Information Processing Letters 118 (2017) 83–90

Contents lists available at ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

Group-based collective keyword querying in road networks

Sen Su [a,*], Sen Zhao [a], Xiang Cheng [a], Rong Bi [a], Xin Cao [b], Jie Wang [c]

[a] Beijing University of Posts and Telecommunications, Beijing, China
[b] Queen' University Belfast, UK
[c] University of Massachusetts Lowell, USA

ARTICLE INFO

Article history:
Received 6 March 2016
Received in revised form 31 July 2016
Accepted 25 October 2016
Available online 31 October 2016
Communicated by Tsan-sheng Hsu

Keywords:
Design of algorithms
Collective keyword query
Road networks
Approximate algorithms

ABSTRACT

This paper addresses a group-based collective keyword (GBCK) query problem in road networks. We model the road network as an undirected graph, where each node locating in a two-dimensional space represents a road intersection or a Points of Interest (POI), and each edge with weight represents a road segment. The GBCK query aims to find a region containing a set of POIs that covers all the query keywords and these POIs are close to the group of users and are close to each other. We show the problem of answering the GBCK query is NP-hard. To solve this problem, we develop a series of query processing algorithms. We first propose an efficient algorithm, which gives a 5-factor approximation to find a feasible region. The cost of this region is further used to limit the search space in the other algorithms. We then propose an exact algorithm, and an approximate algorithm with a $\frac{15}{7}$-factor approximation. Extensive performance studies using real datasets confirm the efficiency and accuracy of the proposed algorithms.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

With the popularization of GPS-enabled devices there is an increasing interest for location-based queries. Recently, new location-based queries need to target not only individual users but also user groups. For example, consider three users at different locations want to have dinner at the nearest restaurant to them. The processing of such query requires taking into account the location of each user. Many studies considering multiple query points for a group of users (e.g., [6,10,12,14]) have received much attention.

In some cases, a group of users may want to find a region containing multiple objects that can satisfy their requirements collectively instead of a single object. Consider some close friends living at different places who intend to have a good relaxing time together. They are going to spend some time at a park, and do shopping next and then go to a restaurant for dinner. The required region should meet the following needs: (1) the region must contain a park, a shopping mall, and a restaurant; (2) the region is close to all of them such that none of them need to travel a long distance (i.e., the maximum distance any user needs to travel should be minimized); (3) the park, shopping mall, and restaurant in the region are close to each other such that they can visit these locations conveniently. Inspired by this, we introduce a new type of query called **g**roup-**b**ased **c**ollective **k**eyword (GBCK) query, to enable a group of users to find a compact region which can cover all the required keywords around themselves in a road network.

Specifically, a GBCK query is defined over a road network with Points of Interest (POIs) which is modeled as an undirected graph G, where each node locating in a two-dimensional space represents a road intersection or a POI attached with a set of keywords, and each weighted edge

[*] Corresponding author.
E-mail addresses: susen@bupt.edu.cn (S. Su), zhaosen@bupt.edu.cn (S. Zhao), chengxiang@bupt.edu.cn (X. Cheng), birong@bupt.edu.cn (R. Bi), x.cao@qub.ac.uk (X. Cao), wang@cs.uml.edu (J. Wang).

http://dx.doi.org/10.1016/j.ipl.2016.10.008
0020-0190/© 2016 Elsevier B.V. All rights reserved.

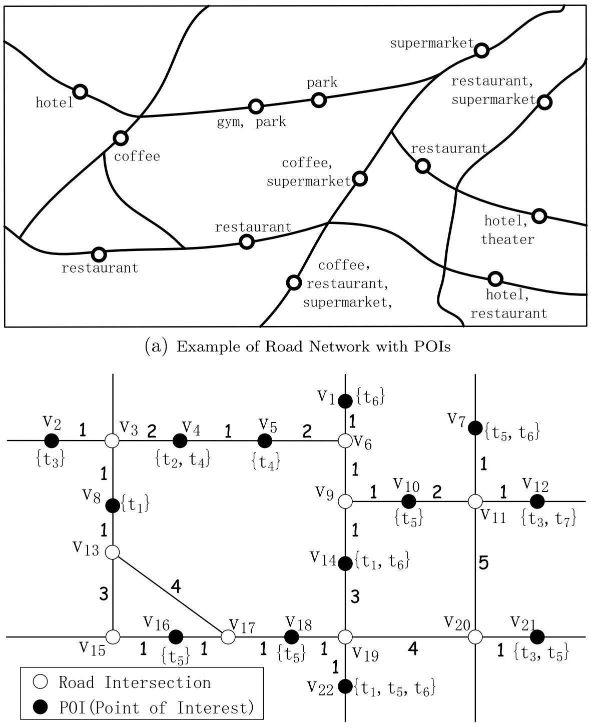
represents a road segment. The input of the query consists of the following three parameters: G, $\mathcal{U}$ and Ψ, where G is the input road network graph, $\mathcal{U}$ is the group of query nodes in G, and Ψ is a set of query keywords. The query returns a set of POIs that covers all the keywords in Ψ, and the cost of these POIs is minimized. In particular, the cost function consists of two parts: (1) the distance cost between the POI nodes and $\mathcal{U}$ (referred to as query distance cost); (2) the distance cost between the POI nodes (referred to as region diameter cost). To the best of our knowledge, none of existing studies in the literature is applicable to the GBCK query.

We show the problem of answering GBCK query is NP-hard. To solve this problem, we first abstract three basic query operations, implemented using the distance oracle based index [15]. Based on these three basic query operations, we design a series of query processing algorithms. Specifically, we first propose an efficient algorithm which gives a 5-factor approximation to find a first feasible region. The cost of this region is considered as an upper bound, and can be used to limit the search space in other algorithms, including an exact algorithm and an approximation algorithm. The exact algorithm is proposed based on the cost function consisting of two parts, which can progressively prune the search space by updating the upper bound with the cost of current best region. To achieve better efficiency, we also propose an approximation algorithm, which gives a $\frac{15}{7}$-factor approximation. We carry out extensive experiments on real datasets: the Florida road network dataset.[1] The experimental results on the datasets show that the proposed algorithms are efficient and scalable.

The contributions of this paper are summarized as follows:

1. We define the group-based collective keyword (GBCK) query problem, and prove that this problem is NP-hard.
2. We present an exact algorithm and an approximation algorithm for answering the GBCK queries. Specifically, the approximation algorithm gives a $\frac{15}{7}$-factor approximation.
3. Extensive experimental results on real data-sets show that the proposed algorithms offer scalability and are capable of excellent performance.

The rest of paper is organized as follows: We formally define the GBCK query and show that the problem of answering GBCK query is NP-hard in Section 2. We then present three basic query operations in Section 3. In Section 4 we present an efficient algorithm for finding the first feasible region, and in Section 5 we propose an exact algorithm and an approximation algorithm for answering the GBCK query. In Section 6 we report on the empirical studies. Finally, we discuss the related work in Section 7 and conclude the paper in Section 8.



(a) Example of Road Network with POIs

(b) Road Network Graph G

Fig. 1. Road network with POIs.

2. Problem statement

We first formally define the group-based collective keyword (GBCK) query problem, and then show the hardness of it.

Definition 1 *(Road network graph).* A road network with POIs (as shown in Fig. 1(a)) is modeled as a road network graph $G = (V, E)$ (as shown in Fig. 1(b)) which is an undirected graph and consists of a set of nodes V and a set of edges $E \subseteq V \times V$. Each node $v \in V$, which locates in a two-dimensional space, represents a road intersection or a POI. If the node is a POI, it is also attached with a set of keywords $v.k$ as its description. Each edge $e \in E$ represents a road segment between two nodes in V, and the edge from v_i to v_j is represented by $e(v_i, v_j)$ which is associated with a weight $e.w$ to show its travel distance. The definition of our model is similar to those used in [3,13,17].

Definition 2 *(Distance cost between a pair of nodes).* Given a pair of Nodes v_i and v_j, the distance cost between them is defined as the shortest distance. Suppose $(e_1, e_2, \cdots, e_n)$ is the shortest path between v_i and v_j, the distance cost between these two nodes can be computed by

$$Dist(v_i, v_j) = \sum_{k=1}^{n} e_k.w.$$

Definition 3 *(Region and its diameter).* A region in G consists of a subset of POI nodes and the shortest paths between

[1] http://www.dis.uniroma1.it/challenge9/download.shtml.

Download English Version:

https://daneshyari.com/en/article/4950926

Download Persian Version:

https://daneshyari.com/article/4950926

Daneshyari.com