

# 地理信息系统属性数据不确定性的研究

周迪民, 段国云

(湖南科技学院, 湖南 永州 425100)

**摘 要:**属性不确定性是 GIS 数据不确定性研究范畴的一个重要组成部分, 属性数据的不确定性在许多侧重于属性分析的领域中直接影响 GIS 分析决策结果的准确性和可靠性。分析了属性不确定性的来源, 介绍了数据生命周期的概念, 研究了 GIS 中属性不确定的主要处理方法, 包括目标模型与域模型、概率论及数理统计、模糊集合、粗集和云理论等, 并对基于模糊逻辑的不确定性传播模型和基于灵敏度分析的不确定性传播模型进行理论上的探讨。

**关键词:**属性不确定性; 地理信息系统; 粗集; 灵敏度分析

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 1673-629X(2009)12-0174-04

## Study on Attribute Data Uncertainty in GIS

ZHOU Di-min, DUAN Guo-yun

(Hunan University of Science and Engineering, Yongzhou 425100, China)

**Abstract:** Attribute uncertainty is an important component of GIS data uncertainty. It can directly affect the quality of GIS-based decision making. In this paper, the origin of the attribute uncertainty is firstly analyzed. The concept on lifecycle of data is introduced. A number of theories and methods to deal with attribute uncertainty are discussed. They may include, for example, object model and field model, probability theory, fuzzy sets, rough sets, cloud theory and so on. The propagation of the attribute uncertainty is discussed.

**Key words:** attribute uncertainty; GIS; rough sets; sensitivity analysis

## 0 引 言

GIS 作为一个空间信息决策支持系统, 其应用日益广泛, 重要性得到了人们的重视, 世界各国都投入了大量的物力和财力对有关 GIS 的理论和应用进行研究。人们发现, GIS 采样获取的数据有时只是对现实世界的一种近似描述, 获取大量空间数据的真值并不容易。同时, 获取的信息在被导入计算机系统并用于空间分析决策的过程中, 又被部分舍弃或删除(如目标模型的地学点、线、面抽象, 制图综合)。这样, 在 GIS 应用过程中, 空间数据质量中的不确定性问题越来越为人重视。

传统的数据处理方法认为空间分布可以用一组离散的点、线和面来表达, GIS 不确定性研究的重点, 是使用数值型的统计方法分析位置的不确定性, 而对属性数据不确定性的研究相对较少。实际上, GIS 包含几何和非几何信息, 它要解决的问题更广阔, 更复杂。

例如 GIS 要解决诸如我国公路铁路的总长度这一几何问题。其精确度不仅仅取决于量测线路长度的几何误差, 更要重视属性(语义)不确定性的误差, 如将采样数据上的水渠当作道路或将道路误辨为水渠。后者产生的影响要大得多<sup>[1]</sup>。因此, 有必要研究属性不确定性及其和位置不确定性的一致框架, 延拓误差为更广义的不确定性概念。由于正确使用不确定性在 GIS 决策支持中具有避免利用错误信息导致决策失误以及度量信息支持度的作用, 故在许多侧重于属性分析的领域中, 属性数据的不确定问题将直接影响 GIS 分析决策结果的准确性和可靠性<sup>[2]</sup>。

## 1 属性不确定性来源

不确定性(uncertainty)是指客观世界或实体本身就具有的变异, 表现为不精确性、随机性和模糊性。GIS 数据的属性不确定性是在采集、描述和分析真实世界中的客观实体的过程中, 实体属性的量测、分析值围绕其属性真值, 随机在时间和空间内的不确定性变化域, 是属性误差的空间延伸<sup>[3]</sup>。

由于人对客观实体与现象认识的局限性和表达的模糊性, 使原始数据本身存在不确定性, 以此数据进行

收稿日期: 2009-04-24; 修回日期: 2009-07-06

基金项目: 湖南省自然科学基金资助项目(08JJ6043); 湖南科技学院青年项目(08XKYTB011)

作者简介: 周迪民(1974-), 男, 湖南永州人, 硕士, 实验师, 研究方向为空间数据库与地理信息系统的应用。

GIS的分析与处理,必然使分析结果也具有不确定性。GIS中的数据的不确定性的主要来源有以下5方面。

(1) 测量数据,即按常规的大地测量、工程测量、GPS测量、遥感测量等方法获得数据,其本身含有误差。

(2) 地图数字化过程中产生误差。

(3) 各种来源数据的误差在分析处理中会以某种方式传播,不仅有算术关系下的误差传播,还有逻辑关系下的误差传播和不精确推理下的误差传播。

(4) 计算机化的空间数据的管理和处理必须对极其复杂的现实世界进行必要的取舍和近似。

(5) GIS空间数据的处理过程中会产生新的误差,如矢量数据转换为栅格数据会产生属性误差和拓扑匹配误差,叠加分析时往往会产生拓扑匹配、位置和属性方面的误差。

Michael Goodchild 提出了数据生命周期的概念(数据从原始观测值到最终文件存档的所有数据处理过程)。在数据的整个生命周期过程中,有多个不同用户使用,当数据从一个用户到另一个用户,其精度很容易被改变,精度是数据的动态的属性。因此,只有在用户之间有效传递元数据、数据模型变换后有效更新或者重新进行精度评价,才能最终保证足够的精度信息和数据的可用性。不确定性研究正是解决这一问题、实现数据生命周期各阶段数据可用性的基础,对于数据的使用与共享具有重要意义。

## 2 属性不确定性的研究方法

不确定性问题是非线性复杂问题。不确定性的复杂性来自于数据之间的相关性,不确定性是属于系统科学和非线性科学范畴,并与其中的复杂性科学有关。因此,用非线性科学的研究方法来研究不确定性问题是比较可行的方法。

GIS不确定性问题的研究处于初步阶段,目前已有的研究集中在数据本身的不确定上,很多方法本身也正处于探索试验阶段,还不是很成熟。除了目标模型与域模型等经典数据处理模型、概率论、数理统计作为研究该问题的理论基础外,还需要证据理论、模糊集合、空间统计学、云理论、粗集理论等非线性科学理论的支持。

### 2.1 基于目标模型和域模型的属性不确定性研究

目标模型(object model)是一种基于目标的标准空间数据模型,认为空间分布可以用一组离散的点、线和面等基本几何实体来表达,并使用点、线和面状目标的位置、属性或拓扑关系特性描述空间实体。在GIS中表达地理信息时,对应于矢量数据<sup>[3]</sup>。目标模型较易

表达明确定义的空间实体目标及其位置不确定性,适用于类似基于遥感图像的明显地物(特征)提取的数据获取过程。

域模型(field model)通过对每个像素赋予属性来描述空间实体,认为空间数据可以用定义在连续空间上的若干单值函数来表示(如大气污染、人口分布)。在GIS中表达地理信息时,域模型对应于栅格数据。域模型中,域的空间分布可以描述许多地理现象的空间差异,对不确定性的研究较为系统,较适于描述异质数据不确定性的空间分布以及渐变区域的不确定性。例如环境污染、植被分布等地理现象通常作为域来研究。在表达GIS不确定性的空间分布时,目标模型和域模型常常是互补的。

### 2.2 基于概率论及数理统计的属性不确定性研究

概率论和数理统计用于处理由于随机误差产生的不确定性。在概率论中,不确定性被描述成给定某些观测值条件下某一假设为真的条件概率。该假设的条件概率表示了一个概率从0到1区间的定量描述。分析不确定性时,概率密度函数较为常用,并辅以计算机模拟该误差分布。史文中<sup>[3]</sup>提出了能够描述属性不确定性空间分布的、基于最大似然分类的概率矢量法:一个像素的概率矢量的零元素被剔除并按降序排列,概率矢量中的第一个元素表示其所属类别,其他元素描述分类的不确定性,兼有描述绝对不确定性、相对不确定性、像素混合度和证据完整性4个参数;提出了基于概率论方法和具有概率解释的确定因子模型,研究并实现了结合位置和属性不确定性的“S-带”模型。

证据理论(evidence theory)又称Dempster-Shafer理论,是概率论的一个扩展,是由可信度函数(度量已有的证据对假设支持的最低程度)和可能函数(衡量根据已有的证据不能否定假设的最高程度)所确定的一个区间。当证据未支持部分为空时,证据理论等同于传统概率论。利用证据理论的结合规则或两两比较法,根据多个带有不确定性的属性进行决策分析。空间统计学(spatial statistic)利用有序模型描述无序事件的不确定性理论,根据不确定性和有限信息,分析、评价和预测空间属性数据,主要运用空间自协方差结构、变异函数或与其相关的自协变量或局部变量值的相似程度来描述空间属性的不确定性。它的随机变量可以根据某概率分布取不同的值,具有平稳性,即任意两点只要距离相等方向相同,它们差值平方的期望值就相等<sup>[2]</sup>。空间统计学能改善GIS对随机过程的处理,估计模拟决策分析的不确定性范围,分析空间模型的误差传播规律,综合空间数据,分析空间过程,预计前景,并为分析连续域的空间相关性提供理论依据和

量化工具等。

### 2.3 基于模糊集合、粗集理论的属性不确定性研究

模糊集合(fuzzy sets)用隶属函数确定的隶属度描述不精确的属性数据,重在处理不精确的概率。模糊集合在 GIS 中把类型、空间实体分别视为模糊集合、集合元素,空间实体对备选类制论域的连续隶属度区间为 $[0,1]$ 。每个空间实体与一组元素的隶属度有关,元素隶属度用于表示实体属于某类型的程度,它越接近于 1,实体就越属于该类型。因反映空间非匀质分布的地理属性不确定性的概率是可变的,类别变量的不确定性主要源自定性数据所固有的主观臆断性、易混淆性、混沌性和模糊性,故没有明确定义的界线的模糊集合论,较传统集合论更适于研究非匀质分布和模糊类别,客观地表述地理现象之间的属性不确定性。在应用模糊集合研究属性不确定性的过程中,Burrough and Frank<sup>[4]</sup>讨论了不确定性数据的模糊布尔逻辑模型;王新洲等<sup>[5]</sup>提出了融模糊综合评判和模糊聚类分析为一体的模糊综合法。

粗集(rough sets)由上近似集和下近似集组成,适于处理不精确、不确定和不完全的数据。粗集从集合论的观点出发,在给定论域中以知识足够与否作为实体分类的标准,并给出划分类型的精度。上近似集中的实体具有足够必要的信息和知识,确定属于该类别;论域全集以内且下近似集以外的实体没有必要的信息和知识,确定不属于该类别;上近似集和下近似集的差集为类别的不确定边界,其中的实体没有足够必要的信息和知识,无法确切地判断是否属于该类别。若两个实体有完全相同的信息,则它们为等价关系,不可区分。粗集可用来描述属性 ROSE 不确定模型,分辨不精确的空间影像和面向目标的软件评估<sup>[2]</sup>等,王树良等把粗集理论进一步延伸完善,提出了地学粗空间理论<sup>[6]</sup>,地学粗空间不是预先给定现实世界中实体的某些特征或属性的概率分布或隶属函数等数量描述,而是利用粗集的一对上下近似集直接刻画给定实体,利用等价类及其等价关系,搜寻现实世界的内在规律。当粗实体的上下近似集相等时,就转变为确定实体。总的说来,粗集理论在研究 GIS 属性不确定性方面是很有价值的,只不过现在研究深度和力度还远远不够。

### 2.4 基于云理论的属性不确定性研究

云理论(cloud theory)是一个分析不确定信息的新理论,由云模型、不确定性推理和云变换三部分构成<sup>[6]</sup>。云在空间由系列云滴组成,具有期望值、熵和超熵三个数字特征。期望值是概念在论域中的中心值,完全隶属于该定性概念;熵是定性概念模糊度的度量,其值越大,概念所接受的数值范围越大,概念越模糊;

超熵反映云滴的离散程度,其值越大,隶属的随机离散度越大。云理论把定性分析和定量计算结合起来,构成定性和定量相互间的映射,用于处理空间关联规则的挖掘、空间数据库的不确定性查询<sup>[6]</sup>等 GIS 中容模糊性和随机性为一体的属性不确定性问题。

除了上述理论和方法,研究 GIS 属性不确定性的还有基于复制、交叉和变异的遗传算法,基于灰色分析的灰色系统,基于信息无序互动的混沌理论,基于倍比概念的未确知数学,等等。当然,这些理论和方法不是孤立的,实际分析某类属性不确定性时,常常要综合予以应用。随机的属性不确定性可以利用概率论和空间统计学研究,不能精确描述的属性不确定性可以考虑模糊集、粗集和云理论。

## 3 属性不确定性的传播分析

原始数据存在着不同程度的误差,经过各种处理、转换后生成的产品也保存着原有的误差,这就是不确定性的传播<sup>[7]</sup>。在 GIS 的应用过程中,人们常利用空间数据库中的基础属性数据派生部分新属性数据。例如,根据某一区域的土壤类型、坡度以及湿度属性数据产生一幅关于在这区域种植某种农作物的适宜性地图<sup>[8]</sup>。由于源数据不可避免地具有不确定性,再加上操作误差,这种 GIS 通过对多个空间属性层进行空间操作得到新属性域的方法导致不确定性的传播。在实践中常表现为逻辑模型和算术模型。传统的布尔逻辑模型,只能处理确定性的属性数据分析,刘文宝等<sup>[9]</sup>把它扩展为基于模糊逻辑的不确定性传播模型。他们把任意位置的属性值满足属性值条件的程度作为隶属度,对得到的  $n$  层单一属性专题图进行逻辑操作,得到具有  $n$  种属性的新专题层。GIS 属性数据分析中的算术模型,是派生新属性数据的另一类模型,如水土流失量 FAO 计算模型<sup>[8]</sup>。

早期 GIS 不确定性传播分析是先假设输入信息中的误差已知,然后讨论输出信息中误差的过程,可是确定理论上与输入信息有关的误差非常困难。为了研究输入输出误差间的函数变化关系,引入了灵敏度分析的概念。GIS 不确定性传播分析的灵敏度分析(sensitivity analysis)通过在地理分析输入中添加模拟理论干扰变量,研究所加输入对输出成果的作用。它主要用于讨论属性不确定性对 GIS 成果的影响规律,分析不能用数学模型表达的属性不确定性,检查和划定 GIS 分类产品的等级。Lodwick 等<sup>[10]</sup>提出了利用灵敏度指标确定输出属性值置信域的两种算法。Bonin 利用灵敏度分析方法研究了属性不确定性在矢量 GIS 中的传播,提出了一个估计矢量 GIS 中属性确定性的

噪声概率模型<sup>[11]</sup>。

#### 4 结束语

属性不确定性是 GIS 数据不确定性研究范畴的一个重要组成部分。文中讨论了属性不确定性的含义、来源,在已有研究成果的基础上总结了研究属性不确定性的几个主要方法,包括基于 GIS 模型的属性不确定性研究方法,基于概率论及数理统计的方法,基于模糊集合与粗集理念的方法,以及基于云理论的方法。同时,对属性不确定性的传播进行了探讨,包括基于模糊逻辑的不确定性传播模型和基于灵敏度分析的不确定性传播模型。由于属性不确定性的复杂性,尽管属性不确定性经过一定时间的研究,随着对属性不确定性研究的不断深入,其内容和方法将会日渐丰富和改善。

#### 参考文献:

- [1] 李德仁. 对空间数据不确定性研究的思考[J]. 测绘科学技术学报, 2006, 23(6): 391-395.
- [2] 史文中. 空间数据误差处理的理论和方法[M]. 北京: 科学出版社, 1998.
- [3] 史文中. 空间数据与空间分析不确定性原理[M]. 北京:

科学出版社, 2005.

- [4] Burrough P A, Frank A U. Geographic Objects with Indeterminate Boundaries [M]. Basingstoke: Taylor and Francis, 1996.
- [5] 王新洲, 王树良. 模糊综合法在土地定级中的应用[J]. 武汉测绘科技大学学报, 1997, 22(1): 42-46.
- [6] 王树良, 李德仁, 史文中, 等. 地学粗空间的理论与应用[J]. 武汉大学学报: 信息科学版, 2002, 27(3): 274-282.
- [7] 牛继强, 徐 丰. 线状要素多尺度表达不确定性的综合分析与评价研究[J]. 测绘科学, 2007, 32(6): 69-71.
- [8] Burrough P A, McDonnell R A. Principles of Geographical Information Systems [M]. New York: Oxford University Press, 1998.
- [9] 刘文宝, 邓 敏, 夏宗国. 矢量 GIS 中属性数据的不确定性分析[J]. 测绘学报, 2000, 29(1): 76-81.
- [10] Lodwick W A, Monson W, Svoboda L. Attribute error and sensitivity analysis of map operations in GIS: suitability analysis[J]. International Journal of Geographical Information Systems, 1990, 4(4): 413-428.
- [11] Bonin O. New advances in error simulation in vector geographical databases [C] // Accuracy 200: Proceedings of the 4th International Symposium on Spatial Accuracy in Natural Resources and Environmental Sciences. Amsterdam: University of Amsterdam, 2000: 59-65.

(上接第 173 页)

用户体验。

该系统中的 JavaScript 调用函数中实现 Ajax 技术的主要函数的介绍如下:

```
<script language="javascript">
    var xmlhttp; //定义一个变量用于存放 XMLHttpRequest 对象

    //该函数用于创建一个 XMLHttpRequest 对象
    function createXMLHttpRequest() { ..... }

    //这是响应隶属列表的 onClick 事件的处理方法
    function updateSelect() { ..... }

    //处理从服务器返回的 XML 文档并更新父类下拉列表
    function processor1() { ..... }

    //这是响应父类列表的 onChange 事件的处理方法
    function updateSelect2() { ..... }

    //处理从服务器返回的 XML 文档并更新子类下拉列表
    function processor2() { ..... }

</script>
```

#### 5 结束语

采用 JSP 与 Javabeen 开发的金保工程门户网站系统,搭建了一个劳动和社会保障综合信息业务平台,实现了五险一金信息的综合管理,提高了办事效率,增加

了政府与群众的沟通渠道,也方便了用户及时获取参保信息。

#### 参考文献:

- [1] 马 强, 宋 玲. 基于 Web 的社保新闻发布系统的设计与实现[J]. 计算机技术与发展, 2007, 17(12): 31-33.
- [2] Avedal K, Ayers D, Briggs T. JSP 编程指南[M]. 黎 文, 等译. 北京: 电子工业出版社, 2001.
- [3] 王连华, 王 华, 鲁大营. 基于 JSP/Servlet 的加油站管理系统的设计与实现[J]. 计算机技术与发展, 2008, 18(6): 236-239.
- [4] 冯 纛. JSP-JavaBean 开发模式研究及在网上答题系统中的应用[J]. 计算机工程与设计, 2006, 27(5): 896-899.
- [5] Loney K, Koch G. Oracle9i: The Complete Reference[M]. [s. l.]: McGraw-Hill Osborne Media, 2002.
- [6] 伍孝金, 郑江波. 基于 ASP 的招生就业网站信息管理系统的设计[J]. 计算机工程与设计, 2006, 27(11): 2066-2069.
- [7] 齐鲲鹏, 顾 宏, 唐 达. JSP 数据库连接技术在构建信息网站中的研究[J]. 控制工程, 2002, 9(5): 17-20.
- [8] Garrett J J. Ajax: A New Approach to Web Applications[EB/OL]. 2005. <http://www.adaptivepath.com/ideas/essays/archives/000385.php>.