

分布式空间对象关系运算的研究

杨超伟 李琦 王京傲

(北京大学遥感与地理信息系统研究所, 北京 100871)

摘要 针对当前面向网络 WebGIS 的特点, 对空间关系运算进行了系统的研究以提供分布式空间分析的需要。采用数据抽取率与运算复杂性两个定性指标来考虑空间关系运算的分布式方案, 设计了空间对象关系运算的分布式方案和相应的空间分析 Web 传输协议。并在 WebGIS 中得到了实现。

关键词 空间对象 空间分析 分布式计算 WebGIS 传输协议

0 引言

空间对象关系运算是地理信息系统中进行空间分析的实质与核心问题。为了发展 GIS 的空间分析能力, 科学家们对此进行了大量的研究, 取得了一系列成果并成功应用于 GIS 中以支持其地理空间分析。在当前面向网络的 WebGIS 中, 空间分析与单机 GIS 系统中的空间分析具有许多由 Internet 本身特点而引出的不同之处。需要对空间对象关系运算进行系统的研究以提供分布式地理空间分析的需要^[1]。本文对空间对象关系运算进行了系统的研究, 依据 Internet 上 WebGIS 的特点, 给出了分布式空间分析方案, 设计了空间对象关系运算的分布式方案和相应的空间分析 Web 传输协议, 并实现于 WebGIS 中。

1 面向 Internet 的 WebGIS 的特点

Internet 与 WWW 的广泛流行, 给地理信息系统软件的发展带来了巨大的机会, GIS 软件开始由桌面转向网络平台, 从而满足了当前大量地理信息的分布式特点和广泛共享的需要。GIS 本身由许多功能模块组成, 在将 GIS 由桌面转向 Web 计算平台时, 就必需要对 GIS 的这些模块进行分布, 以适合当前 Web 计算平台的需要。在 Web 计算环境中, 一般采用 C/S 结构的形式, 就是说采用服务器来存储数据, 进行大量的运算, 而用客户机来进行少量的运算并与用户进行交互。服务器与客户机间则基于 HTTP

协议采用统一资源定位器 (URL) 来进行通信导航, 由于 Web 计算平台的特点和 GIS 运算特点产生了以下 WebGIS 的特点:

(1) WebGIS 处理信息源相对集中。一般 WebGIS 地理信息数据都处于服务器上, 客户机则只是向服务器发出请求来获取满足需要的地理信息并进行相对简单的处理。

(2) 计算量大。由于 GIS 涉及复杂的空间关系运算, 而且空间数据量大, 所以计算量是非常大的。

(3) 独特的网络传输协议, 由于 WebGIS 建立在 Web 计算平台上, 而 Web 采用的 HTTP 协议并不支持空间对象的传输。因而需要特定的传输协议。

(4) 数据传输速度慢。这是由于目前 Internet 本身带宽太窄的原因。

空间关系运算及基于此的空间分析作为提取地理信息的方法, 它们的分布式方案将直接影响到 WebGIS 的响应速度, 本文研究采取合适的空间关系运算分布式方案及相应的传输协议来满足 WebGIS 中空间关系运算的分布式需要。

2 空间关系运算的理论组织模型^[2,3]

空间对象关系运算是建立空间分析的基础, 因而有必要对空间对象及其种类, 空间对象相互间的关系与运算进行定义, 它们之间存在以下模型关系, 即空间分析建立在空间对象关系运算基础之上, 空间对象关系运算建立在空间对象及其相互间的关系

之上。

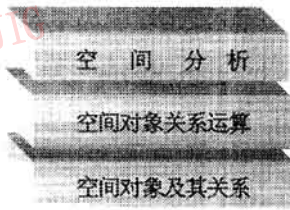


图1 空间分析的实现组织模型

2.1 空间对象

空间对象是指定位于地理空间中具有一定属性的几何体,空间分析大多是基于矢量图形数据及二维空间中的,因此主要研究二维空间中的图形空间对象。它由点、线、面等构成,但在实际处理空间对象时,会遇到一些复杂情况,如表示居民地的所有点对象作为一个对象。因而成为多点组成一个空间对象的情况,按空间对象本身的特点和人们观察空间对象的特点,空间对象可以归结为以下7类:

(1) **单点** 具有坐标空间中唯一一组坐标值的空间对象,可以后式表示: $SP:(x, y)$;

(2) **多点** 具有数个单点组成的点空间对象,可以后式表示: $MP:(SP_1, SP_2, SP_3, SP_4, \dots, SP_n)$;

(3) **简单线** 由一条不自交的线组成的线空间对象,可以后式表示: $SL:(SP_s, SP_e, \dots, SP_e)$, 其中 SP_s, SP_e 分别为线的起点与终点;

(4) **自交线** 由一条与自己相交的线组成的线空间对象,可以后式表示: $SIL:(SP_s, SP_2, P_3, \dots, SP_m, \dots, SP_m, \dots, SP_n)$;

(5) **多部简单线** 由多条简单线组成的空间对象,可以后式表示: $ML:(SL_1, SL_2, SL_3, \dots, SL_n)$;

(6) **单面** 由一片连续的区域组成的空间对象,可表示如后: $SA:(SL_1, SL_2, SL_3, \dots, SL_n)$;

(7) **多部面** 由多个连续的区域组成的空间对象,可表示如后: $MA:(SL_1, SL_2, SL_3, \dots, SL_n)$ 。

2.2 空间对象的关系及空间对象合成运算

空间对象关系及运算是空间分析的基础,基于上节中的空间对象,本节阐述了这些空间对象之间所具有的空间关系及其运算。

以上7种空间对象之间的关系有:

(1) **相等** 当两个空间对象的类型及其表达式完全一样时,这两个空间对象具有相等的关系。

(2) **接触** 若两个空间对象的边界有相互重合的点,且除线对象外只有边界点重合,则两个空间对象具有接触关系。

(3) **包含** 若两个空间对象中的一个完全包含另一个的所有点,则此空间对象包含另一个空间对象。

(4) **被包含** 若两个空间对象中一个空间对象的所有点完全包含于另一个空间对象之中,则此空间对象包含于另一个空间对象。

(5) **相交** 若两个空间对象具有除边界点以外的公共点,且公共点不等于任何一个空间对象时,此二空间对象相交。

(6) **重叠** 若两个同种空间对象具有除边界点以外的公共点,且公共点不等于任何一个空间对象,则此二空间对象重叠。

(7) **不相关** 若两空间对象无公共点,则此二空间对象不相关。

在现实世界中,有时需要将两个或几个空间对象合成为一个空间对象。根据以上的空间对象关系,可以具有以下几种空间对象运算:

(1) **求交** 将两个空间对象的公共部分提取出来作为一个新的空间对象的运算为空间对象求交。

(2) **求余** 将一个空间对象与另一个空间对象相交结果除去后形成的一个新的空间对象的运算为空间对象求余。

(3) **对称求余** 两个空间对象除去它们的公共部分形成一个新的空间对象的运算为空间对象求和。

(4) **联合** 两个空间对象的所有部分之和形成一个新的空间对象的运算为空间对象求和。

(5) **剪裁** 用一个特定的面状空间对象(如矩形、圆)来截取一个空间对象形成一个新的空间对象的运算为空间对象的裁剪运算。

(6) **缓冲区** 将一空间对象边界扩充一定尺度而形成一个新的空间对象的运算。

2.3 空间对象的空间分析类型

空间分析是GIS不同于其他信息系统的主要特点之一,它是建立在以上空间对象的关系运算的基础之上的,参照正在形成的SQL3中的SQL/MM第三部分,主要的空间分析有以下几种:求终点/起点、连接、求中心、求最小外接矩形、重叠、缓冲区、包含、接触、求面积、求长度、求周长、求斜率、求联合、求最近、求最远。这些空间分析都可以根据空间对象及

其关系来进行运算得到。

2.4 空间关系运算的复杂性与数据抽取率

为进行空间关系运算的分布式方案研究,可以定义两个量来衡量空间运算的效率。空间运算的复杂性是指一个空间运算所需耗用的时间,它与运算类型、运算维度等有关,比如乘除运算慢于加减运算,也慢于逻辑比较运算,面的运算比线的运算复杂,线的运算比点的运算复杂。复杂性有三个等级:低、中、高,一般只有比较、遍历、加减等运算的为低,而包含大量乘除运算的则为高。数据抽取率是指在一定的地理信息中抽取有用信息的比例,在相同的数据中抽取出的数据越多,数据抽取率就越高,它也可分为高、中、低三个等级。

由前面对空间对象的定义中可以看出,所有的空间对象最后都实际表示为点,遵循面、线、点的表示递推关系。因此,在实现这些空间对象的关系运算功能时,是基于点运算的。在所有的空间数据中,点的表示都是有一定的精度的,所以在逐点运算时,所取的步长是约等于这一精度的。根据点、线、面的组织方式,具体运算时是以面、线、点来递推运算的。其中有几个重要的基本算法:

线上点的遍历 可由 Bresenham 算法来进行点的遍历,具体参见参考文献。它的运算全由加减、逻辑比较等简单运算组成。

点在面内 可以对面按坐标标定左右位置。对面的边界点按顺时针遍历,当遍历点与所求点有相同的横坐标时,如所求点在此点的右边,对所求点的某一 BOOL 量取反,当遍历后,这一 BOOL 量的值不变,这点不在面内,否则,点在面内。它仅包括加减与比较等简单运算。

线线相交 这可从解析几何中找到运算方法。它的运算包括乘除及开方等复杂运算。

这三种运算的复杂程度依次递增。根据这些运算方法可以实现空间关系运算,同时可大致预测出运算的复杂性与数据抽取率。分别如下:

相等 对两个空间对象顺次取点、比较,若点的个数与对应相同点的坐标都相等,则此二空间对象相等,这一运算简单而且数据抽取率低。

接触 对两个空间对象的边界遍历,看有无重复点,如有,则再遍历边界点,看是否有一个的边界点在另一个内部,如无,则此二空间对象接触,否则,此二空间对象不具有接触关系,所以这一空间关系运算的复杂性低,数据抽取率中等。

包含 对一个空间对象的点进行遍历,看是否都在另一空间对象中,若全在,则存在包含关系,否则,不存在包含关系。这种空间关系运算复杂性中等,数据抽取率也为中等。

被包含 与前面的运算相似。性质也类似。

相交 取一空间对象的表示点,逐一取两点组成线段与另一空间对象的所有表示点相邻点组成的线段比较,看是否相交,若相交,则此二空间对象存在相交关系,这种空间关系运算的复杂性高而且数据抽取率高。

重叠 与以上的运算类似。这种空间关系运算的复杂性高而且数据抽取率高。

不相关 对一个空间对象所有点遍历,如没有一个点在另一个空间对象内,则此二空间对象不相关。这种空间关系运算的复杂性中等而且数据抽取率高。

对各种空间对象运算:由于是建立在以上运算基础上的,而且还有重新组合等多种运算,所以,它的空间运算复杂性高而且数据抽取率低。

将以上的分析结果列于表 1。

表 1 空间关系运算复杂性与数据抽取率

空间关系运算	运算复杂性	数据抽取率
相等	低	低
接触	低	中
包含	中	中
被包含	中	中
相交	高	高
重叠	高	高
不相关	中	高
空间对象运算	高	低

3 基于 Web 的空间对象关系运算分布式方案及其传输协议的设计与实现

3.1 空间关系运算分布方案

在 Internet 的 C/S 结构中,服务器端主要提供对数据的组织管理功能和数据提取功能。客户端主要提供用户的数据分析功能并在一定的条件下向服务器发数据请求。空间关系运算在 C/S 结构的 Internet WWW 中主要用于服务器端的数据提取与客户端的

数据分析。空间分析的作用见图 2。

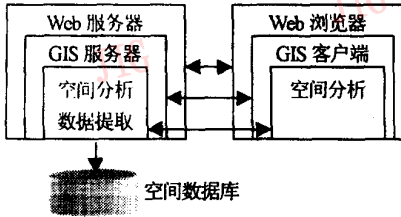


图 2 空间分析在基于 Internet 的 GIS 中的作用

如图,在每次客户端操作时,都需要一定量的满足一定空间关系条件数据。这一数据的数据量相对目前 Internet 上的传输速率限制来说是比较大的,因而应尽量减小这些数据在 Internet 上的传输。它一方面可以从服务器端的数据分析提取模块来减小数据量,只使满足空间关系条件的数据在网上传输。另一方面当客户端已经有前几次传来的数据时,应充分使用这些数据,同时要保证数据不能满足需要时,能向服务器发合适的数据请求。然而,客户端的空间分析能力不能太多,因为要保证在目前信息集成的前提下就必需使空间分析模块尽量小以嵌入到 WWW 浏览器中使用。因此,空间运算复杂性低而数据抽取率高的应分布于客户端(如不相关),而空间运算复杂性高数据抽取率低的应放在服务器端(如空间对象运算),这样可以充分利用服务器的高速运算与存有大量地理信息的特点,而且可以减小网络传输量。按这一思路,可将以下空间关系运算放在客户端:相等、不相关、包含、被包含、接触。

为保证数据对空间分析的需要,对那些客户端不能满足数据要求的空间分析应重定位到服务器端,也就是说在客户端加一个重定向远程数据提取模块。

服务器端进行数据的提取,为增加提取效率与命中率,将所有的空间关系运算都在服务器上实现。

基于以上考虑,空间对象关系运算的分布及其在 C/S 模式中的作用可表示如图 3。

3.2 空间对象关系运算的传输协议

Web 基于 HTTP 协议,而 HTTP 协议并不支持空间对象,但支持字符流,所以可据此建立空间对象关系运算的传输协议。

在 GIS 中空间对象的组织形式是按照图层集、图层、空间对象来组织的,一般一个 GIS 应用系统采用图层集来自动表示空间信息集。因此,在运算请

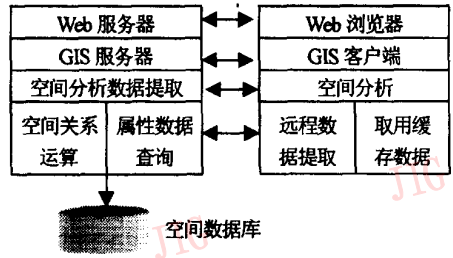


图 3 空间关系运算在基于 Internet 的 WebGIS 中的分布式方案

求协议中应指明空间关系运算的对象,即图层及相应的空间对象。而在运算返回协议中应详细指明图层、空间对象名称、坐标等信息。

根据以上特点,我们采用的协议如下:

空间运算请求 SELECT <目标对象层名>
WHERE <空间操作码> <缓冲区大小> <提交层名>
<提交空间对象 ID>

空间运算返回 <层名> <空间对象数 n> <空间对象 1 名称> <空间对象 2 名称> ... <空间对象 n 名称> <空间对象 1 坐标> <空间对象 2 坐标> ... <空间对象 n 坐标>

3.3 WebGIS 中空间分析的实现

基于 WebGIS 的图形显示模块,实现了 WebGIS 中的空间分析。如图 4,用户可根据地图集中图层对象及查询方法来任意组合自己的查询请求,图中选择为:SELECT districts WHERE 10 rings 1 表示查询与 rings 层中外环线相交的所有 districts 层中县空间对象。图 5 显示了相应的查询分析结果。其中表中显示了空间对象的 ID 及名称,地图则显示了空间对象在图上的位置。



图 4 WWW browser 中基于空间关系运算的空间分析查询提交界面

4 总 结

本文根据 GIS 由桌面到 Internet 的转变过程中空间分析的特点。对空间关系运算进行了系统的分析,并据此提出了空间关系运算的分布式方案及其相应的传输协议,在 WebGIS 中得到了实现。Internet 上 GIS 的研究刚刚起步,对基于 Internet 的空间分析研究也是全新的,对它的研究还有待于进一步深入和扩展。



图 5 WWW browser 中基于空间关系运算的空间分析查询返回界面

参 考 文 献

- 1 Newton P W. Networking Spatial Information System. Belhaven Press, 1992.
- 2 Mark Ashworth. Spatial Database Standards. A MapInfo White Paper, 1997.
- 3 ESRI Coporation. SDE C Developer's Guide. ESRI Corporation, 1997.

李 琦 现任北京大学遥感与 GIS 研究所及中国可持续发展研究中心数字地球软件工作室教授、博士生导师。已在国际、国内刊物上发表论文 50 多篇。研究领域:空间信息科学。



杨超伟 北京大学遥感与地理信息系统研究所博士生,已在国际国内刊物上发表论文数篇。研究领域:空间信息科学。

王京傲 北京大学遥感与地理信息系统研究所硕士毕业,研究领域:空间信息科学。

Research on Distributed Spatial Object Relation Computation

Yang Chaowei, Li Qi, Wang Jingao

(Institute of Remote Sensing and Geographic Information System, Peking University, Beijing 110006)

Abstract Based on the characters of spatial relation computation of Network-Oriented WebGIS, the authors made a systematic study on spatial relation computation to meet the demand of distributed spatial analysis. The computation complexity and data abstract probability were used to consider the distributed schema of spatial relation computation, and a respective schema and spatial analysis web-based transportation protocol were proposed and implemented in WebGIS.

Keywords Spatial object, Spatial analysis, Distributed computation, WebGIS, Transportation protocol