A method for partitioning very small targets that accounts for crossing point constraints

Yong Yin
Shandong University of Science and Technology. Geomatics Colleget
Qingdao, Shandong, China, 266590
yinyong@casm.ac.cn

Wei Wu
Chinese Academy of Surveying and Mapping
Beijing, China, 100830
weiwu@casm.ac.cn.

Abstract

Very small targets (VSTs) are common elements of national geographical condition data, and the integration of these targets directly affects the quality of results synthesized from these data. Most conventional methods use amalgamation or aggregation to merge VSTs with their proximal patches, but these approaches tend to neglect the competitiveness of each proximal patch. To address this gap, we propose a method of partitioning VSTs that accounts for crossing point constraints. We first analyze how surface area, semantic proximity, length of shared edges, and regional importance affect the splitting ability of a proximal patch. Then, we use the analytic hierarchy process to construct a hierarchical model of these factors, in which the weights of each factor are calculated. Finally, a comprehensive assessment of the splitting ability of each proximal element is performed, and the skeletons of VSTs are amended accordingly, thus realizing the partitioning of VSTs. The viability and effectiveness of the method proposed in this work is validated in experiments using real data.

1 Introduction

The increasingly intricate usage of national geographical condition survey data has led to the integration of large-scale national geographical condition patches with various smaller-scaled data. In particular, the demand for data that provides a stratified and holistic reflection of regional geography and space usage is growing rapidly. Very small targets (VSTs) refer to fragmented patches in national geographical condition data that are morphologically complex, presenting elongated, circular, or irregular polygonal shapes. They are among the most important elements of national geographical condition data, due to their numerical abundance and widespread distribution, despite being individually small in area and sporadic in distribution. Therefore, the integration of VSTs directly impacts the quality of results derived from the data.

Presently, integration of VSTs is usually performed via aggregation or amalgamation, in which distance thresholds are defined for the aggregation of VSTs into patches within the distance threshold, which are semantically proximal. VSTs that do not possess semantically proximal patches within the distance threshold are, instead, amalgamated within topologically proximal patches. There are two scenarios that are relevant for amalgamation. If the very small target only has a single topologically proximal patch within the distance threshold, it is then directly amalgamated within that patch. If there are multiple topologically proximal patches within the distance threshold, the very small target is amalgamated in a partitioned manner via the construction of a skeleton. Therefore, conventional methods provide a holistic account for the effects of spatial distance, semantic similarity maps, and topological relationships about the integration of very small patches. Thus, these methods are capable of rationally maximizing the merging and removal of very small patches, while ensuring full patch coverage without overlaps, before and after the merging of patches. The analytic hierarchy process (AHP) is a systematic and hierarchical analytical method that accounts for quantitative and qualitative properties. Based on the aforementioned considerations, we incorporate AHP in the process of partitioning and amalgamating VSTs by constructing a hierarchical model for determining the weights of each factor that influences the splitting process. These weights are then used to amend the partitioning skeletons, thus achieving the partitioning and integration of VSTs.

2 Definition of VSTs

VSTs refer to patches with surface areas smaller than the smallest area delineated in a map; they are distributed widely among the patches. Table 1 lists the smallest delineated areas for each type of land in maps exceeding 1:500,000 scale, per the technical specifications of the recent national geographical conditions survey. We use this survey data as the judging standard for VSTs. The judging standards for VSTs in other map scales are adjusted per Table 1.

Table 1: Smallest delineated areas on a map

Land type	Map area (mm^2) (Map scales larger than 1:500000)	Actual surface area corresponding to the smallest patches (m^2)
Farmland	4	400
Parkland	4	400
Forested land	4	Large areas: 1,600 Others: 400
Grasslands	4	Grassland areas: 1,600 Others: 400
Building construction (zones)	2	Building construction zone: 1,600 Standalone houses: 200
Structural buildings	4	Impervious surfaces: 1,600 Parking aprons and runways: 5,000 Sand barriers: 5,000 Others: 4,00
Artificially excavated land	4	1,600
Desert land and exposed surfaces	4	Desert land: 10,000 Others: 1,600
Lakes, reservoirs, ponds	2	400

3 Factors affecting the splitting ability of proximal polygons

National geographical condition data integration is mainly performed on VSTs, and the fundamental principles of this process include simplifying geometric features, multi-level merging of semantic types, and conserving the percentages of each type of land, before and after integration. Ai (2002) proposes that the merging of VSTs be interpreted as a competition for survival during the competitive division of very small targeted proximal retained polygons. The splitting ability of a proximal polygon (i.e., its ability to absorb a very small target) will be proportional to its level of importance. The importance of a proximal polygon for a very small target is determined by its semantic similarity, the length of its shared side, the surface area of the proximal patch, and the regional importance of the type of land used in the proximal patch.

3.1 Surface area of proximal patches

The surface area of a proximal patch is the most intuitive factor for determining its splitting ability. In principle, the splitting ability of a patch is proportional to its surface area. The area of a proximal patch is usually calculated via coordinate analysis, and the corresponding mathematical model may be expressed as

$$S = \frac{1}{2} \sum_{i=1}^{n} x_i (y_{i+1} - y_{i-1}) \tag{1}$$

Here, i is the clockwise numbering of each node in a proximal polygon. When $i = 1$, then $i - 1 = n$; when $i = n$, then $i + 1 = 1$. x_i is the horizontal coordinate of each node in the proximal polygon, while y_{i+1} and y_{i-1} are the vertical coordinates of the proximal polygon.

3.2 Semantic proximity between a proximal patch and a very small patch

From the national geographical conditions survey, the geographic coverage of China is divided into 10 primary types, 58 secondary types, and 135 tertiary types. The total number of primary and secondary land types is often the focus of national geographical condition data, but patch classification and management is usually conducted via the more discriminating tertiary land types. In this work, we assume that semantic proximity is only possible between land types having the same primary type and class (e.g., primary/secondary/tertiary).

To calculate the semantic proximity between one land type and another land type in the same class, a semantically proximal set is first constructed and sorted (i.e., $Y = \{Y_1, Y_2, \ldots, Y_m\}$). The order of this set is based on the classification concepts used in the survey, where proximity relationships between land types having the same parent are prioritized over those with land types having different parents. The semantic distance between proximal elements is defined as one unit, and the semantic proximity between a patch at the i^th^ position, Y_i, and X is defined as

$$SemNei(X, Y_i) = 1 - \frac{i}{m} \tag{2}$$

When two land types are semantically unrelated, $SemNei\ (X,\ Y) = 0$. For primary land types, their proximity relationships are based on the method of classification shown in Figure 1. For example, the semantic proximity of other primary land types with respect to grasslands (also a primary land type) in decreasing order, is forested land, parkland, farmland, housing construction (zones), roads, structural buildings, artificially excavated land, desert and exposed land surfaces, and water bodies. Among the tertiary land types, high-coverage grasslands, medium-coverage grasslands, and low-coverage grasslands have a high degree of proximity to grasslands, as compared to pasture lands, lawns, sand-fixation shrubs, and slope stabilization shrubs.

3.3 Length of shared edges between a proximal patch and a very small patch

The length of the edges shared between two patches is an important indicator for judging their spatial proximity. In landscape ecology, matter transfer, energy transfer, and transitional capacity between two patches increase as the length of their shared edge increases. Therefore, proximal patches with longer shared edges have greater competitiveness, or splitting ability. The recognition of shared edges is performed by attaching semantic information to topological structures. If the nodes of some arc segment contain two different types of semantic data, it is then a shared edge. The distance between two nodes of a shared edge, d, between a proximal patch and a very small patch is usually calculated using the Euclidean method for calculating distances, whose mathematical model is as follows

$$d = \sqrt{(x_i - x_{i+1})^2 - (y_i - y_{i+1})^2}\ , \quad i = 1,2,\ldots,n \tag{3}$$

Here, i is the numbering of the nodes on the shared edge; x_i and x_{i+1} are the horizontal coordinates of each node on the shared edge; and y_i and y_{i+1} are the vertical coordinates of each node on the shared edge.

3.4 Importance of a proximal patch in the region

The importance of a proximal patch in a region is defined by its land-use type. In this work, the importance of each type of land use in a region is quantified via its total area as a proportion of the region's surface area. The mathematical model for calculating importance is

$$Imp(\ patch) = \frac{Area_{patch}}{Area_{total}} \times 100\%\ . \tag{4}$$

Here, $Imp\ (patch)$ signifies the importance of some type of land-use in the region, whereas $Area_{patch}$ is the total area of all patches within the region belonging to that category of land-use, and $Area_{total}$ is the total area of the region. All patches with the same type of land-use have the same degree of importance.

4 Extraction and refining of skeletons

The partitioning and splitting of VSTs is founded upon the use of Delaunay triangulation for the extraction of target patch skeletons. Whereas, competitiveness-based splitting and merging of proximal patches is crucial for the seamless splitting of VSTs.

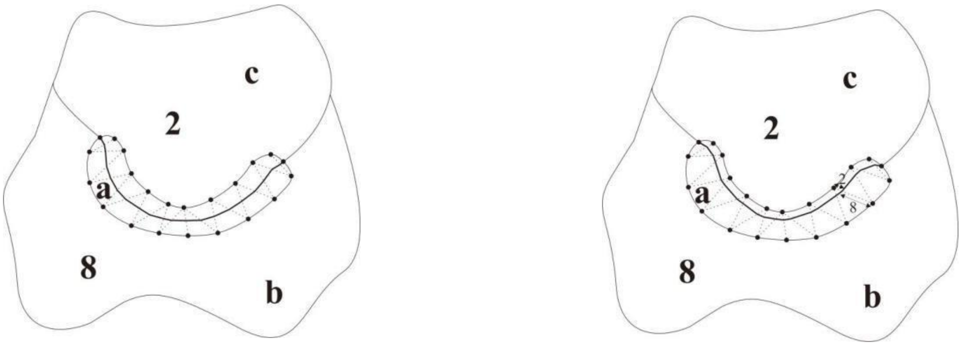
4.1 Skeleton refinement that accounts for spatial competitiveness

Because each proximal patch has different levels of competitiveness during the splitting of VSTs, the skeleton for these splits may be refined through the assignment of weights to the proximal patches. We define a splitting ability function (SAF) to calculate the spatial competitiveness of a proximal patch, **b**, on some very small target, **a**, expressed as follows

$$SAF(a,b) = w_1 S_{Area(b)} + w_2 S_{SemNei(a,b)} + w_3 S_{SharedEdge(a,b)} + w_4 S_{Imp(b)} \tag{5}$$

Here, $S_{Area(b)}$ is the area of the proximal patch, **b**; $S_{SemNei(a,b)}$ is the semantic proximity between the very small target, **a** and the proximal patch, **b**; $S_{SharedEdge(a,b)}$ is the length of the shared edge between the very small target, **a**, and the proximal patch, **b**; and $S_{Imp(b)}$ is the importance of the land-use type of proximal patch, **b**, in the region. The calculation of each parameter is described in Section 3. w_1, w_2, w_3, and w_4 are the weights calculated via the AHP for each parameter.

The calculation of *SAF* (*a*,*b*) values for all of the patches proximal to the very small target, **a**, therefore, gives the splitting ability of each proximal patch. At this point, the skeleton inside **a** may be refined per the splitting ability of each proximal patch. The refinement procedure is as follows. The two patches adjacent to **a** are determined via the edges of the class II triangles within the triangular mesh of **a**. The partitioning of the target via the skeleton (i.e., a bisection through the edges of the adjacent triangles) is then refined per the splitting ability of each proximal patch. As shown in Figure 1(a), **b** and **c** are the proximal patches of the very small target, **a**. Each have splitting abilities of 8 and 2, respectively. The refinement of the skeleton inside **a** results in the **b** and **c** partitions of **a** accounting for 80 % and 20 % of its area, respectively, as shown in Figure 1(b).



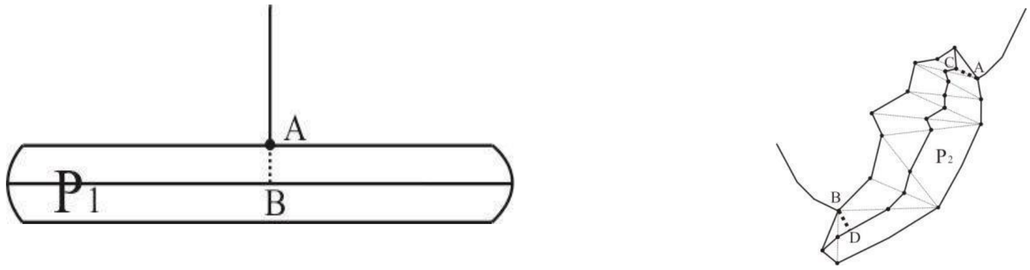
(a) Original skeleton (b) Refined skeleton

Figure 1: Skeleton refinement process accounting for spatial competitiveness.

4.2 Skeleton refinement that accounts for crossing point constraints

Ai (2002) proposed the classic method for mesh generation and a reference for boundary-constrained Delaunay triangulation. The triangles within a polygon were divided into three types. However, the direct extraction of skeletons from Delaunay triangular meshes within polygons often resulted in skeletons that do not extend to the contour of a patch, or discontinuities between outer and inner skeletons, as shown in Figures 2(a) and 2(b). To achieve the seamless partitioning of VSTs, the constraints imposed by the connection nodes of proximal polygon edges on the partitioning of patches must be accounted for. This includes

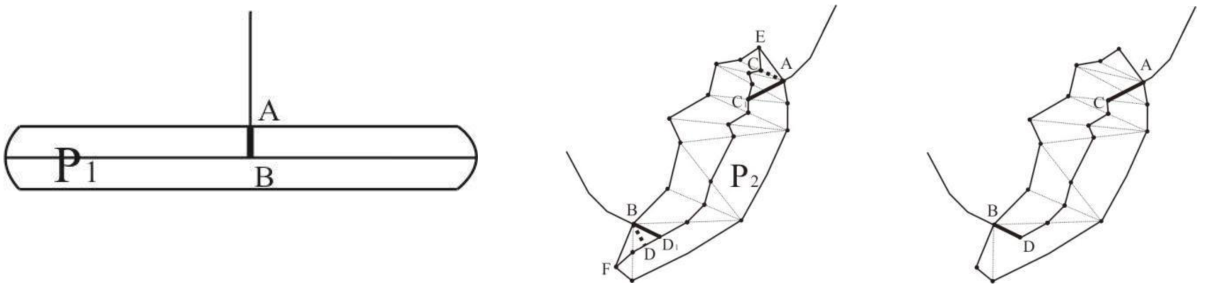
(1) The endpoints of each skeleton must be exactly located at the edge node of an adjacent polygon.
(2) The edge nodes of each adjacent polygon must exactly coincide with the endpoint of a skeleton.



(a) Partitioning lines that are perpendicular to an adjacent polygon's boundary (b) Partitioning lines that exhibit a certain angle with the boundary of an adjacent polygon

Figure 2: Flaws in the partition lines

Because Delaunay triangulation is based on boundary constraints, the boundary-crossing points between a proximal polygon and a candidate polygon for partitioning (i.e., point A in Figure 2(a) and points A and B in Figure 2(b)) must be the vertex of a Delaunay triangle. However, these crossing points are not necessarily connected to partitioning lines. Due to the constraints imposed by crossing points, the appendment of skeletons is therefore necessary for connecting crossing points. The appendment of partitioning lines is usually conducted via the nearest-neighbor method (Ai et al., 2010), but this method places stringent demands on the angular relationships between skeletons and boundary lines. If these lines are perpendicular to each other, the appendment of skeletons is relatively effective (i.e., the AB segment in Figure 2(a)). However, if the lines have any other angular relationship, the appended skeletons will exhibit hard corners around the crossing points (i.e., AC and BD in Figure 2(b)). Therefore, we have proposed the reversed extension method for augmenting the appendment of skeletons via the nearest-neighbor method. Via this method, the positional relationship between two proximal polygon nodes near the patch is used to determine the direction of extension of the polygon's boundary line. The boundary is then extended in the reverse direction so that it connects with the patch's skeleton. For example, extending the boundary of the polygon in Figure 3(a) from A to B yields the appended skeleton, AB, shown in Figure 2(a). In Figure 3(b), the boundary lines of the proximal polygons are extended from A and B to C1 and D1, respectively, giving us the appended skeletons, AC1 and BD1, shown in Figure 2(b). The hanging segments in skeleton topology are cyclically deleted (e.g., CE and DF in Figure 3(b)), except for appended skeletons, until none remain. This gives the refined skeleton shown in Figure 3(c).



(a) Nearest neighbor method (b) Reversed extension method (c)Final partition line

Figure 3: Partition line compensation for a single node

5 Experiments and analysis

The method for partitioning VSTs, which we developed, is imported into the WJ-III mapping workstation, developed by the Chinese Academy of Surveying and Mapping, to validate the rationality and efficacy of our method. This experiment was performed using the national geographical condition survey data of a city in the Jiangsu Province. The scale of the raw data is 1:10,000, and it contains 5,214 patches that mainly consist of natural features, such as farmland, forested land, and grassland. Artificial features, such as roads and houses are sporadically distributed throughout the region. The scale used for the simplification of targets is 1:100,000.

5.1 Procedure for the partitioning of VSTs based on the proposed method

The procedure for partitioning VSTs using the method proposed in this work is as follows.

(1) Construct a hierarchical structure of the splitting abilities of proximal patches

The splitting abilities of the patches proximal to each very small target play an important role in the refinement of skeletons inside the VSTs. Based on the principles of AHP, a two-level hierarchical structure is constructed for the splitting ability of proximal patches and its influencing factors, as shown in Figure 4.

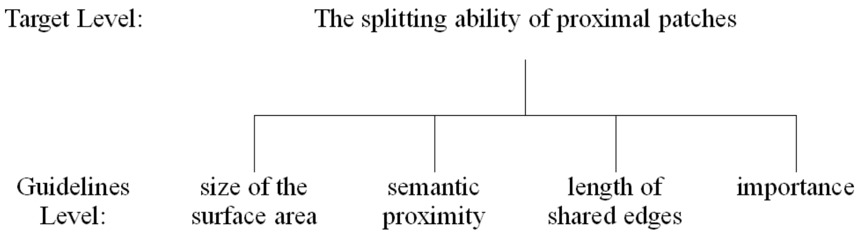


Figure 4: Hierarchical structure for the splitting ability of proximal patches

(2) Construct a judgment matrix for determining the importance of each factor

In the target layer, the splitting ability of a proximal patch is represented by A. In the criterion layer, factors such as the size of the surface area, semantic proximity, length of shared edges, and importance are represented by B_1, B_2, B_3, and B_4, respectively. The splitting ability of proximal patches is most strongly affected by the importance of the proximal patch within the region. Thus, land-use types that are relatively important in the region should have a higher level of splitting ability, whereas unimportant land-use types within the region (e.g., exposed land) will be gradually merged into other types of land. Additionally, to ensure that the overall percentages of each type of land-use in the region will not change significantly after integration, VSTs should be merged into patches that have more semantic proximity with the target. If there are several proximal patches with the same level of semantic proximity, the surface area and length of shared edges of each patch then should be taken into consideration. Surface area particularly has a greater impact on the polygonal partitioning of VSTs, geometrically. Therefore, the importance of each factor in decreasing order is as follows. The regional importance of the proximal patch's land type, its semantic proximity, surface area, and the length of shared edges. The relative importance of each factor is scored on a scale of 1 to 9, and the judgment matrix, **A**, obtained from this process is shown in Table 2.

Table 2: The judgment matrix of each factor

A	B_1	B_2	B_3	B_4
B_1	1	1/3	3	1/5
B_2	3	1	5	1/3
B_3	1/3	1/5	1	1/7
B_4	5	3	7	1

(3) Calculate the weight vectors to determine the weights of each factor

The eigenvector of judgment matrix, **A**, is calculated by summing the entries, giving the eigenvector W = (0.122, 0.263, 0.057, 0.558) and the maximum eigenvalue, $\lambda_{max} = 4.118$.

(4) Consistency test

Acodding to analytic Hierarchy Process, canculates the consistency index CI = 0.039. and the consistency ratio (CR): $CR = 0.044 < 0.1$. Hence, the judgment matrix satisfies the requirements of the consistency index.

(5) Perform boundary-constrained Delaunay triangulation

To ensure that the triangular mesh will not cross-over the patch boundaries of the very small target, the boundary-constrained Delaunay triangular mesh is constructed using the incremental insertion algorithm on the discrete points of the patch's internal border.

(6) Skeleton refinement that accounts for spatial competitiveness

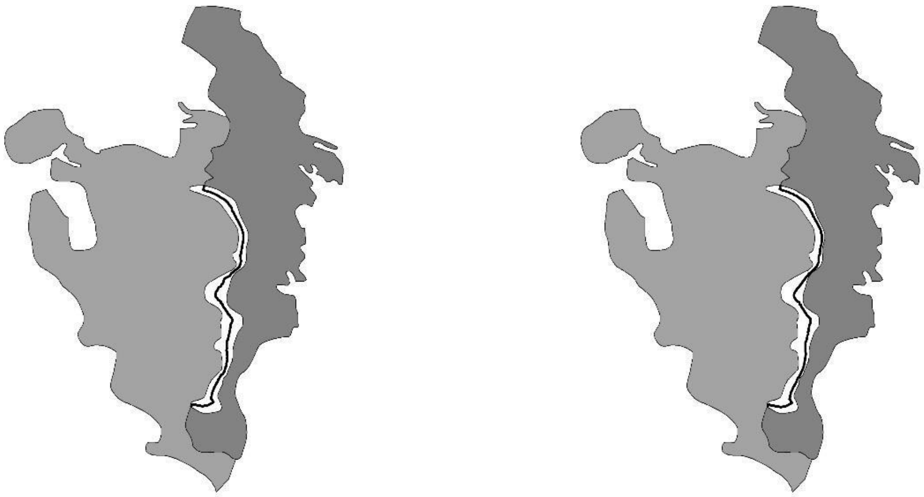
Equation (5) is used to calculate the spatial competitiveness (i.e., SAF) of the VSTs' proximal patches, where $w_1 = 0.122$, $w_2 = 0.263$, $w_3 = 0.057$, and $w_4 = 0.558$. The skeleton formed via bisection of the triangles connecting two proximal patches is then refined by altering the partition per the SAF of each proximal patch.

(7) Skeleton refinement with the consideration of crossing point constraints

The reversed extension method is finally used to connect the crossing points and skeletons of proximal patches and VSTs, thus realizing the seamless splitting of VSTs.

5.2 Experimental analysis

To validate the effectiveness of the method proposed in this work, we compared our method with commonly used centerline bisection-based methods in the partitioning of experimental data. The results of this comparison are shown in Fig 5.



(a) centerline bisection-based method (b) the method proposed in this work

Figure 5: The comparison of two methods.

Using the method proposed in this work, the partitioning line of the VSTs can better reflect the ability of subdivision of map spot, because the left side of figure spot area is larger, sharing is longer, it will get more subdivision of the area, division of results is more advantageous to maintain the relative percentage of different types of land use area before and after the comprehensive change obviously.

6 Concluding remarks

The partitioning of VSTs has a direct impact on the quality of results derived from national geographical condition data, and the seamless splitting of VSTs based on the competitiveness of proximal patches is a crucial (and problematic) aspect of this process. In this work, we proposed a partitioning method for VSTs that accounts for crossing point constraints, in which AHP was used to construct a hierarchical model for the factors influencing the competitiveness of proximal patches. The corresponding weights of each factor were then calculated to assess the spatial competitiveness of each patch in proximity with the VSTs. This is combined with crossing point constraints for skeleton refinement, thus providing an excellent solution for the rational partitioning of VSTs. In our experiments, it was shown that the method proposed in this work outperforms conventional partitioning methods in minimizing changes in the percentages of each type of land-use associated with the integration of VSTs, and satisfies the requirements for very small target integration.

Acknowledgements

This research was supported by the National Surveying and Mapping Projects (A1713).

References

AI Tinghua, LIU Yaolin (2002). Aggregation and Amalgamation in Land-use Data Generalization. Geomatics and Information Science of Wuhan University, 27(5): 486-492.

AI Tinghua, YANG Fan, LI Jingzhong. (2010) Land-use Data Generalization for the Database Construction of the Second Land Resource Survey . Geomatics and Information Science of Wuhan University, (8): 887-891.

AI Tinghua, YANG Fan, LI Jingzhong. (2010). Land-use Data Generalization for the Database Construction of the Second Land Resource Survey. Geomatics and Information Science of Wuhan University, (8): 887-891.

DENG Xue, LI Jiaming, ZENG Haojian, et al. (2012). Research on Computation Methods of AHP Weight and Its Applications.Mathematics in Practice and Theory, 42(7):93-100.

GDPJ 17-2016. Content and Index of National Census Geography. Leading Group Office of the First National Census Geography of the State Council, Beijing:2013.

GDPJ 17-2016. Technical Specification for Resulting Map of National Census Geography. Leading Group Office of the First National Census Geography of the State Council, Beijing:2016.

HU Huiming, QIAN Haizhong, HE Haiwei, et al. (2016). Auto-selection of Areal Habitation Based on Analytic Hierarchy Process . Acta Geodaetica et Cartographica Sinica, 45(6):740-746.

LIU Hailong, QIAN Haizhong, WANG Xiao, et al. (2015). Road Networks Global Matching Method Using Analytical Hierarchy Process. Geomatics and Information Science of Wuhan University, 40(5):644-651.

SAATY T L. (1980). The Analytic Hierarchy Process. New York:McGraw Hill.

YANG Jun, XI Jianchao, KONG Fanqiang, et al. (2013). Generalization of Land Use Patch Based on Semantic Priority —A Case of Beihai Sub-district of Lushun Port. Scientia Geographica Sinica, 33(8): 949-956.

ZHU Changqing, SHI Wenzhong (2006). Modeling and Principle of Spatial Analysis. Being: SciencePress.