

文章编号:1671-8860(2011)05-0588-05

文献标志码:A

# GIS信息综合中一种改进的粗集选取方法

李雯静<sup>1</sup> 邱佳<sup>1</sup> 吕垒<sup>1</sup>

(1 武汉科技大学资源与环境工程学院,武汉市和平大道947号,430081)

**摘要:**在现有的基于粗集的GIS信息综合的基础上,针对GIS信息表的特殊性,提出了先离散化再添加决策属性的粗集计算思路,采取动态评价目标重要性并选取的方法对目标进行综合。实验表明,该离散化思路可以很好地解决GIS信息表的离散化效率问题,动态评价选取的方法不仅能综合考虑目标的空间及属性信息,而且可以保持目标综合前后空间特征的一致性。

**关键词:**GIS信息综合;粗集;动态选取;离散化

中图法分类号:P208

选取问题是GIS信息综合的基本问题之一,关于定量选取的研究已有了很多成果,但定性分析仍是选取研究中的一个难点。目前对于选取问题,已有许多地图学和GIS方面的学者进行了研究<sup>[1-6]</sup>。文献[7, 8]所提出的基于粗集的选取方法虽然可以很好地兼顾目标的空间信息及属性信息,但仍然存在以下问题:①在描述与相关要素间的协同关系信息时,没有明确被综合要素的主体地位。②对构建的GIS属性决策表进行离散化的过程中,由于一般情况下没有特定的决策属性,采取的是假定决策表中每个对象的决策属性均不相同的方法。贪心算法<sup>[9]</sup>虽然能得到合理的离散化结果,但需反复构建二维表以确定当前区分能力最强的断点,时间复杂度及空间复杂度较高,当空间实体较多时,计算甚至无法进行。③在对空间目标进行排序并选取的过程中,一般是结合数量化指标一次性舍去重要性程度低的目标(本文称之为静态评价目标重要性并选取的方法)。然而实际上空间中各要素不是孤立存在的,因此,若静态地评价点的重要性并选取,可能会使点群分布较密集区域的点被整体去掉。

针对上述不足,本文将基于粗集的选取方法扩展到不同要素多目标协同的方式,对传统方法进行了改进,并提出了先离散化再添加决策属性及动态目标重要性评价、目标选取的方法。

## 1 多目标协同空间/属性一体化模型

目前采用的GIS信息综合粗集方法大多只考虑了空间结构信息,对于属性信息考虑得较少,或者所考虑的那些属性信息并没有对目标的重要性产生一定的影响,本文选择的属性信息不仅来自被综合要素本身,而且还来自与其产生协同作用关系的相关要素。

多目标协同空间/属性一体化模型的总体框架如图1所示。首先,挖掘被综合要素的空间拓扑信息及几何信息,利用属性化的方法对其进行表达;其次,找出对目标重要性产生影响的属性信息并量化表示;最后对产生协同关联的要素间的信息进行提取。如本文对居民地选取的过程中,考虑与其产生协同的道路对其产生的影响。

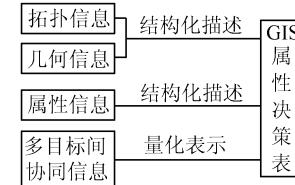


图1 多目标协同空间/属性一体化模型

Fig. 1 Multi-elemental Cooperative and Spatial/Attribute Integration Model

## 2 GIS 数据离散化方法的改进

在许多情况下, GIS 属性决策表中没有哪个属性项可以明显地作为决策属性, 或者数据的应用目标还不明确, 没有针对任何一个专题进行分级分类的目标。文献[7]中先假定每个目标的决策都不相同, 再对属性决策表中连续型字段进行离散化, 并确保决策表中无不相容数据。本文提出了一种先离散化, 再添加决策属性的方法。具体操作过程如下: 先对各连续型条件属性进行无监督离散化, 如可以利用人为参与的等距离划分法或等频率划分法进行离散化, 再根据各属性项的取值异同对各实体添加决策属性, 若各条件属性项取值均相同, 则添加相同的决策属性; 若相异, 则添加不同的决策属性。该方法不仅可以兼顾时间效率, 并且可以确保属性决策表中无不相容数据。其实质是对每个目标进行分类, 当目标之间具有相近的空间特征及属性信息, 则认为这些目标隶属于同一类。按此思路, 利用等距离划分法对原始属性值进行离散化。从表 1 可以看出, 离散化结果不仅很好地保留了条件属性的区分能力, 且表中无不相容信息。

## 3 动态评价目标重要性并选取的方法

空间中各目标不是孤立存在的, 要素内部与要素之间的协同关系会随着综合的过程实时发生变化, 因此, 要充分考虑目标重要性的变化问题, 使得选取的结果更加合理。基于这一思路, 本文采取了动态评价目标重要性的方法对目标进行选取。

表 1 原始属性表和等距离离散化结果

Tab. 1 Original Attribute Table and Equal Interval Discretization

ID	原始属性值/等距离离散化结果				
	A	B	C	D	E
1	10/4	0.066 9/1	14.582 5/2	0.088 1/1	1/7
2	11/4	0.050 4/1	33.475 1/4	0.060 3/1	2/10
3	11/4	0.052 0/1	24.485 3/3	0.741 8/4	3/9
4	9/3	0.169 5/2	1.341 8/1	0.083 8/1	4/5
5	10/4	0.089 4/1	12.881 9/2	0.234 2/2	5/8
6	8/2	0.310 1/4	0.443 5/1	0.064 2/1	6/3
7	9/3	0.101 7/1	5.529 3/1	0.175 5/1	7/4
8	9/3	0.210 7/3	0.126 2/1	0.078 4/1	8/6
9	6/1	0.096 8/1	17.623 8/3	0.070 7/1	9/1
10	7/1	0.157 9/2	1.461 7/1	0.591 3/4	10/2

### 3.1 相关定义

相邻点。如图 2 所示, 对于由离散点群生成的 Delaunay 三角网, 空间点 P 的相邻点为以点 P 为顶点的三角形的其他顶点的集合, 如点 P 的相邻点为 A、B、C、D、E。

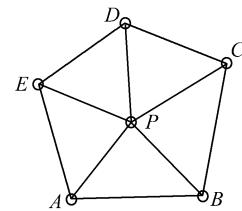


图 2 相邻点

Fig. 2 Adjacent Points

综合阈值。根据开方根规律式(1)计算出综合后保留的目标数, 按式(2)定义综合阈值  $\alpha$ 。可见,  $\alpha$  是一个动态的变量, 随着综合过程的进行而改变:

$$N_2 = N_1 (M_1/M_2)^{1/2} \quad (1)$$

$$\alpha = (N_1 - N_2 - n)/(N_1 - n) \quad (2)$$

式中,  $N_1$ 、 $N_2$  分别为综合前后的目标个数;  $M_1$ 、 $M_2$  分别为综合前后的比例尺分母;  $\alpha$  为综合阈值;  $n$  为当前被综合的目标数。

### 3.2 目标动态评价与选取

目前, GIS 信息综合中进行排序时往往采用静态排序的方法, 即根据参加排序的影响因子的重要性程度和取值一次性计算出所有目标的排列顺序。这种方法适合在空间目标数量不变的情况下对目标的重要性进行判断, 但在对空间信息进行综合时, 当采取逐个删除重要性程度最小的目标的手段进行综合时, 则空间目标数量会随着综合的进程而不断减少。由于空间中目标不是孤立存在的, 一个目标的状态变化, 如目标被舍去, 很可能会对相邻的目标及其相关的各种空间特征产生影响, 包括目标所控制的面积、目标所处位置的空间密度等。对于存在层次或等级关系的空间目标(如树状河系), 一个目标的取舍对这种等级或层次关系同样会产生影响。因此, 空间目标之间的各种关系随着目标状态的变化实时地发生变化。基于此, 本文采用目标动态排序的方法, 在判断目标重要性的过程中, 充分考虑空间目标动态变化的过程。首先给地图中的空间目标定义三种状态, 包括未被标记、已舍去、暂不舍去。该动态评价目标重要性并选取的过程实质上是个一边舍去、一边标记的过程, 如图 3 所示。

图 3 中,  $n$  为计数器, 表示当前舍去的点数,  $k$

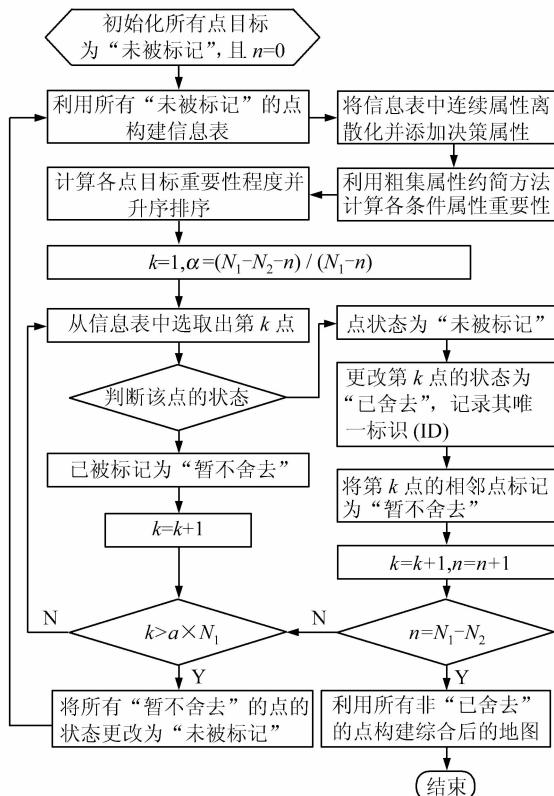


图3 目标动态评价并选取流程

Fig. 3 Dynamic Assessment then Selection of Points

为静态升序排序后属性决策表中点的次序。当删除某一点时,强制性地暂时保留其相邻点,这是动态评价目标重要性的关键所在,它可以有效地避免处在密度较高区域的点被整体删除的情况,并且这些点只是暂时保留,在下一次构建信息表的循环中,还会参与排序。每次搜索待删除点目标的次序严格限制在已按升序排序的属性决策表的前 $\alpha \times N_1$ 部分内,其目的是避免静态排序的重要性在综合阈值外的点目标被过早地删除。

## 4 实验

以图4所示的城镇居民地的分布为实验对象,按本文所提出的改进方法对代表居民地的点群进行选取。首先对点目标空间特征进行结构化描述,分别利用点群凸壳嵌套层次、点的Voronoi图面积等方法描述点群的轮廓形状、点所控制的范围,并在Voronoi图面积的基础上计算点的一阶邻近密度,以描述点群的分布密度关系。对初始点群按上述操作,如图5、图6所示。

此外,影响居民地重要因素不仅有其本身的空间分布特征,还与其所处的位置相关,本文采用居民地与市内主干道的距离来反映道路对居民地重要性的影响,如图7所示。



图4 综合前的点群

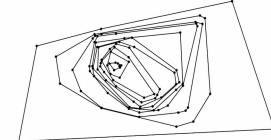
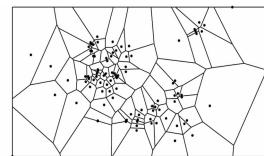
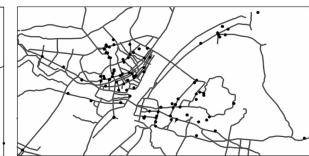
图5 点群凸壳嵌套图  
Fig. 5 Convex Hull of Points

图6 点群的 Voronoi 图

图7 点群与道路协同关系  
Fig. 7 Cooperative Relationship Between Points

完成空间特征属性化描述以后,再对目标自身的属性信息进行提取。居民地的等级可以从建筑物的装潢、设备、设施条件和维修保养等情况的数据来反映,它是评价居民地重要性的一个重要指标。居民地的人口数可以反映居民地的规模,同样是影响居民地重要性的指标,该数据可以由建筑物所含有的房间套数作为估计值。本文选取这两个指标作为评价居民地重要性的属性信息。综合以上信息,构建无决策属性的GIS属性决策表,表2列出了其部分记录。对表2按本文提出的先离散化再添加决策属性的方法进行离散化,根据属性重要性的计算方法<sup>[10]</sup>,计算出各条件属性的重要性(凸壳层次、一阶邻近密度、Voronoi图面积、与道路距离、居民地等级、房间数量的重要性)。

表2 某市居民地属性决策表和离散化后的属性决策表

Tab. 2 Decision-Making of Residential Area and Discretized Decision-Table

ID	原始属性决策表/离散化后的属性决策表						
	凸壳 层次	一阶邻近 密度	Voronoi图 面积	与主干道 距离	等 级	房 间 数 量	决 策 属 性
1	10/2	0.066 9/4	14.582 5/2	0.088 1/4	3	281/3	51
2	11/2	0.050 4/4	33.475 1/3	0.060 3/4	0	63/1	54
3	11/2	0.052 0/4	24.485 3/2	0.741 8/2	0	64/1	52
4	9/2	0.169 5/4	1.341 8/1	0.083 8/4	3	86/1	35
5	10/2	0.089 4/4	12.881 9/1	0.234 2/4	3	326/3	36
6	8/2	0.310 1/4	0.443 5/1	0.064 2/4	0	52/1	33
7	9/2	0.101 7/4	5.529 3/1	0.175 5/4	0	34/1	33
...	...	...	...	...	...	...	...
94	10/2	0.044 8/4	28.262 6/3	0.154 6/4	4	200/2	56
95	10/2	0.056 9/4	18.403 6/2	0.262 9/4	0	80/1	50
96	11/2	0.028 4/4	47.466 7/4	0.142 7/4	3	162/2	57

要性分别为 0.333 3、0.510 4、0.343 8、0.572 9、0.635 4、0.218 8), 再按式(3)进行重要性排序, 结果为 8.250 00、6.250 00、4.760 42、7.468 75、7.906 25、5.562 50、5.562 50、…、9.010 42、5.906 25、8.718 75。

$$W_i = \sum_{k=1}^m (c_{ik} \times v_k) \quad (3)$$

式中,  $i$  为点序号;  $W_i$  为第  $i$  点的重要性;  $m$  为条件属性项数目;  $k$  为条件属性项序号;  $c_{ik}$  为离散化后第  $i$  点第  $k$  个条件属性项的取值;  $v_k$  为第  $k$  个条件属性项的重要性。

在离散化过程中, 各条件属性应离散化为大致相等的区间数, 因为若某条件属性离散化的区间数明显大于其他条件属性, 即使该条件属性的重要性程度不高, 但与该条件属性项的取值相乘后, 会较大地影响点的重要性。本实验将连续型条件属性按等距离离散化算法离散为 4 个区间, 由于凸壳层次较多, 这里将其看作连续型属性, 合并为两个区间。

依据前述的动态评价目标重要性并选取的方法对点群进行选取, 将比例尺综合到原比例尺的  $1/2$ , 保留点数为 68, 经过 5 次构建属性决策表得到最终的选取结果, 如图 8(a) 所示; 将比例尺综合到原比例尺的  $1/4$ , 保留点数为 48, 经过 5 次构建属性决策表得到综合的结果, 如图 8(b) 所示。可以看到, 动态评价目标重要性并选取的结果比静态一次性选取所得的结果更合理, 主要表现在保持点群的空间特征上。动态评价方法得到的选取结果图可以很好地保持原来点群的分布轮廓, 符合空间形态在综合前后一致性的要求。另外, 点群的分布密度关系没有发生大的改变, 原来密度大的地方, 选取后还是大, 反之亦然。

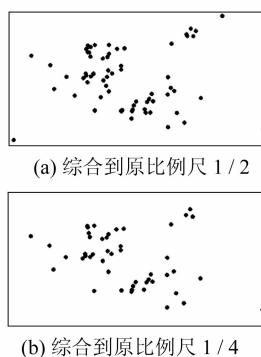


图 8 比例尺综合到原比例尺的  $1/2$  和  $1/4$

Fig. 8 Generalize Scale to  $1/2$  and  $1/4$  of Original

## 5 结语

本文所提出的动态评价目标重要性并选取的思想基于原有的粗集选取方法, 不仅能综合地考虑到目标的空间/属性信息, 能顾及到多要素之间的协同关系及相互影响, 而且可以很好地平衡目标内部相互协调的关系。GIS 信息综合是一个复杂而重要的问题, 本文对现有方法进行了一定的改进, 取得了较好的结果, 但如何进一步挖掘多要素之间的协同关系, 找出对被综合要素产生影响的信息, 以及在属性决策表的离散化过程中, 怎样根据信息表的自身特征找到合理的离散区间数而无需人为干涉等问题需要进一步研究。

## 参 考 文 献

- [1] 艾廷华, 刘耀林. 保持空间分布特征的群点化简方法[J]. 测绘学报, 2002, 31(2): 175-181
- [2] 邓红艳, 武芳, 钱海忠, 等. 基于遗传算法的点群目标选取模型[J]. 中国图像图形学报, 2003, 8(8): 970-976
- [3] 钱海忠, 武芳, 邓红艳. 基于 CIRCLE 特征变换的点群选取算法[J]. 测绘科学, 2005, 30(3): 83-85
- [4] 蔡永香, 郭庆胜. 基于 Kohonen 网络的点群综合研究[J]. 武汉大学学报·信息科学版, 2007, 32(7): 626-629
- [5] 高三营, 闫浩文, 陈静静, 等. 基于圆增长特征的点状要素群选取算法[J]. 测绘工程, 2008, 17(6): 20-23
- [6] 闫浩文, 王家耀. 基于 Voronoi 图的点群目标普适综合算法[J]. 中国图像图形学报, 2005, 10(5): 633-636
- [7] 李雯静, 龙毅, 林志勇. 粗集分类思想在 GIS 点群综合中的应用[J]. 武汉大学学报·信息科学版, 2008, 33(9): 896-899
- [8] 宋鹰, 何宗宜, 栗卫民. 基于 Rough 集的居民地属性知识约简与结构化选取[J]. 武汉大学学报·信息科学版, 2005, 30(4): 329-332
- [9] 王国胤. Rough 集理论与知识获取[M]. 西安: 西安交通大学出版社, 2001
- [10] Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning About Data[M]. Norwell: Kluwer Academic Publishers, 1992

**第一作者简介:** 李雯静, 博士, 副教授, 现主要从事粗集、分形理论在 GIS 与地图综合中的应用研究。

E-mail: wtusm\_lwj@126.com

# An Improvement Selection Method of GIS Information Generalization Based on Rough Set

LI Wenjing<sup>1</sup> QIU Jia<sup>1</sup> LUE Lei<sup>1</sup>

(1 School of Resources and Environmental Engineering, Wuhan University of Science and Technology,  
947 Heping Road, Wuhan 430081, China)

**Abstract:** On the basis of the existing GIS information generalization method based on rough set, a computing idea is put forward in view of the particularity of GIS information table. Firstly, we discretize the continuous attributes by unsupervised discretization. Secondly, we add the decision attribute according to the value of conditional attributes. Thirdly, we calculate the importance of conditional attributes by attribute reduction of rough set. Lastly, we generalize the objects by dynamic assessment. The experimental results show that the procedure can solve the discretization efficiency problem of GIS information table; and generalize objects by dynamic assessment not only take into account the spatial information as well as the attribute information of objects, but also keep the main spatial feature of the objects after generalization.

**Key words:** GIS information generalization; rough set; dynamic selection; discretization

---

**About the first author:** LI Wenjing , Ph. D, associate professor, majors in the application of rough set and fractal geometry to GIS and digital cartography.

E-mail: wtusm\_lwj@126.com

---

(上接第 587 页)

building polygons, and discuss a progressive simplification method based on the hierarchical structure of the concave. In the method, a building polygon is partitioned by constrained Delaunay TIN, and as shape primitives, triangles are used to build the concave triangle tree and to extract triangle sequence. The basic patterns of concaves are recognized by the matching of the typical sequences, which confirm and adopt the corresponding concave simplification methods. On the basis of this, the progressive simplification of complicated concaves of buildings is realized through repeating the process of recognition and simplification. The experimental results show that our method have the features of structure and progressive generalization.

**Key words:** map generalization; constrained Delaunay TIN; building polygon; structure recognition; progressive simplification

---

**About the first author:** CHEN Wenhan, postgraduate, majors in automated map generalization.

E-mail: chenwhleo@gmail.com