A Novel Approach to find out the Nearest Object Location

B.Sowmya
PG Scholar,CSE,
Sri Vasavi Engineering College,
Tadepalligudem.

A.Venkanna Babu,
Asst. Professor,CSE,
Sri Vasavi Engineering College,
Tadepalligudem.

Abstract-- **The nearest neighbor search is one of the most primitive operations in machine learning, pattern recognition, and information retrieval. It is also a crucial task in a lot of computer vision applications such as image clustering, object classification, image search, visual tracking, computational photography, and so on. Recently, the efficiency of the nearest neighbor search algorithm in both speed and space becomes a critical issue in computer vision research due to the demand for large-scale problems involving high dimensional multimedia data. We propose an efficient algorithm to find the exact nearest neighbor based on the Euclidean distance for large scale computer vision problems. We embed data points nonlinearly onto a low-dimensional space by simple computations and prove that the distance between two points in the embedded space is bounded by the distance in the original space. Instead of computing the distances in the high-dimensional original space to find the nearest neighbor, a lot of candidates are to be rejected based on the distances in the low-dimensional embedded space; due to this property, our algorithm is well-suited for high-dimensional and large-scale problems. We also show that our algorithm is improved further by partitioning input vectors recursively. Contrary to most of existing fast nearest neighbor search algorithms, our technique reports the exact nearest neighbor—not an approximate one—and requires a very simple preprocessing with no sophisticated data structures. We provide the theoretical analysis of our algorithm and evaluate its performance in synthetic and real data.**

Keywords: Information Retrieval Tree, Keyword Search, Spatial Inverted

I.INTRODUCTION

Nearest neighbor search (NNS), also known as closest point search, similarity search. It is an optimization problem for finding closest (or most similar) points. Nearest neighbor search which returns the nearest neighbor of a query point in a set of points, is an important and widely studied problem in many fields, and it has wide range of applications. We can search closest point by giving keywords as input; it can be spatial or textual. A spatial database use to manage multidimensional objects i.e. points, rectangles, etc. Some spatial databases handle more complex structures such as 3D objects, topological coverage's, linear networks. While typical databases are designed to manage various NUMERIC'S and character types of data, additional functionality needs to be added for databases to process spatial data type's efficiently and it provides fast access to those objects based on different selection criteria. Keyword search in document performed with various

Keyword search is the most popular information discovery method because the user does not need to know either a query language or the underlying structure of the data. The search engines available today provide keyword search on top of sets of documents. When a set of query keywords is provided by the user, the search engine returns all documents that are associated with these query keywords. Solution to such queries is based on the IR2-tree, but IR2-tree having some drawbacks. Efficiency of IR2-tree badly is impacted because of some drawbacks in it. The solution for overcoming this problem should be searched. Spatial inverted index is the technique which will be the solution for this problem. Spatial database manages multidimensional data that is points, rectangles.

This paper gives importance spatial queries with keywords [5] [6] [9] . Spatial queries with keywords take arguments like location and specified keywords and provide web objects that are arranged depending upon spatial proximity and text relevancy. Some other approaches take keywords as Boolean predicates [1] [2], searching out web objects that contain keywords and rearranging objects based on their spatial proximity. Some approaches use a linear ranking function [7] [8] to combine spatial proximity and textual relevance. Earlier study of keyword search in relational databases is gaining importance. Recently this attention is diverted to multidimensional data [3] [4] . N. Rishe, V. Hristidis and D. Felipe has proposed best method to develop neighbor search with keywords. For keyword-based retrieval, they have integrated R-tree with spatial index and signature file . By combining R-tree and signature they have developed a structure called the IR2-tree . IR2-tree has merits of both R-trees and signature files. The IR2-tree preserves object's spatial proximity which important for solving spatial queries.

II. RELATED WORKS

IR2 – Tree The IR2 – Tree combines the R-Tree and signature file. First we will review Signature files. Then IR2-trees are discussed. Consider the knowledge of R-trees and the best- first algorithm for Near Neighbor Search. Signature file is known as a hashing-based framework and hashing -based framework is which is known as superimposed coding (SC).

Drawbacks of the IR2-Tree

IR2-Tree is first access method to answer nearest neighbour queries. IR2-tree is popular technique for indexing data but it having some drawbacks, which impacted on its efficiency. The disadvantage called as false

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICACC - 2016 Conference Proceedings

hit affecting it seriously. The number of 940 false positive ratio is large when the aim of the final result is far away from the query point and also when the result is simply empty. In these cases, the query algorithm will load the documents of many objects; as each loading necessitates a random access, it acquires costly overhead .

Keyword search on spatial databases This work, mainly focus on finding top-k Nearest Neighbors, in this method each node has to match the whole querying keywords. As this method match the whole query to each node, it does not consider the density of data objects in the spatial space. When number of queries increases then it leads to lower the efficiency and speed. They present an efficient method to answer top-k spatial keyword queries. This work has the following contributions: 1) the problem of top-k spatial keyword search is defined. 2) The IR2-Tree is proposed as an efficient indexing structure to store spatial and textual information for a set of objects. There are efficient algorithms are used to maintain the IR2-tree, that is, insert and delete objects. 3) An efficient incremental algorithm is presented to answer top-k spatial keyword queries using the IR2-Tree. Its performance is estimated and compared to the current approaches. Real datasets are used in our experiments that show the significant improvement in execution times.

Disadvantages:

1. Each node has to match with querying keyword. So it affects on performance also it becomes time consuming and maximizing searching space.
2. IR2-tree has some drawbacks.

Processing Spatial-Keyword (SK) Queries in Geographic Information Retrieval (GIR) Systems.

Location based information stored in GIS database. These information entities of such databases have both spatial and textual descriptions. This paper proposes a framework for GIR system and focus on indexing strategies that can process spatial keyword query. The following contributions in this paper: 1) It gives framework for query processing in Geo- graphic Information Retrieval (GIR) Systems. 2) Develop a novel indexing structure called KR*-tree that captures the joint distribution of keywords in space and significantly improves performance over existing index structures. 3) This method have conducted experiments on real GIS datasets showing the effectiveness of our techniques compared to the existing solutions. It introduces two index structures to store spatial and textual information.

Advantages:

1. Easy of maintaining two separate indices.
2. Performance bottleneck lies in the number of candidate object generated during the filtering stage.

Disadvantages:

1. If spatial filtering is done first, many objects may lie within a query is spatial extent, but very few of them are relevant to query keywords. This increases the disk access cost by generating a large number of candidate objects. The subsequent stage of keyword filtering becomes expensive.

Hybrid index

Advantages and limitations:

1. When query contains keywords that closely correlated in space, this approach suffer from paying extra disk cost accessing R*-tree and high overhead in subsequent merging process.

Hybrid Index Structures for Location-based Web Search.

There is more and more research interest in location-based web search, i.e. searching web content whose topic is related to a particular place or region. This type of search contains location information; it should be indexed as well as text information. text search engine is set-oriented where as location information is two-dimensional and in Euclidean space. In previous paper we see same two indexes for spatial as well as text information. This creates new problem, i.e. how to combine two types of indexes. This paper uses hybrid index structure, to handle textual and location based queries, with help of inverted files and R*-trees. It considered three strategies to combine these indexes namely: 1) inverted file and R*-tree double index.2) first inverted file then R*-tree.3) first R*-tree then inverted file.

It implements search engine to check performance of hybrid structure that contains four parts:

1. An extractor which detects geographical scopes of web pages and represents geographical scopes as multiple MBRs based on geographical coordinates.
2. The work of indexer is use to build hybrid index structures integrate text and location information.
3. The work of ranker is to ranks The results by geographical relevance as well as non-geographical relevance.
4. An interface which is friendly for users to input location-based search queries and to obtain geographical and textual relevant results.

Advantages:

1. Instead of using two indexes for textual and spatial information. this paper gives hybrid index structures that integrate text indexes and spatial indexes for location based web search.

Disadvantages:

1. Indexer wants to build hybrid index structures to integrate text and location information of web pages. To textually index web pages, inverted files are a good. To spatially index web pages, two-dimensional spatial indexes are used, both include different approaches, this cause to degrading performance of indexer.
2. In ranking phase, it combine geographical ranking and non-geographical ranking, combination of two

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICACC - 2016 Conference Proceedings

rankings and the computation of geographical relevance may affects on performance of ranking.

III. PROPOSED SYSTEM

To overcome the drawbacks of previous applications, we proposed an application for users. In our system we are mainly dealing with searching and nearer location issues and database manage multidimensional objects which resulted in failure of previous systems [2]. To deal with spatial index as searching the entered keyword and from that find the nearest location having that keyword available and showing the location.

For example the user will enter the keyword searching for menus available in restaurant which will nearer from its position. Whenever user will enter keyword (menu name) it will match data with the hotel database server and find the nearest restaurant with the available entered menu by customer. For nearest restaurant we are using IR2tree & compression. The IR2-Tree is a combination of an R-Tree and signature files. In particular, each node of an IR2-Tree contains both spatial and keyword information; the former in the form of a minimum bounding area and the latter in the form of a signature. An IR2-Tree facilitates both top-k spatial queries and top-k spatial keyword queries as we explain below. More formally, an IR2-Tree R is a height-balanced tree data structure, where each leaf node has entries of the form (Obj Ptr, A, S). Obj Ptr and A are defined as in the R-Tree while S is the signature of the object referred by Obj Ptr. Anon-leaf node has entries of the form (Node Ptr, A, S). Node Ptr and A are defined as in the R-Tree while S is the signature of the node. The signature of a node is the superimposition (OR-ing) of all the signatures of its entries. Thus a signature of a node is equivalent to a signature for all the documents in its sub tree.

Algorithm:

Input:-Menu list, Input query

{
1. Let Hl be the list of hotels along with their latitude and longitude Let Hm be the menu available in respective hotel
2. Let Sm be the signature set for each menu,
   For each (menu in Hl)
   {
       //signature generation
           Hashing (menu);
       Add hash menu in Sm;
   }
3. Let Sh be the signature for hotel list
   For each (Hl in List)
   {
       //Generation of signature for each Hotel
       Hashing (menu (i) OR menu (i+1));
       Add hashing to Sh
   }
4. Declare HLc is list of hotel having all menu which satisfies query
   For each (val in Sh)

{
    Check if (q = val)
    Add hotel in HLc
}
5. Declare SHl sorted list of hotelDeclare Dmin is min distance from list of sorted hotels
//List along with distances
Hl=Sorted list (HLc);
Dmin=min(Hl)
}
6. Sorting

User can search the hotel by sorting the names of hotels Inverted indexes (I-index) have proved to be an effective access method for keyword-based document retrieval[1]. In the spatial context, nothing prevents us from treating the text description Wp of a point p as a document, and then, building an I-index. Each word in the vocabulary has an inverted list, enumerating the ids of the points that have the word in their documents.

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] I. De Felipe, V. Hristidis, and N. Rishe. Keyword search on spatial databases. In ICDE, pp. 656–665, 2008.

[2] D. Zhang, Y. M. Chee, A. Mondal, A. K. H. Tung, and M. Kitsuregawa. Keyword search in spatial databases: Towards searching by document. In ICDE, pp. 688– 699, 2009

[3] R. Hariharan, B. Hore, C. Li, and S. Mehrotra, "Processing Spatial- Keyword (SK) Queries in Geographic Information Retrieval (GIR) Systems," Proc. Scientific and Statistical Database Management (SSDBM), 2007.

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

[5] Y.-Y. Chen, T. Suel, and A. Markowetz. Efficient query processing in geographic web search engines. In SIGMOD, pp. 277–288, 2006.

[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] I. De Felipe, V. Hristidis, and N. Rishe. Keyword search on spatial databases. In ICDE, pp. 656–665, 2008.

[8] Y. Zhou, X. Xie, C. Wang, Y. Gong, and W.-Y. Ma, "Hybrid Index Structures for Location-Based Web Search," Proc. Conf. Information and Knowledge Management (CIKM), pp. 155-162, 2005.

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