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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

A Survey on Keyword Cover Search Method

Mehjbeen N Tamboli, Prof.Suvarna Pawar

M.E. Student, Dept. of Computer Engineering, TCOER, Pune, Maharashtra, India

Assistant Professor, Dept. of Information Technology, TCOER, Pune, Maharashtra, India

ABSTRACT: Spatial databases are stores the information about the spatial objects which are associated with the keywords to indicate the information such as its business/services/features. Very important problem known as closest keywords search is to query objects, called keyword cover. In closest keyword search, it covers a set of query keywords and minimum distance between objects. From last few years, keyword rating increases its availability and importance in object evaluation for the decision making. This is the main reason for developing this new algorithm called Best keyword cover which is considers inter-distance as well as the rating provided by the customers through the online business review sites. Closest keyword search algorithm combines the objects from different query keywords to generate candidate keyword covers. Baseline algorithm and keyword nearest neighbor expansion algorithms are used to find the best keyword cover. The performance of the closest keyword algorithm drops dramatically, when the number of query keyword increases. To solve this problem of the existing algorithm, this work proposes generic version called keyword nearest neighbour expansion which reduces the resulted candidate keyword covers.

KEYWORDS: Spatial database, point of interests, keywords, keyword rating, keyword cover.

I. INTRODUCTION

Now a days, use of mobile computing increases[5]. Inspired by the mobile computing, the spatial keywords search problem has attracted much attention recently because of location-based services and wide availability of extensive digital maps and satellite imagery. So the number of users using the location based services has been also increased to large extend. Spatial objects indicates the information such as its business/services/features which are associated to keyword(s)[7][8]. In spatial database, each tuple represents a spatial object. The main idea behind the spatial keywords search is to identify spatial object(s) which are associated with keywords relevant to a set of query keywords which are close to each other and/or close to the query location. This problem has unique value in variousapplications because users' requirements are often expressed as multiple keywords. In existing, spatial keyword search problem have been studied because of the value of the special keyword search in practice. This paper investigates a generic version of mCK query, called Best Keyword Cover (BKC) query, which considers inter-objects distance as well as keyword rating. It is motivated by the observation of increasing availability and importance of keyword rating in decision making. Millions of businesses/services/features around the world have been rated by customers through online business review sites such as Yelp, Citysearch, ZAGAT and Dianping, etc. For example, a restaurant is rated 65 out of 100 (ZAGAT.com) and a hotel is rated 3.9 out of 5 (hotels.com). According to a survey in 2013 conducted by Dimensional Research (dimensionalresearch.com), an overwhelming 90 percent of respondents claimed that buying decisions are influenced by online business review/rating. Due to the consideration of keyword rating, the solution of BKC query can be very different from that of mCK query).

This work develops two BKC query processing algorithms, baseline and keyword-NNE. The baseline algorithm is inspired by the mCK query processing methods Both the baseline algorithm and keyword-NNE algorithm are supported by indexing the objects with an R*-tree like index, called KRR*-tree. In the baseline algorithm, the idea is to combine nodes in higher hierarchical levels of KRR*-trees to generate candidate keyword covers. Then, the most promising candidate is assessed in priority by combining their child nodes to generate new candidates. Even though BKC query can be effectively resolved, when the number of query keywords increases, the performance drops dramatically as a result of massive candidate keyword covers generated. To overcome this critical drawback, we developed much scalable keyword nearest neighbour expansion (keyword-NNE) algorithm which applies a different strategy. Keyword-NNE selects one query keyword as principal query keyword. The objects associated with the principal query keyword are principal objects. For each principal object, the local best solution (known as local best

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

keyword cover (lbkc)) is computed. Among them, the lbkc with the highest evaluation is the solution of BKC query. Given a principal object, its lbkc can be identified by simply retrieving a few nearby and highly rated objects in each non-principal query keyword (two-four objects in average as illustrated in experiments). Compared to the baseline algorithm, the number of candidate keyword covers generated in keyword-NNE algorithm is significantly reduced. The in-depth analysis reveals that the number of candidate keyword covers further processed in keyword-NNE algorithm is optimal, and each keyword candidate cover processing generates much less new candidate keyword covers than that in the baseline algorithm.

II. RELATED WORK

A Now a days, mobile users widely uses the Location-Based Services (LBS). Location-based systems can work efficient if user enters the complete keyword, otherwise it was shown incomplete output. It is difficult to enter the complete keyword on mobile devices for getting the correct relative result. To avoid this problem, proposed system studied in [1] about the location-aware search. It can be returned search answers as the user enters in queries letter by letter. In this paper, the main challenge is to provide the relevant answers speedily. Author uses a new index structure, prefix-region tree (called PR-tree), that can be support to provide the speedily results to the users. PR-Tree is a tree-based index structure which seamlessly integrates the textual description and spatial information to indexthe spatial data. Using the PRTree, authors develop efficient algorithms to support single prefix queries and multi-keyword queries. Experiments show that our method achieves high performance and significantly outperforms state-of-the-art methods.

The location-aware keyword query returns ranked objects that are near a query location and that have textual descriptions that match query keywords[2][3]. There are many mobile applications and traditional services uses this type of query, e.g. Yellow pages and Maps services. In previous work, ranked query returns independent potential results. Ranking is very important in decision making. However, a relevant result object with nearby objects that are also relevant to the query is likely to be preferable over a relevant object without relevant nearby objects. The paper proposes [2] the concept of prestige-based relevance to capture both the textual relevance of an object to a query and the effects of nearby objects. Based on this, a new type of query, the Location-aware top-k Prestige-based Text retrieval (LkPT) query, is proposed that retrieves the top-k spatial web objects ranked according to both prestige-based relevance and location proximity. They propose two algorithms that compute LkPT queries. Empirical studies with real-world spatial data demonstrate that LkPT queries are more effective in retrieving web objects than a previous approach that does not consider the effects of nearby objects; and they show that the proposed algorithms are scalable and outperform a baseline approach significantly.

There are lots of applications that finds the objects nearest to the specified location which contains a set of keywords[5][7][8]. Yellow pages required address and a set of keywords to get the results. Yellow pages returns a list of business/features/services whose description contains entered keywords, ordered by their inter-object distance from the specified location. In this paper [3], author studied problems of nearest neighbor search on location data and keyword search on text data separately. There is no any method that returns answer to the best for spatial and keyword queries which is related to the same. In this paper, author proposed an efficient algorithm that returns top-k spatial keyword queries. Proposed system introduces indexing structure called Information retrieval R-Tree which is combination of R-tree with superimposed text signatures. Algorithm returns the answer from IR2-tree which construct and maintain by the algorithm to the keyword queries. Proposed algorithms are superior performance and excellent scalability to the previous work experimentally.

Geographic search engine returns documents which are very close textually and spatially to the query keywords. Retrieved documents are ranked according to their joint textual and spatial relevance to the entered query. Existing indexing scheme inefficient in answering spatial queries because of lacking in index simultaneously handle both the textual and location aspect. In this proposed system [4], author proposes new index called IR-tree that combines with a top-k document facilitates four major tasks in document searches, textual filtering, spatial filtering, relevance computation, and document ranking. These four tasks are used in this algorithm in a fully integrated manner. Also, this algorithm adopt different ranks on textual and spatial relevance of documents at the run time. Hence, it can be used in variety of applications. Experimentally a set of comprehensive experiments over a wide range of scenarios has been conducted and the experiment results demonstrate that IR-tree outperforms the state-of-the-art approaches for geographic document searches.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

The aim is always to find out multiple objects which altogether covers the query keyword.this is because no one object individually satisfies all query keywords [7][8].the approach studied in [7][9] is similar to the mKC queries.The m closest keyword cover (mKC) aims to find the objects which are close to each other or with minimum distance.In proposed system we are going to consider not only the iner object distance bu as well as the keyword rating of the object.

III. PROPOSED ALGORITHM

A. *Keyword-NNE:*

In previous work, BKC algorithm drops its performance when the number of query keywords are increases. To solve this problem, here developed a more efficient keyword nearest neighbour expansion (keyword-NNE) which uses the different strategy. In this algorithm, one query is considered as a principal query keyword. Those objects are associated with principal query keyword are considered as principal objects. Keyword-NNE computes local best solution for each principal object. BKC algorithm returns the lbkc with having highest evaluation. For each principal object, its lbkc can be simply selects few nearby and highly rated objects by the viewer/customer. Compared with the baseline algorithm, the keyword covers significantly reduced. These keyword covers further processed in keyword-NNE algorithm that will be optimal, and each keyword candidate cover processing generates very less new candidate keyword covers.

B. *Preliminary:*

In spatial database, each object present in database may be associated with either one or multiple keywords. In this object with multiple keywords are directly transformed to multiple objects located at the same location without loss of generality. These objects are in the form of <id, x, y, keyword, rating> where location of the objects in two dimensional geographical space represented by x and y.

Definition 1 (Diameter): Let O be a set of objects {o1,…….,on}. For oi; oj € O, dist(oi, oj) is the euclidean distance between oi, oj in the two-dimensional geographical space. The diameter of O is

Diam(O)=max dist(oi, oj). eq.(1)

Each objects has its score with respect to diameter of object and keyword rating of objects in O. Interest of the user may be different in keyword ratings of the objects.

Definition 2 (keyword Cover): Let T be a set of keywords {k1, . . . , kn} and O a set of objects {o1, . . . , on} O is a keyword cover of T if one object in O is associated with one and only one keyword in T.

Definition 3 (Best Keyword Cover Query): Given a spatial database D and a set of query keywords T, BKC query returns a keyword cover O of T (O subset D) such that O.score ≥ O'.score for any keyword cover O’ of T (O’ subset D).

In keyword-NNE algorithm, instead of individually processing principal objects are processed in blocks. Suppose k be the principal query keyword. KRR*k-tree used for indexing principal objects. Given principal node Nk in KRR*k-tree, and lbkcNk consider as local keyword cover of Nk, that consists of Nk and other corresponding nodes of Nk in each non-principal query keyword.

IV. CONCLUSION AND FUTURE WORK

The proposed system is provides more sensible decision making than the mCK query. Baseline algorithm which is inspired by the mCK query. The main problem of baseline algorithm is that it reduces the performance when number of query keywords increases. Keyword-NNE algorithm applies a different strategy that searches the best solution in query keyword for each object. It reduces the generated candidate keyword covers. Baseline keyword covers are passed to keyword-NNE algorithm for further processing which is optimal and generates less new candidate keyword covers than the baseline algorithm.

REFERENCES

[1] Ruicheng Zhong, Ju Fan, Guoliang Li, Kian-Lee Tan and Lizhu Zhou, ‘Location-Aware Instant Search’ CIKM’12, October 29–November 2, 2012, Maui, HI, USA

[2] Xin Caoy, Gao Cong, Christian S. Jensenz, ’ Retrieving Top-k Prestige Based Relevant Spatial Web Objects’, Proceedings of the VLDB Endowment, Vol. 3, No.1 , 2010.

[3] Ian De Felipe, Vagelis Hristidis, Naphtali Rishe, ’ Keyword Search on Spatial Databases’, in Proc. IEEE 24th Int. Conf. Data Eng., 2008, pp. 656–665.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

[4] Z. Li, K. C. Lee, B. Zheng, W.-C. Lee, D. Lee, and X. Wang,' IRTree: An efficient index for geographic document search', IEEE Trans. Knowl. Data Eng., vol. 99, no. 4, pp. 585–599, Apr. 2010.
[5] G. Cong, C. Jensen, and D. Wu,' Efficient retrieval of the top-k most relevant spatial web objects', Proc. VLDB Endowment, vol. 2, no. 1, pp. 337–348, Aug. 2009.
[6] D.Amutha Priya, Dr. T.Manigandan,' Fast Accurate Mining on Spatial Database Using Keywords', International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 3, Special Issue-1, May 2015.
[7] Ke Deng, Xin Li, Jiaheng Lu, and Xiaofang Zhou ,' Best Keyword Cover Search', IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 1, JANUARY 2015.
[8] X .cao, G Cong, C Jensen "COLLECTIVE SPATIAL KEYWORD QUERYING' IN Proc.ACM SIGMOD Int. Conf manage. Data , 2011 ,pp. 373-384.
[9] J .fan , G Li , L . Zhou, S Chen , and J. Hu Seal : "SPATIO- TEXTUAL SIMILARITY SEARCH ". PVLDB, 5(9): 824-835, 2012.
[10] W Huang , G .Li , K-L. tan and J Feng. "EFFICIENT SAFE-REGION CONSTRUCTION FOR MOVING TOP-K SPATIAL KEYWORD QUERIES" in CIKM 2012.

BIOGRAPHY

Ms.Mehjbeen N Tamboliis ME, Student at Computer Dept. Trinity College of Engineering,Pune, and Maharashtra, India.

Prof. Suvarna Pawar is an Assistant Professor at, Department of Computer Engineering, Trinity college of Engineering , Pune, Maharashtra, India.