

# 空间 OLAP 技术研究

李一军, 樊博

(哈尔滨工业大学管理学院, 哈尔滨 150001)

**摘要:**针对决策支持系统中空间数据可视化应用趋势,研究基于空间数据仓库的一种决策分析工具——在线分析处理,通过改进空间数据立方体的物化视图选择方法,并进行了实例分析,验证该方法的有效性和优越性,更好地解决空间 OLAP 响应时间延迟问题.进而扩展空间 OLAP 应用模型和模式,实现更加完备、灵活的空间 OLAP 操作功能.

**关键词:**在线分析处理;空间数据立方体;空间数据仓库;物化视图选择;空间在线分析处理  
**中图分类号:**C934 **文献标识码:**A **文章编号:**1007-9807(2003)04-0009-08

## 0 引言

研究表明,目前在各类决策问题使用的数据中有 80% 带有与空间位置有关的特性,如企业位置、供应商与客户地址、交通路线、设施分布等.在通常的信息管理系统中,这些空间数据只能用简单的属性字段来存储和表示,缺乏可视化和空间化的信息分析与表达.

目前,多数的空间决策支持主要通过地理信息系统(GIS)与应用模型库的耦合来实现基于空间信息的决策分析,空间与非空间数据的存储是分离的.其主要特点是空间数据操作面向工作流程的事务处理.如果将不同应用领域的的数据按相应主题集成到一起,形成统一的信息视图,就能更全面地反映客观世界.因此,数据仓库(data warehouse)及其分析工具——在线分析处理技术(OLAP)和数据挖掘技术(data mining, DM)作为有效的决策支持技术得到研究者的重视,逐渐成为决策支持系统(DSS)研究的一个热点.可以把空间决策信息集成起来构成数据仓库,形成面向主题的、集成的、随时间变化的、稳定的空间和非空间数据的集合<sup>[1]</sup>.空间数据作为整体纳入到统一的空间数据仓库中,能够克服传统 GIS 数据处理

的过程化和存储的分离化,提升空间数据的分析能力.它是决策支持系统集成空间分析能力的必然趋势.

本文将从上述角度出发,结合空间数据处理的特点,研究基于空间数据仓库的空间在线分析处理(spatial OLAP).主要解决空间 OLAP 操作存在的响应时间延迟问题,实现快速在线分析,并进一步扩充空间 OLAP 的应用模型,实现更加完备的空间 OLAP 功能.

## 1 空间 OLAP 模型

空间数据仓库的物理存储方式有 ROLAP 和 MOLAP 两种,两种实现方法各有优点,由于空间数据存储量大,以及业已成熟的关系数据库技术优势,通常采用 ROLAP 方式.

### 1.1 数据立方体

数据仓库和 OLAP 工具是基于数据立方体(data cube),由维和事实度量组成<sup>[2]</sup>.数据立方体为 OLAP 提供快速的多维分析条件.

数据立方体的形式化表达:

$$MD = (D, H, M, )$$

其中: $D = \{D_1, D_2, \dots, D_n\}$ ——数据立方体的

收稿日期:2002-07-02;修订日期:2003-05-29.

基金项目:国家自然科学基金资助项目(70171013);国家 863 计划资助项目(2001AA136010-08).

作者简介:李一军(1957-),男,博士,教授,博士生导师.

维,如时间维、商品维等,每个维与一个维表关联, $D_i = (i_{i,1}, \dots, i_{i,t(i)})$ 为维的具体值集合, $i_{i,t(i)}$ 是维  $D_i$  的一个具体值, $t(i)$  是  $D_i$  的属性个数.

$H = \{ H_1, H_2, \dots, H_n \}$  ——维的层次关系集合.  $D_i \supseteq H_i, H_i$  是  $D_i$  的对应的概念层次,如  $H_i$ : year month day.

$M = (M_1, M_2, \dots, M_n)$  ——数据立方体事实度量,是一组数值函数,如销售额、销售量等.

$F = (F_1(M), \dots, F_S(M))$  ——事实度量的聚集函数,如  $Sum(M), Count(M)$  等,聚集函数可以是分布的、代数的、整体的三种类型,由于  $M$  和  $F$  的变化,一个多维数据库可以衍生出若干多维视图.

### 1.2 数据立方体的格

数据立方体的格的形式化表达:

方体格 (lattice) ——不同分组或方体间形成的偏序关系构成的格  $(L, \leq)$ .  $\leq$  为偏序操作符,

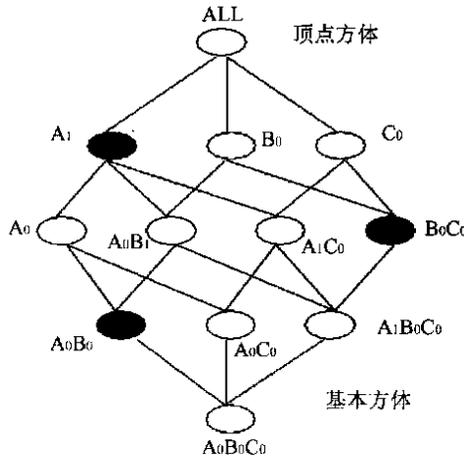


图 1 数据立方体的格

数据仓库是基于多维数据模型,多维数据分析的核心是有效计算任意维组合的聚集.由图 1 可知<sup>[3]</sup>,三维数据立方体的方体格是对  $A、B、C$  三个维的任意分组,其中  $C$  维有两个层次,聚集计算度量  $M$ .按式(1)计算方体总数为  $3 \times 2 \times 2 = 12$ .如果一个数据立方体有 10 维,每维有 4 个层次,则它的 cuboid 总数达到  $5^{10}$  之巨.因此,系统难以提供预先计算(物化)全部的方体(cuboid)需要的海量存储空间,而不预先计算任何方体可能导致联机分析时昂贵的多维聚集计算,OLAP 操作的响应时间难以忍受.解决这个问题的最有效的技术是选择性的预先计算一部分方体,在联机分析代价和存储空间之间作出平衡选择<sup>[2]</sup>.

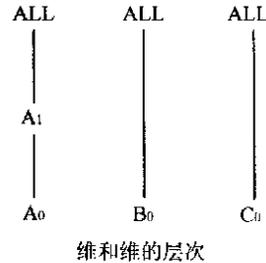
如图 1 中  $B_0 \leq B_0C_0$  是指方体  $B_0$  可以由方体  $B_0C_0$  的聚集计算得到.在格中,所有的方体都是由基本方体  $(A_0B_0C_0)$  计算出来的,ALL 为顶点方体.

其中:分组(又称 cuboid,方体) ——维及其层次集合的任意组合,在 SQL 语句中称为分组,分组的结果产生方体.如(年,地区,电器)为一个分组.

单元(cell) ——给定方体中各个维的具体值  $(i, j)$ ,则唯一确定度量的具体取值,称为 cell,例如:(1999 年,哈尔滨,电冰箱) (销售额).方体总数的计算公式<sup>[2]</sup>

$$T = \prod_{i=1}^n (L_i + 1) \tag{1}$$

式中: $T$  为数据立方体中 cuboid 的总数; $n$  为维数; $L_i$  为维  $i$  的层次数; $L_i + 1$  是指维  $i$  泛化到 ALL 层(即去掉维  $i$ ).



计算方体的数目是由存储空间的大小唯一决定的.存储空间越大,可以预先计算和存储的方体就越多,OLAP 的响应速度就越快.但系统的存储空间难以满足数据仓库中不断膨胀的数据量,所以如何在有限的存储空间约束下选择更加合理的一部分方体来预先计算,使 OLAP 的响应速度加快,成为研究的焦点.方体的选择问题(物化视图的选取)是一个 NP-Hard 问题.解决方法主要有 HRU 贪心算法<sup>[4]</sup>、遗传算法<sup>[5,6]</sup>等.例如图 1 中的黑色方体为文献[4]的贪心算法选出的预计算方体,该算法每次选出一个物化增益最大的方体(由于方体之间存在的偏序关系,选择的物化方体的原则是要对方体格中的其他方体的计算最有

利),直至填满存储空间. 本文在以上普通数据立方体的物化视图选择的基础上,研究更加复杂的空间数据立方体的物化视图选择问题.

### 1.3 空间数据立方体的多维数据模型

普通数据立方体的度量是一个数值函数,它对数据立方体的每一个点求值. 通过给定点的各‘维 - 度量’对聚集数据,计算该点的度量聚集值  $F_i(M_i)$ , 例如  $Sum()$ ,  $Count()$  ... 空间数据立方体由许多维和大量度量(空间度量和数值度量)组成,但不同于普通数据立方体的数值度量,空间度量是指向空间对象集合的一组指针. 空间 OLAP 操作作用于空间度量,需要计算大量空间对象的合并、叠加、联接等.

- $= (F_1(M), \dots, F_s(M)) = Sum(M),$   
 $Count(M), Avg(M), \dots$   
——数值度量的聚集函数
- $= (F_1(M), \dots, F_s(M)) = Am\ lg\ m\ te(M),$   
 $Overl\ y(M), Join(M), Intersect(M), \dots$   
——空间度量的聚集函数

图 2 中有数以千计的气象探测器形成的观测站点覆盖着整张地图,和针对气象分析建立的空间数据仓库的星型模式<sup>[7,8]</sup>,它包括四个维:时间维,温度维,降水维,区域维. 其中前三个是非空间

维,区域维是空间维. 每一个维都有相应的概念层次,事实表中包括两个度量:区域指针,面积. 面积是数值度量,区域指针是空间度量. 空间 OLAP 操作要求快速精确的空间区域合并显示. 如 roll-up 操作可以将泛化在同一个抽象层次(如相同温度或降水量)的大量空间指针合并成同一单元,并计算合并后的面积,然后将合并结果快速显现在地图上. 例如查询 1998 年某省的高温地区.

有效的空间 OLAP 在于快速、灵活地实现基于数据立方体的在线决策支持查询. 空间在线分析处理和空间度量的计算无疑具有更高复杂度和巨大的计算开销. 每个数字度量的聚集值仅需要几个字节的存储空间,而存储地区合并图要上兆的空间<sup>[7]</sup>. 将所有空间 OLAP 操作形成的地图存储起来是不可能的,在线合并时间更是不可忍受的. 人们仍然面临着空间在线分析代价和存储空间开销的两难选择. 最有效的解决方法是预计算部分空间度量,但具有空间度量的数据立方体的物化不能仅仅停留在选择 cuboid 上,必须进一步深入到空间数据立方体的 cell 级<sup>[8]</sup>,根据空间度量的访问频度、可共享性等因素来预先计算一部分可以合并的空间对象. 因此,空间数据立方体的物化是从 lattice 到 cuboid 到 cell 的逐渐深入的过程,如图 3.

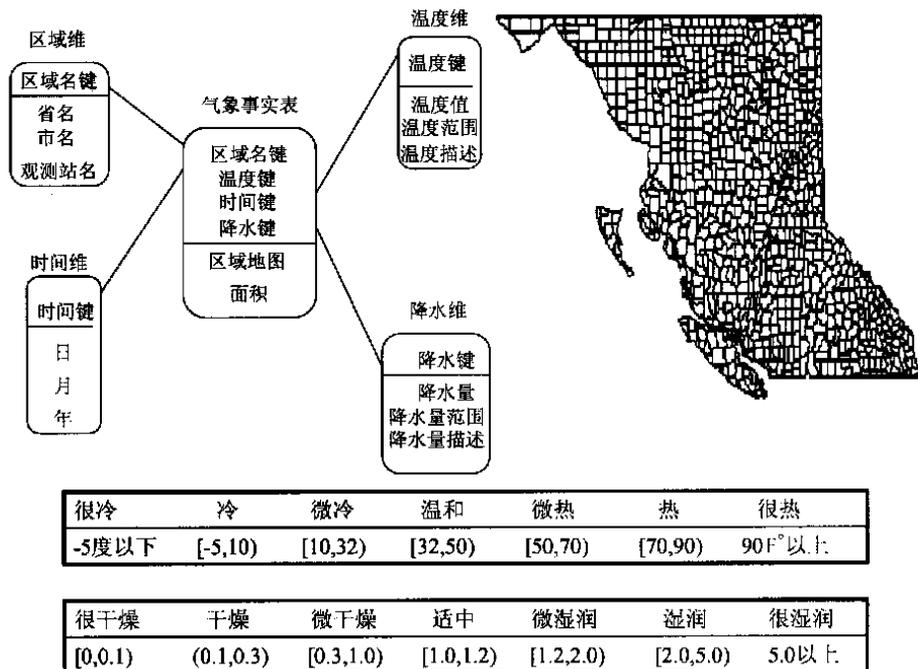


图 2 空间立方体的星型模式及维层次的划分

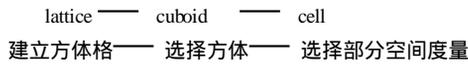


图 3 空间数据立方体的物化选择过程

## 2 基于修正式遗传算法的空间度量选择性物化

空间度量的选择性物化也是 NP-Hard 问题. 空间贪心算法 (spatial greedy)<sup>[9]</sup> 用于解决空间对象的选择问题, 是从单个个体开始, 逐步选出每一步的最优解, 所选择的解只是局部最优解. 当所需要选择的物化视图个数增多或搜索空间很大时, Greedy 算法解搜索能力和解的质量将下降. 遗传算法从问题解的一个集合开始, 具有全局的和并行的解搜索能力. 将遗传算法用于空间度量的选择性物化能改进解的质量, 在存储空间的约束下选出合理的物化解, 但传统遗传算法<sup>[5,6]</sup> 并没有解决这个问题. 本文提出将修正式遗传算法应用到空间数据立方体的单元 (cell) 级, 充分利用传统遗传算法求解质量高的优点, 并采用修正算子去除超出存储空间的不可行解, 使空间数据立方体的物化计算方法更为合理, 最终更好地解决空间 OLAP 的响应时间延迟问题. 下面以选择空间对象集合的合并为例, 设计系列算法如下:

Input :

1. 根据数据立方体的格 (cube lattice), 运行 HRU 算法或遗传算法选择出待物化的方体 (cuboid).
2. 方体的访问频度, 每个可合并的空间对象集合的频度, 是指包含的该空间度量集合所有 cuboid 的访问频度之和.
3. 所选中方体的每个单元 (cell) 中的空间指针集合.
4. 表明空间对象邻接关系的邻接信息表.
5. 存储每组空间对象的空间开销值.
6. 给定空间物化视图的总存储空间.

Output :

需要物化的空间对象集合.

### 2.1 选择可以合并的空间对象集合

Algorithm 1<sup>[3]</sup>

$G = \emptyset;$

For each cuboid that HRU algorithm selected

For each cell that the cuboid contains / 对于选定方体中的每个单元

$remain$  all the spatial objects pointers in cells / 所有指向空间对象集合的指针赋给  $remain$

while  $remain \neq \emptyset$  do

test connectivity of  $g$  / 用邻接信息表验证空间对象集合的邻接性

$G = G + g$

$remain = remain - g$

End while

End for

遗传算法的适应度函数

如果一组可合并空间对象的访问频率  $f(g)$  比其它组高, 则合并这组空间对象比较有利; 如果一组可合并空间对象的个数  $N(g)$  多, 则合并这组空间对象较有利. 因为这可以使在线分析时磁盘的访问次数减少. 根据这些启发策略, 得到物化一组可合并空间对象的利益公式

$$B(g, M) = f(g) \times N(g) + \sum_{(g < A) \in (A - G - M)} f(A) \times (N(g) - 1) \quad (2)$$

$A$  是尚未选出的空间对象集合. 公式右端第一项表示  $g$  的预计算给该对象集合自身带来的直接利益; 第二项表示  $g$  的预计算为包含  $g$  的对象集合带来的间接利益. 利益主要来源于在线分析时磁盘访问次数的减少, 间接利益的计算准则见文献<sup>[9]</sup>. 因此, 适应度函数就是要找到  $G$  的一个物化子集, 使其在存储空间  $S$  的约束下达到总体利益最大.  $c/g$  为物化的一组空间对象所占的存储空间.

$$F = \sum_{g=1}^n B(g, M) \quad (3)$$

$$c_M \mid c \mid S \quad (4)$$

### 2.2 遗传算法

Algorithm 2

Begin

Generation a population  $p$  randomly; / 随机产生一组初始种群  $p$

generation 1;

While generation max. gen do / 给定最大繁殖代数

$P' = \emptyset;$

Use a fitness function to evaluate each

individual in  $p$ ; / 利用适应度函数评估  $p$  中的每个个体(见式(3))

While |  $p'$  | population. size do / 在存储空间约束(式(4))的条件下执行循环

Select two parents from  $p$ ; / 选出适应度大的两个个体

Perform crossover; / 交叉

Perform mutation; / 变异

Place the offspring into  $p'$ ; / 将上两步的结果写入  $P'$

Endwhile;

$p = p'$  /

generation generation + 1; / 下一代

Endwhile

End;

### 2.3 不可行解的修正算法

由于简单的遗传算法没有将存储空间约束嵌入遗传编码中,使用单一的遗传算法将产生超过存储空间约束的不可行解.例如:两个待交叉的基因代表的空间对象集合的存储空间小于总的存储空间约束,但执行交叉后,两个新基因代表的空间对象集合的存储空间却都大于总存储空间,因此必须采用相应的算法以修正不可行解.本文设计的修正算法也是基于贪婪思想,不同的是这个修正算法是逆向贪婪算法,从 Algorithm 2 所选择的物化集合中逐一去除引起利益损失最小的空间对象集合,直到物化集合满足存储空间约束.最终得到的空间对象集合为待物化的结果集.其启发式算子如下

$$D(c_i, M) = B(c_i, M) -$$

$$\text{Indirect. Benefit. Change}(c_i) \quad (5)$$

其中:右式第一项为空间对象集合  $c_i$  物化带来的总利益;右式第二项为撤消物化对象  $c_i$  后,  $c_i$  的子结点或兄弟结点引起的间接利益的增加;左式为去除物化对象  $c_i$  所导致利益的损失.

Algorithm 3

$$X = (x_1, x_2, \dots, x_m) \text{ and } M = \{c_i \mid x_i = 1, 1 \leq i \leq m\}$$

while  $\sum_{c \in M} |c| > S$  do

for each  $c_i \in M$  do

calculate the detriment  $D(c, M)$  / 式(5)

choose the min  $D(c_j, M)$ ; /  $c_j$  是被去除的物化对象

$x_j = 0$ ;  $M = M - \{c_j\}$ ; / 修正基因

Endwhile

### 2.4 实例分析

本文从气象分析空间数据仓库中抽取 2 GB 的数据为实验数据,模拟建立空间数据立方体.假设每个空间方体的访问频率在 0 与 1 之间随机取值;选取空间数据立方体存储空间的 10%、20%、30%、40%、50%、60%、70%、80% 为存储空间约束;由于空间度量物化(空间对象的合并)的目的是尽量减少磁盘的访问次数,使空间 OLAP 操作能够快速响应.因此,应以节省磁盘访问次数作为衡量算法有效性的标准.图 4 是空间数据立方体计算的贪心算法<sup>[9]</sup>与本文的修正式遗传算法的运行结果的比较.

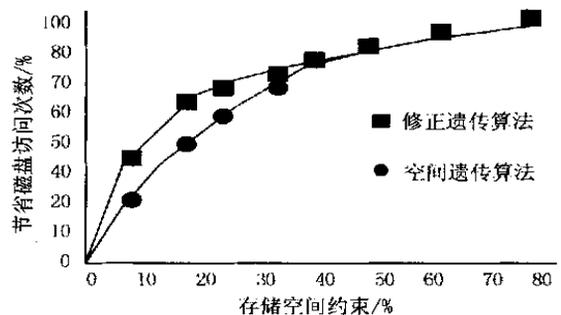


图 4 两种算法的实验结果比较

可以看出,随着存储空间的增大,两种算法节省的磁盘访问数趋于一致.但空间数据仓库存储海量的数据,而且数据量在不断膨胀,系统通常只能提供空间数据立方体 20%~30% 左右的存储空间.因此,修正式遗传算法要明显优于空间贪心算法.

## 3 空间 OLAP 模型和模式的扩展

上述分析的空间 OLAP 是针对气象分析模式的,有局限性.实现比较完备的空间 OLAP 分析应基于:

1) 空间 OLAP 的操作应该是联动的,用户不为预先设计的分组和层次所限定,在地图上选择任意形状的兴趣区域,要求得到数值度量和空间度量的聚集值.

2) 空间 OLAP 的操作应该可以在多层主题图

上执行.例如:“住在某商场周围 3 km 的所有顾客的消费总额”,需要在顾客主题图层和商场主题图层的叠加基础上进行聚集分析.“显示高、中、低消费额度区域”,要求在地区主题图和顾客位置主题图叠加的基础上,进行以区域对象指针为空间度量的合并.

3) 空间维的分组和维层次的划分是不清晰的<sup>[10]</sup>.空间数据仓库的 OLAP 操作是建立在清晰的

的维分组和维层次的基础上的,空间数据立方体的物化计算也是以此为基础的.然而,很多分布在主题图层的空间对象的分层并没有明晰的领域知识,空间维分组和维层次划分并不像非空间维那样清晰.

### 3.1 空间 OLAP 的模型扩充

根据以上空间分析,建立基于客户事实的事实星座模式,如图 5 所示.

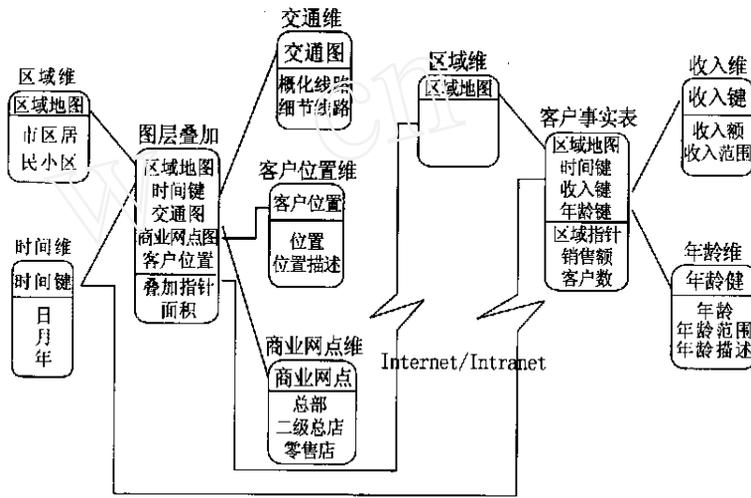


图 5 基于客户分析的星型事实星座

其中,叠加分析和客户分析分属于不同的分析主题,图层叠加分析为客户分析提供背景图层,叠加指针、区域指针是空间度量,销售额、客户数是数值度量.在泛化操作中,落在相同销售额区间的地域可以合并.收入维、年龄维、时间维是非空间维,它们的分组和层次关系清晰.而作为空间维的区域维,则是多个空间图层叠加的逻辑并操作.当决策者需要进行客户分析时,可以灵活地选择基于该分析主题的、不同粒度的背景图,基于地图的空间 OLAP 操作应该充分利用空间索引,例如可以选择空间维客户位置图、商店位置图、地区图为分析背景图,  $R_d$ 、 $R_k$ 、 $R_s$  分别为索引空间区域、客户位置、商场位置的最小边界矩形(MBR).

### 3.2 空间 OLAP 应用模式扩展

基于空间数据立方体构建的 MBR 矩形的形式化描述

{ Rectangle- ID, Type, min- X, max- X, min- Y, max- Y } { Spatial measure, Numerical measure }

其中: Rectangle- ID ——空间矩形标识

Type ——表示该矩形是虚拟空间对象还是

实际空间对象

min- X, max- X, min- Y, max- Y ——表示二维空间对象的最大最小范围

Spatial measure ——空间度量

Numerical measure ——数字度量

每个 MBR 矩形和一组空间度量和数字度量的聚集值存在映射关系.例如,用户要查询“沿公路 20 km 的客户数量”,其基本原理如图 6 所示.

三号区域(a3)的 MBR 如图所示,沿公路涉及的空间区域的索引结构可以用 MBR A1 和 MBR A2 表示. A1 是包括 MBR a2, MBR a4, MBR a5 的更大的 MBR; A2 包括区域 8、9、11、15. 基于空间索引的空间数据立方体<sup>[10]</sup>有以下特点:

1) a 与 A 存在着层次关系,可以作为空间维的概念层次.

2) 度量值与 R. Tree 的有机整合、泛化操作能够沿索引层次的爬升而聚集度量.

3) 索引数据结构的搜索策略是广度优先搜索,这是分组操作和累加计算(如 A1 的聚集值  $15 = 4 + 7 + 4$ )所必须的.

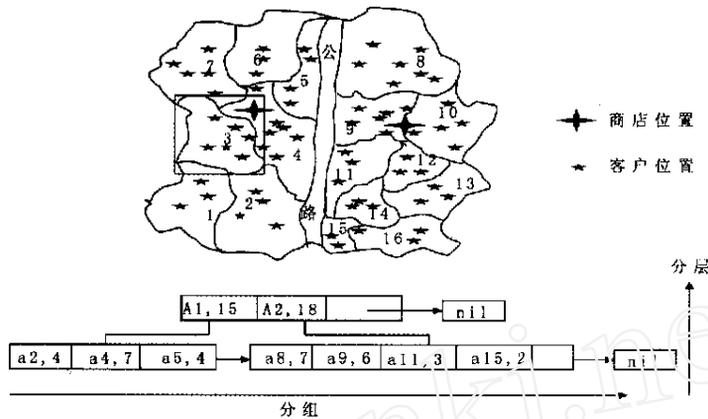


图 6 基于空间索引的 OLAP 工作原理

通过 MBR 构建的 R-Tree 提供了空间维的层次结构、分组和度量的聚集机制,所以,可以把空间维和非空间维结合起来,构建基于空间索引的空间数据立方体,然后建立空间数据仓库的立方体如图 1,最终按照图 3 的设计顺序实施空间数据立方体的物化视图选择。

### 3.3 物化视图的有效利用

针对上述多维数据模型的建立和物化视图的设计,可进一步探讨基于物化视图的利用问题<sup>[10]</sup>。假设用户在地图上提出一个查询窗口(如计算窗口区域的销售总额,并按销售额高、中、低三个层次显示地区),窗口既可以完全包含 MBR 矩形中的空间对象,也可以与 MBR 中的空间对象部分相交。访问窗口从所包括的索引树的根结点开始递归地访问叶结点,递归算法如下

Algorithm 4

Function *Aggregation* ( $X_i, q$ )

For every object  $x \in X_i$  do {

If  $q$  contains  $x$  then  $result := AG(result, x. agr) / x. agr$  是数值度量或空间度量的聚集值,如果  $q$  完全包含  $x$ ,则结果 ( $result$ ) 可能直接利用已经预计算的空间和数值度量的聚集值累计计算,而不必访问具体的叶结点。

Else if  $q$  intersects  $x$  then { / 如果窗口  $q$  与空间对象  $x$  部分相交

$Patial\_result := Aggregation(x.subtree, q) /$   
递归访问  $x$  的子树,并递归调用算法

$result := AG(result, patial\_result) /$  累计计算聚集值,如果存在已物化的聚集值,则直接利用其累计计算。如果没有可利用的物化结果,则带有聚集功能的 R-Tree 作为一般的空间索引访问基本的空间数据库,将查询窗口、空间区域、客户位置相连接,得出精确结果。

} end if

} end for

return  $result$

## 4 结论与展望

对于空间 OLAP 的研究,本文只解决集成空间数据的 OLAP 的快速响应问题,以及空间 OLAP 的应用模型和模式。空间 OLAP 作为空间决策支持系统的一种分析工具,仅仅是一种较浅层次的空间数据挖掘和知识发现工具。可以在空间 OLAP 的基础上,进一步研究有效的空间在线分析挖掘 (OLAM)<sup>[11,12]</sup>,将空间信息的不确定性、模糊性等特征结合到空间 OLAM 中,以期实现集成空间数据的、科学的、深层次的决策分析和知识发现。

### 参考文献:

[1] Haier Baazaoui, Zghal Samifciiz, Henda Ben Ghezala. Exploration techniques of the spatial data warehouses: Overview and application to incendiary domain[J]. IEEE Conference on Computer System, 2001, 7(3): 21—23

- [2] Han Jiawei, Kamber. Data Mining: Concepts and Technique[M]. 北京: 机械工业出版社, 2001
- [3] 童云海, 谢昆青, 唐世渭. 空间 Cube 计算方法[J]. 计算机工程与应用, 2002, 6(1): 3
- [4] Venky Harinarayan, Anand Rajaraman. Implementing data cubes efficiently[J]. Communications of ACM SIGMOD International Conference, 1996, 5(1): 205—216
- [5] Zhang Chuan, Yao Xin, Yang Jian. An evolutionary approach to materialized views selection in a data warehouse environment[J]. IEEE Transaction on System, Man, and Cybernetics, 2001, 8(31): 3—8
- [6] Lin Wen-yang, Kuo I-chung. OLAP data cubes configuration with genetic algorithms[J]. IEEE Transaction on Knowledge and Data Engineering, 2000, 2(5): 132—143
- [7] Han Jiawei, Stefanovic Nebojsa. Selective materialization: A efficient method for spatial data cube construction[J]. Communications of ACM, 1998, 4(1): 7—12
- [8] Stefanovic Nibojsa, Han Jiawei, Kooerski Krzysztof. Object-based selective materialization for efficient implementation of spatial data cubes[J]. IEEE Transactions on Knowledge and Data Engineering, 2000, 12(6): 2—7
- [9] Stefanovic Nebojsa. Design and implementation of on-line analytical processing(OLAP) of spatial data[D]. Canada: Computer Science, Simon Fraser University, 1997
- [10] Papadias Dimitris, Kalnis Panos, Tao Yufei. Efficient OLAP Operations in Spatial Data Warehouses[R]. HKUST-CS01-01, 2001
- [11] Koperski Krzystaf. Mining knowledge in geographical data[J]. Communications of ACM, 1998, 4(2): 1—3
- [12] Han Jiawei, Stefanovic Nebojsa. Geominer: A system prototype for spatial data mining description of geominer system[J]. Communications of ACM, 1997, 4(9): 553—554

## Research on technology of spatial OLAP

LI Yi-jun, FAN Bo

Management School of Harbin Institute of Technology, Harbin 150001, China

**Abstract:** There now appears urgent need for embedding spatial data into decision support system(DSS) which provides spatial analytical function and visualization to user. This article investigates a decision analytical tool based on spatial data warehouse—On-line analytical processing(OLAP). We mainly solve the serious problem of slow response time in spatial data cube operation by improving materialization view selection method. And then, an analysis example is given to test the efficiency of our work. Furthermore we enrich the spatial OLAP model and its application mode to realized more powerful and flexible spatial analytical functions.

**Key words:** on-line analytical processing(OLAP); spatial data cube; spatial data warehouse; materialization view selection; spatial OLAP