

地图注记质量评价模型的研究

樊红, 张祖勋, 杜道生

(武汉大学 测绘遥感信息工程国家重点实验室, 湖北 武汉 430079)

The Study of the Evaluation Model for Map Labeling

FAN Hong, ZHANG Zu-xun, DU Dao-sheng

(LIESMARS, Wuhan University, Wuhan 430079, China)

Abstract: According to some researches, map labeling can be viewed as a combinatorial optimal problem with more than one objectives. The key step of solving it is to define an abstract goal evaluation function, which can be used to compare the placing results and to obtain the labeling algorithms. For this purpose, a labelling quality evaluation model should be given out.

An evaluation model of high-quality map labelling is proposed for obtaining a value representing the quality. There are four aspects relevant to the labelling quality which we call the overlap for a label with label, the overlap for label with the background features, position optimal level and the definite attachment for a label to the feature, represented in evaluation model. The map labelling system MapLabel developed by the author uses this quality model to define the fitness function of Genetic Algorithm to guarantee the good quality of labelling.

The model presented in this paper just takes the most important criteria into consideration so it is still a rough model. The map labelling problem is far more complex than we realized. To improve the map labelling quality, this model can be extended to incorporate more criteria and more complex factors in the future.

Key words: map labeling; quality evaluation model

摘要: 抽象出冲突、压盖、位置优先级和关联性4个基本因素,通过建立评分体系,构造地图注记的4因素形式化质量评价模型。同时,介绍该质量评价模型在地图自动注记系统MapLabel中的应用情况和实验结果,说明此模型的合理性和有效性。

关键词: 地图注记; 质量评价模型

1 前言

从Yoeli进行自动地图注记研究^[1]算起,自动地名注记问题的研究已有三十多年历史,这一研究领域取得的理论成果已经证明地名注记自动配置问题是一个多目标优化问题,即满足一定的约束条件下,目标函数取最大(极大)或最小(极小)值的问题^[2]。优化问题需用优化算法来求

解,为此,需要建立合理而有效的目标评价函数。目标评价函数在爬山法和模拟退火法、神经网络算法、遗传算法中分别被称为目标函数、能量函数和适应度函数。其实质是对地图注记质量的一种定量表达,要建立如此的一种定量表达,其基础和前提便是建立合理的注记质量评价形式化系统或模型。

2 质量评价模型考虑的因素

高质量的地图注记应该遵循的一些准则可以概括为: 清晰性; 美观性; 和谐性; 无二义性; 符合阅读习惯; 不影响地图内容和提示要素的位置、方向、形状和范围等^[3]。它们应该如何定量地描述并在模型中进行表达是需要首要考虑的问题。

在提出的模型中, 选取 4 个独立(很少重叠)的因素: 冲突、压盖、位置优先级和要素注记关联性来表达注记质量, 这 4 个影响注记质量的重要因素, 不仅在概念上有区别, 而且评估它们的参数也彼此独立。

1. 冲突: 注记与注记的重叠称为冲突。在地图注记的配置中, 冲突是最严重的问题。地图设计、地图美观性等方面的缺点, 有时不可避免, 但不会引起信息传达的障碍, 可注记冲突却会妨碍信息的传播。

2. 压盖: 注记与要素的重叠称为压盖。注记不是孤立存在的, 它与要素有机地结合在一起。地形图包含丰富的地物类型, 一幅地形图对其地物分类有时可达 19 层之多。压盖分为两种情形: 第一类压盖是不允许的, 如点状要素和一些重要的交叉点(如道路交叉处、河流汇合处、居民地入口处等)的压盖; 第二类压盖是不可避免的, 如密集地图上居民地注记对道路的压盖。

3. 位置的优先级: 注记的备选位置在易读性和美观性方面存在差别, 无论是点状注记还是线状注记, 注记与被注记要素的位置关系有优劣之分。

4. 要素—注记关联性: 高质量的注记的另一个重要条件是注记和被注记要素的关联性清楚, 无二义。

3 注记质量评价模型

笔者提出的质量评价模型采用评分体系描述质量评价函数, 质量评价函数取值为整数, 范围为 0~99。为便于遗传算法等优化算法应用该模型,

$$E_{\text{冲突}}(L_i) = \begin{cases} 1 & (\forall j, 0 < j < n, j \neq i, du(L_i, L_j) > 0) \\ 0 & \text{其他} \end{cases} \quad (1)$$

3.2 压盖评价函数

压盖应该尽量避免、减缓。压盖评价函数除了对压盖行为本身进行评价, 同时需要对被压盖要素的重要性程度、压盖量的大小进行评价。根据应用要求的不同, 可以定义简单的压盖评价函

考虑适应度函数的要求, 取值愈大, 代表质量愈好, 取值愈小, 代表质量愈差。同时, 评价函数值只有相对意义, 没有绝对意义, 它的基本内容如下。

1. 分别定义和计算单个注记配置的冲突、压盖、位置优先、关联性 4 个评价函数;

2. 由单个注记配置的 4 个评价函数值复合成单个注记配置的质量评价函数;

3. 将每个注记配置的质量评价函数累加(或再取平均值)得到注记配置的总的质量评价函数。

为了表达的方便, 先定义一组谓词的含义如下。

1. $B(L_i)$ 表示注记 L_i 的区域, 一个注记矩形框或一组注记矩形框所占据的区域。其中, L_i 代表第 i 个注记。

2. $A(F_i)$ 表示要素 F_i 的区域, 其中, F_i 代表第 i 个要素。

3. $d_p(p_i, p_j)$ 表示点 p_i, p_j 的欧几里得距离。

4. $d_u(L_i, L_j)$ 表示注记 L_i, L_j 的距离, $d_u(L_i, L_j)$ 定义如下:

$$d_u(L_i, L_j) = \min\{d_p(p_i, p_j) \mid p_i \in B(L_i) \wedge p_j \in B(L_j)\}$$

5. $d_f(L_i, F_j)$ 表示注记 L_i 与要素 F_j 的距离, $d_f(L_i, F_j)$ 定义如下:

$$d_f(L_i, F_j) = \min\{d_p(p_i, p_j) \mid p_i \in B(L_i) \wedge p_j \in A(F_j)\}$$

3.1 冲突评价函数

冲突应该消除。当两个注记发生重叠时, 重叠量大小并不重要。所以, 冲突评价函数只对冲突发生的行为本身进行评价, 而并不度量压盖量的大小或区分是两个注记发生冲突还是多个注记同时发生冲突。

我们将冲突评价函数的取值范围定义为 $\{0, 1\}$, 即用 0, 1 来给冲突评分, 无冲突评分为 1, 有冲突评分为 0。该冲突评价函数反映了地图未发生冲突的注记的个数。称为 0-1 冲突评价函数。即

数或复杂的压盖评价函数。前者仅对被压盖要素的重要性等级加以考虑, 后者除了考虑被压盖要素的重要性等级之外, 同时兼顾压盖量的大小。

为此, 需要为背景要素定义重要性评价函数(其值称为重要性等级或重要性权重) $W(BF_i)$ 。

同样采用 0~ 99(取值为整数)的评分体系。对禁止压盖的要素评分为 99, 重要性程度越低, 评分越低, 最不重要的要素评分为 0。所以有

$$W(BF_i) = \begin{cases} 0 & \text{重要性最低} \\ 1 \sim 98 & \text{重要性居中} \\ 99 & \text{重要性最高} \end{cases}$$

最简单的压盖评价函数定义为被某注记所压

$$E_{\text{压盖}}(L_i, BF) = \begin{cases} 99 & \text{无压盖} \\ 99 - \max\{W(BF_i) | \text{overlap}(L_i, BF_i) \wedge BF_i \in BF\} & \text{有压盖} \end{cases} \quad (2)$$

复杂的压盖评价函数考虑被压盖要素的重要性程度和压盖量的大小。同样, 无压盖时, 评分为最高分 99 分, 有压盖时, 如果压盖要素的重要性

盖要素中的最高重要性权重, 它考虑压盖地物的重要性程度, 忽略压盖量的大小。无压盖时, 评分为最高分 99 分, 被压盖要素的重要性等级越高, 则认为压盖越严重, 压盖评分越低。令 $Area(R)$ 表示区域 R 的面积; $Overlap(O_1, O_2)$ 表示两个对象 O_1, O_2 重叠。定义如下

程度越高、压盖重要性要素的面积越大, 则认为压盖越严重, 压盖评分越低。为此, 可用下式表示压盖的严重程度:

$$\text{压盖严重程度指数}(L_i) = \sum_{j=1}^M (\text{要素 } BF_j \text{ 重要性等级} \times \text{对要素 } BF_j \text{ 的压盖量}) / L_i \text{ 的面积} =$$

$$\sum_{j=1}^M (\text{要素 } BF_j \text{ 重要性等级} \times (\text{对要素 } BF_j \text{ 的压盖量}) / L_i \text{ 的面积})$$

令对要素 BF_j 的相对压盖量 = 对要素 BF_j / L_i 的面积, 那么

$$\text{压盖严重程度指数}(L_i) = \sum_{j=1}^M (\text{要素 } BF_j \text{ 重要性等级} \times \text{对要素 } BF_j \text{ 的相对压盖量})$$

即

$$\text{压盖严重程度指数}(L_i) = \sum_{j=1}^M (W(BF_j) * \text{Area}(B(L_j) \wedge A(BF_j)) / \text{Area}(B(B_i)))$$

其中, $Overlap(L_i, BF_j) \cap BF_j \in BF, 0 < j < M$

压盖严重程度指数是介于 0~ 99 之间的一个整数, 较好地量化了注记压盖。压盖严重程度越

高, 压盖评分越低, 相反, 压盖评分越高。所以, 压盖评价函数定义为 99 与压盖严重程度指数之差。亦即

$$E_{\text{压盖}}(L_i) = E_{\text{压盖}}(L_i, BF) = \begin{cases} 99 - \text{压盖严重程度指数}(L_i) & \text{有压盖} \\ 99 & \text{无压盖} \end{cases} \quad (3)$$

为了后面引用的方便, 我们称式(2)为简单压盖评价函数, 式(3)为复杂压盖评价函数。

3.3 位置优先级模型

位置优先级模型很多, 这里, 提出如下的排序模型。当注记的备选位置是有限多个, 而且数目不大可以枚举出来, 例如点的 4 位置注记模式或 8 位置注记模式, 我们将其按优先级由高到低的顺序排列, 令 $Pos_j(L_i)$ 表示 L_i 的第 j 个注记位置, $Order(Pos_j(L_i))$, 表示备选位置排序后的序号, 可将位置评价函数定义为 99 与位置序号之差, 即优先级最高的位置评分最高为 99, 其余位置依序递减。我们称式(4)为排序位置评价函数

$$E_{\text{位置}}(L_i) = 99 - Order(Pos_j(L_i)) \quad (4)$$

3.4 注记-要素关联性评价函数

按照质量评价准则, 好的关联性应该满足 3 个条件:

1. 注记与关联要素的距离在一个最小距离 δ_{\min} 和最大距离 δ_{\max} 之间, 这个条件也保证注记和要素不会重叠。不同的注记类型, 这个距离值会有所不同, 因而该距离可以用函数如 $\delta_{\min}(L_i)$ 和 $\delta_{\max}(L_i)$ 来表示。
2. 注记与关联要素的距离应小于其他注记与该要素的距离。
3. 注记与关联要素的距离应小于该注记与其他同类要素的距离。

如果注记违反了上述任何一个条件, 它的关联性评分为最低分 0; 否则, 如果注记满足这些条件, 认为注记离要素越近越好。当距离为

$\delta_{\min}(L_i)$ 时, 得最高评分 99, 距离为 $\delta_{\max}(L_i)$ 时, 得最低评分 1, 介于两者之间, 与 δ_{\max} 相距越大,

得分越高。有

$$E_{\text{关联}}(L_i, F_i) = \begin{cases} 99(\delta_{\max}(L_i) - d_{\text{f}}(L_i, F_i)) / (\delta_{\max}(L_i) - \delta_{\min}(L_i)) & ((\delta_{\min}(L_i) \leq d_{\text{f}}(L_i, F_i) \leq \delta_{\max}(L_i)) \wedge (\forall k, 0 < k < n, k \neq i, d_{\text{f}}(L_i, F_i) < d_{\text{f}}(L_k, F_i)) \wedge (\forall k, 0 < l < m, l \neq j, d_{\text{f}}(L_i, F_i) < d_{\text{f}}(L_i, F_i))) \\ 0 & \text{其他} \end{cases} \quad (5)$$

3.5 多个评价函数的复合

优化算法要求使用目标函数或适应度函数来表示注记质量的优劣。在进行注记自动优化配置时, 通常考虑的因素不止一个, 需要有一种方式将这 4 个质量评价函数复合成 1 个总的质量评价函数(它将用作优化算法中的目标函数或适应度函数)。其中, 我们的模型采用的复合的方式是先将单个注记配置的各种质量评价价值通过乘上权重因子再相加得到一个总的质量评价价值, 再对单个注记的质量评价价值进行累加(必要时可取平均值)得到整个注记配置的评价价值(用浮点数表示)。即

$$E(L_i) = W_{\text{冲突}}E_{\text{冲突}}(L_i) + W_{\text{压盖}}E_{\text{压盖}}(L_i, BF) + W_{\text{位置}}E_{\text{位置}}(L_i) + W_{\text{关联性}}E_{\text{关联性}}(L_i)$$

$$E(L) = \sum_{i=1}^N E(L_i) / N$$

其中, 各个权重因子定义如下:

$W_{\text{冲突}}$ 注记与注记冲突因子, 定义注记冲突因素在整个注记评价方案中所占的比重。

$W_{\text{压盖}}$ 注记与要素压盖因子, 定义注记与要素压盖因素在整个注记评价方案中所占的比重。

$W_{\text{位置}}$ 注记位置优先级因子, 定义注记位置优先级因素在整个注记评价方案中所占的比重。

$W_{\text{关联性}}$ 注记关联性评价因子, 定义注记关联性因素在整个注记评价方案中所占的比重。

每个权重因子为一个实数, 每个权重因子根据各个因素在整个注记评价方案中所占的比重确定它们的数值。例如, 从经验我们知道, $W_{\text{冲突}} \gg W_{\text{压盖}} \gg W_{\text{位置}}$, 因为:

1. 没有任何重叠、但有其他缺陷的注记, 如没有遵守美学原则的注记是可以接受的, 即使注记放置得很好, 如果存在冲突, 这样的地图仍会被制图员拒绝。

2. 没有任何压盖、注记位置优先级较低的注记, 仍要好于优先级较高、但对地图要素有压盖的注记。

3. 通过调整各个权重因子可以适应不同情况的地图的注记。例如, 对于注记密度较低注记难度较小的地图, 适当提高 $W_{\text{冲突}}$ 之外其余几个权重; 对于注记密度较高注记难度较大的地图, 显著提高 $W_{\text{冲突}}$ 的权重。

4 MapLabel 中评价模型的实现

MapLabel 是作者开发的一个地图自动注记系统, 它通过规则库提供对要素重要性权重和位置优先级计算的支持, 采用遗传算法求取注记的整体最优解, 遗传算法的适应度函数采用上述函数模型定义, 通过大量与其他优化算法的对比实验证明了采用上述适应度函数的遗传算法取得了良好的注记结果, 有关遗传算法注记的实验及结果详见参考文献[4]。

为提高效率, MapLabel 对注记评价模型进行了一些简化, 它的质量评价函数考虑了冲突、压盖和位置优先级 3 个因素, 为此, 遗传算法适应度函数定义为:

$$W_{\text{冲突}} = 10\ 000.0; W_{\text{压盖}} = 100.0; W_{\text{位置}} = 1.0$$

$$E(L_i) = 10\ 000E_{\text{冲突}}(L_i) + 100E_{\text{压盖}}(L_i, BF) + E_{\text{位置}}(L_i)$$

$$E(L) = \sum_{i=1}^N E(L_i) / N$$

为了进一步提高计算效率, 实验中还将上述模型进一步简化。考虑冲突的作用时, 采用“一票否决”的方案, 定义当注记出现冲突时的评价价值为 0。当对无冲突注记进行评价时, 计算评价价值可以忽略评价函数中的冲突项(因均无冲突)。得适应度函数为:

$$E(L_i) = \begin{cases} 100 * E_{\text{压盖}}(L_i, BF) + 1 * E_{\text{位置}}(L_i) & \text{无冲突} \\ 0 & \text{其他} \end{cases}$$

$$E(L) = \sum_{i=1}^N E(L_i) / N$$

其中, N 为注记的总数。

由于关联性的评价函数计算量较大, 故 MapLabel 未将它包括在质量评价函数中, 而是采取下列算法方案来保证注记和要素的良好关联性:

1. MapLabel 通过定位参数表为每个注记定义 $\delta_{\min}(L_i)$ 和 $\delta_{\max}(L_i)$ 。通过算法强制注记与要素的距离落在 $\delta_{\min}(L_i)$ 和 $\delta_{\max}(L_i)$ 之间。

2. 从前面的分析中我们知道, 如果 (L_i, F_i) 关联性清楚, 则必须满足下列 2 个条件:

条件 1: $\forall k, 0 < k < n, k \neq i, d_f(L_k, F_i) > d_f(L_i, F_i)$

条件 2: $\forall l, 0 < l < m, l \neq i, d_f(L_i, F_l) > d_f(L_i, F_i)$

MapLabel 在算法中限定每个注记满足“与其他要素之间的距离都大于 δ_{\max} ”的条件, 这里所有其它要素特指所有同类的其它要素。其中,

$\delta_{\max} := \max\{\delta_{\max}(L_i) | 0 < i < n\}$

$\forall k, l, 0 < k < n, 0 < l < m, k \neq l, d_f(L_k, F_l) > \delta_{\max}$

当每个注记满足上述条件时, 也自然使 (L_i, F_i) 关联性清楚的条件 1, 2 能得到满足。

5 结束语

本文提出了一个考虑 4 种因素的注记质量评

价的形式化模型, 同时在作者开发的地图自动注记系统中, 采用利用该模型定义适应度函数的遗传算法来求解自动注记最优整体配置, 取得了很好的注记效果[4]。实际制图作业中, 影响注记质量的因素很多, 因素之间的关系也很复杂, 如何形式化地描述这些因素及其错综复杂关系以及它们对注记质量的影响以进一步完善本文所提模型, 是作者需要进一步研究和解决的问题。

参考文献:

- [1] DIJK S V, KREVELD M V, *et al.* Towards an Evaluation of Quality for Label Placement Methods[A]. Proceedings of the 19th International Cartographic Conference (ICA' 99) [C]. Ottawa: Int Cartographic Association, 1999. 905-913.
- [2] YOELI P. The Logic of Automated Map Lettering [J]. The Cartographic Journal, 1972, (9): 99-108.
- [3] MARKS J, SHIEBER S. The Computational Complexity of Cartographic Label Placement[R]. [s. l.]: Harvard, 1991.
- [4] FAN Hong, LIU Kai-jun, ZHANG Zu-xun. An Robust Genetic Algorithm for Automated Map Name Placement[J]. Geomatics and Information Science of Wuhan University, 2002, 27(6). (in Chinese)
- [5] FAN Hong. On Map Labeling[D]. Wuhan: Wuhan University, 2001. (in Chinese)

欢迎订阅《测绘通报》

《测绘通报》创刊于 1955 年, 是由国家测绘局主管, 中国地图出版社主办的反映我国测绘科技发展现状和指导全国测绘生产业务的综合性科技刊物。

《测绘通报》致力于宣传国家测绘科技方针、政策及法律、法规, 公布新的测绘科技成就, 传播测绘科技信息, 交流学术思想, 促进科技成果的商品化、产业化。其内容广, 信息全, 周期短, 受到了广大读者的信任和喜爱。

《测绘通报》设有: 学术研究、技术交流、测绘市场、国外测绘、行业管理、知识窗、企业之窗、测绘教学等栏目。主要内容包括: 大地测量、GPS、摄影测量、RS、地图制图、GIS、工程测量、矿山测量、地籍测绘、海洋测绘、测绘仪器、信息传输、行业管理、测绘教学; 计算机、通讯等相关理论技术在测绘领域里的应用; 国内外测绘学术动态及有关测绘科技信息。

《测绘通报》(月刊), 定价: 5.00 元, 邮发代号: 2-223。

编辑部地址: 北京复外三里河路 50 号, 邮编: 100045, 电话: 010-68531317(兼传真), 68531192(金英)。