

# 空间点集自动概括方法的优化设计

彭 唬<sup>1</sup>, 齐清文<sup>1</sup>, 刘兆礼<sup>2</sup>, 姜莉莉<sup>1</sup>

(1 中国科学院地理科学与资源研究所, 北京 100101; 2 中国科学院东北地理与农业生态研究所, 长春 130012)

**摘要:**在考虑空间点集整体结构的前提下,从系统的观点和人类视觉的角度出发,把空间点集划分成三个子集:边界点子集、聚集中子集和内部点子集。对于边界点子集采用并改进 Delaunay 三角网方法确定边界点以及边界点的取舍;对于聚集中子集,采用模糊聚类分析方法确定中心点子集的组成。最后在确定边界点子集和聚集中子集的前提下,设计内部点子集自动化简的最优化方法。

**关键词:**空间点集;自动概括方法;优化设计

**中图分类号:** P282

## 1 引言

在数字环境下,地球表面的各种地理实体和地理现象都被抽象概括为点、线和面的集合,所以数字环境下制图概括实质上就是对这些抽象了的点、线、面概括研究。点状特征是空间分布分析的一种重要对象;而空间点集表现为复杂目标,空间点的大小、形状已经失去了意义,因此,空间点集自动概括的重点在于空间点的选取,除了参考空间点本身的某些“资格”外,反映空间点集分布和整体结构特征,就成为选取考虑的主要问题<sup>[1]</sup>。总的来说,地理空间的点实体表明那些在空间分辨率上已经没有大小和形状的差别,而仅仅因它自身特性、存在位置及相关情况而成为引人关注的地理事物。对它进行概括时,人们关心的是一些重要性较高的单体点,以及空间点集的整体特征。早期的空间点集概括方法,如:统计学方法、空间占有比方法、重力模型方法、分布相关系数控制模型、圆增长方法等没有或者很少考虑空间点集的整体特征<sup>[2-4]</sup>;后来的凸壳模型、Delaunay 三角网方法等部分地考虑了这一点<sup>[5,6]</sup>。

一般来说,为了使系统达到最优目标所提出的各种求解方法统称为最优化方法,它是完成最优化问题的一种手段。首先把空间点集看作是一个整体(系统),然后把自动综合问题用适当的数学形式表达出来,这样就可以采用最优化数学方法来完成这

种工作,给出一种自动概括优化方案。为了充分保持点集的空间结构,本文对空间点集进行特征拆分,最后结合各部分的特点对空间点集自动概括方法进行优化设计。

## 2 空间点集的几何参数分析

(1)空间分布是从整体的、全局的角度来描述空间点集的特性,空间分布方式决定于空间点在空间的排列特点以及空间点之间的相互关系。空间分布方式决定了空间点集分布边界的走向。空间点集的分布边界是个笼统、概略、模糊的概念,隐含了一个人们对该空间点集认可的平均影响范围或活动半径,由分布边界形成范围内、外的概念。

空间点集不像线状和面状地理特征那样具有实际的边界线,它并没有真实的分布边界。对于不同的观测者来说,可能会得出不同边界范围。空间点集分布边界的模糊性对于计算机自动化处理的定量性来说是不相符的,因此为了有效地实施空间点集的自动化简,首先应该强制性地定义一个描述空间点集的分布边界,然后再以此边界为约束条件进行其他点的化简。

在计算几何中,群点分布的空间结构可以通过空间邻近关系来体现<sup>[7]</sup>。Delaunay 三角网是计算几何中邻近分析的强有力的模型工具。对空间点集生成 Delaunay 三角网,能够把离散的空间点关联起

收稿日期:2004-05-13; 修回日期:2004-12-12.

作者简介:彭唬(1975-),男,博士研究生,主要从事遥感技术应用、制图综合和 GIS 应用,E-mail:pengx@igsnrr.ac.cn

来,实现“点”到“面”的转化。然后依据 Gestalt 原则,即只有两点相距小于视觉邻近距离时,其连接才充当范围边界 $[a]$ ,求算后的 Delaunay 三角网的分布边界就是空间点集的分布边界。

(2) 空间点集分布特征包括两个方面的内容:一是点集的分布形心和分布轴线,即用少量的形心代表整个点集,分布轴线是指空间点集体现一个什么样的线状分布,是轴线、树形线还是环线;二是点集内部的聚集状况。艾廷华在 2002 年提出:只要空间点集的分布范围不变,其分布轴线便不会发生改变;只要空间点集的相对密度得以保持,其分布中心就不会发生偏移<sup>[9]</sup>。为此保持空间点集分布特征的自动概括就简化为保持群点分布范围及相对密度两个约束条件。

空间点集的分布中心并不完全是以空间点集的边界点所围成的多边形的几何中心,有时空间点集可能存在多个分布中心。由于比例的缩小,人眼分辨能力的限制,使得距离小于人眼最小分辨尺寸的空间点集特征聚集成一个面状特征,从视觉上来说,此时的空间点集就可以简单地被看作是几个面状特征的组合。可以认为此面状特征是由一个聚集中心和其周围相邻若干点集积聚而成。所以我们使用几个点勾画出此“面状”特征的聚集中心来代表整体点集的大致分布中心。选取聚集中心点作为整个点集分布特征的骨架结构点。

聚类中心的求算选择了模糊聚类分析方法,模糊聚类分析是将空间点集置于一个多维空间中,按照它们空间关系的亲疏程度进行分类。事实上聚类分析是一种非监督的分类。模糊聚类得到了空间点属于各个类别的不确定性程度,表达了样本类属的中介性,即建立起了样本对于类别的不确定性的描述,能更客观地反映空间点集的实际分布情况,并易于计算机实现。

### 3 空间点集自动概括方法的优化设计

#### 3.1 空间点集分布范围的计算

空间点集分布范围的确定,是利用三角剖分中的 Delaunay 三角网方法来实现的:第一步对所有的空间点进行三角剖分, Delaunay 三角网生成算法很多,可以选用其中的任何一种,关心的只是 Delaunay 的生成以及边界点和边界线的确定。第二步是

边界线的确定,根据 Delaunay 三角网构建的原则及 Delaunay 三角网的性质可以得出如下结论:如果连接空间两点的边有两个三角形来共用,此边一定是三角网内部的边,此边的两个端点也是空间点集的内部点。见图 1, 边  $P_1P_2$  被三角形  $P_1P_2P_3$  和三角形  $P_1P_2P_5$  所共用,可见边  $P_1P_2$  是空间点集的内部边,而点  $P_1$  和点  $P_2$  则是空间点集的内部点;相反,如果连接空间两点的边只属于一个三角形,则此边一定是空间点集的边界线,该边的两个空间点就是空间点集的边界点,图 1 中边  $P_3P_4$  只被三角形  $P_2P_3P_4$  所拥有,所以此边肯定是空间点集的边界线,而点  $P_3$  和点  $P_4$  则是空间点集的边界点。按照上述的判断准则,同时结合“剥皮”算法,设定“剥皮”阈值对三角剖分后的三角网进行判断、取舍、标号,分别建立边界点集结构、边界线集结构和边界三角形结构,以备计算机自动处理使用。第三步就是对空间点集的边界点实施概括。对利用“剥皮”算法所求得的边界点概括时采用局部算法<sup>[9]</sup>,即对任意一个边界点,选取与它左右相邻的两个边界点,由三者作为一个局部小团体参与当前边界点的取舍。分别求算它与相邻两个边界点连线的长度  $L_1$  和  $L_2$ ,并算出二者的比值  $Q$ ; 同时还要算出以当前边界点为顶点,  $L_1$  和  $L_2$  为边的夹角  $\alpha$ 。

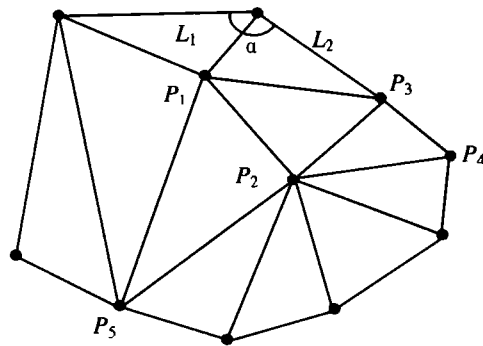


图 1 Delaunay 三角网边界点、边界线

Fig.1 The boundary points and lines in the delaunay triangle

由于空间边界点经过“剥皮”处理后,边界点的连线都比较符合人的视觉感知,同时也考虑到聚集效应,得出如下的取舍方法:如果  $Q$  值远大于 1,则当前的边界点是聚集的边缘点,所以给予保留,反过来,如果接近于 1,则当前三个边界点对整个边界点结构的贡献相差不多,所以舍去当前操作点;再者,考虑夹角  $\alpha$  的缘故是:当前边界点有可能是整

个边界点轮廓的视觉拐弯处,所以设定  $\alpha$  是锐角应给予保留考虑。综合上面的论述,总结边界点概括从以下两个判断条件考虑:

- ①  $Q$  值大于给定的比例阈值;
  - ②  $\alpha$  为钝角或者大于给定夹角阈值;
- 满足上述的边界点舍去。

### 3.2 聚集中心的确定

关于聚集中心的求算,我们采用模糊聚类方法。设空间点集的内部点集合  $X$  中含有  $n$  个样本,表示为  $x_k, k=1, \dots, n$ 。空间点集聚集中心的求算就是要将  $\{x_1, x_2, \dots, x_n\}$  区分为  $X$  中的  $c$  个子集,  $2 \leq c \leq n$ , 要求距离相近的样本应尽量在同一类,  $c$  为聚类数。模糊聚类分析的基本思路为: 设  $X$  中的任意样本  $x_k$  对第  $i$  类的隶属度为  $\mu_{ik}$ , 式中  $0 \leq \mu_{ik} \leq 1$ 。所以该分类结果可以用一个  $n \times c$  阶矩阵  $U$  来表示, 该矩阵称为模糊矩阵。为了在众多可能的分类中寻求合理的分类结果, 为此, 就要确定合理的聚类准则。定义目标函数  $J(U, V)$  为

$$J(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m (d_{ik})^2 \quad (1)$$

式中:  $U = u_{ik}$  为模糊分类矩阵,  $i = 1, 2, \dots, c; k =$

$1, 2, \dots, n$ , 且满足  $u_{ik} \in [0, 1]$ , 和  $\sum_{i=1}^c u_{ik} = 1 \quad \forall k, V = \{v_1, v_2, \dots, v_c\}$  为  $c$  个聚类中心集合,  $v \in R^p, m \in [1, \infty]$  为加权指数,  $d_{ik}$  为第  $k$  个样本到第  $i$  类的距离。

$J(U, V)$  表示为各类中空间点到聚集中心的加权距离平方和, 权重是样本  $x_k$  对第  $i$  类隶属度的  $m$  次方, 聚类准则取为求  $J(U, V)$  的极小值:  $\min \{J(U, V)\}$ 。由于矩阵  $U$  中各列都是独立的, 因此有:

$$\begin{aligned} \min \{J(U, V)\} &= \min \left\{ \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m (d_{ik})^2 \right\} \\ &= \sum_{k=1}^n \left[ \min \left\{ \sum_{i=1}^c (u_{ik})^m (d_{ik})^2 \right\} \right] \quad (2) \end{aligned}$$

上述极值的约束条件为等式  $\sum_{i=1}^c u_{ik} = 1$ , 可用拉格朗日乘法进行求解。则使得  $J(U, V)$  为最小的  $u_{ik}$  值和  $V_i$  值分别为(3)式所示:

若数据集  $X$ , 聚类类别数  $c$  和权重  $m$  已知, 就能用式(1)、式(2)和式(3)确定最佳分类矩阵和聚类中心。此时的聚类类别数  $c$  和权重  $m$  可以根据经验或者通过缩放原始数据给定。

$$\mu_{ik} = 1 / \sum_{j=1}^c \left( \frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}}, \quad v_i = \frac{1}{\sum_{k=1}^n (\mu_{ik})^m} \sum_{k=1}^n (\mu_{ik})^m x_k \quad (3)$$

### 3.3 空间点集内部空间点自动概括的优化设计

在很多实际问题中, 常常要选用一组“设计参数”, 在一定条件下使那些不希望出现的“效应”尽可能的小, 这就是最优化的数学方法<sup>[9]</sup>。最优法求解实际问题时, 首先需要把实际问题用适当的数学方法表达出来。要明确哪些是变量, 哪些是已给常数。由于空间点集自动制图概括方案有多种可能, 故可以采用最优化的数学方法来完成这种取舍工作, 给出一种最优设计方案。下面给出空间点集内部点的自动取舍的最优设计方法。

首先对空间点集作如下的分析: 从系统论角度出发, 把空间点集看成一个系统, 系统的外力由边界点产生, 系统的内力由聚集中心产生, 整个空间点集就在内外力的综合作用下稳定存在。然后根据著名的开方根定律确定比例尺缩小后应该保留的空间点的数目, 开方根定律的基本公式如下:

$$n_f = n_a \sqrt{\frac{M_a}{M_f}}$$

其中表示待求小比例尺地图上地物的数量,  $M_a$  表示原始比例尺地图上地物的数量,  $M_f$  表示原始地图的比例尺, 表示小比例尺地图上地物的数量。利用上面的公式, 有  $n_a, M_a, M_f$  三值已知, 所以可以求出小比例尺的空间点的数量  $n_f$ 。根据上述对边界点的数目和聚集中心的数目以及利用开方根定律求出的小比例尺地图上空间点的数目, 我们就可以确定空间点集中内部点的保留数量。

设空间点集中的边界点集合为  $B = \{B_1, B_2, \dots, B_m\}$ , 空间点集中的聚集中心集合为  $C = \{C_1, C_2, \dots, C_n\}$ 。利用最优方法制定内部点的选取方案最关键的就是确定目标函数, 我们用  $x_{ij}$  表示空间点集中点  $i$  对点  $j$  的作用力, 由于无法确定各点之间的作用力的大小, 所以把  $x_{ij}$  看做是目标函数的变量; 用  $DB_{ij}$  表示边界点对内部点的距离,  $DC_{ij}$  表示聚集中心点对内部点的距离, 用  $SB_{ij}$  表示边界点与内部点的面积比, 用  $SC_{ij}$  表示聚集中心点对内部点的面积比, 这四个参数都可以由空间点属性求算出来, 故此可以看作是常量。这样目标函数的变量与常量就确定下来。其边界点对内部点的作用力为:

$$\sum_{j=1}^m x_{ij} = b_j \quad (j = 1, 2, \dots, k)$$

其中  $k$  是待求内部点的数量, 设空间点集中空间点的总数量为  $Q$ , 则  $k = Q - m - n$ 。空间点集聚集中心对内部点的作用力为:

$$\sum_{i=1}^n x_{ij} = a_j \quad (j = 1, 2, \dots, k)$$

由于边界点对内部点的作用力被看作是外作用力, 所以作用力值看作为负值, 考虑到距离及面积的影响, 所以有外部作用力最终表示为:

$$F_{out} = - \sum_{i=1}^m \sum_{j=1}^k DB_{ij} \cdot SB_{ij} \cdot x_{ij}$$

而聚集中心对内部点的作用力被看作是内作用力, 所以作用力值看作为正值, 最终的内部作用力表示为:

$$F_{in} = \sum_{i=1}^n \sum_{j=1}^k DC_{ij} \cdot SC_{ij} \cdot x_{ij}$$

则边界点与聚集中心点对内部点的作用力就可以表示为:

$$F(x) = F_{in} - F_{out} = \sum_{i=1}^n \sum_{j=1}^k DC_{ij} \cdot SC_{ij} \cdot x_{ij} - \sum_{i=1}^m \sum_{j=1}^k DB_{ij} \cdot SB_{ij} \cdot x_{ij}$$

则内部点取舍的目标函数就表示为:

$$\min |F(x)| = \min |F_{in} - F_{out}|$$

$$= \min \left| \sum_{i=1}^n \sum_{j=1}^k DC_{ij} \cdot SC_{ij} \cdot x_{ij} - \sum_{i=1}^m \sum_{j=1}^k DB_{ij} \cdot SB_{ij} \cdot x_{ij} \right|$$

考虑到内部点的取舍, 则被保留的点才会存在作用力, 而被舍去的点作用力则不会存在, 即作用力为 0, 上述的函数满足下面的约束条件:

(1)  $Res = \{r_1, r_2, \dots, r_k\}$ , 其中有  $Q - m - n$  个元素为 1, 而其他值为 0;

(2)  $x_{ij} \geq 0$ ;

这样, 优化后的空间点集自动概括方法就可以用图 2 表示。

## 4 结语

空间点集的空间结构特征是空间点排列特点以及空间个体之间的相互关系的体现<sup>[9]</sup>。只有充分明确空间结构, 才能真正从整体的角度来保证空间点集制图概括的有效实施。优化设计就是充分考虑了这方面的因素, 首先从空间点集组成结构和空间形状出发, 把空间点集划分为边界点子集、聚集中

心子集和内部点子集几部分; 然后再根据各部分的相关关系设计的一种最优化数学方法。本文只是从边界点、中心点和内部点 3 个方面对空间点集进行了剖析, 对空间点集内部点子集进行概括优化设计时, 并没有单独对内部点的结构进行进一步分析, 尤其是分布轴线, 因此内部点子集空间结构的保持还有待于论证。空间点集自动制图概括方法依旧停留在几何操作的水平上, 在面向地理特征方面并没有太多的突破, 原因一是地理特征的区域性及自身的复杂性; 二是利用几何特征能够部分地体现空间点集的结构特征。因此对于空间点集自动制图概括来说, 如何利用有限的几何操作来实现保持空间结构是今后有待研究的目标。当然最终解决空间点集的自动制图概括应该是面向地理特征的制图概括专家系统<sup>[10]</sup>, 而其前提是自动概括的几何算法应相当完善, 这也是追求完美几何算法最重要的原因。

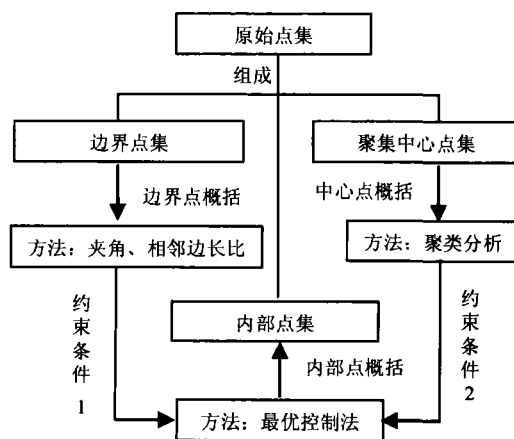


图 2 空间点集自动概括优化流程

Fig.2 Flow chart of optimum generalization of point cluster

## 参考文献

- [1] 王 桥. 地图信息的分形描述与自动综合研究. 武汉: 武汉测绘科技大学出版社, 1998.
- [2] Topfer F T, Pillewizer W. The principles of selection. The Cartographic Journal, 1966, (3): 10~16.
- [3] Langran G E, Poiker T K. Integration of name selection and name placement. In Proc. 2nd Int. Symp. on Spatial Data Handling, 1986, 50~64.
- [4] Kreveld M J van, Oostrum R W van, Snoeyink J. Efficient settlement selection for interactive display. In Proceedings Auto -Carto 13: ACSM/ASPRS Annual Convention Technical Papers. 1997, 287~296.
- [5] 毋河海. 凸壳原理在点群目标综合中的应用. 测绘工程,

- 1997, (3): 1~6.
- [6] 艾廷华, 刘耀林. 保持空间分布特征的群点化简方法. 测绘学报, 2002, 31(2): 175~181.
- [7] F.P. 普霍帕拉塔, M.I. 沙莫斯. 计算几何导论. 北京: 科学出版社, 1992.
- [8] 袁亚湘, 孙文瑜. 最优化方法与理论. 北京: 科学出版社, 1997.
- [9] 郭仁忠. 空间分析(Spatial Andysis). 武汉: 武汉测绘科技大学出版社, 1997.
- [10] 齐清文. 面向地理特征的制图概括. 博士学位论文, 中国科学院地理科学与资源研究所, 1996.

## The Optimum Method for Automated Generalization of Spatial Point Clusters

PENG Hu<sup>1</sup>, QI Qingwen<sup>1</sup>, LIU Zhaoli<sup>2</sup>, JIANG Lili<sup>1</sup>

(1 Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;

2 Northeast Institute of Geography and Agricultural Ecology, CAS, Changchun 130012, China)

**Abstract:** In the field of computer-assisted cartography, cartographic generalization is one of the basic theories and methods of mapping. It is not only an abstract cognition of objective world, but also an important means of spatial information transformation and an important step of mapping. In digital environment, especially in GIS, all kinds of geographic entities and phenomena on the earth's surface are abstracted as points, lines, and polygons. Point cluster is one of the most important objects in the spatial analysis. In GIS, size, direction and shape of spatial points are not vital, however the holistic spatial configuration of spatial point cluster can represent the overall characteristics of point cluster. Thus, the whole spatial configuration of point cluster must be taken into account in automatic cartographic generalization. For the former methods, the reduction of spatial points is settled commendably, but there are still insufficiencies of maintaining the boundary and spatial structure of spatial points. In this paper, the boundary and the clustering center of point clusters are defined in detail. For the point clusters, the delaunay triangle is built and the points on the boundary are located. Then a method to extract the key points on the boundary is set forth. For the center points, the cluster method is quoted to locate them. Finally, the optimum method to improve the maintenance of the structure of spatial point clusters is identified and the content of this method is designed at length.

**Key words:** automated generalization; spatial point cluster; optimum method