

# 对象关系型空间数据库 MMP 过滤算法及其优越性

孔冬艳<sup>1</sup>, 刘俊<sup>2</sup>, 王宏斌<sup>3</sup>

(1. 中国地质大学(北京)地球科学与资源学院,北京 100083; 2. 北京大学地球空间学院,北京 100081; 3. 大庆石油管理局物探公司,大庆 163000)

**摘要:** MMP( Minimum Maximum Points ) 过滤算法是对象关系型空间数据库中空间数据的高效过滤算法。通过对 MMP 算法原理的阐述及应用效果分析,说明 MMP 算法在实现海量空间数据检索中的优越性,即采用 MMP 过滤算法的速度比函数操作至少提高几十倍。

**关键词:** 地理信息系统;对象关系型空间数据库;ZEUS 2000;MMP 过滤算法

**中图分类号:** TP 311.132.3 **文献标识码:** A **文章编号:** 1001-070X(2006)01-0046-05

## 0 引言

空间索引在地理信息系统中具有重要的地位,决定整个系统的效率。空间数据索引方法可归纳为两种:以数据驱动的索引和以空间驱动的索引。以数据驱动的索引是以对象数据为单位建立索引,对每一个对象数据建立唯一的 ID 号,建立索引,如 R-Tree、R<sup>+</sup>-Tree、R<sup>\*</sup>-Tree、X-Tree、WTV-Tree 及 Hilbert Tree 等;以空间驱动的索引是通过对空间进行分割建立索引,即将空间按照规定的网格或其它分层结构分割成单元,每个单元分配一个号码,然后将每一个空间对象与它重叠单元的号码建立联系,如 Linear quad trees(线性四叉树)和 Grid file(网格文件)。例如,Oracle Spatial 使用线性四叉树方法实现索引,采用 B-Tree 访问结构;InfoRmix 的 Spatial DataBlade 利用数据片结合大小不同的 MBR 组成树结构,访问结构也采用 B-Tree;ESRI 的 SDE 用二层(或三层)网格文件把地理对象画成有规则的格,实现索引机制。

关系型数据库所提供的标准索引机制只适用于处理标量(GIS 中的属性数据)数据,而不适用于处理空间数据。空间索引是基于某空间条件,以缩小数据搜索范围的机制。空间数据索引的主要目的是方便空间数据的筛选,有效处理查询。空间数据索引也是减少处理数据的磁盘访问次数的主要方法。

## 1 R-Tree 与 R<sup>\*</sup>-Tree 空间数据索引方式

空间数据查询一般经过两个过滤步骤:第一步,利用空间数据索引找出满足查询条件的近似数据集;第二步,比较分析这些近似数据集,得出准确的查询结果。前者的目的是快速创建最符合查询要求的近似数据子集,以减少第二阶段的处理负荷,因此,过滤时需要借助基于数据特征而建立的空间数据索引,帮助缩小搜索范围。

### 1.1 R-Tree 索引

R-Tree 索引自 1984 年提出后一直是研究热点,人们对其不断改进,提出了多种 R-Tree 变形,形成了 R-Tree 类索引体系。其中,具代表性的有 R<sup>\*</sup>-Tree、Hilbert R-Tree 和 Shift-Optimal R-Tree。

R-Tree 是以数据驱动的索引方法,是空间数据索引结构中最重要的一种层次结构,它是一维 B-Tree 在空间数据的一种扩展,目前已成为很多空间数据索引方法的基础。R-Tree 类空间索引是空间索引发展的主要方向。它具有较强的灵活性,能够满足数据的动态要求,与传统关系型数据库兼容性强。采用 R-Tree 索引可以用于任何 3D 或 4D 的数据索引,这对于解决石油勘探、建筑、工程和很多其它科学应用都很重要。

R-Tree 是一棵由目录节点和叶节点构成的平衡树,它用最小的边界矩形(MBR)来近似每一个几

何图形。如图 1 中最小边界矩形 1~14, 在图中作为叶节点, 一个空间对象只能被分到一个 MBR 中去,

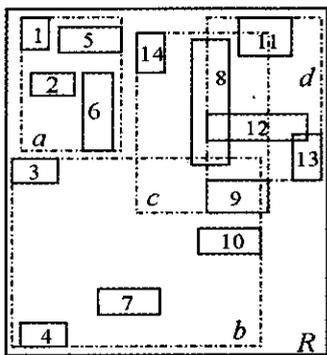


图 1 R-Tree 是由目录节点和叶节点构成的平衡树

图 1 的空间对象 a、b、c、d 划分成矩形即为 MBR。但由于矩形之间有重叠, 查找时可能需要多个路径, 如对象 9 可以化在矩形 b 中, 也可以化在矩形 c 中。若用动态插入算法构建 R-Tree, 则在树的节点上会引起过多的空间重叠和死区, 使算法性能降低。

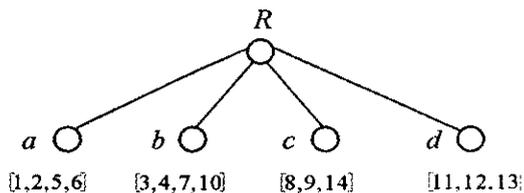
R-Tree 插入算法主要是选择合适的叶节点来插入数据, 并解决叶节点的溢出问题。R-Tree 的分裂算法按照复杂度分为线性、二次和指数分裂 3 种方式。实验证明, 二次分裂方式在性能和计算复杂度上具有比较好的综合指标。该结构的 MBR 之间允许重叠, 一方面保证了 R-Tree 至少 50% 的空间利用率, 但另一方面, 这种无约束的重叠, 在维数增高时很可能会导致索引次数和存储空间的大量增加, 严重影响查询效率。

### 1.2 R\* - Tree 索引

R\* - Tree 和 R-Tree 采用同样的方法划分空间数据集, 但它通过修改插入、分裂算法, 并通过引入强制重插机制对 R-Tree 的性能进行改进, 在选择插入路径的同时考虑矩形的面积、空白区域和重叠大小, 而 R-Tree 只考虑面积的大小。R\* - Tree 不仅对叶子区域体积优化, 而且还优化了叶子节点区域之间的覆盖、叶子节点区域的表面及存储利用率。

与 R-Tree 的插入算法不同, R\* - Tree 在为一个数据目标选择合适的插入叶子节点时, 当点未被包括在任何叶子节点区域内时, 对于叶子节点和目录节点的处理方法是不同的。对于叶子节点, 选择使覆盖产生最小扩大值的区域, 扩大值相同时, 进一步比较叶子节点的体积扩大值和自身体积大小。对于目录节点, 先判断最小体积扩大区域, 若不能判定时, 则由自身的体积决定。

其基本结构是由大小不同的 MBR 排成树状, 上一级的 MBR 由多个下一级的 MBR 组成。



行再插入”的概念, 即如果一个节点溢出, 就删除一定百分比的远离中心区域的目标, 再按插入方法重新插入这些目标, 使存储利用率达 71% ~ 76%。

通过分析, 虽然 R\* - Tree 的“再插入”算法会增加运算复杂度和访问磁盘次数, 但却避免了很多节点的分裂。通过实验, R\* - Tree 各方面的性能都超过 R-Tree(提高了 10% ~ 75%), 而且 R\* - Tree 的鲁棒性也很强, 适于多种数据分布的情况。

## 2 MMP 过滤算法——一种改进的算法

在使用 R\* - Tree 处理空间查询时, 由于第一步对划分的最小边界矩形 MBR 过滤较粗, 导致第二步过滤时, 检查过多的候选对象消耗大量 CPU 时间。为了减少运算时间, 在进行 R\* - Tree 处理前, 采用 MMP 过滤方法, 并且将具有更高过滤效率的空间过滤器以层叠方式把经过 MBR 过滤的数据再做多重过滤, 然后从 R\* - Tree 结构中提取满足查询条件的 MBR 数据作为预处理对象集合。通过采用 MMP 过滤器, 对提取的预处理对象集合进行过滤, 大大地缩小最终参加数据比较运算的目标对象数量。MMP 过滤器适合处理面对象和线对象。通过对 MMP 算法原理的研究及应用效果分析, MMP 算法在实现海量空间数据检索方面具有显著的优越性。

### 2.1 MMP 过滤算法原理及处理过程

由于空间数据检索涉及大量读写操作, 需要很多处理时间, 因此, 在 R\* - Tree 处理中, 采用 MBR 来近似每一个几何图形, 以便查询。查询分为两步: 第一步, 利用空间数据索引找出与查询吻合的近似矩形数据集; 第二步, 分析这个近似矩形数据集, 得出准确的空间检索结果。如图 2 所示。

在空间查询处理时, 在 R\* - Tree 的第一步和第

R\* - Tree 分裂算法的特点在于提出了一种“强

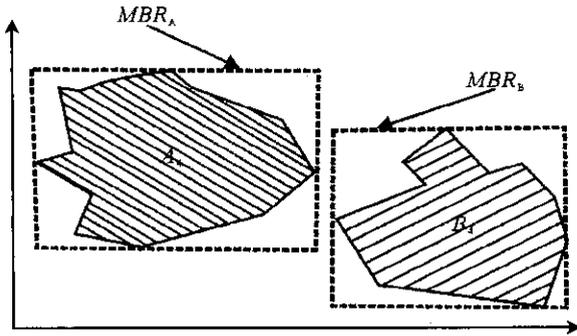


图 2 候选区域最小边界矩形  $MBR_A$  与  $MBR_B$

二步之间增加 MMP 过滤。

第一步过滤是利用  $R^*$ -Tree 取得满足第一步条件的 MBR, 并将 MBR 作为索引键。这一步只检索最小边界矩形 MBR, 确定候选 MBR 对象集, 并不检索实际的空间对象。

第二步, 采用 MMP 过滤器, 对候选 MBR 对象集过滤, 过滤掉不满足查询条件的 MBR( 无效数据), 确定满足查询条件的 MBR 集, 以减少第三步提取阶段要处理对象的数量。

第三步, 对经过 MMP 过滤器确定为满足条件的 MBR 空间对象进行实际检索运算。在这一阶段, 要对复杂的空间对象实际形态进行运算( 例如图 2 中  $A_4$  或  $B_4$  )。如果  $MBR_A$  和  $MBR_B$  被确定为目标对象, 还需要检查二者是否交叉, 如果交叉, 再检查空间对象  $A_4$  和  $B_4$  是否交叉等, 因此最为消耗时间。实验表明, 采用 MMP 过滤速度比函数操作至少快几十倍。

在空间数据检索中, 点对象的检索实现简单, 重要的是如何高效地检索线对象与面对象。MMP 算法不仅适用于面状对象, 也适用于线对象, 可以对非空间数据进行  $B^+$ -Tree 索引结构定义, 也可对空间数据进行空间索引, 即  $R^*$ -Tree 索引结构的定义。

MMP 算法具有在生成空间过滤器或进行管理时消耗资源少、保存时所占存储空间少及不仅适用于面对象, 也适用于线对象的特点, 但 MMP 算法并不适用于所有的空间运算。

### 2.2 多重过滤处理器的结构

多重过滤处理器由 MMP 算法的  $R^*$ -Tree、MBR 过滤处理器和 MMP 过滤处理器构成。 $R^*$ -Tree 弥补了在空间访问方法中  $R$ -Tree 的缺点, 并改善了  $R$ -Tree 的性能。在生成  $R^*$ -Tree 时, 生成空间对象的 MMP 算法, 并添加在  $R^*$ -Tree 叶节点上; 在删除  $R^*$ -Tree 时, 修改  $R^*$ -Tree, 完成删除 MMP 过滤处理。MMP 过滤器通过 MBR 过滤处理器得到候选对象, 在对各点的查询中获取区域内对象

的信息, 并根据位置信息完成对无效数据的滤除, 确定处理目标对象。目标对象在提取阶段(对复杂的空间对象实际形态进行运算)进行处理, 如图 3 所示。

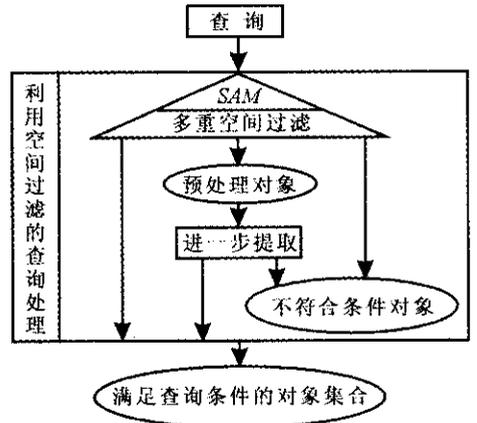


图 3 多重过滤处理过程

所要检索的对象如果是点或单纯的线时, 由于提取阶段所需时间不多, 无须利用 MMP 算法进行过滤处理。即使要处理的运算始终没有适当的目标对象( Always False :AF ), 也无需使用 MMP 算法进行过滤处理。要检索的对象为复合线时, 如果区域对象的包含、覆盖、交迭运算和待检索的对象为多边形、圆形和四边形时, 由于区域对象的交叉运算为 AF, 无须进行多重过滤处理。表 1 是采用 MMP 算法的多重过滤处理器时, 判断无效数据和目标数据的情况。

表 1 无效数据、目标数据及目标数据检索类型<sup>①</sup>

目标数据检索类型的运算 & 检索	无效数据	目标数据
包含 多边形/圆 矩形	接触点 > 0	-
被包含 多边形/圆 矩形	$n_{Contained} > 0$	- $n_{Contained} = nMMP$
覆盖	No Filtering	
被覆盖 多边形/圆 矩形	$n_{Disjoint} > 0$ $n_{Contained} = nMMP$	-
交迭 多边形/圆 矩形	- $n_{Disjoint} = 0$	$(0 < n_{Disjoint} < nMMP)$ && $(n_{Contained} > 0)$
交叉 多边形/圆 矩形	- $n_{Contained} = nMMP$	$(0 < n_{Contained} < nMMP)$ && $(n_{Disjoint} > 0)$
接触 多边形/圆 矩形	$n_{Contained} > 0$	-

<sup>①</sup>  $nMMP$  表示构成 MMP 算法点的数量;  $n_{Contained}$  表示区域对象内满足被包含关系 MMP 算法点的数量;  $n_{Disjoint}$  表示区域对象内满足相离关系 MMP 算法点的数量

### 3 MMP 算法的优越性

为了验证采用 MMP 算法多重过滤处理器的优越性,针对相同的实验数据,采用多重过滤处理器的查询处理和只用 MBR 查询处理做性能实验,并且对测定的处理时间和 MMP 算法的过滤率进行比较。

测量的处理时间为采用 MBR 确定候选对象直至得到满足查询条件的对象集合这段时间。MMP 算法的过滤率是指在 MBR 过滤处理中确定的候选对象与提取阶段无须处理的对象数的比率。计算公式:过滤率=(根据 MMP 确定的目标对象数+MMP 运算确定的无效对象数)/MBR 候选对象数×100,图 4 所示。

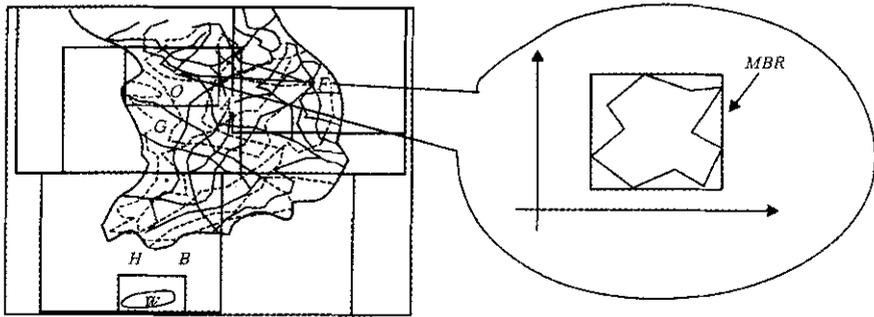


图 4 四边形区域对象的复合线检索实验图形

表 2、表 3 分别是对四边形区域对象内道路和面边界数据的检索结果。

表 2 对四边形区域对象内一般道路数据检索结果

运算	MBR 的候选对象	MMP 的目标对象	MMP 的无效对象	MMP 的候选对象	检索对象数量	所需时间/s		MMP 过滤率/%
						用 MMP	不用 MMP	
被包含	1 199	1 199	0	0	1 199	0.05	0.25	100
被覆盖	1 199	0	1 199	0	0	0.05	26.13	100
交叉	1 256	57	1 199	0	57	0.05	29.16	100
接触	1 256	0	1 256	0	0	0.05	27.59	100

表 3 对四边形区域对象面边界数据检索结果

运算	MBR 的候选对象	MMP 的目标对象	MMP 的无效对象	MMP 的候选对象	检索对象数量	所需时间/s		MMP 过滤率/%
						用 MMP	不用 MMP	
被包含	411	411	0	0	411	0.01	0.09	100
被覆盖	411	0	411	0	0	0.02	10.9	100
交叉	444	32	411	1	32	0.04	13.14	99.77
接触	444	0	411	1	0	0.03	11.69	99.77

在以区域对象为目标的检索运算当中,使用多重过滤处理器时,由于 MMP 算法的过滤率接近 100%,与作为运算对象的 MBR 候选对象数量无关,检索速度小于 0.1 s。只利用 MBR 进行查询处理时,检索速度与候选对象数量和复合线构成的点数按比例改变。实验表明,采用 MMP 过滤速度比函数操作至少快几十倍。

ZEUS 2000 的实际应用检测中效果突出,比常用函数操作检索空间数据方法速度提高几十倍。

### 4 结语

MMP 过滤算法是一种效率较高的空间数据过滤算法,经过理论论证与实际运用,该方法能提高空间数据库检索速度。在对象关系型空间数据库

### 参考文献

[ 1 ] Kim won. Bringing Object Relational Down To Earth[ EB/OL ]. http://www.ktdata.com 2002-11-16.  
 [ 2 ] 孔冬艳,刘斌,Margrit. ZEUS2000 技术白皮书[ EB/OL ]. http://www.geosmap.com 2004-11-8.  
 [ 3 ] 宋关福. 面向对象整体 GIS 数据模型的设计与实现[ EB/OL ]. http://www.surpermap.com 2004-10-15.  
 [ 4 ] 陈俊华,宋关福,李绍俊. 基于 RDBMS 的空间数据库的设计与实现[ EB/OL ]. http://www.supermap.com 2002-11-16.  
 [ 5 ] 解云琳. 空间数据与关系数据库集成方法研究[ A ]. 中国地理信息系统会第二次团体会员代表大会暨海峡两岸地理信息系统发展研讨会论文集[ C ]. 北京:地理信息协会.

[ 6 ] Oracle Spatial. 技术白皮书[ EB/OL ]. <http://www.oracle.com> , 2003 - 11 - 8.

[ 7 ] 周新忠, 赵俊. GIS 中应用 Oracle Spatial 对象模型进行空间数据存储的分析[ J ]. 矿山测量, 2003( 8 ) 39 - 42.

[ 8 ] 罗平, 杜清运. 基于关系数据库的 CA 模型扩展和时空模拟实验研究[ J ]. 地理学与国土研究, 2002 ( 8 ) 8 - 11.

[ 9 ] 梅承力, 周源华. 高维数据空间索引的研究[ J ]. 图像通信与信息处理研究, 2002 ( 1 ) 22 - 25.

[ 10 ] 马修军, 周贵云, 邱伦. GIS 空间动态模型组件的设计实现地理学与国土研究[ A ]. 中国地理信息系统会第二次团体会员代表大会暨海峡两岸地理信息系统发展研讨会论文集[ C ]. 2003.

[ 11 ] 马常杰, 陈守余. 基于 MapGIS 的勘探剖面数据库设计与实现[ J ]. 地理与地理信息科学, 2003 ,19( 5 ) 32 - 34.

## THE MMP FILTER ALGORITHM IN THE OBJECT - RELATIONAL SPETIAL DATABASE AND ITS ADVANTAGES

KONG Dong - yan<sup>1</sup> , LIU Jun<sup>2</sup> , WANG Hong - bin<sup>3</sup>

( 1. China University of Geosciences , Beijing 100083 , China ; 2. Peking University , Beijing 100081 , China ; 3. Daqing Oil-field Geophysical Company , Daqing 163000 , China )

**Abstract :** The MMP ( Minimum Maximum Points ) filter algorithm is a highly - efficient filter algorithm for the spatial data in the object - relational spatial database. The description and the application result analysis of the MMP algorithm demonstrate the advantages of the MMP algorithm in the retrieval of mass volume of spatial data , i. e. , the speed of the MMP algorithm is at least scores of times higher than that of the function operation.

**Key words :** GIS ; Object - relational spatial database ; ZEUS 2000 ; MMP filter algorithm

第一作者简介: 孔冬艳( 1969 - )女, 博士, 计算机、自控、地图制图学与地理信息专业, 目前研究方向为图形处理与地理信息。

( 责任编辑: 刁淑娟 )

=====  
( 上接第 42 页 )

[ 6 ] Wu Tzong - Dar , Chen K S. A Reappraisal of the Validity of the IEM Model for Backscattering From Rough Surfaces[ J ]. IEEE Trans. Geosci. Remote Sensing , 2004 42( 4 ) 743 - 753.

[ 7 ] Kendra John R. Microwave Remote Sensing on Snow : An Empirical/ Theoretical Scattering Model for Dense Random Media[ D ]. University of Michigan , 1995.

## THE APPLICATION OF THE AIEM MODEL TO THE SNOW COVERED GROUND SCATTERING SIMULATION

WU Sheng - li , WANG Jian - ming , LIU Wei , YU Qin

( State Key Lab of Remote Sensing Science , Institute of Remote Sensing Application , Chinese Academy of Sciences , Beijing 100101 , China )

**Abstract :** A method of snow covered ground simulation using the first - order radioactive transfer model is described in this paper. In this method , three parts of the echo are taken into account , i. e. , top surface scattering , volume scattering and subsurface scattering. A new surface scattering model——AIEM , instead of the IEM model , is used to deal with the surface scattering problem. Finally , the improved method is verified using the measured data of 1995 from the University of Michigan.

**Key words :** Radioactive transfer model ; Surface scattering ; Volume scattering ; Snow ; AIEM

第一作者简介: 武胜利( 1980 - )男, 中国科学院遥感应用研究所博士研究生, 主要从事主动微波遥感方面研究。

( 责任编辑: 刁淑娟 )