Chapter 11
C2S:
A Spatial Skyline Algorithm for Changing Data

Marlene Goncalves
Universidad Simón Bolívar, Venezuela

Fabiola Di Bartolo
Universidad Simón Bolívar, Venezuela

ABSTRACT

Skyline queries may be used to filter interesting data from a broad range of data. A Skyline query selects those data that are the best according to multiple user-defined criteria. A special case of Skyline queries are the Spatial Skyline Queries (SSQ). SSQ allow users to express preferences on the closeness between a set of data points and a set of query points. We study the problem of answering SSQ in presence of changing data, i.e., data whose values regularly change over a period of time. In this chapter, it is proposed an algorithm to evaluate SSQ on changing data. The proposed algorithm is able to avoid recomputation of the whole Skyline with each update on the data. Also, the performance of the proposed algorithm against state-of-the-art algorithms was empirically studied. The experimental study shows that the proposed algorithm may become 3 times faster than state-of-the-art algorithms.

INTRODUCTION

Skyline queries are particularly relevant in the area of decision support or data visualization. They may be used to filter interesting data from a broad range of data. Intuitively, a Skyline query selects those data that are the best according to multiple user's criteria; user's criteria consist of minimizing or maximizing the attributes of an input dataset simultaneously.

A special case of Skyline queries is the Spatial Skyline Queries (SSQ); SSQ focus on geometric or spatial data and they allow expressing preferences on the closeness between a set of data points and a set of query points (Sharifzadeh & Shahabi, 2006).

DOI: 10.4018/978-1-4666-8767-7.ch011

The Skyline query evaluation is a costly process for a large dataset. The worst case complexity is $O\left(|P|^2\right)$ where P is the dataset (Godfrey, Shipley, & Gryz, 2005). Clearly, the Spatial Skyline query evaluation is still more expensive because distance functions must be computed. In this sense, performing an exhaustive search increases the search space to $O\left(|P|^2 |Q|\right)$ where Q is the set of query points (Sharifzadeh & Shahabi, 2006). Sharifzadeh and Shahabi (2006) exploit geometric properties in order to reduce the search space to $O\left(|S|^2 |C| + \sqrt{|P|}\right)$, where S is the Skyline set and C is the vertex set belonging to the convex hull of the query points in Q. The convex hull of Q is the unique smallest convex polytope which contains all the points in Q (Sharifzadeh & Shahabi, 2006).

In this chapter, we study the problem of answering Spatial Skyline Queries (SSQ) in presence of both spatial and non-spatial changing data. Changing data are data whose values regularly change over a period of time. Suppose a recommendation system that is able to suggest the best parking spaces in a parking lot. A Spatial Skyline query may be evaluated to identify the best parking spaces. In this example, the data are changing, e.g., the driver's vehicle may move continuously or the availability of parking spaces may change frequently. Therefore, the Spatial Skyline set changes depending on the vehicle location and/or the availability of spaces. Particularly, the availability of a parking space is a changing non-spatial datum and the location of the vehicle is a moving query point.

SSQ on changing data may have several applications. They may be used in logistics of services and parking lots of festivals and events such as Rock in Rio, Pinkpop Festival, Lollapalooza, Glastonbury Festival, Olympic Games, FIFA World Cup and Super Bowl to which attend thousands of citizens and tourists from different regions and countries (Lande & Lande, 2008). In these events the parking time could be critical if customers have to find a parking space by themselves without any help or suggestion.

In addition, SSQ on changing data may be applied in case of emergency or natural disaster to identify the highly vulnerable regions near the most endangered areas in order to plan and organize a rescue; the affected regions change their status among attended and urgent attention and have an impact level, so the most urgent and vulnerable regions are attended first. The SSQ on changing data may also be used in reservation systems able to recommend on the basis of criteria that meet the user's preferences. For example, a user may be interested in booking an event preferring a lower price seat which is close to the stage, to the exit and to the initial seat of any row; the seats availability may change because other customers canceled a reservation or made a new one. Another use could be in a mobile application to reserve a table for n people in a restaurant which is open, well rated and is on the way home, this means close to the user in motion and close to home. Lastly, the problem of answering SSQ on changing data is hard to solve because data is constantly changing while the Spatial Skyline set is been computed in polynomial time.

Additionally, the state-of-the-art algorithms may be adapted to evaluate SSQ on changing data. However, the Skyline set must be continuously refreshed according to data changes if these algorithms are executed. For example, the algorithms for SSQ introduced in (Sharifzadeh, Shahabi, & Kazemi, 2009) allow multiple query points, one of which is a moving query point. Using these algorithms, any change to non-spatial data will cause a recomputation of all the Skyline set.

Our contribution is to provide a solution to the problem of evaluating SSQ on changing data. We propose an algorithm to calculate the Spatial Skyline when data are changing. Our proposed algorithm is able to identify the Spatial Skyline based on changing characteristics of the data and prunes the search space using spatial properties, in order to avoid exhaustive searches. We have empirically studied the

Related Content

A Hybrid Approach to Content-Based Image Retrieval
Görkem Asilioglu, Emine Merve Kaya, Duygu Sarikaya, Shang Gao, Tansel Ozyer, Jamal Jidaand Reda Alhajj (2012). *Intelligent Multimedia Databases and Information Retrieval: Advancing Applications and Technologies (pp. 91-104).*
www.irma-international.org/chapter/hybrid-approach-content-based-image/59954

Fuzzy Opinion: Detection of Opinion Based on SentiWordNet Dictionary by Using Fuzzy Logic
Mohamed Amine Boudia, Reda Mohamed Hamouand Abdelmalek Amine (2016). *International Journal of Information Retrieval Research (pp. 1-18).*
www.irma-international.org/article/fuzzy-opinion/161658

Schema Independent XML Compressor
Baydaa Al-Hamadani, Zhongyu (Joan) Luand Raad F. Alwan (2011). *International Journal of Information Retrieval Research (pp. 18-38).*
www.irma-international.org/article/schema-independent-xml-compressor/58889

Efficient Pixel-Value Differencing Based Hybrid Steganographic Method Using Modulus Function
Aruna Malikand Sonal Gandhi (2020). *International Journal of Information Retrieval Research (pp. 51-62).*
www.irma-international.org/article/efficient-pixel-value-differencing-based-hybrid-steganographic-method-using-modulus-function/262177

Towards a Statistical Approach to the Analysis, the Indexing, and the Semantic Search of Medical Videoconferences
Ameni Yenguiand Mahmoud Neji (2018). *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications (pp. 1274-1299).*
www.irma-international.org/chapter/towards-a-statistical-approach-to-the-analysis-the-indexing-and-the-semantic-search-of-medical-videoconferences/198599