Scientific
Research
Publishing

Rural Habitation Multistage Nature Boundary Extraction Based on Geographic Name Database

Binbin Hu[1], Hong Wang[1], Wei Zhang[2]

[1]Faculty of Resources and Environment, Hubei University, Wuhan, China
[2]School of Urban and Environmental Science, Central China Normal University, Wuhan, China
Email: 656819705@qq.com

Abstract

In order to extract the boundary of rural habitation, based on geographic name data and basic geographic information data, an extraction method that use polygon aggregation is raised, it can extract the boundary of three levels of rural habitation consists of town, administrative village and nature village. The method first extracts the boundary of nature village by aggregating the resident polygon, then extracts the boundary of administrative village by aggregating the boundary of nature village, and last extracts the boundary of town by aggregating the boundary of administrative village. The related methods of extracting the boundary of those three levels rural habitation has been given in detail during the experiment with basic geographic information data and geographic name data. Experimental results show the method can be a reference for boundary extraction of rural habitation.

1. Introduction

Rural habitation is an important part of habitation, compared with urban area, rural area has lower population density, and its administrative boundary is much larger than the boundary consist of its residential features. Town and administrative village have clear administrative boundary in basic geographic information database; the boundary of nature village is not included in basic geographic information database. The nature boundary of habitation has obvious orientation and social application value [1], as for this the paper uses geographic name data and basic geographic information data to extract the nature boundary consists of residential features belonging to the annotation of rural habitation. The result can make up for the lack of rural habitation's nature boundary, and be helpful to mastering the rural land use status quo.

There are some existent algorithm for residential polygons' generalization, such as the algorithm based on Delaunay triangulation structure and the model of Voronoi, the polygon generalization based on grid model, the polygon simplification based on mathematical morphology and syntactic pattern. The algorithm based on Delaunay triangulation structure and the model of Voronoi [2] can get a great generalizing result, but the program is too complex and the efficiency remains to be further improved. The polygon generalization based on grid model [3] has obvious effect just for the polygon having obvious rectangle feature under large scale. The polygon simplification based on mathematical morphology and syntactic pattern [4] has a complete theory, but the program is difficult to be realized as the establishment of neural network is too complicated and the complete grammar model set is hard to be exported. The paper uses geographic name data and basic geographic information data to obtain the residential features belonging to the annotation of rural habitation, and then generalizes the features by polygon aggregation. Polygon aggregation can merge the features in specified distance, all the habitation can get good merging results that the original habitation's spatial structure and spatial distribution have been retained by giving different aggregation distance. The paper, with the experiment with the 1:50 000 geographic name data and basic geographic information data as its experimental data, clearly describes the experimental result and the extraction process.

2. The Method of Nature Boundary Extraction

2.1. The Process of Nature Boundary Extraction

Nature village used to be one or more settlements. It is the place where villagers live for a long time [5]. Administrative village consists of nature villages. It has villagers committee and other organs of power, and is the most basic layer of China's administrative system [6]. Township consists of administrative villages. The boundary in this paper is the outline of residential features. In order to get the boundary the residential features belonging to the annotation should be extracted first, and then aggregated. The basic geographic information data contains three kinds of residential features, which are point, line and polygon. The nature boundary of nature village can be extracted by aggregating residential features of the three kinds with appropriate aggregation distance. Aggregating the nature boundary of nature village can get the nature boundary of administrative village. Aggregating the nature boundary of administrative village can get the nature boundary of township. The key point of this method is building the relation between nature village annotation and residential features, and then aggregating the related features. The process of boundary extraction is as follows:

1) Creating thiessen polygons with all the annotation of nature villages.

2) Selecting the thiessen polygon where the annotation locates in by the method of selecting by location.

3) Selecting the residential features in the thiessen polygon by the method of selecting by location, the relation between annotation and residential feature has been build.

4) Calculating aggregation distance of different annotations according to their related residential features, extracting nature boundary by aggregating the features.

5) Extracting the nature boundary of administrative village by aggregating the nature boundary of nature village belong to the administrative village in administration with the aggregation distance according to the distribution of nature villages' boundary.

6) Extracting the nature boundary of township by aggregating the nature boundary of administrative village belong to the township in administration with the aggregation distance according to the distribution of administration villages' boundary.

2.2. Aggregate Polygon

The nature boundary of residential annotation can be extracted by aggregating the residential features belonging to the annotation. **Figure 1** shows three different kinds of' methods of generalizing polygons, that are bounding rectangle, minimal bounding rectangle and convex hull [7]. The spatial distribution of buildings has been broken for those methods because the generalizing degree is too large. Polygon aggregation is a kind of cartographic generalization operation. The principle of polygon aggregation is converting vector polygon features to grid features, merging the polygons within the distance of specified distance by grid expansion algorithm, converting the result to vector data in the end. Polygon aggregation can get a new polygon feature by merging the polygons
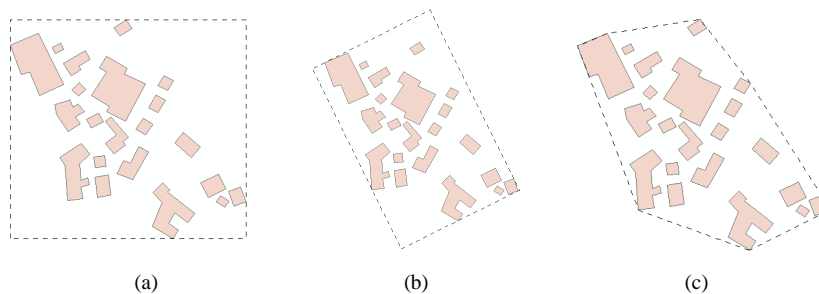
Figure 1. Generalization methods of polygons. (a) Bounding rectangle; (b) minimal bounding rectangle; (c) convex hull.

within specified distance. The result can keep the original spatial distribution of polygon features and filling blank area [8].

In comparison to the normal generalization methods of bounding rectangle, minimal bounding rectangle and convex hull, polygon aggregation has a universal applicability because it can set different aggregation distance to maintain the spatial structure and spatial distribution of the original residential features. Aggregation distance is the kernel parameter of polygon aggregation; different aggregation distance will directly affect the result. **Figure 2** shows the result of different aggregation distance. The aggregation distance in **Figure 2(a)** is 30 m, the result is multi-block boundaries, which means the aggregation distance is too small and the result is unreasonable. The aggregation distance in **Figure 2(b)** is 40 m, the result correctly reflects the outline of residential features, which means the aggregation distance is reasonable. The aggregation distance in **Figure 2(c)** is 80 m, the result leaves out part of the important turning point comparing with **Figure 2(b)**, which means the aggregation distance is a little too big. The aggregation distance in **Figure 2(d)** is 100 m; the result is the same with the result of convex hull, which means the aggregation distance is much too big. The committed step of polygon aggregation is the setting of aggregation distance. Only with appropriate aggregation distance, the result of polygon aggregation can correctly reflect the outline of residential features.

3. General Situation of the Experimental Area and Data

The experimental area is a county-level city consist of 6 towns and 3 townships, there are 62 administrative villages and 719 nature villages in total. This county-level city has an area of 1166 square kilometers and a population of 170 000, it is a rural area.

The experimental data consists of geographic name data and basic geographic information data, the mainly used geographic name data are the annotations of townships, administrative villages and nature villages, the mainly used basic geographic information data are the residential features of point, line and polygon. The attribute of the annotation contains the name, type and the administrative membership information; the attribute of the residential features do not include name or membership information.

4. The Extraction of Nature Villages' Boundary

The precondition of boundary extraction is extracting the residential features belonging to the target annotation. There are complete administrative membership information between the annotations of nature village, administrative village and township. However, the annotation of nature village does not have the administrative membership information with residential feature; residential feature in basic geographic information data does not have the information of name or administrative membership neither. The traditional method selecting by attribute cannot be used. According to the situation of data and the hash distribution of rural habitation, the paper extracts the nature boundary of nature village by aggregating the residential features of specified annotation extracted by the relation between annotation and residential features established by creating thiessen polygons from nature villages' annotation.

4.1. Establishing Relation between Nature Villages' Annotation and Residential Features

The method to establish relation between nature villages' annotation and residential features is as follows:
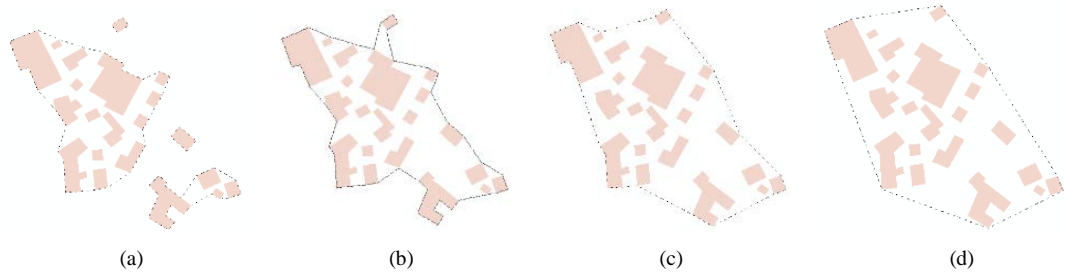
Figure 2. Result of different aggregation distance. (a) 30 m; (b) 40 m; (c) 80 m; (d) 100 m.

1) The annotation of habitation includes four levels of nature village, administrative village, township, county-city in geographic name data. Creating thiessen polygons with nature villages' annotation selected by the method of selecting by attribute. **Figure 3** is the thiessen polygon.

2) Extracting the thiessen polygon where the nature village's annotation is located by the method of selecting by location. Thiessen polygon will get the same name with annotation by giving nature village annotation's attribute of name to thiessen polygon. Just like **Figure 4**.

3) Extracting the residential features locates in thiessen polygon by the method of selecting by location, the extracted features belonging to corresponding the annotation of the thiessen polygon. According to the sparse rural houses, the method of selecting by location can extract the residential features belong to the annotation in most cases like **Figure 5**. However, in a few areas where buildings are close to each other, the residential features may cross the boundary of thiessen polygon, just like the feature of 1, 2 and 5 in **Figure 6**, the annotation these features belonging to can't be judged directly.

Aiming at the situation of **Figure 6**, the annotation of the features crossing the boundary can be judged according to the annotation of the closest feature not crossing the boundary.

4.2. The Decision Method of Nature Villages' Aggregation Distance

Hash style and assemble style are two main styles of rural habitation [9]. **Figure 7** shows different habitations have different style of distribution. In order to keep the original style of habitation and make the extracted boundary an abstractly generalization of the habitation, aggregation distance should be different according to the distribution of residential features of different annotations.

The aggregation distance is decided by the following method:

a) Getting the nearest distance between annotations , a list of distance for the annotation and the nearest annotation can be obtained, $(L_1, L_2, L_3,..., L_n)$.

b) Taking the center of residential features as the center of circle, taking L_n as radius of the circle. Counting the number of the intersected circle, a list of the integer order sequence can be obtained, $(N_1, N_2, N_3, ..., N_{n-1}, N_n)$. $N_n = 1$ in the list. If only $N_n = 1$, it means the residential feature with L_n is furthest from other residential features. L_{n-1} should be the aggregation distance to separate the relatively independent feature.

c) If not only $N_n = 1$, taking L_{n-1} as the radius of circle to create circle. A list of integer order sequence can be obtained, $(N_1, N_2, N_3, ..., N_{n-1}, N_n)$. $N_n = 0$, $N_{n-1} = 1$ in the list. If only $N_{n-1} = 1$, taking L_{n-2} as the aggregation distance. If not only $N_{n-1} = 1$, keep trying the process until only $N_{n-a} = 1$ $(0 < a < n)$, taking L_{n-a-1} as aggregation distance.

Deciding aggregation distance with this method can keep the extracted boundary correctly reflects the outline and density of residential features by separating the independent features. **Figure 8** shows the result.

5. The Extraction of Administrative Village and Townships' Boundary

There are clear administrative membership information between township, administrative village and nature village. The boundary of administrative village can be extracted by aggregating the boundary of nature village. The boundary of township can be extracted by aggregating the boundary of administrative village.

In order to ensure that all the boundaries are taken into calculating the aggregation distance should be big enough when aggregating the boundaries. At the same time, the aggregation distance cannot be too big to make
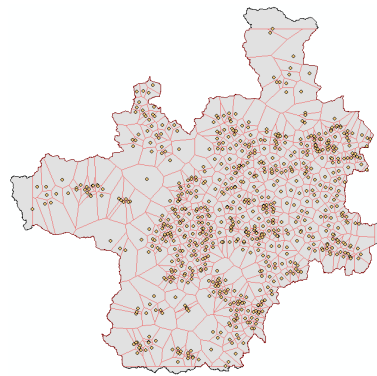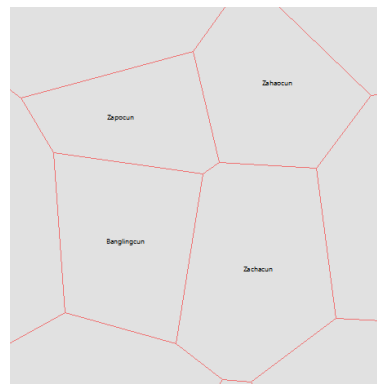
Figure 3. Result of creating Thiessen polygon.

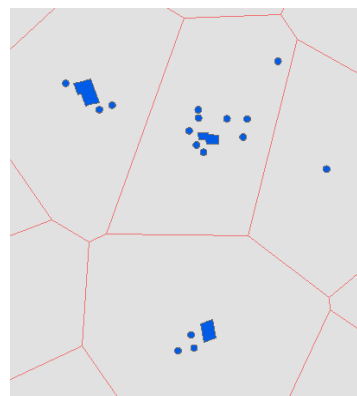
Figure 4. Attribute of Thiessen polygon.
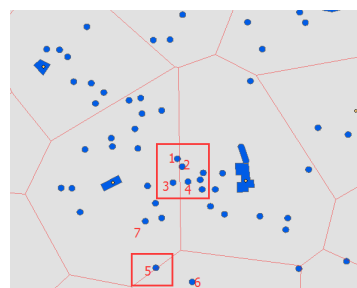

Figure 5. Normal situation.

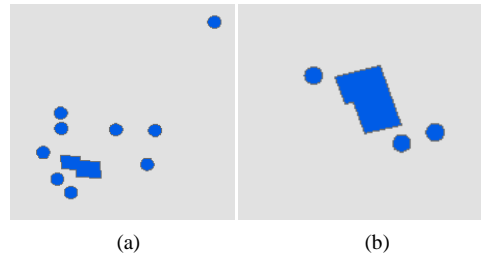**Figure 6.** Resident features cross boundary.

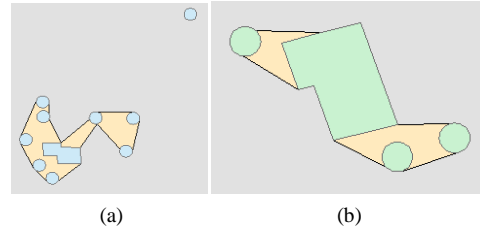**Figure 7.** Rural habitation. (a) Hash style; (b) assemble Style.

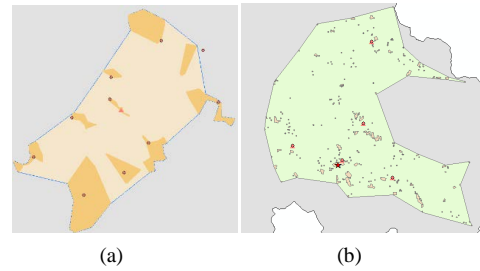**Figure 8.** Result of nature village's extraction. (a) Hash style; (b) assemble style.

Figure 9. Result of extraction of administrative village and township. (a) Administrative village; (b) township.

the generalization degree too large. The aggregation distance is decided by the following method:

Taking 1000 m as the original aggregation distance and 100m as the interval change according to the general situation of the experimental data to aggregate the residential features. If the result of aggregation contains many blocks, the aggregation distance should be bigger. Increasing the aggregation distance n times with 100 m every time until the result does not divided into many pieces. Taking 1000 + 100n as aggregation distance. If the result contains only one block with the original distance, the aggregation distance should be smaller. Decreasing the aggregation distance n times with 100m every time until the result divided into pieces. Taking 1000 − 100 (n − 1) as aggregation distance. Deciding aggregation distance by the method taking critical value of result divided as aggregation distance can ensure the aggregation distance is big enough that all the boundaries belonging to annotation may be taken into calculating and the aggregation distance not too big to ignore the details. The result of extraction is as following **Figure 9**.

6. Conclusion

Rural habitation is an important part of habitation, but the study of it is relatively rare currently. The paper uses basic geographic information data and geographic name data to extract the boundary of rural habitation by the method of aggregation polygon. The result of the experimental area is good, and the distribution of original residential features has been well kept. It has certain reference value for the same kind problems.

Boundary extraction is a complex question, the method in this paper has the following problems need to be studied further: the relation between annotation and features would be more reasonable if taking terrain, drai-

nage and road as constraint condition to establish thiessen polygon. The kinds of habitation is of great diversity, but the test samples are limited. The result will be more reasonable if quality control method is added after the extraction. The efficiency need to be improved as the polygon aggregation involves grid vector transformation.

Acknowledgment

References

[1] Zhang, X. (2014) Modeling Method of Vague Place Names Based on Rough Sets. Wuhan University.

[2] Ai, T.H. and Guo, R.Z. (2001) Simplification and Aggregation of Polygon Object Supported by Delaunay Triangulation Structure. *Journal of Image and Graphics*, **6(A) 7**, 709-709.

[3] Ai, T.H. and Guo, R.Z. (2000) Simplification and Aggregation of Building Polygon in Automatic Map Generalization. *Journal of Wuhan Technical University of Surveying and Mapping*, **25**, 25-30.

[4] Wang, H.L., Wu, F., Zhang, L.L., *et al.* (2005) The Application of Mathematical Morphology and Pattern Recognition to Building Polygon Simplification. *Acta Geodaetica et Cartographica Sinica*, **34**, 269-276.

[5] http://baike.so.com/doc/5683577-5896256.html

[6] http://baike.so.com/doc/350003-370758.html

[7] Papadias, D. and Theodridis, Y. (1997) Spatial Relations, Minimum Bounding Rectangles, and Spatial Data Structure. *International Journal of Geographical Information Systems*, **11**, 111-138. http://dx.doi.org/10.1080/136588197242428

[8] Yan, H.W., Chu, Y.D., Yang, S.W., *et al.* (2007) The Principle and Algorithm of Computer Mapping. Science Press, Beijing.

[9] Lu, X.Q. (2016) Dispersed and Clustered: Rural Settlement Patterns and Its Evolution in Traditional China. *Journal of Huazhong Normal University* (*Humanities and Social Sciences*), **52**, 113-130.