

Research on Optimization of Top-k algorithm for continuous query of spatial objects in road network

Zhen Wang

The Information Technology Institute, Liaodong University, China

ldxywz@163.com

Keywords: Road network; Top-k; Spatial data query; STM;LBS

Abstract: In the network environment, the query efficiency of spatial object Top-k query algorithm is an important index to measure the performance of the query. Because in the practical application of query location is often mobile, using the traditional snapshot query to deal with the efficiency of mobile query point Top-k query is poor. The server often receive a large number of query requests, the traditional query algorithm to deal with this situation is also relatively low efficiency. Due to the above reasons, this paper based on the traditional snapshot query method based on Top-k spatial data based on continuous query method is improved, in the algorithm used bottom snapshot query algorithm to solve the problem of space on a single keyword query, use the expansion tree T to traverse the network edge. Finally, through experiments in real road network data set. Finally, through experiments in real road network data set, the efficiency of the proposed algorithm is verified under different parameter conditions.

1 Introduction

At present, spatial data query technology can be roughly points into two categories: one is query based on Euclidean space; second is based on the road network query. In the Euclidean space, the motion of the object is not constrained, and the distance between objects is Euclidean distance; but in real life, who is on a specific road network movement, the distance between two points is usually road network distance between the two sites can't free to the straight line communication, need to follow the existing road network. Nearest neighbor query is one of the most important technology of query, spatial location, close to the space between the object and the Top-k query problem is the evaluation of the degree of factors, according to user generated spatial query requests to find the nearest object position information.

2 related research

In 2009, the Y.-K. Huang [1] proposed multiple moving objects in road network for continuous monitoring method, including pruning and refined in two stages. In 2013, R. Zhong[2] proposed in the network environment of space object k nearest neighbor query method. And put forward a balanced tree index structure G tree. Most of the existing research based on road network is to query the nearest neighbor problem, only to return to the nearest object. Such problems can be attributed to the way online Top-k query. That is, the need to consider the relevant text space and the degree of polymerization. Rocha-Junior [3]in 2012 first proposed Top-k static network environment query method.

The research of mobile object indexing technology in the network, need to complete the road network modeling and moving object modeling. Shahabi [4]proposed to solve the UBA (Upper Bound Algorithm) algorithm of Continuous nearest neighbor query in spatial network. Firstly,

according to the query returns the position of a space object, then calculate an upper bound to exclude a position between the query results invariant nearest neighbor query computation. In addition, there is a method of using snapshot nearest neighbor query in the query path intersection, and the method of continuous query based on Top-k.

3 Top-k continuous query optimization algorithm for road network space

3.1 problem definition. A road network is represented by an undirected connected planar graph $G(N,E)$. In which N and E are represented by a group of nodes and edges in the space network.

O is used to represent a set of spatial text objects in the network G . Each of these objects $o \in O$ has two properties, Space position $o.l$ and text description $o.d$. $|o,n|$ and $|o,n'|$ represent the object O to its side of the two end nodes n 、 n' distance. In addition, the shortest distance between any two objects o and o' in the space network G is represented as $d(o,o')$.

3.2 spatial query based on Top-k. A query in a spatial network can be defined as $q = \langle q.l, q.d, q.k, q.r \rangle$. Among them, $q.l$ represents the location of the query, $q.d$ said that the query text description (a group of keywords), $q.k$ said the number of results returned by the query, $q.r$ represents the direction of query point movement. For a set of spatial text objects in space O , Top-k based spatial query is from the object set O , return to the integration of the τ value of the lowest $q.k$ before the object of the process. The comprehensive score τ can be expressed as follows

$$\tau(q,o) = \alpha \cdot \delta(q.l, o.l) + \beta[1 - \varphi(o_r)] + (1 - \alpha - \beta)[1 - \theta(q.d, o.d)] \quad (1)$$

Where $\delta(q.l, o.l)$ is the spatial proximity between the object location $o.l$ and the query location $q.l$ 、 $\theta(q.d, o.d)$ is the text correlation between the query key $q.d$ and the object text description $o.d$ 、 $\varphi(o_r)$ is the object of the query space density. Spatial proximity (δ): The measurement method of spatial proximity can be defined as

$$\delta(q.l, o.l) = \frac{d(q.l, o.l)}{d_{\max}} \quad (2)$$

d_{\max} , which represents the maximum network distance of any two objects in the space network G . The δ range is $[0,1]$.

Text relevance (θ): text relevance can be obtained through a variety of different information retrieval models. The measurement method can be defined as follows

$$\theta(q.d, o.d) = \frac{\sum_{t \in q.d} \omega_{t,q.d} \cdot \omega_{t,o.d}}{\sqrt{\sum_{t \in q.d} (\omega_{t,q.d})^2 \cdot \sum_{t \in o.d} (\omega_{t,o.d})^2}} \quad (3)$$

The weight of $\omega_{t,q.d} = \ln\left(1 + \frac{|O|}{df_t}\right)$, $|O|$ said the number of objects in space, df_t said the text

contains a number of t text objects (text frequency); The weight of $\omega_{t,o.d} = 1 + \ln(f_{t,o.d})$, $f_{t,o.d}$ represents the number of times (frequency) of the t in the object text description $o.d$. The range of θ is also $[0,1]$.

Intensity (φ): Calculation of a related object o_i concentration process first in a related object edges where the two endpoints found and closer to the endpoint n' ; after calculation the endpoint

of close center of.

3.3 space continuous query based on Top-k under the environment of road network.

Continuous query Top-k and traditional fixed point query based spatial query difference method is based on the network environment, according to the change of position of the need of real-time query Top-k query results to the user. Under the road network environment based on the Top-k spatial continuous query is the biggest problem faced by the method, with the query location of mobile need frequent query location updates to the server delivery, resulting in the tedious Top-k result set calculation process.

Used in this algorithm with the enhanced[3] algorithm is similar to the snapshot query keyword query algorithm to solve a single space at the bottom of the problem, and gradually expand the scope of coverage from a query node, it uses a pair of upper and lower bounds for the pruning algorithm for joint development of space and text fields. In addition, T this algorithm uses extended tree traversing the network. Expand the tree in the efficient handling of continuous query problems play an important role.

Algorithm 2 describes the detailed steps to obtain the results of the snapshot query. Need to calculate the set of query nodes s and where it edge e as input to the algorithm, calculated according to the score of τ in ascending order k object as a result set R_q . With a priority queue N (available minimum heap Implementation), to increase the distance from the s to the order of the organization to meet the node. First of all, the algorithm updates the result set R_q the k of the first of the object score ε . One end of the positioning side e node n' , s and T for the root node expansion tree, set to its child nodes. Insert SS and NN into the queue n' and mark it as an access (fourth rows). Later in the edge of the $e(s, n')$ query scores less than ε of the candidate object C , and update the R_q and ε . After from the priority queue N column and s recent node n , and the processing unvisited neighbor node. The stopping condition of the algorithm for all network has been extended, or generated with the minimum distance of the remaining objects of the lower limit value is greater than or equal to the current k has been searching for objects score (line eighth). The lower limit value of the node n can be obtained by the network distance of s and n and the maximum value of text correlation ($\theta = 1$). Thus, $\delta(s, o.l) \geq \delta(s, n)$.

Algorithm 1 is described in detail to find the candidate set of steps. First, the algorithm by mapping each text $t \in q.d$ module stores the maximum degree of impact value λ_{\max} calculates the maximum text θ_{\max} . After that, the lower bound of the value of the τ_{\min} is calculated using the shortest distance between the edge and the query nodes and the maximum text correlation value θ_{\max} . When the lower bound value τ_{\min} is less than e , the exact value of the object on the edge is calculated and the candidate object set C is returned to the score below e .

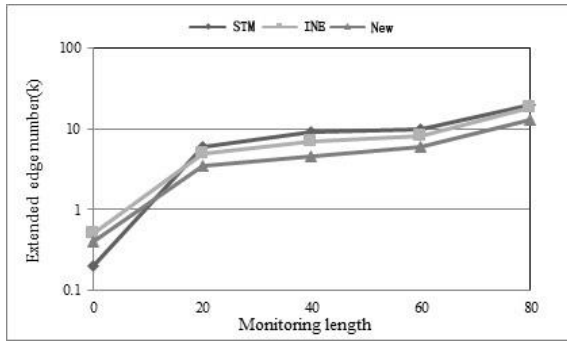
4 Simulation Results

All the algorithms in the experiment include: i5-6300HQ 8GB, CPU memory and 256GBSSD hard disk; Windows Microsoft 8.1 operating system; programming language: Java.

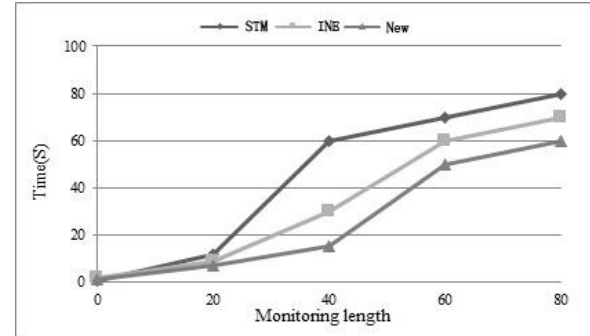
Experiments using the France and Germany road network as a real data set, this paper uses the Brinkhoff generator[5] to generate the trajectory of the mobile query point. The specific process is an object appears in a network node, to a randomly generated destination by the shortest path to move after the disappearance of. The trajectory generated by each object is represented by 100 points, each time stamp (s) generates a point, that is, the need to use 100 seconds to monitor the formation of a trajectory. Each experiment consisted of 100 such queries and the average results of trajectory.

This experiment analyzes the influence of the monitoring length on the algorithm. As shown in

Fig. 1, monitor the change in length between 0~80. When the monitoring length is 0, the equivalent of the snapshot query algorithm. Fig. 1(a) and Fig. 1(b) can be seen in the snapshot query environment INE, STM, New three algorithm performance is almost. However, with the increase of the length of the monitoring, the execution time and the extended variables of the three algorithms continue to increase. Among them, the performance of INE algorithm is better than STM, this algorithm is better than INE algorithm.



(a) Effect of L on execution time



(b) The effect of L on the extended edge number

Fig. 1 effect of monitoring length L on the performance of the algorithm

Performance analysis of space continuous query algorithm based on Top-k. By comparing with the original algorithm, the performance of the query system is compared under different monitoring length parameter environment. Through the experimental results, that the algorithm is lower than the traditional algorithm in the communication consumption and communication frequency.

5 Conclusion

In this paper, we propose a continuous query method based on Top-k for moving query points in the road network. Advantages are mainly reflected in: in order to search results in the space, in addition to considering the spatial proximity and text relevance, but also special consideration of the density of the query object. In terms of the efficiency of the algorithm, the algorithm is superior to the traditional algorithm in terms of execution time and the number of edges. In terms of the communication efficiency between the client and the server, the communication consumption and the communication frequency of this algorithm are both less than STM algorithm.

References

- [1]Y.-K. Huang, Z.W. Chen and C. Lee, Continuous K-nearest neighbor query over moving objects in road networks, APWeb/WAIM2009, LNCS5446, 2009, pp.28–38..
- [2]Zhong R, Li G, Tan K L, et al. G-tree: an efficient index for KNN search on road networks[C]//Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. ACM, 2013: 39-48.
- [3]Rocha-Junior J B, Nøravåg K. Top-k spatial keyword queries on road networks[C]//Proceedings of the 15th international conference on extending database technology. ACM, 2012: 168-179.
- [4]Kolahdouzan M R, Shahabi C. Continuous K-Nearest Neighbor Queries in Spatial Network Databases[C]//STDBM. 2004: 33-40.
- [5]Brinkhoff T. A framework for generating network-based moving objects[J]. GeoInformatica, 2002, 6(2): 153-180.