AGGREGATION IN LAND-COVER DATA GENERALIZATION CONSIDERING SPATIAL STRUCTURE CHARACTERISTICS

P.D. Wu 1, Y. Yin 1, C.M. Li 1, X.L. Liu 1

1 Chinese Academy of Surveying and Mapping, 100830 Beijing, China - (wupd, yinyong, cmli, liuxl)@casm.ac.cn

KEY WORDS: Map generalization, Aggregation operation, Structured geographic object, Linear pattern, Grid pattern

ABSTRACT:

Aggregation is an important operation for the generalization of land-cover data. However, current research often entails aggregation on a global perspective, which is not conducive to capturing the spatial characteristics of geographic objects with significant spatial structures, i.e., structured geographic objects. Hence this paper proposes an area aggregation method that can maintain the boundary characteristics of the structured geographic objects. First, we identify the structured geographic objects based on the description parameters of the spatial structure. Second, a Miter-type buffer transformation is introduced to extract the boundary of each structured geographic object, and area elements inside the boundary are processed with corresponding aggregation operations. Finally, the boundary of the structured geographic objects and the aggregation result of the area elements are inserted back into the aggregated result of the original land-cover data using the NOT operation. The proposed approach is experimentally validated using geographical condition census data for a city in southern China. The experimental result indicates that the proposed approach not only reasonably identify the typical characteristics of structured geographic objects but also effectively maintains the boundary characteristics of these objects.

1. INTRODUCTION

Land-cover thematic data are a specific spatial tessellation with no gap or overlap. As a map is transformed from a larger scale into a smaller one, long and narrow areas in the larger-scale map are difficult to present in the smaller-scale map. Under such circumstances, an area aggregation operation is required. This aggregation can be divided into two types, namely, amalgamation and aggregation. Amalgamation refers to aggregation small areas into topologically adjacent areas with different semantic information, whereas aggregation refers to aggregation areas with homogenous semantic information that are topologically separated by a narrow area (Ai & Liu 2002). Previous studies note that such aggregation operations must not only consider the natural boundary shape of each area but also account for the structural pattern of various land-cover classes or the spatial distribution pattern (Brassel & Weibel 1988, Steiniger& Weibel 2007). They must also ensure that the relative percentage of each land-cover class remains the same on a general statistical basis (Sester 2005, Haunert & Wolff 2010).

Van Oosterom (1995) proposed a classic iterative algorithm in which the least-important small-area from the data set of the land-cover map is iteratively merged into its adjacent area. However, this method, which is essentially an amalgamation operation, can deal only with adjacent areas, and the land-cover class patterns of the map change significantly after the generalization. Ai et al. (2002) introduced boundary-constrained Delaunay triangulation into the aggregation operation, in which the skeleton edge in a small area is picked up through triangulation, and the area is then divided and assigned to multiple topologically adjacent areas for further amalgamation. Moreover, the “bridging” area between the topologically separated areas is extracted via triangulation, and area aggregation is enabled using the OR operation. Weng et al. (2012) later described a modified area aggregation algorithm that introduces the concept of a buffer to effectively accomplish the aggregation of the bridging area. The modified method better maintains the natural curved shape of the area boundary. Nevertheless, all of the above methods fail to address the problem of considerable area variation of land-cover classes after aggregation. To this end, some scholars have proposed an area aggregation approach based on the global optimum. In addition, Haunert & Wolff (2006) attempted to achieve optimization in land-cover data aggregation via mixed integer programming by considering both the minimum variation in area of each land-cover class and the compactness of the aggregation results.

In summary, the present aggregation methods adequately maintain the natural shapes of areas and effectively keep the sum of area variation of each land-cover class as low as possible. However, these methods often proceed on a global basis and consider each land-cover class as a unit to account for the characteristics of “no gap” and “no overlap” in such land-cover data. They seldom consider the characteristics of area features that have their own inherent patterns in terms of the spatial distribution, such as buildings and pit ponds. Therefore, during aggregation, the boundaries of the area features with unique spatial structures are altered, and the spatial structure characteristic is partially or fully lost. On the basis of the previous studies, this paper focuses on following the regularity of the original spatial distribution and proposes an area aggregation method that can maintain the boundary characteristics of structured geographic objects.

The remainder of the paper is organized as follows: Section 2 presents two main area aggregation methods and discusses their pros and cons. Section 3 introduces the proposed approach, including characteristic identification of the typical pattern, boundary extraction of typical patterns, different aggregation operations for area objects arranged in typical patterns and conflict adjustment during the restoration process. Section 4
presents a series of experiments that were conducted to validate the effectiveness and reliability of the proposed method. Section 5 summarizes the findings and discusses future research avenues.

2. RELATED WORK

2.1 Two main area aggregation methods

2.1.1 The classic aggregation method: Van Oosterom (1995) proposed the classic iterative aggregation algorithm, which can also be interpreted as an area "growth" algorithm. The key procedure of the algorithm is as follows:

\[
\text{S} \leftarrow \text{set of areas below a threshold for the target scale}
\]

\[
\text{while } \text{S} \neq \emptyset \text{ do}
\]

\[
\text{a} \leftarrow \text{smallest area in S}
\]

\[
\text{merge a to the most compatible neighbor}
\]

\[
\text{update S}
\]

\[
\text{end while}
\]

The neighboring area is chosen by calculating the local optimum, for example, the degree of semantic similarity in the land-cover class between a small area and its neighboring area (semantic distance) and the area of the neighboring area or the length of the shared boundary (geometric distance) (Podrenek et al., 2002; Van Smaalen et al., 2003). The classic iterative aggregation algorithm is a greedy algorithm and serves to simply fulfill the aggregation of neighboring areas. Although it has a high implementation efficiency, significant area variation is observed in land-cover classes after generalization, and the resultant aggregation quality is low (Haunert & Wolff 2006).

2.1.2 Area aggregation method based on the global optimum: Given the disadvantages found of the classic iterative aggregation algorithm, Haunert & Wolff (2006, 2010) comprehensively accounted for the area variation of each land-cover class and the compactness of the geometry and tried to address the optimization of the area aggregation operation using mixed integer programming. In their work, the area is described on the basis of graph theory. That is, an area is represented by nodes, the size of the area is measured by the node weight, land-cover classes are represented by colors, and adjacency relationships are represented by edges. As such, the area aggregation problem is converted into a node grouping problem.

Haunert & Wolff (2010) introduced mixed integer programming into processing for regions with fewer data (a small number of nodes), whereas they adopted a center-distance heuristic approach for regions with more data (many nodes). The area data aggregation approach based on the global optimum considers the geometric distance and semantic and topological relationships between areas, which obtained better compactness of the merged area and minimum area variation for each land-cover class after generalization. As is presented in Fig. 1(a), multiple buildings in an aggregated distribution are merged into one.

2.2 Shortcomings of current area aggregation methods

From the above analysis, it is clear that the present approaches are adequate for area features that are in an aggregated distribution and randomly arranged, such as natural geographic objects, e.g. farmlands and woodlands. However, in regard to artificial area objects that are characterized by inherent regularity and unique complexity in their spatial distribution, such as buildings that are arranged in a grid or linear pattern and pit ponds, which are often adjacent to each other, after aggregation via current approaches, the spatial distribution pattern is compromised (Figs. 2).

Buildings that are arranged in a linear pattern are illustrated in Fig. 2(a). As the scale is changed from larger to smaller, the typical structural characteristics of the buildings and pit ponds are expected to be maintained, and small areas around them should be merged into other land-cover classes. Unfortunately, because current approaches mostly address the problem on a global basis and ignore the spatial distribution patterns of the area features, the small areas are merged into adjacent areas of the same classes, and the original spatial structure of buildings and pit ponds are broken, as depicted in Figs. 2(b).

3. METHODOLOGY

This paper proposes an area aggregation method that can maintain the boundary characteristic of structured geographic objects, the basic principles of which are as follows. First, the typical spatial distribution pattern for the structured geographic objects are identified on the basis of descriptive parameters of the spatial structure, and the objects are extracted from original land-cover data. Second, the boundary of each typical pattern is
extracted, and the structured geographic objects aggregated within the boundary are merged using a corresponding aggregation operation. Third, aggregation of the original land-cover area data is performed through the aggregation approach proposed by Haunert & Wolff (2010). Finally, the complementary set of aggregation results of the typical distribution patterns were obtained using the NOT operation (Margalit & Knott 1989) for the extracted boundary and the aggregation results of the original land-cover area data. Then, the aggregation results with typical spatial distribution patterns are embedded into the complementary set using the spatial Insert operation (Song et al. 2015) to reform it into complete area data with no gaps or overlap. The detailed workflow of the proposed method is shown in Fig. 3.

### 3.1 Identification of structural characteristics

The structured geographic objects present a macroscopic group distribution, and within the group, the area element has a similar size and shape, as well as a regular distance and directional and topological relationships.

#### 3.1.1 Parameters of the spatial structure and their calculation

Based on previous studies (Yan et al. 2008, Liu 2013, Yan et al. 2017, Li et al. 2018), this paper adopts the following parameters as the description factors for the spatial structure of the aggregated area objects, including the principal direction difference \( O_{\text{diff}} \), the path direction difference \( \text{PO}_{\text{diff}} \), the size similarity \( S_{\text{size}} \), the shape similarity \( S_{\text{shape}} \), the width of the bridging area \( B_{\text{distance}} \), the effective connection index \( ECI \), and the distribution pattern index \( DPI \).

Detailed definitions of the parameters are shown in Fig. 4, the parameter of \( ECI \) which are not defined in the figure are calculated as follows:
In terms of area elements approaching each other, the two elements can be recognized as a unit in the case that some part of one area element visually connects with the other, according to the Gestalt principle of continuation. The ratio of the connective region between the two neighboring elements to the unit is a key indicator for identifying an agglomerate region. This paper uses ECI to capture this indicator, which can be calculated as detailed below:

Step 1: Identify the Type-II triangles, which have two neighboring triangles in the boundary-constrained Delaunay triangulation (Ai & Liu. 2002). Then, define the region covered by the Type-II triangles as the connection region between the two elements, as illustrated in yellow and green in Fig. 5.

Step 2: Calculate the inner angles of each triangle. Triangles with no obtuse inner angles and whose neighboring triangles also contain no obtuse angles are recognized as effective triangles. The region covered by these effective triangles is referred to as the effective connectable area (ECA), as illustrated in green in Fig. 6. Other regions are defined as the invalid connectable area, as shown in yellow in Fig. 5.

Step 3: Calculate ECI based on the ratio of the area of the connectable area to that of the total area, the expression of which is shown below as follows:

$$ECI_{X_i,X_j} = \frac{CA_{X_i,X_j}}{TA_{X_i,X_j}}$$

where $ECI_{X_i,X_j}$ is the ECI of two neighboring area elements $X_i$ and $X_j$, $CA_{X_i,X_j}$ refers to the effective connectable area and $TA_{X_i,X_j}$ refers to the total connectable area, including the effective connectable area and invalid connectable areas.
3.1.2 Typical pattern identification: Linear and grid patterns are two typical structures for aggregated area objects and are identified according to the abovementioned spatial structure description factors combined with the Gestalt principles of closure, extensibility and element connectedness (Zhang et al. 2013a, Zhang et al. 2018). The controlling parameters for the pattern identification are listed in Table 1.

Table 1. Controlling parameters for the identification of structural characteristics of structured geographic objects.

<table>
<thead>
<tr>
<th>Typical Pattern</th>
<th>Pattern Description</th>
<th>Identification Parameter &amp; Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear pattern</td>
<td>(1) Each object in the pattern group has similar shape and size; (2) The principal directions of all objects in the pattern group are almost identical, and the global direction of the pattern is basically identical or perpendicular to the principal direction of each object. (Yan et al. 2008)</td>
<td>$O_{\text{diff}} &lt; \text{Maximum principal direction difference } \delta_{O_{\text{diff}}}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$B_{\text{distance}} &lt; \text{Maximum width of the bridging area } \delta_{B_{\text{distance}}}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$S_{\text{size}} &gt; \text{Minimum size similarity } \delta_{S_{\text{size}}}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$S_{\text{shape}} &gt; \text{Minimum shape similarity } \delta_{S_{\text{shape}}}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$ECI &gt; \text{Minimum ECI } \delta_{ECI}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$PO_{\text{diff}} &lt; \text{Maximum path direction difference } \delta_{PO_{\text{diff}}}$</td>
</tr>
</tbody>
</table>

Grid pattern

| (1) Two sets of linear patterns coexist; (2) Linear patterns are nearly parallel to each other in each set; (3) Two sets of linear patterns are nearly orthogonal to each other; (4) Two sets of linear patterns are connected to each other. (Zhang et al. 2013b) |

The boundary-constrained Delaunay triangulation is first constructed across the region where the structured geographic objects are located, and an initial area binary group is built on the basis of the spatial adjacency of each area object captured with respect to areas connected by triangle edges. Then, each group structural description parameter between elements in the binary group is calculated. Furthermore, a linear pattern is identified, and the grid pattern is characterized through the identified linear pattern.
3.2 Boundary extraction of structured geographic objects

Maintaining the pattern of the structured geographic objects relies on the boundary of their external boundary. Given this and based on the concept of morphological transformation, this paper introduces a Miter-type buffer treatment (Park et al. 2003, Yi et al. 2008) to extract the boundary of structured geographic objects (Li et al. 2018). Fig. 6(a) presents an example of how this operation proceeds.

Step 1 is implementing a dilation-erosion transformation of the structured geographic objects. First, dilation transformation of the original aggregated area group is performed with a buffer distance of \( L \), and the overlapped regions among dilated areas are amalgamated, resulting in Polygon \( P_1 \), as depicted in Fig. 6(b). Then, erosion transformation of Polygon \( P_1 \) is conducted with a buffer distance of \( L \), which leads to Polygon \( P_2 \), as depicted in Fig. 6(c).

![Figure 6. Dilation and erosion transformations. (a) Original pattern; (b) dilation transformation; (c) erosion transformation.](image)

As shown in Fig. 6, the dilation-erosion transformation is characterized by remaining “convex” and “flat” while eliminating “concave” parts. Comparison of the graph before and after the transformation indicates that the general morphology of the graph remains the same, and the convex and straight parts remain the same, namely, “convex and flat remaining”. However, the concave part of the graph is amalgamated during the transformation, and the overall graph morphology tends to be smoother, which is called “concave eliminating”. Certainly, the degree of “concave eliminating” depends on the buffer distance \( L \).

Step 2 is restoring the concave structure. First, the concave part of the area must be identified, and the principle of this operation is as follows: a unified topology of Polygon \( P_2 \) and the original area element group is constructed, and the semantic information of the area element is assigned to a corresponding arc section; if the arc section of an area is composed of a certain arc section with semantic information and another with no semantic information, then the area is identified as the concave part eliminated by the erosion transformation. The arc section with no semantic information is replaced by that with semantic information, and thus, the concave part is restored. The resultant boundary Polygon \( P \) of the area element group after this operation is the smallest envelope polygon of the aggregated area element group, and its boundary is thus the boundary of the aggregated area element group. For example, in Fig. 7a, the topological Polygon \( O \) consists of Arc Sections \( L_1 \) and \( L_2 \). \( L_2 \) contains the semantic information of Polygon \( A \), whereas \( L_1 \) of Polygon \( P_2 \) contains no semantic information. Therefore, Polygon \( O \) is a concave part, and the boundary \( P \) is obtained by replacing \( L_1 \) with \( L_2 \) as the arc section of the boundary Polygon \( P \), as depicted in Fig. 7(b).

![Figure 7. Restoration of the concave structure. (a) Concave identification; (b) boundary of the structured geographic objects.](image)

For aggregated area objects within the external boundary, with respect to their structural characteristics, area objects in the linear pattern are often aggregated and amalgamated, but the aggregation operation for area objects in the grid pattern depends on the object density.

Supposing the grid pattern of the aggregated area objects is defined as \( G = \{O_1, O_2, ..., O_n\} \), its external boundary is defined as \( BC_G \); the area of the polygon formed by the external boundary is defined as \( S_G \), and the area of area objects is defined as \( \{S_{O_1}, S_{O_2}, ..., S_{O_n}\} \). Accordingly, the area object density \( D_G \) within the grid pattern can be calculated using Eq. (3):

\[
D_G = \frac{S_{O_1} + S_{O_2} + ... + S_{O_n}}{S_G}
\]
If $D_{ij} < 0.7$, it is thought that the objects within the grid pattern are sparsely arranged with large gaps. The polygon defined by the external boundary is not sufficiently filled, and under such circumstances, typification is preferred. If $D_{ij} \geq 0.7$, it is believed that the gap between the area objects within the grid pattern is small, that the objects are closely arranged and that they sufficiently fill in the polygon that is defined by the external boundary. In such cases, agglomeration is preferred for waters and aggregation is preferred for buildings.

4. EXPERIMENTS AND ANALYSIS

4.1 Experimental data and experimental environment

To validate the effectiveness and reliability of the proposed method, this paper tests our method on a dataset derived from the 1:10,000 geographical census data of a city of southern China. Relying on the WJ-III map workstation developed by the Chinese Academy of Surveying and Mapping, the method proposed in this paper is embedded. The experimental data cover an area of 699.38 km², containing 124237 patches and including Level-1 land-cover classes such as farmland, woodlands, buildings and water. The buildings in the test area are dense, and the adjacent buildings of different directions and sizes are staggered, showing obvious linear and grid patterns. The system operation environment of the software is a 64-bit Windows 7 system, with a CPU with 8 cores of Intel Core i7-3770 3.2 GHz, a memory of 16 GB and a solid-state disk of 1024 GB.

4.2 Effective Validation of maintaining the boundary characteristics

Zoomed-in figures of typical places of the two patterns found in the experiment region are shown in Figs. 8-11. Fig. 8 presents the aggregation results of the simple linear pattern of regularly arranged structured geographic objects using the mixed integer programming method and the method proposed in this paper. The aggregation results of the complex linear pattern with disturbance from adjacent area objects based on the two abovementioned approaches are shown in Fig. 9. The aggregation results of the simple grid pattern of regularly arranged area objects based on the two abovementioned approaches are shown in Fig. 10, whereas those of the complex grid pattern with branches based on the two abovementioned approaches are included in Fig. 11.

Figure 8. Comparison of aggregation results for the simple linear pattern. (a) Original data; (b) result of mixed integer programming; (c) result of the method proposed in this paper.

Figure 9. Comparison of aggregation results for the complex linear pattern. (a) Original data; (b) result of mixed integer programming; (c) result of the method proposed in this paper.

As shown in Fig. 8, the aggregation results of the simple linear pattern produced by mixed integer programming method and the approach presented in this paper are basically consistent. The two approaches both sufficiently merge the structured geographic objects, and the boundary accurately reflects the characteristic of the structured geographic objects. Moreover, Fig. 9 demonstrates that the mixed integer programming method is disturbed by the neighboring area object of the same land-cover class, preserving only the aggregation characteristic and losing the linear characteristic of the aggregated area object. In contrast, the approach proposed by this paper maintains the linear characteristic of the aggregated area object well, and all disturbing adjacent area objects are merged into the other area object that is defined by a different land-cover class.

Figure 10. Comparison of aggregation results for the simple grid pattern. (a) Original data; (b) result of mixed integer programming; and (c) result of the method proposed in this paper.

Figure 11. Comparison of aggregation results for the complex grid pattern. (a) Original data; (b) result of mixed integer programming; (c) result of the method proposed in this paper.

Given that the traditional area aggregation approach for land-cover data fails to effectively maintain the spatial characteristics of structured geographic objects, this paper proposes an area aggregation approach that can identify, merge and maintain two typical aggregated distribution patterns of structured geographic objects, namely, linear and grid patterns. The proposed approach is experimentally validated based on geographical census data for a city in southern China. The main conclusions are as follows:

5. CONCLUSIONS
(1) The proposed approach not only reasonably identify the typical characteristics of structured geographic objects but also effectively maintains the boundary characteristics of these objects.

(2) The proposed approach is superior to the traditional approach in terms of addressing complex spatial structure. The traditional mixed integer programming method can effectively handle simple linear and grid patterns of regularly arranged area objects, whereas the proposed approach can sufficiently work in the case of a complex linear pattern with disturbances from adjacent same-class area objects or a complex grid pattern with branch structures.

The focus of our future research is to elaborate typical aggregated distribution patterns of structured geographic objects and thus refine the aggregation results.

REFERENCES


Li, Z., 2006. Algorithmic Foundation of Multi-Scale Spatial Representation; CRC Press: Bacon Raton, FL, USA.


