

地理信息理论的探讨

M. 莫伦那

很多空间数据处理的专业领域正处于过渡状态。许多学科都在探寻和发展GIS的能力。象其它学科一样,地图学也正在发现由现代信息提供的越来越多的新机会。

信息学的发展已经历了几个阶段。在它的早期阶段,重点放在机算法能力的使用上。第二阶段是处理大数据流;管理操作如数据贮存和提取等也成为独立的研究领域。

现在是第二阶段向第三阶段的过渡时期。试图用机器效仿人的思想。研究人员正在探讨利用数据以及它们的关系进行推理的规则。人工智能这个新学科还有许多问题,而且我们也不知道将会怎么样。但是经过研究已经形成了一些新技术,可以把人的知识转化成机器能掌握的知识。地理信息系统已经广泛使用了信息学的前两阶段成果,能进行空间数据存贮、处理和图形显示等操作。

许多系统都是为特定应用领域而建立的。虽然它们也能放宽或缩小使用范围,但是它们总是存在一些可能还没体会到的限制。而且可能会产生预料不到的结果。当你想把知识从某一领域转到另一领域时,就会强烈地感受到这一点。

我们面对的问题是复杂的自然界。需要多学科的综合知识来解决它。它已经超出了从数据到知识的转化。这些事实迫使我们重新考虑地理信息系统的特征和用途。“知识”这个术语指的是数据和数据之间的关系以及从它们派生其它数据和关系的原则。

从以上的讨论中就会产生一个问题,即能否建立地理信息处理领域的理论。这个问题早在1985—1986年期间,荷兰几所大学的代表已进行了讨论并发表了题目为“为GIS服务的研究工作”的报告。其它国家也提出过类似的问题。

范围的特点。

更进一步,信息系统的输出还显示了一些有趣的特点。地区级输出只能通过地区数据和地区模型得到,而国家级输出可以从很多方面来评价,参看图7所示;例如,从地区数据通过多区域模型可得到国家级输出。在花费—效率原则上,所有组成有关输出的方法的结合对于具体的真实世界的规划问题的深入研究都是有价值的。还应该注意到不是所有的数据都需要在地方尺度上可观察。这样,图7主要包括了达到一预期输出的最优化方法,对已有数据有一些限制条件。

5. 结 论 以前的观念是:信息爆炸和现代计算机软硬件技术为开发使用各种信息系统提供了极大的可能性。然而,信息科学在规划上的实际作用并不十分明显。在这方面,城市和区域规划部门可从多部门多区域的联合机构中学到很多有意义的关于处理大型的多样化数据组织方面的经验。确实,在规划机构内部的通讯组织分散了人们的很多精力,特别是在实现阶段。

进一步,还必须建立统计学、数值分析、计量地理和评价方法的新的知识和功能,以此为基础为GIS应用于规划部门开发有效的工具。将来几年内的一个主要难题不是扩展我们的工具箱,而是保证模型和GIS使用的一致性。

章燕燕译自《IJGIS》1993,第7卷,第1期

在大部分的这些建议中,系统发展、实现和使用与基本理论混合在一起。在许多国际会议如 ISPRS、ICA 和 FIG 中都是这样的。但空间数据处理的国际研讨会是个例外,它把地理信息处理的理论问题作为主要关心对象。但也没给出这领域的研究框架。

荷兰大地测量委员会专门成立了研究 GIS 理论的分委员会。提出了地理信息理论应主要解决的五个中心问题。下面先讨论背景知识。

——GIS 的应用领域

——地理信息获取和处理的主要特征

——地理信息与环境的关系

——地理信息模糊性和精确度的限制要素

一、GIS 的应用领域 地理信息收集和处理的目如下:

1. 空间分析 为了管理环境,应该了解环境的组织过程以及各个过程的作用。各学科如地质、土壤、植被和自然地理都是对环境不同方面进行研究。由于人对环境的重要作用,行为科学如社会学、经济学和社会地理学等也是空间分析的一部分。它们的第一步工作是描述特定时间的空间条件,如区分地质构造、土壤类型、植被类型、失业高或低的地区、农业区、工业区以及城市化程度等。

根据需求和规则对各要素进行区分并列出清单。在大部分例子中,这些清单只给出状态描述。当你想知道这状态将如何发展时,这些描述是不充分的。当前状态只是发展中的一个阶段。只有对整个过程进行研究,才能预测将来新状态能否合乎需要和被接受。

地质结构经常是作为过程产物来定义和理解的。生态系统则是按相互影响因素的平衡状态(暂时)来定义的。其它学科也类似。时间在环境分析中起着重要作用。

2. 空间规划 空间分析将表明人类是否最优使用了地球表面、土地利用的分布和基本结构是否需要改变。这样就引出了空间规划,即从给定的条件和希望的发展出发,作出是否调整影响输出结果过程的决定。当我们试图改变土地利用的优化再分布或重新建立基础结构时,可以完全重新建立空间结构,如城市规划、重新租种土地、土地保护项目等都是其中的例子。

规划以后应设计具体的操作过程。包括调整环境及调整以后的结果。如设计道路、灌溉系统,种植植物和建造房子等。

3. 规划的实施 实现规划的各项工。如建房子、修灌溉和流域系统、种植植被等。

4. 环境监测和管理 如果不需重建环境,至少也应对环境进行监测和管理。

这些活动都使用空间信息而且产生新的数据。在许多情况下,不同的活动可使用相同的数据。但也经常需要为它们收集特定的数据,使这些不同类型的空间信息能满足数据处理方法的比较。

二、地理信息处理的主要特征 地理数据集处理之前称为调查阶段。其主要工作是区分土地特征、评价专题属性和形状、大小、位置的几何描述。调查结果贮存在用户容易获取的数据库中。原始数据经常必须进行预处理才能为用户提取有关信息。例如土地测量得到的待计算的角度和距离、通过“自由点”抽样获得的土壤数据等,必须经过内插才得到全部土壤矢量或派生等值线。

预处理以后的数据仍存贮在数据库中,有待于进一步处理获得有关信息。如从两点坐标计算距离、从高度矢量得到土地坡度、从土壤和坡度数据得到土地适应因素等。这种处理结果可用图形形式或字母、数字形式提供给用户,也可把它们存贮在另一数据库中。

图 1 给出了信息流程示意图。图中表明了调查阶段和处理阶段是受学科对现实世界解释

影响的。土地调查把现实世界翻译成几何术语,把地球表面压缩成用来决定相对位置的特征点。土壤学家根据土壤单元描述景观,社会地理学家则关心人口统计单元。

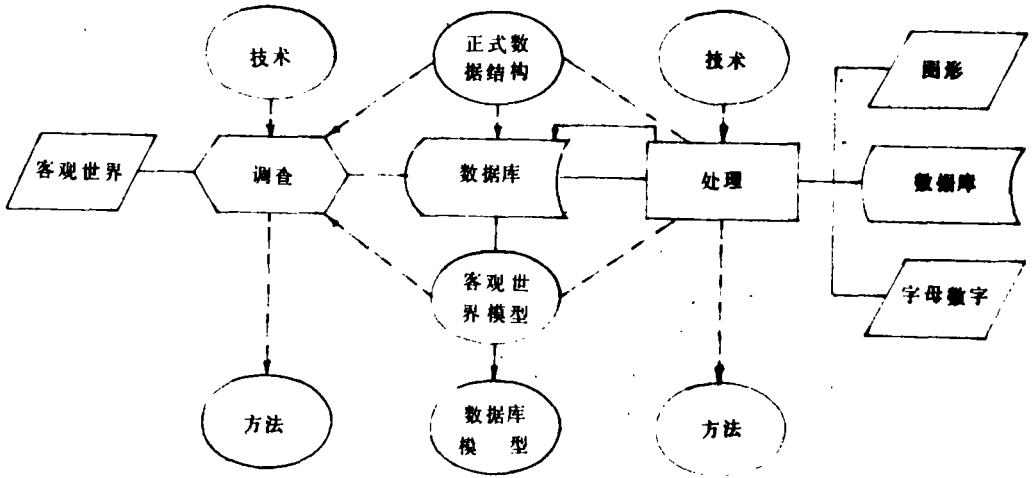


图1 地理信息处理过程及影响因素框图

每一学科定义各自观察到的实体。在调查阶段,观测值被赋予这些实体及属性——土地调查中的角和距离、土壤学中的土壤类型等。调查方法取决于科学家的现实世界模型和有效的测量工具和技术。因此调查方法常是根据模型需要的信息和可能的调查技术确定的。

数据处理阶段也一样,处理方法的确定取决于学科现实世界模型的需求和技术设备。

在图1中,数据库在调查和处理之间,说明调查结果并不直接进入处理阶段,而是先贮存。数据的存贮需要一定的数据结构如数据记录和编码的方式等。同时也要一定的数据库结构,如数据的组织方式可使我们从数据库中得到数据之间的基本关系。

数据结构和数据库结构是与调查阶段成果密切相关的。因此为了使处理时能容易提取数据,必须对数据进行转化和重建的预处理。它不仅为了重建数据而且也能派生其它数据(如从角度和距离得到坐标)。数据和数据库结构的最后选择决定于数据获取、数据处理和数据语义等方面。多边形数据与光栅数据结构不同。

三、正式数据结构(FDS)与语义 FDS这个术语是一系列数据类型及它们之间的基本关系类型。FDS有几个不同抽象程度。较高程度是较低水平的数据和关系类型的组合。数据类型和关系类型贮存在指定的数据库中。特定的数据库模型是有特定的FDS的

对于地理信息系统,主要有两种FDS,即栅格数据结构和矢量数据结构。

1. 栅格结构 当使用栅格结构时,地表是由(二维的)点或小单元的栅格组成(见图2)。栅格单元是在两垂直方向(X和Y轴)按一定的距离(ΔX 和 ΔY)分布它们位置的。因此栅格单元可用(i, j)坐标来表示。

如果地表是由一系列的专题属性描述的,每一栅格单元都将有一定的属性

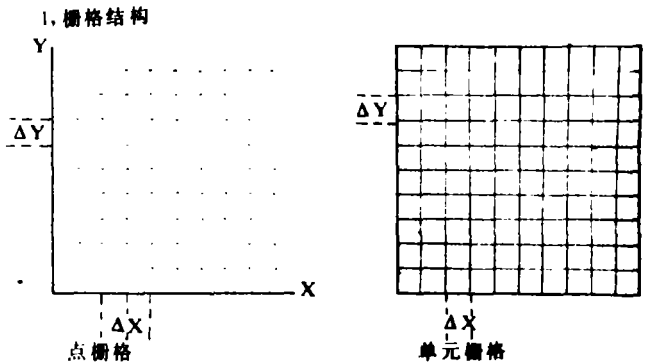


图2 点和单元栅格

值,如评价每一单元的土地利用、土壤类型、植被类型、高度和坡度等。因此在栅格结构中,土地的专题内容直接与每一单元的位置相连。在点栅格中只指一点的位置,而在规则单元栅格中,整个地表由规则单元覆盖。

2. 矢量结构 矢量结构与栅格结构对地表描述的不同点在于使用不同的几何关系以及专题内容与几何关系的联系方式。在矢量结构中,先分出地表主题或特征如道路、河流、森林区域、城镇等。用普通的术语来说就是区分点、线、面特征。然后通过边的链(线部分)或点进行几何描述。区域特征由一闭合链表示。线状特征经常由一开链描述。

然后,地表的专题内容被处理成地表要素的专题内容。因此要素具有专题属性及属性值。专题属性与几何之间的联系是直接的(见图4)。

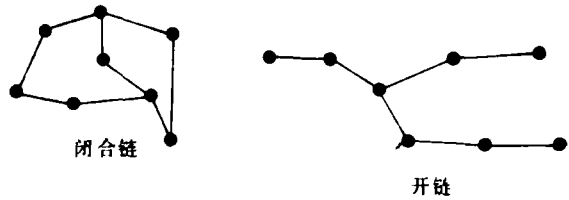


图3 开链和闭合链

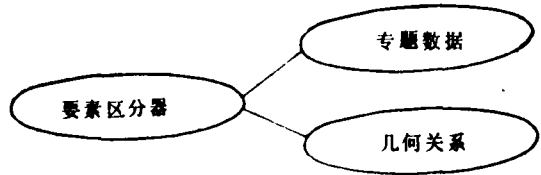


图4 矢量结构数据库中数据类型的联系

栅格和矢量结构都是FDS的较高抽象程度的数据结构。从以上讨论可以看出,FDS和数据的语义内容有很强的相互依赖性,如描述地表特征的数据和基本关系类型的方式。在区域调查中,第一件事是要建立数据与地表之间的联系。当然在数据处理时还可重新建立它们之间的关系。

语义和FDS之间的关系类似于语言学中的语义和语法之间的关系。因此FDS可以看成建立句子的规则。利用这些规则可以表示建议,也可产生新的建议。要想知道在FDS的结构中能形成哪一类型的建议,必须知道放入FDS中的数据及基本关系。

四、查询类型与正式数据结构(FDS) 查询类型是由FDS决定的,而数据库模型和有关查询语言决定了怎样去查询。矢量数据比栅格数据的潜在信息丰富,因此矢量数据也能产生较多的查询。

栅格数据是以专题属性值和位置为关键码进行查询的。如具备植被和高程属性的可查询:

- 森林区域(属性“植被”为关键码)
- 窗口中的植被类型(关键码是“位置”)
- 高程大于500米的森林区域(关键码是“植被”和“高程”)
- 窗口中高程大于500米的森林区域(关键码是“植被”、“高程”和“位置”)

在矢量结构数据库中,根据特征标识符、专题属性和几何数据等关键码进行查询。例如:

- 给出所有房子(“专题属性”为关键码)
- 有多少人住在阿姆斯特丹(“标识符阿姆斯特丹”)
- 在特定区域内的公司(“专题属性”和“位置”)
- 所有大于100公顷的森林(“专题属性”和“几何特征”)

这些例子说明两种数据的查询是不同的。从原理上来说,它们都可以进行更复杂的查询,如查询莱茵河流经哪些国家等。这查询需涉及到地表特征的拓扑关系,需要更复杂的FDS去实现。这说明FDS与查询之间的密切关系。事实上每个FDS都有一套查询即查询间。查询空间是由数据库来实现的。它受存贮在数据库中的数据及基本关系限制。

在设计数据库时,必须把用户所需的查询包括在系统的查询范围中,并使常用的查询能

得到最便利的使用。按查询类型把用户分组,然后每一组给定相应的FDS。而GIS又按照FDS服务的用户进行分类。

五、模糊数据 语义并不是总可以清晰定义的,这种事实称为“模糊”。位置坐标的精确性是受测量过程、方法和工具影响的。地理信息系统经常处理这种在几何上不能确定的数据。栅格数据主要取决于位置变量,在地理规划中,对城市和农村的划分是明显的,它涉及到区域的经济、人口统计和结构等。但农村和城市的分界线也只能在一定的空间精度内描述。因为区域划分涉及到许多要素及它们之间的相互作用。这说明这些要素只能在一定的集合程度上被处理。改变集合程度,区域定义可能会失去相关的东西。

用地图语言来说,这要素只能在一定范围内表示。当范围缩小或扩大时,就要对其它很多要素进行定义。

以上讨论表明在GIS中,很多概念只在一定的环境中有用。环境是通过制图规划、制图空间分析的目的和集合程度来确定的。如果想把概念从一个环境转到另一个环境,而意义又不能变模糊是非常困难的。另外许多概念和值也只能在一定时间内被理解。如“农业土壤适应性”、“景观”等概念都随时间而改变意义。

目标的专题类型定义与空间定义的模糊性之间有很强的联系。分类在有明显差别的地区是不困难的,但是当差别不明显时,要从专题内容一致的观点去划分区域就不一样了。

前面已经叙述了地理数据值不确定的影响方面。即概念的有效相关环境和测量仪器、技术的限制。这里要提出的第三方面是使用算法输出的数据精度不仅受输入数据质量的影响,而且也受算法本身的影响。如栅格和矢量数据之间的转换或遥感中使用的分类算法等。

六、地理信息理论的五个基本问题 地理信息处理大多需要人机接口,使人和机器能共同完成信息处理。我们可以描述人在这个过程中的作用,但要分析人怎样完成工作就困难了。机器的作用可以精确描述,机器如何工作也可以通过正规的结构来分析。这就是为什么把地理信息理论的讨论重点放在机器的作用上。因此地理信息理论用在解释机器的有关问题。它的五个中心问题是:

1. GIS能处理哪些知识?许多学科都使用GIS去描述地表现象。每个学科都使用相应的专题和几何特征来定义自己的目标,并同时调查目标之间的相互作用。但是在一些学科中,GIS有效地处理全部知识是不可能的。只有一部分的知识通过转译和编码能被机器处理。

了解机器对知识处理的限制是非常重要的。因为只有这样才能判断机器形成的描述价值。应该制造出从所有有用知识中抽象出机器知识的规则。这意味着要精确描述数据收集测量的过程。另外,应把制定的规则再装在机器输入的知识中。

2. 数据在哪些环境中是有效的?专题类型和目标的定义只有在一定的环境中有效。在这基础上的描述也是正确的。但是把它从一个环境转到另一环境时就要很小心了。这种知识转化的规则也应制定。

3. 如何生成这些知识?这个问题与GIS中的数据结构密切相关。与第一个中心问题也很相关。生成知识的方式与生成什么的解释之间也有很强的相互作用,就象前面“FDS与语义”中叙述的一样。GIS中可以使用的数据结构数量是有限的。但这个限制是可变的。要解决的问题是如何为特定的应用领域选择最优数据结构。

4. 如何解决知识的模糊性?限制GIS处理数据精确度(或模糊性)的原因有几个。当描述数据时,就应认识到这个事实。根据形成的原因,数据的模糊性可以通过几个方式来表达。几个理论系统可以用来测量相应现象。如概率(概率论)、精度和可靠性(数理统计学)、

一个基于知识的地理信息系统: KBGIS-II

T. 史密斯等

1. 简介 地理信息系统 (GIS) 可以看作一个数据库系统, 其中大部分数据有空间索引, 还有一组过程作用其上, 用于回答有关数据库所描述的空间实体的询问。Smith 等 (1987) 认为一个通用的大型 GIS 应该满足五点设计要求, 同时提出并描述了为满足这些要求而应采用的四条通用原则。本文描述的基于知识的 GIS (KBGIS-II) 正是按照这五点要求来设计和实现的。在详细描述 KBGIS-II 之前, 先简要叙述一下这些原则和要求。

设计和实现 GIS 的要求和原则 以往的研究表明, 大多数 GIS 的设计与实现应满足下列通用要求: (a) 能处理由空间索引数据组成的大型多层次异构型数据库; (b) 能在上述数据库中对大范围的空间对象的存在性、位置及属性进行查询; (c) 能处理交互系统的查询; (d) 系统配置灵活, 保证系统能容易地改装成满足各种特殊应用和用户的新系统; (e) 在系统使用过程中, 能用一种方法“学习”知识库中的空间对象。

可能性 (模糊论) 等。对于何时使用何种测量去评价数据的不精确性以及如何在 GIS 中处理它们的问题还有待于进一步研究。

5. 用户如何处理这些知识? 这个问题很大, 但我们可以限制在两个方面。一是与用户查询有关。因为我们应该去分析从数据库中提取数据的选择标准。用户可以按他们的选择标准或查询分组。

另一方面是有关数据处理算法。如数据的转换。因为原始数据常不能用, 必须进行转化。无论特定的转化是有意义的还是与数据的语义无关的, 象这样的算法在原理上取决于 FDS, 而使用的效果取决于算法的结构和使用的数据库结构。

七、地理信息理论的潜在作用 本文虽只提出了五个中心问题, 却引出了很广阔的讨论领域。这里只作了肤浅的解释。这些问题相当于打开潘朵拉之盒的钥匙。当然有些问题已有答案了, 但在这领域中还不多。发展信息理论主要基于以下考虑:

——好的理论能帮助 GIS 的用户了解他们要求的可能性或不可能性。另外理论对系统完成数据处理、或具体地说对选择各自的处理步骤也很有好处。

——许多 GIS 或系统的组成部分在市场上是有用的, 但用户通常不知道选择系统的标准。他们常常几乎任意挑选或根据他们同事选用的系统去挑选。另外也很难对系统能力进行标准评价。

——如果用户不买系统而是自己建立, 那么就必须制定设计标准和设计策略。系统最后必须进行设计标准检验。

——在这篇文章的开头就提到了新一代的地理信息系统不仅能处理数据, 而且要处理知识。因此我们必须清楚地确定怎样使用和应使用什么数据。

很明显, 地理信息理论对建立解决这些问题的方法是很重要的。当然我们不能等理论成熟以后去解决问题。但是理论方面的工作将会开拓我们的视野, 对完成这些工作是很有利的。同时好的理论系统对新一代地理信息专家的培养也是很重要的。

唐小龙, 冀译译自《ITC Journal》, 1989—1