

道路网 stroke 生成问题的形式化表达与新算法

田晶^{1,2} 何青松¹ 颜芬¹

1 武汉大学资源与环境科学学院,湖北 武汉,430079

2 武汉大学地理信息系统教育部重点实验室,湖北 武汉,430079

摘要:首先,将道路网 stroke 生成问题形式化表达为聚类问题;然后,在该表达下提出了基于层次聚类思想的 stroke 生成算法,并进行了算法的时间复杂度分析和算法特性分析。最后,用深圳市 1:5 万道路网对该算法进行了验证。

关键词:道路网; stroke; 层次聚类; 形式化表达

中图法分类号:P208; P283.1 **文献标志码:**A

道路网中的 stroke 是根据特定准则将路段连接而形成的道路链。stroke 的概念源于 Gestalt 认知原则中好的连续律,术语 stroke 从一笔画出的曲线段的思想中产生。通俗地说,就是将一些满足好的连续律的线状要素连接到一起形成若干 stroke,符合人们的认知思维^[1]。同时,stroke 也是一种主要的道路网模式^[2]。stroke 通常作为道路网综合^[1]、拓扑分析^[3-4]、示意性地图生成^[5]、模式识别^[6]等研究的重要步骤。在这些研究中 stroke 的作用主要表现为:① 作为道路网综合的对象,如基于 stroke 重要度排序的道路网综合方法^[7];② 作为确定路段重要程度的依据,如文献[8-9]在基于面合并的道路网综合方法中,通过 stroke 的重要性来确定构成网眼的路段的重要性;③ 作为分析的基本统计单元,如文献[4]对道路网进行分析时,先生成 stroke 集,然后通过计算 stroke 的拓扑参量确定道路网的特性;④ 作为进行下一步变换的预处理结果,如文献[5]在生成示意性路网图时,先生成 stroke,然后在道路网重新组织的基础上生成示意性路网图。再如文献[6]在识别辐射模式与环模式时,将 stroke 作为初步识别单元。

道路网 stroke 的生成主要由路段连接规则和连接策略决定。连接规则主要指几何规则、专题规则以及混合规则^[10]。几何规则由 Gestalt 原则中的“好的延续性”原则导出,主要指两条路段的转折角小于阈值;专题规则主要是指道路同名或道路等级相同;混合规则是指同时考虑几何规则和专题规

则。连接策略主要有每对最大适合策略、自身最大适合策略、自身适合策略等^[11]。每对最大适合策略是满足连接规则、互为最适应的路段相连,如与路段 a 转折角最小的是路段 b ,与路段 b 转折角最小的是路段 a ,则路段 a 和路段 b 相连。文献[12]即采用这种策略。自身最大适合策略是满足连接规则,某条路段每次都与其最适应的路段相连,不需要考虑被连路段是否有更好的选择,如在同一交叉点,与路段 a 转折角最小的是路段 b ,与路段 b 转折角最小的是路段 c ,与路段 c 转折角最小的是路段 b ,假定从路段 a 开始连接,那么路段 a 和路段 b 相连,而路段 b 不与路段 c 相连。文献[4]即采用这种策略。自身适合策略是某一路段随机地与满足连接规则的路段相连的策略,文献[13]即采用这种策略。从理论上,上述策略产生的 stroke 集是不确定的,特别地,在所有交叉点由路段连接准则判断的结果均不相同,由每对最大适合策略生成的 stroke 集是确定的。除了文献[14]提出在复杂道路交叉点连接道路应该满足所有配对路段凸角之和最大之外,已有的 stroke 生成方法几乎都是上述连接规则和连接策略的组合。从计算机实现的角度,主要有道路交叉点配对(点配对)和路段搜索(线搜索)两种实现方式。

道路网 stroke 生成问题是道路网分析与综合研究领域的一个核心问题,然而,由于该问题面向实际应用,所以目前对其研究主要侧重于生成方法的研究,缺乏其概念层次上的探讨。因此,本

文对 stroke 生成问题进行形式化表达,提出相应的 stroke 生成算法。

1 stroke 生成问题的形式化表达

理论上,生成道路网的 stroke 是将 N 条路段连接成 M 条 stroke 的过程($N \geq M$)。现在给出“生成 stroke”问题的形式化定义。令 R 表示道路网中的路段集合, R_i 表示其中的第 i 条路段, S 表示 stroke 集合, S_i 表示第 i 条 stroke,即 $R = \{R_1, R_2, \dots, R_N\}$, $S = \{S_1, S_2, \dots, S_M\}$ 。那么,“生成 stroke”,即将 R 分成 M 个类 S_1, S_2, \dots, S_M , 满足下面三个条件:

$$S_i \neq \emptyset, i = 1, \dots, M$$

$$\bigcup_{i=1}^M S_i = S$$

$$S_i \cap S_j = \emptyset, i \neq j, i = 1, \dots, M, j = 1, \dots, M$$

根据上述形式化表达,从概念层上可将 stroke 生成问题归结为聚类问题,类由路段构成,各个类的形状呈(曲)线状。图 1 给出了一个示例,图 1(a)为道路网 $R = \{R_1, R_2, R_3, R_4, R_5\}$, 图 1(b)和 1(c)是两种 stroke 生成结果。对于图 1(b), $S = \{S_1, S_2, S_3\}$, $S_1 = \{R_1, R_5\}$, $S_2 = \{R_2, R_3\}$, $S_3 = \{R_4\}$ 。对于图 1(c), $S = \{S_1, S_2, S_3\}$, $S_1 = \{R_1, R_5\}$, $S_2 = \{R_3\}$, $S_3 = \{R_2, R_4\}$ 。

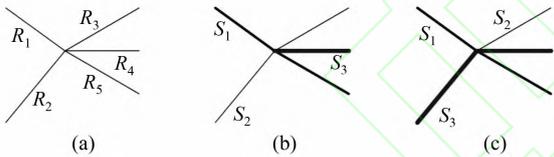


图 1 形式化表达示例

Fig. 1 Illustration of Formalization

2 基于层次聚类思想的 stroke 生成算法

2.1 算法原理

本文将 stroke 生成形式化表达为聚类问题,

那么,很自然地,可以利用已有聚类算法的思想。考虑到 stroke 生成的过程是路段逐步合并的过程,所以,本文基于层次聚类的思想提出了 stroke 生成算法。顾及到生成的类的形状为延伸的(曲)线状,所以采用单连接机制。本文暂不深究两条路段连接时的定量化模型,以几何准则作为连接准则,即判断两条路段之间的夹角,夹角越小,说明两条路段越易被认为连接,根据文献[10]的经验,路段夹角的阈值设为 60° ,当然,这个阈值可以重新自定义。具体算法如下所示。

输入:初始的路段集合 R 。

输出:生成的 stroke 集合 S

1) 初始化:

$$R_0 = \{R_i, i = 1, \dots, N\}$$

令 $t = 0$

2) 重复执行下列步骤:

① $t = t + 1$

② 在 R_{t-1} 的所有可能配对 (R_r, R_s) 中找一组 (R_i, R_j) , 满足:

$$\text{Touch}(R_i, R_j) = \text{TRUE AND Angle}(R_i, R_j) \leq 60 \text{ AND Angle}(R_i, R_j) = \text{Min}_{r,s} \text{Angle}(R_r, R_s)$$

③ 生成 $R_q = R_i \cup R_j$, 并更新 $R_t = (R_{t-1} - \{R_i, R_j\}) \cup R_q$

直到 R_{t-1} 中的所有元素对 (R_q, R_p) 满足:

$$(\text{Touch}(R_q, R_p) = \text{FALSE}) \text{ OR } (\text{Touch}(R_q, R_p) = \text{TRUE AND Angle}(R_q, R_p) \geq 60)$$

3) 令 $S = R_{N_t}$

其中,Touch()是判断路段是否在端点处相连的函数,Angle()是判断路段夹角的函数。

图 2 为算法的一个简单示例,最初的道路网集合包含 16 条路段(图 2(a)),字母为路段交叉点注记,数字为路段注记;图 2(b)为算法运行第 5 步聚类后的结果;图 2(c)是算法处理的最终结果;图 2(d)为对应的层次结构。

2.2 时间复杂度分析

对于 N 条路段聚成 M 条 stroke,在层次 t 上,有 $N-t$ 个聚类。为了确定 $t+1$ 层上要合并

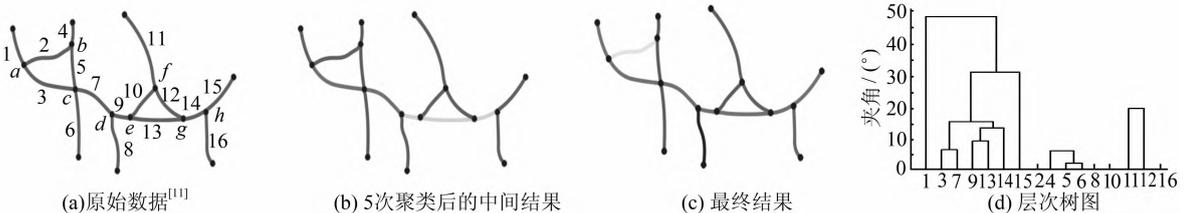


图 2 算法示例

Fig. 2 Illustration of the Algorithm

的聚类对,理论上,须考虑 $C_{N-t}^2 = \frac{(N-t)(N-t-1)}{2}$ 个配对,由于最后生成的聚类数为 M ,聚类过程总共要考虑到的聚类对数量为:

$$\sum_{t=0}^{N-M} C_{N-t}^2 = \frac{(N+1)N(N-1) - M(M-1)(M-2)}{6}$$

通过上式得出算法的全部运算在 N^3 数量级上,此外,算法的时间复杂度与 Touch() 函数和 Angle() 函数的具体实现有关。

2.3 算法特点分析

1) 中止条件方面,与传统的层次聚类算法相比,该算法的中止条件是生成 stroke 集合,集合中的这些 stroke 要么不在端点处相连,要么在端

点处相连但不满足夹角阈值,要么在中间点相交,所以该算法无需设定“距离”阈值作为算法中止条件,同时也无需预先规定聚类的个数。

2) 处理过程方面,算法处理的过程与以往的点配对或线搜索的实现机理不同。点配对实现是以道路交叉点为对象,每次在交叉点将汇集于该点的路段配对,遍历所有交叉点即完成 stroke 的生成;线搜索实现是以道路为对象,搜索与该路段配对的路段,然后递归搜索,直到所有路段均被处理。本文提出的算法并不是沿着一条线或是在某个交叉点处理,而是每次都检测“最易连接”的两条路段,将其连接,直到满足中止条件。表 1 列举了对图 2(a) 的处理过程,假定处理路段交点时按照字母序,处理路段时按照数字序。

表 1 不同实现的处理流程

Tab. 1 Processing of Different Implementation

| 步数 | 点配对实现 ^[1] | 线搜索实现(每对最大适合策略) | 本文的层次聚类实现 |
|----|---|---|---|
| 1 | a 点处 1, 3 连接, 2 孤立 | 处理 1, 1 与 3 连接, 3 与 7 连接, 7 与 9 连接, 9 与 13 连接, 13 与 14 连接, 14 与 15 连接 | 连接 5, 6 |
| 2 | b 点处 4, 5 连接, 2 孤立 | 处理 2, 2 孤立 | 连接 3, 7 |
| 3 | c 点处 3, 7 连接, 5, 6 连接 | 处理 4, 4 与 5 连接, 5 与 6 连接 | 连接(5, 6), 4 |
| 4 | d 点处 7, 9 连接, 8 孤立 | 处理 8, 8 孤立 | 连接 9, 13 |
| 5 | e 点处 9, 13 连接, 10 孤立 | 处理 10, 10 孤立 | 连接(9, 13), 14 |
| 6 | f 点处 11, 12 连接, 10 孤立 | 处理 11, 11 与 12 连接 | 连接(3, 7)(9, 13, 14) |
| 7 | g 点处 13, 14 连接, 12 孤立 | 处理 16, 16 孤立 | 连接(11, 12) |
| 8 | h 点处 14, 15 连接, 16 孤立 | | 连接(3, 7, 9, 13, 14), 15 |
| 9 | | | 连接(3, 7, 9, 13, 14, 15), 1 |
| 结果 | {1, 3, 7, 9, 13, 14, 15} {2} {4, 5, 6} {8} {10} {11, 12} {16} | {1, 3, 7, 9, 13, 14, 15} {2} {4, 5, 6} {8} {10} {11, 12} {16} | {1, 3, 7, 9, 13, 14, 15} {2} {4, 5, 6} {8} {10} {11, 12} {16} |

3) 可扩展性方面,第一,在路段连接准则上,可以很容易地加入专题信息,如可先判断两条路段是否等级相同,假定不同,直接从配对中排除。第二,本文提出的基于层次聚类的算法与每对最大适合策略产生的结果相同。通过对算法执行顺序进行改进,即可轻易实现自身最大适合策略与自身适合策略。第三,由于聚类时采用的是单连接机制,在路段连接时考虑的仍是局部特征,可通过使用平均连接机制或完全连接机制使得算法在运行时考虑单条 stroke 的全局特征。第四,本文采用的是最基本的层次聚类算法的思想,目前针对不同运用,层次聚类算法有许多新的改进版本,这些版本蕴含了优化或针对大数据量处理的思想,如 BIRCH 算法、CURE 算法、ROCK 算法等^[15],可将这些算法的思想运用到本文问题中。

2.4 实例研究

对深圳市 1:5 万道路网(见图 3(a))进行了实验验证,对该道路网进行拓扑检查以及交点处

打断处理,该道路网包含 2 495 条路段,图 3(b)是 1 600 步聚类后的结果,其中,黑色路段表示处理过的路段,灰色表示还未处理的路段,图 3(c)表示最后生成的 stroke,共 528 条。

在实际运用的过程中,可对算法进行一定的优化,因为存在一些必然连接到一起的配对,可对在端点处均仅有一条满足连接准则的路段预先连接,如图 2(a)中的路段 1 和 3,图 4 为对图 2(a)的预先连接,路段 1、3、7,路段 5、6,路段 11、12,路段 14、15 预先连接,路段数从 16 条变为 11 条。

3 结 语

本文在概念层上将 stroke 问题形式化表达为聚类问题,在操作层上提出了相应的 stroke 生成算法。需要进一步研究的问题有:① 本文在判断连接时仅考虑了路段间的转折角,即便加入专题语义信息,如路名、道路等级,也不是创新。关

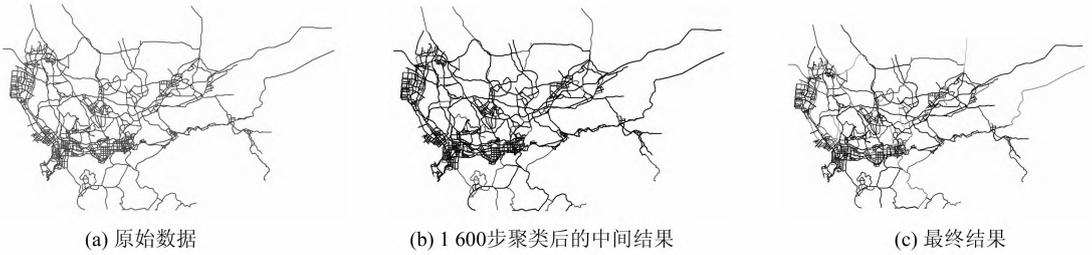


图 3 深圳市道路网的 stroke 生成结果

Fig. 3 Results for Shenzhen Road Network

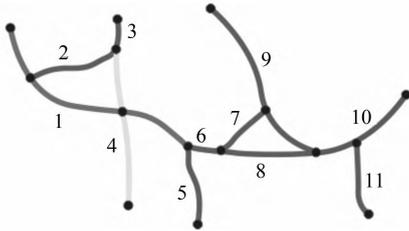


图 4 预连接结果

Fig. 4 Results of Pre-concatenate Process

键是要建立新的认知量化模型来表达 Gestalt 原则中的“好的延续律”,使得更符合人的视觉认知特点。如在连接时,人们还会去考虑连接路段的长度、后续连接的路段与初始路段的关系^[16]。

② 得益于志愿者地理信息的发展,目前本学科的发展处于大规模复杂地理信息的时代,如何对 Openstreetmap 上的大规模城市道路网快速生成 stroke,以满足在线分析与综合的要求,还需要对算法性能进行进一步的优化。

参 考 文 献

- [1] Thomson R C, Richardson D E. The “Good Continuation” Principle of Perceptual Organization Applied to the Generalization of Road Networks[C]. The 19th International Cartographic Conference, Ottawa, 1999
- [2] Mashall S. Streets and Patterns[M]. New York: Spon Press, 2005
- [3] Porta S, Crucitti P, Latora V. The Network Analysis of Urban Streets: A Dual Approach[J]. *Physica A*, 2006, 369: 853-866
- [4] Jiang Bin. A Topological Pattern of Urban Street Networks: Universality and Peculiarity[J]. *Physica A*, 2007, 384: 647-655
- [5] Li Zhilin, Dong Weihua. A Stroke-Based Method for Automated Generation of Schematic Network Maps[J]. *International Journal of Geographical Information Science*, 2010, 24(11): 1 631-1 647
- [6] Heinzle F, Ander K H. Characterising Space via Pattern Recognition Techniques: Identifying Patterns in Road Networks[M]//Mackness W A, Rusas A, Sarjakoski L T. Generalisation of Geographic Information: Cartographic Modelling and Applications. New York: Elsevier, 2007
- [7] Zhang Qingnian. Road Network Generalization Based on Connection Analysis[C]. The 11th International Symposium on Spatial Data Handling, Leicester, UK, 2004
- [8] Chen Jun, Hu Yungang, Li Zhilin, et al. Selective Omission of Road Features Based on Mesh Density for Automatic Map Generalization[J]. *International Journal of Geographical Information Science*, 2009, 23(8): 1 013-1 032
- [9] Touya G. A Road Network Selection Process Based on Data Enrichment and Structure Detection[J]. *Transactions in GIS*, 2010, 14(5):595-614
- [10] Zhou Qi, Li Zhilin. A Comparative Study of Various Strategies to Concatenate Road Segment into Strokes for Map Generalization [J]. *International Journal of Geographical Information Science*, 2012, 26(4): 691-715
- [11] Jiang Bin, Zhao Sijian, Yin Junjun. Self-organized Natural Roads for Predicting Traffic Flow: A Sensitivity Study[J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(7):1-27
- [12] Edwardes A J, Mackness W A. Intelligent Generalization of Urban Road Networks[C]. GIS Research UK Conference, York, 2000
- [13] Liu Xingjian, Zhan B F, Ai Tinghua. Road Selection Based on Voronoi Diagrams and “strokes” in Map Generalization[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2009, 12: 194-202
- [14] Yang Bisheng, Luan Xuechen, Li Qingquan. Generating Hierarchical strokes from Urban Street Networks Based on Spatial Pattern Recognition[J]. *International Journal of Geographical Information Science*, 2011, 25(12): 2 025-2 050
- [15] Theodoridis S, Koutroumbas K. Pattern Recognition (4th Edition)[M]. Amsterdam: Elsevier, 2009

- [16] Chaudhry O, Mackaness W. Rural and Urban Road Network Generalization Deriving 1 : 250 000 from OS MasterMap[EB/OL]. <http://www.era.lib.ed.ac.uk/bitstream/1842/1137/1/ochaudry001.pdf>, 2001

Formalization and New Algorithm of stroke Generation in Road Networks

TIAN Jing^{1,2} HE Qingsong¹ YAN Fen¹

1 School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China

2 Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, Wuhan 430079, China

Abstract: Road network generalization and analysis have been hot issues in the geographical information science. The stroke in a road network is defined as a set of one or more road segment in a non-branching, connected chain, based on the good continuity principle. The stroke plays an important role in road network generalization, analysis, pattern recognition and schematic map generation. Existing studies focus on the stroke generation algorithms. However, the formalization of stroke generation at the conceptual level is absent. This paper first formalizes the stroke generation as clustering problem, and then presents a hierarchical clustering based stroke generation algorithm. Time complexity and some properties of the proposed algorithm are analyzed in detail. Finally, the algorithm is verified using the Shenzhen road network at the 1 : 50 000 scale.

Key words: road network; stroke; hierarchical clustering; formalization

First author: TIAN Jing, PhD, lecturer, specializes in automated map generalization and spatial data mining. E-mail: yutaka-2010@163.com

Foundation support: National Science Foundation for Fostering Talents in Basic Research of the National Natural Science Foundation of China, No. J1103409.