引文格式：郭霄,薛睿,苏艳飞,等. 大比例尺到中小比例尺的居民地聚类简化方法[J]. 地理空间信息, 2025, 23(7): 129-132. doi:10. 3969/j.issn. 1672-4623.2025.07.026

大比例尺到中小比例尺的居民地聚类简化方法

郭　霄[1]，薛　睿[2]，苏艳飞[1]，王创宁[1]，田喜平[1]，杨　磊[1]
（1. 61363部队，陕西 西安 710000；2. 自然资源部第一大地测量队，陕西 西安 710054）

摘　要：针对大比例尺矢量数据简化提取中小比例尺居民地的问题，提出了一种图模型聚类方法。首先对居民地边线加密构建Delaunay三角网，并通过三角网创建最小生成树（MST）；再对MST进行剪枝形成聚类图模型；然后对图模型中的居民地构建Delaunay三角网；最后通过Alpha Shape算法获取凹壳，并优化生成简化的居民地面。由一幅1∶5万数据中居民地矢量数据获取1∶25万居民地，结果表明简化后的居民地满足1∶25万居民地数据精度要求。该方法为大比例尺居民地向中小比例尺简化提供了一条可靠思路。

关键词：中小比例尺；居民地；矢量数据；图；聚类；简化
中图分类号：P283　　**文献标志码：**B　　**文章编号：**1672-4623（2025）07-0129-04

Simplified Method for Clustering Residential Lands from Large Scales to Medium and Small Scales

GUO Xiao[1], XUE Rui[2], SU Yanfei[1], WANG Chuangning[1], TIAN Xiping[1], YANG Lei[1]
(1. 61363 Troops of PLA, Xi'an 710000, China;
2. First Geodetic Survey Team, Ministry of Natural Resources, Xi'an 710054, China)

Abstract: Aiming at the problem of simplified extraction of medium and small scales residential lands from large scales vector data, we proposed a graph model clustering method. Firstly, we encrypted the edges of residential land to construct Delaunay triangular network, and created the minimum spanning tree (MST) by the triangular network. Then, we pruned the MST to form a clustering graph model, and constructed the Delaunay triangular network for residential lands in the graph model. Finally, we used Alpha Shape algorithm to obtain the concave shell, and optimized to generate simplified residential land. We obtained 1∶250 000 residential lands through the vector data of residential lands in 1∶50 000 dataset. The simplified residential lands can meet the accuracy requirements of 1∶250 000 residential land data. The method gives a reliable idea for simplifying the large scale residential lands to medium and small scales.

Key words: medium and small scales, residential land, vector data, graph, clustering, simplification

居民地面作为矢量中最具特点的多边形，已成为地图简化领域常见的研究对象。居民地简化研究主要集中在4个方面：单个居民地形状的简化[1-3]、面的选取[4]、面的聚类合并[5]、布局的模式识别[6]。目前研究多集中在大小、分布、形状均匀的大比例尺居民地，但实际生产中在1∶10万以上1∶5万以下尺度的数据中大量居民地是不均匀的，而针对大比例尺矢量数据向中小比例尺简化的研究较少。鉴于此，本文以大比例尺中分布不规则的居民地为实验对象，通过研究目前聚类简化算法找到适合该类对象的算法，并验证其可行性。

1　聚类方法

目前常用的聚类方法包括K-means、密度聚类[7]、层次聚类、网格聚类、最大期望模型聚类和图论聚类。K-means需制定初始化簇数量且聚类结果偏向圆形扩张；密度聚类更关注密度，未考虑其他性质更适合简单无属性不均匀分布的点聚类；层次聚类的算法较复杂，需要多次遍历数据且受奇异值影响较大；网格聚类将数据范围划分为网格，只有网格密度足够大才能形成一个簇，将影响密度较小处数据的聚类；最大期望模型聚类也需确定初始簇数量；图论聚类又分为基于多级图划分的聚类[8]和基于最小生成树（MST）剪枝[9]的聚类。上述方法不适用于聚类簇数量不明、分布不均匀的居民地面数据聚类。

大比例尺居民地向中小比例尺简化可抽象为大比例尺居民地面数据网络的简化。这种网络通过某种规则搭建，一个具体的网络问题可抽象为点集合V和边

项目来源：国家自然科学基金资助项目（41501498）；陕西省自然科学基础研究计划资助项目（2018JM4041）。
第一作者简介：郭霄（1995—），硕士，助理工程师，研究方向为地图制图学与地理信息工程，E-mail：2426588269@qq.com。

集合 E 组成的图 $G=(V,E)$。因此，本文问题可抽象为对 G 的聚合简化。图论聚类与居民地聚类具有很好的适配性，其中基于多级图划分的聚类是将层次聚类的思想引入有权图结构，对图进行相似度多级点边融合，并在最终融合的粗图上细化调整数据，具体的簇数量由面相似度阈值决定，但在多级划分与细化过程中需多次遍历数据，算法复杂度较高[8]。通过有权图构建MST的算法既不受密度影响又可增加除距离外的其他决策因素，且在构建有权图的基础上直接剪裁算法较简洁，因此本文采用构建MST的方法进行聚类。

1.1 Delaunay三角网构建

对大比例尺居民地面构建三角网，进而形成居民地面数据间关系的网络骨架基础。在建立约束Delaunay三角网[10]之前需对面的边线进行点加密[11]。为更好地适应不均匀面的问题，加密时计算面的每一条边，若长度超过整个面边线均值的1/3，则插入点个数为边长与1/3均值的比；反之则不插入点，这样可在一定程度上保证构建的Delaunay三角网不会出现过短边，且加快了构建速度。边线加密后，首先基于构成面的点生成凸壳，并对凸壳点构建Delaunay三角网；再逐点插入非凸壳的其他点。计算待插入点所在的三角形，若在某个确定的三角形内则删除该三角形，并将其3点与插入点构建生成3个新三角形；若待插入点在两个三角形的边上，则删除两个三角形，并将两个三角形的4个点与待插入点连接形成4个新三角形；若待插入点正好在边界三角形的边线上则删除该三角形，并连接插入点与边界三角形的顶点形成两个新三角形，对于处于一条线上的3个点，不连接顶点与插入点。

需要构建两个点在一个面、另外一个点在其他面的三角形才能通过三角形获得两个面之间邻接关系；但通过上述方法构建的三角形存在由单个面内顶点构成的三角形、单个角度≥170°的狭长三角形以及3个顶点分别属于不同面的三角形，因此需要遍历创建的三角形，删除不满足需求的个体，实现约束三角网的构建[12]。

1.2 MST构建

若图中任意两个顶点之间都存在至少一条路径，则称该图为连通图。一个包含 N 个顶点的连通图 G 至少有 $N-1$ 条边，若该图只有 $N-1$ 条边，则称其为树。如果一个连通图本身不是树，也可看作是在树的基础上添加一些边而形成的，连通图包括多个子树，其中权值之和最小的树即为MST[13]。在形成的约束三角网中，若两个面之间包含三角形则认为它们是相邻的，可以得到面之间的邻接关系图。获得MST需对邻接图赋权，当距离小于阈值时，权值只关注距离；当大于阈值时，边权值信息包含面面之间的方向和密度差异，即提高街区规划走向与密度相似街区间的权值，使聚类更加符合街区的规划[14]。距离可通过三角形的高或中间距离获取，当三角形中两个点在同一个面构成的边与另外一个点形成的角度>90°时，以中间距离作为权值；反之，则以高度作为距离权值（图1）。

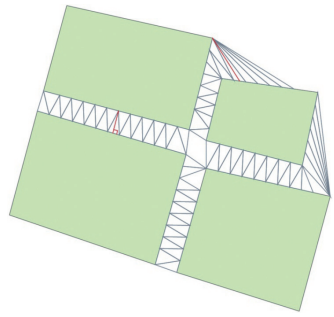


图1 距离的定义

面的方向通过最小外接矩形方法获取，若两个建筑物的方向值之比<0.25，则连接这两个面的边的方向权值为2；反之，则等于2减去方向之比。面的密度是面的面积与面的面积加其周边全部Delaunay三角网面积之比，若两个建筑物的密度比<0.25，则连接这两个面的边的密度权值为2；反之，则等于2减去方向之比。采用加点法处理形成的有权连通图：①在图中任意选择一个顶点 k 存入集合 U 中，不在 U 中的顶点存入 W 中，集合TH中存放MST的边；②找到 U 中顶点与 W 中顶点连接边中权最小的边存入TH集合，并将该边在 W 中的顶点移动到 U 中；③重复步骤②，直至 W 中的顶点全部移动到 U 为止。MST中红线标注数字为权值，每条边都是一个CPEdge对象，记录了红边连接的左右面编号、红边左右点坐标和边的权值。

1.3 剪枝

直接通过设置阈值进行剪裁，阈值与待剪裁的树需相互适应，若权值平均值较大而阈值较小则导致全部边被剪裁。因此，需要对阈值的设置进行归一化处理，将阈值转换为单个边权值与所有边权值平均数之比，可在剪枝时大致了解边的权值分布情况，且针对不同数据阈值不会变化较大，有助于更好地设置阈值。通过比较单个边权值与所有边权值平均数之比和阈值，可以获得大于阈值的边数 m（需剪裁的边数），由于MST的边数为 $N-1$，因此剪裁后的簇数应为 $m+1$。图2中黄色边为阈值为1时剪裁后的剩余结果，最后聚类数量为4，整棵树的边数为9，因此大于阈值的个数为3。

2 通过图模型生成简化面

2.1 图对象结构与深度优先搜索

裁剪后的结果可看作是多个图结构，需要获取这



图2　剪枝

些图结构中节点之间的连接关系，因此需构建裁剪前树的图数据结构。常用的图数据结构记录方法为邻接矩阵与邻接表，由于构建CPEdge对象与邻接表结构更相似，因此本文采用邻接表的形式。邻接表分为常规格式和网络分析格式（Pajek）[15]，Pajek格式由于节点中数据不重复，在数据存储与运算上具有优势，因此本文采用该格式。为了更好地遍历数据，对Pajek格式进行改进拆分，并保证节点编号无重复，通过字典数据结构进行存储，其中节点编号为key、相邻节点为value。通过对最终图结构的递归搜索可将聚类后的图结果中的节点存储在Heshset结构中。面聚类后可获得5个Heshset结构数据，即2，8、4，3，7，0、1、5、6。

2.2　二次构建Delaunay三角网

以改进拆分后的数据为单元再次构建Delaunay三角网，并在三角网的基础上继续插点，插点位置为面积大于所有同一个面内三角形和平均值3倍的三角形内接圆的圆心或大于6倍的圆心与3个点的中点。通过该方式可将大于3倍的三角形剖分为3个三角形，大于6倍的三角形剖分为6个三角形，使内部的点更加密集且均匀，防止生成凹壳时出现内部岛面。

2.3　凹壳生成

在构建聚类面三角网的基础上形成新的外边线，若采用凸壳会丢失很多面特征，因此对其生成凹壳。Alpha Shape算法[16]是常见的凹壳生成方法，需设置圆半径 r，半径阈值应与三角形的边长 L 分布相关，L 服从正太分布，则有：

$$f(L)=\frac{1}{\delta\sqrt{2\pi}}e^{-\frac{(l-\mu)^2}{2\delta^2}}$$

当 $f(L)=x\times\frac{1}{\delta\sqrt{2\pi}}$ 时，$l\approx\mu\pm\sqrt{-\ln^x\times2\times\delta^2}$；

当 x 确定且取正值时，$l=r$，当 x 取值过大时生成的凹壳将无限接近凸壳。为更好地获取面聚类后的形状，r 的取值应在最大值与平均值之间，且为 $l>r$ 的边分布概率不能过低，因此 x 取1/10。r 取值合适时，即可生成保留面聚类后大致外部特征的凹壳。图3中红线为生成的凹壳，即提取结果，在面内插入的内接圆点和边上的加密点相当于图3中的点，这些点分布越均匀，凹壳生成的效果越好。对生成的凹壳进行规则化，获得最终的简化成果。

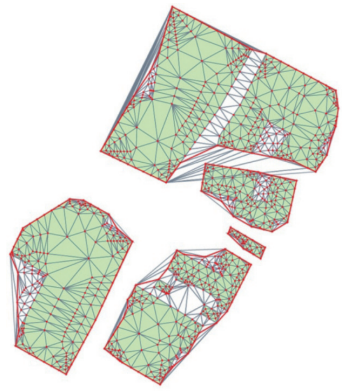


图3　Alpha Shape算法提取边界

3　精度分析

本文以验收合格的1∶5万国标图幅J50E014003中居民地为实验数据，数据范围为经度114.500°~114.750°、纬度37.666°~37.833°，对原始数据中相邻的面进行合并后，采用本文方法对数据进行聚类简化；并与人工采集且通过三级检查、一级验收合格的1∶25万数据进行对比。实验结果见图4，其中蓝色边线为1∶5万数据，红色边线为1∶25万数据，黄色边线为本文方法生成的1∶25万数据，标记数字为黄色边线的实地面积（m^2）。

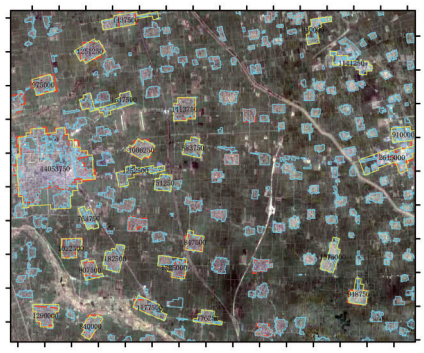


图4　实验结果

由1∶25万图式规定可知，居民地图上面积小于8 mm^2、实地小于500 000 m^2时不予表示，整体误差不应超过±50 m。目前验收合格的1∶25万数据多通过人工采集获取，多个面之间是否合并，面是否保留通过人工判读决定，没有唯一的综合规则，因此同一幅图会随编制人的不同而产生满足图式要求的多种1∶25万居民地综合结果。图4中黄色边线区域大于红色区域，其原因在于，1∶25万数据采集时通过人工判别进行了选取，其结果与影像中整体居民地分布差距不

大，可通过审核验收，但依靠人工判读的标准不够统一；而本文方法的选取规则是首先保留形状为面的居民地，再通过本文方法进行聚类生成凹壳，最后将凹壳面积与500 000 m^2进行对比，大于的保留、小于的剔除，参数都是固定的，这样的综合方法更加规范，黄色边线与居民地图斑套和的效果更好。

本文方法是对原有1∶5万面的聚类简化取舍，因此评价方法应包括两类：①面保留是否正确，包含面的数量与面积对比；②聚类简化后的面与1∶25万数据对比精度是否满足条件。前者可评价整体精度，评价标准可借用交并比和面积比公式，后者可评价数据的细节精度，主要针对属于真正例部分的数据，采用采样点的Δx与Δy的方差与1∶25万规定误差进行评价，采样方法为格网法[17]。交并比混淆矩阵[18]见表1。

表1 混淆矩阵

真实情况	预测结果	
	正例	反例
正例（保留）	*TP*（真正例）25	*FN*（假反例）6
反例（不保留）	*FP*（假正例）33	*TN*（真反例）204

交并比精度公式为：

$$\mathrm{MIoU}=\frac{1}{k+1}\left(\sum_{i=0}^{k}\frac{TP}{FN+FP+TP}\right)$$

根据实验结果，最终预测的MIoU为61.5%。通过实验可知，正确分类的面有229个，错误分类的面有39个，其中正确分类面积占全部面积的81.7%，虽然数量精度不高但归类错误的面积不大，对图面整体影响不大。利用格网法采集数据，共获取399个特征点，将预测结果与1∶25万数据特征点结果进行对比，Δx与Δy的分布见图5，可以看出，实验误差整体符合正态分布，均值约为0，且多集中在±50 m以内，符合1∶25万数据采集要求，简化结果符合图式规范要求。

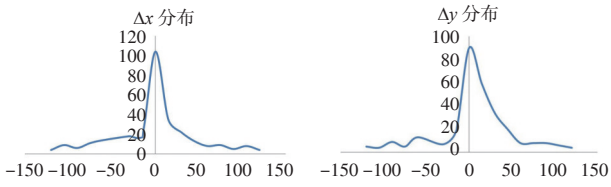


图5 误差分布

4 结 论

目前针对不规则分布的中小比例尺居民地聚类简化研究较少，本文从对不规则分布的适应性与算法的复杂度等角度分析了常见的聚类简化算法，设计了较适用于不规则分布大比例尺居民地的数据简化方法，并对该方法进行了实验验证。评价结果表明，该方法可应用于1∶5万到1∶25万居民地数据的简化，对实际数据生产具有指导意义。

参考文献

[1] Yang M, Yuan T, Yan X, et al. A Hybrid Approach to Building Simplification with an Evaluator from a Backpropagation Neural Network[J]. International Journal of Geographical Information Science, 2021(5):1-30

[2] Buchin K, Meulemans W, Speckmann B. A New Method for Subdivision Simplification with Applications to Urban-area Generalization[C]//Proceedings of the 19th ACM Sigspatial International Conference on Advances in Geographic Information Systems, USA, 2011

[3] Zhou Z, Fu C, Weibel R. Building Simplification of Vector Maps Using Graph Convolutional Neural Networks[C]. European Cartographic Conferece, Austria, 2022

[4] Li W, Yan H, Lu X, et al. A Heuristic Approach for Resolving Spatial Conflicts of Buildings in Urban Villages[J]. ISPRS International Journal of Geo-information, 2023, 12(10):392

[5] 郭庆胜，黎佳宜，曹元晖，等. 建筑物白模多边形的自动合并[J]. 武汉大学学报(信息科学版)，2021，46(1):12-18

[6] 肖天元，艾廷华，余华飞，等. 地图综合图卷积神经网络点群简化方法[J]. 测绘学报，2024，53(1):158-172

[7] 孟妮娜，高晨博，王正阳，等. 面向建筑物数据的密度聚类算法研究[J]. 测绘科学，2022，47(11):204-214

[8] 金澄，安晓亚，陈占龙，等. 矢量居民地多边形多级图划分聚类方法[J]. 武汉大学学报(信息科学版)，2021，46(1):19-29

[9] 闫浩文，应申，李霖. 多因子影响的地图居民地自动聚群与综合研究[J]. 武汉大学学报(信息科学版)，2008，33(1):51-54

[10] 晏雄锋，艾廷华. 中小比例尺下居民地综合及冲突处理[J]. 测绘科学，2016，41(4):124-127

[11] 吕峥，孙群，温伯威，等. 顾及形状相似性的道路与居民地协同化简方法[J]. 地球信息科学学报，2024，26(5):1 270-1 282

[12] 郭沛沛，李成名，殷勇. 建筑物合并的Delaunay三角网分类过滤法[J]. 测绘学报，2016，45(8):1 001-1 007

[13] 蔡娇楠. 顾及格式塔原则的街区建筑物MST聚类[D]. 西安:长安大学，2018

[14] 张振鑫，张维，刘嫔，等. 矢量地图数据简化研究进展[J]. 测绘工程，2016，25(6):10-14

[15] 汪小帆，李翔，陈关荣，等. 网络科学导论[M]. 北京:高等教育出版社，2012:46-47

[16] 高智梅，王竞雪，沈昭宇. 机载LiDAR建筑物点云渐进提取算法[J]. 测绘通报，2021(8):7-13

[17] 张萌，杜为财，郭霄. 一种利用数据集成的基础地理信息数据改化方法[J]. 测绘地理信息，2022，47(3):165-169

[18] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4):640-651