

Oracle Spatial 空间数据操作缺陷 及解决方案

刘南 刘仁义 冯杭建 谢炯 李伟

(浙江大学 GIS 重点实验室, 杭州 310028)

摘要 在进行空间数据叠加分析的实际应用中,发现 Oracle Spatial 处理空间数据的一个缺陷,并得到 Oracle 公司的确认(bug 号:3146244)。针对 Oracle Spatial 处理海量地理空间数据的不足,提出了坐标概化叠加算法,根据空间叠加分析精度要求,选择合适的容许限值,并以此开发了一个扩展模块。另外还给出了应用该扩展模块进行 1:400 万全国市县行政图与全国土地利用图的空间叠加分析的实例。采用 Oracle Spatial 处理出错的数据进行测试,其结果验证了该方法的正确性和有效性。

关键词 Oracle Spatial 缺陷 坐标概化

中图法分类号: P208 **文献标识码:** A **文章编号:** 1006-8961(2004)12-1504-05

A New Limitation of Oracle Spatial for Spatial Data Management and Its Solution

LIU Nan, LIU Ren-yi, FENG Hang-jiang, XIE Jiong, LI Wei

(Keylab of GIS, Zhejiang University, Hangzhou 310028)

Abstract By an application of analyzing for spatial data overlay, a bug was discovered in Oracle Spatial and confirmed by Oracle Corporation. The bug ID is 3146244. In allusion to the disadvantage of Oracle Spatial in processing massive spatial data, an algorithm of "Coordinate-consolidation Overlay" was presented. The appropriate threshold value was selected according to analysis of spatial data overlay. An extended module was developed based on this algorithm in Oracle Spatial. The correctness and validity of the solution were testified by using the spatial data in which bug happened for Oracle Spatial. In the end, an application example in overlay analyses of nationwide municipality-county map and land-use map at scale 1:4000000 was given by using the extended module of Oracle Spatial.

Keywords Oracle Spatial, bug, coordinate-consolidation

1 引言

Oracle 数据库系统内置模块 Oracle Spatial 采用对象关系模型和高效的空间索引技术(R-tree),将地图数据存贮到 Oracle 数据库中,实现海量地图数据管理^[1]。Oracle Spatial 空间数据模型由几何元素、几何体、图层 3 个层次组成。几何元素包括点、线和多边形 3 种类型,以此构成几何体,多个几何体组成一个图

层。利用 Oracle Spatial 提供的空间算子及 SQL 语言,可对地图数据进行编辑、查询和其他复杂的空间分析。一般情况下,Oracle Spatial 提供的空间操作不会出现异常情况,但当多边形坐标点数达到 6 万个或多边形内环数超过约 200 个时,Oracle Spatial 将会出现错误,系统与 Oracle 数据库的连接将被中断。该错误对于大型 GIS 系统的 TB 级海量空间数据管理无疑是一个非常严重的问题。经与 Oracle 公司技术支持部门的多次研讨和大量测试,最终由 Oracle 公司将此错误

确定为一个新的 bug (bug 号:3146244)。本文在介绍 Oracle Spatial 的 bug 发现过程的基础上,提出了坐标概化叠加解决方案,以此开发了一个功能模块,并通过实际应用得到验证^[2,3]。

2 Bug 的确定及产生原因

2.1 测试环境与数据

系统测试采用操作系统 Red Hat AS 2.1,数据库服务器 Oracle 9.2.0.4,客户端 Red Hat 7.1;测试数据为 1:25 万土地利用图(图幅号 gh4804),全国 1:400 万土地利用图,全国 1:400 万市县图。

从 1:25 万土地利用图幅 gh4804 中剪切出一块矩形空间范围(如图版 I 图 1):X 方向(106.726 11, 107.271 7),Y 方向(31.644 558, 32.002 064),并进行编辑操作,使用的 Oracle Spatial 函数为 SDO_GEOM.SDO_INTERSECTION,其函数原型如下:

```
SDO_GEOM.SDO_INTERSECTION(
geom1 IN MDSYS.SDO_GEOMETRY,
geom2 IN MDSYS.SDO_GEOMETRY,
tol IN NUMBER
) RETURN MDSYS.SDO_GEOMETRY;
```

参数解释:geom1,参与叠加操作的几何对象 1;geom2,参与叠加操作的几何对象 2;tol,容差;返回值,返回一个 geom1 与 geom2 的交集几何对象,如图 1 所示。

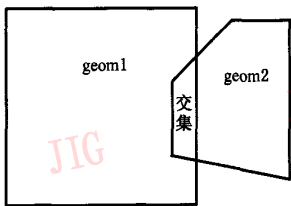


图 1 geom1 与 geom2 的交集几何对象

2.2 剪切操作及系统错误

执行剪切操作的 SQL 语句为

```
SELECT SDO_GEOM.SDO_INTERSECTION(
a.shape,b.shape,0.00000005)
FROM gh4808_landuse a,z_over_geometry b;
```

通过上述 SQL,正常情况应能求出类似图 1 所示空间表 gh4808_landuse 和 z_over_geometry 的相交部分,即从 gh4808_landuse 图幅中剪切出一块矩形区域来。但是,Oracle Spatial 出现了问题,在

SQL Plus 中执行该语句出现的错误信息如下:

```
ORA-21779: duration not active
ORA-06512: at "MDSYS.SDO_3GL", line 439
ORA-06512: at "MDSYS.SDO_GEOM", line 3125
ORA-06512: at "MDSYS.SDO_GEOM", line 3174
```

“ORA-21779”是一个十分严重的错误,导致与 Oracle 的连接被中断,“ORA-06512”错误说明操作的空间数据不符合 Spatial 的 Geometry 定义。

2.3 原因分析

不符合 Spatial 的 Geometry 对象定义规则而导致的空间数据操作失败通常有 3 种情况,一是“ORA-13356”错误原因是坐标点的序列上有重复点的存在,即存在连续两个坐标点相同的情况;二是“ORA-13349”错误原因是多边形边界自相交;三是“ORA-13367”错误是多边形内环的坐标顺序不正确,在 Spatial Release2 中规定多边形外环的坐标顺序为逆时针,如果多边形含有内环,那么内环的坐标顺序为顺时针。通过函数

```
SDO_GEOM.VALIDATE_GEOMETRY
```

检查确认表 gh4804_landuse 和 z_over_geometry 内的几何对象全部符合 Spatial 的 Geometry 对象定义规则。对 gh4804_landuse 中的记录通过二分法过滤,找出如图版 I 图 2 所示的一个 Geometry,单独将该 Geometry 与剪切矩形(红色方框)进行求交操作时,出现异常。

通过分析可知,图版 I 图 2 中出错的 Geometry 是一个“超复杂”的多边形几何对象,有 249 个内环,外环为 89 980 多个坐标点。在图幅 gh4804_landuse 中,比上述出错 Geometry 复杂的几何对象还有许多,有些 Geometry 几何多边形的坐标点数量达到数十万个,这是土地利用图图斑的特点。经分析试验,内环的数量似乎不是数据库与客户端连接中断的主要原因,原因应该在于整个 Geometry 几何对象的坐标点数量“超极大”。任意挑出这些“超复杂”几何对象中的一个进行叠加操作,都出现异常。测试表明,出错 Geometry 的坐标点数量由于系统运行环境的不同而不是一个定值,如果坐标点数量在 6 万个左右,出现异常的概率极高,坐标点数量大于 6 万个,一般都会出错。

3 坐标分别优化叠加算法

产生 3146244 号 bug 的原因为 Geometry 的坐

标数量“超极大”，如果能够通过一种方法，将其坐标数量减少到 6 万个以内再求交，就不会出错。当然，不是去求解每个 Geometry 的坐标点的个数，而是利用异常捕捉方法挑选出异常 Geometry，然后将它们存入到一张新表里，在所有正常 Geometry 做完叠加后，最后处理这些异常 Geometry。坐标分别优化叠加方案的基本思路是：使用异常捕捉机制，将出现异常的 Geometry 存储到新表里，采用坐标分别优化方法，减少这些异常 Geometry 的坐标点数量，然后再采用 SDO_INTERSECTION 函数来求交，最后与直接采用 SDO_INTERSECTION 对“正常”Geometry 求交的结果合并成一个完整的空间数据表。

3.1 坐标分别优化

坐标分别优化方法，如图 2 所示，根据空间叠加分析精度要求，选择合适的容许限值 tol_x 和 tol_y ，一般情况下， $tol_x = tol_y = tol$ 。如果 $\Delta X (X_1 - X_0) < tol_x$ ，并且 $\Delta Y (Y_1 - Y_0) < tol_y$ ，则去除 A 坐标点。

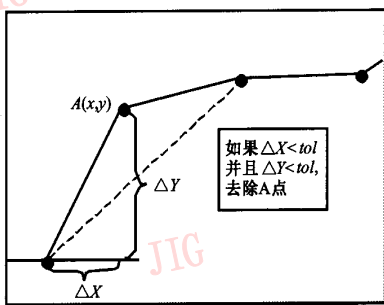


图 2 坐标概化算法

具体算法如下：

```

CG_GeoPolygon * pGeoPolygon =
(CG_GeoPolygon *) pObj;
If (pGeoPolygon->GetPolygon()->
GetElementType() == ET_Polygon)
{
  CG_Line * pLine = (CG_Line *) pPolygon->
  GetRing(i); //取出一个环
  CG_Line * pNewLine = new CG_Line();
  CG_ElementPoint * pPt =
  new CG_ElementPoint();
  pLine->GetPoint(0)->Clone(pPt); //line 的 第
  一点
  Ox = pPt->GetX();
  Oy = pPt->GetY();
  pNewLine->AddPoint(pPt);
  int ptCount = pLine->GetPointCount();
  for(int j = 1; j < ptCount - 1; j++)
  {
    double x = pLine->GetPoint(j)->GetX();
  
```

```

double y = pLine->GetPoint(j)->GetY();
if (fabs(x-Ox) < tol && fabs(y-Oy) < tol) //小
于容差去除该点
{
  Generalized = true;
  continue;
}
else
{
  pPt = new CG_ElementPoint(x,y);
  pNewLine->AddPoint(pPt);
  Ox = x;
  Oy = y;
}
}
pPt = new CG_ElementPoint();
pLine->GetPoint(ptCount-1)->Clone(pPt);
pNewLine->AddPoint(pPt); //添加 line 的 最
后一点
  
```

由于采用面向对象的几何模型，数据分别优化以环(多边形)而不以弧段作为单元进行存储计算^[4,5]。

3.2 分别优化叠加流程

坐标分别优化叠加可以归纳为 8 个步骤，如图 3 所示。

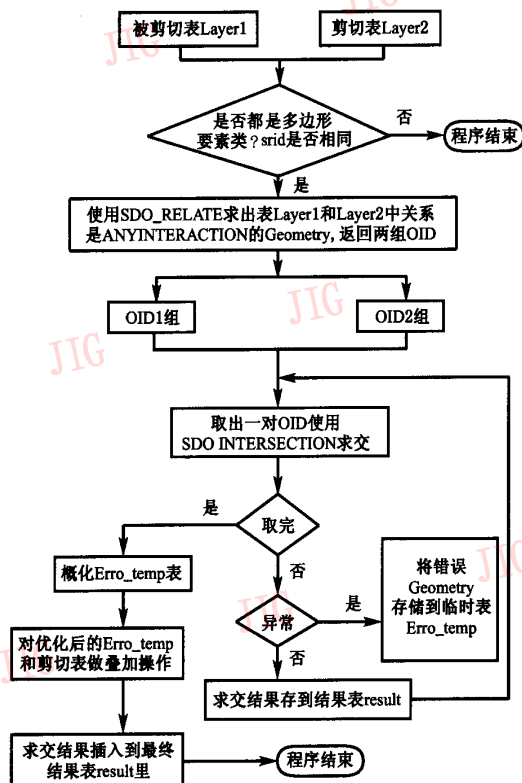


图 3 分别优化叠加解决方案流程

(1) 判断两个 featureClass 的 shape 类型，两个

featureClass 类型都是多边形,且空间参考 SRID 相同;

(2) 根据用户提交的属性字段建立新表,即结果表 result,属性字段的信息来自两个源 featureClass;

通过 sdo_relate 运算求出两个表中空间关系是 ANYINTERACT 的 GEOMETRY 的 OID,返回的记录集中有两个字段:OID1 和 OID2,任何一对 OID1、OID2 的空间关系都是 ANYINTERACT。通过该步骤,大大缩小了需要进行 SDO_INTERSECTION 空间叠加操作的 GEOMETRY 的数量,提高了效率。

(3) 将相交部分(即两个多边形的交集)存入新表,即结果表里。在此步骤中,需要将出现异常的“超复杂”多边形记录下来。需要记录“一对”OID,一个是“超复杂”多边形的 OID,另一个是剪切几何的 OID。该对 OID 对应多边形使用 SDO_INTERSECTION 函数进行求交时,将会出现异常。通常情况下,剪切图形一般是比较简单的多边形,即非“超复杂”对象,如规则多边形、县市的边界等,“超复杂”几何多边形一般出现在土地利用要素类中。

(4) 取出异常多边形表 ERROR_GEOMETRY 中的每一对几何多边形,按照一定的规则做概化操作,具体算法如图 2 所示。

(5) 对概化后的 ERROR_GEOMETRY 表中几何多边形进行求交操作,并将结果插入到最终的结果表 result 中;

(6) 删除由于 SDO_GEOM.SDO_INTERSECTION 运算产生的点、线、多边形点、多边形线;

(7) 将最终结果表 result 注册到 SDM 中。

4 测试结果与应用

基于坐标分别优化叠加算法开发了一个 Oracle Spatial 空间叠加分析扩展模块。对图版 II 图 2 中的图层 gh4804_landuse 与剪切矩形进行空间叠加运算,结果如图版 II 图 3 所示,可以看出叠加精度高、效果好,实现了“超复杂”几何对象的空间操作。该扩展模块成功应用到 1:400 万全国市县行政图(图版 II 图 4(a))与 1:400 万全国土地利用图(图版 II 图 4(b))的空间叠加分析中,图版 II 图 4(c)为叠加结果。

5 结 论

Oracle Spatial 是 Oracle 数据库系统处理地理空间数据的专用模块,随着 Oracle Spatial 应用的普及和深入,对其海量数据(TB 级)处理能力和性能的要求越来越高。坐标分别优化叠加方法扩展了 Spatial 对“超复杂”几何对象处理的能力,具有极强的实际应用意义。由于坐标分别优化叠加方法没有脱离 Spatial,而是将 Spatial 处理出错的几何多边形采用异常处理机制挑选出来,在叠加精度范围内通过坐标分别优化,仍使用 SDO_INTERSECTION 几何函数进行求交运算,属于精确计算,精度较高,但是降低了效率,速度受到了影响。针对此不足,进一步的工作是对其进行完善。

参 考 文 献

- 1 Oracle. Oracle Spatial[DB/OL]. <http://www.Oracle.com/products/spatial>. 2003-05-15.
- 2 ESRI. ArcSDE[EB/OL]. <http://www.esri.com>. 2003-02-17.
- 3 IBM. Informix Spatial DataBlade [EB/OL]. <http://www.ibm.com/products/spatial>. 2002-02-10.
- 4 刘仁义,刘南.一种基于 RDBMS 的空间数据管理模型及应用[J].中国图形图象学报,2000,5(10):825~829.
- 5 郭菁,郭薇,胡志勇.大型 GIS 空间数据库的有效索引结构 QR-树[J].武汉大学学报,2003,28(3):306~310.



刘 南 1944 年生。教授、博士生导师。中国地理学会常务理事,浙江省地理学会理事长。主要研究领域为面向对象的空间数据库理论。



刘仁义 1960 年生。教授、GIS 博士。主要研究领域为面向对象的空间数据库理论、时空数据模型及 WebGIS 平台。
E-mail:liuren yi@mail.hz.zj.cn



冯建安 1979 年生。2004 年于浙江大学获 GIS 硕士学位。研究方向为海量空间数据处理。



谢 炯 1977年生。现在浙江大学攻读GIS专业博士研究生。研究方向为海量空间数据库理论和技术。



李 伟 1976年生。现在浙江大学攻读GIS专业博士研究生。研究方向为海量空间数据库理论和技术。

第三届全国信息获取与处理学术会议

征文通知

第三届全国信息获取与处理学术会议将于2005年8月在浙江省金华市召开,本次会议由中国仪器仪表学术主办,沈阳市仪器仪表与自动化学会和浙江师范大学联合承办。

征文范围:(1)各种电量和非电量的测量方法(如机械量、热工量、物性和化学成分量、状态量等);(2)各种测量装置(如传感器、敏感元件、敏感材料、测量电路、处理电路、显示电路等);(3)测量过程中的信号传输(如现场总线等);(4)模式识别(如图像、语音等);(5)误差与数据处理;(6)故障诊断;(7)过程控制方法与装置;(8)其他。

征文要求:(1)论文未公开发表;(2)来稿请严格按照《沈阳测控信息网》提供的排版格式要求排版,打印两份,包括中英文摘要,参考文献,全文原则上控制在二个版面,摘要的撰写请参考《Ei数据库文摘要求》;(3)受各类基金资助的论文,请注明项目名称及编号;(4)请用A4纸打印论文题目,作者姓名,简历,单位,通讯地址,邮编,电话,传呼,手机,E-mail等;(5)来稿请通过电子邮件和邮局同时投寄;(6)版权纠纷,作者自负。

有关会议具体时间,日程安排,金华概况等会议相关信息请关注《沈阳测控信息网》。

重要日期:论文截稿日期为2005年3月15日,论文录用通知发出日期为2005年4月1日。

投稿地址:沈阳市铁西区南十三路1号29信箱(邮编:110023) 沈阳市仪器仪表与自动化学会 贾冬娜收。电话:024-25691417 传真:024-25415320 电子信箱:webmaster@meeting.org.cn,或 webmaster@syias.org.cn。《沈阳测控信息网》网址:<http://www.syias.org.cn>。

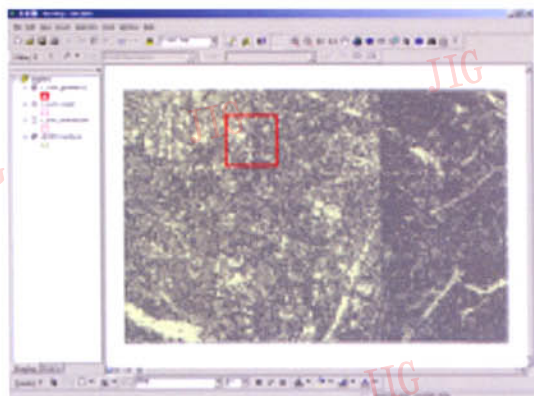
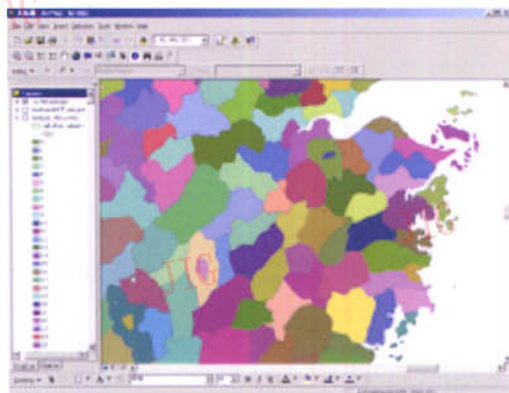


图1 剪切矩形在图幅gh4804中的位置



(a) 全国市县1:400万行政图(放大到浙江)

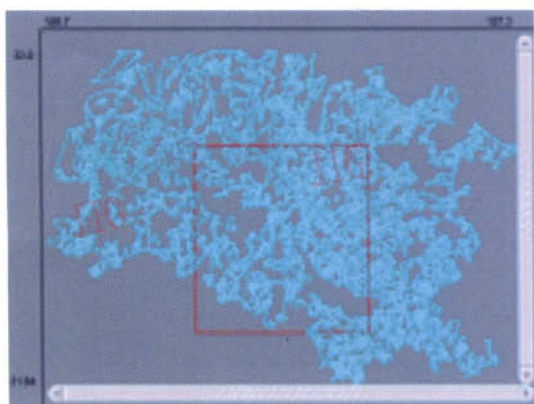
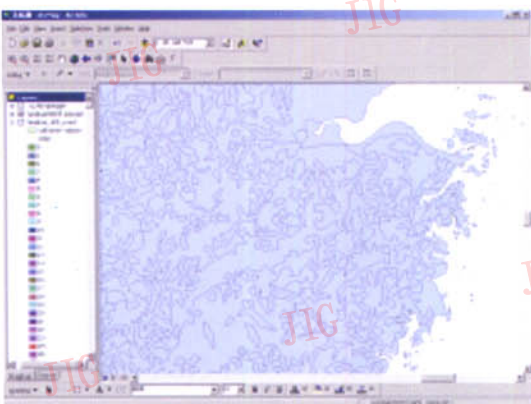


图2 求交操作错误的Geometry



(b) 全国1:400万土地利用图(放大到浙江)

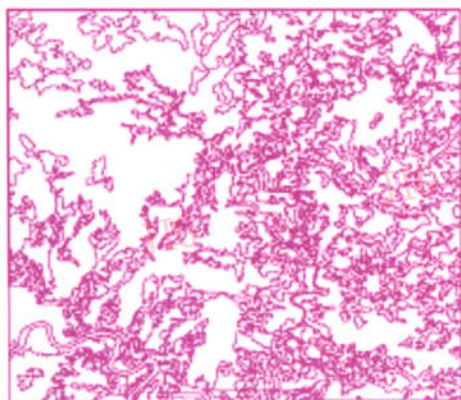
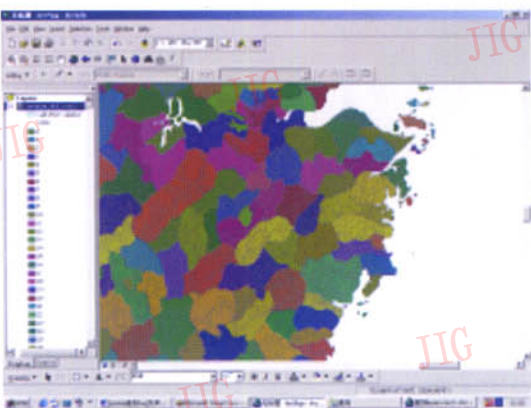


图3 gh4804_landuse与剪切矩形叠加结果



(c) 叠加结果(放大到浙江)

图4 测试结果与应用