The Study on the quality Evaluation Model for Map Labeling

Fan Hong Zhang zuxun Du Daosheng

(The State key Lab. of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, No.129 Luoyu Road, Wuhan, China, Zip code:430079)

Abstract-the quality evaluation model for map labeling is the quantitative evaluation system for the optimization goal of labeling quality, which is set up in the course of automatic labeling disposition. A rational evaluation system is the foundation and premise of the automatic labeling disposition with high quality. This article discusses and sums up the basic criterions of guaranteeing the labeling quality, and on this basis author further abstracts the four basic factors including the conflict, overlay, position's priority and relativity, through establishing the mark system, construct the formalized four-factors quality evaluation model, meanwhile, author introduces this the application and experimental result of this quality evaluation model in the automatic map labeling system-MapLabel, thus prove the rationality and validity of this model.

Keywords Map Labeling, Quality Evaluation Model

Ⅰ.Introduction

Since the research of Yoeli [2], the study on automatic toponym labeling question has passed more than thirty years, the theory achievements in this field has proved the automatic toponym labeling question is a multi-goal optimization question belonging to NP-complete [5]. Generally the automatic toponym labeling question can be formalized expressed the objective function acquire the maximum (extremely great) or minimum (extremely little) under certain restraint condition [9], the optimization question need to solving with optimization algorithm. In actual solution, establishing a rational and effective object evaluation function is the key procedure. The object evaluation function is 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 function is the quantitative expression of the map labeling quality, in order to set up such a expression, need to establish rational the formalized evaluation system (model) of labeling quality.

It is a difficult course to establish a rational evaluation model of labeling quality., Some domestic and international scholars have done some relevant research and made certain achievement already [1,2,3]. This article discusses and analyzes several important factors which influence the labeling quality at first, and then introduce a kind of practical formalized quality evaluation model of map labeling put forward by the author, explain its design foundation and rationality, finally introduce the application of this model in MapLabel.

Ⅱ.the quality evaluation criterions

Before establishing the quality evaluation model should first some criterions for quality evaluation. Generally the map labeling work still adopts handwork manner, which need 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, 2, 3, 6, 7].

(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 surface feature, distance between labels, label's density, interval and direction of dispersed labels.

(2)Beautiful: Should consider the label's font, label's line-style, polygon-label's shape and the labels' density fully.

(3)Harmonious: The good map disposition should use uniform font. The label' color should be compatible with the color of surface features, make the whole map harmonious.

(4)Without ambiguity: the relativities between the labels and elements should be clear.

(5)Accord with the reading habit: the labels disposition should accord with the reading habit as much as possible, for example the Chinese reading habit is from left to right, so should label from to right horizontally, when label vertically or along a curve, should arrange from top to bottom.

(6)Do not influence the content of maps: the label should avoid overlaying other elements, when the overlay is inevitable, overlay those unimportant elements.

(7)Highlight the position, direction, shape and range of the element: show important city with bigger font to indicate its importance; the point label and the labeled point feature should lies in the same side of border, road; great ocean and lake should be labeled along the skeleton line in the plane; single-line-river should be labeled along its curve, when necessary, label in subsections repeatedly.

Ⅲ.the quality evaluation model

3.1the main factors of influencing the label's quality

from the above analysis about the quality evaluation criterions we can know that many factors influence the label disposition's quality, some of them can be depicted with formalized expression, some factors are difficult to depict in 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 relativity between elements' labels, to express the label's quality, and call the model based on this as the label evaluation model of conflict, overlay, position and relativity. The conflict, overlay, position's priority and the relativity between elements' labels are the important factors of influencing the label's quality, not only their concepts are different, but also the evaluation parameters of the four factors are independent of each other. Now discuss them respectively.

(1)Conflict: the overlay between labels is called conflict. In map label disposition, the conflict is the most serious problem. Sometimes the shortcoming on map's design and beauty 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 mark as to whether the labels are conflicted each other.

(2) Overlay: the overlap between label and element is called overlay. The label is isolated, and it must link with element close. The topographic map includes abundant surface feature types, sometimes the classifications of surface feature in a topographic map can up to more than 19 layers [10]. The overlay is classified into two kinds: the first kind of overlay is forbidden, for example the overlay to point elements and some important intersections (such as the intersection of roads, junction of rivers, entrance to resident area, etc.); the second kind overlay is inevitable, such as the overlay of the residences' labels to road in intensive map. The evaluation model will mark as to the type of overlaid element and the overlay degree.

(3) The position's priority: the backup positions of label have difference in legibility and beauty, no matter point label or line label, the positions of labels and labeled elements are different

in suitability. The evaluation model will mark as to the suitability of the backup positions of label.

(4) The relativity between element and label: another important condition of high-quality label is that the relativity between element and label are clear without ambiguity. The evaluation model will mark as to the quality of relativity.

Before introduce the quality evaluation model, illuminate two things.

1. The quality evaluation model discussed here is used to mark the backup label solution as the basis of selection. Certainly, it can be used to compare the quality of different labels (not discuss here) too.

2. Express the automatic label question [9] with the eight-member model $\{UL, UF, BF, G, E, R, A, LP\}$ put forward by author. Among them $\{UL, UF, BF, R, E\}$ is the input of automatic label, and LP is the output of automatic label, namely the label's position. 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. Further introduce a mark L to represent the positioned label or positioning label, $L_{i,j}$ represents the i-th label of the j-th backup position, sometimes $L_{i,j}$ is abbreviated to L_i, which represents the i-th label; L^i represents the i-th label solution (solve).

3.2the quality evaluation model of label

The quality evaluation model put forward by author adopts the mark system to express the quality evaluation function. The value of quality evaluation function is an integer, and the value domain is 0~99. To apply some optimization algorithms, such as genetic algorithm, in this model, consider the demand of adaptive function, the bigger the value is, the better the quality is, and the smaller the value is, the worse better the quality is. Meanwhile the value of evaluation function has only relative meanings, but has no absolute meaning. Now introduce its basic content.

(1) Separately define and calculate the four evaluation functions, including conflict, overlay, position priority, relativity, of individual label;

(2) compose 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.

Now discuss them separately.

For the convenience of expression, first define the meanings of a group of predications as follows.

(1) $B(L_i)$: The area of label L_i, represents the area of a label's rectangular frame or a group of label's rectangular frames.

(2) $A(F_i)$: The area of element F_i.

(3) $d_p(p_i, p_j)$ represents the Euclidean distance between points p_i, p_j, namely

$$d_p(p_i, p_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

(4) $d_{ll}(L_i, L_j)$ represents the distance between labels L_i, L_j, the definition of $d_{ll}(L_i, L_j)$ is as follows:

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

(5) $d_{lf}(L_i, F_j)$ represents the distance between label L_i and element F_j, the definition of $d_{lf}(l_i, f_j)$ is as follows:

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

3.3the 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 only 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 mark the conflict with 0 and 1, the mark without conflict is 1, the mark with conflict is 0. Such a conflict evaluation function reflects the number of labels which are not with conflict. It is called 0-1 conflict evaluation function. Namely

$$E_{\text{conflict}}(L_i) = \begin{cases} 1 & \text{if } (\forall j, 0 < j < n, j \neq i, d_{ll}(L_i, L_j) > 0) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

3.4the 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, can define simple overlay evaluation function or complicated overlay evaluation function. The former just only considers the importance of overlaid element, the latter to keep and overlay the former importance grade of key element consider, the latter considers the size of overlay besides the importance of overlaid elements.

Therefore need to define importance evaluation function $W(BF_i)$ (its value is called the grade or weight of importance) for background element. Adopt the 0～99 mark system (the mark is integer). The mark of element which can’t be overlaid is 99, the little the importance of element is, the smaller its mark is, and the element with the minimal importance is marked 0. So

$$W(BF_i) = \begin{cases} 0 & \text{the minimal importance} \\ 1 \sim 98 & \text{the median importance} \\ 99 & \text{the maximal importance} \end{cases}$$

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 surface feature, but neglects the size of overlay. When there is no overlay, the mark is 99. The greater the importance of overlaid elements, the more serious the overlay is, and the fewer the mark is. Suppose *Area (R)* represents the area of region R and $Overlap(O_1, O_2)$ represents two objects overlap each other. The definition is as follows.

$$E_{压盖}(L_i, BF) = \begin{cases} 99 & \text{without overlay} \\ 99 - \max\{W(BF_i) \mid overlap(L_i, BF_i) \wedge BF_i \in BF\} & \text{with overlay} \end{cases}$$

(2)

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 overlaid element and the overlaid area of element are, the more serious the overlay is, and the fewer the mark is. So represent the serious degree of overlay with the following expression:

$$TheSevious\ \ IndexofOve\ \ rlay\left(L_i\right)$$
$$= \sum_{j=1}^{M} ((IMPORTANCE\ \ Level\ (BF_j) \times OvelayArea\ \ (L_i, BF_j)) / TotalArea\ \ (L_i)$$
$$= \sum_{j=1}^{M} ((IMPORTANCE\ \ Level\ (BF_j) \times (OvelayArea\ \ (L_i, BF_j) / TotalArea\ \ (L_i))$$
$$Let\quad RELATIVEOv\ \ elayArea\ \ (L_i, BF_j) = OvelayArea\ \ (L_i, BF_j) / TotalArea\ \ (L_i),$$
$$then$$
$$TheSevious\ \ IndexofOve\ \ rlay\left(L_i\right)$$
$$= \sum_{j=1}^{M} (IMPORTANCE\ \ Level\ (BF_j) \times RELATIVEOv\ \ elayArea\ \ (L_i, BF_j) \quad , \quad So$$
$$TheSevious\ \ IndexofOve\ \ rlay\left(L_i\right) = \sum_{j=1}^{M} (W\left(BF_j\right) * Area\left(B\left(L_j\right) \wedge A\left(BF_j\right)\right) / Area\left(B\left(L_i\right)\right))$$
$$having\quad Overlap\ \ (L_i, BF_j) \cap BF_j \in BF, 0 < j < M$$

The serious degree index of overlay is an integer between 0 and 99, which quantizes the overlay of labels well. The high the serious degree is, the fewer the mark of overlay is, contrarily, the great the mark of overlay is. So define the overlay evaluation function as the difference between 99 and the serious degree index of overlay. Namely

$$E_{overlay}(L_i) = E_{overlay}(L_i, BF) = \begin{cases} 99 - \text{the serious degree index of overlay}\left(L_i\right) & \text{with overlay} \\ 99 & \text{without overlay} \end{cases} \qquad (3)$$

For the convenience of citing behind, we called (2) as the simple overlay evaluation function, and call (3) as the complicated overlay evaluation function. For example suppose the number of background elements is 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 β_i, the overlay amount of labels is calculated with the following expression. The serious degree index of overlay PV equals the total overlay amount V is divided by the total area of labels A. So

$$V = \sum_{i=1}^{NUN} N_i * \beta_i; \qquad PV = \sum_{i=1}^{NUM} N_i * B_i / A;$$
$$E_{Overlay}(L_i) = 99 - PV$$

3.5 The position priority model

There are many kinds of position priority models, here introduce the following sort model.

When the backup label positions are 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 descending of priority. Let $Pos_j(L_i)$ represents the j-th label position L_i , $Order(Pos_j(L_i))$ represents the serial number of backup positions via sort, 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 expression 4 as the position sort evaluation function. Then

$$E_{\text{position}}(L_i) = 99 - Order(Pos_j(L_i)) \qquad (4)$$

3.6 The evaluation function of the relativities between labels and elements

According to the quality evaluation criterion, a good relativity should meet the following three conditions:

(1) The distance between label and element should be between a minimal distance $\delta_{\min}$ and a maximal distance $\delta_{\max}$, this condition guarantees the label and element don't overlap each other. As to different label types, this distance are different, now represent the distances with functions, such as $\delta_{\min}(L_i)$ and $\delta_{\min}(L_i)$.

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

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

If a label disobeys any of the above conditions, its relativity mark is 0; otherwise the closer the label is to the element, the more well. When the distance is $\delta_{\min}(L_i)$ the mark is 99 and when the distance is $\delta_{\max}$ the mark is 1. When the distance between the two values, the bigger the difference between it and $\delta_{\max}$ is, the bigger the mark is. So

$$E_{\text{relativity}}(L_i, F_i) = \begin{cases} 99*(\delta_{\max}(L_i) - d_{lf}(L_i, F_i)) \quad if((\delta_{\min}(L_i) \le d_{lf}(L_i, F_i) \le \delta_{\max}(L_i)) \wedge \\ /(\delta_{\max}(L_i) - \delta_{\min}(L_i)) (\forall k, 0 < k < n, k \ne i, d_{lf}(L_i, F_i) < d_{lf}(L_k, F_i)) \wedge \\ \qquad (\forall k, 0 < l < m, l \ne j, d_{lf}(L_i, F_i) < d_{lf}(L_i, F_l)) \\ 0 \qquad \text{otherwise} \end{cases} \qquad (5)$$

3.7 The combination of several evaluation functions

The optimization algorithm demands to represent the label' quality with object function (or adaptive function). At the time of optimizing the label disposition automatically, according to the requirement of map-making and quality, generally need to consider more than one factor, in our model we consider four factors. So need to combine the four quality evaluation functions into a whole quality evaluation function in some way (it is used as the object function or adaptive function of optimization algorithm). One way of combining is that: first multiply the evaluation functions of every label by their weights and calculate the sum to get the total evaluation value, an then sum (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

$$E(L_i) = W_{\text{conflict}} E_{\text{conflict}}(L_i) + W_{\text{overlay}} E_{\text{overlay}}(L_i, BF) + W_{\text{position}} E_{\text{position}}(L_i) + W_{\text{relativity}} E_{\text{relativity}}(L_i)$$

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

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

W_{conflict} The conflict factor among labels: represent the weight of the conflict factor to the whole evaluation solution.

W_{overlay} The overlay factor between label and element: represent the weight of the overlay factor to the whole evaluation solution.

W_{position} The position priority factor: represent the weight of the priority factor to the whole evaluation solution.

$W_{\text{relativity}}$ The relativity factor: represent the weight of the relativity 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, we know that $W_{\text{conflict}} >> W_{\text{overlay}} >> W_{\text{position}}$, because:

(1) The label without overlay has other deficiencies, such as disobeys the aesthetic principle, it is acceptable. If it has conflict, whatever well the label disposition is, such a map still will be refused drawer.

(2) The label of low position priority without overlay is better than the one of higher priority with overlay.

3.8 The realization of the evaluation model in MapLabel

MapLabel is an automatic map label system developed by author. It offers the support of the weight of element's importance and the mechanism of position priority, and adopts the above evaluation function model, in order to improve the efficiency of realization author adjust and modify some contents.

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

$$W_{\text{conflict}} = 10000.0; \quad W_{\text{overlay}} = 100.0; \quad W_{\text{position}} = 1.0;$$

$$E(L_i) = 10000 * E_{\text{conflict}}(L_i) + 100 * E_{\text{conflict}}(L_i, BF) + E_{\text{position}}(L_i);$$

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

For the convenience of calculation, MapLabel simplifies the above model further. About conflict adopt such a scheme: define the evaluation value with conflict as 0. When there is no conflict, the evaluation value doesn't contain the conflict item (because no conflict). So
In the above expression, N is the number of labels.
3. The treatment of relativity

$$E(L_i)=\begin{cases}100*E_{\mathrm{overlay}}(L_i,BF)+1*E_{\mathrm{position}}(L_i) & \text{without conflict}\\ 0 & \text{otherwise}\end{cases}$$

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

Because the calculation amount of relativity evaluation function is relatively great, the quality evaluation function of MapLabel doesn't contain it, and adopt the following algorithm scheme to guarantee the well relativities among labels and elements:

(1)MapLabel defines $\delta_{\min}(L_i)$ and $\delta_{\max}(L_i)$ for each label by position parameter table, constraint the distance of label and element is between $\delta_{\min}(L_i)$ and $\delta_{\max}(L_i)$.

(2) From the above analysis, we know that if the relativity (Li, Fi) is clear, it must meet the following conditions:

$$\text{condition one}: \forall k, 0<k<n, k\neq i, d_{lf}(L_k,F_i)>d_{lf}(L_i,F_i)$$
$$\text{condition two}: \forall l, 0<l<m, l\neq i, d_{lf}(L_i,F_l)>d_{lf}(L_i,F_i)$$

MapLabel constraint each label meet the condition of "the distances from all other elements are greater than $\delta_{\max}$ in algorithm", here the all other elements is referred to other elements of the same category. Among them,

$$\delta_{\max}:=\max\{\delta_{\max}(L_i)\mid 0<i<n\}$$
$$\forall k,l,0<k<n,0<l<m,k\neq l,d_{if}(L_k,F_l)>\delta_{\max}$$

When each label meets the above condition, the two conditions of guaranteeing (L_i,F_i) clear are tenable. Namely

$$\forall k,l,0<k<n,0<l<m,k\neq l,d_{if}(L_k,F_l)>\delta_{\max}$$
$$\Downarrow$$
$$\text{condition one}: \forall i \quad (\forall k,0<k<n,k\neq i,d_{lf}(L_k,F_i)>d_{lf}(L_i,F_i))$$
$$\text{condition two}: \forall i \quad (\forall l,0<l<m,l\neq i,d_{lf}(L_i,F_l)>d_{lf}(L_i,F_i))$$

Ⅳ.conclusion

This article summarizes the criterions of well labels, and put forward a formalized quality evaluation model considering the conflict, overlay, position priority and label-element relativity, at the same time realize the model in Maplabel and get a well result (figure 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, need to consider more factors. How to express these factors and the complicated relations among them in order to perfect the model in this article further, this is a question need to study further.

Ⅴ. Thanks

The study about this thesis is funded by state natural science of this gets the fund project of national natural fund (serial number 40001019) and open research fund (serial number(99) 0101)of State Key Laboratory of Information Engineering in Surveying of Wuhan University.

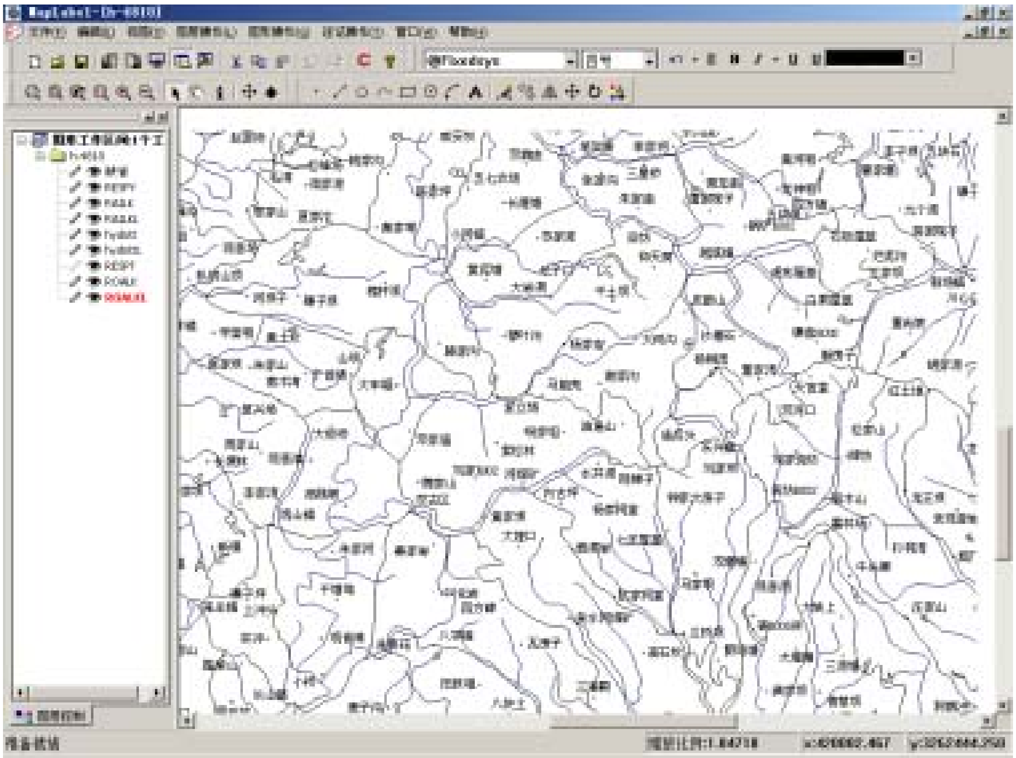

Figure 1 the label result of the quality evaluation model in MapLabel

References:

[1] Steven van Dijk, Marc van Kreveld, Tycho Strijk, and Alexander Wolff. Towards an evaluation of quality for label placement methods. In Proceedings of the 19th International Cartographic Conference (ICA'99), pages 905-913, Ottawa, 14-21 August 1999. Int. Cartographic Association.

[2] Pinhas Yoeli. The logic of automated map lettering. The Cartographic Journal, 9:99-108, 1972.

[3] Imhof, E. " Positioning Names on Maps. The American Cartographer, vol 2., No.2, pp 128-144,1975.

[4]Ingo Petzold, Lutz Plümer, and Markus Heber. Label placement for dynamically generated screen maps. In Proceedings of the 19th International Cartographic Conference (ICA'99), pages 893-903, 1999.

[5] Joe Marks and Stuart Shieber. The computational complexity of cartographic label placement. Technical Report TR-05-91, Harvard CS, 1991.

[6] ESRI,1998,Maplex——Automatic Cartographic Name Placement Software.

[7] A H Ronbinson，1989，The principle of 地图学原理，The Fifth Edition，The Press ofSurveying and Mapping.

［8］Fan Hong, Du Daosheng, Zhang Zuxun, Zhang Jianqing. the regulations and strate-gies for map name placement, Jour. of Wuhan Technical Univ. of Surveying and Mapping,vol.24, No.2, 1999,p154-157

[9] Fan Hong，On Automatic Map Labeling，2001，Dissertation，Wuhan University

[10] The Schema1 of 1：25 topogaphicm map geodatabase，1998，the National Bureau of Surveying and Mapping