

doi:10.3799/dqkx.2010.049

基于多重近似索引的空间距离半连接

林伟华¹, 谈晓军², 余艳³, 毛典辉⁴

1. 中国地质大学信息工程学院, 湖北武汉 430074

2. 华中科技大学仿真与数字化工程中心, 湖北武汉 430074

3. 华中科技大学文华学院, 湖北武汉 430074

4. 北京工商大学计算机与信息工程学院, 北京 100048

摘要: 为了更有效地解决基于外部近似索引空间距离半连接效率较低问题, 提出一种基于多重近似索引的空间距离半连接处理方法. 该方法在充分利用多重近似索引结构特征基础上, 推导出在半连接处理中的距离和空间对象数量约束关系, 并在距离半连接算法中利用这些约束关系进行连接处理, 从而减少了进行精过滤处理空间对象的数量. 通过实验分析表明, 基于多重近似索引的距离半连接算法有效, 并且基于多重近似索引比基于外部近似索引的距离半连接效率要高.

关键词: 半连接; 空间距离; 空间索引; 多重近似索引; 空间数据库.

中图分类号: TP311

文章编号: 1000-2383(2010)03-0415-06

收稿日期: 2010-01-15

Spatial Distance Semi-Join Based on Multi-Approximate Spatial Index

LIN Wei-hua¹, TAN Xiao-jun², YU Yan³, MAO Dian-hui⁴

1. Faculty of Information Engineering, China University of Geosciences, Wuhan 430074, China

2. Digital Engineering Research Center, Huazhong University of Science and Technology, Wuhan 430074, China

3. Wenhua College, Huazhong University of Science and Technology, Wuhan 430074, China

4. School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China

Abstract: To improve the efficiency for space distance semi-join based on external approximation index, a method of spatial distance semi-join based on multi-approximate index is proposed. The constraint relationship of distance and the number of spatial objects during processing semi-join are deduced based on taking advantage of characteristics of multi-approximate index structure. And these are used during spatial distance semi-join to reduce the number of spatial objects needed to process during refine filter step. A series of tests and verifications indicate that the new method of spatial distance semi-join is valid and the performance of index based on multi-approximation is more effective than the index based on external approximation during processing spatial distance semi-join.

Key words: semi-join; spatial distance; spatial index; multi-approximate index; spatial database.

0 引言

空间连接是空间数据库中一种重要的空间查询处理, 按照数据集的规模, 可以分为二路连接和多路连接, 按照连接处理形式可以分为连接和半连接等. 通常的空间连接主要是针对空间拓扑关系, 并且主要是指二路连接, 即从两个空间数据集中检索出满足一定空间拓扑关系谓词(如相交、相邻及包含

等)的空间对象. 而空间关系一般有空间拓扑关系、距离关系和方向关系等; 另外, 半连接在空间数据库中应用非常广泛, 它是指两个空间数据集按照某种空间关系连接后, 返回部分连接结果数据, 即在其中一个空间数据集上进行了某种“投影”. 因此, 研究二路距离关系半连接具有一定的理论和现实意义, 如“给定两图层分别是超市和居民区, 找出距离超市最邻近或 k 邻近的居民区”即为典型的距离关

系半连接查询问题。

目前,国内外众多学者对有关距离连接进行了许多研究,并取得了一些成果。这些学者的研究主要有以下几类:(1)基于增量距离连接技术。如 Shin *et al.* (2003)提出了一种自适应多级算法用于 K 距离连接查询和增量距离连接查询,该算法使用双向结点和平面扫描技术对两个距离对象进行快速剪枝。(2)基于非增量距离连接技术。如 Corral *et al.* (2004)提出了 SDR (sorted distances recursive)、PSR(plane-sweep recursive)和 PSI(plane-sweep iterative)3种非增量的分支界限算法用于解决 k 邻近对问题,并指出 SDR 算法关于磁盘访问次数的性能最好,PSR 算法在 k 值较小时处理速度最快,PSI 算法在主存空间较大时性能优于其他两种算法;梁银和张虹(2008)采用深度优先递归搜索策略,用基于距离的平面扫描技术对基于非增量距离连接算法进行了优化。(3)基于最邻近对的连接查询。如 Corral *et al.* (2000)提出了最邻近对的查询算法;Papadopoulos *et al.* (2006)针对约束最近对查询问题进行了讨论,并指出其中基于堆的方法是性能最佳的。(4)基于连接索引的距离连接查询。如 Yeh(1999)提出了基于连接索引的距离连接方法,即仅存储空间对象连接虚点的距离而避免存储所有连接距离,从而减少存储和处理的耗费;肖予钦等(2003)提出了基于距离连接索引的空间距离连接查询方法,该方法考虑了查询处理时的计算费用和存储费用,并采用分步的方式进行实现。(5)基于一种约束条件下的距离连接查询。如 Sankaranarayanan *et al.* (2006)提出在空间网络中距离连接查询算法。

但是,这些用于解决距离连接方法主要存在以下问题:(1)用于在距离连接过程中要么没有利用空间索引,或者利用了空间索引也是基于外部近似索引方式,在海量的空间数据连接处理中,连接处理效率很难进一步提高;(2)半连接处理中针对 top- k 一类的距离半连接处理效率较低,特别是当 k 较大时,效率较低。

因此,本文针对当前主要基于外部近似索引下的距离半连接方法进行改进,提出一种能提高距离半连接查询效率的新方法。该方法的索引结构是基于多种近似的索引,即在原来包含空间对象的外部近似,如最小外包矩形(minimum bounding rectangle, MBR)基础上增加空间对象的内部近似,如最大内接圆(maximum enclosed circle, MEC),这种既包含外部近似表达又包含内部近似表达的索引方

法称为多重近似索引;该方法的距离半连接算法是在 Zhu *et al.* (2005)提出的一种 top- k 空间连接算法基础上进行改进,即将其针对空间拓扑关系连接处理移植到针对基于多重近似索引下距离关系连接处理。该方法在进行距离半连接时,避免了传统算法,并对连接结果空间对象对进行统计排序;同时,由于采用多重近似索引结构,减少了详查阶段访问外存的次数,通过分析验证,基于多重近似索引的距离半连接算法有效,并且基于多重近似索引比基于外部近似索引的距离半连接效率要高。

1 多重近似索引结构

在空间数据索引中 R-tree(Guttman, 1984)是一种比较常用的索引结构,这里以 R-tree 为基本索引原型构建基于多重近似索引(multi-approximate R-tree, MR-tree)(Lin *et al.*, 2008)。对于一棵 m 阶的基于 R-tree 多重近似索引 MR-tree,其结点结构经改进后如下:

叶子结点: $(Count, Level, \langle OI_1, MEC_1, MBR_1 \rangle, \dots, \langle OI_m, MEC_m, MBR_m \rangle)$, (1)

非叶结点: $(Count, Level, \langle CP_1, MBR_1 \rangle, \langle CP_2, MBR_2 \rangle, \dots, \langle CP_m, MBR_m \rangle)$, (2)

式(1)、(2)中,在叶子结点中, OI_i ($i=1, 2, \dots, m$) 为空间对象的标识, MBR_i ($i=1, 2, \dots, m$) 为该对象在 k 维空间中的最小外包矩形, MEC_i ($i=1, 2, \dots, m$) 为空间数据对象的最大内接圆,在非叶子结点中, CP_i ($i=1, 2, \dots, m$) 为指向子结点的指针, MBR_i ($i=1, 2, \dots, m$) 为其子树索引空间范围,即是其所有子结点 MBR 的最小外包矩形。另外, m ($M/2 \leq m \leq M$) 为当前结点的分支数或记录数, M 为每个结点存放的最大记录数, $M/2$ 为结点存放的最小记录数。 $Count$ ($Count \leq m$) 标识当前结点孩子的数量, $Level$ ($Level \geq 0$) 标识当前结点在树中的层数,其中, $Level=0$ 表示叶子结点。其他参数含义与 R-tree 结构中相同。如图 1 所示为一棵 4 阶 MR-tree。

从图 1 的 MR-tree 的索引结构可以看出,基于多重近似 MR-tree 与基于外部近似 R-tree 索引在结构上不同主要是:MR-tree 索引结构的叶子结点除了空间对象的 MBR 和空间对象标识之外,还包含有空间对象的 MEC;而 R-tree 索引结构的叶子结点中只包含有空间对象的 MBR 和空间对象标识。

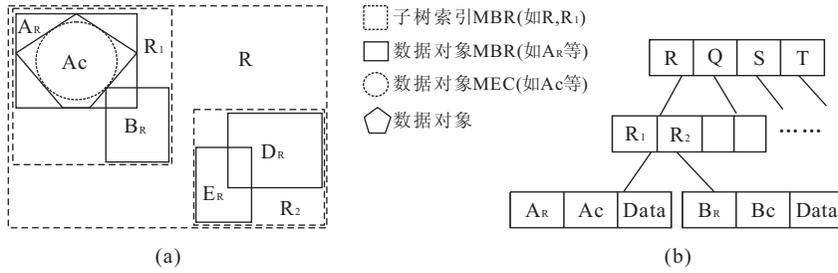


图 1 MR-tree 索引结构示意图

Fig. 1 The index architecture of MR-tree

(a)为空间数据的平面示意;(b)为其对应的 MR-tree 结构

2 半连接的约束关系

2.1 距离关系约束

(1)基本距离定义. 在基于多重近似索引结构中,由于空间对象 O_i 、空间对象 MEC 和空间对象 MBR 之间隐含着一定的距离约束关系,如它们之间的最小距离、最大距离和最小最大距离. 利用这些距离进行剪枝处理时,可以过滤掉大部分无效对象,从而减少遍历索引空间范围和从外存中提取空间对象数据次数.

定义 1: 在 n 维欧氏空间中,有两个矩形 $R_1(A, B)$ 和 $R_2(C, D)$, 其中 $A = (a_1, a_2, \dots, a_n)$, $B = (b_1, b_2, \dots, b_n)$, $C = (c_1, c_2, \dots, c_n)$, $D = (d_1, d_2, \dots, d_n)$, A, B 分别是 R_1 主对角线上两个端点, C, D 分别是 R_2 主对角线上两个端点, 则 R_1 到 R_2 间最小距离为:

$$\min_{RR}(R_1, R_2) = \sqrt{\sum_{i=1}^n r_i^2},$$

$$\text{其中 } r_i = \begin{cases} c_i - b_i, & c_i > b_i, \\ a_i - d_i, & a_i > d_i, \\ 0, & \text{其他.} \end{cases} \quad (3)$$

定义 2: 在 n 维欧氏空间中,有两个矩形 R_1 和 R_2 , 设矩形 R_1 的顶点为 (P_1, P_2, \dots, P_n) , 矩形 R_2 的顶点为 (Q_1, Q_2, \dots, Q_n) , $Dist_{PQ}(P_i, Q_j)$ ($i, j = 1, 2, \dots, n$) 为两点 P_i 与 Q_j 间距离, 则 R_1 到 R_2 间最大距离为:

$$\max_{RR}(R_1, R_2) = \max(Dist_{PQ}(P_i, Q_j)),$$

$$i, j = 1, 2, \dots, n. \quad (4)$$

定义 3: 在 n 维欧氏空间中,有两个矩形 $R_1(A, B)$ 和 $R_2(C, D)$, 设矩形 R_1 和 R_2 任意一边 L_i 和 L_j , $\max Dist_{LL}(L_i, L_j)$ 为两边 L_i 与 L_j 间的最大距离, 则 R_1 到 R_2 间最小最大距离为:

$$\min \max_{RR}(R_1, R_2) = \min(\max Dist_{LL}(L_i, L_j)). \quad (5)$$

定义 4: 在 n 维欧氏空间中,设两空间对象的矩形为 R_1 和 R_2 , R_1 的 MEC C 的圆心 O , 半径为 r , $\min \max Dist_{PR}(O, R)$ 为 O 到 R 的最小最大距离(张奋等, 2007), 则 MEC $C(O, r)$ 到 MBR R_2 最小最大距离可定义为:

$$\min \max_{CR}(C, R_2) = \begin{cases} \min \max_{RR}(R_1, R_2), & R_1 \text{ 与 } R_2 \text{ 相交,} \\ \min \max Dist_{OR}(O, R_2) - r, & R_1 \text{ 与 } R_2 \text{ 不相交.} \end{cases} \quad (6)$$

定义 5: 在 n 维欧氏空间中,设两空间对象的 MEC C_1 与 C_2 的圆心分别为 O_1, O_2 , 半径分别为 r_1, r_2 , MEC $C_1(O_1, r_1)$ 到 MEC $C_2(O_2, r_2)$ 最小最大距离可定义为:

$$\min \max_{CC}(C_1, C_2) = \begin{cases} 0, & C_1 \text{ 与 } C_2 \text{ 相交或包含} \\ Dist_{PP}(O_1, O_2) - (r_1 + r_2), & \text{其他.} \end{cases} \quad (7)$$

定义 6: 给定空间对象 O_1 及其 MBR R_1 、MEC C_1 , 空间对象 O_2 及其 MBR R_2 、MEC C_2 , 令 O_1 与 O_2 间的最小最大距离:

$$\min \max_{O_1, O_2} = \min(\min \max_{RR}(R_1, R_2), \min \max_{CR}(C_1, R_2), \min \max_{CR}(C_2, R_1), \min \max_{CC}(C_1, C_2)). \quad (8)$$

(2)距离关系约束. 一般地, 给定一距离范围 $[Dmin, Dmax]$, 在进行距离半连接时, 需要查询满足该距离范围内的空间连接对. 若面空间数据集 S_1 中任一对象为 O_1 , 面空间数据集 S_2 中任一对象为 O_2 , $\min Dist_{RR}$ 和 $\max Dist_{RR}$ 分别为空间对象间最小距离与最大距离的定义, 空间对象 O_1 和 O_2 间需满足的距离关系:

$$Dmin \leq \min Dist_{RR}(O_1, O_2) \leq \max Dist_{RR}(O_1, O_2) \leq Dmax. \quad (9)$$

在距离半连接查询中,针对两个空间数据集均有多重近似索引的情况,即需要通过索引树来约束或减小搜索空间范围,其中,在索引树结构的叶节点的数据域中主要有空间对象的 MBR、MEC 信息,根据空间对象及其 MBR、MEC 间的距离关系,两个空间对象的 MBR、MEC 间需满足的距离关系:

$$\begin{cases} D_{\min\max} \geq D_{\min}, \\ \min_{RR}(R_1, R_2) \leq D_{\max}. \end{cases} \quad (10)$$

在两个空间数据集中,索引树结构中各结点的 MBR 包络于它各子结点的 MBR,因此,这些结点的父结点 MBR 与子结点 MBR 之间也隐含满足一定的距离关系.若多边形空间数据集 S_1 中多重近似索引树一结点 MBR 为 R_1 ,多边形空间数据集 S_2 中多重近似索引树一结点 MBR 为 R_2 ,其父结点的 MBR 分别为 R_1^p 和 R_2^p ,它们之间存在的距离关系为:

$$\begin{cases} \min_{RR}(R_1^p, R_2^p) \leq \min_{RR}(R_1, R_2), \\ \max_{RR}(R_1^p, R_2^p) \geq \max_{RR}(R_1, R_2). \end{cases} \quad (11)$$

另外,由空间对象与其 MBR 间距离关系知:

$$\begin{cases} \min_{RR}(R_1, R_2) \leq \min_{Dist_{RR}}(O_1, O_2) \leq D_{\max} \\ \max_{RR}(R_1, R_2) \geq \max_{Dist_{RR}}(O_1, O_2) \geq \\ D_{\min\max} \geq D_{\min}. \end{cases} \quad (12)$$

因此,可以得到基于多重近似索引各结点的父结点 MBR 应满足的距离关系是:

$$\begin{cases} \min_{RR}(R_1^p, R_2^p) \leq D_{\max}, \\ \max_{RR}(R_1^p, R_2^p) \geq D_{\min}. \end{cases} \quad (13)$$

这样,在基于多重近似索引的两个空间数据集中,根据式(13),通过计算和比较各父结点 MBR 间距离关系,可以减小距离半连接查询中索引树的搜索空间;同时,根据式(10),通过计算和比较各空间对象间的 MBR、MEC 之间距离关系,许多空间对象不需要通过从外存中调入空间对象的实际数据即可判断是否满足距离条件,因此可减少 I/O 次数来提高连接查询效率;最后,在精过滤处理阶段,根据式(9),通过从外存中调入空间对象的实际数据进行计算,最终判断两空间对象是否满足距离条件.

2.2 空间对象数据约束

为了尽早剪掉不可能是连接结果的分支,提高连接查询效率,需要对两空间对象集的索引树进行连接查询时就能提供空间对象数量的约束信息.

定义 7: 设 e 为空间数据集 A 的索引树 R_a 上一中间结点, C 为该结点的度, $e.level$ 为 e 在索引树 R_a 上的高度(如 e 为叶结点,则 $e.level=0$),则 e 的

子树中包含空间对象数量的上限为(Zhu *et al.*, 2005):

$$\maxnum(e) = C^{e.level}. \quad (14)$$

定义 8: 若 e 为空间数据集 A 的索引树 R_a 上叶结点,令 $count(e)$ 为 e 与空间数据集 B 满足某种关系 θ (如距离关系等)在 B 中对象的数量;若 e 为一中间结点, $count(e)$ 为 e 与空间数据集 B 满足某种关系 θ (如距离关系等)在 B 中对象的数量上限,其定义形式如下(Zhu *et al.*, 2005):

$$count(e) = \sum_{e_i \in R_b, \text{且 } e_i \theta e} \maxnum(e_i). \quad (15)$$

若有空间数据集 A 和空间数据集 B , 设 e 为空间数据集 A 的数据对象, $count(e)$ 为 e 与空间数据集 B 满足某种关系 θ (如距离关系等)在 B 中对象的数量, e' 为空间数据集 A 的索引树中结点 MBR, $count(e')$ 为 e' 与空间数据集 B 满足某种关系 θ (如距离关系等)在 B 中对象或其索引树中结点 MBR 的数量,如果满足如下不等式 $count(e') < count(e)$, 则满足与 e' 在空间数据集 B 中某种关系 θ 在 B 中对象连接对或其索引树中结点 MBR 连接对均可删除而不用进行遍历.

3 距离半连接算法

在传统的空间半连接算法中,由于没有充分利用启发信息进行剪枝以减少搜索路径,其算法效率较低.因此,改进的距离半连接算法主要是在遍历过程中利用分支界限的方法尽量减少搜索路径.其算法是在 Zhu *et al.* (2005) 的拓扑空间连接算法基础上进行改进.

给定两空间数据集 A 和 B , 其索引树根结点分别为 R_a 和 R_b , k 为 A 与 B 最多满足距离半连接中 B 的空间对象数目,若 a 为 A 中一空间对象, $[b_1, b_2, b_3, b_4]$ 为 B 中 4 个对象,且 a 与 $[b_1, b_2, b_3, b_4]$ 满足距离关系 θ ,若 a 是距离半连接的一个结果,其用一个三元组表示为 $(a, a.count, a.IL)$, 即 $(a, 4, [b_1, b_2, b_3, b_4])$. 基于多重近似索引的距离半连接算法过程如下:

Step 1: 连接两空间数据集 A 和 B 的索引树的根结点 R_a 和 R_b , 计算其各子结点在满足距离关系 θ 的入口列表,并将含有 R_a 中结点入口的连接三元组按照 $count()$ 大小顺序插入到一列表堆(Heap)中.

Step 2: 从列表堆中取出一三元组 $(e, e.count, e.IL)$, 如果 e 和 $e.IL$ 中数据都是实际对象,若 k 个

对象已经输出,则程序退出,否则输出 $(e, e.count, e.IL)$;如果 e 和 $e.IL$ 中数据有非实际对象,则转到 Step 3.

Step 3: 设 e 指向子结点其中一数据域 n (若 e 为实际对象,则 $n=e$); e 和 $e.IL$ 中任意一数据域为 e_i , 设 e_i 指向子结点其中一数据域 n_i (若 e_i 为实际对象,则 $n_i=e_i$); 则对每一对 (n, n_i) 进行在距离关系 θ 连接计算;对每一满足距离连接条件的入口对 (e', e'_i) 进行统计计算出 $e'.count$ 和 $e'.IL$, 其中, e' 和 e'_i 分别是 n 和 n_i 子结点中一数据域.

Step 4: 如果计算出来的某一 $e'.count$ 值大于至今为止已找到的第 k 个对象的 $count$ 值,则把三元组 $(e', e'.count, e'.IL)$ 按大小顺序插入到列表堆(Heap)中,若 e' 和 $e'.IL$ 中数据都是实际对象,则更新剪枝条件并转到 Step 2; 如果小于则直接转到 Step 2.

4 验证分析

为了验证与评估基于多重近似索引与外部近似索引在距离半连接上的效率,通过建立基于不同索引结构的距离半连接查询算法进行实验测试. 实验的硬件和软件环境是:所有的实验在 PC 机上进行,其中 PC 机的配置是 Pentium 4 & 2.4GHz 的 CPU, 256 MB 内存, Windows XP professional 2002 操作系统;基于两种索引结构下的不同距离半连接算法均用 C++ 在编程环境 Microsoft Visual C++ 6.0 上实现. 实验效果的评价主要是通过测算 CPU 计算时间和 I/O 的耗费, CPU 耗费时间采用 MBR 到 MBR、MBR 到 MEC 以及 MEC 到 MEC 进行距离计算的次数来进行衡量;同样, I/O 的耗费采用需要进行精处理的空间连接对的数量来进行衡量.

为了充分体现基于不同索引下的距离半连接性能,随机产生包含 MBR 和 MEC 的两空间数据对象集进行距离半连接实验,两数据集包含的空间对象数分别为 200 与 250, 两数据对象集的坐标 X, Y 分布范围分别 $(0 \sim 1500)$ 和 $(0 \sim 500)$, 两数据对象集的密度分别为 0.002 51 和 0.088 75, 两数据对象集中空间对象的长宽比均在 1~2 之间(表 1 和表 2).

基于多重近似索引的 MR-tree 和基于外部近似的 R-tree 的距离半连接算法在连接结果数量 k 为 1、不同距离上限的距离计算次数及需精过滤对

象对数比较情况如表 1 所示. 从表 1 中可以看出,随着连接距离的上限不断增大,基于 R-tree 和 MR-tree 索引下连接查询的距离计算次数也不断增大,这是由于距离上限增大后,能满足距离条件的 MBR 间连接对就相应地增多,因而连接距离计算的次数也相应增多;在相同连接距离上限的情况下,基于多重近似的 MR-tree 索引比基于外部近似的 R-tree 索引的距离计算次数要多. 但是,由于基于多重近似的 MR-tree 索引提供了更严格的连接条件限制,在

表 1 不同距离上限和索引结构下距离半连接性能比较
Table 1 The performance comparison of distance semi-join by the difference distance upper-limit and index structure

距离上限	距离计算次数(次)		需精过滤对象对数(对)	
	R-tree	MR-tree	R-tree	MR-tree
100	8 380	18 844	84	54
200	13 754	32 027	141	90
400	26 036	61 940	228	119
600	41 188	99 088	239	137
800	58 382	141 635	312	171
1 000	77 438	189 170	326	178
1 300	97 344	238 935	276	178
1 500	102 026	250 640	183	104
1 700	102 950	252 950	43	32
2 000	102 950	252 950	0	0

表 2 不同连接结果 k 和索引结构下距离半连接性能比较
Table 2 The performance comparison of distance semi-join by the difference join result k and index structure

k	距离计算次数(次)		需精过滤对象对数(对)	
	R-tree	MR-tree	R-tree	MR-tree
1	16 692	39 060	3	2
2	16 692	39 060	3	2
3	16 692	39 060	3	2
4	16 692	39 060	3	2
5	16 692	39 060	3	2
6	16 692	39 060	5	3
7	19 124	45 140	6	4
8	19 124	45 140	13	8
9	19 124	45 140	22	13
10	21 322	50 635	30	17
11	21 322	50 635	37	18
12	21 322	50 635	41	19
13	21 322	50 635	47	20
14	21 322	50 635	51	23
15	21 322	50 635	56	25
16	21 642	50 955	61	27
17	21 642	50 955	62	28
18	22 992	54 330	68	32
19	22 992	54 330	77	37
20	22 992	54 330	81	40

粗过滤阶段即可判定空间对象连接对是否满足距离条件,因而在同一连接距离上限情况下,从外存中提取连接的空间实体对象对数在基于多重近似的 MR-tree 索引下比在基于外部近似的 R-tree 索引下要少。

基于多重近似索引的 MR-tree 和基于外部近似的 R-tree 的距离半连接算法在同一距离上限为 400、不同连接结果数量 k 、不同距离上限的距离计算次数和需精过滤对象对数比较情况如表 2 所示。从表 2 中可以看出, k 越大,距离计算次数、需精过滤对象对数在不同索引下均在增大,并且 k 越大,基于 R-tree 索引下的需精过滤对象对数的增幅明显比基于 MR-tree 索引下要大。虽然基于 MR-tree 的距离半连接距离计算次数要比基于 R-tree 的距离半连接要大,耗费了一定的 CPU 时间,但 CPU 所耗费的时间与需精过滤对象 I/O 所耗费的时间相比要小得多,因此,在处理空间数据量较大的 top- k 一类的距离半连接时,当 k 较大时,基于多重近似处理效率较基于外部近似索引要高。

5 结语

讨论了空间距离半连接的基本含义,并指出当前有关距离连接研究存在的缺陷,然后提出了一种基于多重近似索引的空间距离半连接处理方法。该方法一是避免了直接对连接结果空间对象对进行统计排序;二是充分利用多重近似索引结构特征,即该索引结构既包含空间对象的外部近似又包含空间对象的内部近似,并在此基础上推导出在半连接处理中的距离和空间对象数据约束关系,然后利用这些约束关系在距离半连接中减少进行精过滤处理空间对象的数量,从而降低了 I/O 次数,提高了连接处理效率。

References

- Corral, A., Manolopoulos, Y., Theodoridis, Y., et al., 2000. Closest pair queries in spatial databases. *ACM SIGMOD Record*, 29(2): 189–200. doi: 10.1145/335191.335414
- Corral, A., Manolopoulos, Y., Theodoridis, Y., et al., 2004. Algorithms for processing k -closest-pair queries in spatial databases. *Data and Knowledge Engineering*, 49(1): 67–104. doi: 10.1016/j.datak.2003.08.007
- Guttman, A., 1984. R-trees: a dynamic index structure for

- spatial searching. The ACM SIGMOD Int. Conf. on Management of Data Boston, Massachusetts, 47–57.
- Liang, Y., Zhang, H., 2008. Method for multi-way spatial distance join query processing. *Computer Applications*, 28(1): 155–158 (in Chinese with English abstract).
- Lin, W. H., Wu, Y. G., Tan, X. J., et al., 2008. Multi-approximate index based on R-tree for massive spatial data. *Proceedings of Information Technology and Environmental System Science, Jiaozuo*, 1: 574–579.
- Papadopoulos, A. N., Nanopoulos, A., Manolopoulos, Y., 2006. Processing distance join queries with constraints. *Computer Journal*, 49(3): 281–296. doi: 10.1093/comjnl/bxl002
- Sankaranarayanan, J., Alborzi, H., Samet, H., 2006. Distance join queries on spatial networks. In: ACM, ed., proceedings of the 14th annual ACM international symposium on advances in GIS, New York, U. S. A., 211–218. doi: 10.1145/1183471.1183506
- Shin, H., Moon, B., Lee, S., 2003. Adaptive and incremental processing for distance join queries. *IEEE Transactions on Knowledge and Data Engineering*, 15(6): 1561–1578. doi: 10.1109/TKDE.2003.1245293
- Xiao, Y. Q., Zhang, J., Chen, L., et al., 2003. Online spatial distance queries processing based on the multi-step implementation of DJI. *Journal of National University of Defense Technology*, 25(6): 5–9 (in Chinese with English abstract).
- Yeh, T. S., 1999. Spot: distance based join indices for spatial data. In: ACM, ed., proceedings of the 7th ACM international symposium on advance in GIS, New York, U. S. A., 103–109. doi: 10.1145/320134.320161
- Zhang, F., Pan, M. S., Zou, B. J., 2007. Nearest neighbor queries of spatial object based on SR-tree. *Computer Engineering and Applications*, 43(4): 173–175 (in Chinese with English abstract).
- Zhu, M. L., Papadias, D., Zhang, J., et al., 2005. Top- k spatial joins. *IEEE Transaction on Knowledge and Data Engineering*, 17(4): 567–579. doi: 10.1109/TKDE.2005.65

附中文参考文献

- 梁银,张虹,2008.一种多路空间距离连接查询处理方法. *计算机应用*, 28(1): 155–158.
- 肖予钦,张巨,陈萃,等,2003.基于 DJI 分步实现的联机空间距离查询处理. *国防科技大学学报*, 25(6): 5–9.
- 张奋,潘梅生,邹北骥,2007.基于 SR-树的空间对象最近邻查询. *计算机工程与应用*, 43(4): 173–175.