Taxicab Geometry Based Analysis on Skyline for Business Intelligence

Partha Ghosh, Kingston School of Management & Science, West Bengal, India
Takaaki Goto, Ryutsu Keizai University, Chiba, Japan
Soumya Sen, University of Calcutta, West Bengal, India

ABSTRACT

This article describes how multi-criteria decision making problems are difficult to handle in normal SQL query processing. Skyline computation is generally used to solve these types of requirements by using dominance analysis and finding shortest distance with respect to a prime interesting point. However, in real life scenarios shortest distance may not be applicable in most of the cases due to different obstacles or barriers exist between the point of interests or places. In order to consider the presence of obstacles for geographically dispersed data, this research work uses Taxicab geometry for distance calculation, which is a simple Non-Euclidian geometry with minimum time complexity. Another limitation of previous skyline based works are that they only focus upon a single interesting point and can’t be apply for multiple interesting points. This research article focuses upon multiple visiting points for the travelers in an optimized way. In addition to this, the article also selects areas for setting up of new business properties considering the constraints.

KEYWORDS
Dominance Analysis, Multi-criteria Decision Making, Multiple Interesting Points, Obstacle, Skyline, Taxicab Distance

1. INTRODUCTION

Queries are executed on the database to fetch results based on business requirements. The predicate of the queries could be simple in nature or may get complex due to the requirements of multi-criteria decision makings. Sometimes this requirement further complex if the criterions are inversely proportional. In this context the traditional query optimization techniques do not work well as these are not supposed to optimize multi-criteria decision making. The skyline computation (Borzsonyi, Kossmann & Stocker, 2001; Tan, Eng & Ooi, 2001; Kossmann, Ramsak & Rost 2002) was introduced to solve these type of optimization problem. Dominance analysis (Borzsonyi, Kossmann & Stocker, 2001; Ghosh & Sen, 2015; Ghosh & Sen, 2016) is being performed to extract skyline points from a set of k-dimensional n points. It is defined as “A point that is not dominated by any other point in the system is called skyline point” (Borzsonyi, Kossmann & Stocker, 2001; Chomicki, Godfrey, Gryz & Liang, 2003; Ghosh & Sen, 2015; Bartolini, Ciaccia, & Patella, 2006). For a set of k-dimensional points, a point dominates another point if it is at least equally good in all the (k – 1) dimensions and better in at least one dimension (Borzsonyi, Kossmann & Stocker, 2001). After extracting the relevant
skyline points from the given problem domain, the common practice is to identify an optimal point (Ghosh, Goto & Sen, 2017; Tan, Eng & Ooi, 2001) from the problem domain and rank all the other skyline points according to the merit of that optimal point. One of the alternative solutions of this type of problem is to identify the most important dimension and convert all the other dimensions in to that in some format and finally rank all the other points according to that (Ghosh & Sen, 2016).

Skyline computation is widely used in “Travelling and Tourism Industry” to generate an optimal tour plan. The general criteria for the travelers are to select a hotel that is cheapest in its own class but closest to the Point of Interest (visiting point). Unfortunately, these two criteria are inversely proportional to each other, as the hotels near to the point of interest tend to be more costly. Hence traditional SQL (Rahman, N., 2016). query can’t optimize these types of requirements. Most of the existing methodologies identify an optimal point within the skyline region (Ghosh & Sen, 2016) and compute shortest distance from that optimal point using the formula provided by the Euclidean geometry. In a two dimensional space, Euclidean geometry measures distance between two points A and b using the formula:

\[ d(A,B) = \sqrt{(X_A - X_B)^2 + (Y_A - Y_B)^2}. \]

This formula works perfectly in the space where all types of movements are possible (without any physical constraints). However, it is not suitable for travelling and tourism Industry, as the two points are connected mainly through horizontal and vertical roads (in the form of straight line in certain direction). Moreover, due to the presence of different physical obstacles and barriers, all type of horizontal and vertical movements are also not possible. Another limitation of the Euclidean geometry is its high complexity when more than two dimensions are taken into computation. Therefore, an improved methodology is required that can adapt according to the situations.

One of the alternatives of Euclidean geometry is Taxicab geometry (Tang, Kerber, Huang, & Guibas, 2013; Ekici, Sevinc, & Cengiz, 2012; Ghosh, Goto & Sen, 2017). It is a simple non-Euclidean geometry that uses modular differences of the coordinates for finding distance between points. To measure distance between two points in a two dimensional space, it uses the formula:

\[ \text{dist}(A,B) = |X_A - X_B| + |Y_A - Y_B|. \]

It is a simple non-Euclidean geometry formula with much lesser complexity. This type of formula is suitable especially in case of measuring distance between two points within a city, where mostly horizontal and vertical movements are possible (as maximum street are).

Consider the following example, If we measure the distance between two points A and B in a two dimensional plane having coordinates A(20, 50) and B(30, 100) then the shortest distance will be \[ \sqrt{(20 - 30)^2 + (50 - 100)^2} = 50.99 \] unit.

In real life scenarios, due to the presence of some physical obstacles upon the shortest distance path like as shown in Figure 2, paths can’t be straight line like as shown in the Figure 1 rather the paths are generally horizontal and vertical and often looks like as shown in Figure 3 and Figure 4 as one of the instances.

Hence traditional Euclidean geometry formula for shortest distance calculation is not applicable in many of the real-life applications. In this context, Taxicab Geometry is more applicable. The formula that is being used to calculate the distance between two points in a two-dimensional plane using taxicab geometry is as follows.
Vulnerability Discovery Modeling for Open and Closed Source Software
www.igi-global.com/article/vulnerability-discovery-modeling-for-open-and-closed-source-software/176399?camid=4v1a

LAKE: Using Log Files Recorded during Program Execution
www.igi-global.com/article/lake/120515?camid=4v1a