OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

SURVEY ON NEAREST NEIGHBOUR SEARCH OVER SPATIAL KEYWORDS DATABASE

Miss. Snehal Patil[1], Miss. Monika Nemade[2], Miss. Kalyani Birajdar[3], Miss. Reshma Magar[4], Prof.Vinayak Kadam[5]

B. E. Students, Dept. of Computer Science, JSPM'S BSIOTR, Wagholi, Pune, Maharashtra, India[1,2,3,4]

Assistant Professor, Dept. of Computer Science, JSPM'S BSIOTR, Wagholi, Pune, Maharashtra, India[5]

Abstract: **Conventional spatial queries, such as range search and nearest neighbor retrieval, involve only conditions on objects geometric properties. Today, many modern applications call for novel forms of queries that aim to find objects satisfying both a spatial predicate, and a predicate on their associated texts. For example, instead of considering all the restaurants, a nearest neighbor query would instead ask for the restaurant that is the closest among those whose menus contain "steak, spaghetti, brandy" all at the same time. A search engine is able to efficiently support novel forms of spatial queries which are integrated with keyword explore. The accessible solutions to such queries also acquire prohibitive space consumption or are unable to give real time answers. As today's need is smart search from search engine not just what they query but relevant to query and similar to that query and where they actually find that product, place or person in real world. To provide such smart search results fast nearest neighbor search with keywords technique by using spatial inverted index (SI-index) has been proposed by researchers. This technique has great efficiency to provide results but processes huge data to fulfill queries. To overcome huge data pre-processing, proposed algorithm reduces no of objects processed by this technique to minimize the memory and processing cost. Proposed technique uses limits to fetch limited objects from the dataset or database.**

Keywords: Superimposed Coding, nearest neighbor search, keyword search, spatial index, FNNSK, AFNNSK.

----------∴∴∴----------

I INTRODUCTION

A spatial database manages multidimensional objects (such as points, rectangles, etc.), and provides fast access to those objects based on different selection criteria. The importance of spatial databases is reflected by the convenience of modelling entities of reality in a geometric manner. For example, locations of restaurants, hotels, hospitals and so on are often represented as points in a map, while larger extents such as parks, lakes, and landscapes often as a combination of rectangles. Many functionalities of a spatial database are useful in various ways in specific contexts. For instance, in a geography information system, range search can be deployed to find all restaurants in a certain area, while nearest neighbour retrieval can discover the restaurant closest to a given address. There are easy ways to support queries that combine spatial and text features. For example, for the above query, we could first fetch all the restaurants whose contain the set of keywords {steak, spaghetti, brandy}, and then from the retrieved restaurants, find the nearest one. Similarly, one could also do it reversely by targeting first the spatial conditions—browse all the restaurants in ascending order of their distances to the query point until encountering one whose menu has all the keywords. The major drawback of these straightforward approaches is that they will fail to provide real time answers on difficult inputs.

The problem of neighbour search is as follow:

A. Pattern Recognition:

It is largely synonymous to machine learning. This branch of artificial intelligence focuses on the recognition of patterns and regularities in data [1]. In many cases, these patterns are learned from labelled "training" data (supervised learning), but when no labelled data is available other algorithms can be used to discover previously unknown patterns (unsupervised learning).

B. Statistical Classification:

In machine learning and statistics, classification is that the drawback of distinguishing to that of which of classes (Sub-populations) a replacement observation belongs, on the idea of a training set of information containing observations

(or instances) whose class membership is thought [2]. The individual observations are analysed into a collection of quantitative properties, referred to as varied various variables, features, etc.

C. Online Advertising

This is also called Internet advertising, uses the Internet to deliver promotional marketing messages to consumers [3].It includes email marketing, search engine marketing, social media marketing, many types of display advertising(including web banner advertising), and mobile advertising.

A typical example is that the real nearest neighbour lies quite far away from the query point, while all the closer neighbours are missing at least one of the query keywords. Spatial queries with keywords have not been extensively explored. In the past years, the community has sparked enthusiasm in studying keyword search in relational databases. It is until recently that attention was diverted to multidimensional data .The best method to date for nearest neighbour search with keywords is due to Felipe et al.

II RELATED WORK

Anjum and Saktel [2] have presented a survey of such various techniques as collective spatial keyword query, the combined notion of keyword search with reverse nearest neighbour query, hybrid indexing structure bR*-tree, efficient method to answer top-k spatial keyword query, computing the relevance between the documents of an object and a query, spatial inverted index, etc., for nearest neighbour search for spatial database, The authors have proposed, to overcome the drawbacks of previous methods, like, expensive space consumption, unable to give real time answer, etc., a new method based on variant of inverted index and R-tree and algorithm of minimum bounding method to reduce the search space. The approach accommodates a query with both spatial data and associated text. For this type of query, a variant of inverted index is used that is effective for multidimensional points and comes with an R-tree built on every inverted list, and uses the algorithm of minimum bounding method that can answer the nearest neighbour queries with keywords in real time.

The spatial inverted index list and enhanced search [4]. The spatial inverted index is persecution each capability of the R-Tree and also the process of signature files. In his implemented work, the proposed thought of the list merging and distance alignment are used to facilitate for searching, and also the compression scheme is used to produce the effectiveness of the quick search. The improved search is used for locating the objects supported the users priority level. In this system uses efficient algorithmic program to search out the precise nearest neighbour based on the geometrician distance for large-scale computer vision issues [5]. There algorithmic program is well suited for high dimensional and large-scale issues and also show that his algorithm is improved additional by partitioning input vectors recursively. Contrary to most of existing promptly nearest neighbour search algorithms, his technique reports the precise nearest neighbour not associate approximate one and needs a very straightforward pre-processing with no complicated information structures. They contribute the theoretical analysis of his algorithm and appraise its performance in artificial and real information.

Yufei Tao and Cheng Sheng [3] have proposed a variant of inverted index that optimizes multidimensional points called the spatial inverted index (SI-index). IR2 -tree inherits a drawback of signature files: file hits, e.g., a signature file by virtue of its conservative nature may still direct to search to some objects, though they do not posses all the keywords. Thus, only in full text description, it becomes possible to verify if a query is satisfied or not. False hit problem may exist in other methods too for approximate set membership tests with compact storage. The SI-index approach successfully incorporates point's coordinates into a conventional inverted index with small extra space, owing to a delicate compact storage scheme. This is achieved in two competing ways for query processing, namely, sequentially merge multiple lists very much like merging traditional inverted lists by ids or alternatively, leverage the R-tree to browse the points of all relevant lists in descending order of their distance to the query point. It is claimed that SI-index significantly outperforms the IR2 -tree in query efficiency usually by a factor of magnitude.

Sophiya.K et al [8] proposed the spatial inverted index list and enhanced search. The spatial inverted index is persecution each capability of the R-Tree and also the process of signature files. In his implemented work, the proposed thought of the list merging and distance alignment are used to facilitate for searching, and also the compression scheme is used to produce the effectiveness of the quick search. The improved search is used for locating the objects supported the users priority level. Rajkumar R et al [5] proposed a system uses efficient algorithmic program to search out the precise nearest neighbour based on the geometrician distance for large-scale computer vision issues. There algorithmic program is well suited for high-dimensional and large-scale issues and also show that his algorithm is improved additional by partitioning input vectors recursively. Contrary to most of existing promptly nearest neighbour search algorithms, his technique reports the precise nearest neighbour not associate approximate one and needs a very straightforward pre-processing with no complicated information structures. They contribute the theoretical analysis of his algorithm and appraise its performance in artificial and real information. In this paper [6], they propose to use the classification procedure of thin

illustration to enhance NNC. The designed technique has the following basic idea: the training samples aren't unrelated and also the distance between the test sample and a training sample shouldn't be severally calculated and should take under consideration the relationship between totally different training samples.

The proposed technique initial uses a linear combination of all the training samples to represent the analysis sample and so exploits modified distance to classify the test sample. The strategy obtains the coefficients of the linear combination by resolution a linear system. The strategy then calculates the gap between the test sample and also the results of multiplying every training sample by the corresponding constant and assumes that the test sample is from constant class because the training sample that has the minimum distance. The method in a complicated way modifies NNC and considers the relationship between totally different training samples, thus it's able to turn out higher classification accuracy.

III PROPOSED SYSTEM

We have proposed Advanced Fast Nearest Neighbor Search with Keywords to overcome the limitless search of FNNSK. To filter objects from dataset, proposed technique uses lower limit and upper limits which are respectively minimum and maximum distance from query location. AFNNSK retrieves only those object which satisfy lower limit and upper limit, calculate the relative positions of each object from query location and update relative location of object. Now use subset(filtered objects) of objects to search nearest neighbors using previously proposed FNNSK. First search keyword from limited area bounded by minimum distance 1unit and maximum distance limit 50 units, process only those objects and associated keywords which satisfy limits and search nearest neighbor using FNNSK. There are two possibility of result, first nearest neighbor found and second no neighbor found which satisfy query. In first case where neighbor found no further processing required. In second case where no result found we have to update limits, to search in next limited area, where lower limit 50 unit and upper limit 75 units and search again using AFNNSK. In second search in area will not process any objects in A, which are already processed in first search to reduce overhead of processing and memory.

In proposed , we design a variant of inverted index that is optimized for multidimensional points, and is thus named the spatial inverted index (SI-index).This access method successfully incorporates point coordinates into a conventional inverted index with small extra space, owing to a delicate compact storage scheme.

Meanwhile, an SI-index preserves the spatial locality of data points, and comes with an R-tree built on every inverted list at little space overhead. As a result, it offers two competing ways for query processing.

We can (sequentially) merge multiple lists very much like merging traditional inverted lists by ids. Alternatively, we can also leverage the R-trees to browse the points of all relevant lists in ascending order of their distances to the query point. As demonstrated by experiments, the SI-index significantly out performs the IR2 -tree in query efficiency, often by a factor of orders of magnitude

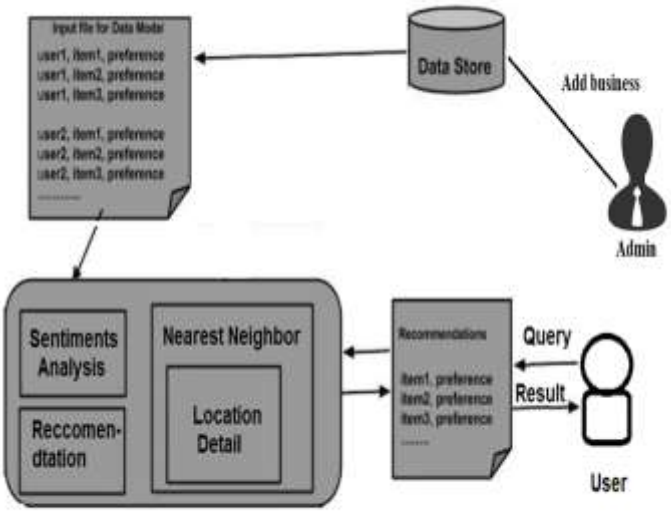

Figure 1 System Architecture

IV CONCLUSION

In advanced fast nearest neighbor search with keyword have improvement over FNNSK. AFNNSK is reduces processing of number of objects and associated keywords with them using limited circular area search bounded by upper limit and lower limits. The proposed technique will definitely reduce the processing load and memory load on live query server. The proposed technique seems to a solution for low end server to provide fast nearest neighbor search with keywords due to low processing requirement and low memory requirements.

REFERENCES

[1] X. Cao, G. Cong, and C. S. Jensen. Retrieving top-k prestige-based relevant spatial web objects. PVLDB, 3(1):373–384, 2010.

[2] X. Cao, G. Cong, C. S. Jensen, and B. C. Ooi. Collective spatial keyword querying. In Proc. of ACM Management of Data (SIG-MOD) , pages 373–384, 2011.

[3] B. Chazelle, J. Kilian, R. Rubinfeld, and A. Tal. The bloomier filter: an efficient data structure for static support lookup tables. In Proc. of the Annual ACM-

SIAM Symposium on Discrete Algorithms (SODA), pages 30–39, 2004.

[4] Y.-Y. Chen, T. Suel, and A. Markowetz. Efficient query processing in geographic web search engines. In Proc. of ACM Management of Data (SIGMOD), pages 277–288, 2006.

[5] E. Chu, A. Baid, X. Chai, A. Doan, and J. Naughton. Combining keyword search and forms for ad hoc querying of databases. In Proc. of ACM Management of Data (SIGMOD), 2009.

[6] G. Cong, C. S. Jensen, and D. Wu. Efficient retrieval of the top-k most relevant spatial web objects. PVLDB, 2(1):337–348, 2009.

[7] C. Faloutsos and S. Christodoulakis. Signature files: An access method for documents and its analytical performance evaluation.ACM Transactions on Information Systems (TOIS), 2(4):267–288,1984.

[8] I. D. Felipe, V. Hristidis, and N. Rishe. Keyword search on spatial databases. In Proc. of International Conference on Data Engineering (ICDE), pages 656–665, 2008.

[9] R. Hariharan, B. Hore, C. Li, and S. Mehrotra. Processin g spatial-keyword (SK) queries in geographic information retrieval (GIR)systems. InProc. of Scientific and Statistical Database Management(SSDBM), 2007.

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

[11] V. Hristidis and Y. Papakonstantinou. Discover: Keyword search in relational databases. In Proc. of Very Large Data Bases (VLDB),pages 670–681, 2002.

[12] I. Kamel and C. Faloutsos. Hilbert R-tree: An improved r-tree using fractals. In Proc. of Very Large Data Bases (VLDB),pages500–509, 1994.

[13] J. Lu, Y. Lu, and G. Cong. Reverse spatial and textual k nearest neighbour search. In Proc. of ACM Management of Data (SIGMOD),pages 349–360, 2011.

[14] S. Stiassny. mathematical analysis of various superimposed coding methods. Am. Doc., 11(2):155–169, 1960.

[15] J. S. Vitter. Algorithms and data structures for external memory. Foundation and Trends in Theoretical Computer Science , 2(4):305–474,2006.

[16] D. Zhang, Y. M. Chee, A. Mondal, A. K. H. Tung, and M. Kitsure-gawa. Keyword search in spatial databases: Towards searching by document. In Proc. of International Conference on Data Engineering (ICDE), pages 688–699, 2009.

[17] Y. Zhou, X. Xie, C. Wang, Y. Gong, and W.-Y. Ma. Hybrid index structures for location-based web search. In Proc. of Conference on Information and Knowledge Management (CIKM), page