

基于 4CDRS 的空间连接查询

潘红岩¹, 郝忠孝^{1,2}

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

2. 哈尔滨工业大学 计算机科学与技术学院, 黑龙江 哈尔滨 150001)

摘要: 提高空间查询效率是空间数据库方面备受关注的-个问题,而空间操作交运算又是最常用的谓词. 空间操作交运算的主要对象是多边形和线,针对空间线对象,通过存储线的偏移方向,来保存线的走向的这样一种对象压缩的近似栅格表示,并且用不同颜色代表了对象和单元格之间不同的相交类型. 这种新的栅格近似即四色偏移栅格标识图(4CDRS). 将4CDRS用作查询处理中的过滤器,并设计出相应的测试算法.

关键词: 四色偏移栅格标识图(4CDRS); 空间连接; 空间数据库; 查询

中图分类号: TP311 **文献标志码:** A **文章编号:** 1007-2683(2010)04-0045-04

Spatial Join Query Based on 4CDRS

PAN Hong-yan¹, HAO Zhong-xiao^{1,2}

(1. School of Computer Science & Technology, Harbin University of Science and Technology, Harbin 150080, China;

2. School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Efficient evaluation of spatial queries is an important issue in spatial database. Among spatial operations, spatial join is very useful, intersection being the most common predicate. The main objects of spatial joins are polygons and polylines, special for polylines, saving the direction of the polylines, and using different colors to represent different types of spatial join that the objects join with the cells. The new raster approximation is Four-Color Directional Raster Signature (4CDRS). It is used as a filter in the Multi Step Query Processor, and we design the appropriate algorithm.

Key words: Four-Color Directional Raster Signature (4CDRS); spatial joins; spatial database; query

0 引言

地理信息系统的一个重要部分就是数据. 从数据结构上来说,栅格数据结构是描述空间地理信息的最简单的结构,又称为网格结构(raster或grid cell)或象元结构(pixel). 栅格数据结构实际就是像元阵列,每个像元由行列确定它的位置. 在栅格结构中,点用一个栅格单元表示;线状地物则用沿线走向

的一组相邻栅格单元表示;面或区域用记有区域属性的相邻栅格单元的集合表示. 任何以面状分布的对象(土地利用、土壤类型、地势起伏、环境污染等),都可以用栅格数据逼近表示.

本文提出了一种新的栅格近似,这种新的近似方法在多步查询处理中作为一种过滤器来处理空间连接,包括2种数据类型和它们之间不同的空间连接. 已经有很多专家和学者研究了空间连接,也有很多方法来处理空间连接操作. 考虑到点、线、面是空

收稿日期: 2009-06-03

基金项目: 黑龙江省自然科学基金(F2007-01)

作者简介: 潘红岩(1983—),女,硕士研究生,E-mail: phy311@163.com;

郝忠孝(1940—),男,教授,博士生导师.

间数据库中最常用的3种数据类型,有9种不同类型的空间连接.由于实用性和复杂性,多边形连接已经有很多研究,但线的空间连接相对来说研究很少.为了提高效率,文[1]为空间连接处理提出了一种三步体系(多步查询处理器)记为MSQP.

其主要目标是通过过滤器减少剩余需要比较的空间对象的数量.本文在此基础上提出一种四色偏移栅格标识图(4CDRS).

1 基本概念

在实际应用中,空间数据也常常被称为图形数据.空间数据有以下三种基本类型,即点、线、面[2].而本文的主要研究对象是线,线不仅表示线上各点在空间的位置,而且还有长度,即表示它在空间中延伸的范围.

空间数据查询的实质是按照一定条件对空间对象的空间数据和属性数据进行查询,以形成一个新的数据子集[3].

传统的处理空间连接操作分两步处理.第一步,使用一个空间访问方法(SAM)来减少搜索空间.其中,最小边界矩形(MBR)是SAM方法经常使用的.MBR提供一个候选对集是所求结果的一个超集.第二步,从磁盘读取第一步得到的候选集,并处理它们.在以上两步过程中,搜索和从磁盘上读取空间对象会耗费I/O时间,计算出具体结果会耗费CPU时间.

多步查询处理(MSQP)处理空间连接操作分三步处理.第一步,候选选择.第二步,过滤器应用.第三步,精确几何比较[4].

2 四色偏移栅格标识图(4CDRS)算法设计

以下介绍4CDRS的主要特征、产生4CDRS的算法,还提出一种转换为4CDRS的简略算法,以及使用4CDRS判断两个对象是否重叠的算法.

2.1 4CDRS的特点

4CDRS定义:用不同颜色表示空间对象与单元格的相交类型,并通过存储空间对象的偏移方向来保存空间对象的大致走向的近似栅格表示.

表1列出4CDRS单元类型.图1为一个用4CDRS表示线的例子.事实上,就是将对象进行分段,之后判断分段与单元的相交类型.当分段与单元横向相交时,用水平单元类型来表示;当分段与单元

纵向相交时,用垂直单元类型来表示;当分段与单元不相交时,用空单元类型来表示;而其余的分段与单元相交的情况,用不确定单元类型来表示.当计算一个4CDRS时,仅需要判断单元与对象相交的类型.因此,每一种单元类型都很快被计算,并且象点、线这样的对象都可以用4CDRS来表示.

表1 4CDRS单元类型

单元类型	描述
Horizontal	与对象横向相交
Vertical	与对象纵向相交
Inconclusive	与对象其它类型的相交
Empty	与对象不相交

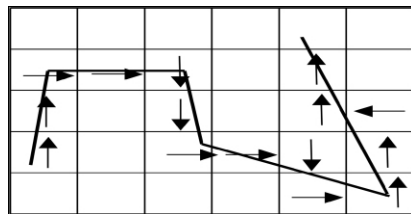


图1 4CDRS表示线的例子

2.2 4CDRS的生成算法

1) 计算包围对象的 $MBR/2^n$.其中 $MBR/2^n$ 是坐标,是 2^n 倍数的MBR的扩展,如文[5]所示.将这个区域拆分成 $p \times q$ 个单元格.单元格由横轴和纵轴上一系列的点共同表示,如横轴由 $2^n a_0, 2^n a_1, \dots, 2^n a_p$ 一系列点来表示;纵轴由 $2^n b_0, 2^n b_1, \dots, 2^n b_q$ 一系列点来表示.单元格创建的过程中,同时也实现了对象的分段.

2) 单元格已建立好,单元格中每段可能有四个方向,即上、下、左、右.下面通过使用生成算法计算每段的偏移向量,进而生成4CDRS.

该生成算法的核心思想,首先将每个单元格都初始化为空类型,之后根据单元格内组成该段的每个点的坐标,以及它与它相邻的各点坐标之间的换算,从而确定该单元格的类型.由于栅格中单元的最大数量是一个定值 $p \times q$,该生成算法的时间复杂度为 $O(p \times q)$.

4CDRS生成算法描述如下:

输入:待处理的单元格

输出:处理后确定的该单元格的类型

Begin

for each cell do

begin

```

    cell. type = empty;
  end
  for each approximation vector do
    begin
      if vector. type = up or vector. type =
down then
        cell. type = vertical;
      else if vector. type = right or vector. type =
left then
        cell. type = horizontal;
      else
        cell. type = inclusive;
      end
    end
  End

```

2.3 使用4CDRS 测试相交

当判断两个4CDRS图时,两个图要有相同的单元大小. 否则要作一个面积转换. 当必须作面积转换的时候,通过 2^m 个单元来实现. 文[6]中给出了一个面积转换算法.

如果需要则进行面积转换^[6], 然后开始判断两个单元的4CDRS是否相互重叠. 仅仅位于图的MBR内部的单元才被处理^[8]. 当比较的两个单元中一个是横向类型,另一个是纵向类型时,则相交的结果为 yes; 当比较的两个单元中有一个是空类型时,则测试结果为 no; 其他情形,测试结果均为 inconclusive. 在比较过程中,一旦得到 yes,则测试停止且返回测试相交的结果; 如果所有结果均为 no, 则返回不相交的结果; 其他情况,均为不确定. 表2显示两个单元之间的比较结果. 图2显示比较结果的实例. 图3显示两个对象测试相交的图.

表2 比较两个单元格可能的结果

	Empty	Inconclusive	Vertical	Horizontal
Empty	no	no	no	no
Inconclusive	no	inconclusive	inconclusive	
Vertical	no	inconclusive	inconclusive	yes
Horizontal	no	inconclusive	yes	inconclusive

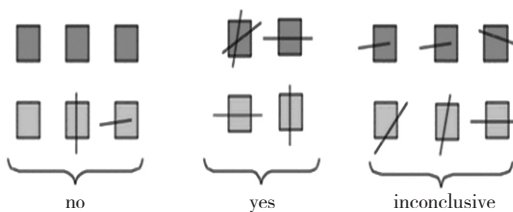
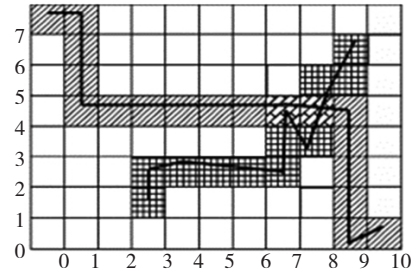
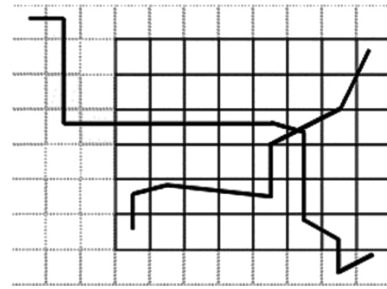


图2 比较结果的实例



(a) 测试前生成4CDRS



(b) 测试后相交 MBR

图3 两对象测试相交实例

4CDRS 测试相交算法描述如下:

输入: 已经确定类型的单元格

输出: 是否相交的结果

Begin

result = no;

for each pair of the MBR intersection cells do

begin

if cell1. type $\cap$ cell2. type $\neq \phi$ then

if ((cell1. type = = inconclusive) or
(cell2. type = = inconclusive))

or ((cell1. type = = horizontal) and
(cell2. type = = horizontal))

or ((cell1. type = = vertical) and
(cell2. type = = vertical)) then

result = inconclusive;

else

result = yes;

exit

end

return result;

End

3 结 语

空间数据库技术正逐步取代传统文件,空间查询优化则成为研究的焦点. 本文提出了一种新的栅格近似(四色偏移栅格标识图 - 4CDRS) 是针对线

的空间连接设计的,本文优点是减少了访问磁盘的次数、以及连接查询结果产生的时间,而且还可以用于表示点、线.同时,同样的算法可以用来评估包含这 2 种数据类型的连接谓词.将 4CDRS 用作查询处理中的过滤器.实验结果表明了 4CDRS 的效率,平均来说,与原始数据集相比,4CDRS 的磁盘访问次数降低了 44%,而整体的运行时间也降低了 38%.其查询的磁盘访问时间和磁盘访问次数比没有图像的查询处理有相应的降低^[10].

参 考 文 献:

- [1] ORENSTEIN J A. Spatial Query Processing in an Object Oriented Database System [C] //Proceedings of the 1986 ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 1986: 326 - 336.
- [2] NEWMAN W M, SPROULL R F. Principles of Interactive Computer Graphics [M]. 2th ed. New York: McGraw-Hill Book Company, 1979.
- [3] BR INKHOFF T, KR IEDEL H P, SCHNEIDER R. Comparison of Approximations of Complex Objects Used for Approximation based Query Processing in Spatial Database Systems [C] //Proceeding International Conference on Data Engineering, Washington: IEEE Computer Society, 1993: 40 - 49.
- [4] BR INKHOFF T, KR IEDEL H P, SCHNEIDER R, et al. Multi-step Processing of Spatial Joins [C] //Proceedings of the 1994 ACM SIGMOD International Conference on Management of Data, New York: ACM Press, 1994: 197 - 208.
- [5] LEONARDO GUERREIRO AZEVEDO, RODRIGO SALVADOR MONTEIRO. Polyline Spatial Join Evaluation Using Raster Approximation [C] //Kluwer Academic Publishers, 2003: 315 - 336.
- [6] ZIMBRAO G, SOUZA J M. A Raster Approximation for Processing of Spatial Joins [C] //Proceedings of the 24th International Conference on Very Large Databases Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 1998: 558 - 569.
- [7] KOTHUR IR K, RAVADA S. Efficient Processing of Large Spatial Queries Using Interior Approximations [C] //Proceedings of the 7th International Symposium on Spatial and Temporal Database, London: Springer-Verlag, 2001: 404 - 424.
- [8] 陈 琳,杜友福,王元珍. MBR: 基于 MBR 的空间关系模型 [J]. 计算机工程与应用, 2002, 18(5): 76 - 78.
- [9] 楚 放. 空间查询语言和查询优化 [D]. 北京: 北京大学, 2007: 57 - 63.
- [10] 王能斌. 数据库系统原理 [M]. 北京: 电子工业出版社, 2004: 232 - 244.

(编辑: 温泽宇)

(上接第 44 页)

参 考 文 献:

- [1] 乌 兰. 缓释化肥的研究现状与进展 [J]. 西北民族大学学报: 自然科学版, 2004, 25(54): 64 - 67.
- [2] 李浙江,王金信,连玉朱,等. 壳聚糖在农药领域中的应用和前景 [J]. 农药研究, 2005, 26(6): 28 - 31.
- [3] 杨 郈,杨铁金. 多孔壳聚糖膜的制备研究 [J]. 齐齐哈尔大学学报, 2007, 23(1): 30 - 32.
- [4] 吴国杰,黎碧娜. 壳聚糖膜的研制 [J]. 现代化工, 2000, 20(6): 34 - 37.
- [5] 陈 强,吕伟娇,张文清. 醛交联剂对壳聚糖复合膜性能的影响 [J]. 华东理工大学学报: 自然科学版, 2005, 31(3): 398 - 402.
- [6] 强志斌. 壳聚糖膜的制备与应用初步 [D]. 上海: 东华大学, 2006: 24 - 27.
- [7] 徐 蔚,曹 毅,张 纪. 壳聚糖微球释药机制的研究 [J]. 昆明医学院学报, 2002, 23(1): 7 - 10.
- [8] 朱华跃,肖 玲. 戊二醛交联对壳聚糖 PVA 共混膜结构和性能的影响 [J]. 浙江海洋学院学报: 自然科学版, 2005, 24(2): 126 - 129.
- [9] 陈 强,吕伟娇. 壳聚糖缓释肥料包膜的制备和结构表征 [J]. 高分子材料科学与工程, 2005, 21(4): 216 - 219.

(编辑: 温泽宇)