

# 多维空间分析的关键技术——空间数据立方体

邹逸江

(宁波大学建筑工程与环境学院, 浙江 宁波 315211)

**摘要:** 介绍空间数据立方体在“数字城市”建设中的作用以及国内外研究现状; 讨论非空间维、空间维、数字度量、空间度量的基本概念和结构, 给出空间数据立方体的完整描述。空间数据立方体的分析操作主要由概括分析、局部分析、全局分析和旋转分析组成, 介绍这些分析操作的功能和结构; 结合具体实例数据介绍多维缓冲区空间分析。

**关键词:** 地理信息系统; 空间数据立方体; 空间数据仓库; 多维空间分析

**中图分类号:** P208      **文献标识码:** A      **文章编号:** 1672-0504(2006)01-0012-05

## 0 引言

构筑“数字城市”首先是构建基础地理空间信息平台; 其次是把城市的各种专题信息和时间信息加载上去, 构筑一个基于多维数据模型(地理空间信息维、多种专题信息维、时间信息维)的多维信息框架; 第三, 提供基于地理空间定位的多维信息查询和信息图形图像显示, 为政府和社会提供信息化服务; 第四, 提供多维信息空间分析, 进一步为政府提供决策支持。因此, 多维信息空间分析是构筑“数字城市”的关键技术之一。

但是, 目前的 GIS 空间分析很难满足多维信息空间分析的需求, 表现在: 1) 多维信息空间分析能力不足, 随着专题信息维的增多, GIS 常规空间分析方法与多维信息空间分析方法存在本质区别, 不能简单地前者用于后者, 否则分析性能会下降。2) 多维信息的概括性分析能力不足。在 GIS 实际应用中, 用户不仅需要得到详细的分析结果, 也需要得到概括性的决策型分析结果。此时, 概括性分析结果不仅体现在属性信息上, 还必须体现在地图图形的形状聚集上(如多边形合并、多边形叠加等)。

近年来国外新兴了一门软件技术称为 SOLAP, 即空间数据立方体。它来自不同领域的地理空间信息、专题信息 1…专题信息 n、时间信息按维的形式组成一个易理解的超数据立方体, 用地理空间维、专题维、时间维或地理空间维、专题维 1…专题维 n、时间维来描述空间对象, 每个维彼此垂直, 用户所需的分析结果就发生在维的交叉点上, 通过维的不同操作可产生不同的空间分析结果, 以满足多维信息空间分析和概括性分析的需求。空间数据立方

体是当前 GIS 界研究的热点之一, 是解决上述 GIS 面临问题的有效途径。

## 1 国内外研究现状

目前国内对空间数据立方体的研究几乎为空白, 国外的研究也刚刚起步, 关于空间数据立方体的论文数量有限, 且局限于加拿大、美国、澳大利亚、奥地利、德国和中国香港等少数发达国家和地区。

Stefanovic 等<sup>[1]</sup> 提出进行空间数据立方体高效分析操作的目标选择视图算法, 探讨由空间维、非空间维和度量组成的空间数据立方体模型, 研究空间数据立方体计算方法和分析处理过程, 研究重点是基于目标的选择实现视图化。Shekhar 等<sup>[2]</sup> 提出地图数据立方体作为空间数据仓库可视化工具的概念, 探讨在地图数据立方体上进行上钻、下翻和其他分析操作的原理, 并利用 1990 年的人口普查数据论述了地图数据立方体的概念。Zhou 等<sup>[3]</sup> 提出空间数据立方体中高效的多边形合并算法, 因为在空间数据立方体中, 用户常常聚集或重新分类空间目标, 因此代表不同空间目标的多边形需要随着标准的变化而进行合并。若处理大量的多边形, 这种操作处理的代价太高, 因而研究了两种多边形合并算法。Papadias 等<sup>[4]</sup> 提出空间数据立方体高效操作的原理和空间数据库存储空间目标的详细信息, 然而在许多应用中, 用户仅仅需要概括性的数据。这些信息能从空间数据库中获取, 但计算代价太高, 致使在线处理不实用。为解决这个问题, 专门构建了一个基于最佳空间粒度的空间索引分类分层算法, 将分层用空间数据立方体的格模型描述, 以便能处理任意的概括性数据聚集。Kamp 等<sup>[5]</sup> 提出用于传染病数

收稿日期: 2005-06-29; 修订日期: 2005-09-22

基金项目: 国家自然科学基金资助项目(40273042)

作者简介: 邹逸江(1963—), 男, 教授, 博士, 从事空间数据仓库和空间数据立方体的研究与建立。E-mail: zvjw@163.com

据分析的空间数据立方体原理与方法, 探讨利用空间数据立方体概念进行传染病数据的统计分析和地图显示。Wang 等<sup>[6]</sup>提出基于 Peano 树的高效空间数据 OLAP 操作算法, 探讨建立一种 PD 数据立方体, 在此基础上利用 Peano 树进行快速分析操作的原理和方法。Prasher 等<sup>[7]</sup>提出利用空间数据立方体进行空间目标多分辨率合并的原理和方法, 由于空间目标数量很大, 在线实时聚合并空间目标耗时太多。为解决这个问题, 采取预先存储多种不同分辨率的合并好的空间目标, 以满足不同用户的实时需求。Raubert 等<sup>[8]</sup>提出利用四叉树集成地理空间数据到 OLAP 系统中的原理与方法等。

总之, 上述研究虽然对空间数据立方体的某些方面进行了初步研究和论述, 但在概念、原理、结构、操作与算法等方面缺乏深入而细致的研究, 没有形成较完整的空间数据立方体概念体系。

2 基本概念

2.1 维

维是人们观察对象的角度, 如土地部门(土地类维)关心土地面积变化情况, 所以时间是一个维(时间维); 同时关心土地的分布情况, 所以地理分布是一个空间维(地区维)。由此可见, 土地类维、时间维、地区维决定土地面积的大小和变更情况, 其数据立方体如图 1 所示。

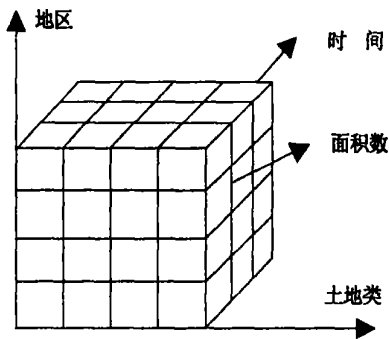


图 1 空间数据立方体结构  
Fig 1 The cube structure of spatial data

2.1.1 非空间维 它是一种仅包含非空间数据的维, 在空间数据立方体中, 把与地理要素关联的各种专题信息和时间信息作为非空间维, 如气象条件、经济结构、民族文化等。因此, (非空间维 1, 非空间维 2, ..., 非空间维 n)构成空间数据立方体的非空间维世界。

(1)非空间维可以存在细节程度不同的多个描述方面, 称为非空间维的概念分层, 它是对非空间维的细节描述<sup>[1]</sup>。如可以从年、月、日、周、时、分、秒等层次来描述时间维概念分层。因此非空间维的概念

分层是按维的实际意义进行树的分层。

(2)非空间维的格模型分层是空间数据立方体中非空间维所有组合分层排列。假如某个空间数据立方体有 n 个非空间维, 那么维的格模型有 n+1 个不同等级的分层, 2<sup>n</sup> 种维的排列组合。图 1 空间数据立方体(假设地区维是非空间维)非空间维的格模型分层如图 2 所示。

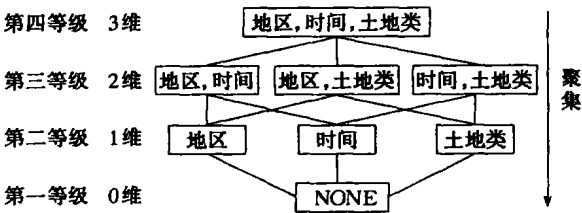


图 2 非空间维的格模型分层  
Fig 2 The grades of non-spatial dimensions for lattice model

非空间维格模型分层是一种直图, 代表空间数据立方体维聚集结果之间的相互依赖关系。直线从上向下(即沿着聚集的方向)代表从一种聚集结果可导出另一种聚集结果, 如: 聚集结果<地区, 时间>可由<地区, 时间, 土地类>导出, 此时土地类维被聚集掉了。非空间维的概念分层与格模型分层有本质区别, 一般格模型分层强调的是若干维的聚集结果以及聚集结果间的相互关系, 而概念分层强调的是维的成员取值路径。

(3)非空间维的组成元素称为维成员, 维成员是维的概念分层从树根节点到叶节点路径上所有节点的组合, 表示为: 非空间维成员: (维. 成员 1. 成员 2. ... 成员 n), 其中 1, 2, ..., n 代表维的层次。时间维成员的形象化表示是: “时间维. 1995 年. 2 月. 18 日”。

2.1.2 空间维 地理要素中的空间信息以及属性信息构成空间数据立方体的空间维, 两者所起的作用不同。属性信息构成非空间维, 由它和其他非空间维确定所描述的对象; 空间信息仅作为所描述对象结果的地图图形显示背景, 其本身的平面形状、空间位置及空间关系不会改变。

(1)空间维与非空间维的概念分层相似, 本文主要探讨空间维空间信息和属性信息的概念分层。空间维的空间信息可按地理要素的几何特征进行概念分层, 以图 3 为例进行说明。

空间维的属性信息可按地理要素的属性特征进行概念分层(图 4), 属性特征可划分主码、识别码、描述码、质量码、参数码、名称等层次。

(2)对于空间数据立方体, 空间维不能参与格模型分层, 只有非空间维才能进行格模型分层, 因为没有空间维的数据立方体就不是空间数据立方体。

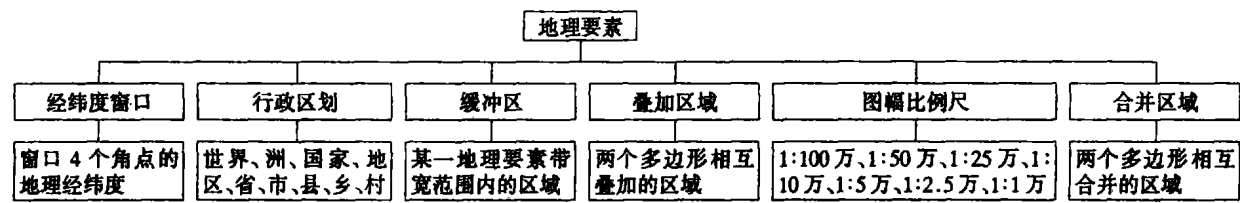


图 3 空间信息的概念分层  
Fig. 3 The conceptual hierarchy of spatial information

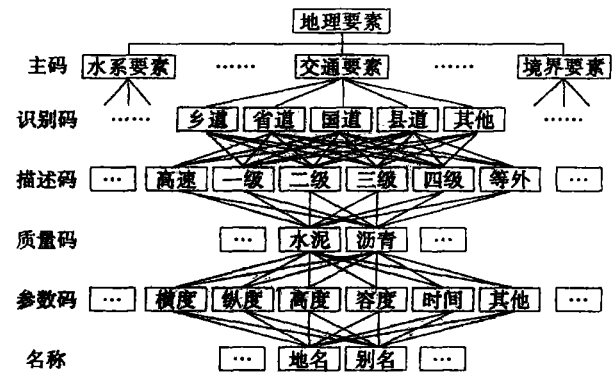


图 4 空间维属性信息的概念分层  
Fig. 4 The conceptual hierarchy of attributive information

(3)空间维成员是空间维属性信息概念分层中树根节点到叶节点路径上节点的组合,可表示为:空间维成员:: (空间维.成员 1.成员 2. ...成员 n),其中 1, 2, ..., n 代表维层次。在空间数据立方体中,空间维成员的通用表现形式为:空间维成员:: (空间维.主码.识别码.描述码.质量码.参数码.名称),例如:空间维.交通要素.国道.高速公路.路宽.铺面质量.107 国道。

2.2 度量

度量是由全部维或若干维确定的某对象属性值,度量的测量值都发生在维的交叉点上,度量往往不局限于一个,(度量 1,度量 2, ..., 度量 n)共同构成度量世界。空间数据立方体不同于数据立方体,它具有数字度量和空间度量<sup>[1]</sup>: 1)全部维或若干维确定的某个对象往往是一非地理空间对象,描述该聚集结果需用数字度量来表述。数字度量是一种包含数字值的度量,进一步划分为分布式的、代数的和整体的数字度量<sup>[3]</sup>。2)全部维或若干维确定的某个对象往往是一地理空间对象,描述该聚集结果需用空间度量来表述<sup>[4]</sup>。例如,地图上相邻多边形具有不同的数据值,当用户进行概括性操作时,相邻多边形具有相同描述数据值的可能性会增加,此时具有相同描述数据值的相邻多边形可合并为一个较大的多边形,该多边形就是地理空间对象的聚集结果。

2.3 空间数据立方体

空间数据立方体的通用表达式为:空间数据立方体:: (非空间维 1,非空间维 2, ..., 非空间维 n,

空间维 1,空间维 2, ..., 空间维 m,数字度量 1,数字度量 2, ..., 数字度量 i,空间度量 1,空间度量 2, ..., 空间度量 j),n,m,i,j 为正整数。

本文只讨论两种常用的空间数据立方体,其通用表达式为:空间维数据立方体:: (非空间维 1,非空间维 2, ..., 非空间维 n,空间维,数字度量 1,数字度量 2, ..., 数字度量 m);空间度量数据立方体:: (非空间维 1,非空间维 2, ..., 非空间维 n,数字度量 1,数字度量 2, ..., 数字度量 m,空间度量)。n,m 为正整数。

由于维的性质和度量的性质不同,图 1 可以是空间度量数据立方体,此时时间、土地类和地区名都是非空间维,度量为面积数和合并多边形地图,其中面积数是数字度量,合并图是空间度量;也可以是空间维数据立方体,此时时间、土地类为非空间维,地区名是空间维,面积数是一个数字度量。

3 基本概念和操作过程

为支持空间数据立方体的分析操作,文献[5]和[9]叙述了空间数据立方体概括分析、局部分析、全局分析和旋转分析的基本概念和操作过程。

3.1 概括分析

概括分析就是改变空间数据立方体的维数或维的层次,变换分析的概括度和粒度,主要满足用户基于多维数据的综合分析。概括分析主要基于钻取分析,它包括维上钻、维下翻和维层次上钻、维层次下翻<sup>[1,2,4]</sup>。

(1)维上钻操作是在空间数据立方体上综合一个或更多的维(空间维除外),聚集出相应度量值的过程。对于非空间度量,维上钻操作的结果是聚集出更概括性的数字度量;对于空间度量,就是执行地理空间对象聚集操作(存储聚集前的地理空间对象)。

(2)维下翻操作是维上钻操作的逆操作,是在已上钻的基础上,额外产生一个或更多的维(空间维除外)并计算相应度量的过程。对于非空间度量,维下翻操作的结果是得出更细致的数字度量;对于空间度量,就是得出上钻操作前的地理空间对象(调出存储的地理空间对象)。

(3)维层次上钻操作是在空间数据立方体上不综合任何维,在某维的层次链上按照其维成员从低等级到高等级进行概括赋值操作(空间维也不例外)聚集出相应度量值的过程。对于非空间度量,维层次上钻操作的结果是得出比较概括性的数字度量;对于空间度量,就是执行地理空间对象聚集操作(存储聚集前的各种地理空间对象)。

(4)维层次下翻操作是维层次上钻操作的逆操作。在维层次上钻的基础上,不增加任何维,在该维的层次链上按照其维成员从高等级到低等级赋值操作(空间维也不例外)计算出相应度量的过程。对于非空间度量,维层次下翻操作的结果就是得出较细致的数字度量;对于空间度量,就是得出维层次上钻操作前的地理空间对象(调出存储的地理空间对象)。

3.2 局部分析

当空间数据立方体的维数多于 3 个时,人们往往不易观察度量随维的变化规律,此时固定一部分维,关心度量数据在剩余维上的分布是局部分析。局部分析往往用于对事物的局部状态分析,如果所选定的部分维只有一个,则是维切片,如果所选定的部分维有两个或两个以上,则是维切块<sup>[3]</sup>。1)在空间数据立方体中选定某一维*i*并取维成员(*V<sub>i</sub>*),所得的子集并计算出相应度量的过程称为在维*i*上的一个切片操作。一次切片固定一个维,切片的作用就是舍取一些观察角度,使人们能在更少的维上观察对象,关心度量数据在剩余维上的分布。2)维切块则是在两个或两个以上的维上进行切片。

3.3 全局分析

选定维*i*, ..., 维*j*, 并取相应的维成员(*V<sub>i</sub>*, ..., *V<sub>j</sub>*), 所得子集并计算出相应度量的过程称为全局分析,全局分析往往用于对事物的细致分析。

3.4 旋转分析

旋转分析是变换空间数据立方体维的角度,从不同的方向关心度量数据在维上的分布<sup>[4]</sup>。空间数据立方体的旋转并不改变度量值。

4 分析实例

利用空间数据立方体的概念和原理可解决 GIS 多维信息空间分析中的分布分析、连接分析、缓冲区分析、叠加分析、网络分析和合并分析等,本文仅介绍 GIS 中使用频繁的缓冲区分析。多维信息结构的缓冲区分析可用空间维数据立方体<sup>[7]</sup>的概念和原理来解决。假设空间维数据立方体由 8 个非空间维

(气象条件、经济结构、民族文化、通信设施、部队结构、通行能力、科学教育、医疗卫生)、一个空间维(交通要素中的高速公路)、数字度量(居民地数)和空间数字度量(点状居民地位置)组成(图 5)。

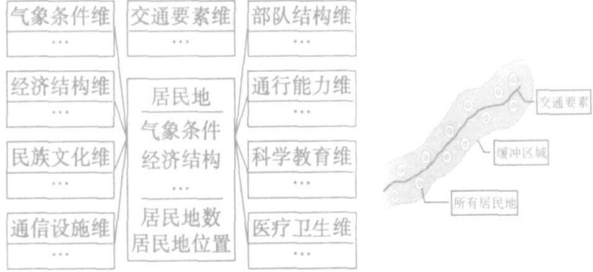


图 5 空间维数据立方体星形/雪片模型  
Fig. 5 The star/ snowflake model of spatial dimension data cube

图 4 是空间维交通的概念分层,图 6 是该空间维数据立方体非空间维的概念分层,表 1 是该空间维数据立方体非空间维、空间维、数字度量、空间数字度量值的组成。表 1 中的“取值”代表某维的成员值,  $\{X, Y\} = \{x_1y_1, x_2y_2, \dots, x_ny_n\}$ , *n* 是图 1 缓冲区内满足某些条件的居民地数即数字度量,  $\{X, Y\}$  是图 1 缓冲区内相应的居民地位置即空间数字度量,“固定值”代表某维的成员值固定,“NULL”指该维被综合或对数字度量不起作用。

气象条件维	⊃(平均降雨量、年平均气温、年平均日照量、降雨量最多季节、降雨量最少季节、降雪季节)
经济结构维	⊃(工业总产值、农业总产值、商贸总产值、军工总产值、信息总产值、人均收入)
民族文化维	⊃(人口数量、男女比例、宗教信仰、文化程度、后备预力量、主要民族、科学教育、医疗卫生)
通信设施维	⊃(有线通信能力、无线通信能力、通信枢纽、通信台站数、主要通信设备、通信器材厂数)
部队结构维	⊃(部队番号、部队级别、人员数量、兵种性质、战斗力、武器装备、机动性、所属军区)
通行能力维	⊃(铁路、公路、水路、航空、交通枢纽、车站吞吐量、港口吞吐量、机场吞吐量)
科学教育维	⊃(小学数量、中学数量、大学数量、科研机构数量、民办教育数量、高科技公司数量)
医疗卫生维	⊃(医科大学数量、医院数量、医生数量、总床位数)

图 6 非空间维的概念分层  
Fig. 6 The conceptual hierarchy of non-spatial dimension

由图 7(图中黑圈代表选中的居民地,白圈代表未选中的居民地)可以看出:空间维数据立方体的概括分析产生表 1 中第一行和第二行数据以及图 7 中图形 A 和 B;局部分析产生表 1 中第三行数据和图 7 中图形 C,此时“经济”和“通行”非空间维成员值固定;全局分析产生表 1 中第四行数据和图 7 中图形 D。图形 A、B、C 分别是图形 B、C、D 上钻操作的结果,图形 D、C、B 分别是图形 C、B、A 下翻操作的结果。由图 7 可知,图形 A、B、C、D 的数字度量分别是 12、6、4、3,空间数字度量分别是黑色圈形居民地。

表 1 空间维数据立方体数据组成  
Table 1 Data content of spatial dimension data cube

气象	经济	民族	通信	部队	通行	科教	医疗	交通	居民地数	居民地位置
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	高速公路	12	{X, Y}
NULL	取值	NULL	NULL	取值	NULL	NULL	取值	高速公路	6	{X, Y}
取值	固定值	取值	取值	取值	固定值	取值	取值	高速公路	4	{X, Y}
取值	取值	取值	取值	取值	取值	取值	取值	高速公路	3	{X, Y}
...	...	...	...	...	...	...	...	...	...	...
NULL	取值	NULL	NULL	取值	NULL	NULL	取值	高速省道	9	{X, Y}
...	...	...	...	...	...	...	...	...	...	...
NULL	取值	NULL	NULL	取值	NULL	NULL	取值	一级县道	9	{X, Y}
...	...	...	...	...	...	...	...	...	...	...
NULL	取值	NULL	NULL	取值	NULL	NULL	取值	二级乡道	9	{X, Y}
...	...	...	...	...	...	...	...	...	...	...

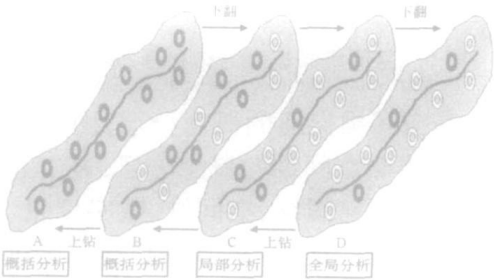


图 7 分析操作过程  
Fig. 7 The process of analysis operation

5 结语

为了比较多维信息空间分析的优劣和效率, 用气象条件、经济结构、民族文化、通信设施、部队结构、通行能力、附属设施、科学教育、医疗卫生 9 个非空间维的详细数据与地理空间维和时间维数据组成多维数据框架, 进行常规 GIS 缓冲区分析和空间数据立方体缓冲区分析, 得出如下结论: 1) 当只有 0 个非空间维时, 两者空间分析的效率一样; 2) 当小于 4 个非空间维时, 两者空间分析的效率基本一样; 3) 当大于 4 个小于 9 个非空间维时, 两者空间分析效率出现较大差别; 4) 当大于 9 个非空间维时, 两者空间分析效率出现明显差别, 此时常规 GIS 系统的效率明显低于空间数据立方体, 而且随着非空间维数的增加, 常规 GIS 系统的分析效率呈几何级数下降趋势。

空间数据立方体提高了多维空间数据分析的速度, 更符合用户的实际需求, 尤其是“数字城市”的构

建。但是, 空间数据立方体是一个新兴领域, 其复杂的多维数据模型设计、有效的多维数据组织存储、高效操作的立方体选择视图和较大的存储空间需求等是今后急需研究解决的问题。因此, 应结合“数字城市”的建设, 大力开展这方面的研究。

参考文献:

[ 1 ] STEFANOVIC N, HAN J W. Object-based selective materialization for efficient implementation of spatial data cubes[ J]. IEEE Transactions on Knowledge and Data Engineering, 2000, 12(6): 1-21.

[ 2 ] SHEKHARS, TAN X. Map Cube: A Visualization Tool for Spatial Data Warehouses. <http://www.cs.umn.edu/research/shashi-group/>, 2005.

[ 3 ] ZHOU X F, HAN J W. Efficient Polygon Amalgamation Methods for Spatial OLAP and Spatial Data Mining[ C]. USA: Sixth SSD, 1999. 121-125.

[ 4 ] PAPADIAS D. Efficient OLAP Operations in Spatial Data Warehouse[ R]. HongKong: Technical Report HKUST- CS01- 01, 2001. 65-69.

[ 5 ] KAMP V, SITZMANN L. A Spatial Data Cube Concept to Support Data Analysis in Environmental Epidemiology[ C]. Canada: Proc Ninth SSDBM' 97, 1997. 234-238.

[ 6 ] WANG B Y, PAN F. Efficient OLAP Operations for Spatial Data Using Peano Trees[ C]. San Diego: DMKD' 03, 2003. 126-130.

[ 7 ] PRASHER S, ZHOU X F. Multi-Resolution Amalgamation: Dynamic Spatial Data Cube Generation[ C]. Dunedin: 15th Australasian Database Conference(ADC2004), Conference in Research and Practice in Information Technology, 2004. 345-349.

[ 8 ] RAUBER A, TOMSICK P, RIEDEL H, et al. Integration Geo-Spatial Data into OLAP Systems Using a Set-Based Quad-Tree Representation[ C]. Canada: Proc Ninth SSDBM' 97, 1997. 256-260.

[ 9 ] 邹逸江. 空间数据立方体分析操作原理[ J]. 武汉大学学报(信息科学版), 2004, 29(9): 822-825.

The Key Technology of Multi-Dimension Spatial Analyses: Spatial Data Cube

ZOU Yi-jiang

(College of Architectural Civil Engineering and Environment, Ningbo University, Ningbo 315211, China)

**Abstract** This paper introduces researches on spatial data cube on abroad and home, and explains the concept and structure of non-spatial dimension, spatial dimension, digital measure and spatial measure of spatial data cube, and gives definition of spatial data cube. This paper states basic concept and principle of analytical operation on spatial data cubes, and also indicates that analytical operation comprises recapitulative analysis, local analysis, global analysis and pivoting analysis. According to these, this paper recites completely function and structure of above analysis and detail process of analytical operation for multi-dimension analyses. In the process of analysis the methods that show example for these analysis functions are adopted in order to make reader more intelligible.

**Key words:** GIS; spatial data cube; spatial data warehouse; multi-dimension spatial analyses