

网络环境中近邻对的查询方法

张丽平¹, 李 松²

(1. 辽宁工程技术大学 电子与信息工程学院, 辽宁 葫芦岛 125105;

2. 哈尔滨理工大学 计算机科学与技术学院, 黑龙江 哈尔滨 150080)

摘要:网络环境下的数据集中的近邻对查询在地理信息系统、网络查询和空间数据库等领域有着重要的应用。为了对网络环境下的近邻对进行有效查询,基于 Voronoi 图对数据集中近邻对问题进行了详细研究,给出了网络环境下查询数据点集中近邻对的定理和算法;为了利用计算机对网络环境下的近邻对进行查询处理,设计了相应的数据存储结构;对在网络环境下的查询数据集中的近邻对问题进行了实验分析。该方法可较好地解决网络环境下的数据集中近邻对的查询问题,相应的维护代价较低。

关键词:Voronoi 图;最近邻;最近对;近邻对

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2008)06-0119-03

Methods of Near Neighbor Pair Query in Network

ZHANG Li-ping¹, LI Song²

(1. School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China;

2. School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China)

Abstract:The near neighbor pair query of the datasets in the network has important application in geographical information systems, query in the network and spatial database. To search the near neighbor pairs effectively in the network, the near neighbor pair query was studied detailedly based on the Voronoi diagram. The theorems and algorithms of searching the near neighbor pair of the datasets in the network were proposed. To search the points in the database, the corresponding structures of the database were constructed and the experiments for the algorithms were also given. The methods in this paper can deal with the near neighbor pair query of the datasets in the network and the costs of the maintenance are low.

Key words:Voronoi diagram;nearest neighbor;closest pair;near neighbor pair

0 引言

网络环境中有关数据集中的近邻对查询问题在实际应用中有着重要意义。所谓近邻对查询就是在数据集中查找数据点和它的最近邻、数据集中互为最近邻的数据点对及数据集中的最近对。近邻对查询可被应用到地理信息系统、设施定位、图像处理、交通管制、网络查询、空间数据库和多媒体数据库等多种领域。

已有的相关工作^[1-5]主要集中在查询数据集中数据点的最近邻和计算数据集中最近对等方面,没有对网络环境下的数据集中近邻对进行分析,不能有效得出数据集中互为最近邻的数据点对和最近对。为了弥补已有方法的不足,文中基于 Voronoi 图^[6]的概念和

性质对网络环境下有关数据集中近邻对查询问题进行了详细分析,对近邻对进行了形式化定义,给出了基于 Voronoi 图的查询策略和查询算法。该方法可较好地解决网络环境下数据集中近邻对的查询问题。

1 定义与性质

定义 1 给定空间数据对象点集 D 和 P 。 $d \in D$, $p \in P$, $D(d, p)$ 表示点 d 到 p 的最短距离。 $\forall o \in P$, 若 $D(d, p) \leq D(d, o)$, 称 p 是 d 的最近邻, 记为 $p = NN(d)$; 若 $NN(d) = p$ 且 $NN(p) \neq d$, 则称点对 (d, p) 为单向近邻对, 记为 $< d, p \rangle$, d 称为近邻基点; 若 $NN(d) = p$ 且 $NN(p) = d$, 则称点对 (d, p) 为对偶近邻对, 记为 $< d, p >$; $\forall d' \in D, p' \in P$, 若 $D(d, p) \leq D(d', p')$, 则称 (d, p) 为最近对, 记为 $[d, p]$ 。单向近邻对、对偶近邻对和最近对统称为近邻对, 若点对 (d, p) 为近邻对, 则有 $NN(d) = p$ 或 $NN(p) = d$ 。在 D 和 P 中寻找近邻对的查询称为近邻对查询。

收稿日期: 2007-09-10

基金项目: 黑龙江省创新科研基金资助项目(YJSCX 2006-13HLJ)

作者简介: 张丽平(1976-), 女, 辽宁铁岭人, 讲师, 硕士, 研究方向为数据结构和算法设计、数据库。

由定义 1 可知最近对必是对偶近邻对, 对偶近邻对不一定是最近对。

定义 2 设一空间网络为 SPN, 若 P 为空间网络 SPN 上沿边运动的动态点集, D 是分布在网络边上或边外的静态点集, 则称在 P 和 D 中寻找近邻对的查询为网络环境中的近邻对查询。

如图 1 所示, p_1, p_2, \dots, p_{36} 是网络边上的动态点, d_1, \dots, d_{34} 是静态点。

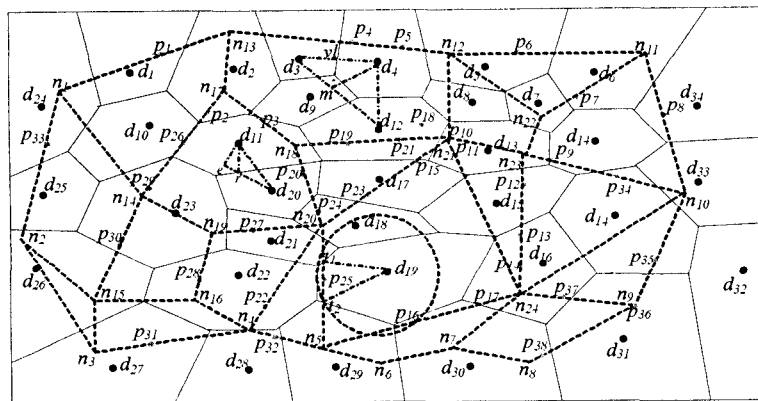


图 1 Voronoi 图和网络环境下的数据点集示例

定义 3^[6] 给定一组空间数据点 $D = \{d_1, \dots, d_n\} \subset R^2$, 其中 $2 < n < \infty$, 当 $(i \neq j)$ 时, $d_i \neq d_j$. Voronoi 区域由以下公式给出: $VP(d_i) = \{d \mid D(d, d_i) \leq D(d, d_j)\}$. $D(d, d_i)$ 为 d 与 d_i 间的最小距离. d_i 称为 Voronoi 生成点, 由 d_i 所决定的 Voronoi 区域 $VP(d_i)$ 称为 Voronoi 多边形, Voronoi 多边形的棱记为 $VL(d_i)$. 由 $V(D) = \{VP(d_1), \dots, VP(d_n)\}$ 所定义的图形被称为 Voronoi 图. 共享相同棱的 Voronoi 多边形被称为邻接多边形, 它们的生成点被称为邻接生成点. 图 1 展示了部分 Voronoi 图, 其中, d_1, \dots, d_{34} 是生成点。

性质 1 离 Voronoi 生成点 d_i 最近的生成点 d_j 必是 d_i 的邻接生成点。

证明 连接 Voronoi 生成点 d_{12} 与 d_3 (如图 1 所示), 线段 $d_{12}d_3$ 经过 d_{12} 的邻接 Voronoi 多边形, 且与棱 v_l 的反向延长线交于 m 点, 连接 d_4 与 m . 因为 v_l 是线段 d_3d_4 的垂直平分线, 所以得出 $|d_4m| = |d_3m|$, 由三角形的性质可知 $|d_4m| + |d_{12}m| > |d_{12}d_4|$, 又因为 $|d_{12}d_3| = |d_3m| + |d_{12}m|$, 所以可得出 $|d_{12}d_3| = |d_3m| + |d_{12}m| = |d_4m| + |d_{12}m| > |d_{12}d_4|$. 证毕。

2 网络环境中近邻对查询

2.1 基于 Voronoi 图的查询方法

为了利用 Voronoi 图进行网络环境中的近邻对查

询, 本节先给出定理 1 及其推论。

定理 1 一个 Voronoi 多边形 $V(d)$ 内任何点 p ($p \in D$) 到 d 的距离必小于 p 到其他 Voronoi 生成点的距离。

证明 设 e 点为 Voronoi 多边形 $V(d_{11})$ 内的任一点 (如图 1 所示), 连接点 e 与生成点 d_{11} , 再连接与 $V(d_{11})$ 共享棱的任一生成点例如 d_{20} , 线段 ed_{20} 与共享棱交于点 r , 连接点 d_{11} 与 r . 因为 $|rd_{11}| = |rd_{20}|$, 又因为 $|ed_{20}| = |er| + |rd_{20}|$, 即 $|ed_{20}| = |er| + |rd_{11}|$, 由三角形的性质可知 $|er| + |rd_{11}| > |ed_{11}|$, 即 $|ed_{20}| > |ed_{11}|$. 证毕。

推论 1 设对象点集 D 和 P , $V(D)$ 是 D 的 Voronoi 图. 若 D 和 P 内的单向近邻对为 $\langle d_m, p_n \rangle$, 则 p_n 必在 $VP(d_m)$ 内或 $VL(d_m)$ 上; 若 D 和 P 内的对偶近邻对为 $\langle d_i, p_j \rangle$, 则 p_j 必在 $VP(d_i)$ 内或 $VL(d_i)$ 上; 若 D 和 P 内的最近对为 $[d_s, p_t]$, 则 p_t 必在 $VP(d_s)$ 内或 $VL(d_s)$ 上。

推论 2 设空间网络为 SPN, 静态对象点集 D 的 Voronoi 图为 $V(D)$, 则 SPN 被 Voronoi 多边形 $VP(d_i)$ 分割成各个子部分, 文中将 $VP(d_i)$ 内的网络部分记为 $NVP(d_i)$, 由定理 1 可知, 在 $NVP(d_i)$ 上运动的动点 p_j 的最近邻即为 d_i , 即 $NN(p_j) = d_i$ 。

设 p_j 为 $NVP(d_i)$ 上的动态对象点, $\langle d_i, p_j \rangle$ 为对偶近邻对, 则以 d_i 为圆心, $|d_i, p_j|$ 为半径做圆 $\text{Circle}(d_i, |d_i, p_j|)$, 设 $\text{CNVP}(d_i) = \text{Circle}(d_i, |d_i, p_j|) \cap NVP(d_i)$, 则称 $\text{CNVP}(d_i)$ 为数据集中最近对的监测区, 由定理 1 及其推论可知, 下一时刻出现在监测区内的动态对象点 p_k 最可能成为 d_i 的新的最近邻, 即 (d_i, p_k) 最有可能成为新的最近对. 如图 1 所示, 下一时刻在 t_1, t_2 段上运动的动态对象点将成为重点计算对象。

根据定理 1 及其推论可得出在网络环境中进行近邻对查询的算法如下:

算法: NVO_SEARCH(D, P)

输入: 空间网络 SPN, SPN 上的动态对象点集 P , 静态对象点集 D

输出: 单向近邻对集 SC , 对偶近邻对集 DC , 最近对集 NC

begin

$SC = \emptyset, DC = \emptyset, NC = \emptyset, SSC = \emptyset$; // 初始化

if D 集的 Voronoi 图没生成 then

生成 $V(D)$;

else

```

for  $i = 1$  to  $m$  do
     $VPD \leftarrow NVP(d_i)$ ;
    //SPN 被  $V(D)$  分割的子部分集合
    if  $NVP(d_i)$  上有动态对象点集  $P_i$  then
        for  $s = 1$  to  $|P_i|$  do
             $SSC \leftarrow \langle p_{is}, d_i \rangle$ ;
            计算  $d_i$  的最近邻  $p_{ix}$ ;
            if  $p_{ix} \in P_i$  then
                 $DC \leftarrow \langle p_{ix}, d_i \rangle$ ; // 得出对偶近邻对
                 $SC \leftarrow SSC - DC$ ;
            else  $SC \leftarrow \langle d_i, p_{ix} \rangle$ ;
        for  $j = 1$  to  $|DC|$  do
            计算  $DC$  中各对偶近邻对的距离;
            求距离最小值  $|d_w, p_z|$ ;
             $NC \leftarrow [d_w, p_z]$ ;
        return  $SC, DC, NC$ ;
end

```

NVO_SEARCH()算法的效率主要决定于数据集的大小及 P 集数据对象点的空间分布情况。

2.2 数据存储结构

为了完成网络环境中对数据集 D 集和 P 集中近邻对的查询,文中设计了相应的数据存储结构。图 2 是部分数据存储结构图。

图 2 中,空间网络 SPN 表存储的是空间网络结点和边的信息;数据对象点集 D 的 $V(D)$ 表存储的是数据集 D 生成的 Voronoi 图信息,其生成点之间的邻接关系由多级指针进行链接索引;NVP(d_i) 表存储的是空间网络 SPN 在 $VP(d_i)$ 内的网络部分信息;数据对象点集 P 表存储的是 P 集的数据点信息。单向近邻对表、对偶近邻对表和最近对表是由 NVO_SEARCH() 算法基于数据对象点集 D 的 $V(D)$ 表、NVP(d_i) 表和 数据对象点集 P 表而得出的。

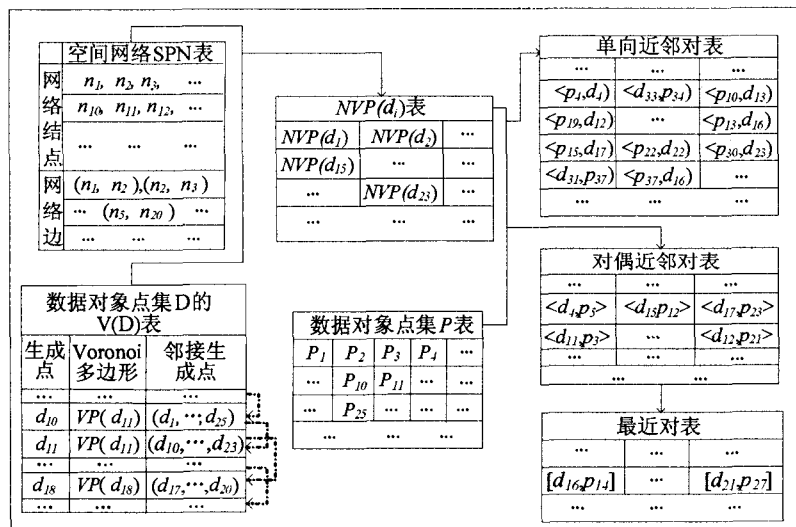


图 2 部分数据点的存储结构

3 方法讨论和实验分析

已有的方法^[1~5]主要针对特定空间内的数据集进行处理,没有给出网络环境中处理近邻对的数学模型和数据存储结构,解决具体问题时具有很大的局限性。由于 Voronoi 图可在各种类型的复杂曲面上构建,因此文中的方法可扩展到在各类曲面上进行网络环境中的数据集的近邻对的计算和查询。利用 Voronoi 图可预先对点对进行筛选和分析,不需要进行大量的距离其相应的维护代价较低。

文中在 Pentium4, 1.8G CPU, 512M 内存, Windows XP 环境下,利用 C++ builder 语言对文中的算法进行了实验。图 3 给出了详细的实验分析结果。纵轴表示在曲面上查询相对于平面上查询代价的比率,横轴表示数据集 D 和 P 的空间数据对象点的数量。由于网络环境下的空间数据对象点的位置分布具有相当的不均匀性,算法的查询效率受数据点的分布和数据集的大小双重影响较大。实际应用中,对空间数据点的有效处理和筛选是查询算法优化的一个重要方面。

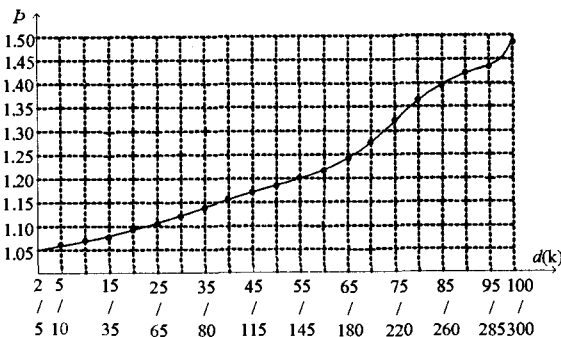


图 3 查询效率分析

4 结束语

文中基于 Voronoi 图对网络环境下有关数据集中的近邻对问题进行了详细研究。给出了在网络环境下查询数据集集中近邻对的定理和算法并设计了相应的数据存储结构,针对不同的数据集进行了实验分析。文中的方法可较好地解决网络环境下数据集中近邻对的查询问题。下一步的研究重点主要在于网络环境下的近邻对的近似查询和受限查询等方面。

参考文献:

[1] Kyriakos M, Man L Y, Dimitris P, et al.

(下转第 127 页)

相应的隶属度分别是:

$$\mu_{y_1} = 0.7755, \mu_{y_2} = 0.8817, \mu_{y_3} = 0.3864$$

可以看出,得到的隶属度很好地满足优先级结构。

现在考虑对输出变量 $y_i (i = 1, 2, 3)$ 的隶属度进行放松,以便按照优先级结构对操作变量 $u_i, i = 1, \dots, 5$ 进行优化。假设放松的条件为: $|y'_1| \leq 0.3, |y'_2| \leq 0.15, |y'_3| \leq 0.8$, 得到如下的表达式:

$$|y'_1| \leq 0.3, |y'_2| \leq 0.15, |y'_3| \leq 0.8, u_1, u_5 \in [-1, 1]$$

$\lambda = 3$ 时,得到按照优先级结构排列的解:

$$(u_1, u_4, u_5, u_2, u_3) = (-0.7658, 0.9402, -0.7535, -0.3035, 0.3802)$$

相应的隶属度分别是:

$$(\mu_{u_1}, \mu_{u_4}, \mu_{u_5}, \mu_{u_2}, \mu_{u_3}) = (0.2342, 0.0598, 0.2465, 0.6965, 0.6198)$$

可以看出优先级结构得到了很好的满足。

注释1:在第二层的操作变量进行优化时,必须考虑操作者的要求和第一层优化得到的解(包括通过调整参数 λ 而得到的解),防止优化的不可行性。

注释2:在进行了第一层的优化之后,仅仅通过调整参数 λ , 输出变量和操作变量的优先级结构就可以得到满足,而且这种方法可以节省计算量。但是这样做可能使得某些操作变量到达取值的上界或者下界,不利于优化。

(上接第121页)

Continuous nearest neighbor monitoring in road Networks[C]//Very Large Data Bases (VLDB). Proceeding of the 32nd international conference on very large data bases. Seoul, Korea: [s. n.], 2006: 43-54.

- [2] Xiaopeng X, Mokbel F M, Walid A G. SEA-CNN: Scalable processing of continuous K-nearest neighbor queries in spatio-temporal databases[C]//ICDE, 21st International Conference on Data Engineering (ICDE'05). Tokyo, Japan: [s. n.], 2005: 643-654.

(上接第124页)

过调整并发线程的最大数量,能够有效控制并发用户;
(2)通过限制 EJB 实例池,限制 EJB 构件实例的个数。总之,为了对基于 J2EE 应用系统进行优化设计,需要充分考虑系统的各个方面和环节,进行周密设计,这样才能构建出性能良好的应用系统。

参考文献:

- [1] 苗晓辉. 基于 J2EE 的数据持久化的研究与实现[J]. 计算

2 结束语

对于稳态的 CMMMO 问题,考虑到输出变量的优先级大大高于操作变量的优先级,因此采用宽容完全分层方法将优化过程分为两层进行,在每一层都使用变可行域优化方法进行问题的求解。由此得到的解可以使得较高的优先级有较高的满意度。

参考文献:

- [1] 席裕庚. 复杂工业过程的满意控制[J]. 信息与控制, 1995, 24(1): 14-20.
[2] Sakawa M, Yano H. An interactive fuzzy satisficing method for multi-objective nonlinear programming with fuzzy members [J]. Fuzzy sets and systems, 1989, 30(3): 221-238.
[3] 李少远, 席裕庚. 基于模糊目标和模糊约束的满意控制[J]. 控制与决策, 2000, 15(6): 674-677.
[4] 胡疏达. 实用多目标最优化[M]. 上海: 上海科学技术出版社, 1987.
[5] Li Shaoyuan, YANG Yipeng, Teng Changjun. Fuzzy Goal Programming with Multiple Priorities via Generalized Varying-domain Optimization Method[J]. IEEE Trans. on Fuzzy Systems, 2004, 12(5): 596-605.
[6] Brdy's M, Chen S, Roberts P D. An extension to the modified two-step algorithm for steady-state system optimization and parameter estimation[J]. Int. J. Systems Science, 1986, 17(8): 1229-1243.

- [3] Corral A, Manolopoulos Y, Theodoridis Y, et al. Algorithms for processing k-closest-pair queries in spatial databases [J]. Data and Knowledge Engineering, 2004, 49: 67-104.
[4] 廖巍, 雄, 王钧, 等. 可伸缩的增量连续 k 近邻查询处理[J]. 软件学报, 2007, 18(2): 268-278.
[5] 刘灿, 张德贤. 一种在 KNN 查询处理中预估剪枝阈值的方法[J]. 计算机技术与发展, 2007, 17(2): 89-91.
[6] Sad J R, Urrutia J. Voronoi diagrams [M]. Handbook on Computational Geometry. Ottawa: Elsevier Science, 2000.

机工程, 2007(3): 272-274.

- [2] 成典勤, 崔杜武. J2EE 架构下数据库访问的性能优化[J]. 计算机应用研究, 2006(4): 64-66.
[3] Roman E, Sriganesh R P, Brose G. 精通 EJB[M]. 第3版. 罗时飞译. 北京: 电子工业出版社, 2006.
[4] Spell B. Java 专业编程指南[M]. 邱仲潘, 等译. 北京: 电子工业出版社, 2001.
[5] 杨瑞. J2EE 中提高数据库应用性能的方法[J]. 计算机系统应用, 2003(11): 60-62.