

# 基于 Kohonen 网络的点群综合研究

蔡永香<sup>1,2,4</sup> 郭庆胜<sup>1,2,3</sup>

- (1 武汉大学资源与环境科学学院, 武汉市珞喻路 129 号, 430079)  
(2 武汉大学地理信息系统教育部重点实验室, 武汉市珞喻路 129 号, 430079)  
(3 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)  
(4 长江大学地球科学学院, 荆州市南环路 1 号, 434102)

**摘要:**基于 Kohonen 网络模型,对标准的 SOM(self-organizing feature map)算法进行了改进,在保持点群原有空间分布特征的情况下研究点群的选取和典型化。实践表明,该方法适合任意空间分布类型的点群综合。

**关键词:**点群;地图综合;Kohonen 网络;空间分布

**中图法分类号:**P283.1

地图上具有区域景观特征的点群目标很多,如呈点状分布的散列式居民地、密集分布的点状表示的小岛屿群或小湖泊群等。点群目标的综合是制图综合中最基本的问题之一,一般只能选取,不能合并,在减少相似目标的数量时,保持点群的空间分布特征。众多的专家学者提出了一些点群选取方法<sup>[1-7]</sup>。本文基于 SOM 算法对点群综合问题进行了分析和算法改进,该方法能自动保持点群的空间分布特征。

## 1 点群的分布类型及综合原则

本文所研究的点群指点状表示的同类地理要素构成的群体,从视觉上看呈离散分布,但就人的感知来说,这些点构成的群体存在一定的分布范围,呈现一定的分布结构、排列方式,在局部环境中形成了一个相对的整体,在整个研究区域中,也可以是一个“子群”,并且在研究过程中对点的具体性质暂不考虑,即认为点群中的点具有同等重要性,并且无大小之分。点群的分布主要有 3 种:沿线状要素离散分布、面域上离散分布、均匀规则分布。

点群综合的基本原则是<sup>[1,2,8]</sup>:① 轮廓形状保持。随着抽象程度的增加,点群在空间认知上会构成一个虚拟的区域,这个区域的边界就是点群的轮廓。② 密度对比、纹理结构特征保持。点群

的空间分布特征在综合后应当保留,所用的主要指标是密度和纹理。

## 2 SOM 的基本原理

自组织特征映射(SOM)是模拟感官输入在人脑中的皮质映射,将任意维数的大数据集在保留数据样本初始模式的情况下转变为低维映射(通常是一维或者二维),在输入空间彼此接近的样本在低维空间也彼此接近。SOM 网络结构一般为两层:输入层和输出层,输出层的神经元可以被放置在一维、二维或者更高维的网格节点上,输入层和输出层之间实行全互连接。在输入层有  $D$  个输入节点,对应输入样本的  $D$  维属性,用输入向量  $\mathbf{X}=[x_1, x_2, \dots, x_D]^T$  表示,输出层有  $M$  个节点,对应着  $M$  个神经元,通过权值  $w_{ij}$  ( $i=1, 2, \dots, M; j=1, 2, \dots, D$ )和输入节点相连,输出层中的任一神经元  $i$ ,其输入为  $u_i = \sum_{j=1}^D w_{ij} x_j$ ,输出为  $V_i = f(u_i)$ ,其中  $f(\cdot)$  为神经元激励函数,一般取线性函数<sup>[9]</sup>。

SOM 的学习算法由包含竞争、合作和更新三个过程的反复迭代完成。在竞争过程中,具有与输入向量距离最近权值向量的神经元获胜。根据应用的不同,关注的兴趣点可能是获胜神经元的标号(输出层格网中的位置),或者是在 Euclid 距

离意义下输入向量最近的突触权值向量。生物神经元在接受刺激并进行竞争产生获胜神经元后,获胜神经元不但加强自身,而且带动周围的邻近神经元得以相应地加强,同时抑制周围距离较远的神经元,即所谓“on-center/off-surround”现象,这就是合作。获胜神经元位于合作神经元的拓扑邻域的中心,邻域范围内的神经元是兴奋神经元,拓扑邻域函数的典型选择是高斯函数<sup>[9]</sup>。初始化时,包括以获胜神经元为中心的几乎所有神经元随着迭代次数的增加以指数方式衰减。更新的过程也是突触自适应的过程,采用 Hebb 学习规则的改变形式对网络上获胜神经元拓扑邻域内的神经元进行权值向量更新。

学习迭代执行,直到遇到终止阈值,迭代的结果是使获胜神经元  $j$  的突触权值向量  $\mathbf{W}_j$  向输入向量  $\mathbf{X}$  移动。随着训练数据的重复出现,由于邻域更新使得突触权值向量趋于服从输入向量的分布,并具有保持邻近关系不变的特点。

### 3 基于 SOM 的点群综合

一旦 SOM 算法收敛,其所计算的特征映射能显示输入空间的重要统计特性<sup>[9]</sup>:① 输入空间的近似。输出空间的突触权值向量  $\{\mathbf{W}_j\}$  的集合表示的特征映射对输入空间提供了一个良好的近似。② 密度匹配。特征映射反映了输入分布在统计上的变化,输入空间中以高概率抽取的区域在输出空间映射为更大的区域,比以低概率抽取的区域拥有更好的分辨率。表现在突触权值向量空间,高输入密度区域仍是高密度表示,低输入密度区域仍是低密度表示,即能够保持密度对比特征。

SOM 的这两个重要统计特性正好能够满足点群综合中的密度对比、纹理结构特征保持的要求。因此,考虑用 SOM 来进行点群的综合。

点群的综合还需要保持其分布轮廓。SOM 的学习机制决定了当样本空间中多个样本映射到输出空间的同一个单元时,这个神经元的权值向量会在这多个样本向量间移动,最后收敛于一个中间位置,反映在点群的综合中就是点群的典型化可以保持其密度对比、纹理结构特征,但其轮廓范围很难保持,存在“内缩”,即向点群内部移动的现象。

艾廷华等按照 Gestalt 邻近性原则,在点群 Delaunay 三角网上进行“剥皮”操作,获取点群分布范围多边形<sup>[2]</sup>,这种方法能够有效地保持点群的外部轮廓特征。鉴于此,本文通过建立点群的外部轮廓多边形来表达点群的分布范围,轮廓线

采用 Douglas-Peucker 算法化简。Kohonen 特征映射能保持点群密度对比、纹理结构特征,外部轮廓多边形化简能够保持点群的外部轮廓范围,可以将二者结合起来,用于点群的综合。具体方法如下:① 建立原始点群的外部轮廓多边形,分成外部轮廓点集和内部点集。② 外部轮廓点集用 Douglas-Peucker 算法化简。③ 内部点集作为输入样本,用 Kohonen 网络进行特征映射。④ 二者的并集便是综合后的点群,外部轮廓点和内部点的删减数目比例相同。

内部点的化简过程如下。

1) 初始化过程,设置各项变量和参量。

样本数据集:设内部点点集中有  $N$  个点作为输入样本,由于只考虑位置属性,则样本空间维数为 2 维,即输入节点只有两个,为点的  $x$ 、 $y$  坐标。

输出层:输出层的神经元的个数为  $M$ ,由开方根定律计算确定,神经元呈二维网格排列,排列的方式可以任意指定,神经元在网格中的位置决定了其拓扑邻域。

权向量空间:输出层的每个神经元通过权向量与输入层实行全互连接,所有的权向量构成了权向量空间,同样本数据空间一样,也是 2 维空间,  $\mathbf{W}_i = \{w_{i1}, w_{i2}\}, i = 1, 2, \dots, M$ 。

权向量在学习开始时要进行初始化,随机给定输入向量分布范围内的值作为权向量的初始值,或者直接在样本空间选择  $M$  个样本,将其向量值作为初始权值。

另外,根据经验设定初始学习率  $\eta_0$ 、最大迭代次数、权值的变化量平方和阈值、时间常数  $\tau_1$ 、 $\tau_2$  以及初始拓扑邻域  $\sigma_0$  的值。

2) 训练过程如下。

① 初始化。从样本数据集中随机选取  $M$  个样本作为权向量初始值  $\mathbf{W}_j(0), j = 1, \dots, M$ 。

② 取样。随机从样本数据集中取样本  $\mathbf{X}$ ,其坐标值作为输入层节点的输入。

③ 相似性匹配。计算输出层每个神经元权向量与样本  $\mathbf{X}$  的欧氏距离,找出获胜神经元  $c$ 。

④ 更新。动态调整获胜神经元邻域内所有神经元的权向量。

⑤ 继续步骤②,直到满足终止条件为止。

在迭代过程中,如果权值的变化量平方和  $\sum_{i=1}^M \Delta \mathbf{W}_i^2$  小于一定阈值,说明权值已趋于稳定,达到收敛状态,迭代终止,也可直接设定一个最大迭代次数作为终止条件。

映射完成后,输出空间的突触权值向量  $\{\mathbf{W}_j\}$

的集合就是输入空间向量的典型化的综合结果,如果将权值向量用距离最近的样本点代替,就是选取的综合结果。

### 4 点群综合的应用实例

图 1 是利用上述方法进行面域上任意分布点群的综合实例。图 1(a)是原始数据集,共有 51 个点,综合时点数减少至 70%时,保留 36 个点;图 1(b)是利用 Kohonen 特征映射综合的结果和原始点集的比较图,圈是综合后的点,黑色三角形点是原始点。可以看出,虽然密度对比能够保持,但综合后点的分布轮廓存在一定的畸变。

按照前面所述的方法,将原始数据集分成凸壳点集和内部点集分别化简,图 1(c)显示了提取出来的凸壳点集;采用 Douglas-Peucker 算法化简,保留 70%的点,见图 1(d);图 1(e)是内部点典型化的结果与原始数据点集的比较图;图 1(f)是内部点直接选取的结果与原始数据点集的比较图。内部点集的综合结果与凸壳点集的化简结果合并,得到最终的综合结果,图 1(g)是最终的典型化结果图;图 1(h)是最终选取的结果图,二者之间有差别。从视觉上可看出,图 1(g)相对于图 1(b)有很大的改善,轮廓范围和密度对比特征都能很好地保持;图 1(i)是综合后点数减少至 70%的结果图;图 1(j)是综合后点数减少至 40%的结果图;图 1(k)是综合后点数减少至 20%的结果图。可见,综合的结果都能保持原始点集的空间分布特征。

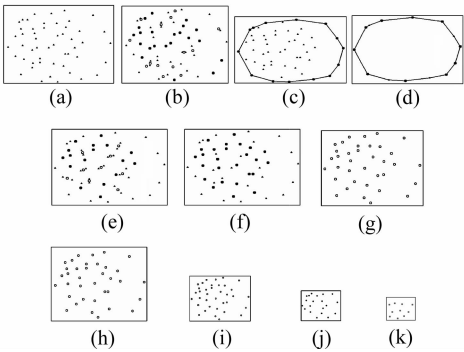


图 1 面域上离散分布的点群综合  
Fig. 1 Generalization of Points Group Scattered in Plane

为了验证综合前后点群的相对密度是否保持,可计算出点综合前后的密度进行对比。保持相对密度是指原来密度大的区域,综合后密度仍然相对较大;原来密度小的区域,综合后密度仍然相对较小。某点的密度用包围该点的 Voronoi 图

元多边形面积的倒数表示<sup>[2]</sup>,点的选取可直接计算点选取前后的密度进行对比,典型化的点综合后的密度可直接计算,综合前的密度用该点所代表的原始区域内点的平均密度代表。图 2 是本例中选取点的综合前后密度对比图(图 1(h)与图 1(a)的对比),共选取点 36 个。从图中可以看出,选取前后的密度高低对比仍能保持,说明本方法能够保持相对密度。

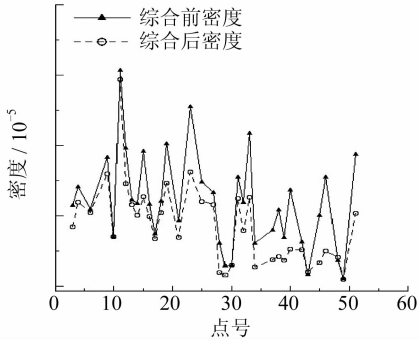


图 2 点的选取前后密度对比图  
Fig. 2 Contrast Diagram Between Original and Reserved Points Density

另外,还选了一个沿特定方向分布的点群进行综合试验。如图 3(a)是原始点集,有 36 个点,点群沿线分布,综合后,只需保持其疏密对比特征,可以直接用 Kohonen 特征映射进行典型化。当综合后的点数保留 70%时,需选取 25 个点;图 3(b)是选取综合的结果,图 3(c)是典型化综合的结果,二者在分布形态上没有大的差别,只是在有些点之间的距离上有微小的偏移;当综合后保留原始点群 45%的点数,即 16 个点时,典型化的结果如图 3(d)所示,图 3(e)是选取的结果。两者对比可以看出,当比例尺进一步缩小时,综合的结果趋于一致。实例表明,综合后的点的分布保持了原始数据的密度对比、分布走向等特征。

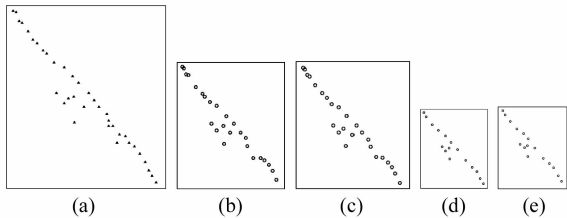


图 3 沿特定方向分布的点群的综合  
Fig. 3 Generalization of Points Group Distributing Along a Special Orientation

图 4 是一个数据量相对较大的点群选取试验。图 4(a)是原始点集,有 213 个点,图 4(b)是选取约 70%点的结果,即保留了 144 个点,图

4(c)为比例尺缩小后的结果。从图中可看出,分布范围仍保持了原来多边形的形状,视觉上的相对密度也得到了较好的保持。

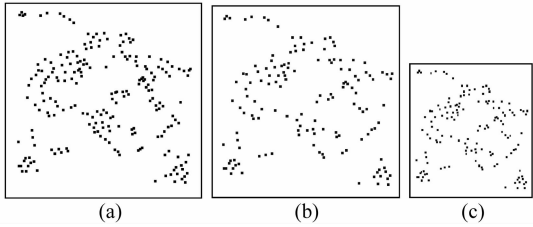


图 4 保持空间分布特征的点群选取  
Fig. 4 Selection of Points Group with Spatial Distribution Remaining

5 结 语

综合实例表明,基于 Kohonen 网络的点群综合化简方法能够保持原始点群的空间分布特征。但本文只考虑了单个点群的综合,并且假定点与点具有同等重要性。当考虑点目标的语义特征以及多个子群同时存在时,就需要考虑点群的聚类,并针对具体问题考虑点目标的地理语义特征及上下文关系,这有待进一步研究。

参 考 文 献

[1] 毋河海. 凸壳原理在点群综合中的应用[J]. 测绘工程, 1997, 6(1):1-6

[2] 艾廷华,刘耀林. 保持空间分布特征的点群化简方法[J]. 测绘学报,2002,31(2): 175-180

[3] Openshaw S, Blake M, Wymer C. Using Neuro-computing Methods to Classify Britain's Residential Areas[OL]. Http://www. geog. leeds. ac. uk/papers/95-1/,1995

[4] Allouche M K, Moulin B. Amalgamation in Cartographic Generalization Using Kohonen's Feature Nets[J]. International Journal of Geographical Information Science, 2005, 19(819): 899-914

[5] Jiang Bin, Lars H. Selection of Streets from a Network Using Self-organizing Maps[J]. Transactions in GIS, 2004,8(3): 335-350

[6] Højholt P. Generalization of Build-up Areas Using Kohonen-networks[C]. Eurocarto XIII, Ispra, Italy, 1995

[7] Sester M. Optimization Approaches for Generalization and Data Abstraction[J]. International Journal of Geographical Information Science, 2005, 19(8/9): 871-897

[8] 郭庆胜. 地图自动综合理论与方法[M]. 北京: 测绘出版社, 2002: 131-135

[9] Haykin S S. Neural Networks a Comprehensive Foundation[M]. 2nd ed. Canada: Prentice-Hall Inc, 2001

第一作者简介:蔡永香,博士生。研究方向为地理信息智能化处理及其应用。  
E-mail:caiyx2002cn@yahoo. com. cn

Points Group Generalization Based on Konhonen Net

CAI Yongxiang<sup>1,2,4</sup> GUO Qingsheng<sup>1,2,3</sup>

- (1 School of Resource and Environment Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)
- (2 Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)
- (3 State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)
- (4 School of Geoscience, Yangtze University,1 Nanhuan Road, Jingzhou 434102, China)

**Abstract:** Selection and typification of points group is discussed based on Kohonen net modal, standard SOM algorithm is improved so as to preserve the spatial distribution properties of original points group. Examples illustrate that this method can be used to generalization of points group with any spatial distribution.

**Key words:** points group; map generalization; Kohonen net; spatial distribution