ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

Spatial Index Keyword Search in Multi-Dimensional Database

Pradeep Panchal, Prof. Amol. R. Dhakne

Department of Computer Engineering, Savitribai Phule Pune University, Maharashtra, India, Flora Institute of Technology, Pune, India

Department of Computer Engineering, Savitribai Phule Pune University, Maharashtra, India, Flora Institute of Technology, Pune, India

ABSTRACT: Nearest neighbor search in multimedia databases needs more support from similarity search in query processing. Range search and nearest neighbor search depends mostly on the geometric properties of the objects satisfying both spatial predicate and a predicate on their associated texts. We do have many mobile applications that can locate desired objects by conventional spatial queries. Current best solution for the nearest neighbor search is IR2 trees which have many performance bottlenecks and deficiencies. So, a novel method is introduced in this paper in order to increase the efficiency of the search called as Spatial Inverter Index. This new SI index method enhances the conventional inverted index scheme to cope up with high multidimensional data and along with algorithms that's compatible with the real time keyword search.

KEYWORDS: Querying, multi-dimensional data, indexing, hashing.

I. INTRODUCTION

a. BACKGROUND

Multi-dimensional objects such as points, rectangles managed by spatial databases provide fast access to those objects based on different selection criteria. For example, location of hospitals, hotels and theatres are represented as points whereas parks, lakes and shopping malls are represented as rectangles [1]. For instance, GIS range search gives all the cafes in certain area and nearest neighbor gives location of café near to our geometrical location. Today, the search engine optimization has made a realistic approach to write a spatial query in a brand new style. Some of may have few applications which finds the objects in a huge multidimensional data along with its geometrical locations and associated texts. There are easy ways to support queries that combine spatial and text features. For example, if we want to search a café whose menu contains keywords {Mocha, Espresso, Cappuccino} it would fetch all the restaurants with the keywords and from that list gives the nearest one. This approach can also be in another way but this straight forward approach has a drawback, which they will fail to provide real time answers on difficult inputs.

b. MOTIVATIONS

Objects (e.g., images, chemical compounds, or documents) are often characterized by a collection of relevant features, and are commonly represented as points in a multi-dimensional attribute space. For example, images (chemical compounds) are represented using color (molecule) feature vectors. These objects also very often have descriptive text information associated with them, e.g., images are tagged with locations. In this paper, we consider multi-dimensional datasets where each data point has a set of keywords. The presence of keywords allows for the development of new tools for querying and exploring these multi-dimensional datasets. A typical example, while all the closer neighbors are missing at least one of the query keywords, that the real nearest neighbor lies quite far away from the query point. The introduction of internet has given rise to an ever increasing amount of text data associated with multiple dimensions (attributes), for example customer feedbacks in online shopping website like flip kart as they are always associated with

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

the price, specifications and product model. Keyword query, one of the most popular and easy-to-use ways retrieves useful data from plain text documents. Given a set of keywords, existing methods aim to find joins or all the relevant items that contains a few or all the keywords. Spatial queries with keywords have not been explored. Recently, attention was diverted to multimedia databases [8]. The integration of two well-known concepts: R-tree [2], a popular spatial index, and signature file [4], an effective method for keyword-based document retrieval. This makes to develop a structure called IR2 trees, which has strengths of both signature files and R-Trees. Like R-Trees, IR2 -Tree has object spatial proximity that solves spatial queries efficiently. On the other side, the IR2 -tree is able to filter a considerable portion of the objects that do not contain all the query keywords, like signature files

II. LITERATURE SURVEY

Shilpa et al [12]A variety of queries, semantically different from our NKS queries, have been studied in literature on text-rich spatial datasets. Location-specific keyword queries on the web and in the GIS systems [9], [10] were earlier answered using a combination of R-Tree [3] and inverted index. Felipe et al. [4] developed IR2 -Tree to rank objects from spatial datasets based on a combination of their distances to the query locations and the relevance of their text descriptions to the query keywords. Cong et al. [5] integrated R-tree and inverted file to answer a query similar to Felipe et al. [4] using a different ranking function. Martins et al. [6] computed text relevancy and location proximity independently, and then combined the two ranking scores.

Cao et al. [7] recently proposed a method to retrieve a group of spatial web objects such that the group's keywords cover the query's keywords and the objects in the group are nearest to the query location and have the lowest inter-object distances. Other keyword-based queries on spatial datasets are aggregate nearest keywordsearch in spatial databases [8], top-k preferential query [9], and finding top-k sites in a spatial data based on their influence on feature points [2], and optimal location queries [2].

Our NKS query is similar to the m-closest keywords query of Zhang et al. [7]. They designed bR*-Tree based on a R*-tree [3] that also stores bitmaps and minimum boundingrectangles (MBRs) of keywords in every node along with points MBRs. The candidates are generated by the apriorialgorithm [4]. They prune unwanted candidates based on the distances between MBRs of points or keywords and the best found diameter. Their pruning techniques become ineffective with an increase in the dataset dimension as there is a large overlap between MBRs due to the curse of dimensionality. This leads to an exponential number of candidates and large query times.

A poor estimation of starting diameter further worsens the performance of their algorithm. bR*-Tree also suffered from a high storage cost, therefore Zhang et al. modified bR*-Tree to create Virtual bR*-Tree [2] in memory at run time. Virtual bR*-Tree is created from a pre-stored R*-Tree which indexes all the points, and an inverted index which stores keyword information and path from the root node in R*-Tree for each point. BothbR*-Tree and Virtual bR*-Tree, are structurally similar, and use similar candidate generation and pruning techniques. Therefore, Virtual bR*-Tree shares similar performance weaknesses as bR*-Tree. Tree-based indices, e.g., R-Tree [3] and M-Tree [5], have been researched extensively for an efficient near neigh-bor search in high-dimensional spaces. These indices fail toscale to dimensions greater than 10 because of the curse of dimensionality [6].

VA-file [6] and iDistance [7] provide better scalability with the dataset dimension. However, the task of designing an efficient method for solving NKS queries by adapting VA-file or iDistance is not obvious. Random projections [8] with hashing [2] have come to be the state-of-the-art method for an efficient near neighbor search in high-dimensional datasets. Datar et al. [8] used random vectors constructed from p-stable distributions to project points, and then computed hash keys for the points by splitting the line of projected values into disjoint bins. They concatenated hash keys obtained for a point from m random vectors to create a final hash key for the point. All points were indexed into a hashtable using their hash keys.

Our index structure is inspired from the same. Multi-way distance joins of a set of multi-dimensional datasets, each of which is indexed into a R-Tree, have been studied in literature [3], [4]. As discussed above, a tree-based index fails to

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

scale with the dimension of the dataset.Further, it is not straightforward to adapt these algorithms if every query requires a multi-way distance join only on a subset of the points of each dataset.ProMiSH-E uses a set of hashtables and inverted indices to perform a localized search of the results. ProMiSH-E hashta-bles are inspired from Locality Sensitive Hashing (LSH) [8],which is a state-of-the-art method for the nearest neighbor search in high-dimensional spaces.

The index structure of ProMiSH-E supports accurate search, unlike LSH-based meth-ods that allow only approximate search with probabilistic guar-antees. ProMiSH-E creates hashtables at multiple bin-widths,called scales. A search in a hashtable yields subsets of points that contain query results.ProMiSH-E explores each subset using a novel pruning based strategy. An optimal strategy is NP-Hard; therefore, ProMiSH-E uses a greedy approach. ProMiSH-A is an approximate variation of ProMiSH-E to achieve even more space and time efficiency.[12]

III. EXISTING SYSTEM APPROACH

Present system gives the real nearest neighbor that lies quite far away from the query location, while all the closer objects missing one or any of the keywords. This system mainly focuses on finding the nearest neighbor where each node satisfies all the query keywords. This leads to low efficiency for incremental query. The problem is Implement k nearest neighbor search algorithm using for given data set and to find out closest point from give query also analyzes.The result fetches time and accuracy. Implement the Inverted index algorithm by extending point the k nearest neighbor and forming R tree to find closest point from given set of query and also analyze the result against time and result accuracy.

Disadvantages:-
The IR2-tree is the first access method for answering NN queries with keywords. Although IR2 Trees gives pioneering solutions, it also has few drawbacks that affect its efficiency. The most important drawback is the result set may be empty or the number of false hits can be very large when the object of the final result is far away from query point. The query algorithm would need to load the documents of many objects, incurring expensive overhead.The R-trees allow us to remedy awkwardness in the way NN queries are processed with an I-index. Recall that, to answer a query, currently we have to first get all the points carrying all the query words.

IV. PROPOSED SYSTEM ARCHITECTURE

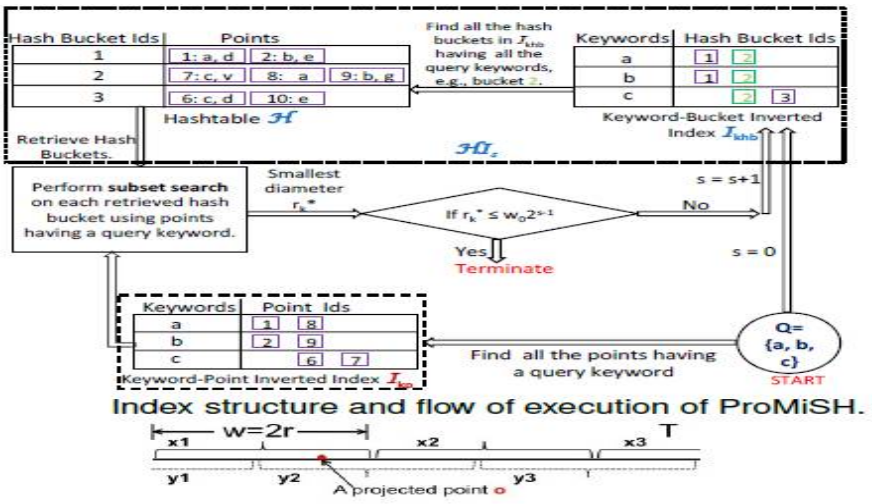

Fig No 01 Proposed System Architecture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

The drawbacks of R-Trees and inverted index can be overcome by designing a variant of inverted index that supports compressed coordinate embedding. This system deals with searching and nearer location issues and database manage multidimensional objects which resulted in failure of previous systems. 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 of restaurant having menus available in map. So easier to find the location of nearer restaurant in map having the available keyword. Spatial databases manage multidimensional objects and provide quick access to those objects. The importance of spatial databases is mirrored by the convenience of modeling entities of reality in an exceedingly geometric manner.The Inverted Index is compressed by coordinate encoding which makes Spatial Inverted Index (SI-Index). Query processing with an SI-index can be done either by together or merging with R-Trees in distance browsing manner. The inverted index compress eliminates the defect of a conventional index such that an SI-index consumes much less space.

Advantages:-

1. Compression is already wide used technology to reduce the space of an inverted index where each inverted list contains only ids.
2. So, the effective approach is used to record gaps with consecutive ids. Compressing an SI index is less straightforward than other approaches.
3. For example, if we decide to sort the list by ids, gap-keeping on ids may lead to good space saving, but its application on the x- and y-coordinates would not have much effect.

V. CONCLUSION

From the consideration all the above points we conclude that nearest neighbor search in multimedia databases needs more support from similarity search in query processing. Range search and nearest neighbor search depends mostly on the geometric properties of the objects satisfying both spatial predicate and a predicate on their associated texts. In existing system The IR2-tree is the first access method for answering NN queries with keywords. Although IR2 Trees gives pioneering solutions, it also has few drawbacks that affect its efficiency. The most important drawback is the result set may be empty or the number of false hits can be very large when the object of the final result is far away from query point. The query algorithm would need to load the documents of many objects, incurring expensive overhead. In proposed spatialdatabases manages multidimensional objects and provide quick access to those objects. The importance of spatial databases is mirrored by the convenience of modeling entities of reality in an exceedingly geometric manner.

REFERENCES

[1] D. Zhang, B. C. Ooi, and A. K. H. Tung, "Locating mapped resources in web 2.0", in Proc. IEEE 26th Int. Conf. Data Eng., 2010, pp. 521–532.
[2] V. Singh, S. Venkatesha, and A. K. Singh, "Geo-clustering of images with missing geotags", in Proc. IEEE Int. Conf. Granular Comput., 2010, pp. 420–425.
[3] V. Singh, A. Bhattacharya, and A. K. Singh, "Querying spatial patterns", in Proc. 13th Int. Conf. Extending Database Technol.: Adv. Database Technol., 2010, pp. 418–429.
[4] X. Cao, G. Cong, C. S. Jensen, and B. C. Ooi, "Collective spatial keyword querying", in Proc. ACM SIGMOD Int. Conf. Manage Data, 2011, pp. 373–384.
[5] C. Long, R. C.-W. Wong, K. Wang, and A. W.-C. Fu, "Collective spatial keyword queries: A distance owner-driven approach", in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2013, pp. 689–700.
[6] A. Khodaei, C. Shahabi, and C. Li, "Hybrid indexing and seamless ranking of spatial and textual features of web documents", in Proc. 21st Int. Conf. Database Expert Syst. Appl., 2010, pp. 450–466.
[7] V. Singh and A. K. Singh, "SIMP: Accurate and efficient near neighbor search in high dimensional spaces", in Proc. 15th Int. Conf. Extending Database Technol., 2012, pp. 492–503.
[8] I. De Felipe, V. Hristidis, and N. Rishe, "Keyword search on spatial databases", in Proc. IEEE 24th Int. Conf. Data Eng., 2008, pp. 656–665.
[9] D. Zhang, Y. M. Chee, A. Mondal, A. K. H. Tung, and M. Kitsuregawa, "Keyword search in spatial databases: Towards searching by document," in Proc. IEEE 25th Int. Conf. Data Eng., 2009, pp. 688–699.
[10] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality sensitive hashing scheme based on p-stable distributions," in Proc. 20th Annu.Symp.Comput. Geometry, 2004, pp. 253–262.
[11] ShilpaThakare" Novel Method for NKS Search in Multidimensional Dataset Using Advance Promish& Ranking Function" International Journal for Research in Engineering Application & Management (IJREAM) ISSN : 2494-9150 Vol-02, Issue 08, Nov 2016
[12] K.Pujitha " A SURVEY ON SEMANTIC BASED SEARCH ENGINE FOR REAL IMAGES AND WEB URL'S USING HYPERGRAPH DISTANCE MEASURE ALGORITHM" ijptonline ISSN: 0975-766X