

文章编号: 1007-4619 (2002)05-0393-08

GIS 中属性不确定性的处理方法及其发展

史文中¹, 王树良^{1,2}

(1. 香港理工大学土地测量与地理资讯学系, 九龙 香港; 2 武汉大学测绘遥感信息工程国家重点实验室, 武汉 湖北 430079)

摘 要: 属性数据的不确定直接影响决策的准确性和可靠性, 特别是对侧重于属性分析的领域。在研究属性不确定性的基础上, 分析了 GIS 中的主要处理方法及其研究进展, 具体地就基于 GIS 的模型、概率论及数理统计学、模糊集合、云理论、粗集等方法及进展进行讨论。

关键词: 属性不确定性; GIS; 研究进展

中图分类号: P208 **文献标识码:** A

1 引 言

GIS 作为一个空间信息决策支持系统, 其应用日益广泛(图 1)。在现实世界中, 复杂的地理现象并非全都是空间匀质分布的, 且实体多相互混杂, 甚至很少界限分明。纯几何意义上的点、线和面在现实世界并不存在。GIS 采样获取数据有时不易明确地将其分类, 只是对现实世界的一种近似描述。同时, 误差一般指测量值或计算值与真实数据或假定真实数据间的差值, 可是获取大量空间数据的真值并不容易, 甚至有些严格或绝对意义上的真值往往并不存在。由于现实世界的位置和属性信息的定位、储存、处理、分析、管理大部分都和 GIS 有关, 故在 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。但是多年来, 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, 通常假设已经检核了属性数据, 并把属性不确定性和位置不确定性隔离讨论。这造成 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^[1]。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。

实际上, 在以属性数据分析为主的土地评估学、土壤化学、环境科学和农林科学等许多 GIS 应

用领域中, 对属性数据质量的要求甚至高于位置精度。如土地的价格属性评估的整个过程基本上都是在做属性不确定性分析。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和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^[2]。文献[3]讨论了属性不确定性的含义、来源、传播, 以及它和位置不确定性的关系, 但对已有 GIS 数据属性不确定性研究方法、可视化方法缺乏深入分析。

2 属性不确定性

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。它度量被测量对象知识缺乏的程度, 可表现为空间属性数据的误差、不精确性、随机性和模糊性, 且受尺度、分辨率、抽样等因素影响。与仅指属性观测值与其真值间差异的属性误差^[3]相比, 属性不确定性的含义更广。

地学信息的基本单元是一个连接时空位置和系列属性的数组^[4], 边界不确定性是属性与空间不确定性紧密相关的典型表现。研究属性不确定性可概括为探讨不确定性的产生、传播和控制等问题, 基本

收稿日期: 2001-05-22; 修订日期: 2001-12-06

基金项目: 香港理工大学科研基金项目 No. 1. 34. 9709、国家高技术研究发展计划(863 计划)项目 No. 2001AA 135081 和国家自然科学基金项目 No. 40023004 资助。

作者简介: 史文中(1963—), 男, 副教授, 香港理工大学, 目前主要从事 GIS 遥感技术应用研究。发表论文 150 篇, 著作 3 部。

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

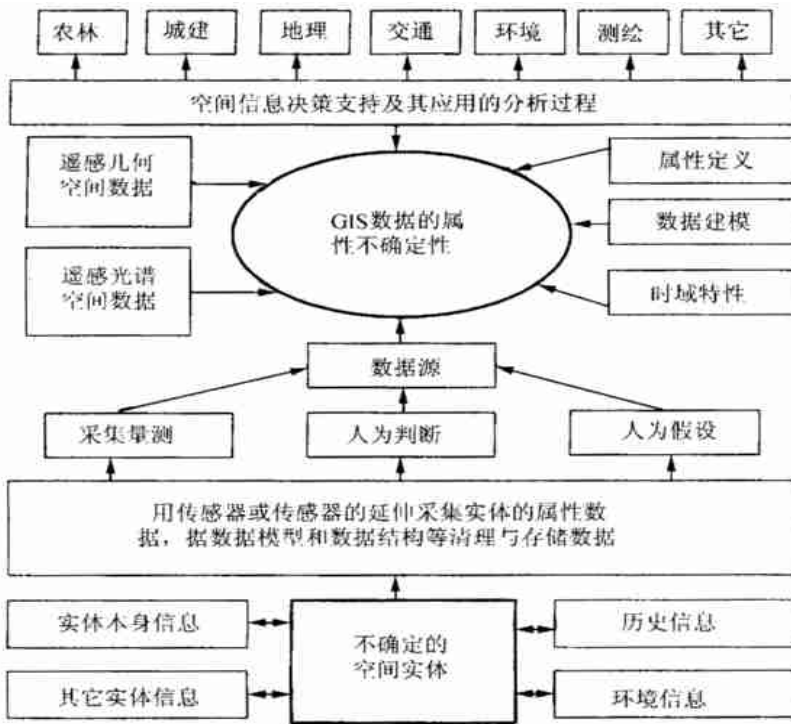


图 1 GIS 属性不确定性的来源及对应用之影响

Fig. 1 Sources of attribute uncertainties in GIS and their influences to the applications

包括不确定的实体定义、边界、关系、分析和查询等。属性不确定性的来源有多种(图1)。

3 主要研究领域及进展

在可用于属性不确定性分析的方法中, GIS 的目标模型和域模型是经典的数据处理模型, 概率论及数理统计学具有统计意义, 粗集、模糊集合和云理论实行软计算(图2)。

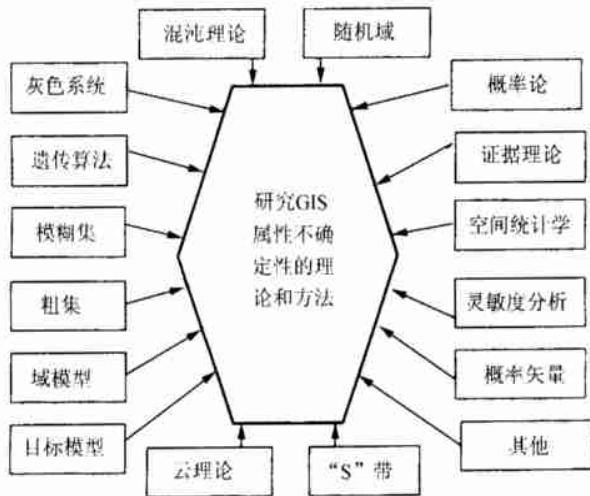


图 2 研究 GIS 属性不确定性的理论和方法

Fig. 2 Theories and methods to study attribute uncertainties in GIS

3.1 基于使用不同 GIS 模型而产生的属性不确定性

3.1.1 基于目标模型的研究

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^[5]。

目标模型是 GIS 中的常用模型, 也较为成熟。虽然在理论上, 明确定义的目标位置和属性可以精确测定和获得, 但是目标模型的适用性是相对的。纯几何意义上的点、线和面在现实世界中并不存在。由于实用性及经济因素的制约, 空间实体定位数据的采样事实上只是一种近似。而且, 用一系列离散的采样点来描述现实世界复杂空间目标的连续属性时, 本身就有近似性和取舍问题, 从而导致不确定性的产生。

相对而言, 人们对于目标误差模型的位置不确定性研究较多。在建立目标误差模型, 用位置和属性数据研究有明确定义的空间实体的不确定性时, 其充要条件是位置和属性的真值(即参考数据)。分别用误差椭圆和 ϵ -误差带描述点和线的不确定性, 使用的度量指标是数值型的均方根差、标准离差、误差带宽等。基于目标的数据库把这些误差指标视为

目标的扩展属性,用地学关系数据结构来存储^[4]。但是,目标模型割裂二者,重点研究位置的不确定性,并认为数据的采集、录入和编辑不含粗差,属性数据只是位置数据的补充。

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,可是不容易表达异质专题数据不确定性的空间分布区域内的渐变特性^[9]。例如,GIS 中的空间目标界址点、道路和宗地,其位置由某个合适的坐标系来描述,相应的误差一般由测量平差的随机模型和误差模型处理。这些空间目标的属性描述它们的某些定性和/或定量特性,可以是界址点使用了的时间,道路的宽度,以及宗地的权属、拓扑关系、与相邻宗地的关系等,如来自遥感影像的土地覆盖分类数据。相应的属性不确定性,目标模型则难以完全胜任。

目标模型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^[7]。然而,自然界的地理属性,并非都是空间匀质分布,而多是界限相互混杂有的区域不易明确地划分为属性类型(如植被),当GIS 利用遥感图像对这样的区域作资源环境调查或专题制图时,目标模型显然不是十分合适。与目标模型相对的是域模型,它在研究不确定性时具有不同的特点。

3.1.2 基于域模型的研究

域模型(field model)通过对每个像素赋予属性来描述空间实体,而不是通过对目标的抽象或描述它们之间的拓扑关系来实现的^[4],又称为场模型^[7]。在GIS 中表达地理信息时,域模型对应于栅格数据。域模型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(如环境污染、农作物分布),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,其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。例如大气污染、人口分布等地理现象通常作为域来研究。遥感图像记录了空间实体域的光谱信息,在利用它调查资源环境或制作专题图时,域模型比目标模型更适合讨论其属性的不确定性。

因函数值与真值一般不吻合,故不确定性存在于域模型。描述域模型的变量有定性类别型(如土壤分类)和定量数值型(如土地价格),它把不确定性的研究归结为所涉及变量的不确定性问题。在度量域模型在多大程度上反映真值时,可分别利用方差(或标准离差)和概率对数值型和类别型变量进行不确定性分析^[9]。类别型变量的概率指某一位置上所观测得到的类别为正确的概率, n 个类别概率域需 n 个地图层来存贮。在建立域误差模型,生成连续域时,一个属性由

随机模拟方法在一个域内建立等概率的空间分布,所产生的一组模型数据即为一个等概率分布的误差“实现”,并遵从数据点位的数值。对于有着明确定义的空间实体,参考文献[7]用场的概念和模型统一了位置不确定性与属性不确定性,把不确定性的研究归结为所涉及变量的不确定性问题,分别利用方差和概率对数值型和类别型变量进行不确定性分析。

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。可是,由于备选类别的个数与存贮图层的个数成正比,类别域需要付出为此存储和查询的代价。理想的方法应该既最大限度地保留类别原有的信息,又能尽量节省数据存贮层。

与GIS 模型有关的属性不确定性研究还有若干方面的问题值得进一步探讨。例如,在基于目标的模型中,多维GIS 模型中所引出的属性不确定性;目标之间的逻辑不一致性对属性不确定性的影响,属性不确定性的抽样检验方法,建立模型不确定性的描述指标体系等等。对于基于域模型的属性不确定性研究,则有属性不确定性的空间分布、域之间属性不确定性的相关性研究、空间分析中域属性不确定性的积累与传递等等。这些,皆值得进一步探讨。

3.2 基于概率论及数理统计的属性不确定性模型

3.2.1 概率论

概率论(probability theory)可用于分析由属性随机误差造成的不确定性^[2]。概率论是传统的数据处理方法,它在获得大量的观测值后,把不确定性表示成给定条件下、某一假设为真的条件概率。如用Bayes 定律讨论GIS 属性数据从0到1的概率不确定性定量描述。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。当使用高斯概率密度函数分析点位误差时,如果误差分布不遵从高斯密度函数,那么更一般的概率密度函数可以通过点位坐标值的反复量测而得到。在得知某一点的误差椭圆参数后,可以根据误差椭圆概率密度函数,利用计算机模拟该误差分布的若干“实现”。这些不同的“实现”作为样本输入到某一空间过程,可用来探求点的位置误差在该空间过程中的传播以及该误差对某一派生数据产品的影响。线的位置误差较为复杂。虽然 ϵ -误差带度量位置误差直观和简易,但仅由带宽参数还不足以获得线目标的不同“实现”。现有的误差模型多为某种形式的概率密度函数解析式,如包含一线段两端点的正态分布函数。为了建立某个空间运算(过程)的误差特性,模拟线

段的不同“实现”，可能更为有效地产生反映该过程的统计量，以据此模拟数据。

根据栅格的结构化土地覆盖分类模型，Canters and Genst^[8]在评价输入数据的不确定性对输出结果的影响时，提到的基于概率的影像隶属矢量认为，一个像素矢量中的所有元素共同影响和决定该像素的类别。张景雄等^[7]的概率场认为类别型变量的概率指某一位置上所观测得到的类别为正确的概率，即假定所考虑的类别个数为 n ，则任意一点可以定义 n 个类别概率场，需 n 个地图层来存贮。此外，还有随机影像的不确定性研究^[9]。可是，相对于概率矢量，Canters and Genst 的影像隶属概率矢量、张景雄等的概率场和 Journel 随机影像研究，都没有剔除零元素和排序，存贮空间和查询时间均较大。

3.2.2 概率矢量的提出

概率矢量(probability vector)基于最大似然分类法能够描述属性不确定性的空间分布^[4]。在一个像素的概率矢量中，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的四个参数。通常用误差矩阵描述遥感分类结果的不确定性，可是它不能表示不确定性的空间分布，因这种属性不确定性产生于分类决策根据的不足。在此基础上，史文中^[27]提出了基于概率论方法和具有概率解释的确定因子模型，研究和实现了结合位置和属性不确定性的“S-带”模型。

概率矢量的最大优点就是能够表示不确定性的空间分布，但所占的存储空间仍可以相对较大，而所提出的参数描述方法，则可以减少不确定描述数据的质量。

3.2.3 证据理论

证据理论(evidence theory)是概率论的一个扩展(又称 Dempster-Shafer 理论)，是由可信度函数和可能函数所确定的一个区间^[10]。其中，可信度函数度量已有的证据对假设支持的最低程度，而可能函数衡量根据已有的证据不能否定假设的最高程度。当证据未支持部分为空时，证据理论等同于传统概率论。证据理论将实体分为确定部分和不确定部分，可以用于属性不确定性决策分析。利用证据理论的结合规则，可以根据多个带有不确定性的属性进行决策分析^[4]。两两比较法也用于属性不确定性的分析^[11]。证据理论发展了更一般性的概率论，却不能解决矛盾证据或微弱假设支持等问题。

3.2.4 空间统计学

空间统计学(spatial statistics)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^[12]。空间统计学是利用有序模型描述无序事件的不确定性理论，其发展是从非空间、空间而到时空域，它的随机变量可以根据某概率分布取不同的值(如污染物集中和人口密度)，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^[4]。空间统计学的变异函数量化空间相关特性并引导空间插值，实验变异函数能从抽样数据估算得到。

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。Cressie^[12]利用地理统计数据(描述连接空间内的统计过程)、栅格数据(表示通过空间中的栅格内的空间过程)和点数据(表达空间点过程)三种空间数据描述现实世界，并提出了一个适用于地理统计数据、栅格数据、点模式和物体的通用模型。应用空间统计学的克吕格方法，由一组已分类的观测点直接估计未观测点位的属于各类别的验后概率，求得类别变量在任一位置上所观测到的各类别的概率，就可以从影像上获取模糊分类信息。

可是，当知道非匀质实体的什么属性可能发生与否，但却不知道也难以构建其概率分布模型时，模糊集理论比空间统计学更利于处理这种不确定性。

在概率论及数理统计(特别是空间统计)的框架下，以下问题值得进一步探讨：(1)空间目标在 GIS 中描述的可靠性；(2)GIS 空间分析中不确定性的积累；和(3)对 GIS 中属性不确定性的统计检验。其中，抽样方法是一个重要的问题。

3.3 基于模糊集合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^[13]。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，空间实体对备选类论域连续隶属度区间为 $[0, 1]$ 。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。具有类型混合、居间或渐变不确定性的实体可用元素隶属度描述，如一块含有土壤和植被的土地，可以由两个元素隶属度表示。

基于传统的集合理论，每一个空间实体都与单

一的属性说明有关,属性之间被表示为清晰的边界。确定的集合具有精确定义的界线,为0、1二值逻辑。给定一个元素,要么完全属于集合,要么完全不属于。显然,这种方法很难处理一个属性不能精确描述的情况。例如,由于遥感数据采集的空间分辨率偏低(细节太粗略)而产生的类别模糊性。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。对于遥感图像的计算机分类处理,模糊类别域的生成可籍所使用的分类器不同而输出不同的中间结果,如统计分类器中有某像素隶属于各备选类别的似然值,神经网络分类器中的类别激活水平值^[7]。

在应用模糊集合研究属性不确定性的过程中,文献[1]讨论了不确定性数据的模糊布尔逻辑模型;文献[14]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,基于绝对均值距离的模糊聚类分析,分别用于评估土地的地价和级别属性不确定性,探讨了GIS空间数据的属性不确定性,借助模糊关系数据模型评价了土地适宜性;根据模糊土地植被的分类评估面积的不确定性^[15];模糊隶属度也用于表达遥感影像中的不确定相邻边界的像素类别^[3]。

当模糊隶属度一旦确定,模糊集合的后续数值计算实际上已经把不确定性抛开,并没有继续向前传送到结果。而且模糊集合主要处理具有模糊性的属性不确定性,对于同时含有模糊性和随机性的不确定性,它只能丢弃随机性,这是不合适的。

在空间分析中,例如GIS的叠置(overlap)分析中,应用模糊数学讨论属性不确定性的传播与积累,是一个值得进一步深入研究的问题。此外,基于模糊数学的属性不确定性的空间分布也值得进一步探讨。

3.4 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大部分构成^[16]。云理论把定性分析和定量计算结合起来,可以用于处理GIS中容随机性和模糊性为一体的属性不确定性。云在空间由系列云滴组成,远观象云,近视无边。

云具有期望值、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。云理论构成定性和定量相互间的映射,处理GIS中容模糊性(定性概念的亦此亦彼性)和随机性(隶属度的随机性)为一体的属性不确定性,解决了作为模糊集理论基石的隶属函数的固有缺陷。云理论已经用于空间关联规则的挖掘、空间数据库的不确定性查询^[14]。

实验结果证明,在处理属性不确定性时,云理论具有一定的潜力。一个值得进一步深入研究的问题是,在云理论中顾及属性不确定性。例如,处理多源属性不确定性的多维非正态云、属性不确定性的分离为聚合的逆的证明等。

3.5 基于粗集的属性不确定性研究

3.5.1 粗集

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^[17]。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,并给出划分类别的精度。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。根据利用统计信息与否,现存的粗集模型及其延伸可以分为代数型和概率型两大类^[18]。粗集的基本单位为等价类,类似于栅格数据的栅格、矢量数据的点或影像的像素。等价类划分越细,粗集描述实体越精确,但数据量和计算时间也越大。粗集不排斥不确定性,力求按照实体的原形来研究实体,为GIS的属性不确定性分析提供了一个新的理论基础。

粗集可用于不确定影像分类,模糊边界划分,分析GIS属性数据库中的属性不确定性、属性表的一致性和属性的可靠性,研究属性可靠性对GIS决策的影响,评定属性的绝对不确定性和相对不确定性,简化属性依赖和属性表,生成最小决策利分类算法等。

现在,粗集已被用于基于属性不确定性的银行粗选址,从数据库中发现不确定属性的知识^[16],集成多源不确定的属性数据,实现定性和定量语言值的粗转化,结合模糊隶属函数的遥感影像粗分类、粗邻域和粗属性精度^[19],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^[20]等。目前,这些成果仅限于实验室阶段,具体应用的适用性尚待证明。而且,由于粗集不是源于地球空间信息学的专门理论,它给出了概念定义的衍生拓扑关系,却没有涉及空间关系的研究,空间计算分析的工具(如多个图层的叠置)也尚未具备。因此,有必要在地球空间信息学领域内,把粗集进一步延伸完善,如王树良的地质粗空间理论。

3.5.2 地质粗空间理论的提出

地质粗空间(geo-tough space)由粗实体、粗关系和粗算子组成。其基本单位是粗元。其中,粗实体通过一对下近似集、上近似集逼近地质空间实体,是同时包容空间的位置不确定性和属性不确定性的近似域。粗关系是粗实体之间的相互作用规律,粗元关系为其基础。地质粗空间不同于粗集,它具有粒度和尺度,粒度的大小描述粗元的精细,如遥感影像的空间分辨率;尺度表达粗实体的比例尺,决定地质粗空间的置信粗度。

地质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。地质粗空间根据GIS已有信息对空间实体的支持程度,将地质空间划分为肯定支持域、肯定不支持域和可能支持域。通过简化条件属性,地质粗空间又将GIS决策分析定义为不同简化层次上的子问题决策,对应不同的数据不完备情况,并给出置信粗度。

利用地质粗空间研究属性不确定性,(1)可以对已有属性不确定性理论或模型给出一种新的数学解释。如当上、下近似集相等时,地质粗空间可以表示目标模型;当下近似集和上近似集合并时,地质粗空间可以表示域模型;当上近似集变化时,地质粗空间可以表示误差带、 ϵ 带和“S”带等模型,置信粗度表示其不确定性的大小。(2)能够对空间世界作全域覆盖式划分,包容其他空间理论。如下近似集、上近似集和负域将地质粗空间的论域划分为“是”、“可能”和“非”三个不同的置信域。当“可能”域为空时,地质粗空间变为确定空间,而当“非”域为空时,它就成为了模糊空间。(3)兼顾了空间实体的确定性和不确定性,具有保真性。地质粗空间承认地质信息的复杂性和多变性,按照现实世界的固有存在形式研究地质规律,可以最大可能地逼近地质实体的真实存在状态。利用粗关系和粗算子计算分析粗

实体得到的计算结果,信息量丰富,而并非确定空间的二值信息或模糊空间的“似是而非”信息。(4)可以较为全面地分析属性观测值的不确定性。在地质粗空间使用下近似集和上近似集描述现实世界的实体时,下近似集中的元素全为真值,上近似集是对真值的观测延伸,其中的元素是观测值对真值的部分偏离。由于观测只是对真值的逼近,难以得到完备的信息,因此当一个空间实体的属性被观测多次后,观测值就围绕其真值形成一个延伸。导致延伸的信息不完备性,可能来自设备、人或数学函数。当粗实体的下近似集和上近似集相等时,可能实体就转变为确定实体。等等。

3.6 灵敏度分析

传统GIS误差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,理论误差被模拟引入,即为灵敏度分析。

GIS的属性不确定性的灵敏度分析(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^[21],在理论上属于不确定性传播分析。属性不确定性的灵敏度分析只考虑人为影响,可视为理论上的误差传播分析。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。在属性不确定性的灵敏度分析中,加入干扰前的数据为参考数据,不同的干扰变量输入得到不同的分析成果输出,含有不同的误差。它需要大量数据,适宜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。

属性不确定性有三个基本参数:不足率(真实存在而属性没有)、过量率(真实没有而属性出现)和紊乱率(属性拥有和真实不符的错误值)。对于GIS属性不确定性的灵敏度分析的理论和应用,Lodwick等^[21]作了较为系统的回顾,定义了属性、位置、地图删除、多边形和面积五个统计灵敏度分析指标,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。为了观察和分析误差对最后成果的作用,必要时通常故意把误差强加于真实数据造成数据误分类和划错等级。Bonin^[22]利用灵敏度分析方法,以道路属性的不确定性为例,研究了属性不确定性在矢量GIS中的传播;后来又提出了一个估计矢量GIS中

属性确定性的噪声概率模型^[23]。灵敏度分析方法应用于栅格数据和地图叠置,发现包括连续和离散变量在内的许多问题并不呈现各向同性,符号语义对结果有各种等级的影响^[24]。

灵敏度分析对属性不确定性的研究,尚需进一步深化。例如,不确定性估计本身的理论误差,衡量参数不确定性的不确定指标,属性灵敏度强弱的不确定性度量等。

3.7 其他

除了上述理论和方法,变换区域模型描述实体类别属性的不确定性时,把不同的地区和像素赋以不同的可靠性,在多边形的中心可靠性较高,而边界附近较低^[1]。文献[25]抽象了一个加权平均后得到的典型稳定的降雨的土壤渗透状态,利用分布式随机域模型把这种状态描述成降雨强度、汇水大小、饱和渗透率和流失方式的函数,研究了降雨对土壤的不确定性作用,文献[26]认为农业数据不必总是依附在地理坐标上,土壤的特性可以从GIS系统的土壤图中衍生出来。农业生态系统中的土地利用的可持续性跟土壤中重金属的富集度有关,可通过土壤中重金属的输入输出流量是否平衡来评价。所用的带有土壤特性的农业数据在结果中可能带来相当数量的不确定性。于是他们提出了一种视输入为输出的随机模型,并使用 Latin Hypercube Sampling Strategy 在模型变量中传播不确定性,来保证该模型成果的质量。此外,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。其中,常用于表达遥感分类结果的不确定工具是误差矩阵,却不能描述不确定性的空间分布。概率矢量方法有描述不确定性的空间分布的潜在可能。目标模型是基于目标的空间数据模型,重在用传统的统计理论讨论位置不确定性,采用域模型讨论属性的不确定性问题优于使用目标模型。目前用于表示空间数据的方法多数是基于传统的集合理论,每一个空间实体都与单一的属性说明有关,属性之间被表示为清晰的边界。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。在属性自变量集 x 和类别因变量集 $\mu(x)$ 之间,模糊数学是一对一关

系,即对一个特定的 x_k ,只存在惟一的隶属度 $\mu(x_k)$,且 $\mu(x_k) \in [0, 1]$;粗集是一对区域关系,即对一个特定的 x_k ,存在隶属区域 $\{\{\min^{-\mu}(x_k)\}, \{\max^{-\mu}(x_k)\}\}$;云理论则是一对多云滴关系,这个云滴在空间随机离散分布,当云滴聚集到一定程度,概念空间就成了一朵没有边界的云。

4 总结

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。本文总结了处理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描述方法,基于概率论及数理统计的方法,基于模糊数学的方法,基于云理论的方法,以及基于粗集理论的方法。这些基于不同理论的方法,在研究属性不确定性方面各具特点,有些是相互补充的。因此,在使用时,应该考虑综合应用。同时,属性不确定性研究还要充分借鉴其他领域已经成熟的传统技术方法(如测绘数据处理对位置误差的估计方法),并与新兴的学科(如空间数据挖掘)紧密联系。

尽管属性不确定性已经经历了一定时间的研究,但是仍有相当多的问题值得进一步深入探讨。例如属性不确定性的空间分布;属性不确定性的积累和传递;属性不确定性的抽样检验;属性不确定性对决策支持的影响程度;以及属性不确定性描述指标的建立等。这些问题的研究,有赖于以上理论及方法、或几种理论及方法的综合使用,及其本身的进一步发展。因此,GIS数据中的属性不确定性,是一个值得进一步深入探讨的领域。

参考文献 (References)

- [1] Burrough P A, FRANK A U. Geographic Objects with Indeterminate Boundaries [M]. Basingstoke: Taylor and Francis 1996.
- [2] SHI W Z. Theory and Methods for Handling Errors in Spatial Data [M]. Beijing: Science Press, 1998. [史文中,空间误差处理的理论和方法[M].北京:科学出版社 1998.]
- [3] SHI W Z, WANG S L. State of the art of research on the attribute uncertainty in GIS data [J]. *Journal of Image and graphics*, 2001, 6(9): 918-924. [史文中,王树良, GIS 数据之属性不确定性研究[J].中国图象图形学报, 2001, 6(9): 918-924.]
- [4] Goodchild M F. Modeling Error in Objects and Fields [A]. In *Accuracy of Spatial Databases* (London: Taylor & Francis), 1989, 107-113.
- [5] Goodchild M F. Attribute accuracy Elements of Spatial Data Quality [C]. New York: Elsevier, 1995: 139-151.
- [6] Burrough P. A. GIS and Geostatistics: Essential partners for spatial analysis Proceedings of the International Symposium on Spatial Data

- Quality' 99[C]. Hong Kong: Department of Land Surveying & Geo-Infomatics, The Hong Kong Polytechnic University, 1999; 10—20.
- [7] Zhang Jingxiong, Du Daosheng. Field-based Models for Positional and Attribute Uncertainty [J]. *Acta Geodaetica et Cartographica Sinica*, 1999, 28(3), 244—249. [张景雄, 杜道生. 位置不确定性与属性不确定性的场模型[J]. 测绘学报, 1999, 28(3) 244—249.]
- [8] Canters F., Genst W D. Handling Uncertainty Propagation through the Buffer Operation in a Raster Environment [C]. In Accuracy 200: Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 2000, 145—152.
- [9] Journel A. G. Modelling uncertainty and spatial dependence: stochastic imaging [J]. *International Journal of Geographical Information System*[J]. 1996, 10(5): 517—522.
- [10] Shafer G. A Mathematical Theory of Evidence [M]. Princeton: Princeton University Press, 1976.
- [11] Chrisman N. C. Exploring geographic information systems [M]. New York: Wiley & Sons, 1997.
- [12] Cressie N. Statistics for Spatial Data [M]. John Wiley and Sons, Inc. 1991.
- [13] Zadeh L. A. Fuzzy sets [J]. *Information and Control*, 1965, 8: 338—353.
- [14] Wang X Z., Wang S L. Fuzzy Comprehensive Method and Its Application in Land Grading [J]. *Journal of Wuhan Technical University of Surveying and Mapping*, 1997, 22(1): 42—46. [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.]
- [15] Canters F., Evaluating the uncertainty of area estimates derived from fuzzy land—cover classification [J]. *Photogrammetric Engineering and Remote Sensing*, 1997, 63: 403—414.
- [16] Di K C. The Theory and Methods of Spatial Data Mining and Knowledge Discovery [D]. Wuhan: Wuhan Technical University of Surveying and Mapping, 1999. [邱凯昌. 空间数据挖掘和知识发现的理论和方法[D]. 武汉: 武汉测绘科技大学, 1999.]
- [17] Pawlak Z. Rough sets [J]. *International Journal of Computer and Information Sciences*, 1982, 11(5): 341—356.
- [18] Yao Y. Y., Wong S. K. M., LIN T. Y., A review of rough set models. *Rough Sets and Data Mining Analysis for Imprecise Data* [C]. Norwell: Kluwer Academic Publishers, 1997; 47—75.
- [19] Ahkqvist O., Keukeelaar J. Oukbir K. Rough classification and accuracy assessment [J]. *International Journal of Geographical Information Science*, 2000, 14(5): 475—496.
- [20] Gunther R. Rough set-based analysis in goal-oriented software measurement Proceedings of METRICS' 96 [C], 1996; 10—19.
- [21] Lodwick W A *et al.*, Attribute error and sensitivity analysis of map operations in GIS: suitability analysis [J]. *International Journal of Geographical Information Systems*, 1990, 4(4): 413—428.
- [22] Borin O. Attribute Uncertainty Propagation in Vector Geographic Information Systems: Sensitivity Analysis. In Proceedings of the Tenth International Conference on Scientific and Statistical Database Management, edited by Kristine KELLY (Capri, Italy: IEEE Computer Society), 1998, 254—259.
- [23] Borin O. New advances in error simulation in vector geographical databases. Accuracy 200: Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences [C]. Amsterdam: University of Amsterdam, 2000; 59—65.
- [24] Fisher P F. Modeling Soil map-unit Inclusions by Montecarlo Simulation [J]. *International Journal of Geographical Information Systems*, 1991, 5(2), 193—208.
- [25] Karszenberg D. *et al.* Uncertainty of Field Scale Steady State Infiltration Derived from Rainfall Plot Experiments [A]. In Accuracy 200: Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences [C], edited by G. B. M. HEUVELINK, M. J. P. M. LEMMENS (Amsterdam: The Netherlands: University of Amsterdam), 2000; 343—354.
- [26] Keller, A. *et al.* Uncertainty Assessment in Modelling Cadmium and Zinc Accumulation in Agricultural Soils. In Accuracy 200: Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 2000; 347—354.
- [27] Shi W. Z. Modeling Positional and Thematic Uncertainties in Integration of Remote Sensing and Geographic Information Systems [M] (Erschede: ITC Publication), 1994.

Further Development of Theories and Methods on Attribute Uncertainty in GIS

SHI Wen-zhong¹, WANG Shu-liang^{1, 2}

(1. Department of Land Surveying & Geo-Infomatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong;

2. National Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, 430079)

Abstract: Attribute uncertainty can directly affect the quality of GIS—based decision making. In this paper, the theories and methods to study the attribute uncertainty and their developments are discussed. And a number of theories and methods have been developed to deal with attribute uncertainties. These may include, for example, GIS—based models; probability theory and statistics—based statistical theories; fuzzy sets, rough sets and cloud theory based soft computing. These theories and methods for describing attribute uncertainties in GIS data are reviewed and analyzed. Furthermore, the future developments of the research in these areas are also briefly summarized.

Key words: Attribute uncertainty, GIS, research development