized association rules,” in *Proc. of the 22nd VLDB Conference*, 1996, pp. 146–155.

[4] 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.

[5] J. Kleinberg, C. Papadimitriou, and P. Raghavan, “Segmentation problem,” in *Proceedings STOC*, 1998.

[6] R. Wong, J. Pei, A. Fu, and K. Wang, “Online skyline analysis with dynamic preferences on nominal attributes,” *IEEE Transactions on*, pp. 1–1, 2003.

[7] C. Li, B. C. Ooi, A. Tung, and W. Shan, “Dada: A data cube for dominant relationship analysis,” in *Proceedings SIGMOD 06*, ser. ACM SIGMOD, 2006.

[8] M. Preparata, F.P. Shamos, *Computational Geometry, an Introduction* Springer-Verlag, 1985.

[9] S. Borzsony, D. Kossmann, and K. Stocker, “The skyline operator,” in *Proc. 17th International Conference on Data Engineering*, 2–6 April 2001, pp. 421–430.

[10] D. Kossmann, F. Ramsak, and S. Rost, “Shooting stars in the sky: an online algorithm for skyline queries,” in *Proceedings VLDB*, 2002.

[11] K.-L. Tan, P.-K. Eng, and C.-Y. Chan, “Efficient progressive skyline computation,” in *Proceedings of VLDB*, 2001.

[12] D. Papadias, Y. Tao, and B. Seeger, “An optimal and progressive algorithm for skyline queries,” in *Proceedings of SIGMOD*, 2003.

[13] J. Chomicki, P. Godfrey, J. Gryz, and D. Liang, “Skyline with presorting,” in *Proc. 19th International Conference on Data Engineering*, 5–8 March 2003, pp. 717–719.

[14] Z. Huang, H. Lu, B. C. Ooi, and A. Tung, “Continuous

skyline queries for moving objects," vol. 18, no. 12, pp. 1645–1658, Dec. 2006.

[15] U. Balke, W.-T. Guntzer, "Multi-objective query processing for database system," in *Proceedings of the 30th VLDB Conference, Toronto Canada*, 2004, pp. 936–747.

[16] X. Wang, H. Liu, and J. Han, "Finding frequent items in data streams using hierarchical information," in *Proc. ISIC Systems, Man and Cybernetics IEEE International Conference on*, 7–10 Oct. 2007, pp. 431–436.

[17] Y. Yuan, X. Lin, Q. Liu, W. Wang, J. Yu, and Q. Zhang, "Efficient computation of skyline cube," in *Proceedings VLDB*, 2005.

[18] T. Xia, D. Zhang, and Y. Tao, "On skylining with flexible dominance relation," in *Proc. IEEE The 24th International Conference on Data Engineering ICDE 2008*, 7–12 April 2008, pp. 1397–1399.

[19] C. Zhang, J. Gao, Z. Liu, and F. Deng, "Nash equilibrium strategy of two-person bilinear-quadratic differential game: a recursive approach," in *Proc. Sixth World Congress on Intelligent Control and Automation WCICA 2006*, vol. 1, 2006, pp. 1075–1079.

[20] C.-Y. Chan, P.-K. Eng, K.-L. tan, A. Tung, and Z. Zhang, "Finding k-dominant skylines in high dimensional space," in *Proceedings of SIGMOD 503-514*, 2006.

[21] C.-Y. Chan, P.-K. Eng, and K.-L. Tan, "Efficient processing of skyline queries with partially-ordered domains," in *Proc. Of The 21st International Conference on Data Engineering ICDE 2005*, 5–8 April 2005, pp. 190–191.

[22] C. Sharifzadeh, M. Shahabi, "The spatial skyline queries," in *Proceedings of The 32th VLDB Conference*, 2006, pp. 751–762.

[23] A. Asuncion and D. Newman. "UCI machine learning repository". 2007 [online] Available: http://www.ics.uci.edu/~mlearn/MLRepository.html

[24] P. Godfrey, R. Shipley, and J. Gryz. Maximal vector computation in large data sets. In *VLDB 2005*.

Taufik Djatna received the B.Sc. and M.S. degrees in Agro-industrial Technology from Bogor Agriculture University (BAU), Indonesia in 1994 and 2004, respectively. Since 1995-present, he is an academic and researcher staff in BAU. Since 2005-2009 he studied data mining and query optimization for PhD degree in Hiroshima University, Japan. His research interest includes attribute selection in data mining and Pareto query optimization.

Yasuhiko Morimoto is an Associate Professor at Hiroshima University. He received B. E., M. E., and Ph. D. from Hiroshima University in 1989, 1991, and 2002, respectively. From 1991 to 2002, he had been with IBM Tokyo Research Laboratory where he worked for data mining project and multimedia database project. Since 2002, he has been with Hiroshima University. His current research interests include data mining, machine learning, geographic information system, and privacy preserving information retrieval.