Finding the most potential customer in bundling products by using reverse k-skyband query: current research progress and challenges

L A Gumilar¹, Annisa¹ and T Djatna²

¹ Department of Computer Science, Faculty of Mathematics and Natural Science, IPB University (Bogor Agricultural University)
² Department of Agroindustrial Technology, Faculty of Agricultural Technology and Engineering, IPB University (Bogor Agricultural University)

e-mail: lizza_gumilarlizza@apps.ipb.ac.id, annisa@apps.ipb.ac.id, taufikdjatna@apps.ipb.ac.id

Abstract. Online shopping nowadays is one of the most popular ways to make transactions. Customer transactions recorded on the online shopping application can be used by a retail company's Customer Relationship Management (CRM) manager to make sales strategies and increase customer loyalty. Sales strategies are needed to identify products that will be of interest in the market as well as potential customers who will be interested in products. The implementation of skyline queries produces the most interesting and preferred set of products. Whereas to search for the potential set of customers who are interested in certain products, a reverse skyline query can be used. To find a larger set of customers, a reverse k-skyband query can be implemented. Reverse k-skyband queries can also be applied to determine the level of interest of each customer in each product in the product bundling strategy. Product bundling is a sale of two or more products in one package commonly used by CRM manager to increase customer loyalty. This research proposes reverse k-skyband query to find the most potential customers who like bundling products. This study evaluates the performance of the reverse k-skyband query in terms of computation time.

1. Introduction
As online shopping develops through electronic commerce (e-commerce), several retail companies have begun developing special applications for the shopping services in their stores. With this application, retail companies can obtain periodic customer transaction records, which they can use to determine the preferences of each customer. Some products may be of interest to the same customer, but not all products have the same appeal to all customers [1]. In order to increase customer loyalty, the Customer Relationship Management (CRM) of a company should develop a sales strategy. CRM can be defined as an integration of coordinated strategies of sales, marketing and customer service [2]. CRM managers at retail companies are tasked with identifying products that will be of interest in the market as well as potential customers who will be interested in products from the retail companies.

One method from previous studies to select the best products is by applying skyline queries. An object can be called a skyline object if it is not dominated by other objects, i.e., when the value of the object is better on all criteria or the same on all criteria but better at least on one criterion [3]. For example, there are three snacks A, B, and C, with prices of $2, $3 and $4, the number of calories is 2 cal, 1 cal and 3 cal respectively. If customers prefer snacks that have low prices and smaller calories, it...
can be said that product C will be dominated by snacks A and B, because they have the lowest price and the smallest calories. Snack A has a lower price than B, while Product B has smaller calories. Thus, products A and B are said to be attractive products (skyline) for customers. If the customer prefers low prices, then product A will be his choice, if the customer prefers low calories, then product B will be chosen.

The concept of skyline query was then developed into reverse skyline query [4][5], which can be used to generate a set of customers who are potentially interested in a product. However, reverse skyline query only produces a few number of potential customers, so to expand it, reverse k-skyband method [5] was proposed to produce greater numbers of potential customers who are interested in one product. Reverse k-skyband queries can also be applied to determine the level of interest of many customers in several products that are bundled as one package (product bundling strategy). Product bundling is commonly used by CRM manager to increase customer loyalty. This research proposes reverse k-skyband query to find the most potential customers who like bundling products. This study evaluates the performance of the reverse k-skyband query in terms of computation time.

1.1 Research purposes
The purpose of this study is to:
- analyze customer shopping transaction data into customer preferences, which are then needed for the search for potential customers of product bundles;
- implement a reverse k-skyband query to find the most potential customers who like bundling products;
- evaluate the performance of the reverse k-skyband query in terms of computation time.

1.2 The scope of research
The scope of this research includes synthetic data, which are data of products in bundles and criteria for each product, and data of potential customers and criteria preferences of each product that the customers are interested in. All products in bundles are limited to products with the same attributes (characteristics) of a retail company.

1.3 Benefits of research
This research is expected to implement an efficient method in terms of computing time for CRM managers of retail companies to find potential customers of product bundles.

2. Literature review

2.1 Skyline
Skyline objects are a collection of objects that are not dominated by other objects based on certain criteria or what can be classified as objects of interest (point of interest). An object is called a skyline object when it is not dominated by other objects, i.e. when its value is better on all criteria or the same on all criteria but better at least on one criterion [3]. In Figure 1, \(x_1, x_2, x_3, x_4, x_5, x_6\) are assumed to be snack data. Snack with low prices and calories is preferred in the market. In Figure 1, \(x_1\) dominates \(x_4\) because \(1<3\) and \(4=4\), \(x_2\) dominates \(x_5\) because \(2<4\) and \(2<3\), and \(x_3\) dominates \(x_6\) because \(3<5\) and \(1=1\). The set of skyline objects are \(x_1, x_2,\) and \(x_3\), which are the most interesting snacks on the market.
2.2 Reverse skyline
There is a data set containing customer $X$ data that consists of all customer $x \in X$ and product target $A$ as a query point. Reverse skyline is a set of points $(x)$ that find query points $(A)$ interesting [4]. In Figure 2, $A$ is assumed as snack data, and reverse skyline query will look for a set of customers $(x)$ who think that $A$ is an interesting snack. Snack $A$ appeals to customers $x_3$, $x_{12}$, and $x_{13}$.

2.3 Reverse $k$-skyband
Conceptually, in the reverse $k$-skyband, $k$ represents the thickness of the reverse skyline layer which can extend the scope of the query point $A$ [6]. Therefore, the case $k = 0$ (reverse 0-skyband) is a conventional reverse skyline. Figure 2 shows that points $x_3$, $x_{12}$, and $x_{13}$ are reverse skyline of $A$, while Figure 3 shows that $x_3$, $x_9$, $x_{10}$, $x_4$, $x_5$, $x_{12}$, and $x_{13}$ are reverse 1-skyband of $A$. 

Figure 1. Example of skyline objects.

Figure 2. Example of reverse skyline objects.
3. Methodology

3.1 Data pre-processing and initialization
Data of customer preference is obtained from the analysis of shopping transactions of each customer in retail company applications (in this study it is synthetic). The analysis uses market basket analysis, which is searching for relationships of several products in purchase transactions [7]. This technique looks for the most likely (frequent) product to appear from a set of products. This customer preference data is combined with the data of products in bundles to be used as a data set in this study. Table 1 shown the example of data after pre-processing and initialization process was done. A, B, and C are products in one bundling packet. X1 until X13 are customer preferences.

| ID  | Calorie | Price |
|-----|---------|-------|
| A   | 50      | 80    |
| B   | 70      | 40    |
| C   | 30      | 45    |
| X1  | 10      | 10    |
| X2  | 20      | 30    |
| X3  | 60      | 20    |
| X4  | 10      | 60    |
| X5  | 40      | 50    |
| X6  | 30      | 90    |
| X7  | 60      | 60    |
| X8  | 55      | 45    |
| X9  | 60      | 50    |
| X10 | 45      | 70    |
| X11 | 80      | 90    |
| X12 | 90      | 10    |

3.2 Implementation of reverse k-skyband
In the example case of Figure 4, the query points A, B, C are new snack products that are combined in a bundle package, and the other points are the customer preferences for snacks (point Xi). The implementation of reverse k-skyband aims to find the most potential customers who are most interested in the three new snack products (A, B, C). Searching for potential customers from the product bundle can be done by applying the reverse k-skyband method to determine the level of interest of each customer in each product in the bundle package as shown in Figure 4.
Figure 4. Implementation of reverse 1-skyband to point A and point B.

The steps of reverse k-skyband implementation are shown in Figure 5. In this process, to see the entire level of interest of each customer (X1 to X13) towards products A, B, C, the k on the reverse k-skyband method is carried out until k = 8. The result of the reverse k-skyband implementation is a list of customers sorted by the potential level of product purchase and the most potential customers, which in the case of Figure 5 is X8.

Figure 5. Steps of reverse k-skyband implementation.

4. Results and discussion
This research evaluates reverse k-skyband method based on time complexity. The current progress of this study is conducted with several experiments that have data sets with different parameters such as in Table 2.
Table 2. Current progress experiment parameters.

| Parameter                                                  | Value                  |
|------------------------------------------------------------|------------------------|
| Number of Products                                        | 1                      |
| Number of Objects (customer preferences)                   | 13, 100, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10K |
| The number of characteristic attributes of each object     | 2D, 5D                 |

The results of current progress based on computational time shown in Figure 6:

Figure 6. Result of reverse skyline for product A.

Figure 6 shows that more dimension and more number of customer preferences increase the computational time. Experiment with 10,000 customer preferences and 5 dimensions of product take the highest computational time. The most influential parameter in this current progress based on the result above is dimension of products. In this experiment, we take product A as an example. After we implemented the reverse skyline, we got the most potential customers (X6, X10, and X11) that are attracted to product A as we can see in Table 3.

Table 3. Potential customers for product A.

|       | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|-------|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|
| Reverse Skyline (Reverse 0-Skyband)                      | -  | -  | -  | -  | -  | A  | -  | -  | A  | A   | -   | -   |     |

5. Conclusion

This current progress need more experiments to be done. Furthermore, the study should continue with more experiments that have datasets with different parameters such as higher number of attributes. The implementation of reverse k-skyband algorithm with different parameters aims to see the relationship of various parameters and computational time.
6. Reference

[1] Islam M S and Chengfei L 2016 *VLDB J.* 25 545–570
[2] Tsiptsis and Chorianopoulos 2009 *Data Mining Techniques in CRM Inside Customer Segmentation* (West Sussex: Wiley)
[3] Börzsönyi S, Kossmann D and Stocker K 2001 The skyline operator *Proc. IEEE Int. Conf. on Data Engineering* pp 421–430
[4] Dellis B and Seeger 2007 Efficient computation of reverse skyline queries *Proc. of the 33rd Int. Conf. on Very Large Databases VLDB Endowment* pp 291–302
[5] Yin B, Gu K, Wei X, Zhou S and Liu Y 2018 *Knowl. Based Syst.* 152 117–135
[6] Gao Y, Liu Q, Zheng B, Mou L, Chen G and Li Q 2015 *Inf. Sci.* 293 11–34
[7] Han J and Kamber M 2006 *Data Mining: Concept and Techniques* (San Francisco (US): Morgan Kaufmann Publishers)