Fast Query Retrieval using High Utility Item sets from Transactional Database

[1]Ms.M.Usharani., [2]Mrs.V.Annapoorani ., MCA ., M.Phil., PH.D, [3]Mrs. M. Ranjani., MCA., M.PHIL.,

[1]Research scholar, [2]Assistant professor, [3]Assistant professor,
Department of computer science, Pavai Arts and Science College for women

ABSTRACT

Now-a-days the large amount of data is stored in the databases. A spatial database is used to access multidimensional objects like points, rectangles, etc. Spatial predicate, and a predicate on their associated texts are need to be satisfied for finding the required objects with fast access. For example, instead of finding all the restaurants, object grouping method is used to find the restaurants that are closest among those whose menus contain the specified keywords. IR^2-tree is used in the existing system for providing best solution for finding nearest neighbour. This method has few deficiencies. So we implement the new method called spatial inverted index with object grouping to improve the space and query efficiency. And priority level search is used to search the objects based on the users priority. Thus the proposed algorithm is scalable to find the required objects using object grouping.

1. INTRODUCTION

Spatial data mining is a special kind of data mining. The main difference between data mining and spatial data mining is that in spatial data mining tasks we use not only non-spatial attributes (as it is usual in datamining in non-spatial data), but also spatial attributes.Spatial data mining is the application of data mining methods to spatial data.

The objective of spatial data mining is to find patterns in data with respect to geography. So far, data mining and Geographic Information Systems (GIS) have existed as two separate technologies, each with its own methods, traditions, and approaches to visualization and data analysis. The immense explosion in geographically referenced data occasioned by developments in IT, digital mapping, remote sensing, and the global diffusion of GIS emphasizes the importance of developing data-driven inductive approaches to geographical analysis and modeling.

2. BACKGROUNT WORK

2.1 IR^2-TREE

The older system IR^2-Tree follows the two kinds of strategies

- R trees,
- Signature files.

The R tree strategy wants the more no of keywords to search the user specification. The signature files are loading the more no text to matching the object for user specification. Here we discuss the drawback of IR^2 trees where it has the advantages of both R trees and signature files. The IR^2 trees does not contains the all the query keywords. It will direct the search to some objects those does not contain all keywords.

Signature file in general refers to a hashing-based framework, whose instantiation in is known as superimposed coding (SC), which is shown to be more effective than other instantiations. It is designed to perform membership tests: determine whether a query word w exists in a set W of words. SC is conservative, in the sense that if it says "no", then w is definitely not in W. If, on the other hand, SC returns "yes", the true answer can be either way, in which case the whole W must be scanned to avoid a false hit.

word	*hashed bit string*
a	00101
b	01001
c	00011
d	00110
e	10010

Figure 3.1 Example of bit string computation

For example, in thebit string h(a) of a, the 3rd and 5th (counting from left)bits are set to 1. As mentioned earlier, the bit signatureof a set W of words simply ORs the bit strings of all themembers of W. For instance, the signature of a set {a, b}equals 01101, while that of {b, d} equals 01111. Given a query keyword w, SC performs the membershiptest in W by checking whether all the 1's of h(w)appear at the same positions in the signature of W. If not,it is guaranteed that w cannot belong to W. Otherwise,the test cannot be resolved using only the signature, anda scan of W follows. A false hit occurs if the scan revealsthat W actually does not contain w.

3. METHOD

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 modeling 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.

Today, the widespread use of search engines has made it realistic to write spatial queries in a brand new way. Conventionally, queries focus on objects' geometric properties only, such as whether a point is in a rectangle, or how close two points are from each other. We have seen some modern applications that call for the ability to select objects based on both of their geometric coordinates and their associated texts. For example, it would be fairly useful if a search engine can be used to find the nearest restaurant that offers "steak, spaghetti, and brandy" all at the same time. Note that this is not the "globally" nearest restaurant (which would have been returned by a traditional nearest neighbour query), but the nearest restaurant among only those providing all the demanded foods and drinks. 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 menus 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. 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. This access method successfully incorporates point coordinates into a conventional inverted index with small extra space, owing to a delicate compact storage scheme.

3.1 SI-INDEX

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. Object grouping is used to group the objects and finding the required object those containing specified keywords. It contains the set of points and the points are related to the set of keywords and the keywords are related to derive the set of documents. Here using the new concepts of the merge list and the distance alignment to retrieve the documents from the user requirement.

3.2 PRIORITY LEVEL SEARCH

The enhanced search is used to effectively search the data's based on the priority level. By using this search we get the data's depending upon the user priority levels. It means user enter the key words for search data's at the same time we include the priority box.After entering the priority the database ranking the documents to display the required documents.

CONCLUSION

The spatial inverted index list with object grouping and priority level search is proposed. The spatial inverted index is using both capacity of the R-Tree and the processing of signature files. Compared with the previous work the existing systems are not efficient to provide the real time answers. In this implemented work, the proposed concept of the list merging and distance alignment are used to help for searching, and the object grouping method is used to provide the effectiveness of the quick search. The priority level search is used for finding the objects based on the users priority level.

REFERENCES

1. Yufei Tao and Cheng Sheng (2013) 'Fast NearestNeighbour Search with Keywords', In Proc. Of IEEE Transactions On Knowledge And Data Engineering.
2. Agrawal, S. and Chaudhuri, S. and Das, G. (2002) 'Dbxplorer: A system for keyword-based search over relational databases' , In Proc. of International Conference on Data Engineering (ICDE), pages 5–16.
3. Anandhi R J andNatarajan and Subramanyam (2009) ' Efficient Consensus Function for Spatial Cluster Ensembles: An Heuristic Layered Approach', International Symposium on Computing, Communication, and Control (ISCCC).
4. Bhalotia, G. and Hulgeri, A. Nakhe, C. and Chakrabarti, S. andSudarshan, S. (2002) 'Keyword searching and browsing in databases using banks', In Proc. of International Conference on Data Engineering (ICDE), pages 431–440.
5. Cao, X. and Chen, L.and Cong, G. and Jensen, C. S. andQu, Q. and Skovsgaard, A. and Wu, D. andYiu, M. L (2012) 'Spatial keyword querying', In ER, pages 16–29.
6. Cao, X. and Cong, G. and Jensen, C. S. (2010) 'Retrieving top-k prestige-based relevant spatial web objects', PVLDB, 3(1):373–384.
7. Cao, X. and Cong, G. and Jensen, C. S. and Ooi, B. C. (2011) 'Collective spatial keyword querying', In Proc. of ACM Management of Data (SIGMOD), pages 373–384.
8. Chazelle, B. and Kilian, J. and Rubinfeld, R. and Tal, A.(2004) '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.
9. Chen, Y. and Suel, T. and Markowetz, A.(2006) 'Efficient query processing in geographic web search engines', In Proc. of ACM Management of Data (SIGMOD), pages 277–288.
10. Chu, E. and Baid, A. and Chai, X. and Doan, A. andNaughton, J.(2009) 'Combining keyword search and forms for ad hoc querying of databases', In Proc. of ACM Management of Data (SIGMOD).
11. Cong, G. and Jensen, C. S. and Wu, D.(2009) 'Efficient retrieval of the top-k most relevant spatial web objects' , PVLDB, 2(1):337–348.
12. Debing Zhang and Genmao Yang and Yao Hu and Zhongming Jin and Deng Cai and Xiaofei He (2013), 'A Unified Approximate Nearest Neighbor Search Scheme by Combining Data Structure and Hashing' , Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence.
13. Hariharan, R. and Hore, B. and Li, C. and Mehrotra, S.(2007) 'Processing spatial keyword.(SK) queries in geographic information retrieval (GIR) systems', In Proc. of Scientific and Statistical Database Management (SSDBM).
14. Hjaltason, G. R. and Samet, H.(1999) 'Distance browsing in spatial databases' ,In proc. of ACM Transactions on Database Systems (TODS), 24(2):265–318.
15. Haibo Hu, DikLun Lee1, and JianliangXu(2006) 'Fast Nearest Neighbor Search on Road Networks' , Research Grants Council, Hong Kong SAR under grant HKUST6277/04E.
16. Kamel, I. and Faloutsos, C. (1994) 'Hilbert R-tree: An improved r-tree using fractals', In Proc. of Very Large Data Bases (VLDB), pages 500–509.
17. Wei Wang and JiongYangand and Richard Muntz (2000), 'An Approach To Active Spatial Data mining Based on Statistical Information', IEEE Transactions on Knowledge and Data Engineering, vol. 12, no. 5.
18. Yufei Tao, Jun Zhang, DimitrisPapadias, and Nikos Mamoulis (2004) 'An Efficient Cost Model for Optimization of Nearest Neighbor Search in Low and Medium Dimensional Spaces' In Proc. Of IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 10.
19. Yoonho Hwang and Bohyung Han and Hee-KapAhn (2012) 'A Fast Nearest Neighbor search algorithm by nonlinearEmbedding', www.postech.ac.kr/ bhhan/papers/ cvpr2012_fnn.pdf Online.