International Journal of Research
Available at
https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 10
September 2017

Adaptive Keyword Cover Search Algorithm

P. Satyanarayana & Mrs. M.Swapna

[1,2]Aurora's Scientific Technological and Research Academy, Bandlaguda, Hyderabad – 500005.

narayanapaul@gmail.com & swapna0801@gmail.com

Abstract:

It is common that the objects in a spatial database (e.g., restaurants/hotels) are associated with keyword(s) to indicate their businesses/ services/features. An interesting problem known as Closest Keywords search is to query objects, called keyword cover, which together cover a set of query keywords and have the minimum inter-objects distance. In recent years, we observe the increasing availability and importance of keyword rating in object evaluation for the better decision making. This motivates us to investigate a generic version of Closest Keywords search called Best Keyword Cover which considers inter-objects distance as well as the keyword rating of objects. The baseline algorithm is inspired by the methods of Closest Keywords search which is based on exhaustively combining objects from different query keywords to generate candidate keyword covers. When the number of query keywords increases, the performance of the baseline algorithm drops dramatically as a result of massive candidate keyword covers generated. To attack this drawback, this work proposes a much more scalable algorithm called keyword nearest neighbor expansion (keyword-NNE). Compared to the baseline algorithm, keyword-NNE algorithm significantly reduces the number of candidate keyword covers generated. The in-depth analysis and extensive experiments on real data sets have justified the superiority of our keyword-NNE algorithm. .

Keywords

Spatial database, Point of Interests, Keywords, Keyword Rating, Keyword Cover..

1. Introduction

Driven by mobile computing, location-based services and wide availability of extensive digital maps and satellite imagery (e.g., Google Maps and Microsoft Virtual Earth services), the spatial keywords search problem has attracted much attention recently. In a spatial database, each tuple represents a spatial object which is associated with keyword(s) to indicate the information such as its businesses/services/features. Given a set of query keywords, an essential task of spatial keywords search is to identify spatial object(s) which are associated with keywords relevant to a set of query keywords, and have desirable spatial relationships (e.g., close to each other and/or close to a query location).

This problem has unique value in various applications because users' requirements are often expressed as multiple keywords. For example, a tourist who plans to visit a city may have particular shopping, dining and accommodation needs. It is desirable that all these needs can be satisfied without long distance travelling. Due to the remarkable value in practice, several variants of spatial keyword search problem have been studied.

The works aim to find a number of individual objects, each of which is close to a query location and the associated keywords (or called document) are very relevant to a set of query keywords (or called query document). The document similarity is applied to measure the relevance between two sets of keywords. Since it is likely none of individual objects is associated with all query keywords, this motivates the studies to retrieve multiple objects, called keyword cover, which together cover (i.e., associated with) all query keywords and are close to each other. This problem is known as m Closest Keywords (mCK) query in.

The problem studied in additionally requires the retrieved objects close to a query location. It 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.

2. Design Constraints

Some existing works focus on retrieving individual objects by specifying a query consisting of a query location and a set of query keywords (or known as document in some context). Each retrieved object is associated with keywords relevant to the query keywords and is close to the query location. The approaches proposed by Cong et al. and Li et al.

employ a hybrid index that augments nodes in non-leaf nodes of an R/R*-tree with inverted indexes. In virtual bR*-tree based method, an R*-tree is used to index locations of objects and an inverted index is used to label the leaf nodes in the R*-tree associated with each keyword. Since only leaf nodes have keyword information the m Closest Keywords (mCK) query is processed by browsing index bottom-up.

This motivates us to investigate a generic version of Closest Keywords search called Best Keyword Cover which considers inter-objects distance as well as the keyword rating of objects. The baseline algorithm is inspired by the methods of Closest Keywords search which is based on exhaustively combining objects from different query keywords to generate candidate keyword covers. When the number of query keywords increases, the performance of the baseline algorithm drops dramatically as a result of massive candidate keyword covers generated.

Compared to the most relevant mCK query, BKC query provides an additional dimension to support more sensible decision making. The introduced baseline algorithm is inspired by the methods for processing mCK query. The baseline algorithm generates a large number of candidate keyword covers which leads to dramatic performance drop when more query keywords are given. The proposed keyword-NNE algorithm applies a different processing strategy, i.e., searching local best solution for each object in a certain query keyword. As a consequence, the number of candidate keyword covers generated is significantly reduced. The analysis reveals that the number of candidate keyword covers which need to be further processed in keyword-NNE algorithm is optimal and processing each keyword candidate cover typically generates much less new candidate keyword covers in keyword-NNE algorithm than in the baseline algorithm.

3. Design and Implementation

A. *Existing system*

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.

Disadvantages

- Without query coordinates, it is difficult to adapt existing techniques to our problem.
- A simple reduction that treats the coordinates of each data point as possible query coordinates suffers poor scalability.

We develop a novel index structure based on random projection with hashing. Unlike tree-like indexes adopted in existing works, our index is less sensitive to the increase of dimensions and scales well with multi-dimensional data.

B. *Proposed System*

To overcome these critical drawbacks in existing system, we developed much scalable keyword nearest neighbor 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 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.

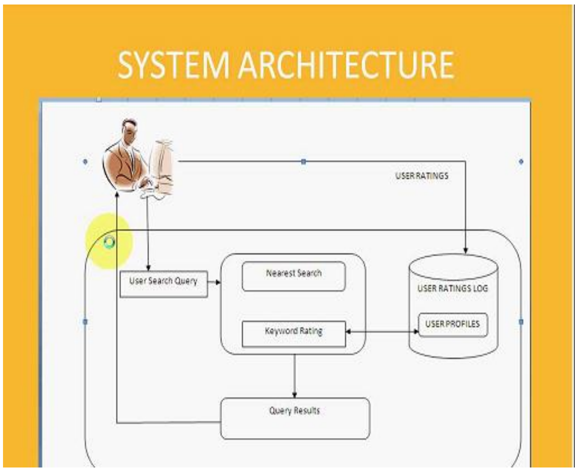

Fig 1: System Architecture

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.

Advantages of Proposed System:

1. We can reduce the number of candidate keyword covers.

Modules

1. Keyword-NNE algorithm
2. Baseline Algorithm

Module Description:

Keyword-Nearest Neighbor Expansion (NNE):

Keyword-NNE algorithm applies a different processing strategy, i.e., searching local best solution for each object in a certain query keyword. As a consequence, the number of candidate keyword covers generated is significantly reduced.

Baseline Algorithm:

The baseline algorithm is inspired by the methods of Closest Keywords search which is based on exhaustively combining objects from different query keywords to generate candidate keyword covers.

4. Conclusion and Future Scope

Compared to the most relevant mCK query, BKC query provides an additional dimension to support more sensible decision making. The introduced baseline algorithm is inspired by the methods for processing mCK query. The baseline algorithm generates a large number of candidate keyword covers which leads to dramatic performance drop when more query keywords are given. The proposed keyword-NNE algorithm applies a different processing strategy, i.e., searching local best solution for each object in a certain query keyword. As a consequence, the number of candidate keyword covers generated is significantly reduced. The analysis reveals that the number of candidate keyword covers which need to be further processed in keyword-NNE algorithm is optimal and processing each keyword candidate cover typically generates much less new candidate keyword covers in keyword-NNE algorithm than in the baseline algorithm.

The proposed system provides a flexible approach and a very sensible decision making than the existing approach. The bKC query provides the result on the basis of not only the inter-object distance but also with the keyword rating of that object. The keyword rating of the object is provided by the user on his personal experience while using the system. So as the keyword rating is important in decision making this approach gives the optimized result than the mCK query given in existing approach. The KNN algorithm provides optimized approach for the system in which the generated candidate cover set is minimized. The future work with this system is adding the concept of personalized search. The personalized search is gaining popularity due to its benefits. So the use of personalized search will increase the flexibility of the system. The future work is to provide the methods which automatically provide the methods for detecting the keyword rating than provided by the user.

5. 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.

[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.