ISSN 2229-5518

Optimal results with consumer rated keyword Search using WK-NN algorithms in Spatial database

Rakshi Iyappan, Dr.Savithri.V

Abstract: Generally, most of the objects in Spatial database are tied with various 'Indicating Text' which would describe the concern object's properties / features such as its business, products, services, features etc. As a result, the users are not limited in searching of simple queries, they are extended to complex keyword queries that present in numerous spatial input becomes more common. In spatial database there are several algorithms being proposed for Keyword search, which are limited only to work exists on searching keywords. In this work WK-NN algorithm is used to search optimum results on spatial data along with the customer rating from highest to lowest.

Index terms: keyword search,WK-NN,spatial database,keyword,customer rating.

1. Introduction

Data mining is the means of extracting data from a dataset for users to use it for various purposes. The purpose of such data plays a significant role in keyword searching. Searching is a common activity happening in data mining. Searching for spatial objects from a spatial database has recently sparked enthusiasm among researchers. This motivated to develop methods to retrieve spatial objects. Spatial objects consist of objects associated with spatial features. In other words, spatial objects involve spatial data along with longitude and latitude of the location. Querying such data is called best keyword cover querying. The search is called best keyword cover search. The existing method of such data considers either minimum inter objective distance and keyword search. As a result, new methods for best keyword cover search was developed. Traditional nearest neighbor search computes nearest neighbor by considering the distance as a feature. In this context, nearest neighbor search focus on finding nearest neighbors where keywords and spatial data plays a major impact. It comes with algorithms to answer such query.

A spatial database is a collection of data which is optimized to store and query data that represents objects defined in a geometric space. These objects represented in spatial database are called Spatial Objects. Most spatial database permits geometric objects such as points, lines and polygons. Hence spatial data are also referred as Geo database. Few spatial databases handle more complex structures such as three dimensional objects, topological coverages, linear networks, and Triangulated Irregular Network (TINs). While typical databases are designed to manage various non- spatial data such as numbers and characters, additional functionality needs to be added to databases to process spatial data types efficiently. These are called geometry. The Open Geospatial Consortium created the open Geospatial simple features [21]

The advantageous features of 'Spatial database Search' such as lateral registration, wide accessibility, easy availability of digital maps and satellite pictures (like Google maps, MS Earth, etc.) made it as the most preferred one among the present Computer generation. Each data object in spatial database represents a spatial object combined with other relevant details such as its concern, interest, Business, Service, Products etc. The main task of Optimum results with consumer rated keyword Search is to obtain desired information with

Consumer rating within desired location. For example, a person wants to dine in a Continental hotel nearest to his location. With Optimum results with consumer rated keyword Search [1]

In a larger database, the search of hidden patterns is termed as Data Mining. This kind of Data mining technique's application performed on spatial data is called Spatial Data Mining. The discovery of interesting relationship and characteristics that may exist indirectly in Spatial Database is Spatial Data Mining. Usually spatial data will be in a huge amount of data (terabytes) which would have been obtained from satellite images, Medical equipment's, video cameras, etc. Hence, it is high in cost and often impractical for users to examine it in detail. The main aim of Spatial Data Mining is Automation such as Knowledge Discovery process. Hence this plays an important role in the following functions:

Extraction of Spatial patterns and its interesting features. Identifying intrinsic relationship exists between Spatial & non spatial data Presenting data regularity concisely and at a higher conceptual level.

To reorganize spatial database to accommodate data semantics and also to achieve Better performance. Usually Spatial database stores a large amount of space related data, such as maps, preprocessed remote sensing or medical imaging data and VLSI chip layout data. It has many features distinguishing them from relational databases. They carry topological and/or distance information, usually organized by sophisticated, multi-dimensional Spatial indexing structures that are accessed by Spatial data access methods and often require spatial reasoning, geometric computation, and spatial knowledge representation techniques. [2]

Database systems use indexes to quickly find values and the way that most database index data is not optimal for spatial queries. Instead, to speed up database operations, spatial databases use a spatial index. [21]

To enhance spatial queries, spatial indices are used in spatial databases. Usual index types are not being handled spatial queries like how far two lines differ, or whether lines fall within a spatial area of interest. Common spatial index methods such as Grid (spatial index), Z-order (curve), Quad tree, Octree, UB-tree, R-tree,R+ tree,R* tree, Hilbert R-tree ,X-tree,kd-tree , m-tree , Binary space partitioning (BSP-Tree). Among the above mentioned index methods, R-tree- the typically preferred method for indexing spatial data objects (shapes, lines and points) R-trees is nothing but a tree data structure, which are being used to access spatial database for indexing Multi-dimensional information.Multidimensional information includes geographical coordinates, rectangles or polygons. In spatial database tree like data structures are used such as R-trees. [22]

2. Related work

It has been observed that the issue of identifying association rules between items in a larger database of business transactions. By introducing effective two new algorithms in place of existing algorithm would pave solution for this problem. It was evident that upon various trails with hypothetical data as well as real data, which shows that the newly introduced algorithms are more effective than the usual algorithm. This outstanding performance of newly introduced algorithms is influenced by the factors ranging from three for small problems to more than an order of magnitude for large problems. The main features of the newly introduced algorithms can be integrated into a hybrid algorithm called AprioriHybrid. The result parameters of experiment with reference to the transaction size and the number of items in

the database shows excellent scale-up properties. [5]

Earlier spatial queries (like range search, NNR etc.) were involved only on conditions on objects location properties. Rather in today's world, applications are running with objects containing both spatial properties as well as relevant texts associated with the properties of the object. For example, instead of considering all the beauty parlor for a girl, the nearest neighbor query would provide beauty parlors located nearest as well as with the customer rating. Presently appropriate result for such queries is based on IR2-tree as shown in this paper is having few drawbacks that would affect efficiency. With reference to this, a new access method called spatial inverted index has been introduced.[3]

Spatial joins are one among the very essential operations for integrating spatial objects of spatial relations. It is very important that the spatial joint should function efficiently in any given system because the execution time of the system is super linear in the number of spatial objects of the participating relations and this number of objects will be very high. In this stance, a detailed analysis carried on spatial join processing using R-trees particular tree R*-trees.The R-tree is most preferred for supporting spatial queries. Among the R-tree family members R*-tree stands as a most efficient partner. Various techniques including straightforward approach have been presented to improve its execution time with respect to CPU and I/O time. By an order of magnitude, execution time got drastically improved over the first approach with an algorithm. I/O time varies directly proportional to the time for reading each required page of the relation exactly once. By carrying out an experimental performance comparison the performance of various approaches was experimented in an Experimental performance comparison where several large data sets from real applications are used. [6]

The location identifiable keyword query returns results with rated objects whose location is approximately near to query location. These resulting objects are associated with a textual description matching to the queried keywords. This query occurs inherently in various types of mobile & conventional Web services and applications, ex.: Yellow Pages, Classified websites, Map services, etc. As per new type of query, the location identified top-k Prestige based keyword retrieval (LkPT) query is proposed which retrieves the top-k location based web objects (close location proximity) with ratings. [7]

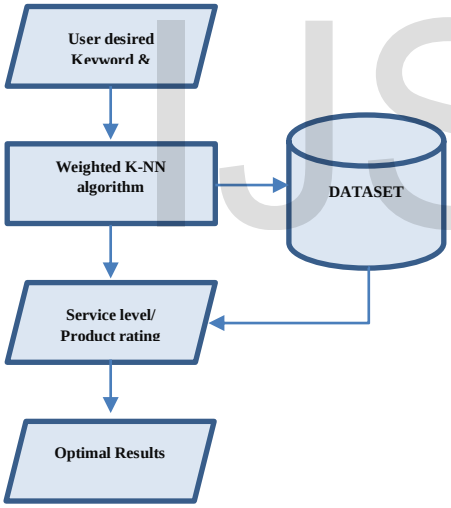
Mining-Spatial information has a way of representing thematic and location knowledge within queries. Relevant information to the query is processed through Information retrieval method such as keyword matching geographic information is given by the names of places. The location names are then converted to be saved in spatial database as Geometric coordinates. Then similarities among the queries are calculated by Euclidean distance. [9]

Initially, the system considers a collection of Points of Interest (POIs) which are linked with location information ad its textual representations. Hence, a location aware spatial keyword query is given by a set of keywords. The location aware search locates the k nearest neighbor of the query. This is in the direction of search and accommodates all the keywords as Input [10] Present spatial cum textual objects helps Users to retrieve latest objects whose location have an overlap with the area specified by the user and the texts associated with the object. Similar to this both the keywords (Location input & Keyword text) would act as filters. This system still has few drawbacks such as

User may tend to receive either few matching location-textual objects or very large number of matching results based on the input location & keyword text. The Geo-textual objects rank-ordering will reduce the problem by returning only the top-ranked objects. [11]

3. Methodology

Refer to the below mentioned flow chart the following methodology has been developed for the user desired results. This methodology will explain the functions and features of our system and its requirements. The system procures location details from the dataset, and then it provides results with predefined customer ratings and minimum inter object distance.



Input of keywords and location is being obtained first from the users. System (consists of spatial database) processes user provided input with WK-NN algorithm. This algorithm identifies a sum of independent objects which are nearest to the user required location along with matching of keywords with object associated texts[3][5]. Hence output resulted from this process is efficient in terms of desired keywords as well as objects nearest to the desired location. Output will be delivered to the user as well as other system as applicable. Effective and efficient results can be obtained through a well-defined Output-plan.

Initially, a query is entered by the user which is given to the system. The system consists of a spatial database and uses the Keyword- Nearest Neighbor Expansion algorithm. The K-NNE algorithm finds a number of independent objects, in which every object is nearer to the location of the query and the keywords associated are related to the collection of query keywords [3] [5]. An output which represents the information distinctly and meets the end user demands is efficient. The results of processing are conveyed to the users and to other system through outputs, in every system an efficient and smart output plan helps in improving the user's relationship with the system which further helps in better decision-making.

4. Conclusion

The important objective of this project is to define a system which has to deliver results with customer rating as well as results nearest to the user desired location. The above implemented baseline algorithm is inspired by the application of WK-NN algorithm which is derived from exhaustively bringing together the objects (result) as desired by the user. At the time of applying an increased number of keywords, the working of Baseline algorithm gets down drastically as a result of massive candidate keyword covers generated. To overcome this drawback a much more scalable algorithm called WK-NN algorithm is applied. This algorithm minimizes the number of candidate keyword covers generated compared with that of baseline algorithm.

REFERENCES

[1] Ke Deng, Xin Li, Jiaheng Lu, and Xiaofang Zhou, Senior Member,"Best Keyword Cover Search" ,in IEEE Jan 2015
[2] Manjula Aakunuri#, Dr.G.Narasimha*, Sudhakar Katherapaka $ "Spatial Data Mining: A Recent Survey and New Discussions," in / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (4) , 2011, 1501-1504.
[3] Yufei Tao Cheng Sheng, "Fast Nearest Neighbor Search with Keywords" IEEE IEEE Trans Knowl. Data Eng., Nov 29, 2014 .
[4] N. Beckmann, H. Kriegel, R. Schneider, and B. Seeger. The R*-tree: An efficient and robust access method for points and rectangles.In Proc. of ACM Management of Data (SIGMOD), pages 322–331, 1990.
[5] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in Proc. 20th Int. Conf. Very Large Data Bases, 1994, pp. 487–499.
[6] T. Brinkhoff, H. Kriegel, and B. Seeger, "Efficient processing of spatial joins using r-trees," in Proc. ACM SIGMOD Int. Conf. Manage.Data, 1993, pp. 237–246.
[7] X. Cao, G. Cong, and C. Jensen, "Retrieving top-k prestige-based relevant spatial web objects," Proc. VLDB Endowment, vol. 3, nos.1/2, pp. 373–384, Sep. 2010.
[8] Beckmamr N., Kriegel H.-P., Schneider R., Seeger B.: 'The J?*-tree; ,4n Efficient and Robust Access Method for Poirrtr and Rectanglts', Proc.ACM SIGMOD Int. Coaf. on Management of Data, Atlantic City, N.J., 1990, pp. 322-331.
[9] Lei, Liu, Yong Gao, Xing Lin , Xiao Guo , Haoran Li, "A Framework and Implementation for Qualitative Geographic Information Retrieval".
[10] Guoliang Li, Jianhua Feng, Jingxu , "DESKS: Direction-Aware Spatial Keyword Search", ICDE Conference 2015.
[11] Lisi Chen, Gao Cong, Xin Cao, Kian-Lee Tan, "Temporal Spatial-Keyword Top-k Publish/Subscribe", 2015, IEEE
[12] X. Cao, G. Cong, C. Jensen, and B. Ooi, "Collective spatial keyword querying," in Proc. ACM SIGMOD Int. Con.
[13] I. D. Felipe, V. Hristidis, and N. Rishe, "Keyword search on spatial databases," in Proc. IEEE 24th Int. Conf. Data Eng., 2008, pp. 656–665.
Web sites:
[21]https://en.wikipedia.org/wiki/Spatial_database
[22] https://en.wikipedia.org/wiki/R-tree