

GIS 中的数据转换与代数同构分析

李国建¹ 胡 鹏²

(1 中国四维公司, 北京市紫竹院百胜村 1 号, 100044)

(2 武汉测绘科技大学土地科学学院, 武汉市珞喻路 129 号, 430079)

摘 要 对 GIS 中常见的数据转换进行了细致的研究, 为数据转换中的质量保证提供了一种理论上严谨的检验手段, 首次提出了用代数同构的思想来检测网络环境下数据共享的可信度。

关键词 数据转换; 地理信息系统; 代数同构; 代数同态

分类号 P208; TP311.12 文献标识码 A

随着 GIS 在实际工程中的迅速发展, GIS 在理论和应用上都有了很大进步, 尤其是 Web-GIS 的突起和以开放共享型 OpenGIS 的研究为标志的体系框架的突破^[1], 为 GIS 的发展打开了新思路, 但同时也暴露了 GIS 存在的一些技术弱点。

不同的 GIS 系统对各种时空现象的理解、描述方式、概念模型、数据结构、实现手段等也互不相同, 各有优势和局限。当用户需要使用新系统时, 首先考虑的是能否利用已有的数据, 这就涉及到数据的转换问题。GIS 软件系统是管理数据的一种工具, 数据是根本, 因此, 数据转换的质量和信息损耗是用户最关心的。为了便捷而高效地实现数据的传输和交换, 必须采用一定的规则对不同数据源的数据进行转化, 且必须制定和遵循地理信息数据转化的标准。在栅格标准中, 特征图像文件格式 TIFF 是目前常规空间数据最为重要的栅格标准。相关组织积极推动标准的确定, 在数据转换工作中发挥着重要作用。英国地理学会推荐 NTF 格式为空间数据交换的优选矢量标准格式; DGIWG 开发了数字地理信息交换标准格式; DIGEST 以及美国地质调查局开发了空间数据交换标准格式 SDTS; 中国也推出了数据交换格式 CNSDTF。数据转换中客观存在着信息损失和变形, 虽然人们对数据转换的可靠性存在疑问, 但若没有严谨和科学的方法来检验, 只根据设计的例子做实验验证或在实际应用中去逐渐发现错误, 其损耗、代价都是巨大的。因此有必要深入研究一些数据转换的问题。

1 数据转换原因

数据需要转换的根本原因是由于数据所处的保存、管理和使用的环境差异造成的。一般理解是, 早期 GIS 采用各自独立的理解方式开发自己的数据存储格式及令用户深感困难的系统之间的数据格式转换问题, 形成了数据间的语义差别。实际上, 出于各种考虑, 数据在一个系统内部也需要经过多次的格式转换。但除系统开发人员以外, 一般人不关心这种数据转换, 大多数人关心的还是异类系统之间在存储介质上发生的格式转换。

例如, 实现矢量数据向栅格数据的转换, 要把图形上的面元素转化为线元素, 再用适当的程序把线元素转换成某种分辨率的栅格数据格式。栅格数据到矢量数据的转化要复杂一些, 首先要对扫描所得到的密集像元形成的线划用细化处理算法予以细化, 同时, 还要用合适的算法剔除细化处理时产生的多余毛刺。另外, 要处理线划间断、重叠等问题, 数据转换过程中不可避免地要损失部分信息。要解决上述问题, 需依赖于计算机技术及 GIS 技术的发展。

需要数据转换的应用有几种类型:

1) 汇总。数据采集汇总的来源多样化。目前存在着各种地理空间数据的采集方式和相应的数据格式, 将不同的采集方式采集的数据转换到一个主系统中时, 就必须进行数据转换。如车载监控系统, 需要同时处理基础地形图、车载 GPS 信息、基站差分信息、道路路况等不同格式的数据。

2)继承。应用的主系统发生变化,需继承原系统数据。在使用一个新系统时,会先利用已有的数据来建立,而这些管理系统往往因为版本、厂家等原因,造成数据格式变化,必须转换。地理信息涉及历史数据的存储和管理的领域,如地籍信息系统,随着技术进步和系统的升级,数据格式也应变化。

3)共享。数据以多种格式存在,需要交流共享。在使用不同系统的各个部门和领域,自己所需的部分数据往往可以利用别人已采集管理的数据,因此需要实现数据共享交流,但数据转换经常成为一个瓶颈。这是目前建立 GIS 应用系统最普遍的问题。

4)集成。网络环境下,对异构多源(multi-source)数据系统进行远程浏览查询分析时,必须正确识别和操纵远端数据,达到数据格式对前端透明集成的目的。多源数据在系统中随着数据迁移和计算迁移^[1]而产生的数据转换是很频繁的。

研究多源数据的信息保持与丢失问题,对于提取 GIS 的信息、压缩数据容量、保证网络地理数据的迁移质量等应用有着非常重要的意义。

汇总和共享主要是发生在单系统中外部数据源进入系统时的转换,集成和继承是涉及多系统协同的数据转换。

数据转换的具体方式有两种:

1)静态转换。系统不能直接操作外部的异构数据,需要某种中介,从一种存储介质格式转换到另一种存储介质格式,如从 Maptitude CDF 文件转换到 Autodesk Mapguide 的 SDF 文件,在汇总、共享、继承方式下这是最常使用的方法。其优点是一次移植后便于以后的使用和进一步加工,缺点是有时需要保留格式转换前后的多种数据,造成同步维护困难,一致性难。特别是在共享方式中,原始数据有更新,则需要重新转换,而此时如果已经对上次转换过的数据做过紧密相关的工作就会引起很大麻烦,尤其是对于图形共享且属性进行各自专业关联的情况。

2)动态转换。系统可以直接操作外部系统的异构数据,不需要事先转换,格式转换在操作时动态完成,此时数据转换对用户透明,由后台借助内存和缓存完成。一般是仅对当前操作的数据集完成局部转换。特别是在逐渐兴起的 GIS 构件化模型体系中,这种动态转换的方式将有很大的应用前景。随着分布式 GIS 的研究推广,数据格式的动态转换将成为主流。

动态转换有的是单向转换,仅能读其他格式

数据;有的是双向转换,不仅读出后能以自己的格式操作,还能写回原格式。使用双向转换比较典型的软件是 Autodesk World,采用数据格式引擎可以直接读取异种数据,采用 Coverage、MID、DWG、DGN 等格式即可直接读原始格式,也可按原始格式存储,内存操作则采用独立格式。动态单向转换的典型应用是在网络环境下。网络 GIS 的结构大多是采用 Client/Server 模型,Multi-Tie 和 Browser/Web 等实际也属于传统双层 Client/Server 的扩展。前端用户看到的数据格式往往与服务器端存储的数据格式有很大差别。大部分服务器把管理的格式转换成栅格图像格式,也有的转成通用或专用的矢量格式,前端用户无论用何种通用浏览器还是用专业 Client 软件,接收操纵的数据都已经是被后台转换过的。这种动态单向转换主要方便于查询检索等领域,不能同时维护原始格式数据。

从体系结构上分析,静态转换中大多数采用以中间交换格式为中介的方式实现,数据是 $M:1:N$ 类型的流动,中间格式数据充当转换的数据中转站,目前制定的很多 GIS 交换标准就是这种数据。动态转换没有明显的数据中心,一般数据目标单一,属 $M:1$ 类型。很多动态转换有一个与格式有关的功能中心,集中了对各类格式的互相解译,充当了交换中介,形成系统的功能结构也是 $M:1:N$ 方式。这在构件化系统中最明显。

静态转换多采用外部标准交换数据文件。标准交换格式充当了不同 GIS 系统的公共接口的角色,标准的完备性、通用性、伸缩性、扩展性等是关键。动态转换大多采用内存或临时文件做转换数据的中介载体。

2 转换的质量评定标准

虽然数据转换的应用环境和方式有很多,但是如何评价这些转换是比较困难的。这涉及到信息期望和语义传递等复杂问题,目前研究的并不透彻。要解决该问题,不但要依赖于计算机技术及 GIS 技术的发展,还要依赖于对 GIS 理论的进一步研究。

在分布式地理信息系统中存在着大量的数据。如果数据是自己采集的,那么对数据的精度和可用性有一定把握。但是在分布式地理信息系统中存在着各种类型、各种来源、各种格式的数据,也有不知来源和可用性的数据。如加拿大的 DeltaX 计划,用户可以从其全球分布式系统中取

到任何地区的地图数据,这些数据可以由授权用户自由添加进去。即将建成的全球地图服务器^[2]将把全球的地形、植被、土地利用、水系、交通、行政区划、海岸线、土质植被等地图数据上载到因特网,依赖元数据等技术,对内容、范围、现势性等要求有时容易判断。相比之下,网络上提取或迁移过来的数据大多是格式发生变化的数据,如何判定和相信这些数据是否满足计算的要求,则是一个非常迫切需要解决的问题。

以往的格式转换大都是凭经验感觉和一定的试验验证来解决。软件开发人员使用白盒测试案例,用户使用黑盒测试案例,没有科学准确的方法来判定这种格式转换的成功程度。尤其对于用户,即使采用同样的数据、同样的软件转换出同样的结果,但由于对转换结果所关心的重点不同,其所要求的标准自然不同,这种转换的成功率也就不同。例如,数据从采用双精度数据的系统转到采用整数的系统,这种转换虽然存在信息精度的严重损失,但对于用户来讲可能是成功的,因为它只使用整数数据。Web-GIS中无论是由矢量数据转换为栅格数据,还是由栅格数据转换为矢量数据,转换过程中所要采用的数据模型有着显著的区别,但用户能查到关心的信息就算转换成功。

数据格式转换的过程必然是一种信息变形的过程。信息由一种形态转化为另一种形态必定发生了某种变化,因此信息的变化与否与我们所关心的信息类别内涵是密切相关的。一般所说的数据格式转换成功与否,是指用户所关心的那些信息有无变化,即与用户对数据信息的期望值有关。

信息的期望值又和语义有密切关系。人们所希望得到的不损失的信息是以各种数据语义的方式表现的。例如,对于道路的连通性,转换前后的道路数据应该是连通的,则即使原系统的拓扑结构数据不能转换,在新系统中对道路拓扑的重建应该在连通性上与转换前数据没有差别,拓扑不变量是相同的。这种连通语义表现在转换前后图形坐标数据本身满足一定精度容差范围,从而形成图形线段弧段的相邻关系。文献[6~8]用误差理论分析了数据精度的误差传递,但没有涉及数据格式转换造成的语义误差。例如两条道路临近的两个节点(335.467 8, 123.999 9)和(335.467 9, 124.000 1)在容差0.000 5的标准下是连通的,数据转换为整数系统后的(335, 123)和(335, 124)在容差1的标准下也是连通的。可见,这种连通语义与信息期望的精度有密切关系。

上节提到的引起数据转换的根本原因,在实

际应用中也是主要涉及到地理信息数据的语义变化。在不同系统之间,对同样的地理现象构造的数据结构模型不同,对信息的表述和侧重点也是不一样的,即语义表达有区别。在共享交流方式下,关注的也是信息语义是否相同。虽然信息载体的格式不同,但在数据转换过程中的语义传递不应有超过限差的差别。不同数据模型的语义描述差别造成了转换的困难和信息损失。

目前,由网络热潮所激发的对数据精度、目录数据、元数据等的研究,目的是解决数据共享的问题,其中也涉及到为访问异构地理信息数据服务的应用,它们对数据的描述其实也是语义表达的一部分内容。如分布于网络上的数据和程序,涉及有关空间检索的数据集语义的信息交换。构件化GIS颇受重视,如果不解决这些信息期望和语义传递等评价问题,没有一个共识,构件化的普及只能是局部的发展,不能形成地理信息化的主体。地理语义数据模型的研究包括:①地理数据语义模型设计的方法论研究;②地理对象体系的建立和实现;③空间关系的语义表达;④语义数据库的开发和地理数据描述语言及系统的开发等。

目前,规划资源—环境—人口—经济的协调发展是研究的大环境,资源管理、环境监测、灾害评估、自然区划、区域开发、土地荒漠化、植被变化、森林病虫害监测、水土流失、成矿预测、道路交通管理、土地利用管理、旅游资源管理、水利工程管理以及污染源控制、城市人口流动等都在建立各自行业、部门的应用区域性地理信息系统,这也是数字地球扩展的基本应用领域。这些领域对于地理信息数据的共享交流等存在迫切需求,动态转换和静态转换的需求都很强烈,尤其对数据的动态转换的可靠性应严格评价。因此,研究高效的数据传输和交换,需要找出数据转换信息的评价方法和准则。

3 代数同构分析

数据的转换方式有多种,究其实质,高效地实现数据的传输和交换只是信息载体的变化,转换本身涉及两个数据集中信息的迁移问题。转换前后信息的改变与否与两个数据集的标准、用途有关。转换前后的环境不同,就无法对数据实现直接的对比,依靠两个数据集使用各自的功能推出各自的结论是否一致来判断信息是否发生变化。如果转换前后两个数据集能导出同样的结论,则证明各自所含的信息没有发生变化。如果得到的

结论有差别,则差别的大小也就是数据转换所带来的。这正是科学的质量评定准则。这种数据集构造的对比正是现代数学所研究的主要问题之一。数据集以及各自数据集的功能的映射关系组成了代数同态的基本条件。这种对比是数学化的,其严密性、可靠性得到保障。

定义1 设 A, B 是两个给定的集合。如果通过某种对应法则 f , 使 A 中的每一个元素 a 对应于 B 中某个唯一的元素 b , 则称 f 为一个从 A 到 B 的映射, 记为 $f: A \rightarrow B$ 。 b 称为 a 在 f 下的像, 记为 $b = f(a)$, a 是 b 的原像。

f_1, f_2 称为相等的两个映射, 若它们都是 A 到 B 的映射, 且对每个 $a \in A$, 有 $f_1(a) = f_2(a)$ 。如果 a, a' 均属于 A , $a \neq a'$, $f(a) \neq f(a')$, 则映射 f 称为单射; 如果 $b \in B$, 皆存在 $a \in A$ 使 $b = f(a)$, 映射 f 称为满射; 若 f 既是单射又是满射, 则称 f 为一个一一对应。

定义2 设 A, B 是两个分别带有代数运算 \oplus 和 \otimes 的集合。若 $f: A \rightarrow B$ 是映射, 且 $a_1, a_2 \in A$, $a_1 \rightarrow b_1, a_2 \rightarrow b_2$, 即 $f(a_1) = b_1, f(a_2) = b_2$, 总有 $a_1 \oplus a_2 \rightarrow b_1 \otimes b_2$, 即 $f(a_1 \oplus a_2) = f(a_1) \otimes f(a_2) = b_1 \otimes b_2$, 则称 f 是 A 到 B 的代数同态。 f 为单射、满射时分别称为单同态、满同态。 f 为一一对应时, 称 f 为 A 到 B 的代数同构, 表示为 $A \cong B$, 这时也称 A 与 B 是代数同构的集合。从 A 到 A 的代数同构称为自同构。各种运算 \oplus 和 \otimes 组成各自所在集合的代数运算系统 (\oplus) 和 (\otimes) 。

从抽象观点看, 代数同构的集合在代数构造上没有什么区别。它们具有同样的关于代数运算的性质。因此, 如果 A 有一个性质, 这个性质是完全可以用代数运算计算出来的, 那么 B 就对应一个完全类似的性质, 也可以用运算得到。这就是说, 若是仅就代数运算对 A, B 所产生的影响来看, A 与 B 只有形式上的不同, 而没有本质上的区别。这种性质就是信息内涵没有发生变化。

对于 GIS 的数据转换来讲, A 是转换前系统管理下的原始数据, 称为原数据 A ; B 是经过对 A 的转换后系统所管理的数据, 称为现数据 B 。在转换前的数据管理系统下的功能操作的集合体就是 A 的代数运算系统 (\oplus) , 转换后的数据管理平台系统所进行的功能分析操作的集合就是 B 的代数运算系统 (\otimes) , 运算 (\oplus) 和 (\otimes) 是对等的单射, 这个转换可以看成是 A 到 B 的一个代数映射 f 。如果这个转换达到理想状态的无信息损失, 原数据 A 中包含的信息完全等同于现数据 B 中包含的信息, 那么在原数据 A 中经过查询分析能得到的结

论成果, 在现数据 B 中经过相应的查询分析也一定能得到一致的结果。因此, 可以称这个数据转换满足同态映射或同构映射, 称原数据 A 和现数据 B 在这个数据转换的映射下对于各自的查询分析系统是代数同态的, 甚至是代数同构的。

这种数据转换所带来的代数同构正是 GIS 数据转换所追求的理想目标。但实际上这个理想是很难实现的。由于各地理信息系统对于数据的组织结构、模型表达、覆盖范围等方面的差异, 造成了 A 数据集的信息在转到 B 数据集中必然有所变化, 因此原数据 A 在原系统下进行查询分析的结果, 不一定能在现系统中对现数据 B 经过查询分析得到同样的结果。也就是说, 在 GIS 两个系统、两种格式的数据转换完成以后, 数据集 A 和 B 一般不是代数同构的, 两个系统中所应该对应的运算可能得不到相同的结论, 这时这个数据转换的可行性就值得怀疑。

由于代数同构是针对数据集 A, B 以及相应的运算而言, 而运算是人们根据需要各自定义在各数据集中的, A, B 各自数据集中定义的运算是相当多的, 对某一对运算构成代数同构并不表示对所有运算均同构。但是运算种类毕竟是极有限的, 同构中对于运算的要求又是相当宽松、广义的, 也是绝大多数函数运算能够满足的。同时, 它们也是一些特定运算是否符合同构严密的判定条件。除上述代数同构外, 还有一种图同构也十分重要。所谓图同构, 也即数据集要素一一对应且关系相同。一般的 GIS 数据转换都是在保证一部分信息不损失的前提下完成的。对于两个 GIS 系统的一部分数据, 对应功能的查询分析结果可以是一致的。例如, A 系统为 GPS 接收系统, B 系统为 GIS 中车辆导航系统, A, B 系统中关于车辆的数据集的处理系统是完成各自的任务。GPS 接收机采集记录的数据, 内部格式信息丰富, 有球面三维坐标 (B, L, H) 、接收时刻 T 、速度 v 及星历数据等, 当它作为某一个车载监控系统的数据采集端的一部分以后, 就必然要进行数据格式转换, 此时, GIS 系统需要从 GPS 系统得到的只有平面坐标位置 x, y 和时刻 t 。在两端系统中功能对应的只有两点间的距离, 某点的向量速度 V , 也即求距和求速均为定义在 A 和 B 系统中的相应运算 \oplus_1, \oplus_2 和 \otimes_1 及 \otimes_2 。在数据格式转换前后, 两点的距离、单点速度在两个系统中相同。此时, 两个集合表示为:

$$A = \{B, L, T\}, B = \{x, y, t\}$$

A 与 B 的映射关系为:

①投影变换 $f(B, L) = (x, y)$;

②时间系统变换 $\phi(T) = t$ 。

集合 A 的代数运算:

$$D_a = \text{Distance}(B_1, L_1, B_2, L_2)$$

$$V_a = \text{Velocity}(B_1, L_1, B_2, L_2, T_1, T_2)$$

集合 B 的代数运算:

$$D_b = \text{Distance}(x_1, y_1, x_2, y_2)$$

$$V_b = \text{Velocity}(x_1, y_1, x_2, y_2, t_1, t_2)$$

证明: 因为投影变换 f 和时间系统变换 ϕ 都是一一对应, 故 A, B 间的组合变换 (f, ϕ) 也为一一对应。

对于 A , 存在 $a_1 = (B_1, L_1, T_1)$ 和 $a_2 = (B_2, L_2, T_2)$; 在 B 中, 存在 $b_1 = (x_1, y_1, t_1)$ 和 $b_2 = (x_2, y_2, t_2)$, 使得在 f 和 ϕ 下: $(B_1, L_1, T_1) \rightarrow (x_1, y_1, t_1)$, $(B_2, L_2, T_2) \rightarrow (x_2, y_2, t_2)$, 一定存在 $D_a(a_1, a_2) \rightarrow D_b(b_1, b_2)$, $V_a(a_1, a_2) \rightarrow V_b(b_1, b_2)$

故原数据 A 和现数据 B , 对于各自的代数运算 (D_a, V_a) 、 (D_b, V_b) 来说, 存在一个满足同构的映射 (f, ϕ) , 因此 GPS 数据 A 和导航监控 GIS 数据 B 是代数同构的, 表示为 $A \cong B$ 。

上式表明, 在 GPS 的 A 系统中, 如果有一个性质是采用距离运算 DA 和速度运算 VA 得到的, 那么在 GIS 的 B 系统中, 也可通过该系统下的距离运算 DB 和速度运算 VB 得到类似的性质, 也即仅就各自形式的距离和速度运算而言, A 与 B 仅形式不一, 实质相同。

例如, 1:5 万比例尺 DEM 的建立有两种构网方案: 一是高斯平面上 25m 间距构网, 另一种是经纬差 1" 间距方案。显然, 两种方案中数据集不是一一对应的, 因而不同构。可以确定, 两种方案尽管数据可转换, 但本质是不相容的, 二者必取其一。具体分析可知, 前者大区域具有大的断裂, 不连续, 各子系统格网不统一, 是不可取的。

前已述及, 在一个系统的运行过程中, 数据实际要经历多次的各种格式转换, 系统开发人员的各种转换处理和相应编程过程实际上是保证了这种转换的代数同构性。无论矢量之间的转换还是矢量与栅格的转换, 甚至图形库与关系数据库等存储容器的变化, 评价的关键是数据集的功能必须效果对应, 即用户的应用期望得到满足。保障语义精度, 即是在数据集 A 中通过某个运算得到的一个性质, 那么在同构的数据集 B 中也有一个类似的性质。至于现数据的进一步分析利用, 则是另外衍生的东西。例如, 专业系统利用转换

的基础地理信息数据, 建模分析达到的结论是原系统得不到的, 这已超出了数据转换讨论的范围。这一方法是具体的, 内容广泛。另一种即可采用形式化的理论上的代数同构、同态的判别方法, 这一方法目前可认为是必要条件。

对于数据精度问题, 从理论上讲, 在丧失精度的同时, 该精度有限集上也丧失了一一对应特性; 相反, 若在某精度上有效, 也即说明有一个该精度上的有限集 A , 与之同构的相应有限集 B , 由于元素的一一对应性, 也是相应于该精度的, 相对应某运算的, 这样在 B 上能得到与 A 上相类似的性质将毫无疑问。必需明确数据集本身精度以及定义运算所保证的精度均是保证一一对应性质和同构所必需的。因此, 数据精度这一广泛的实际工程问题, 理论上解决是十分简单而又明确的, 也进一步印证了本文对所讨论问题的简洁性和有效性。

4 结 论

现代数学的趋势就是公理化和结构化。数据集同构的进一步研究涉及到群、同胚以及拓扑空间。群是针对带有运算的集合的问题研究。地理信息系统中的拓扑是广泛使用的概念, 它是代数拓扑的具体发展。数据转换涉及的拓扑和两个数据集间的同胚与连续映射等是更复杂的问题。

数据转化的标准和其他 GIS 技术标准一样, 虽然 ISO、CEN、OGC 等组织有专门的机构进行探讨, 也推动了 GIS 本身的发展和地理信息产业的发展, 但是实践总是先行于标准, 数据转换并不是简单的具体技术开发。一方面涉及很多的利益问题, 另一方面, 随着分布式地理信息系统的发展, 应用数据转换技术也越来越普遍, 对可靠性、精确性和完整性的要求也越来越高。

从单机到网络, 从经典的客户/服务器计算模型转变到以构件开发为基础的分布式计算模型等, 这些应用环境的多样化, 促使对 GIS 数据转换的需求更趋强烈, 涉及的技术将更复杂化。虽然, 在 GIS 的数据模型和数据标准化方面还存在着严重的不足, 但是在市场、厂商和用户的多重推动下, 对它的研究兴趣还是有增无减, 尤其在网络环境下有着重要的研究意义。例如, 数据转换对于制图领域和建模分析领域有着完全不同的内涵, 但有时就是需要在网络下动态共享。基础地理信息与各种应用模型(如环境模型、降雨模型)的结合, GIS 与 GPS、RS 等的集成, 采用并行技术处理各类地理空间信息等研究热点, 均涉及到数

据转换。因此,对新环境下的数据转换等问题进行理论的整理和解释,对促进 GIS 技术的深入发展和进步有重要意义。

参 考 文 献

- 1 李国建. 分布式地理信息系统. 中国地理信息系统协会年会第四届年会, 北京, 1998
- 2 吴宝成. 中国积极支持全球地图开发. 中国测绘报, 1998-02-13
- 3 陆启韶. 现代数学基础. 北京: 北京航空航天大学出版社, 1997
- 4 张禾瑞. 近世代数基础. 北京: 高等教育出版社, 1985
- 5 陈清明, 徐建刚. GIS 数据集中的坐标变换模型应用. 测绘通报, 1999(9): 16~18

- 6 史文中. GIS 空间数据库系统误差控制的黑箱方法. 遥感学报, 1997, 3(8)
- 7 李德仁, 彭美云, 张菊清. 估计 GIS 中面要素定位误差的精度指标. 武汉测绘科技大学学报, 1996, 21(2): 134~138
- 8 朱光, 张保钢. 关于 GIS 属性数据逻辑运算的精度研究. 测绘学报, 1997, 26(1)
- 9 陈常松. 地理信息分类体系在 GIS 语义数据模型设计中的作用. 测绘通报, 1998(8)
- 10 钟业勋. 地图图像系统为布尔代数系的证明. 武汉测绘科技大学学报, 1999, 24(1): 53~55

李国建, 男, 30 岁, 博士生. 现从事分布式地理信息系统、车辆自主导航系统、计算机网络安全等研究。

E-mail: lgj@home.com

Data Transformation and Algebra Isomorphism in GIS

LI Guojian¹ HU Peng²

(1 China Siwei Surveying & Mapping Tech. Corp., 1 Baishengcun, Zizhuyuan, Beijing, China, 100044)

(2 School of Land Science, WTUSM, 129 Luoyu Road, Wuhan, China, 430079)

Abstract This paper presents a rigorous method to check the data transformation of GIS, and first proposes the algebra isomorphism or homomorphism to protect the availability of data for sharing in the environment of net work.

The data transformation in GIS is a high frequency and trouble work, especially in distributed GIS. It roots in the four kinds of requirements to collecting, inheriting, sharing, and integrating multisource data. The GIS data are managed by normal DBMS and graph DB, and accessed by the commands, interactive tools, API, CGI, SQL, and so on. The space information saved in all kinds of structures is different each other in their semantic methods. When the format of data is changing, we should observe the transfer between two kinds of semantic information in the data, which is the center of the transform. Although many researchers have done many works, the question still exist, even more exacerbating. First, it is difficult to evaluate the change between both semantics. The static transformation is from files to files, and the dynamic one needs to access the source file formats directly. On the other hand, there isn't any standard for the quality of the transformation. Now the authors uses the concept of algebra isomorphism in the transformation to build a way to evaluate. If the transform between the both GIS data is not an algebra isomorphism, then, the semantics between them must vary, and so this transform is not successful. Using the feature of the algebra isomorphism in the GIS format transform is efficient in GIS. It is also a basic method to resolve the same difficulty in distributed GIS.

Key words format transformation; GIS; algebra isomorphism; algebra homomorphism

LI Guojian, male, 30, Ph. D candidate. Recently research the distributed GIS, navigation system in driving, and security in computer net, etc.

Email: lgj@home.com