第 36 卷 第 6 期 计 算 机 工 程 2010 年 3 月
Vol.36 No.6 Computer Engineering March 2010
·开发研究与设计技术· 文章编号：1000—3428(2010)06—0250—03 文献标识码：A 中图分类号：TP391

GIS 属性数据不确定性及其传播研究

周迪民[1]，林依勤[2]
(1. 湖南科技学院现代教育技术中心，永州 425100；2. 湖南科技学院数学系，永州 425100)

摘 要：属性数据的不确定性直接影响地理信息系统(GIS)分析决策结果的准确性和可靠性。将 GIS 属性数据的整个生命周期分成数据采集与准备、数字化信息提取、信息综合和属性数据表达等阶段，研究属性数据不确定性在各阶段的传播，构造其传播模型，提出属性数据不确定性的合成算法与传播算法，实验结果验证了该模型的可用性与有效性。

关键词：属性数据；不确定性；地理信息系统；传播模型；灵敏度分析

Study on Attribute Data Uncertainty in GIS and Its Propagation

ZHOU Di-min[1], LIN Yi-qin[2]
(1. Modern Education Technology Center, Hunan University of Science and Engineering, Yongzhou 425100;
2. Dept. of Mathematics, Hunan University of Science and Engineering, Yongzhou 425100)

【Abstract】 Attribute data uncertainty can directly affect the quality of Geographic Information System(GIS)-based decision-making. The lifecycle of GIS attribute data is divided into a lot of stages, such as data collection and preparation, digital information extraction, information integration and the expression of attribute data. This paper explores the propagation of the attribute uncertainty at all stages and constructs propagation model of attribute uncertainty. It concludes synthesis and propagation algorithm of attribute uncertainty and tests the availability and effectiveness of the model.

【Key words】 attribute data; uncertainty; Geographic Information System(GIS); propagation model; sensitivity analysis

在地理信息系统(Geographic Information System, GIS)中嵌入 HTTP 和 TCP/IP 标准的综合应用技术体系，利用 Internet 在 Web 上发布空间数据，为用户提供空间数据的浏览、查询、分析等功能，已成为 GIS 发展的必然趋势[1]。对空间数据而言，GIS 采样获取的数据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，甚至有些严格或绝对意义上的真值往往并不存在。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)。因此，空间数据质量中的不确定性问题越来越为人重视，地理信息不确定性模拟问题已被美国国家地理信息与分析中心(NCGIA)列为优先研究的课题[2]。

1 相关概念

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，其内涵随着科学的进步，随着人们认识的逐步深入而不断增加和丰富。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。

传统 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测量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[3]。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，因此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。

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，在 GIS 中，这种不确定性通过空间分析传播，也许还会进一步被放大。而数据的可用性和足够的精度信息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后才能最终保证。不确定性研究正是解决这一问题、实现数据传递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

GIS 不确定性问题的研究目前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，离实际应用还有一定差距。除了以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

基金项目：国家自然科学基金资助项目(10801048)；湖南省自然科学基金资助项目(08JJ6043)；湖南科技学院青年基金资助项目(08XKYTB011)
作者简介：周迪民(1974－)，男，实验师、博士研究生，主研方向：空间数据库，地理信息系统；林依勤，副教授、博士
收稿日期：2009-08-06 **E-mail**：yz_zdm@163.com

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 模糊集合及粗集理论方法

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，用[0, l]中的某个值表示空间实体隶属于某类型的可信度，即元素隶属度，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。与传统的集合不同，部分属于关系在模糊集中是允许的。这意味着类别之间是可以重叠的，而隶属函数应该反映类别间的重叠或转化[4]。

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 2 个实体有完全相同的信息，则它们为等价关系，不可区分。粗糙集可以用来分析和解决由信息不完整而造成的不精确和不确定问题[5]，可描述属性 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等。粗集理论在研究 GIS 属性不确定性方面是很有价值的，只是目前的研究深度和力度还远远不够。

2.2 概率论及数理统计方法

概率论和数理统计用于处理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在 0~1 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。

空间统计学(Spatial Statistic)利用有序的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。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 云理论方法

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 3 个部分构成。云在空间由系列云滴组成，具有期望值、熵和超熵 3 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等 GIS 中容模糊性和随机性为一体的属性不确定性问题。

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。在此基础上，文献[6]提出基于数学度量的计算机可视化技术，从视觉感知角度表现属性不确定性信息。

3 属性不确定性的传播

原始数据存在不同程度的误差，经过各种处理、转换后生成的产品也保存着原有的误差，即属性不确定性的传播[7]。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。由于源数据不可避免的具有不确定性，再加上操作误差，因此这种 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。

3.1 传播机理分析

不确定性传播是原始误差或数据中的误差在数据处理或应用过程中，各种误差之间相互影响，所引起不确定性在处理结果中的积累。假设表示含有不确定性空间数据为 $U(D, S, T, \alpha, \beta, \gamma, \delta, \varepsilon)$，其中，$U$ 表示空间数据集合；D 表示空间数据的值；S 表示数据集合在空间上的延伸；T 表示空间数据集合在时间上的延伸。D, S, T 这 3 种数据都含有不确定性。α, β, γ, δ, ε 分别表示空间数据的值、空间的、时间的、逻辑一致性以及完整性方面的不确定性，这些不确定性之间相互影响，使得空间数据的不确定性呈现出复杂性。

根据对空间数据不确定性传播的分类，表示空间数据不确定性的积累过程如下：

$$U(D, S, T, \alpha, \beta, \gamma, \delta, \varepsilon) \xrightarrow{\text{model}} U'(D', S', T', \alpha, \beta, \gamma, \delta, \varepsilon)$$

在对空间数据进行处理过程中所出现的不确定性生成现象表示如下：

$$U(D, S, T, \alpha, \beta, \gamma, \delta, \varepsilon) \xrightarrow{\text{model}} U'(D', S', T', \alpha+\Delta\alpha, \beta+\Delta\beta, \gamma+\Delta\gamma, \delta+\Delta\delta, \varepsilon+\Delta\varepsilon)$$

每种数据都存在于一定的模型当中，表示为 $U(D, S, T, \alpha, \beta, \gamma, \delta, \varepsilon):(CX)$，其中，$CX$ 表示为一种概念模型。

属性不确定性传播表示为 $Z=M(u)(D_1, D_2, \cdots, D_n)$，其中，$Z$ 为 GIS 属性数据的不确定性；$M(u)$为数据操作过程；D_i 为数据集。

(1)如果 Z 为线性函数，则可用误差传播定律计算属性数据的不确定性，用方差表示为 $Z=\sum_{i=1}^{n} k_i x_i$，其中，K_i 为常数；X_i 为观测值。其方差和协方差分别用 δ_i , δ_{ij} 表示，根据方差运算的加法定理可得：

$$\delta_Z^2 = D(\sum_{i=1}^{n} k_i x_i) = \sum_{i=1}^{n} k_i^2 D(X_i) = \sum_{i=1}^{n} k_i \delta_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} k_i k_j \delta_{ij}$$ 。

(2)如果 Z 为非线性函数且可导，则 Z 可转化成线性函数，再利用误差传播定律计算：$Z=f(x_1, x_2, \cdots, x_n)$，若已知 $\underset{n\times 1}{X}$ 的协方差 D_{xx}，若函数 Z 可导，则可得：

$$\Delta Z = \frac{\partial f}{\partial x_1} x_1 + \frac{\partial f}{\partial x_2} x_2 + \cdots + \frac{\partial f}{\partial x_n} x_n$$

(3)绝大多数情况下 Z 不是连续、可导或由误差传播定律引起的近似误差是不可接受的，因此，不能直接用解析方法计算属性数据的不确定性。

3.2 传播模型

不确定性传播在人工智能(AI)领域是研究的前沿和热点，在 AI 中称不确定性推理。然而，目前尚无描述属性数据的不确定性传播模型。由于 GIS 属性数据的不确定性是整个属性数据生命周期所有不确定性的综合结果，因此对于不确定性的分析和处理都是分阶段的，不确定性在各个阶段之间

的传播规律是一个还未解决的难题。初步认为，GIS 属性数据整个生命周期可以看作是由多个阶段组成的复杂系统，可分为数据采集与准备、数字化信息提取、数据化简(模型、图形综合)和属性数据表达(图形显示)等环节，可以写出每一过程的显式或隐式传递函数，将复杂问题简单化，并逐个解决。在此理论构思下，初步提出一个属性不确定性传播模型，基本模式如图 1 所示。

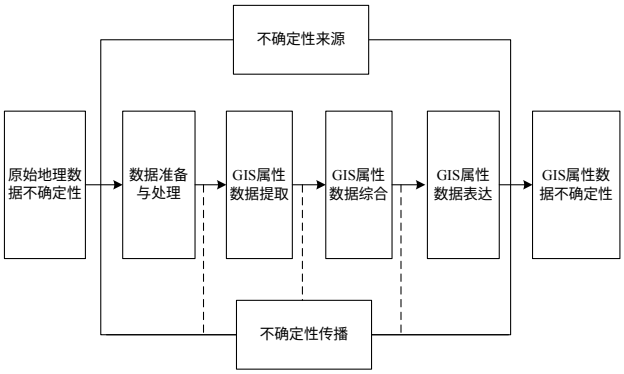

图 1　GIS 属性数据不确定性传播模型

从图 1 可知，导致属性不确定性的根源是原始地理数据的不确定性。原始地理数据的不确定性主要是由于地理数据本身的复杂性和变化多样性导致的，其次是由于人类认识的不完备性导致的。原始地理数据的不确定性会在 GIS 整个过程中积累与传播，而且前一阶段的不确定性又会传播给后一阶段，从而导致相当数量的不确定性积累。

(1)不确定性的合成算法

设数据集 D_A, D_B 的不确定度分别为 TU_A, TU_B，结果数据集为 D_C，不确定度为 TU_C。当合取属性数据的不确定性时，$D_C=D_A$ and D_B，则 $TUc=\min(TU_A, TU_B)$，推广到 n 个数据集：

$D_C=D_1$ and D_2 and $\cdots$ D_{n-1} and D_n

$\bigcap_{i=1}^{n} TU_i = \min\{TU_1, TU_2, \cdots, TU_n\}$

$TU_C=\min\{TU_1, TU_2, \cdots, TU_n\}$

当析取属性数据的不确定性时，$D_C=D_A$ or D_B，则 $TU_C=\max(TU_A, TU_B)$，推广到 n 个数据集：

$D_C=D_1$ or D_2 or $\cdots$ D_{n-1} or D_n

$\bigcup_{i=1}^{n} TU_i = \max\{TU_1, TU_2, \cdots, TU_n\}$

$TU_C=\max\{TU_1, TU_2, \cdots, TU_n\}$

(2)不确定性的传播算法

设操作前数据集为 D_A，不确定度为 TU_A，操作后结果数据集为 D_C，不确定度为 TU_C，操作 $D_A \rightarrow D_C$ 的不确定度为 TU_M。

当 TU_A, TUc 已知，求操作不确定度 TU_M：

$$TU_M = \begin{cases} \dfrac{TU_A - TU_C}{1 - TU_C} & \text{当} TU_A \geq TU_C \\ \dfrac{TU_C - TU_A}{TU_C} & \text{当} TU_A < TU_C \end{cases}$$

当 TU_A, TU_M 已知，求结果不确定度 TU_c：

$TUc=TU_M \times TU_A$

3.3　灵敏度分析

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。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。这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。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。

灵敏度分析在栅格数据和地图叠置应用中，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。灵敏度分析对属性不确定性传播的研究还处于摸索阶段，对于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问题还需进一步深化研究。

4　结束语

属性不确定性是 GIS 数据不确定性研究范畴的重要组成部分。国内外对属性数据不确定性研究已取得较多成果，但仍有如下一些问题需要进一步研究：

(1)GIS 每个阶段产生的属性不确定性之间的相互作用，如地理数据源的差异和比例尺的差异是如何通过空间分析传播和放大的。研究各种不确定性原因之间的关系，以便更加精确地评价 GIS 最终结果中存在的不确定性。

(2)扩展现有的数据模型，使它能在一个地理数据集中描述属性不确定性的空间变化；研究开发用于表示地理数据属性不确定性的有效的可视化工具，以反映属性不确定性的时空变化；研究地理数据不确定性的评价标准，规范地理数据不确定性的量化和可视化。

(3)如何将当前的属性数据不确定的各种表达式的分析结果变为用户可操作的确定性标量(如最大风险值，平均风险值等)，使用户得到空间数据质量和空间信息服务质量的确定性指标，这样才可能使空间数据不确定性研究走向实用。

参考文献

[1] 张俊耀, 成　筠, 郑丙辉. 基于 WebGIS 的河口水环境管理信息系统[J]. 计算机工程, 2008, 34(24): 279-281.

[2] 舒　红, 齐翠红. 地理信息时态不确定性的语义与计算[J]. 武汉大学学报: 信息科学版, 2007, 32(7): 633-636.

[3] 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.

[4] 程　涛, 邓　敏, 李志林. 空间目标不确定性的表达方法及其在 GIS 中的应用分析[J]. 武汉大学学报: 信息科学版, 2007, 32(5): 389-393.

[5] 李利伟, 马建文, 欧阳赟. 卫星遥感数据分类不确定性的容差粗糙集处理[J]. 计算机工程, 2008, 34(6): 1-2.

[6] 葛　咏, 李三平. 遥感分类信息不确定性的可视化表达方法[J]. 地球信息科学, 2008, 10(1): 88-96.

[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.

编辑　金胡考