

地学粗空间的理论与应用

王树良¹ 李德仁¹ 史文中² 王新洲³

(1 武汉大学测绘遥感信息工程国家重点实验室, 武汉市珞喻路 129 号, 430079)

(2 香港理工大学土地测量与地理资讯学系, 香港, 九龙红石坳)

(3 武汉大学测绘学院, 武汉市珞喻路 129 号, 430079)

摘要: 提出了由粗实体、粗关系和粗算子组成的地学粗空间。在分析通过粗集研究地球空间信息学不足的基础上, 提出了简化的实用粗符号系统; 定义了地学粗空间的内涵和外延, 把空间论域全域划分为下近似集、边界集和非集, 将空间实体以上、下近似集表示为粗实体, 并用粗度衡量其逼近程度; 给出了兼顾位置和属性的不确定性的粗关系及其矩阵表示; 研究了地学粗空间的粗算子, 重点讨论了粗交集近似通用矩阵、基于影像灰度值的粗隶属函数和冗余属性剔除等粗算子; 以基于地学粗空间的河流专题遥感影像分类为例, 检验了地学粗空间的实用性。结果表明, 地学粗空间可以较大可能地逼近地学实体的真实存在形式, 决策信息较为丰富、可靠。

关键词: 地学粗空间; 粗实体; 粗关系; 粗算子; GIS

中图法分类号: P208; TP 18

计算机化的 GIS 是对空间现实的近似表达。在复杂多变的现实世界中, 空间实体多相互混杂, 界线有时不很分明。人们在利用 GIS 认识和改造自然的过程中, 常不易明确地将空间实体分类, 所获信息也不可能穷尽现实世界的全部, 而且已获取的信息在存储和分析中又被部分舍弃或删除(如制图综合)^[1]。所以研究逼近地学实体的现实存在形式的理论和方法, 具有现实意义。

逼近地学实体的理论和方法的研究正为人们所重视。为了避免利用错误信息导致决策失误, 人们提出或发展了多种近似表达地学实体的理论和方法。例如, GIS 的目标模型和域模型、概率论、概率矢量、证据理论、空间统计学、粗集、模糊集、云理论、遗传算法、混沌理论、灰色系统理论、未确知数学、GIS 灵敏度分析、可视化技术、类别变换区域模型等。其中, 粗集是 Pawlak 提出的一种根据不完备信息分析决策的方法^[2], 已经被应用在医疗诊断、机器学习、人工智能、模式识别、近似推理、知识发现等领域。在此过程中, 粗集日臻完善, 从初始的偏重定性分析(如上、下近似集的

描述、最小决策集的生成)发展到现在的定性定量并重, 并且与模糊集、概率论和证据理论等互相交叉。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^[3]。

但是, 粗集的下近似集是上近似集的子集, 当利用粗集描述空间现实世界, 计算分析, 推导决策规则、存储实体及其处理的过程和结果时, 下近似集中的元素都把计算资源的静态存储空间和缓冲存储空间占有或使用了两次, 造成计算资源的浪费。同时, 粗集毕竟不是研究地学的专门理论, 没有充分考虑空间的关系、分布、粒度、测度等特性, 以及属性数据的不确定性等, 如某些研究空间关系的工具尚未具备。而且, 由于应用的目的或领域的差异, 出现了各种粗集表示符号, 即使创始人 Pawlak 也几乎在不同的文章中使用了不同的符号。Pawlak 曾注意到问题的严重性, 认为“就发展和延伸粗集理论而言, 各种各样的粗集提法及其表示, 已经到了不得不予以统一的地步”, 并在《粗集元素》一文中试图给以统一。遗憾的是, 他在统一的过程中又再次创造了新的符号系统, 而且在

收稿日期: 2001-10-12。

项目来源: 国家重点基础研究发展规划(973 计划)资助项目(G19980305084); 国家高技术研究发展计划(863 计划)资助项目(2001AA135081); 国家自然科学基金资助项目(40023004; 49874002); 香港理工大学科研基金资助项目(1. 34. 37. 9709); 国家测绘局测绘科技发展基金资助项目(98017); 教育部博士点基金资助项目(98049801)。

随后的研究中也并不遵守自己的统一。这些,在一定程度上阻碍了粗集理论的发展^[4~11]。

可见,利用GIS研究地球空间信息世界的理论和方法,数学假设应该承认其不确定性,以及信息采集的不完备性^[2],符号表示也最好标准统一。因此,本文基于粗集,在地球空间信息学中提出了由粗实体、粗关系和粗算子组成的地学粗空间理论。

1 地学粗空间

地学粗空间(geo-rough space)由粗实体、粗关系和粗算子组成,基本单位是粗元。它是经典集合的一种扩展,适于处理不精确、不确定和不完整的数据。在地学粗空间中,粗实体利用上近似集和下近似集同时刻画空间实体的确定性和不确定性;粗关系表示粗实体之间的相互作用规律;粗算子是粗实体及其粗关系的数学模型和算法。地学粗空间的测度是经典测度的继承和发展。地学粗空间的粗元,在二维空间中类似于遥感影像的栅格,在多维空间内则是如同多维积木的单位体。

地学粗空间把粗集发展为一种针对地球空间信息学特性的专业理论。在空间分析时,地学粗空间重在研究不完备的属性信息,其数学基础为上、下近似集,不同于模糊集的模糊隶属度、证据理论的证据函数、云理论的隶属度概率空间分布等。粗集符号系统被地学粗空间简化统一后,每个符号含有粗集意义和应用专业领域意义两部分,如在 $L_r(X)$ 中, L_r 表示粗集下近似集(lower approximation of rough set); X 表示地学实体。如果设 U 为非空的有限论域,那么 U 被 $R(R \subseteq U \times U)$ 分割成的不相交子集,称为粗元 R_e (rough element), U 和 R 则共同构成一个地学粗空间 (U, R) 。对于任意的集合 $X \subseteq U(x \in X)$,以 $[x]_{R_e}$ 表示包括地学元素 x 的所有粗元 R_e 。因为地学粗空间是粗集在地球空间信息学内的应用和发展,所以它的上、下近似集的定义仍然与粗集基本相同。

定义1 X 在 U 中的下近似集为 $L_r(X) = \{x \in U | [x]_{R_e} \subseteq X\}$,上近似集(upper approximation of rough set)为 $U_r(X) = \{x \in U | [x]_{R_e} \cap X \neq \Phi\}$,边界集(boundary approxiamtion of rough set)为 $B_r(X) = U_r(X) - L_r(X)$ 。

显然, $L_r(X) \subseteq X \subseteq U_r(X) \subseteq U$ 。一般地,空间现实世界中的实体及其属性,确定的占绝大部分,不确定的只是少数。因粗集总是对下近似集

计算和存储两次,为了避免计算资源的浪费,在实际计算分析中传递地学实体 X 的不确定性时,地学粗空间不是像粗集一样使用一对上、下近似集,而是以 X 的完全不确定的边界集 $B_r(X)$ 代替 $U_r(X)$ 。也就是说,在地学粗空间中参加计算和分析的是 X 的一对下近似集和边界集。

1.1 粗实体

在地学粗空间的论域 U 中,用下近似集 $L_r(X)$ 和上近似集 $U_r(X)$ 定义的实体 X 称为粗实体,记作 $R(X)$ 。当 $L_r(X) \neq U_r(X)$ 时, $R(X)$ 包含不确定性,不确定的部分为边界集 $B_r(X)$ 。例如遥感影像中的河流实体,其冬季枯水季节和夏季洪水季节的水涯线是不同的,致使河流和河滩的边界不确定。如果河岸再有森林,那么遥感影像中河流和河滩土壤、森林和河滩土壤、森林和河流的含混关系,若无先验知识作分类指导,则可能使河流的专题提取含更多不确定性。当 $L_r(X) = U_r(X)$ 时, $R(X)$ 没有不确定性,就变为确定空间实体 X 。例如行政区划实体就是确定的,边界集为空。因此,确定实体 X 是粗实体 $R(X)$ 的特殊形式,地学粗空间是经典确定集合的扩展。

定义2 粗实体 $R(X)$ 的非集为 $N_r(X) = U - U_r(X)$ 。

至此,论域 U 就被地学粗空间全域划分了,即 $U = L_r(X) \cup B_r(X) \cup N_r(X)$ 。明显地,粗实体 $R(X)$ 的非集 $N_r(X)$ 一般不同于实体 X 的补集 X 。因为在地学粗空间中,实体 X 的补集为 $X = U - X = U - U_r(X) + \emptyset$ 除非 $X = U_r(X)$ 且 $\emptyset = 0$ 。对于地学元素 $x \in U$,下近似集 $L_r(X)$ 是确定属于实体 X 真值的 x 集合,上近似集 $U_r(X)$ 是可能属于实体 X 的 x 集合, X 的不确定集合为边界 $B_r(X)$ 。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。还可以用粗元作进一步的解释,下近似集中的粗元具有足够必要的信息和知识,确定属于该实体类别;非集中的实体没有必要的信息和知识,确定不属于该实体类别;边界集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,为类别的边界。下近似集和边界集的并集为上近似集。为了度量粗实体 $R(X)$ 对实体 X 的逼近程度,再给出一个粗实体 $R(X)$ 的粗度 R_d (rough degree)。

定义3 粗实体 $R(X)$ 的粗度为:

$$R_d(X) = R_{\text{card}}(B_r(X)) / R_{\text{card}}(X) \times 100\% \quad (1)$$

其中, $R_{\text{card}}(X)$ 表示集合 X 的粗元基数。一

般地,难以得到 $R_{card}(X)$, 如果知道 X 的确定边界, 那么 $R(X)$ 可能就是不含不确定性的 X 。在实验中发现^[2], $R_{card}(X) \approx R_{card}(U_r(X))$, 因此可以使用 $R_{card}(U_r(X))$ 近似地代替 $R_{card}(X)$ 。当 $L_r(X) = U_r(X)$ 时, $R(X)$ 就褪变为用经典确定集合表示的实体 X , 其粗度 $R_d(X) = 0$ 。使用影像栅格数据描述的 GIS 空间实体, 就是粗实体的一个特例。一幅空间影像由分辨率一定的离散像素(即规则栅格)组成, 分辨率决定了像素的大小。影像分辨率越高, 栅格空间表示的实体 X 的粗度越小; 相反, 影像分辨率越低, 粗度越大。如果分辨率足够大, 或者栅格足够小, 那么粗实体的一对上、下近似集就会相等, 粗实体也褪变为没有粗度的现实实体。可是, 影像占用的计算机存储空间也随之增大, 这也是地学粗空间对随分辨率变化的遥感影像的一种可供选择的数学新解释。

图 1 描述了粗实体的上近似集、下近似集、非集等概念。其中, 图 1(b) 的空间分辨率大于图 1(a), 图 1(c) 是图 1(a) 的 3D(即三维)表示。从图 1 中不难看出, $L_r(X) \subseteq X \subseteq U_r(X) \subseteq U$, $U_r(X) = U - N_r(X)$, $B_r(X) = U_r(X) - L_r(X)$, $U = L_r(X) \cup B_r(X) \cup N_r(X)$ 。图 1(a) 的 $R_{card}(X) \approx R_{card}(U_r(X)) = 20$ 。图 1(a)、图 1(b) 和图 1(c) 的粗度分别为 $R_d(X)_{(a)} = R_d(X)_{(c)} = 16/20 = 0.80$, $R_d(X)_{(b)} = (65 - 33)/65 = 0.49$ 。

由于图 1(b) 的空间分辨率相对于图 1(a) 增大了, 因此图 1(b) 的粗度相对于图 1(a) 也减小了, 即 $R_d(X)_{(a)} > R_d(X)_{(b)}$ 。

粗实体是 GIS 在地学粗空间中分析决策的载体。地学空间信息的基本单元是一个连接空间实体的时空位置 pt 及其系列属性 a 的数组 (pt, a) , 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。在基于计算机化的 GIS 空间分析中, 由于空间真值不可能被精确地获知, 而且可获得的信息也是不完备的, 因此实体 X 不可能被精确地表述, 实体属性的观测值也通常和其真值不一致。当实体属性被重复观测多次后, 观测值就围绕其真值形成一个不确定域。带有不确定边界的地学实体^[12] 可以看作粗实体的雏形, 遥感影像的混合像素则是在地面位置上相邻的不同地物类别属性的综合反映。地学粗空间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。当使用粗实体 $R(X)$ 表示空间实体 X 时, 确定的真值集合为其下近似集, 确定的非值集合是其非集, 不确定的集合是其上近似集, 完全不确定的集合是其边界集。粗实体通过一对上、下近似集逼近地学空间实体, 可以兼容空间实体的位置不确定性和属性不确定性。因此, 地学粗空间可以较为全面地传播地学实体(确定性和不确定性)。

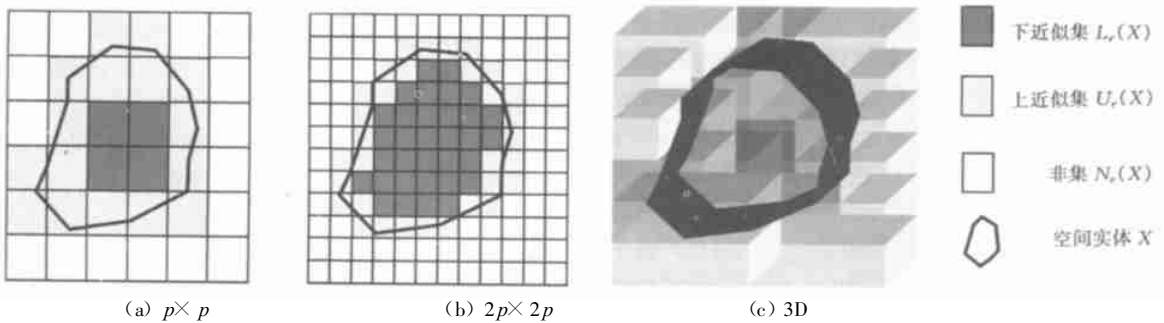


图 1 粗实体及其低分辨率、高分辨率和 3D 表示

Fig. 1 Rough Entity and Illustrations of Its Low Resolution, High Resolution and 3D

借助粗实体的定义, 还可以分析地学粗空间的测度和粗元。在零维、一维、二维、三维、...、 n 维的空间中, 量测实体的数学测度分别为点 O 、数轴 $O-X$ 、平面直角坐标系 $O-X_1X_2$ 、立体直角坐标系 $O-X_1X_2X_3$ 、...、多维坐标系 $O-X_1X_2 \dots X_n$ 。粗空间的数学测度继承了该标准, 但是粗实体在零维、一维、二维、三维、...、 n 维的粗空间中, 分别变成了点集、区间、带(或域)、三维体、...、多维体, 对应的粗元则分别为点、数值、栅格、三维单位体、...、多维单位体。具体而言, 向量

空间中的点、线和面, 栅格空间中的栅格或像素, 多维空间中的单位体构成粗实体的粗元。粗多维空间的多维单位体粗元, 组合起来可以表示各种空间目标, 如同使用多维玩具积木搭建各种自然物或建筑物一样。属于下近似集的单位空间立方体完全被包含在空间目标中, 空间目标的表面穿过的单位立方体构成目标的边界集(一个由粗元组成的中空壳), 不属于下近似集, 下近似集和边界集的并集为上近似集。如图 1(c) 所示的三维体, 其粗元为三维立方体, 组成了该三维体的论

域、下近似集、上近似集、边界集和非集。

1.2 粗关系

粗关系 R_r (rough relationships) 表示在同一地学粗空间中, 不同粗实体之间的空间相互作用规律, 可较为全面地刻画空间关系。其基础为粗等价关系, 基本关系为空间粗拓扑关系。

标准拓扑的含义是由 Munkres 在 1975 年定义的, 初始的空间拓扑关系是简单的点(零维), 线(一维)和面(二维)关系, 使用基于内部 X^0 和边界 ∂X 的 4 个交集模型表达。当其局限性暴露后, 这种模型又被扩充为基于内部 X^0 、边界 ∂X 和外部 X^- 的 9 个交集模型^[13, 14]。后来, 又出现了由内边界和外边界组成的宽边界的区域模型, 把 9 个交集模型的 2^9 个拓扑矩阵减少到 44 个, 并使用 0、1 值表达^[12]。Chen 等提出了基于

$$R_{r9}(A, B) = \begin{pmatrix} L_r(A) \cap L_r(B) & L_r(A) \cap B_r(B) & L_r(A) \cap N_r(B) \\ B_r(A) \cap L_r(B) & B_r(A) \cap B_r(B) & B_r(A) \cap N_r(B) \\ N_r(A) \cap L_r(B) & N_r(A) \cap B_r(B) & N_r(A) \cap N_r(B) \end{pmatrix} \quad (2)$$

可以看出, 式(2)能够区别和传播确定性 $L_r(X)$ 和 $N_r(X)$ 及不确定性 $B_r(X)$, 在 GIS 计算中分别使用 1 和 0 值表示非空和空。当不同的专题图层叠加时, 式(2)是通用的。在同一幅影像地图的粗空间中, $L_r(A) \cap L_r(B) = 0$ 。同时, 在式(2)中, 粗空间关系给出了兼顾确定性和不确定性的信息, 这能够提高影像解译等决策的质量。

因为面来自线, 线来自点, “点-点”、“点-线”、“点-面”、“线-线”、“线-面”的空间粗关系可以看作是“面-面”的特殊情况, 所以本文把“面-面”作为一个实例研究。其中, “粗面-面”拓扑关系主要以二维空间为例研究。图 2 描述了两个粗空间实体的相交关系。其中, $L_r(A)$ 、 $L_r(B)$ 分别是粗实体 A 、 B 的下近似集; $U_r(A)$ 、 $U_r(B)$ 分别是粗实体 A 、 B 的上近似集; $B_r(AB)$ 是 A 和 B 间的粗域, 是最不确定的部分。因为不确定域通常发生在边界, 在空间实体 A 、 B 的下近似集间不存在不确定域, 因此接触关系通常存在于影像分类的不确定过渡区, 由相邻类别的上近似集组成。

Voronoi 的 9 个交集模型, 用实体的 Voronoi 域 X^v 替代其外部 X^- , 矩阵中使用 O (空) 和 \emptyset (非空) 值表达^[15]。

可是, 人们所获信息不可能穷尽现实世界的全部, 在计算机化的 GIS 中存在不确定性, 已是一个不争的事实。根据获取的不完备信息, 通常很难保证空间实体的内部 X^0 、外部 X^- 、边界 ∂X 或 Voronoi 域 X^v 的确切性。同时, “相对于欧氏空间、几何空间和矢量空间, 拓扑属性是最基本的。”^[13] 因此, 粗拓扑学使用粗实体 $R(X)$ 的下近似集、边界集和非集代替 X 的内部、边界、外部或 Voronoi 域, 并主要以粗相交关系为例研究。

假设存在粗实体 $A, B \subseteq U$, 则二者之间的粗关系的矩阵可以表示为:

地学粗空间的粗实体, 除了包含不确定性的空间实体, 也有确切的空间实体。如前所述, 确切空间实体 X 是当 $L_r(X) = U_r(X)$ 时的一种特殊的粗空间实体。因此, 在同一地学粗空间中, 粗空间关系分为三种类型: 确切实体与确切实体间的粗关系(简称 CC 型)、粗实体与确切实体间的粗关系(简称 RC 型)和粗实体与粗实体间的粗关系(简称 RR 型)。在地学粗空间中, 以空间拓扑关系为基础, 可导出用于 GIS 计算的粗相离、粗相触/相遇、粗重叠、粗相等、粗覆盖、粗被覆盖、粗包含和粗被包含等粗关系。它们都可以使用粗矩阵表示, 并兼顾了位置和属性的不确定性。具体的 CC 型、RC 型和 RR 型分别如图 3(a)、图 3(b)和图 3(c)所示。

1.3 粗算子

粗算子是粗实体及其空间粗关系的算法, 即利用地学粗空间解决实际应用问题的数学算法, 也用作地学粗空间的定理证明, 具体表现为逻辑粗算子、代数粗算子和语气粗算子等。逻辑粗算

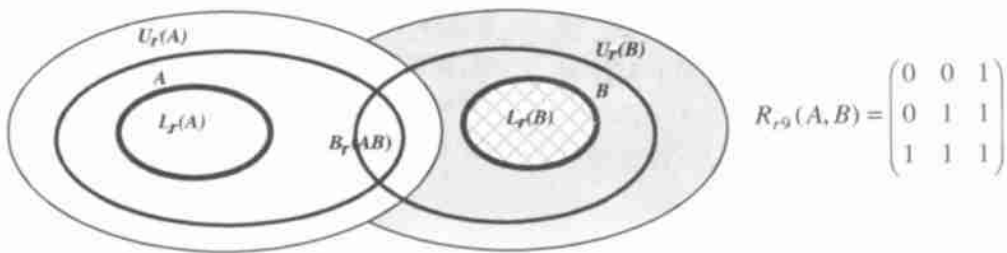
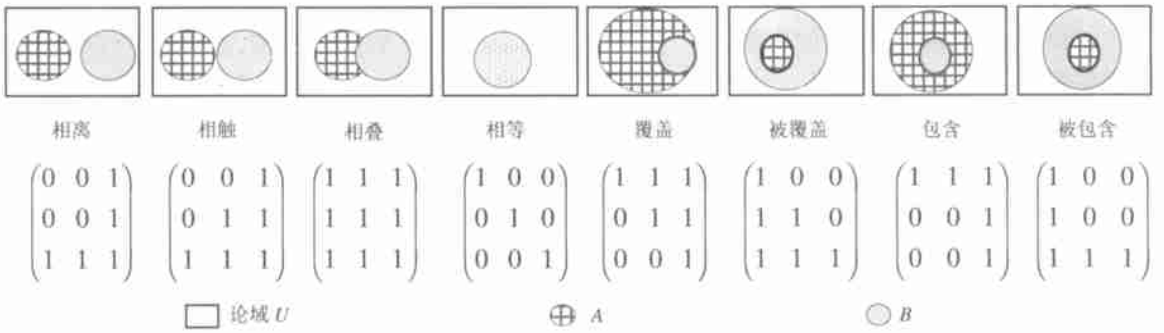
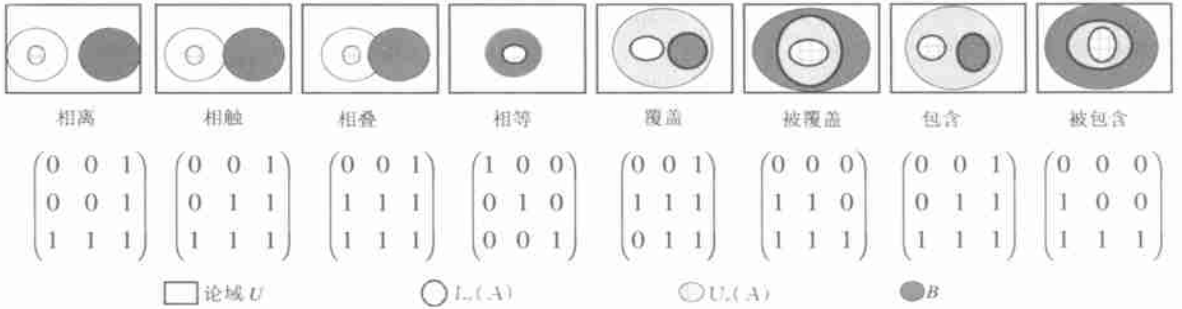


图 2 粗实体 A 、 B 之间的粗拓扑关系及其矩阵表示

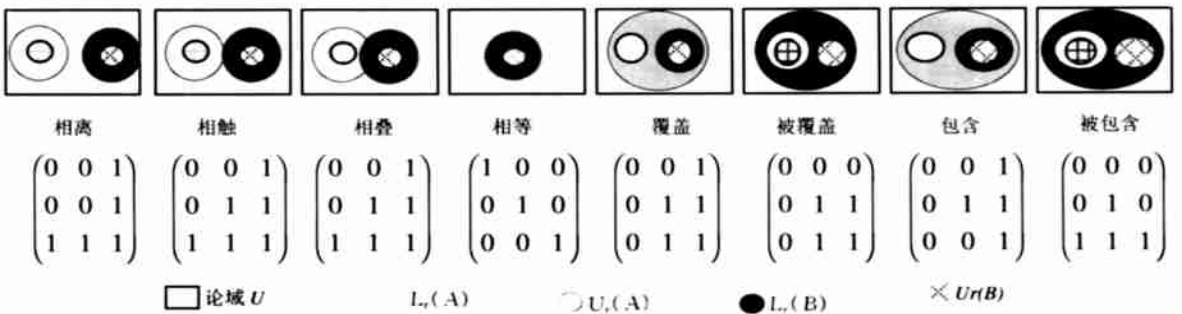
Fig. 2 Rough Spatial Topological Relationship between $R(A)$ and $R(B)$ and Its Rough Matrix



(a) CC型粗拓扑关系



(b) RC型粗拓扑关系



(c) RR型粗拓扑关系

图3 粗空间拓扑关系及其矩阵表示

Fig. 3 Rough Spatial Topological Relationships and Their Rough Matrices

子可以包括粗与、粗或、粗非、粗异或、粗相离、粗相交、粗相等、粗包含等。代数粗算子能够包括加、减、乘、除、数乘、倒数、映射、微积分等。逻辑粗算子和代数粗算子除了具有逻辑运算的一般性质外，还体现在其不确定性表示特征上。语气粗算子是增大或减小粗实体不确定性的运算法则，用以表达粗实体的不确定性大小的变化规律。这对于空间不确定性的“去粗存真”、空间数据挖掘和知识发现的知识精练等很有意义。为了后面的地学粗空间的影像专题提取实例应用，此处重点研究粗关系矩阵和粗隶属函数。

首先，推导通用粗关系矩阵。式(2)表示的是两个粗实体 A 、 B 之间的粗关系矩阵。如果研究粗实体 A 与它自己的关系，那么 $L_r(A)$ 、 $B_r(A)$

和 $N_r(A)$ 与其本身是粗相等关系，即

$$R_{r9}(A, A) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

记作 1，当两个或两个以上的粗空间实体相交时，就有 $R_{r9}(A_i, A_i) = 1$ 。当研究两个或两个以上的粗实体 $A_1, A_2, \dots, A_i, \dots, A_n (n \geq 2)$ 的空间关系时，如果存在所有粗实体的交集，那么可以根据式(2)导出它们相交粗域的通用矩阵 $B_r(A_1, A_2, \dots, A_i, \dots, A_n)$ ，如式(3)：

$$R_{r9}(A_1, A_2, \dots, A_n) =$$

$$\begin{pmatrix} 1 & R_{r9}(A_1, A_2) & \cdots & R_{r9}(A_1, A_n) \\ R_{r9}(A_2, A_1) & 1 & \cdots & R_{r9}(A_2, A_n) \\ \cdots & \cdots & \cdots & \cdots \\ R_{r9}(A_n, A_1) & R_{r9}(A_n, A_2) & \cdots & 1 \end{pmatrix} \quad (3)$$

其次, 分析得到基于影像灰度值和空间地物的逻辑关系的粗隶属函数。概率型粗集对应于粗隶属函数, 粗集可以由粗隶属函数 $\mu_X(x)$ 定义^[7,8]。此处, $\mu_X(x) \in [0, 1]$, 见式(4)。包含 x 的粗元 $R_e(x)$ 对空间实体 X 的粗隶属度, 可以被看作 $R_e(x) \subseteq X$ 的概率, 表示 $x \in X$ 程度的条件概率, 而且 $\mu_X(x) + \mu_{\bar{X}}(x) = 1$ 。如果给定 $P(X|[x]_{R_e}) = \mu_X(x)$, 且置信度 $\alpha \in [0, 1]$, 那么还能够得到基于 α 的概率型的一对上、下近似集定义, 如式(5)。在这种意义下, $\mu_X(x)$ 导致了概率型地学粗空间定义的产生。

$$\mu_X(x) = \frac{R_{\text{card}}(X \cap [x]_{R_e})}{R_{\text{card}}([x]_{R_e})} = \begin{cases} 1, & x \in L_r(x) \\ (0, 1), & x \in B_r(x) \\ 0, & x \in N_r(x) \\ 1 - \mu_{\bar{X}}(x), & x \in X \end{cases} \quad (4)$$

$$\begin{aligned} L_{r\alpha}(X) &= \{x \mid P(X|[x]_{R_e}) \geq 1 - \alpha\} \\ U_{r\alpha}(X) &= \{x \mid P(X|[x]_{R_e}) > \alpha\} \end{aligned} \quad (5)$$

最后, 在决策分析时, 消除多余属性是必要的。一个地学实体由其属性刻画, 属性可分为决策属性 D 和条件属性 C 。假设 $C = (C_1, C_2, \dots, C_k, \dots, C_m)$, 其值为 $V = (V_1, V_2, \dots, V_k, \dots, V_m)$, $D = V_d$, 那么基于地学实体的决策通常以规则(表明 C 和 D 之间的关联程度)的形式给出, 如规则 1、规则 2。

$$\text{规则 1 } (C_1 = V_1) \wedge (C_2 = V_2) \wedge \cdots \wedge (C_k = V_k) \wedge (C_m = V_m) \Rightarrow (D = V_d)$$

当把条件属性 C_k 消除后, 规则 1 变为了规则 2。

$$\text{规则 2 } (C_1 = V_1) \wedge (C_2 = V_2) \wedge \cdots \wedge (C_m = V_m) \Rightarrow (D = V_d)$$

如果“规则 1= 规则 2”成立, 那么条件属性 C 中的 C_k 对于决策属性 D 而言是多余的, 可以剔除。任何多余属性的剔除, 并不影响原始属性的独立关系和最后决策的结果, 却可节省存储, 提高决策的效率。

2 地学粗空间在地球空间信息学中的应用

在现实世界中, 点有大小, 线有粗细, 面有厚薄; 而在经典数学理论中却截然相反, 它常常假设地学实体的属性是连续平滑的, 能够使用一个抽象的平滑数学函数描述, 以致于在使用抽象的数学理论研究具体的世界现实时, 模型实体和世界现实中的实体不一, 丢失了很多信息。因此, 在使用抽象的数学模型研究复杂的实际目标时, 不确定性是难以避免的, 理想的方法是按照实体的本来面目研究实体。

地学粗空间基本保持了实体的本来面目。观测只是对真值的逼近, 难以得到完备信息。当一个地学实体被观测多次后, 观测值就围绕其真值形成一个延伸, 构成兼顾空间的位置和属性不确定性的粗实体。导致延伸的信息不完备性, 可能来自设备、人或数学函数。当基于计算机处理的 GIS 模型表达现实世界的实体 X 时, 误差或某些信息的难以获知性使 $L_r(X) \neq U_r(X)$ 。当粗实体的上、下近似集相等时, 粗实体就转变为确定实体。

在基于 GIS 的应用中, 地学粗空间根据 GIS 已有信息对空间实体的支持程度, 把论域划分为肯定支持域(下近似集)、肯定不支持域(非集)和可能支持域(边界集)。通过简化条件属性, 地学粗空间还能够将 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。因此, 地学粗空间的数学假设承认世界的复杂性、多变性和不确定性, 以及空间信息采集的不完备性。它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, 而是保持现实世界的原有特性, 按照其固有的存在形式, 利用一对上、下近似集直接刻画给定实体, 使用粗空间关系和粗算子搜寻现实世界的内在规律, 以最大可能地逼近地学实体的真实存在状态。例如地学基本元素点、线和面在矢量粗空间和栅格粗空间中的表示(图 4)具有形状和大小。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, 其中的元素可能包含真值, 也有观测值对真值的偏离。这样, 现实世界的实体属性的本来面目就被保持、传播到决策分析的结果之前。虽然增加了存储量, 但是在决策之前, 信息丰富总比信息不足更利于保证决策结果的可靠性。而且地学粗空间还有在最后决策阶

段用于消除多余属性的粗算子作保障,同时计算机的存储容量也已经发展到足够支持地学粗空间的阶段。

地学粗空间不同于已有的地学理论和方法。在可用于研究地学空间信息的理论和方法中,目标模型是基于目标的空间数据模型,重在用传统的测量平差理论讨论空间位置;域模型顾及了实体属性的不确定性,描述现实实体优于目标模型;灵敏度分析是理论上的误差传播分析;GIS 可视化借助计算机技术,把抽象的空间和非空间的属性数据转换为人的视觉可以直接感受的具体图形图像;概率论和空间统计学研究地学问题的随机性,其误差矩阵却不能描述实体不确定性的空间分布;概率矢量能够表示不确定性的空间分布,其 S 带模型兼容了实体的位置和属性的不确定性;证据理论则是概率论的一个扩展。它们都是基于经典的确定集合理论,不能用于研究难以精确描

述的实体或其空间属性。确定集合被扩展后,模糊数学重在模糊性;云理论兼重模糊性和随机性;地学粗空间重在研究地学实体的信息不完备性,基础为一对上、下近似集。它可以解释已有的一般属性不确定性理论或模型^[1]。在地学粗空间中,空间论域 U 由带有属性的地理实体构成,不确定属性是确定属性的外延。 U 被划分为“是”、“非”和“可能”三个不同的置信域。当“可能”域为空时,地学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。目标模型的 $L_r(X) = U_r(X)$, 域、误差带、 ϵ 带和 S 带等模型的 $L_r(X) \neq U_r(X)$ 。地学粗空间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。可以看出,很多模型可视为地学粗空间的一种特殊情况而统一其中。



图 4 地学粗空间中的点、线、面

Fig. 4 Rough Spatial Point, Line and Area in the Geo-rough Space

3 基于地学粗空间的河流专题遥感影像分类

遥感影像分类是研究生态环境的基本技术之一,水域是保持生态平衡的重要因素。因此,本文以基于地学粗空间的河流专题遥感影像分类为例,检验地学粗空间的实用性。原始影像采用的是某地区的一幅 TM 遥感影像(图 5(a))。影响决策属性影像分类的条件属性很多,包括影像灰度、卫星参数、大气折射等,这些条件属性的影响作用,已经在实验前被首先予以修正消除。本文根据影像灰度值来提取影像中的河流类,并使用粗拓扑关系和粗算子处理河流类与影像中其他相邻类别的空间关系,采用的符号是简化统一后的粗符号系统。论域 U 为影像, X 为河流,因河流影像在整幅影像之中,故 $X \subseteq U$ 。假设 G_x 是像素 x 的灰度值, G_X 是河流像素的灰度值,即 $R_{\text{card}}(X \cap [x]_R) = G_x$, $R_{\text{card}}([x]_R) = G_X$, 那么粗隶

属函数可通过式(6)计算:

$$\mu_X(x) = \frac{G_x}{G_X} = \begin{cases} 1 & x \in L_r(X) \\ (0, 1] & x \in U_r(X) \\ 0 & x \in N_r(X) \end{cases} \quad (6)$$

最后结果如图 5(b)、图 5(c)所示。在图中,下近似集 $L_r(X)$ 是确定的河流区域,是最小可能流域,可以解释为枯水季节的最小河道。这时可能拓宽了河流两岸生物圈内生物的数量和活动范围,却增大了它们的用水难度。而上近似集 $U_r(X)$ 是不确定的河流区域,是最大可能流域,可以解释为河流汛期的最大淹没区域,对生态环境的破坏最大。河流粗分类的粗度为 $R_d(X) = R_{\text{card}}(B_r(X)) / R_{\text{card}}(X) \times 100\% = 10.37\%$, 说明不可分辨性或不确定性较小。

为了检验地学粗空间的提取结果,我们又对同一幅影像进行了最大似然分类、模糊分类和实地踏勘。相对于地学粗空间,最大似然分类的结果只是“是”或“非”的二值分类,模糊分类只是介

于0和1的隶属度分类,二者都仅仅得到一种分类结果,远没有地学粗空间的分类结果信息丰富。实地踏勘量测的结果也表明,地学粗空间的影像分类较为逼近该河流的实际水文特征,上近似集 $U_r(X)$ 、下近似集 $L_r(X)$ 分别和夏季、冬季的河流相似。同时,地学粗空间的分类精度也提高了7%(较最大似然分类)或2%(较模糊分类)。上

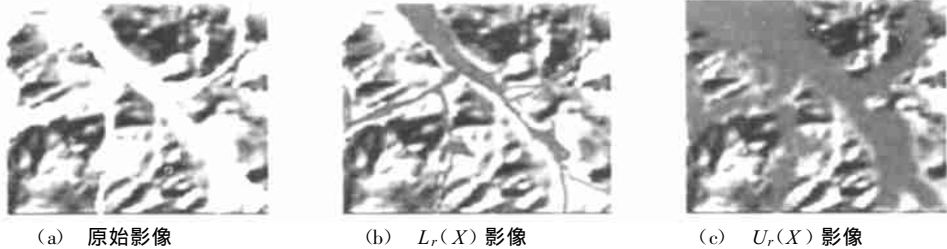


图5 粗河流专题图(连续体)

Fig. 5 Rough Water Thematic Maps(Continuums)

4 结论

1) 地学粗空间兼顾了位置和属性的不确定性,其测度是经典测度的继承和发展。地学粗空间的粗元,在二维空间中类似于遥感影像的栅格,在多维空间内则是如同多维积木的单位体。

2) 在 Pawlak 粗集符号及其变种的基础上,地学粗空间尝试统一简化了目前存在的粗符号。简化后的每个符号含有粗集意义和应用专业领域意义两部分。因粗集总是对下近似集计算和存储两次,故为了节省计算资源,在地学粗空间中实际参加计算的为一对下近似集和边界集。

3) 研究现实世界的理想方法是能够保持实体的本来面目。地学粗空间的数学假设承认世界的不确定性,以及信息采集的不完备性。它使用一对上、下近似集来保持和传播不确定性直到最后的决策,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,其中的元素可能为真值。地学粗空间基本保持了实体的本来面目,为研究地学实体提供了一种可供选择的新数学工具。它虽然不同于已有的地学理论和方法,但是已有的很多理论或模型可以为其解释和统一,对影像分辨率和观测值的不确定性等也有新的数学解释。应用地学粗空间从遥感影像中提取河流的实例表明,地学粗空间可以最大可能地逼近地学实体的存在形式,得到较为可靠的结果。

当然,本文只是提出和研究了地学粗空间的基本框架。粗关系仅仅分析了粗拓扑关系,粗算

述证明,地学粗空间的理论和方法在空间遥感影像的河流资源粗提取时,结果同时具备最大可能流域(上近似集)、最小可能流域(下近似集)和可能误差等多种描述方式。利用地学粗空间研究资源环境,兼顾了位置和属性的不确定性,分类结果同时含有确定性和不确定性的信息,比基于确切集合的结果更为丰富,可以最大可能地逼近地学实体的存在形式。

子也仅仅讨论了基于交集的粗关系矩阵和基于影像灰度值的粗隶属函数,还有很多的粗空间关系、粗逻辑算子、粗代数算子和粗语气算子需要研究、证明和应用。此外,应用实例也只是一幅给定时期的影像,如果把不同时期的同一地区的影像放在一起研究,可能更能显示地学粗空间的优越性。

参 考 文 献

- 1 史文中,王树良.GIS数据之属性不确定性研究.中国图像图形学报,2001,6(A)(9):918~924
- 2 Pawlak Z. Rough Sets. Norwell: Kluwer Academic Publishers 1991
- 3 Yao Y Y, Wong S K M, Lin T Y. A Review of Rough Set Models; In: Lin Y, Cercone N, eds. Rough Sets and Data Mining Analysis for Imprecise Data. London: Kluwer Academic Publishers 1997. 47~75
- 4 Polkowski L, Skowron A. Rough Sets in Knowledge Discovery 1: Methodologies and Applications. In: Studies in Fuzziness and Soft Computing, Vol. 18. Heidelberg, Physica-Verlag 1998
- 5 Polkowski L, Skowron A. Rough Sets in Knowledge Discovery 2: Applications, Case Studies and Software Systems. In: Studies in Fuzziness and Soft Computing, Vol. 19. Heidelberg: Physica-Verlag 1998
- 6 Pawlak Z. Rough Classification. International Journal of Man-machine Studies 1984, 20: 469~483
- 7 Pawlak Z. Rough Sets and Fuzzy Sets. Fuzzy Sets and System, 1985, 17: 99~102
- 8 Pawlak Z. Rough Sets. In: Lin Y, Cercone N, eds. Rough Sets and Data Mining Analysis for Imprecise Data. London: Kluwer Academic Publishers, 1997. 3~7

- 9 Pawlak Z. Rough Set Elements. In: Polkowski L, Skowron A, eds. Rough Sets in Knowledge Discovery 1: Methodologies and Applications. In: Studies in Fuzziness and Soft Computing, Vol. 18. Heidelberg: Physica-Verlag, 1998. 10~30
- 10 Pawlak Z. Rough Sets, Rough Function and Rough Calculus. In: Pal S, Skowron A, eds. Rough-Fuzzy Hybridization: A New Trend in Decision-Making. Singapore: Springer-Verlag, 1999. 99~108
- 11 Pawlak Z, Polkowski L, Skowron A. Rough Sets and Rough Logic: a KDD Perspective. In: Polkowski L, Tsumoto S, Lin T Y, eds. Rough Methods and Applications: New Developments in Knowledge Discovery in Information Systems. Berlin: Physica-Verlag, 2000
- 12 Burrough P A, Frank A U. Geographic Objects with Indeterminate Boundaries. Basingstoke: Taylor and Francis 1996
- 13 Egenhofer M J. Topological Relations between Regions with Holes. International Journal of Geographical Information Systems 1994, 8(2): 129~142
- 14 Egenhofer M J, Franzosa R D. On the Equivalence of Topological Relations. International Journal of Geographical Information Systems, 1995, 9(2): 133~152
- 15 Chen J, Li C M, Li Z L. A Voronoi-based 9-intersection Model for Spatial Relations. International Journal of Geographical Information Systems 2001, 15(3): 201~220

作者简介: 王树良, 博士生。现主要从事空间数据挖掘和知识发现、智能地理信息系统的研究。代表成果: 空间数据挖掘和知识发现、云理论、GIS 属性数据之不确定性研究等, 发表中英文学术论文 36 篇。

E-mail: slwang@liesmars.wtusm.edu.cn

Theory and Application of Geo-rough Space

WANG Shuliang¹ LI Deren² SHI Wenzhong² WANG Xinzhou³

(1 National Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan, China, 430079)

(2 Department of Land Surveying & Geo-Informatics, HongKong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China)

(3 School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan, China, 430079)

Abstract: In this paper, a geo-rough space was proposed as a new approach to handle spatial entities in data analysis. And its basis was rough set. In the geo-rough space, the set of symbols from rough set are simplified and standardized, which includes both rough interpretation and specialized indication. The boundary set was used to take the place of the upper approximation in a practical computation and decision-making. The geo-rough space was characterized by rough entities, rough relationships and rough algorithms in spatial discourse universe, whose element was rough element with a rough degree to measure the roughness. This paper discussed it in Geomatics, paying more attention to the computerized GIS. The result of the case study on thematic image extraction indicated that the geo-rough space was confidential and practical.

Key words: geo-rough space; rough entities; rough relationships; rough algorithms; GIS

About the author: WANG Shuliang, Ph. D candidate. He is concentrated on spatial data mining and knowledge discovery (SDMKD), intelligent GIS and so on. His research interests are in the areas of theories and methods for SDMKD, cloud theory, attribute uncertainties in GIS data. He has published 36 papers in Chinese and English.

E-mail: slwang@liesmars.wtusm.edu.cn