ISSN:0975-9646

Survey on Nearest Neighbor Search for Spatial Database

Ms.**Zarreen Anjum**
M. Tech student, dept of CSE
G.H Raisoni Institute Of Engg & Tech
Nagpur, India

Ms.**Priti Saktel**
Assistant Professor , dept of CSE
G.H Raisoni Institute Of Engg & Tech
Nagpur, India

Abstract **– Conventional spatial queries, nearest neighbor retrieval and range search consists only conditions on objects geometric property only. But today, many modern applications support new form of queries that aim to find objects that satisfies both spatial data and their associated text. For example instead of considering all the hotels, a nearest neighbor query would instead ask for the hotel that is closest to among those who provide services such as pool, internet at the same time. For this type of query a variant of inverted index is used that is effective for multidimensional points and comes with an R-tree which is built on every inverted list, and uses the algorithm of minimum bounding method that can answer the nearest neighbor queries with keywords in real time.**

Keywords: **Spatial database, nearest neighbor search, spatial index, keyword search.**

I. INTRODUCTION

A spatial database is a database that store multidimensional objects such as points, rectangles, and etc. some spatial databases allow representing simple geometric objects such as lines, points and polygons. Some spatial databases handle more complex structures such as 3D objects, topological coverage's, linear networks. Based on different selection criteria spatial database provides fast access to multidimensional objects. In spatial database real entities are modeled in geometric manner, for example location of hotels, hospital, restaurants are represented as points on maps, while larger area such as landscapes, lakes, parks are represented as a combination of rectangles. Spatial database system can used in geographic information system, in this range search can be utilized to find all restaurants in a certain area, while nearest neighbor retrieval can find the restaurant closer to a given address.

Queries in spatial database have become increasingly important in recent years with the increasing popularity of some services such as Google Earth and Yahoo Maps, as well as other geographic applications. Today, widely used of search engines has made it realistic to write spatial queries in a new way. Traditionally, queries focus on objects only geometric properties, for example whether a point is in rectangle or how two points are close from each other. Some new application allows users to browse objects based on both of their geometric coordinates and their associated texts. Such type of queries called as spatial keyword query. For example, if a search engine can be used to find nearest hotel that offer facilities such as pool and internet at the same time. From this query, we could first obtain the entire hotel whose services contain the set of keywords, and then find the nearest one from the retrieved restaurant. The major drawback of this approach is that, on the difficult input they do not provide real time answer. For example, from the query point the real neighbor lies quite far away, while all the closer neighbors are missing at least one of the query keywords. Spatial keyword queries have not been widely explored. In the past years, the group of people has showed interest in studying keyword search in relational databases. Recently the attention has preoccupied to multidimensional data [5][6]. The best method for nearest neighbor search with keywords is because of Felipe et al. [5]. They combine the spatial index R-tree [7] and signature file [8]. So they developed a structure called IR^2-tree. This tree has the ability of both R-tree and signature files. Like R-tree it stores the spatial proximity of object and like signature file it filters those objects that do not include all query keywords.

II. RELATED WORK

Cao et al. [1] proposed collective spatial keyword query, they presented the new problem of retrieving a group of spatial objects, and each associated with a set of keywords. They develop approximation algorithms with provable approximation bounds and exact algorithms to solve the two problems.

Lu et al. [2], combined the notion of keyword search with reverse nearest neighbor queries. They propose a hybrid index tree called IUR-tree (Intersection-Union R-Tree) to answer the Reverse Spatial Textual k Nearest Neighbor (RSTkNN) query that effectively combines location proximity with textual similarity. They design a branch-and-bound search algorithm which is based on the IUR-tree. To further increase the query processing, they proposed an improved variant of the IUR-tree called cluster IUR-tree and two corresponding optimization algorithm

Zhang and Chee[3] introduced hybrid indexing structure bR^*-tree, that combines the R^*-tree and bitmap indexing to process the m-closest keyword query that returns the spatially closest objects matching m keywords. They utilized a priori based search strategy that successfully reduce the search space and also proposed two monotone constraints, distance mutex and keyword mutex to help effective pruning.

Ian De Flipe[4] presented an efficient method to answer top-K spatial keyword query. They proposed an index structure IR^2-tree that combines signature files and R-tree to allow keyword search on spatial data objects that each have limited number of keywords. Using the IR^2-tree an efficient incremental algorithm is presented to answer the spatial keyword queries.

G. Cong, C.S. Jensen, and D. Wu [5] proposed an approach that computes the relevance between the documents of an object and a query. This relevance is then incorporated with the Euclidean distance between object and query to calculate an overall similarity of object to query.

Yufie Tao and Cheng Sheng[6], developed a new access method which is called as spatial inverted index. It extends the conventional inverted index to lay hold on multidimensional data, and uses the algorithms that can answer nearest neighbor queries with keywords in real time. They designed a variant of inverted index called spatial inverted index that is optimized for multidimensional points. This access method successfully includes point coordinates into a conventional inverted index with small space.

III. NEAREST NEIGHBOR SEARCH TECHNIQUE

A. IR-Tree, Approximation algorithm and Exact algorithm:

This method is used to retrieve a group of spatial web objects such that the query's keywords are cover by group's keywords and objects are near to the query location and have the lowest inter object distances. This method addresses the two instantiation of the group keyword query. First is to find the group of objects that cover the keywords such that the sum of their distances to the query is minimized. Second is to find a group of objects that cover the keywords such that sum of the maximum distance among an object in group of objects and query and maximum distance among two objects in group of objects is minimized. Both of these sub problems are NP-complete. Greedy algorithm is used to provide an approximation solution to the problem that utilizes the spatial keyword index IR-tree to reduce the search space. But in some application query does not contain a large number of keywords, for this exact algorithm is used that uses the dynamic programming. [1]

B. IUR-tree (Intersection union R-tree)

Geographic objects associated with descriptive texts are becoming common. This gives importance to spatial keyword queries that take both the location and text description of content. This technique is used to analyze the problem of reverse spatial and textual k nearest neighbor search i.e finding objects that takes the query object as one of their spatial textual similar objects. For this type of search hybrid index structure is used that successfully merge the location proximity with textual similarity. For searching, branch and bound algorithm is used. In addition to increase the speed of query processing a variant of IUR-tree and two optimization algorithm is used. To enhance the IUR-tree text clustering is used, in this objects of all the data base is group into clusters according to their text similarity. Each node of the tree is extended by the cluster information to create a hybrid tree which is called as cluster IUR-tree. To enhance the search performance of this tree two optimization methods is used, first is based on outlier detection and extraction and second method is based on text entropy. [2]

C. BR^*-tree :

This hybrid index structure is used to search m-closest keywords. This technique finds the closest tuples that matches the keywords provided by the user. This structure combines the R^*-tree and bitmap indexing to process the m-closest keyword query that returns the spatially closest objects matching m keywords To reduce the search space a priori based search strategy is used. Two monotone constraints is used as a priori properties to facilitates efficient pruning which is called as distance mutex and keyword mutex. But this approach is not suitable for handling ranking queries and in this number of false hits is large.[3]

D. IR^2-tree :

The growing number of applications requires the efficient execution of nearest neighbor queries which is constrained by the properties of spatial objects. Keyword search is very popular on the internet so these applications allow users to give list of keywords that spatial objects should contain. Such queries called as a spatial keyword query. This is consisted of query area and set of keywords. The IR^2-tree is developed by the combination of R-tree and signature files, where each node of tree has spatial and keyword information. This method is efficiently answering the top-k spatial keyword queries. In this signature is added to the every node of the tree. An able algorithm is used to answer the queries using the tree. Incremental nearest algorithm is used for the tree traversal and if root node signature does not match the query signature then it prunes the whole subtrees. But IR^2-tree has some drawbacks such as false hits where the object of final result is far away from the query or this is not suitable for handling ranking queries.[4]

E. Spatial inverted index and Minimum bounding method:

So, new access method spatial inverted access method is used to remove the drawbacks of previous methods such as false hits. This method is the variant of inverted index using for multidimensional points. This index stores the spatial region of data points and on every inverted list R-tree is built. Minimum bounding method is used for traversing the tree to prune the search space.

IV. CONCLUSION

This paper presents the survey of various techniques for nearest neighbor search for spatial database. As in the previous methods there were many drawbacks. The existing solutions incur too expensive space consumption or they are unable to give real time answer. So to overcome the drawbacks of previous methods, new method is based on variant of inverted index and R-tree and algorithm of minimum bounding method is used to reduce the search space. This method will increase the efficiency of nearest neighbor search too.

REFERENCES

[1] X. Cao, G. Cong, C.S. Jensen, and B.C. Ooi, "Collective Spatial Keyword Querying," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 373-384, 2011.

[2] J. Lu, Y. Lu, and G. Cong, "Reverse Spatial and Textual k Nearest Neighbor Search," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 349-360, 2011.

[3] D. Zhang, Y.M. Chee, A. Mondal, A.K.H. Tung, and M. Kitsuregawa, "Keyword Search in Spatial Databases: Towards Searching by Document," Proc. Int'l Conf. Data Eng. (ICDE), pp. 688-699, 2009.

[4] G. Cong, C.S. Jensen, and D. Wu, "Efficient Retrieval of the Top-k Most Relevant Spatial Web Objects," PVLDB, vol. 2, no. 1, pp. 337- 348, 2009.

[5] I.D. Felipe, V. Hristidis, and N. Rishe, "Keyword Search on Spatial Databases," Proc. Int'l Conf. Data Eng. (ICDE), pp. 656-665, 2008.

[6] Yufei Tao and Cheng Sheng, "Fast Nearest Neighbor Search with Keywords", IEEE transactions on knowledge and data engineering, VOL. 26, NO. 4, APRIL 2014.

[7] N. Beckmann, H. Kriegel, R. Schneider, and B. Seeger, "The R - tree: An Efficient and Robust Access Method for Points and Rectangles," Proc. ACM SIGMOD Int'l Conf. Management of Data,
pp. 322-331, 1990.

[8] C. Faloutsos and S. Christodoulakis, "Signature Files: An Access Method for Documents and Its Analytical Performance Evaluation," ACM Trans. Information Systems, vol. 2, no. 4, pp. 267-288, 1984.

[9] G. R. Hjaltason and H. Samet. Distance browsing in spatial databases. ACM Transactions on Database Systems (TODS), 24(2):265–318, 1999.