Quality Evaluation Model for Map Labeling

FAN Hong ZHANG Zuxun DU Daosheng

ABSTRACT This paper discusses and sums up the basic criterions of guaranteeing the labeling quality and abstracts the four basic factors including the conflict for a label with a label, overlay for label with the features, position’s priority and the association for a label with its feature. By establishing the scoring system, a formalized four-factors quality evaluation model is constructed. Last, this paper introduces the experimental result of the quality evaluation model applied to the automatic map labeling system-MapLabel.

KEYWORDS map labeling; quality evaluation model

CLC NUMBER P283.7

Introduction

Since the research of Yoeli[1], the study on automatic map labeling problem has lasted more than thirty years, the theoretical achievements in this field has proved that the automatic map labeling is a multi-goal optimization problem belonging to NP-complete hard[2]. Generally, the automatic map labeling problem can be formalized as the objective function acquiring the maximum (extremely great) or minimum (extremely small) under certain restraint condition[1]. The optimization problem needs to be solved with optimization algorithm. In actual solving procedure, establishing a rational and effective object evaluation functions are the key step. The object evaluation functions are called objective function, energy function and adaptive function in climbing algorithm and simulated annealing, neural network algorithm and genetic algorithm, respectively. The essence of object evaluation functions is the quantitative expression of the map labeling quality. In order to set up such an expression, a rational and formalized evaluation system (model) of labeling quality is needed.

It is a difficult course to establish a rational evaluation model of labeling quality. Some scholars have done some relevant researches and made certain achievements already[2, 4-6]. This paper discusses and analyzes several important factors which influence the labeling quality then introduces a kind of practical formalized quality evaluation model of map labeling, explains its design foundation and rationality and introduces the application of this model to MapLabel.

1 Quality evaluation criterions

Before establishing the quality evaluation model, some criterions for quality evaluation should first be given. In general, for map labeling the handwork manner is still adopted, which needs to comply with a lot of labeling rules and criterions. Through analyzing these rules and criterions, we can conclude some criterions which should be complied with in high-quality map labeling[1, 4-6].

① Clarity: many factors will influence the clarity, such as the label’s style, label’s size, label’s color, conflict among labels, overlay of label to map feature, distance between labels,
label's density, interval and direction of dispersed labels.

2 Aesthetics: The label's font, label's line-style, polygon-label's shape and the labels' density should be fully considered.

3 Harmonious: A good map labeling practice should use uniform font. A label's color should be compatible with the color of map features, making the whole map harmonious.

4 Unambiguity: The association between the labels and their elements should be clear.

5 The reading habit: the label placement should accord with the reading habit as much as possible.

6 Not disturbing the content of maps: The label should avoid overlaying other elements,

Suggesting the position, direction, shape and range of the element; show an important city with bigger font to indicate its importance; the point label and the labeled point feature should lie in the same side of a road; great oceans and lakes should be labeled along the skeleton line in the plane; single-line-river should be labeled along its curve.

2 Quality evaluation model

2.1 Main factors influencing labeling quality

From the above analysis of the quality evaluation criterions we can know that many factors influence the labeling quality. Some of them can be depicted with formalized expression. In addition, some factors has formalized expression, but the expression is far from completeness and very complicated.

We choose four independent (seldom overlap) factors, including conflict, overlay, position's priority and the association between elements and their labels, to express the labeling quality. The model based on those factors are called as label evaluation model of conflict, overlay, position and association. The conflict, overlay, position's priority and the association between elements and their labels are the important factors influencing the labeling quality. Moreover, their concepts are different, and the evaluation parameters of the four factors are independent of each other.

1) Conflict. The overlay between labels is called conflict. In map labeling, the conflict is the most serious problem. Sometimes the shortcoming of map's design and aesthetics is inevitable, but this will not disturb the transmission of information. But the labels' conflict can hinder the transmission of information. The evaluation model will score by whether the labels are conflicted each other.

2) Overlay. The overlap between label and element is called overlay. The label is not isolated, and it must associate with its element closely. The topographic map includes abundant geographic feature types, sometimes the classifications of features in a topographic map can go up to more than 19 layers. The overlay is classified into two kinds: forbidden and inevitable. The evaluation model will score by the type of overlaid element and the overlay degree.

3) The position's priority. The candidate positions of label have a difference in legibility and beauty, the positions of labels and labeled elements are different in quality. The evaluation model will score the quality of the candidate positions of label.

4) The association between element and label. Another important condition of high-quality label is that the association between element and its label are clear and unambiguity. The evaluation model will score a mark for the quality of association.

Before introducing the quality evaluation model, we will demonstrate two things.

1) The quality evaluation model discussed here is used to evaluate the candidate label solution as the basis for selecting the optimal solution. Certainly, it can be used to compare the quality of different labels (not discuss here), too.

2) Express the automatic label problem with the eight-element model \{(UL, UF, BF, G, E, R, A, LP)\}. Among them \{(UL, UF, BF, R, E)\} is the input of automated labeling, and LP, namely the
label’s position, is the output of automated labeling. The meanings of all symbols are as follows: UL is the label; UF is the element to label; BF is the background elements; G is the reference graph for position; R is the label rule; E is the quality evaluation model; A is the optimization algorithm; LP is the label’s position. Then we introduce a symbol L to represent the positioned label or positioning label, \( L_i \) represents the jth candidate position of the ith label, and sometimes, \( L_i \) is abbreviated to \( L \), which represents the ith label; \( L \) represents the i-cycle labeling solution.

2.2 Quality evaluation model of label

The quality evaluation model adopts the scoring system to express the quality evaluation function. The value of quality evaluation function is an integer, and the value domain is 0-99. In order to apply some optimization algorithms, such as genetic algorithm, this model consider the demand of adaptive function. The bigger the value is, the better the quality is. Meanwhile the value of the evaluation function has only relative meanings, but has no absolute meaning. Now the steps to obtain the value are introduced below.

1) Separately define and calculate the four evaluation functions, including conflict, overlay, position priority, association, of individual label;
2) combine the four evaluation functions of individual label to obtain its quality evaluation function;
3) Sum the quality evaluation functions of all labels to get the total quality evaluation function of label disposition. Moreover, for the convenience of expression, we define the meanings of a group of predications as follows.

\[
\begin{align*}
&1. \quad B(L_i), \text{ it is the area of label } L_i, \text{ which represents the area of a label’s rectangular frame or a group of label’s rectangular frames.} \\
&2. \quad A(F_i), \text{ it is the area of element } F_i. \\
&3. \quad d_p(p_i, p_j), \text{ it represents the Euclidean distance between points } p_i, p_j, \text{ namely, } d_p(p_i, p_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \\
&4. \quad d_E(L_i, L_j), \text{ it represents the distance between labels } L_i, L_j. \\
&5. \quad d_q(L_i, F_j), \text{ it represents the distance between label } L_i \text{ and element } F_j. \\
&6. \quad d_g(L_i, L_j) = \min\{d_p(p_i, p_j) \mid p_i \in B(L_i) \land p_j \in B(L_j)\} \\
&7. \quad d_{qg}(L_i, F_j) = \min\{d_q(p_i, p_j) \mid p_i \in B(L_i) \land p_j \in A(F_j)\}
\end{align*}
\]

2.3 Conflict evaluation function

The conflict should be eliminated. When two labels are overlapping each other, the overlapping amount is unimportant. So the conflict evaluation function just evaluates the conflict itself, but not measures the overlapping amount or distinguishes whether the conflict happens in two or many labels simultaneously.

We define the value domain of the conflict evaluation function as \( \{0, 1\} \), namely marking the conflict with 0 and 1, giving a mark 1 for without conflict, a mark 0 for with conflict. Such a conflict evaluation function reflects the number of labels without conflict. It is called 0-1 conflict evaluation function. Namely

\[
E_{conf}(L_i) = \begin{cases} 
1, & \text{if } \forall j: 0 < j < n, j \neq i, d_g(L_i, L_j) > 0 \\
0, & \text{otherwise}
\end{cases}
\]  

2.4 Overlay evaluation function

The overlay should be avoided as much as possible. Besides evaluating the overlay itself, the overlay evaluation function needs to evaluate the importance of overlaid element and the size of overlay at the same time. According to the different demands in applications, we can define both simple overlay evaluation functions and complicated overlay evaluation functions. The former just consider the importance of overlaid elements, while the latter consider the size of overlay besides the importance of overlaid elements.

Therefore, we need to define an importance evaluation function \( W(BF_i) \) ( whose value is called the grade or weight of importance) for the background element. We adopt the 0-99 mark...
system (the mark is integer). The mark of the element which can not be overlaid is 99. The less the importance of the element is, the smaller its mark is, and the element with the minimal importance is marked by 0. Thus

\[ W(BF_i) = \begin{cases} 0, & \text{the minimal importance} \\ 1.98, & \text{the median importance} \\ 99, & \text{the maximal importance} \end{cases} \]

The simplest overlay evaluation function is defined as the maximal importance weight of the elements overlaid by some label, which considers the importance of overlaid feature, but neglects the size of overlay. When there is no overlay, the mark is 99. The greater the importance of overlaid elements is, the more serious the overlay is, and the fewer the mark is. Suppose \( \text{Area}(R) \) represents the area of region \( R \) and \( \text{Overlap}((O_i, O_j)) \) represents two objects overlapping each other. The definition is as follows.

\[ E_{\text{overlay}}(L, BF) = \begin{cases} 99, & \text{without overlay} \\ 99 - \max\{W(BF_i) \mid \text{Overlap}(L, BF_i) \land BF_i \in BF\}, & \text{with overlay} \end{cases} \]

The complicated overlay evaluation function considers the importance of overlaid element and the size of overlay. In the same way, the mark without overlay is 99. When there is overlay, the greater the importance of the overlaid element and the overlaid area of the element are, the more serious the overlay is, and the fewer the mark is. So we can represent the serious degree of overlay with the following expression:

\[
\text{SeriousIndexofOverlay}(L_i) = \sum_{j=1}^{M} \left( \left( \text{ImportanceLevel}(BF_j) \times \text{RelativeOverlayArea}(L, BF) \right) / \text{TotalArea}(L) \right)
\]

\[
= \sum_{j=1}^{M} \left( \left( \text{ImportanceLevel}(BF_j) \times \text{RelativeOverlayArea}(L, BF) \right) / \text{TotalArea}(L) \right)
\]

Let \( \text{RelativeOverlayArea}(L, BF) = \text{OverlapArea}(L, BF) / \text{TotalArea}(L) \), then

\[
\text{SeriousIndexofOverlay}(L_i) = \sum_{j=1}^{M} \left( \left( \text{ImportanceLevel}(BF_j) \times \text{RelativeOverlayArea}(L, BF) \right) / \text{TotalArea}(L) \right)
\]

The serious degree index of overlay is an integer between 0 and 99, which quantizes the overlay of labels well. The higher the serious degree is, the fewer the mark of overlay is. So we can define the overlay evaluation function as the difference between 99 and the serious degree index of overlay, namely

\[ E_{\text{overlay}}(L, BF) = \begin{cases} 99 - \text{SeriousIndexofOverlay}(L_i), & \text{without overlay} \\ 99, & \text{with overlay} \end{cases} \]

For the convenience of citing below, we call Eq. (2) the simple overlay evaluation function, and Eq. (3) the complicated overlay evaluation function. For example, suppose that the number of background elements is \( \text{NUM} \), the area of label is \( A \) grids, the number of the grids overlaying the \( i \)-th element is \( N_i \), the weight of the \( i \)-th element is \( \beta_i \), and then the overlay amount of labels is calculated with the following expression. The serious degree index of overlay elements overlaid by some label, which considers the importance of overlaid feature, but neglects the size of overlay. When there is no overlay, the mark is 99. The greater the importance of overlaid elements is, the more serious the overlay is, and the fewer the mark is. Suppose \( \text{Area}(R) \) represents the area of region \( R \) and \( \text{Overlap}((O_i, O_j)) \) represents two objects overlapping each other. The definition is as follows.

\[ E_{\text{overlay}}(L, BF) = \begin{cases} 99, & \text{without overlay} \\ 99 - \max\{W(BF_i) \mid \text{Overlap}(L, BF_i) \land BF_i \in BF\}, & \text{with overlay} \end{cases} \]

2.5 Position priority model

There are many kinds of position priority models. Here we introduce the following sorting model. When the candidate label positions are

\[ V = \sum_{i=1}^{\text{NUM}} N_i \cdot \beta_i / A; \quad E_{\text{overlay}}(L) = 99 - PV \]
limited and the number is not great and can be enumerated. For example, the four-position or eight-position label mode of point, we sort them according to the priority descending. Let \( \text{Pos}_j(L) \) represents the \( j \)th label position \( L \), \( \text{Order}(\text{Pos}_j(L)) \) represents the order number of candidate positions via sorting, we can define the position evaluation function as the difference between 99 and the serial number, namely the mark of the position with the highest priority is 99, and the mark of other positions decrease according to the order. We call the following expression (4) as the position sorting evaluation function. Then

\[
E_{\text{position}}(L) = 99 - \text{Order}(\text{Pos}_j(L)) \quad (4)
\]

### 2.6 Evaluation function of the association between labels and their elements

According to the quality evaluation criterion, a good association should meet the following three conditions.

1. The distance between label and element should fall into between a minimal distance \( \delta_{\text{min}} \) and a maximal distance \( \delta_{\text{max}} \). This condition guarantees that the label and the element do not overlap each other. As to different label types, this distance are different. We represent the distances with some functions, such as \( \delta_{\text{min}}(L) \) and \( \delta_{\text{max}}(L) \).

2. The distance between some label and its element should be less than the distance between other labels and the element.

3. The distance between some label and its element should be less than the distance between this label and other elements of the same category.

If a label does not meet any of the above conditions, its association gets a mark 0; otherwise the nearer the label is to the element, the better the association is. When the distance is \( \delta_{\text{max}}(L) \) the mark is 99, and when the distance is \( \delta_{\text{max}} \) the mark is 1.

### 2.7 Combination of several evaluation functions

The optimization algorithm demands to represent the labeling quality, with object function (or adaptive function). When optimizing the labeling automatically, according to the requirement of map-making and labeling quality, more than one factor generally needs to be considered. In this model four factors are considered. Thus the four quality evaluation functions need to be combined into a whole quality evaluation function in some way (it is used as the object function or adaptive function of optimization algorithm). One way is that one may first multiply the evaluation functions of every label by their weights and calculate the sum to get the total evaluation value, then sum up (if necessary, the average is ok) the quality evaluation values of all labels to get the evaluation value (expressed with float) of the whole label disposition.

Namely

\[
E(L) = W_{\text{conflict}}E_{\text{conflict}}(L) + W_{\text{overlay}}E_{\text{overlay}}(L, BF) + W_{\text{priority}}E_{\text{priority}}(L) + W_{\text{other}}E_{\text{other}}(L)
\]

\[
E(L) = \frac{1}{N} \sum_{i=1}^{N} E(L_i)
\]

In the above expression, the definition of each weight factor is as follows:

- \( W_{\text{conflict}} \), the conflict factor among labels, represents the weight of the conflict factor to the whole evaluation solution;
- \( W_{\text{overlay}} \), the overlay factor between label and element, represents the weight of the overlay factor to the whole evaluation solution;
- \( W_{\text{priority}} \), the position priority factor, represents
the weight of the priority factor to the whole evaluation solution; 
\( W_{\text{relativity}} \), the association factor, represents the weight of the association factor to the whole evaluation solution.

Each weight factor is a real number, which is determined by the proportion of the factor to the whole evaluation solution. For example, according to experience as follows, we know that 
\( W_{\text{conflict}} \gg W_{\text{overlay}} \gg W_{\text{position}} \).

1. The label without overlay has other deficiencies. If it has conflict, no matter how well the label disposition is, such a map will still be refused by the cartographer.
2. The label of low position priority without overlay is better than the one of higher priority with overlay.

2.8 Realization of the evaluation model in MapLabel

MapLabel is an automatic map label system developed by the authors. It offers the support of the weight of each element according to its importance and the mechanism of position priority, and adopts the above evaluation function model.

MapLabel simplifies the evaluation model. The quality evaluation function of MapLabel takes account of the three factors: conflict, overlay and position priority. The model used to combine the three factors is defined as follows.

\[
W_{\text{conflict}} = 10000.0; W_{\text{overlay}} = 100.0; W_{\text{position}} = 1.0;
\]

\[
E(L_i) = 10000 \cdot E_{\text{conflict}}(L_i) + 100 \cdot E_{\text{overlay}}(L_i, BF) + E_{\text{position}}(L_i); \]

\[
E(L) = \sum_{i=1}^{N} E(L_i)/N, \text{ where } N \text{ is the number of labels.}
\]

For the convenience of calculation, MapLabel simplifies the above model further. For conflict such a scheme is adopted: defining the evaluation value with the conflict equaling to 0. When there is no conflict, the evaluation value does not contain the conflict item.

2.9 Evaluation of association

Because the calculation amount of association evaluation function is relatively enormous, the quality evaluation function of MapLabel does not contain it, and adopt the following algorithm scheme to guarantee good association among labels and their elements.

1. MapLabel defines \( \delta_{\text{min}}(L_i) \) and \( \delta_{\text{max}}(L_i) \) for each label by position parameter table, and constraints that the distance between label and element is between \( \delta_{\text{min}}(L_i) \) and \( \delta_{\text{max}}(L_i) \).
2. From the above analysis, we know that if the association \( (L_i, F_i) \) is clear, it must meet the following conditions.

condition 1:
\[
\forall k, 0 < k > n, k \neq i, d_{ij}(L_k, F_i) > d_{ij}(L_i, F_i)
\]

condition 2:
\[
\forall l, 0 < l < m, l \neq i, d_{ij}(L_l, F_i) > d_{ij}(L_i, F_i).
\]

MapLabel constrains each label to meet the condition “the distances from all other elements are larger than \( \delta_{\text{max}} \) in the algorithm”, here all other elements are referred to other elements of the same category. Among them,

\[
\delta_{\text{max}} = \max \{ \delta_{\text{max}}(L_i) \mid 0 < i < n \}
\]

\[
\forall k, l, 0 < k > n, 0 < l < m, k \neq l, d_{ij}(L_k, F_l) > \delta_{\text{max}}
\]

When each label meets the above condition, the two conditions for guaranteeing association of \( (L_i, F_i) \) clear are tenable, namely

condition 1:
\[
\forall i \forall k, 0 < k < n, k \neq i, d_{ij}(L_k, F_i) > d_{ij}(L_i, F_i)
\]

condition 2:
\[
\forall i \forall l, 0 < l < m, l \neq i, d_{ij}(L_l, F_i) > d_{ij}(L_i, F_i).
\]

3 Conclusions

This paper summarizes the criterions of good
labels and puts forward a formalized quality evaluation model considering the conflict, overlay, position priority and label-element association, then implements the model in Maplabel and obtains a good result (Fig. 1). In actual plotting, there are many factors influencing the label quality, and the relations among them are very complicated. In order to reflect the label quality with a model more accurately, more factors need to be considered.

Fig. 1  Label result of the quality evaluation model in MapLabel

REFERENCES

1 Yoeli P (1972) The logic of automated map lettering. *The Cartographic Journal*, 9, 99-108
2 Marks J., Shieber S (1991) The computational complexity of cartographic label placement. Technical Report TR-05-91, Harvard CS.
3 Fan H (2001) On automatic map labeling, [Ph. D dissertation], Wuhan, Wuhan University. (in Chinese)
4 van Dijk S., van Kreveld M., Strijk T., et al. (1999) Towards an evaluation of quality for label placement methods. The 19th International Cartographic Conference (ICA’99), Ottawa.
5 Imhof E (1975) Positioning names on maps. *The American Cartographer*, 2(2):128-144
6 Robinson A H (1989) The principle of maps (fifth edition). Beijing: Publishing House of Surveying and Mapping.