Information Sciences 310 (2015) 52–68

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Fast optimal aggregate point search for a merged set on road networks ☆

Weiwei Sun [a], Chong Chen [a,*], Baihua Zheng [b], Chunan Chen [a], Liang Zhu [a], Weimo Liu [a], Yan Huang [c]

[a] Shanghai Key Laboratory of Data Science, School of Computer Science, Fudan University, Shanghai, China
[b] School of Information Systems, Singapore Management University, Singapore
[c] Computer Science and Engineering Department, University of North Texas, USA

ARTICLE INFO

Article history:
Received 3 September 2013
Received in revised form 15 February 2015
Accepted 11 March 2015
Available online 19 March 2015

Keywords:
Query processing
Aggregate nearest neighbor
Road networks
Spatial databases

ABSTRACT

Aggregate nearest neighbor query, which returns an optimal target point that minimizes the aggregate distance for a given query point set, is one of the most important operations in spatial databases and their application domains. This paper addresses the problem of finding the aggregate nearest neighbor for a merged set that consists of the given query point set and multiple points needed to be selected from a candidate set, which we name as merged aggregate nearest neighbor(MANN) query. This paper proposes two algorithms to process MANN query on road networks when aggregate function is max. Then, we extend the algorithms to support other aggregate functions (e.g., sum). Extensive experiments are conducted to examine the behaviors of the solutions in terms of five parameters affecting the performance. The overall experiments show that our strategies to minimize the response time are effective.

1. Introduction

Location-based services (LBSs) become more and more important in our everyday life. Worldwide revenues from LBSs are expected to go beyond $6 Billion by 2017, according to ABI Research. This huge and ever-growing market has attracted lots of attentions from both academy and industry. In this paper, we study a new location-based query, namely *Merged Aggregate Nearest Neighbor (MANN)*, on a spatial road network.

Formally, given a target set P, a query set Q, a candidate set C and an integer n, an MANN query returns an optimal target point $p \in P$ and a set of n candidates C_s from C ($C_s \subseteq C$), such that the aggregate distance from target point p to all the points in Q and C_s is minimized, i.e., $q(P,Q,n,C) = \{\langle p, C_s\rangle | p \in P \wedge C_s \in \Gamma(C,n) \wedge \forall p' \in P, \forall C' \in \Gamma(C,n), f(p, C_s \cup Q) \leqslant f(p', C' \cup Q)\}$.

☆ A preliminary version of this work was published in the Proceedings of the 22nd International Conference on Information and Knowledge Management (CIKM 2013). Substantial new technical materials have been added to this journal submission. Specifically, the paper extends the CIKM 2013 paper by contributing (i) two new pruning strategies and a new searching strategy presented in Section 3, (ii) two new algorithms based on the new and old strategies presented in Section 3, and (iii) enhanced experimental evaluation that incorporates more parameters and metrics as presented in Section 4. **Notes:** (i) This manuscript is the authors' original work and has not been published nor has it been submitted simultaneously elsewhere, except for the preliminary version (i.e., [21]) mentioned previously. (ii) The main differences between the conference version and this submission are stated above. (iii) All authors have checked the manuscript and have agreed to the submission.

* Corresponding author.

http://dx.doi.org/10.1016/j.ins.2015.03.028
0020-0255/

Here, $f(p,S)$ is the aggregate distance function and it could be max or sum or others based on application needs. Take max as an example, $f(p,S) = \max_{\forall s \in S}\|p,s\|$ with $\|p,s\|$ returning the network distance between p and s on a given road network; and function $\Gamma(C,n)$ is a function which returns all the subsets C's that are formed by n candidate points of C, i.e., $\forall C' \in \Gamma(C,n), C' \subseteq C$ and $|C'| = n$. As we consider the merged set of Q and C_s and the aggregated distance, we simply name the new query as *Merged Aggregate Nearest Neighbor (MANN)*.

MANN can be fit into many real life applications. For example, three friends want to play basketball. They need to find a basketball court and meanwhile invite seven of their friends to play basketball. Here, the target set P is the set of basketball courts available, the candidate set C is a set of friends, and n is 7. MANN can help to select 7 friends and meanwhile locate a basketball court such that the maximum distance from 10 participants' locations to the court is minimum. Our second example could be doctors offering free health consultations. Assume a group of four doctors decide to offer voluntary health consultation to public. They need to find a place that can reach the public easily, and invite a few nurses (e.g., 10) to facilitate the service. Here, the target set P is the set of public areas that can host the voluntary health consultation, such as parks and subway exits, the candidate set C is the set of nurses that have good working relationship with at least one of those four doctors, and n is 10. MANN can help to select 10 nurses and meanwhile locate a public area such that the total distance from the 4 doctors and 10 nurses to the public area is minimum. Another example could be supermarket chain planning expansion. Assume a supermarket chain currently operates one branch in Shanghai and it plans to open another 3 branches in the coming year. To cut down its operating cost, it also plans to have its own warehouse somewhere in Shanghai. Here, the target set P is the set of available locations for warehouses, the candidate set C is the set of available locations for new branches, and n is 3. MANN can help to select 3 branch locations and meanwhile locate one warehouse such that the maximum distance from the warehouse to any branch is minimized.

Fig. 1 shows a simple example of the MANN query that will be used as the running example throughout this paper. The star points form the target set P, the circle points form the candidate set C, the square points form the query set Q, and the integers next to the edges represent the edges' length. Suppose n is 2 and the aggregate distance function considered is max, we list in Table 1 some potential answers and MANN will return p_1 as the answer target point, and c_1 and c_4 as the corresponding answer candidate points.

As shown in the above example, MANN is complex. It considers the distance from the target point to points in Q and the distance from the same target point to n candidate points, while both the target point and the n candidate points are unknown. To the best of our knowledge, MANN query has not been studied in the literature and the most close one is ANN query [17]. However, ANN only considers the distance from a target point to the query points that are fixed and there is no need to locate n candidate points. Given the definition of MANN, there are two naive solutions to process MANN query, namely *p-oriented algorithm* and *c-oriented algorithm*. p-oriented algorithm considers target point $p \in P$ first and it locates, for each target point $p \in P$, the n points from C that are nearest to p as the candidate points. The one that has the minimum aggregated distance to all the query points and the n candidate points is the answer. On the other hand, c-oriented algorithm considers candidate points first. It enumerates all the potential candidate set C_s and there are in total $\binom{|C|}{n}$ potential C_ss. For each potential C_s, an ANN query is issued with $Q \cup C_s$ as the input, and the one with minimal aggregated distance forms the answer to MANN. Obviously, both approaches are inefficient as they blindly scan either all the points in P or all those $\binom{|C|}{n}$ potential C_ss.

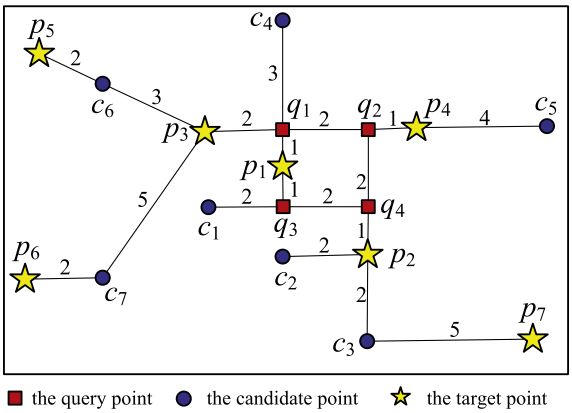

Fig. 1. Example of a MANN query.