

基于 ArcSDE 的空间拓扑关系连接方法研究

文元桥¹,周春辉²

WEN Yuan-qiao¹,ZHOU Chun-hui²

1.武汉理工大学 航运学院,武汉 430063

2.武汉大学 测绘遥感信息工程国家重点实验室,武汉 430079

1.Navigation College,Wuhan University of Technology,Wuhan 430063,China

2.State Key Lab of Information Engineering in Surveying,Mapping and Remote Sensing,Wuhan University,Wuhan 430079,China

E-mail:wenyuanqiao@yahoo.com.cn

WEN Yuan-qiao,ZHOU Chun-hui.Spatial topological relation join methods based on ArcSDE.Computer Engineering and Applications,2010,46(5):117-119.

Abstract: This paper introduces the standard topological relation model,and analyses the characteristics of the spatial topological relation join.According to the spatial query technique and the topological predicates supported by ArcSDE,three methods are proposed for implementing topological join with ArcSDE,including Spatial Constraints Query Topological Join(SCQTJ),Predicates Nested Loop Topological Join(PNLTJ),and Filter-Refining 2 Phases Topological Join(FR2PTJ).Finally the three methods are tested by experiments,and the PNLTJ method shows the best performance.

Key words: topological relation;spatial join;ArcSDE;optimal method

摘 要: 简要介绍了标准的拓扑关系模型,分析了空间拓扑关系连接的特点。根据 ArcSDE 提供的空间查询技术和拓扑关系谓词,提出了在 ArcSDE 平台下实现拓扑连接的三种方法,包括基于空间约束查询的拓扑连接方法(SCQTJ)、基于谓词的嵌套循环拓扑连接方法(PNLTJ)和过滤精化两阶段拓扑连接方法(FR2PTJ)。最后通过试验对三种方法进行检验,试验结果表明 PNLTJ 方法具有最优的性能。

关键词: 拓扑关系;空间连接;ArcSDE;优化方法

DOI:10.3778/j.issn.1002-8331.2010.05.035 文章编号:1002-8331(2010)05-0117-03 文献标识码:A 中图分类号:P208

1 引言

空间数据库具有空间的连续性和很强的空间相关性,使得空间查询比传统的数据库查询更复杂。最基本的空间查询操作包含点查询(Point Query)、窗口查询(Window Query)、最邻近查询(Nearest Neighbor Query)和空间连接(Spatial Join)等,其中空间连接是最常用的一类查询。空间连接是多个数据集之间基于空间谓词的联合查询,其执行代价是空间数据库中最昂贵的^[1],它对于空间数据库的性能影响比窗口查询更加重要。拓扑关系是一类重要的空间关系,基于拓扑关系谓词的空间连接称为拓扑连接。

ArcSDE(Spatial Database Engine)是美国环境系统研究所(Environmental Systems Research Institute Inc,ESRI)推出的空间数据库解决方案,它在现有的关系型或对象关系型数据库管理系统的基础上进行空间扩展,可以将空间数据和非空间数据集成在目前绝大多数商用的数据库管理系统(DBMS)中。在ESRI的产品线中,ArcSDE是一个核心产品,负责其空间数据存储、访问,它在GIS应用程序和空间数据库之间提供了一个开放的接口。主要研究ArcSDE中空间拓扑连接的实现策略,并

对各种策略的性能进行评估,为实现基于ArcSDE的高效的空间拓扑查询与分析提供依据。

2 拓扑模型与空间拓扑连接

2.1 拓扑关系模型

空间查询语言中的空间关系主要包括拓扑关系、方向关系和度量关系等。其中拓扑关系是指那些在拓扑变换(包括平移、旋转、缩放等)中保持不变的性质,主要描述空间对象间是否相邻、包含、重叠。拓扑关系作为一种重要的空间关系,广泛应用于空间分析和查询中。空间邻接、包含关系、穿越、落入、定位、缓冲等空间查询都直接与空间拓扑关系的判断紧密相关。OGC(Open GIS Consortium)在其《简单要素访问规范》^[2]中定义了适合在关系型数据库或对象-关系型数据库中对地理要素进行存储和访问的简单要素模型,以及一组空间拓扑关系算子;在此基础上,OGC还制订了地理信息系统的《空间查询标准》^[3],指出空间数据库应该实现GIS所需要的基本空间运算和分析功能。

有关空间拓扑关系的定义研究较多,主要包括四交模型^[4-6]和九交模型^[7-8]。OGC的《简单要素访问规范》在定义空间关系

基金项目:国家自然科学基金青年科学基金项目(No.40805063/D0512)。

作者简介:文元桥(1975-),博士,副教授,研究方向:海事地理信息系统。

收稿日期:2009-09-16 修回日期:2009-12-08

操作时同时采纳了四交模型^[4](Four-Intersection Model, 4IM)和维度扩展的 9 交模型^[7](Dimensionally Extended Nine-Intersection Model, DE-9IM),并对每个 4IM 中的关系定义给出了 DE-9IM 的等价表示。简单要素访问规范定义的 8 个常用的关系谓词包括 Crosses、Disjoint、Within、Contains、Equals、Touches、Intersects、Overlaps,其中 4IM 中定义的 Disjoint、Touch、Cross、Within 和 Overlap 五个谓词在文献^[5]中被证明是互斥和完备的,为了使用的方便又扩充了另外的三个谓词。所指的“空间对象”,即指符合《简单要素访问规范定义》的地理要素。

2.2 空间拓扑连接

连接是指从两个关系的笛卡尔积中选取属性间满足一定条件的元组。空间连接是从两个空间关系的笛卡尔积中,选取空间属性间满足一定空间关系的元组。如果这里的空间关系是拓扑关系,则这种空间连接称为空间拓扑连接。例如,“查询长江流过的城市”,“查询北京、天津境内的水库”等等,这样的查询涉及两个以上的空间表,而且需要判断空间对象之间的拓扑关系,本质上是一个空间拓扑连接查询。这两个查询可以用 SQL 语句(1)、(2)表示。其中 Crosses、Intersects 都是拓扑关系谓词,shape 表示空间对象的几何属性。

```
SELECT C.Name FROM Cities C,Rivers R
WHERE Crosses(C.shape,R.shape)AND C.Name='长江' (1)
```

```
SELECT R.name FROM Cities C,Reservoirs R
WHERE Intersects(C.shape,R.shape)AND(C.name='天津'OR C.Name='北京') (2)
```

传统的关系连接(Relation Join)算法针对非空间数据(如数值、字符串、时间等)设计,这些数据被排序,数据值相邻的对象在排序序列中也是相邻的。但对于多维空间数据(如点、线、面等),没有一种排序方法能保持空间上相近的对象在序列中也相邻^[6]。因此,传统关系连接技术并不适用于空间数据库中的空间连接。在 ArcSDE 中,提供了多种机制可以用来实现拓扑连接查询,不同的实现方式可以得到不同形式的结果,查询所消耗的代价也不相同。

3 基于 ArcSDE 的空间拓扑连接方法

3.1 ArcSDE 的拓扑关系谓词

ArcSDE 在实现基本的空间关系谓词时遵循了 OGC 的《简单要素访问规范》,该规范中的 8 个空间谓词,ArcSDE 实现了其中的 7 个。Intersects 谓词没有直接实现,但 Intersects 可以用 Disjoint 谓词的否定来实现,即 $\text{Intersects} \Leftrightarrow \text{Not Disjoint}$ 。ArcSDE 中拓扑关系谓词的定义符合 DE-9IM 模型。图 1 是 Overlaps、Crosses、Within、Touches、Disjoint 等谓词所表达的拓扑位置关系。图 1 中在每种位置关系里面,左边部分表达的是线对象之间的拓扑关系,右边部分表达的是面对象之间的拓扑关系。此外还有点、多点、多线、多面等几何形状,它们相互之间均可能存在更复杂的拓扑关系,这里不再赘述。

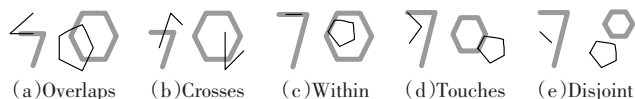


图 1 ArcSDE 中 5 个基本的拓扑谓词表达的空间拓扑关系示意图

3.2 基于空间约束查询的拓扑连接方法

ArcSDE 并未直接提供拓扑连接的实现方法。如何实现拓扑空间连接取决于开发者对空间连接的认识,不同的实现将产

生不同的结果,其效率也可能大不一样。ArcSDE 基于标准的游标模型实现了一个基本的空间数据访问框架,如图 2 所示。ArcSDE 在关系查询模型的基础上,加入了对空间约束的支持。将一个空间关系中所有对象的几何属性作为约束条件,对另一个空间关系进行查询,从而实现拓扑空间连接的方法,称为基于空间约束的拓扑连接方法(Spatial Constraints Query Topological Join, SCQTJ)。

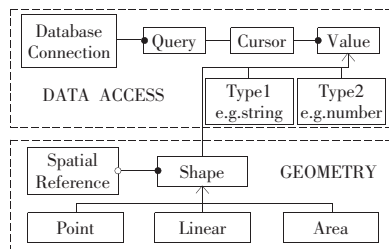


图 2 ArcSDE 中空间数据访问的基本框架图

ArcSDE 实现的空间约束类型有矩形交叠、公共点、公共边、线交叉等共 21 个,SCQTJ 方法可以将一个表中满足和一些空间对象有某种拓扑关系的对象查找出来。其优点是实现简单,缺点则是性能较差且不够灵活。例如,用 SCQTJ 方法实现查询(2),可以知道有哪些水库位于北京、天津两市境内,但如果要知道哪些水库位于天津境内、哪些水库位于北京境内,则 SCQTJ 方法无法用一个查询完成。

3.3 基于谓词的嵌套循环拓扑连接方法

基于谓词的嵌套循环拓扑连接方法(Predicates Nested Loop Topological Join, PNL TJ)是一种更灵活的连接方法。这种方法采用 3.2 节中的数据访问方法获取两个空间表的数据,然后用一个嵌套循环遍历两个空间表中的对象,在内层循环中用拓扑关系谓词对内外层循环中当前的两个空间对象进行测试,如果空间拓扑关系成立,则输出这一对空间对象。可以用于测试的拓扑关系谓词包含 7 个(如 3.1 节所述),实际上涵盖了 OGC 规范要求的所有拓扑关系。如空间关系 R、S,采用 PNL TJ 方法进行连接时主要步聚如算法 1 所示。

算法 1 (PNLTJ)

```
READ(R);READ(S);
For(all r ∈ R)Do
  For(all s ∈ S)Do
    If(r & s satisfy the join condition)Then
      Output(r,s);
    End If
  End For
End For
```

3.4 过滤精化两阶段拓扑连接方法

过滤-精化两层模型^[10]是被广泛使用的空间连接方法,在 Oracle Spatial 中空间查询常常采用这种模型进行优化。借助 ArcSDE 的空间拓扑关系谓词(3.1 节)和空间约束查询方法(3.2 节),可以实现过滤精化两阶段连接方法(Filter-Refining 2 Phases Topological Join, FR2PTJ)。第一步过滤是借助空间表的最小外包矩形(Minimum Boundary Rectangle, MBR)以及空间索引,进行简单的求交(Intersects)判断,以过滤掉不满足条件的空间对象,减小进一步求精时的计算量;第二步从得到的候选结果中再进行利用拓扑关系谓词进行精确的几何计算从而得到最终的结果,如图 3 所示。

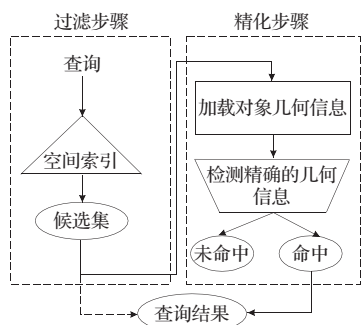


图3 过滤精化两阶段拓扑连接方法

过滤阶段由于使用了空间索引,而且是基于矩形的求交判断,代价很小但可以过滤掉很多不满足条件的对象,因此得到了优化。基于 ArcSDE 的空间约束的查询进行过滤时,对空间约束的选择起着至关重要的作用。SCQTJ 实际上是选择的对象本身,过滤和精化过程没有分开,因而效率较低。空间表 R、S 用 FR2PTJ 方法进行拓扑连接,过滤时可以选择 R 表本身的 MBR 作为过滤器去过滤 S; 也可以选择用 S 作过滤器去过滤 R; 或者选择两个表中各空间对象的 MBR 进行求交判断,达到过滤的效果。

4 试验与分析

4.1 试验环境和数据

试验在一台 PC 机上进行,CPU 为 4 核 Q6600,主频 2.4 GHz,内存 4 GB。操作系统为 Windows 2003 SP2,Oracle Spatial 10g 做为后台数据库,ArcSDE9.2 for oracle10g 为空间数据引擎,开发工具为 VC++2005。试验数据用中国 1:25 万基础地理数据的一部分,包括湖北、湖南两省的县市级行政区划数据,以及湖北省的国家级道路数据共三个空间表。三个空间的数据量为:湖北省县市区界数据 1.9 MB,湖南省县市区界数据 2.0 MB,湖北省的国家级道路数据 0.4 MB。

4.2 试验结果与分析

分别用 SCQTJ,PNLTJ,FR2PTJ 三种方法,实现两个空间连接查询:(1)查询湖北省与湖南省相邻的县市,并指出相邻关系,用 SQL 语句表达为:SELECT hbc.Name,hnc.Name FROM HubeiCounties hbc,HunanCounties hnc WHERE Touches(hbc.Shape,hnc.Shape);(2)查询湖北省内有国道通过的县市,用 SQL 语句表达为:SELECT hbc.Name FROM HubeiCounties hbc,HubeiNatRoad hnr WHERE Intersects(hbc.Shape,hnr.Shape)。

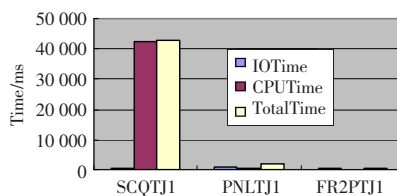
对查询(1),SCQTJ 方法每次用湖北省的一个对象作为约束,采用约束关系为“SM_ET_OR_AI”(Edge Touch or Area Intersect,边相邻或面相交),对湖南省进行带约束的查询,需要执行的查询次数即为湖北省的对象个数;PNLTJ 方法则直接获得两个表的数据,然后采用二重循环,用“SE_shape_is_touching”谓词进行空间关系测试,共需测试次数为两个表的对象个数的乘积;FR2PTJ 在获取两个表的数据时加入空间约束,在取湖北省的数据时用湖南省的 MBR 做为约束,在取湖南省的数据时则反之,精化阶段则同 PNL TJ 方法一样,采用二重循环和“SE_shape_is_touching”谓词进行空间关系测试。对查询(2),SCQTJ,PNLTJ 方法和查询(1)采用的方法类似;FR2PTJ 方法中,过滤阶段则因为湖北省县市区界和湖北省国道两个表的 MBR 范围大致相同,采用带约束的空间查询起不到过滤作用,选择用每一个

对象的 MBR 先进行相交(Intersects)测试的过滤方法。

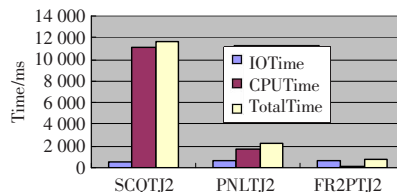
试验中统计每一种方法消耗的 IO 代价、CPU 代价和总代价,分别用 IOTime,CPUTime 和 TotalTime 表示。从空间数据库中读取数据的阶段计入 IO 代价,判断空间拓扑关系的过程计入 CPU 代价,二者的和作为总代价。试验的最终结果如表 1 所示,图 4 采用直方图进行直观表示,其中 SCQTJ1,PNLTJ1,FR2PTJ1 表示查询 1 的三种试验结果,SCQTJ2,PNLTJ2,FR2PTJ2 表示查询 2 的三种试验结果。

表 1 空间拓扑连接试验性能比较表

	SCQTJ1	PNLTJ1	FR2PTJ1	SCQTJ2	PNLTJ2	FR2PTJ2
IOTime	531	1 172	375	546	640	687
CPUTime	42 563	688	110	11 141	1 641	94
TotalTime	43 094	1 860	485	11 687	2 281	781



(a) 三种策略执行查询 1 的性能比较



(a) 三种策略执行查询 2 的性能比较

图 4 空间拓扑连接试验性能比较图

由试验结果可以看到,SCQTJ 在执行空间连接时性能最差,PNLTJ 较好,FR2PTJ 最好。FR2PTJ 方法中,合理利用空间索引,利用空间表和空间对象的 MBR,对拓扑连接中的空间数据进行预处理,提前去除掉大部分不满足条件的元组,这是过滤优化方法的核心思想。

总的来看,SCQTJ 连接只需要空间数据引擎实现窗口查询接口即可实现,不需要调用具体的对象间的拓扑关系判断接口,实现较简单,但性能不够理想;PNLTJ 则不依赖于空间数据引擎的窗口查询接口,只需要对象间的拓扑关系判断接口即可实现;FR2PTJ 则需要同时需要两个接口,性能也最佳。

5 结语

空间数据库引擎技术是目前在系统集成中广泛应用的中层技术,在空间数据库应用解决方案中的一种实现。ArcSDE 是较成熟的、使用最为广泛的空间数据库引擎之一,分析了采用 ArcSDE 提供的技术和接口来实现多空间表之间拓扑连接查询的多种方法,并用试验对三种方法的性能进行了测试,结果表明过滤精化两阶段拓扑连接方法(FR2PTJ)是一种有效的拓扑空间连接方法。进一步地分析,这种方法只要求空间数据库引擎提供较常用的接口和操作,如获取图层或对象的 MBR 和窗口查询,以及空间拓扑谓词算子即可,因此也可以在大多数空间数据库引擎上方便地进行实现。