

移动环境下基于共享客户信息的空间位置查询

孙小培¹, 朱玉全¹, 陈耿², 桑丘¹

(1. 江苏大学 计算机科学与通信工程学院, 江苏 镇江 212013; 2. 南京审计学院 江苏省审计信息工程重点实验室, 南京 210029)
(sunxp_@163.com)

摘要:针对移动环境下空间位置查询不能及时得到响应等问题,提出了一种基于最共享邻近客户信息的空间位置查询算法(SRNN)。该算法充分利用客户本身所具有的通信和计算能力,共享周围客户的信息,减轻了中心服务器的负担,同时降低了移动客户的等待时间。实验结果表明,SRNN算法是可行有效的。

关键词:移动环境;空间查询;缓存;响应时间

中图分类号: TP311.13 **文献标志码:** A

Study on location-based spatial queries with shared client results in mobile environments

SUN Xiao-pei¹, ZHU Yu-quan¹, CHEN Geng², SANG Qiu¹

(1. School of Computer Science and Telecommunications Engineering, Jiangsu University, Zhenjiang Jiangsu 212013, China;
2. Jiangsu Key Laboratory of Audit Information Engineering, Nanjing Audit University, Nanjing Jiangsu 210029, China)

Abstract: For resolving the problem that spatial position query cannot responded in time under mobile environments, a new query algorithm called Share Results Nearest Neighbor (SRNN), based on sharing the information of adjacent mobile clients, was put forward. The proposed method reduces the burden of central server through making full use of the communication and computational capabilities of clients to share the information about their surroundings. Meanwhile, it decreases the waiting time of mobile clients. Experimental results show that SRNN algorithm is feasible and effective.

Key words: mobile environment; spatial query; cache; response time

0 引言

近年来,随着移动通信技术的迅速发展与广泛应用,空间查询已成为移动应用中的一个重要组成部分^[1]。当移动客户需要了解相关数据时,它必须向中心服务器发出请求,一旦客户错过了本次广播周期内它所需要的信息包,就必须等待下一次广播周期^[2]。由于客户的移动性,如果下次广播存在较大的延迟,查询结果将会失去其可用性。诸如回答“找到最近的三所医院”^[3],如用户在高速行驶几分钟后才得到结果,这样查询就变得毫无意义。因而,如何使移动客户及时获得查询结果是空间查询中的一个关键问题。

文献[4]总结了无线广播模式下查询处理技术所面临的挑战,针对备受关注的K最近邻(K-Nearest Neighbor, KNN)和窗口查询(Window Queries, WQs)两类空间查询,利用点对点共享来提高响应时间。该方法涉及参数多达十个以上,应用过于复杂,实际可行性很差。文献[5]利用移动客户响应查询时间间隔,采用距离-响应时间计算算法和KNN查询分段排序算法来简化计算,提高了查询时效性,但其误判率较高。为此,本文提出了一种基于共享最近邻客户信息的查询算法(Share Results Nearest Neighbor, SRNN),该算法充分利用客户本身所具有的通信和计算能力,共享其他客户的信息,减轻了中心服务器的负担,在保证可扩展性的同时,采用较少的参数来降低移动客户的等待时间,同时保证了信息的时效

性。实验结果表明算法SRNN具有较好的性能。

1 相关定义及定理

由于移动客户总是在道路网络上移动的,目标之间的最短网络距离取决于网络之间的连通性^[6],所以应根据连通性确定最邻近客户。除此以外,空间查询结果具有位置性^[7],即两个距离较近的移动客户的空间查询结果可能存在明显的重叠部分。随着距离的缩短,其查询的响应时间也就越短,并且查询结果可共享的可能性就越大。反之,目标离查询点越远,信息失效的可能性越大,共享查询的可能性就越小^[8-9]。

设 $\|P, Q\|$ 表示移动客户 P 和 Q 之间的距离,缓存信息是移动客户通过查询后保存于自身的位置信息。

定义1 有效信息。若 P 中的缓存信息 v_1 是 Q 发出的空间位置查询所需结果,则将 v_1 称为 Q 的有效信息,否则 v_1 为 Q 的无效信息。

定义2 有效区域。设 Q 有 m 个有效信息为 v_1, v_2, \dots, v_m ,对于任意 $v_i (i = 1, 2, \dots, m)$,找到其所位于的移动客户 $S_j (j = 1, 2, \dots, n)$, S_j 根据连通性执行KNN查询后,获得的全部移动客户所形成的位置覆盖区域称为有效区域。

定理1 假设 P 的 k 个最近邻居为 N_1, N_2, \dots, N_k , $\|P, N_1\| \leq \|P, N_2\| \leq \dots \leq \|P, N_k\|$ 。若 $\|Q, N_i\| + \|P, Q\| \leq \|P, N_k\| (i = 1, 2, \dots, k-1)$,则 N_i 中的缓存信息是 Q 的有效信息。

收稿日期:2009-06-22;修回日期:2009-08-11。

基金项目:江苏省“六大人才高峰”项目(07-E-025);江苏省高校自然科学基金资助项目(08KJA520001)。

作者简介:孙小培(1983-),女,河南许昌人,硕士研究生,主要研究方向:移动计算、数据库系统;朱玉全(1966-),男,江苏常州人,教授,博士,主要研究方向:数据挖掘、数据库系统;陈耿(1965-),男,江苏无锡人,教授,博士,主要研究方向:数据库系统、数据挖掘;桑丘(1983-),男,江苏镇江人,硕士研究生,主要研究方向:人工智能、信念修正。

证明 反证法。设 Q 的 k 个最近邻居为 M_1, M_2, \dots, M_k , 若 N_i 中的缓存信息不是 Q 的有效信息, 则对于任意的 $M_j (j = 1, 2, \dots, k)$, 可知 $\|Q, M_j\| < \|Q, N_i\|$ 。

1) 设 M 点是线段 PQ 的延长线和以 Q 为圆心、 $\|Q, N_i\|$ 为半径的圆的交点, 则对于任意的 $M_j (j = 1, 2, \dots, k)$, 有 $\|P, M_j\| < \|P, M\|$ 成立。

2) 若 $\|Q, N_i\| + \|P, Q\| > \|P, N_k\| (i = 1, 2, \dots, k - 1)$, 那么 N_i 中的缓存信息就不一定是 Q 的有效信息。如图 1 所示, 由于 $\|Q, N_2\| + \|P, Q\| > \|P, N_3\|$, 则 N_2 形成的区域部分位于有效区域之外。因此我们只能把 N_2 作为一个不确定的最近的邻居。

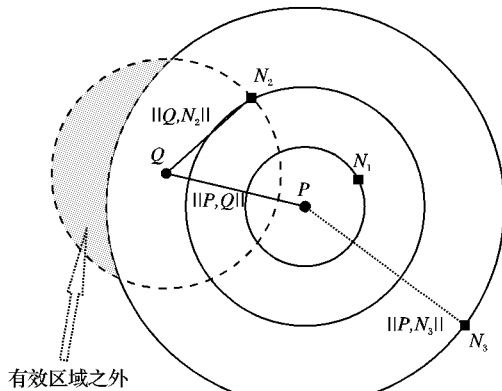


图1 $\|Q, N_2\| + \|P, Q\| > \|P, N_3\|$

3) 由于 $\|Q, N_i\| + \|P, Q\| = \|Q, M\| + \|P, Q\| = \|P, M\|$, 因此圆 CQ (它的圆心是 Q , 半径是 $\|Q, M\|$) 已经完全被圆 CP (圆心是 P , 半径是 $\|P, M\|$) 所覆盖。

由证明 1) ~ 3) 可知, 对任意 Q 的有效信息 $M_j (j = 1, 2, \dots, k)$, 有 $\|P, M_j\| < \|P, N_k\|$ 。因此, N_k 不是 P 的最近邻居之一。然而, 这与条件 N_k 是 P 的最近邻居之一矛盾。因此, N_i 是 Q 的 k 最近邻居之一, 即 N_i 中的缓存信息是 Q 的有效信息。

证毕。

如图 2 所示, $\|Q, N_2\| + \|P, Q\| < \|P, N_3\|$, 则 N_2 中的缓存信息是 Q 的有效信息。

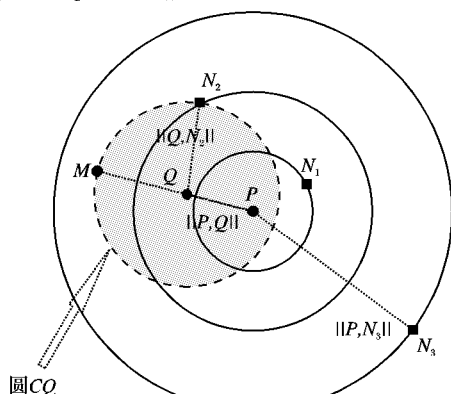


图2 $\|Q, N_2\| + \|P, Q\| < \|P, N_3\|$

定理 2 设 Q 的有效区域用 S 表示, 假设 S 中有 k 个移动客户, 它们分别为 R_1, R_2, \dots, R_k , 对于 S 内任意 R_i , 以 Q 为圆心, Q 和 R_i 的距离为半径画圆 C , 若 C 完全被 S 覆盖, 则 R_i 中的缓存信息是 Q 的有效信息。

证明 反证法。假设 R_i 中的缓存信息是 Q 的有效信息, 但以 R_i 和 Q 为半径创建的圆不被 R 覆盖。则存在 $\|Q, R_i\| < \|Q, R_k\|$, 并且 $\|Q, R_i\|$ 小于最短的距离。由于 $\|Q, R_i\|$ 小于最短的距离, 由定理 1 知 R_i 是一定在确定的有效区域范围内。则 $\|Q, R_i\|$ 小于所有确定的有效区域范围的边的垂线的

长度, 这与 R_i 和 Q 为半径创建的圆不被 R 覆盖矛盾。证毕。

如图 3 所示, Q 到边 E_1 的距离最短, 且 $\|Q, R_1\|$ 小于它的距离, 因此 R_1 中的缓存信息是 Q 的有效信息, 而 $\|Q, R_5\|$ 大于它的距离, 不能完全被 Q 的有效区域覆盖, 则 R_5 中的缓存信息不是 Q 的有效信息。

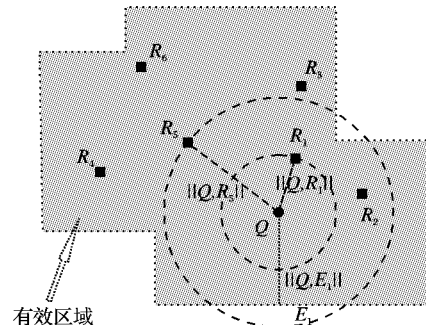


图3 $\|Q, R_1\| < \|Q, E_1\|$ 和 $\|Q, R_5\| > \|Q, E_1\|$

由上可知, 对于任意客户 R_1, R_2 , 若 $\|Q, R_1\| \neq \|Q, R_2\|$, 那么 R_1, R_2 位于不同位置, 且当 $\|Q, R_1\| < \|Q, R_2\|$ 时, 若 R_2 中的缓存信息是 Q 的有效信息, 则 R_1 中的缓存信息也是 Q 的有效信息。

2 基于共享客户信息的空间位置查询 SRNN

假设移动客户能确定自己的位置并有一定的计算能力, 使得等待数据广播信息时, 移动节点能承担部分工作, 并且移动客户之间可以相互通信。

共享客户信息查询 SRNN 算法功能是移动客户 Q 通过共享附近多个移动客户的缓存信息, 判断从其他移动客户获得的信息对自身是否有用。移动客户 Q 根据连通性访问邻近的移动客户 P , 对任意 N_i , 若 $\|Q, N_i\| + \|P, Q\| \leq \|P, N_k\|$, 由定理 1 可知 N_i 中的缓存信息是 Q 的有效信息。若此时得到结果数目不满足移动客户要求, 则求出 Q 的有效区域 S , 并对 S 内任意 R_i , 以 Q 为圆心、 Q 和 R_i 的距离为半径画圆 C , 若 C 完全被 S 覆盖, 则 R_i 中的缓存信息是 Q 的有效信息。为了判断是否被覆盖, 采用首先求 S , 再逐一判断 $\|Q, R_i\|$ 是否小于 Q 到所有顶点之间连线的距离。

算法 共享客户信息查询 $ShareResultsNN(P, Q, PknnResults)$

输入: P, k 个最邻近移动客户之一; Q , 发出查询的移动客户; $PknnResult, P$ 根据网络距离得到的 k 个最邻近移动客户。

输出: 符合移动客户 Q 查询的结果 $QknnResults$ 。

- 1) If ($PknnResults \neq null$) then
- 2) for (each $Pknn$ in $PknnResults$)
- 3) if ($PQDistance + PknnQDistance \leq PknnDistance$) then
// $\|Q, N_i\| + \|P, Q\| \leq \|P, N_k\|$
- 4) AddTo($QknnResults$);
- 5) SortByTime($QknnResults$); // 将结果按响应时间排列
- 6) if (count $QknnResults < k$) then
- 7) TwoResults = MergekNNResult;
// 结果数小于 k , 合并邻近结果
- 8) GetVertices(TwoResults); // 得出外包点
- 9) SortByNeighbor(GetVertices); // 将外包点按顺序排列
- 10) if (DistanceCQ < TwoResults(Point i)) then
// 判断 Q 与当前客户的距离是否小于 Q 到
// 所有相邻外包点之间连线的距离
- 11) AddTo($QknnResults$);
- 12) return $QknnResults$;

3 模拟实验及分析

模拟实验中, 机器配置是 CPU 为 P4 2.66 GHz, 内存 1.43

GB, 硬盘 80 GB。模拟程序采用 Microsoft Visual Studio 2008 的 VC#. Net 语言编写。数据库方面, 采用 Microsoft SQL Server 2005, 采用三组真实数据集^[10], 第一组 Real Datasets for California Road Network, 表 Cnode (NodeID, Longitude, Latitude) 和 Cedge (EdgeID, StartNodeID, EndNodeID, L2Distance); 第二组 Real Datasets for Road Network of North America, 表 NAnode (NodeID, NXCoordinate, NYCoordinate) 和 NAEedge (EdgeID, StartNodeID, EndNodeID, L2Distance); 第三组 Real Datasets for City of Oldenburg Road Network, 表 Onode (NodeID, NXCoordinate, NYCoordinate) 和 Oedge (EdgeID, StartNodeID, EndNodeID, L2Distance)。其中表 Cnode、NAnode、Onode 分别保存 21 048、175 813、6 105 条位置信息, 表 Cedge、NAEedge、Oedge 中分别保存 21 693、179 179、7 035 条相连通的位置信息。受影响的参数有 3 个, 分别为可共享的移动客户 (Mobile Host, MH) 数、 k 值 (移动客户从其他客户共享信息后所需获得有效信息的邻近点个数)、可共享的移动客户距离范围。其中 MH 数取 0~20, k 值分别取 3、5、7、10, 距离范围取 0~200 km 和在国外三个不同地区真实数据集的可行性观察实验结果。图 4 为固定范围大小为 15 km, 讨论当 k 值和移动客户数量不同时, 在移动环境下共享客户信息查询的方法能够帮助解决问题的百分比值。图 5 固定 k 值为 5 和移动客户的数量为 20, 讨论当距离范围不同时, 该方法所能解决问题的百分比值。图 6 固定范围大小为 15 km 和 MH 数为 15 时, 讨论在不同地区数据集, 该方法所能解决问题的可行性。

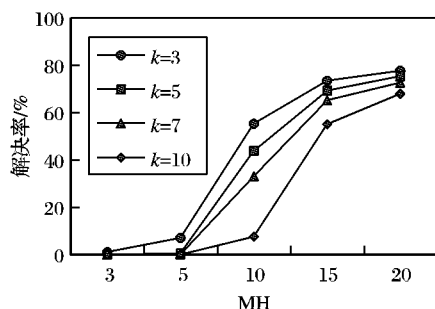


图4 K值和MH值对SRNN查询的影响

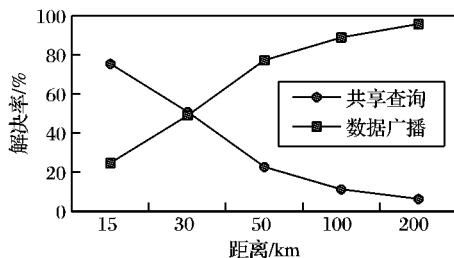


图5 距离范围对SRNN查询的影响

模拟实验结果充分显示, 方法 SRNN 能够更好地解决无线广播模式下查询响应时间长的问题, 其解决的效果主要受以下三个方面的影响: k 值的大小、移动客户数以及查询限定的距离范围。结果表明:

1) 当 k 值越小时, 共享客户信息查询所能解决的百分比就越高。这是由于移动客户所要查询的信息越少, 能够完全满足的可能性越大。

2) 当移动客户数越大时, 共享客户信息查询所能解决的百分比就越高。但是当移动客户数到达一定数值后, 提高效率就会减慢。这是由于移动客户数越大, 连通性越强所获得的信息就越多, 当移动客户数到达一定程度后, 所需查询结果已得到。

3) 需要选择合适的距离范围, 当距离范围扩大到一定程度后, 共享客户信息查询所能解决的百分比就会降低。这是由于如果固定 k 值和移动客户数后, 范围大到一定程度时会出现许多无用信息, 从而使查询解决百分比降低。

4) 客户越集中, 共享客户信息查询所能解决的百分比就越明显。这是由于三个地区中 City of Oldenburg Road Network 相互查询信息的客户端距离近, 从而使得该方法更加有效。

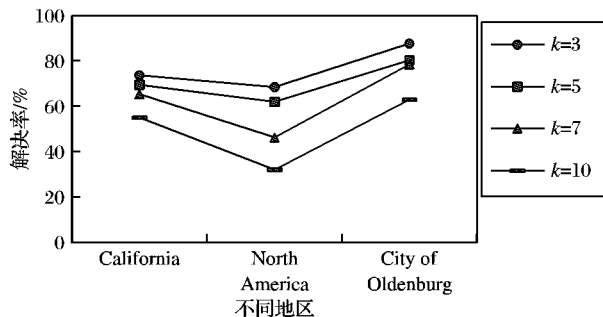


图6 数据集对SRNN查询的影响

4 结语

基于位置空间查询中的时间响应是移动环境下的重要问题之一, 本文通过探讨空间查询结果的位置, 结合无线广播模式和移动客户之间信息的有效性判断, 提出了该环境下共享客户信息查询的算法, 并通过模拟实验验证了此方法的有效性。不过由于其本身受到一些参数的影响, 参数的选定会影响到此方法的效果, 因此需要进一步讨论各个参数值选定的方法, 使其更具有通用性。另外, 需要对共享移动客户信息的安全性加以考虑。

参考文献:

- [1] SHEKHAR S, CHAWLA S. 空间数据库[M]. 谢昆青, 马修军, 杨冬青, 等译. 北京: 机械工业出版社, 2004: 12-22.
- [2] HU H B, XU J L, LEE D L. A generic framework for monitoring continuous spatial queries over moving object[C]// Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 2005: 479-490.
- [3] ROUSSOPOULOS N, KELLEY S, VINCENT F. Nearest neighbor queries[C]// Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 1995: 71-79.
- [4] KU W S, ZIMMERMANN R, WANG H X. Location-based spatial query processing in wireless broadcast environments[J]. IEEE Transactions on Mobile Computing, 2008, 7(6): 778-791.
- [5] 宋晓宇, 王睿, 孙焕良. 基于P2P结构的kNN查询框架[J]. 沈阳建筑大学学报: 自然科学版, 2007, 23(6): 1040-1043.
- [6] ZHENG B H, LEE W C, LEE D L. Spatial queries in wireless broadcast systems[J]. Wireless Networks, 2004, 10(6): 723-736.
- [7] KU W S, ZIMMERMANN R, WAN C N. Location-based spatial queries with data sharing in mobile environments[C]// Proceedings of the 22nd International Conference on Data Engineering Workshops. Washington, DC: IEEE Computer Society, 2005: 1355-1359.
- [8] YIN L Z, CAO G H. Supporting cooperative caching in Ad Hoc networks[J]. IEEE Transactions on mobile computing, 2006, 5(1): 77-89.
- [9] CHOW C Y, LEONG H V, CHAN A. Peer-to-Peer cooperative caching in mobile environment[C]// Proceedings of the 24th International Conference on Distributed Computing Systems Workshops. Washington, DC: IEEE Computer Society, 2004: 528-553.
- [10] LI F F, KOLLIOS G. Real datasets for spatial databases: Road networks and points of interest[EB/OL]. [2008-12-06]. <http://www.cs.fsu.edu/~lifeifei/SpatialDataset.htm>.