

基于改进 Hausdorff 距离的图形注记匹配方法

程胜一

[上海勘察设计研究院(集团)股份有限公司, 上海市 200438]

摘要: AutoCAD 作为常用的计算机辅助制图软件,在工程领域被广泛应用。在实际绘图任务中,常常需要高效且准确地完成图形对象与人工注记文本信息之间的匹配。该匹配问题可抽象为2个集合对象之间距离和最短的组合优化问题。尽管通常采用动态规划、贪心算法、最近邻等构造性算法进行求解,但仍面临着算法复杂度较高的问题。为此,提出了一种基于改进 Hausdorff 距离的模拟退火算法,利用改进后的 Hausdorff 距离计算方法,并结合模拟退火算法的概率搜索特性,以获得匹配问题的近似最优解。通过与一系列算法的对比试验,验证了所提方法的有效性和准确性。试验结果显示,该算法能够明显提高图文匹配的精度和计算效率,适用于实际的工程制图任务。

关键词: 改进 Hausdorff 距离;模拟退火算法;组合优化

中图分类号: TP31

文献标志码: A

文章编号: 1009-7716(2025)08-0382-04

Graphic Annotation Matching Method Based on Modified Hausdorff Distance

CHENG Shengyi

[Shanghai Institute of Survey, Design and Research (Group) Co., Ltd., Shanghai 200438, China]

Abstract: AutoCAD, as a commonly used computer-aided drafting software tool, is widely applied in engineering fields. In practical drafting tasks, it is often necessary to efficiently and accurately match graphical objects with manually annotated textual information. This matching problem can be abstracted as a combinatorial optimization problem of the shortest distance between two sets of objects. Although dynamic programming, greedy algorithms, and nearest neighbor methods are commonly used to solve such problems, the problem of high algorithmic complexity is still faced. Therefore, a simulated annealing algorithm based on a modified Hausdorff distance is proposed. By employing an enhanced calculation method of Hausdorff distance combined with the probabilistic search characteristics of the simulated annealing algorithm, an approximate optimal solution to the matching problem is obtained. Through a series of comparative algorithm experiments, the effectiveness and accuracy of the proposed method are verified. The experimental results indicate that the algorithm can significantly improve the precision and computational efficiency of graphic-text matching, which is suitable for the practical engineering drafting tasks.

Keywords: modified Hausdorff distance; simulated annealing algorithm; combinatorial optimization

0 引言

AutoCAD 作为常用的计算机辅助制图软件工具,因其强大且灵活的绘图功能和处理矢量图形的能力而被广泛应用于建筑、规划、测绘等部门。在一些绘图工作中,需要对图中的图形对象与其对应的注记文本信息进行匹配。在将 AutoCAD 数据转为 GIS 软件支持的数据格式过程中,需要关联匹配图斑及其注记信息的匹配^[1-4]、地形图中高程注记信息与

高程点对象之间的匹配问题^[5-7]。

在实际工作中,总是会遇到各种各样的类似匹配问题,比如在管线绘图工作中,需要对人工注记的节点编号与节点图形对象进行一对一匹配,进而提取相关信息(见图1)。

本文假设文本匹配问题是图形对象与其对应的人工注记文本信息2个集合之间距离和最短问题,从而将其转化成一个类似旅行商问题(TSP)的组合优化问题。这种组合优化问题是一个 NP 完全问题,暴力求解算法复杂度为 $O(n!)$,目前计算能力基本无法求解,通常需采用动态规划、贪心算法、最近邻等构造算法来求解,但算法复杂度仍然很高;也有利用

收稿日期: 2025-02-19

作者简介: 程胜一(1972—),男,学士,高级工程师,从事工程测量工作。

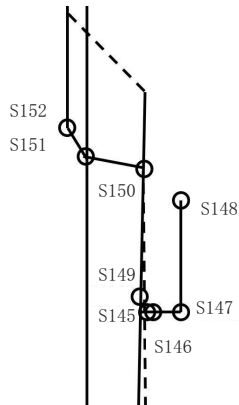


图1 管线节点与点号匹配示例图

遗传算法、粒子群算法、模拟退火算法等启发式算法来快速求该问题的近似解^[8]。

本文通过改进 Hausdorff 距离来描述图形对象距离,利用模拟退火算法求得最短改进 Hausdorff 距离和来匹配图形中图形对象与文本注记信息;同时通过相关对比试验,验证了相关方法在这类匹配问题上的有效性和准确性。

1 基于改进 Hausdorff 距离的匹配方法

由图 1 可知,人工注记的文本信息大致都在图形对象附近。假设人工标注的规律可归纳为图形对象与注记对象之间总距离和最小,则这个匹配问题进行抽象化的数学表示即为:在一个图形绘制空间内,有一个图形对象集合 $G = \{g_i | i = 1, 2, \dots, n\}$ 及对应的注记文本对象集合 $T = \{t_j | j = 1, 2, \dots, n\}$;这 2 个集合的元素匹配关系可表示为无重复元素的整数序列 $K = \{k_i | 1 \leq k_i \leq n, k_i \in Z, i = 1, 2, \dots, n\}$,求解这一整数序列 K ,使得对象之间距离和最短。数学表达式为:

$$\operatorname{argmin}_K \sum_{i=1}^n d(g_i, t_{k_i}) \tag{1}$$

式(1)中图形对象 g 与文本对象 t 之间的距离 $d(g_i, t_{k_i})$ 可以有多种定义方式,最简单的方式是用 2 个对象质心点 $C_g(x_g^c, y_g^c)$ 、 $C_t(x_t^c, y_t^c)$ 之间的距离 $d_c(g, t)$ 来定义,其表达式为:

$$d_c(g, t) = \|C_g - C_t\| \tag{2}$$

但图形对象和注记文本对象一般是一定形状的对象,质心距离可能会导致以式(1)为目标函数的最优解出现偏差。例如图 2 中文本 S113 与其对应的图形对象(左下边方框)质心距离为 11.5 m,而文本 S112 与其对应的图形对象(右上方框)质心距离为 10.0 m,即出现了最短距离匹配错误的匹配图形对象和文本对象。

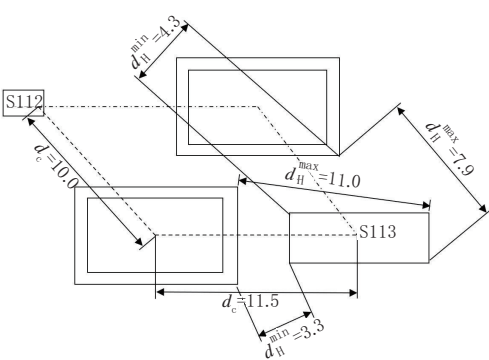


图2 质心距离与 Hausdorff 距离的差异

鉴于以上可能会出现的情况,本文引入 Hausdorff 距离,这是一种衡量几何图形相似度时较为常用的指标^[4-5]。此时图形对象 g 和文本对象 t 轮廓点集有如下定义:

$$\begin{aligned} Pg &= \{p_i^g(x_i, y_i) | i = 1, 2, \dots\}, \\ Pt &= \{p_j^t(x_j, y_j) | j = 1, 2, \dots\} \end{aligned} \tag{3}$$

图形对象与文本对象的标准 Hausdorff 距离定义为式(2)中 $d_H^{\max}(g, t)$ 。文献[9]通过对比不同单向距离选取标准的 Hausdorff 距离,发现其对于图形相似度具有不同的效果表现。基于本文图形相似度问题的特点,常规的 Hausdorff 距离在点匹配问题上与质心距离一样仍然不是最优解(见图 2)。因此,提出一种改进型距离如 $d_H^{\min}(g, t)$ 作为式(1)中的距离定义,其表达式为:

$$\begin{aligned} d_H^{\max}(g, t) &= \max_{p_i^g \in pt} \min_{p_j^t \in pg} \|p_i^g - p_j^t\|, \\ d_H^{\min}(g, t) &= \min_{p_i^g \in pt} \min_{p_j^t \in pg} \|p_i^g - p_j^t\| \end{aligned} \tag{4}$$

将式(1)转化成改进型 Hausdorff 距离和的优化求解:

$$\operatorname{argmin}_K \sum_{i=1}^n d_H^{\min}(g_i, t_{k_i}) \tag{5}$$

本文分别将一种构造算法和一种启发式算法应用到式(5)求解中。Warburton^[10]比较了几种主要启发式算法在组合优化问题上的精度和运行时间表现,发现其中的退火算法在运行时间上略具优势,尤其是在待求序列数量增加时。因此,本文将模拟退火算法作为求解匹配问题近似解的启发式算法。模拟退火算法实现较为容易,许多文献^[9-15]都描述了其算法实现伪代码。本文的改进算法部分如下:

```
Initial hmax = 0 hmin = 0
for every point pi of g
Initial shortest = Inf
for every point pj of t
dij= distance (pi , pj )
```

```
if dij < shortest then
shortest = dij
if shortest > hmax then
hmax = shortest
if shortest < hmin then
hmin = shortest
```

对于构造算法,本文借鉴很早就应用在 TSP 问题中的最近邻算法(Nearest Neighbour Algorithm)^[16-18]的思路,通过距离矩阵排序后,依次抽取最短距离并删除距离矩阵中相关节点的方法来取得式(5)的近似解。

2 算法试验

本文选择实际工作中的 10 幅管线图纸,其中总计有 387 个管线节点和人工注记的节点编号,用此实例来验证算法的精度和有效性。本文算法均采用 AutoCAD 的二次开发语言 Visual Lisp 实现。

表 1 列出了几种方法在匹配上的准确率,其中基于标准 Hausdorff 距离的模拟退火算法在大部分情况下都无法收敛到全局最优,算法运行时间也很短,如图 3(a),因此匹配精度较差;改进最近邻算法的平均

匹配准确率为 81.5%;采用模拟退火算法的平均匹配准确率为 98.7%。无论是质心距离还是改进 Hausdorff 距离,模拟退火算法的准确度都要比单次迭代最近邻算法的准确度高,统计学检验指标 P 值小于 0.000 1,说明模拟退火算法较为显著有效。产生这种现象的主要原因是本文采用的单次迭代最近邻算法并不能得到式(5)的最优解,也从另外一个方面验证了人工标注的规律是采用最小距离和的原则,而模拟退火算法能较为准确地求得该问题的近似解。

基于改进 Hausdorff 距离算法较基于质心距离算法的图形注记匹配准确率分别提升了 3.9% 和 1.6%。在本文选择的试验样本中,2 种方法准确率差异的统计学检验指标 P 值分别为 0.037 和 0.027,均小于 0.05。这说明改进 Hausdorff 距离相对于质心距离在图形注记匹配准确率提升上具备一定的有效性。而与标准 Hausdorff 距离算法相比较,基于质心距离算法和基于改进 Hausdorff 距离算法的图形注记匹配准确率分别高出了 9% 和 12.9%,准确率差异的统计学指标 P 值均小于 0.000 01,说明这 2 种距离算法较未改进的 Hausdorff 距离算法显著有效。图 3 显示了几种算法匹配的实例。

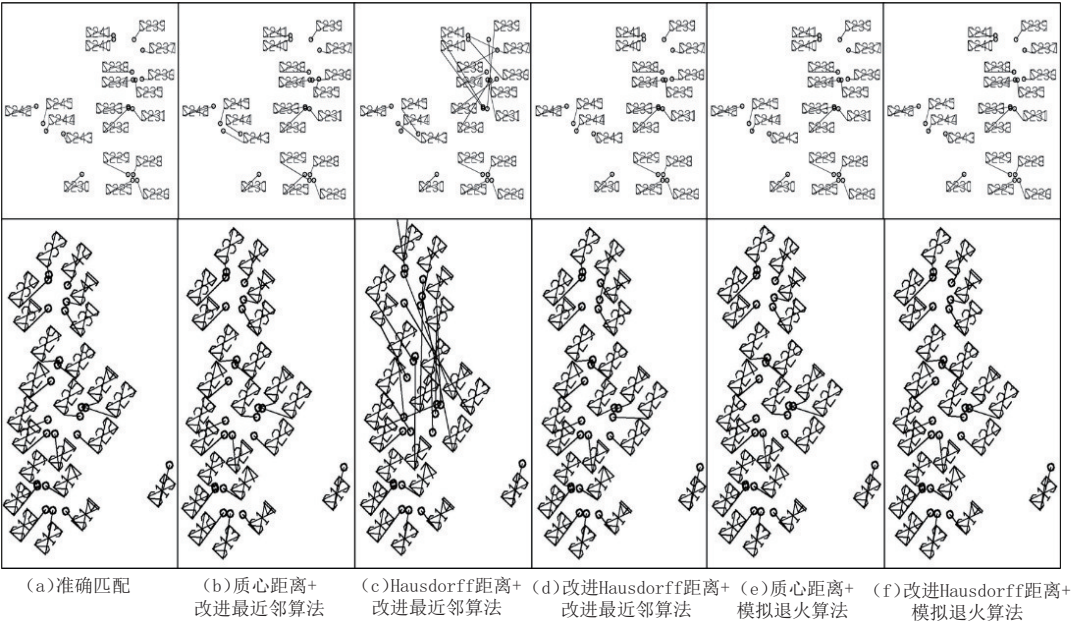


图 3 几种算法匹配结果对比图

另外,本文还在某竣工验收绿化面积图上,针对多线段绿化对象与面积注记匹配验证了本文算法的有效性,其匹配效果如图 4 所示。

几种算法运行时间对比图见图 5。由图 5 可知:

(1)虽然计算改进 Hausdorff 距离需要遍历对象轮廓点集,但并没有较质心距离显著耗时。例如,当

待匹配对象从 10 个提高到 200 个时,质心距离+最近邻算法、改进 Hausdorff 距离+最近邻算法的耗时分别从 32、47 ms 提高到了 562、610 ms。

(2)模拟退火算法是一种随机初始化的迭代求解算法,每次迭代均需要计算当前解关于距离矩阵的代价,其得到最优解依赖于初始值的选择,并不依

表 1 不同方法的图形注记匹配精度比较

匹配方法	总样本数量	准确匹配个数	错误匹配个数	准确率/%
质心距离+改进最近邻算法	387	322	65	83.2
Hausdorff 距离+改进最近邻算法	387	287	100	74.2
改进 Hausdorff 距离+改进最近邻算法	387	337	50	87.1
质心距离+模拟退火算法	387	379	8	97.9
Hausdorff 距离+模拟退火算法	387			
改进 Hausdorff 距离+模拟退火算法	387	385	2	99.5

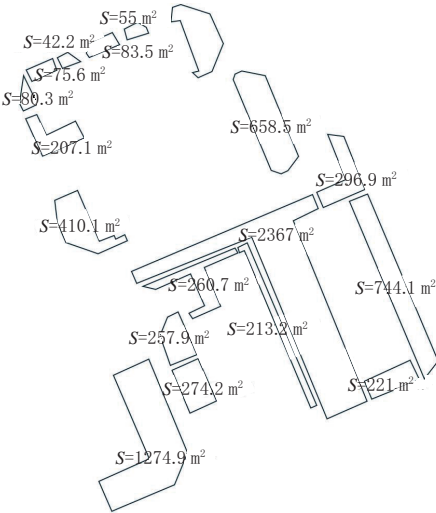


图 4 多线段对象与面积注记文本匹配图

赖于待匹配对象的数量。改进的最近邻算法只需要单次迭代,较模拟退火算法快很多,但其很难得到最优解,准确率也较差。

3 结 语

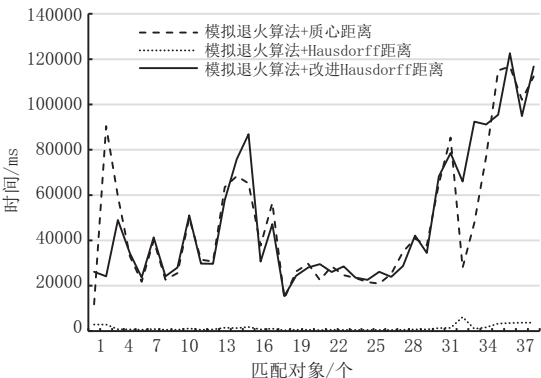
(1)针对图形与注记文本一对一匹配的情况,如管线图中管线节点对象与节点编号对象匹配的问题,转化为了2个集合对象距离和最短的组合优化问题。

(2)将改进 Hausdorff 距离替代质心距离后,提高了匹配精度;同时用实例比较了2种求解最短距离和的算法——改进最近邻算法和模拟退火算法的优劣。改进最邻近算法精度较低,速度较快,而模拟退火算法精度较高,但速度很慢。

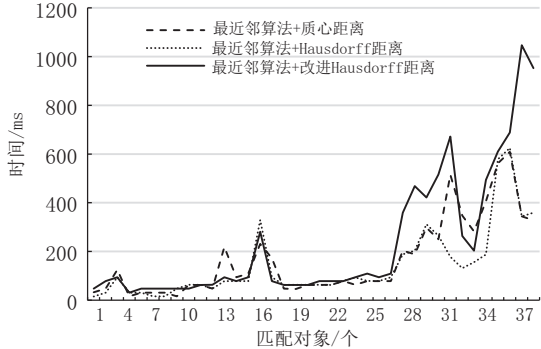
(3)图形与注记匹配是较为复杂的问题,本文算法设计未针对图形与文本注记存在一对多或多对一等其他情况,需要针对实际任务进行算法设计。

参考文献:

[1] 王元超,颜梅春,祁洪霞.从 AutoCAD 到 ARCGIS 快速转换的图斑注记高效匹配关键技术[J].西北水电,2017(167):45-47.



(a) 模拟退火算法+



(b) 最近邻算法+

图 5 几种算法运行时间对比图

[2] 王建华,李明.基于 AutoCAD 的 GIS 数据转换方法研究[J].地理信息科学,2018,20(3):45-50.

[3] 张伟,刘晓东.高程点与高程注记的匹配算法研究[J].测绘科学,2019,44(5):78-84.

[4] 李强,赵晨曦.基于 AutoCAD 的管线数据自动匹配技术[J].测绘工程,2020,29(2):12-18.

[5] 周鹏,孙立明.旅行商问题的启发式优化算法综述[J].计算机应用研究,2017,34(10):2158-2165.

[6] 陈建,王涛.遗传算法在 GIS 空间数据匹配中的应用研究[J].计算机与现代化,2016,12(4):35-40.

[7] 柏捷,顾有兵.高程点与高程注记的匹配及其应用[J].城市勘测,2007(5):89-90.

[8] 张区旺,刘晋虎.基于 VBA 的高程注记点匹配检查实现[J].煤炭工程,2013,45(8):135-136,139.

[9] CIGNONI P, ROCCHINI C, SCOPIGNO R. Metro: Measuring error on simplified surfaces [C]//Computer Graphics Forum. Oxford, UK and Boston, USA: Blackwell Publishers, 1998, 17(2): 167-174.

[10] WARBURTON A R. Worst-case analysis of some convex hull heuristics for the Euclidean travelling salesman problem[J]. Operations research letters, 1993, 13(1): 37-42.

[11] APPLEGATE D L, ROBERT E, BIXBY V C, et al. The traveling salesman problem: A computational study[M]. Princeton: Princeton University Press, 2006.

[12] SCHMIDT F R, BOYKOV Y. Hausdorff distance constraint for multi-surface segmentation[C]//European Conference on Computer Vision. Berlin, Heidelberg: Springer, 2012: 598-611.

[13] DUBUISSON M P, JAIN A K. A modified Hausdorff distance for object matching[C]//Proceedings of 12th International Conference on Pattern Recognition. IEEE, 1994: 566-568.