

圆柱面和锥面上数据集的最近邻查询方法

张丽平^{1a}, 李松^{1a}, 郝晓红^{1b}, 王淼^{1a}, 蔡志涛²

(1. 哈尔滨理工大学 a. 计算机科学与技术学院; b. 计算中心, 哈尔滨 150080;

2. 盐城生物工程高等职业技术学校汽车电子工程系, 江苏 盐城 224051)

摘要: 为处理圆柱面和圆锥面上数据集的最近邻查询问题, 提出利用 Voronoi 图进行查询和曲面转换 2 种解决方法。在圆柱面和锥面上构造 Voronoi 图, 利用 Voronoi 图进行查询处理。将圆柱面和锥面转换映射为二维有界平面, 给出转换规则和查询算法。对 2 种方法进行实验分析, 结果表明, 利用 Voronoi 图的方法适合静态数据集的最近邻查询, 曲面转换方法对动态数据集的最近邻查询更有效。

关键词: 最近邻; 反向最近邻; 圆柱面; 圆锥面; Voronoi 图

Nearest Neighbor Query Method of Datasets on Cylindrical and Cone Surface

ZHANG Li-ping^{1a}, LI Song^{1a}, HAO Xiao-hong^{1b}, WANG Miao^{1a}, CAI Zhi-tao²

(1a. School of Computer Science and Technology; 1b. Computation Center, Harbin University of Science and Technology, Harbin 150080, China;

2. Dept. of Automotive Electronic Engineering, Yancheng Biological Engineering Higher Vocational Technology School, Yancheng 224051, China)

【Abstract】 To handle the Nearest Neighbor(NN) query problem on the cylindrical and cone surface effectively, the two methods are proposed. The methods is respectively the method based on the Voronoi diagram and the method of the curve-divert-plane. The Voronoi diagram can be constructed on the cylindrical and cone surface and the query can be done based on the properties of the Voronoi diagram. In the method of curve-divert-plane, the cylindrical and cone surface can be diverted into the 2D plane and the divert-rule, the query algorithm are given. The performances of the two methods are analyzed by experiment. Experimental results show that the method based on the Voronoi diagram can deal with the nearest neighbor query of the static datasets well and the method of the curve-divert-plane can handle the dynamic datasets effectively.

【Key words】 Nearest Neighbor(NN); reverse Nearest Neighbor(NN); cylindrical surface; cone surface; Voronoi diagram

DOI: 10.3969/j.issn.1000-3428.2012.02.009

1 概述

对海量空间数据对象的最近邻(Nearest Neighbors, NN)查询及其变种查询问题的研究, 成为目前空间数据库领域研究的热点和难点。已有的研究成果^[1-6]主要处理处于理想空间或球面空间中的海量空间数据对象, 但对处于圆柱面和锥面上的空间数据对象却无效。本文针对已有方法无法处理圆柱面和圆锥面上空间数据对象的最近邻查询问题, 对圆柱面和圆锥面上数据集中最近邻查询进行研究与分析。

2 Voronoi 图查询

若在静态数据点集中查询圆柱面或圆锥面上给定点 q 的最近邻, 可引进计算几何中的 Voronoi 图来处理。

定义 1 给定一组数据点 $D=\{d_1, d_2, \dots, d_n\} \subset R^2$, 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: $VP(d_i) = \{d | D(d, d_i) \leq D(d, d_j)\}$, $D(d, d_i)$ 为 d 与 d_i 之间的最小距离, 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 $V(D) = \{VP(d_1), VP(d_2), \dots, VP(d_n)\}$ 所定义的图形被称为 Voronoi 图^[7]。

共享相同的棱的 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。

定义 1 中的 $D(d, d_i)$ 可表示欧氏空间中的直线距离, 也可以为非欧空间中的特殊距离。

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^[7]。

性质 2 一个 Voronoi 多边形 $V(p_i)$ 内任何点 $s(s \notin P)$ 到 p_i 的距离必小于该点到其他 Voronoi 生成点的距离^[7]。

根据定义 1 和性质 1、性质 2, 可在圆柱面或圆锥面上直接构造出静态数据点集的 Voronoi 图, 利用 Voronoi 图的特性来处理空间数据对象的最近邻关系。根据查询点是否参与 Voronoi 图的构造可分 2 种情况处理:

情况 1 若查询点 q 是该 Voronoi 的一个生成点, 则根据性质 1 进一步由 q 的邻接生成点中得出 q 的最近邻。该方法较适合查询点 q 是静态的情况。

情况 2 若查询点 q 不是 Voronoi 图的生成点, 则首先确定查询点 q 处在哪个 Voronoi 图单元内, 再返回该 Voronoi 图单元的生成点 p_i , p_i 即为 q 的最近邻。该方法适合查询点是动态的情况。

利用圆柱面或圆锥面上的 Voronoi 图还可进一步求解给定点的反向最近邻和柱面或锥面上的最近对的问题。在查询最近对时, 根据数据集的种类和特点, 可构造一个或多个 Voronoi 图进行处理。利用 Voronoi 图处理动态数据集中近邻查询问题的效率往往很低, 同时向高维扩展也很不易, 空间

基金项目: 黑龙江省教育厅科学技术研究基金资助项目(11551084)

作者简介: 张丽平(1976—), 女, 讲师、硕士, 主研方向: 数据库技术, 数据结构; 李松, 副教授、博士; 郝晓红, 高级实验师; 王淼, 博士研究生; 蔡志涛, 讲师

收稿日期: 2011-06-17 E-mail: zhanglptg@163.com

索引结构的建立和维护代价较大。

3 曲面转换方法查询

由圆柱面和圆锥面的特点可知：圆柱的侧面展开图是以圆柱的底面周长为一边，圆柱的母线为另一边的矩形；圆锥的侧面展开图是以圆锥的底面周长为弧、母线为半径的扇形。由此，可将圆柱面及圆锥面展开成矩形和扇形(如图1和图2所示)，从而可将圆柱面和圆锥面上的空间数据对象点的关系问题转换成二维有界平面上的数据对象点的关系问题。

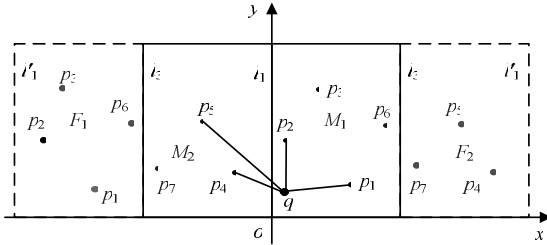


图1 圆柱面的转换平面及扩展平面

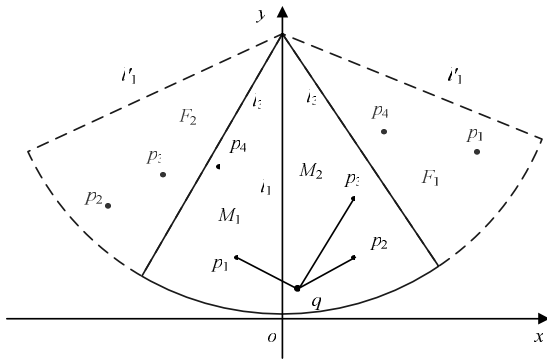


图2 圆锥面的转换平面及扩展平面

定义2 以圆柱面的一条母线为纵轴，以该母线与底面圆周的交点 o 为中心原点，以底面圆周为横轴而构成的坐标系称为柱面坐标系；以圆锥面的垂直于 x 轴的母线为纵轴，以该母线和底面圆周的交点 o 为中心原点，以底面圆周为横轴而构成的坐标系称为锥面坐标系。

转换规则1 由定义2可知，柱面坐标系对应着局部的笛卡儿坐标系。将柱面坐标系中的纵轴(和横轴)对应映射到笛卡儿坐标系的纵轴(和横轴)上，柱面坐标系中的数据点也相应的转换到笛卡儿坐标系中(如图1所示)。

转换规则2 由于锥面坐标系不能完全对应局部笛卡儿坐标系，先以锥面坐标系的纵轴为基线将圆锥面展开成扇形，然后将纵轴投影映射到笛卡儿坐标系的 y 轴上，扇形上的数据点同时做相应映射(如图2所示)。

由于圆柱面及圆锥面都是光滑闭合曲面，当由转换规则将其转成有界平面时，圆柱面及锥面上最近邻关系将不能很好地得到对应。如图1所示，圆柱面上点 p_6 的最近邻是 p_7 ，若单纯转成平面后，点 p_6 的最近邻则成为 p_3 ， p_7 变为 p_6 的最远邻居。为解决此问题，补充以下的定义与规则。

定义3 由圆柱面及圆锥面经过一次转换所成的平面称为实平面，实平面和原始圆柱面及圆锥面具有严格的对应关系。纵轴将实平面分成2个有界平面，称为实半平面。

定义4 以实平面上的2个边界为起始边而向外进行相似性扩展所得的平面称为扩展虚半平面。如图1和图2所示， F_1 是 M_1 的扩展虚半平面， F_2 是 M_2 的扩展虚半平面。

转换规则3 扩展虚半平面和相对象限的实半平面是一致

的；查询点 q 的位置只在实平面内有效， q 的最近邻在实平面和扩展虚半平面内都有效；特殊的，若查询点 q 恒定，数据集中的数据对象动态变化，若多次查询中不需要借助扩展虚半平面，则应保证转换后的纵轴过 q 点。

基于以上讨论，进一步给出近邻关系的查询算法 $CdP_Search()$ 算法如下：

算法1 $CdP_Search()$

输入 圆柱面 Cyc 或圆锥面 Coc ，数据点集 P ，查询点 q
输出 q 的最近邻集 NQ ，反向最近邻集 RNQ

```

begin
     $NQ \leftarrow \emptyset, RNQ \leftarrow \emptyset$  //初始化
    if  $Cyc$  或  $Coc$  没有转换为有界平面  $Cyp$  或  $Cop$  then
        确立坐标系;
         $Cyp \leftarrow Cyc, Cop \leftarrow Coc, P' \leftarrow P, q' \leftarrow q$ ;
    //利用转换规则1、规则2进行信息的转换映射，“ $\leftarrow$ ”表示空
    //间的转换映射。
     $CypA \leftarrow Cyp \cup Cyp'$ ;  $CopA \leftarrow Cop \cup Cop'$ ;
    //将实平面和扩展虚平面合并为有界二维平面
    if 数据点集  $P'$  为静态的 then
        在  $CypA$  或  $CopA$  上以数据点集  $P'$  中的数据点为
        生成点构造 Voronoi 图;
         $NQ \leftarrow Voronoi\_NN(q')$ ;  $RNQ \leftarrow Voronoi\_RNN(q')$ ;
    //利用 Voronoi 图进行最近邻和反向最近邻查询
    if 数据点集  $P'$  为动态的 then
        在  $CypA$  或  $CopA$  上利用 TPR 树组织数据点集  $P'$ 
        中的数据点;
         $NQ \leftarrow TPR\_NN(q')$ ;  $RNQ \leftarrow TPR\_RNN(q')$ ;
    //利用 TPR 树进行最近邻和反向最近邻查询
    return  $NQ, RNQ$ ;
end
    
```

在 $CdP_Search()$ 算法中，当圆柱面或圆锥面给定后，转换过程是一次性的，转换成平面后，查询可根据需要多次进行。故该算法的时间效率主要由在有界平面 $CypA$ 或 $CopA$ 上所采用的具体查询方法所决定。经过转换后，当 P' 为静态数据集时，若在二维有界平面上利用 Voronoi 图进行查询，则构造 Voronoi 图的时间复杂度为 $O(n \log n)$ ，查询最近邻方法(即 $Voronoi_NN(q')$)的时间复杂度为 $O(\log n)$ ，查询反向最近邻方法(即 $Voronoi_RNN(q')$)的时间复杂度为 $O(k \log n)$ 。当 P' 为动态数据集时，若选用 TPR 树进行数据组织和查询，则查询效率主要由动态对象的移动速度和方向、TPR 树的构造和维护代价及查询树的剪枝策略等因素所决定。该算法进一步优化的要点主要在于在有界平面中设计高效率的筛选策略，选取适合不同数据集的查询方法。例如当 P' 为静态数据集，且适合 R 树进行聚类组织时，可利用 R 树取代 Voronoi 图进行最近邻和反向最近邻的查询；当 P' 为动态数据集，且数据点和查询点的动态距离便于计算时，可利用距离函数和时间分裂点进行最近邻的查询和分析。

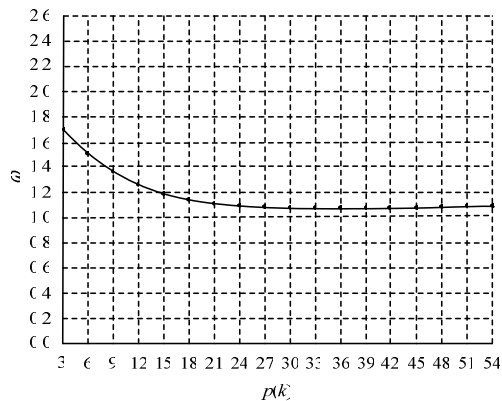
本节的方法经过拓展还可处理圆柱面和圆锥面上数据集中的 k 最近邻查询、最近对查询和组最近邻查询等问题。

4 实验分析

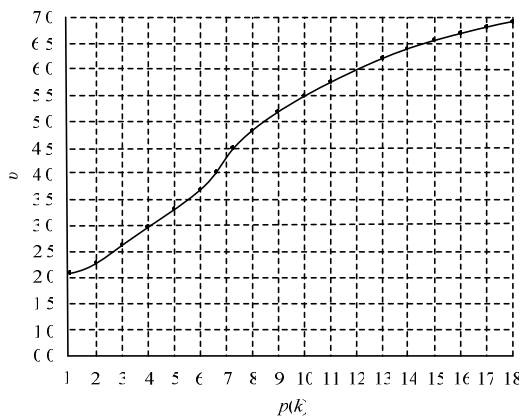
本文提出了利用 Voronoi 图进行查询(简记为 CVs 方法)和曲面转换为平面进行查询(简记为 CDPs 方法)2 种方法。本节着重对这 2 种方法进行性能分析与实验比较。为了分析所提方法的查询效率，本文在 Pentium4, 1.8 GHz CPU, 1 GB 内存, Windows XP 环境下，利用 C++ builder 6.0 软件工具对本文方法进行实验分析。所用实验数据是在圆柱面或圆锥面

上随机生成的模拟数据对象。

图3给出了圆柱面上最近邻查询的实验结果, 圆锥面上的实验结果具有类似结论。



(a)CDPs方法/CVs方法静态数据集



(b)CDPs方法/CVs方法动态数据集

图3 CDPs方法与CVs方法的比较

图3(a)和图3(b)分别表示利用Voronoi图进行查询方法(CVs方法)和曲面转换为平面进行查询方法(CDPs方法), 在静态数据集和动态数据集2种情况下的查询效率比较情况。横坐标表示柱面上数据对象点的数目, 纵坐标 ω 表示CDPs的查询代价和CVs的查询代价的比率, v 表示CDPs的查询效率和CVs的查询效率的比率。由图3(a)可知, 对于静态的

数据集, 当数据量较小时, CDPs方法相对CVs方法的查询代价较大, 这是因为在处理过程中, CDPs方法信息的初始转换较Voronoi图的初始构建代价要大。随着数据量的增大, 2种方法的查询代价逐渐接近, 但CDPs方法较CVs方法的查询代价稍高。由图3(b)可知, 对于动态的数据集, 随着数据量的增多, CDPs方法的查询效率明显优于CVs的查询效率。因此, CDPs更适合处理动态数据集的最近邻查询问题。

5 结束语

已有的方法无法处理圆柱面和圆锥面上的数据集最近邻查询问题, 为了弥补已有方法的空白, 针对柱面和锥面的特点, 提出了利用Voronoi图进行查询和曲面转换为平面进行查询的2种方法。利用Voronoi图的方法还可进一步求解给定点的反向最近邻和圆柱面或圆锥面上的最近对的问题。曲面转换为平面的方法经过拓展还可较好地处理圆柱面和圆锥面上数据集的 k 最近邻查询、最近对查询和组最近邻查询等问题。下一步的研究工作主要为圆柱面或锥面上不确定空间数据对象的最近邻查询。

参考文献

- [1] 李松, 郝忠孝. 移动对象的动态反向最近邻查询技术[J]. 计算机工程, 2008, 34(10): 40-42.
- [2] 李松, 郝忠孝. 基于Voronoi图的反向最近邻查询方法研究[J]. 哈尔滨工程大学学报, 2008, 29(3): 261-265.
- [3] Leong H U, Mamoulis N, Yiu M L. Computation and Monitoring of Exclusive Closest Pairs[J]. IEEE Transactions on Knowledge and Data Engineering, 2008, 20(12): 1641-1654.
- [4] 张丽平, 李松. 数据集中强邻近对的查询方法[J]. 计算机工程与设计, 2008, 29(16): 4353-4355.
- [5] 李松, 郝忠孝. 球面上最近邻空间关系处理方法[J]. 计算机工程, 2010, 36(6): 91-93.
- [6] 张丽平, 李松, 郝晓红. 球面上的 k 最近邻查询算法[J]. 计算机工程, 2011, 37(2): 52-53.
- [7] Sacl J R, Urrutia J. Voronoi Diagrams[M]. Ottawa, Canada: Elsevier Science Press, 2000.

编辑 索书志

(上接第24页)

参考文献

- [1] Bellenguez-Morineau O, Néron E. A Branch-and-bound Method for Solving Multi-skill Project Scheduling Problem[J]. RAIRO-Operations Research, 2007, 41(2): 155-170.
- [2] Bassett M. Assigning Projects to Optimize the Utilization of Employees Time and Expertise[J]. Computers and Chemical Engineering, 2000, 24(2): 1013-1021.
- [3] Valls V, Pérez A, Quintanilla S. Skilled Workforce Scheduling in Service Centers[J]. European Journal of Operational Research, 2009, 193(3): 791-804.

- [4] Ullman J D. NP-complete Scheduling Problems[J]. Journal of Computer and System Sciences, 1975, 10(3): 384-393.
- [5] 陈晶, 潘全科. 求解独立任务调度的离散粒子群优化算法[J]. 计算机工程, 2008, 34(6): 214-215, 218.
- [6] Kashan A H, Karimi B. A Discrete Particle Swarm Optimization Algorithm for Scheduling Parallel Machines[J]. Computers & Industrial Engineering, 2009, 56(1): 216-223.
- [7] Hall N G, Posner M E. Generating Experimental Data for Computational Testing with Machine Scheduling Applications[J]. Operations Research, 2001, 49(6): 854-865.

编辑 张正兴