

文章编号: 1671-8860(2006)07-0569-04

文献标志码: A

# 遥感与地理信息系统数据的信息量及不确定性

林宗坚<sup>1</sup> 张永红<sup>1</sup>

(1 中国测绘科学研究院, 北京市海淀区北太平路 16 号, 100039)

**摘要:** 讨论了遥感、GIS 数据的不确定性与信息论中的不确定性间的联系, 导出了 GIS 图形数据与遥感影像数据的信息量估算式, 提出了位置疑义度和属性疑义度等概念。在统一的数学基础上, 估算几何位置误差和属性正确率不足引起的不确定性, 从而建立起遥感影像与 GIS 图形数据的信息量及不确定性的统一量度。

**关键词:** 遥感; GIS; 信息量; 不确定性

中图法分类号: T P751; P208

十几年来, 关于 GIS 数据不确定性的研究取得了很大的进展<sup>[1-4]</sup>, 尤其在利用误差传播定律估算提供给用户数据的位置误差和利用概率论方法估算属性正确率方面取得了许多有实用价值的成果。但至今为止, 尚未建立起对两种不确定性的统一量度指标体系, 关键问题在于没有用统一的数学基础去描述这两种不确定性。本文通过研究遥感、地理信息系统的不确定性与信息论中的不确定性概念间的联系, 形成在信息理论高度上的统一数学基础, 用以解决遥感与 GIS 数据的信息量及不确定性的定量评估问题。

## 1 不确定度与信息量

设某一事物以  $n$  种状态存在, 其概率分别为  $p_i$ ,  $i = 1, 2, \dots, n$ , 按信息论的观点<sup>[5, 6]</sup>, 它的不确定度可以用熵的形式表达为:

$$E = - \sum_{i=1}^n p_i \lg \frac{1}{p_i} = - \sum_{i=1}^n p_i \lg p_i \quad (1)$$

当取 2 为底数时, 式(1)求得的熵的单位为比特(bit)。

如果对该事物进行观测, 得到它的全部信息, 即完全消除了对该事物的不确定性, 则得到的信息量的表达式同式(1)。如果仅对该事物的  $K$  种状态( $K < n$ )进行观测, 剩下  $n - K$  种状态未知, 则所获得的信息量为:

$$H = - \sum_{i=1}^K p_i \lg p_i \quad (2)$$

对此事物的剩余不确定度为:

$$U = - \sum_{i=K+1}^n p_i \lg p_i = E - H \quad (3)$$

以上是假设观测值没有误差。若观测信号存在误差(即噪声), 则获得的信息量为:

$$H = H_{(s)} - H_{(n)} \quad (4)$$

式中,  $H_{(s)}$  和  $H_{(n)}$  分别表示信号与噪声的熵。

若观测误差服从正态分布, 其方差为  $\delta^2$ , 则此误差带来的疑义度<sup>[7, 8]</sup>为:

$$H_{(n)} = \ln \sqrt{2\pi e \delta^2} \quad (5)$$

因而式(4)可具体化为:

$$H = \sum_{i=1}^n p_i \ln \frac{1}{p_i} - \ln \sqrt{2\pi e \delta^2} \quad (6)$$

## 2 地理空间图形数据的信息量

### 2.1 几何图形的信息量度

图 1 标示出了一个正方形单元, 并涂上某种颜色, 这种颜色是等可能性出现的  $2^C$  种颜色中的一种, 则得到的信息量为  $C$  比特。如果每种颜色代表一种地理目标属性, 设共有  $n$  种地理属性, 按式(1)所示的概率出现, 则该单元所能表示的属性给人们的信息量为:

$$H = \sum_{i=1}^n P_i \lg \frac{1}{P_i} \quad (7)$$

图2所示为平面上的一个正方形单元,且被分割成 $2^k \times 2^k$ 的微小栅格。此时,若被告知某一目标处于栅格网中 $X = i, Y = j$ 的栅格,则得到 $K + K = 2K$ 比特的信息量。这意味着,位置不确定性(度)是空间分辨率的对数函数。由此可知,一个点位的空间分辨率越高,所得到的信息量就越大。

图3所示是地图上某一面积单元的典型图形。其中,点的位置信息量为 $2K$ 比特,属性信息量为 $C_1$ 比特;折线(曲线在此栅格分辨率的近似)由 $m$ 个折点组成,它的位置信息量为 $2Km$ 比特,属性信息量为 $C_2$ 比特;多边形由 $s$ 个角点构成,它具有 $2Ks$ 比特的位置信息量和 $C_3$ 比特的属性信息量。图3说明,地图图形的全部点、线、面要素都可以用分辨率的对数函数的形式计算其信息量,图3中全部图形的信息量为 $H = 2K + 2Km + 2Ks + C_1 + C_2 + C_3$ 。

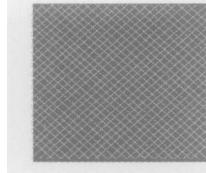


图1 颜色

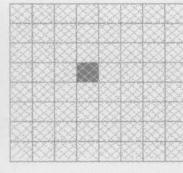
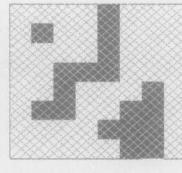


图2 点

图3 点、线、面  
Fig. 1 Color      Fig. 2 Point      Fig. 3 Point, Line  
and Area

## 2.2 位置疑义度

地图图形含有来自测绘过程的误差( $\Delta x, \Delta y$ )。由于大多数测量误差服从正态分布规律,因此,地图图形的位置疑义度为:

$$H_{(\Delta x)} = H_{(\Delta y)} = \ln \sqrt{2\pi e \delta^2} \quad (8)$$

$$H_{(\Delta p)} = H_{(\Delta x)} + H_{(\Delta y)} = 2 \ln \sqrt{2\pi e \delta^2} = 2(1.42 + \ln \delta) \quad (9)$$

式中,  $\delta$ 为地图图形坐标的中误差(标准差),且 $\delta$ 恒为正整数,因为它以最小空间分辨率为单位(见图2); $H_{(\Delta x)}$ 和 $H_{(\Delta y)}$ 为坐标疑义度; $H_{(\Delta p)}$ 为点位疑义度。

依图2的情况,该点位的有效信息量为:

$$H_{(F)} = 2K - 2(1.42 + \ln \delta) + C_1 \quad (10)$$

图3中图形的信息量为:

$$H_{(F)} = [2K - 2(1.42 + \ln \delta)] \cdot [1 + m + s] + C_1 + C_2 + C_3 \quad (11)$$

## 2.3 属性疑义度

设某一地区范围的某一地图中共表达了 $A = 2^a$ 种地类目标,各类地物出现的概率分别为 $P_1, P_2, \dots, P_A$ ,则 $A$ 种图形属性的平均信息量为:

$$H_{(A)} = - \sum_{i=1}^A P_i \lg P_i \leq a \quad (12)$$

若地图表达这 $A$ 类属性的正确率分别为 $q_1, q_2, \dots, q_A$ ,则此地图实际提供的面状地物属性的每个图斑的平均信息量为:

$$H_{(q_A)} = - \sum_{i=1}^A q_i P_i \lg P_i \quad (13)$$

其疑义度为:

$$V = H_{(r_A)} = - \sum_{i=1}^A (1 - q_i) P_i \lg P_i \quad (14)$$

按类似方法可以描述地图上的线状、点状地物图形(符号)属性的信息量及疑义度。考虑到属性疑义度,图3中图形的信息量应在式(11)中减去式(14)所示的疑义度,即

$$H_{(F)} = [2K - 2(1.42 + \ln \delta)](1 + m + s) + C_1 - V_1 + C_2 - V_2 + C_3 - V_3 \quad (15)$$

显然,图3以外的以点、线、面矢量图形表达的地图信息都可以仿此形式描述其信息量。

## 3 遥感影像数据的信息量

设有 $b$ 个波段、每景 $m$ 像元 $\times m$ 像元的遥感影像,像元灰度按 $q$ 比特量化,则每景影像数据的存储总量为:

$$Q = qbm^2 \quad (16)$$

但此数据的信息量远不止此。遥感影像的信息量取决于灰度量化等级、灰度噪声、空间分辨率、几何变形、邻元相关性、波段间相关性等因素。

### 3.1 单像元灰度的信息量

实验证明,遥感影像的灰度信号与噪声皆成正态分布<sup>[7,9]</sup>,设其方差分别为 $\xi^2$ 和 $\eta^2$ ,则有:

$$H_{(\xi)} = \ln \sqrt{2\pi e \xi^2}, H_{(\eta)} = \ln \sqrt{2\pi e \eta^2} \quad (17)$$

因此,单像元灰度的平均信息量可表示为<sup>[8,10]</sup>:

$$H_{(g)} = H_{(\xi)} - H_{(\eta)} = \ln(\xi^2 / \eta^2) \quad (18)$$

这说明单像元灰度的信息量等于其信噪比的对数。

表1列出了在256级灰度情况下几种典型信噪比的像元灰度信息量<sup>[8]</sup>。表1数据说明,采用很简单的单像元编码压缩方法就可能实现约2.3倍的完全无损压缩,亦即解压缩后可以完全恢复原数据。

表1 几种典型信噪比的像元灰度信息量/bit

Tab. 1 Grayscale Image Entropy Under Different

Signal Noise Ratios/ bit

$\delta_\xi$	$\delta_\eta$	$H_{(g)}$
48	4	2.48
48	3	2.77
48	2	3.18
32	1	3.47

### 3.2 邻元相关性

实践证明,遥感影像灰度具有一阶马尔可夫过程的统计特征,其协方差阵可描述为<sup>[11,12]</sup>:

$$C = \delta^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{t-1} \\ \rho & 1 & \rho & & \\ \rho^2 & \rho & 1 & & \vdots \\ \vdots & & & & \\ \rho^{t-1} & \dots & & & 1 \end{bmatrix} \quad (19)$$

式中,  $\delta^2$  为信号方差;  $\rho$  为相邻像元间的自相关系数,  $\rho \leq (1 - \delta^2/\delta_0)$ 。笔者测试过一片航空影像灰度的自相关系数值,当位移的像元数目<sup>[8]</sup>  $\tau$  分别等于 0、1、2、3、4、5 时,该航空影像的自相关系数  $\rho^{\tau}$  分别为 1.000、0.831、0.700、0.575、0.468、0.386。

图 4 所示为按照一阶马尔可夫特性理解的遥感影像像元之间的相关性,图中“1”表示为中心像元,按 8 联通原则,  $\rho$  表示  $\tau=1$  的邻像元与中心像元的相关系数,  $\rho^2$  表示  $\tau=2$  的邻像元与中心像元的相关系数。按信息论原理,相邻像元间的互信息量为<sup>[7]</sup>:

$$I(\tau, \tau+1) = -\ln(1 - \rho) \quad (20)$$

由中心像元增加  $t$  个像元时,所增加的信息量为<sup>[8]</sup>:

$$H_{(t)} = t \ln(\delta_0/\delta_1) + t \ln(1 - \rho) \quad (21)$$

值得注意的是,一般情况下,式(21)的第一项为正值,第二项为负值。因此可认为,由于邻元相关性,平均每像元的信息量较单像元减少  $\ln(1 - \rho)$ 。当  $\rho = 0.9, 0.8, 0.7, 0.65, 0.6$  时,对应的信息减少量  $\ln(1 - \rho) = -2.3, -1.6, -1.2, -1.0, -0.9$ 。

$\rho^2$	$\rho^2$	$\rho$	$\rho$	$\rho$
$\rho^2$	$\rho$	$\rho$	$\rho$	$\rho$
$\rho^2$	$\rho$	1	$\rho$	$\rho$
$\rho^2$	$\rho$	$\rho$	$\rho$	$\rho$
$\rho^2$	$\rho^2$	$\rho$	$\rho$	$\rho$

图 4 邻元相关性

Fig. 4 Correlation of Neighboring Pixels

依此原理容易理解,一幅模糊影像的信息量比一幅清晰影像的信息量要少,一幅水面影像的信息量比一幅城市影像的信息量要少。

### 3.3 几何变形的影响

遥感影像的几何畸变使其像元位置偏离其地面(或地图)目标对应的位置,因此引起的位移疑义度为:

$$H_1 = \ln \sqrt{4\pi e \delta^2 (1 - \rho^2)} \quad (22)$$

式(22)表明,影像自相关系数越小,或灰度方差越大,对几何畸变的容忍能力就越差。因此,高分辨率影像的几何纠正问题十分重要。高精度几何纠正有助于降低疑义度,提高信息量。几何纠正对于多源影像融合尤为重要,位移疑义度应该成为评估融合影像质量的一个指标。

综上所述,式(16)所计的数据量应当按式(18)、(21)、(22)折算成信息量。影像信息量度对评估传感器系统的性能、确定不同地理区域的遥感影像获取的方法、指导数据压缩和图像识别处理算法的研制等都是很有用处的。

### 3.4 多波段遥感影像的信息量

类似于 § 3.2 的原理,可以计算多波段遥感影像的信息量。相邻波段间的互相关性决定了相邻波段间的互信息量。设不考虑互相关性,第  $j$  和  $j+1$  波段影像的信息量分别为  $H_j$  和  $H_{j+1}$ ,则考虑相关性后,两个波段的总信息量为:

$$H = H_j + H_{j+1}(1 - \rho) \quad (23)$$

式中,  $\rho$  为两波段间的互相关系数。依此类推,可以计算多波段影像的信息量。

## 4 结语

通过本文的分析,可以得到如下结论:遥感与地理信息系统数据的不确定性问题可以提升到概率与信息理论的高度进行研究。不确定度、信息量、疑义度都可以用熵的数学形式表达。

当估算信息量时,位置不确定度与属性不确定度可以用统一的信息论方法量度。

## 参 考 文 献

- [1] Dutton G. Handling Positional Uncertainty in Spatial Database[ C ]. International Symposium on Spatial Data Handling, South Caroline, USA, 1992
- [2] Shi Wenzhong. Modeling Appositional and Thematic Uncertainties in Integration of Remote Sensing and Geographic Information Systems [ M ]. Enschede: ITC Publication, 1995
- [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] 承继承, 郭华东, 史文中. 遥感数据的不确定性问题[ M ]. 北京: 科学出版社, 2003
- [5] 沈世镒, 吴忠华. 信息论基础与应用[ M ]. 北京: 高等教育出版社, 2004
- [6] 钟义信. 信息与通信理论[ M ]. 北京: 人民邮电出版社, 1979
- [7] 谢夷杰. 概率论[ M ]. 北京: 人民邮电出版社, 1985

- [8] 林宗坚. 多重信息多重判据影像匹配[D]. 武汉: 武汉测绘科技大学, 1988
- [9] Janes E T. Probability Theory: the Logic of Science [M]. Cambridge: Cambridge University Press, 2003
- [10] Landgrebe D A. Signal Theory Methods in Multispectral Remote Sensing[M]. New York: John Wiley & Sons., 200
- [11] 胡正名. 数字信号处理中的正交变换[M]. 北京:

人民邮电出版社, 1979

- [12] 吴成柯, 樊昌信. 计算机图形显示和图像处理的算法[M]. 北京: 科学出版社, 1987

第一作者简介: 林宗坚, 教授, 博士生导师。主要从事摄影测量与遥感方面的研究。主要研究成果: 多重判据影像匹配、4D技术及应用、无人机低空遥感系统等。发表论文 100 余篇, 著作 2 部。E-mail: lincasm@casm.ac.cn

## Measurement of Information and Uncertainty of Remote Sensing and GIS Data

LIN Zongjian<sup>1</sup> ZHANG Yonghong<sup>1</sup>

(1 Chinese Academy of Surveying and Mapping, 16 Beitaiping Road, Haidian District, Beijing 100039, China)

**Abstract:** The relation between the concept of remote sensing/ GIS uncertainty and the uncertainty of information theory is discussed, to enable quantitative evaluation of the measurement of information and uncertainty of remote sensing and GIS data. The formulation for calculating information content of GIS graphics data is inferred based on the definition of uncertain degree in information theory. With this formulation, the information of geometric position and attribute can be measured in the unit of bit. Furthermore, the information of any geographic target in GIS is measurable by considering the doubtfulness of geometric position and attribute. Besides GIS vector data, the information of remote sensing raster imagery is also discussed.

**Key words:** remote sensing; GIS; information measurement; uncertainty

**About the first author:** LIN Zongjian, professor, Ph. D supervisor, majors in photogrammetry and remote sensing. His main scientific contributions focus on image matching with multi-criteria, 4D technology and applications, and low latitude unmanned remote sensing system. He has authored over 100 papers and 2 books.

E-mail: lincasm@casm.ac.cn

### 下期主要内容预告

- |                          |        |
|--------------------------|--------|
| 种新的基于内容遥感图像检索的图像分块策略     | 李德仁, 等 |
| 种改进的三维点集表面重建的区域生长算法      | 朱庆, 等  |
| 基于角色和任务的层次对象访问控制模型       | 何超英, 等 |
| 面向地理环境主体 GIS 研究          | 龚建华, 等 |
| 基于整体变分模型的影像阴影检测算法研究      | 王树根, 等 |
| 线与面之间基本拓扑关系类型在抽象过程中的等价转换 | 郭庆胜, 等 |