A Methodology for the Assessment of Generalization Quality

Andelina Skopeliti
askop@central.ntua.gr

Lysandros Tsoulos
lysandro@central.ntua.gr

Cartography Laboratory
Faculty of Rural and Surveying Engineering
National Technical University of Athens, Greece
Main Idea

Spatial Data Quality

Linear entities positional accuracy

Shape Fidelity

Assessment of shape change

Measures for shape change

Assessment of horizontal position change

Measures for horizontal position error

Horizontal position accuracy

Horizontal Position Description
Pairs of coordinates

Structure Recognition and Shape Description

Linear Features Partitioning into homogeneous segments

Parametric description of linear segments shape

Cartographic Generalization
Parametric Description of Line Shape

Description of cartographic line shape based on a group of parameters:

- Bibliographic research
- Data set creation
- Experimental analysis
- Factor analysis
- Final set of parameters:
  - The average magnitude angularity at different vertex ranges
  - The error variance
  - The ratio of line length to anchor line length

Classification of line segments into similar shape groups using cluster analysis
Structure Recognition
Partitioning into homogeneous segments

• Methodology applied to natural features based on the variation of the fractal dimension

• Implementation Steps:
  a) Self - similar segments identification along the line to be segmented
  b) Self - similar segments clustering
  c) Selection of a representative segment for each group of segments
  d) Preliminary segmentation
  e) Segmentation refinement
  f) Final Segmentation
Measures for Positional Accuracy [I]
Line shape change

**Quantitative assessment – shape change measure**

- The distance between the original and the generalized line in the parameters’ space implies shape modification.
- The average value of the shape change, for the lines which make up a group, represents the average line shape change for this group.

**Qualitative assessment – classification results**

- When non-hierarchical cluster analysis is applied, using the centers of the original lines groups, the results describe the generalized lines similarity to the original lines.
- When hierarchical cluster analysis is applied, the results show the similarity between the generalized lines.
Measures for Positional Accuracy [II]

Horizontal position

**Distance measures:** Average Euclidean distance, Hausdorff distance

**Areal displacement measures:** the ratio of the area between the original and the generalized line to the length of the original line

\[
\delta_i = \frac{E_i}{\sum K}
\]

\[
E_{\delta_\lambda} = \sum E_i / L
\]

Hausdorff distance

\[
d_H (A, B) = \max(d_{AB}, d_{BA})
\]
Positional Accuracy Study

Stages:

• Linear features partitioning in homogeneous segments
• Parametric description of linear segments shape and clustering in groups with similar shape
• Implementation of several generalization solutions (operators, algorithms and tolerance values)
• Assessment of the positional accuracy
Comparison of Shape Change by Algorithm and Tolerance Value

Simplification

- Euclidean distance
- Lang
- Douglas - Peucker
- Reumann - Witkam

Average Shape Change vs. Tolerance Values (m)
Algorithms Influence on Line Groups of Different Shape

Douglas - Peucker Simplification Algorithm

Average Shape Change

Tolerance Values (m)

VSM
SM
SIN
VSIN
Horizontal Position Error by Algorithm and Line Category

Graph showing the relationship between Tolerance Values (m) and Hausdorff Distance (m) and Average Euclidean Distance (m) for different algorithms and line categories.

- dp Hausdorff
- rw Hausdorff
- ed Hausdorff
- lg Hausdorff
- dp Euclidean
- rw Euclidean
- ed Euclidean
- lg Euclidean
Overall Positional Accuracy

![Graph showing overall positional accuracy with different methods: Euclidean Distance, Lang (5 points), Douglas - Peucker, and Reumann - Witkam. The graph plots average shape change against average Euclidean distance in meters.](image-url)
Line Shape change: Classification results

- The number of groups
- The synthesis of the generalized lines groups

<table>
<thead>
<tr>
<th>Line Code</th>
<th>Group</th>
<th>Original</th>
<th>Simplified</th>
<th>500</th>
<th>1000</th>
<th>200</th>
<th>500</th>
<th>1000</th>
<th>200</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>VSM</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>SM</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>SIN</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>VSIN</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Nu. of Groups
4 4 4 3 4 4 3
Assessment of generalization quality

Assessment Tools

• Measures:
  • The average shape change
  • The average horizontal position
  • The number of groups
  • The synthesis of the generalized lines groups

• Constraints:
  • min shape distortion
  • min hor. position error
  • maintenance of shapes variety
  • preservation of the original line characteristics

Knowledge acquisition

• Comparison of "exemplars” and "automated generalization” results
  • with quantitative measures, without human intervention
  • objective selection
• Shape and structure recognition is applied to the original data
• Procedural knowledge: structural knowledge is related to the selected automatic generalization solution
Shape Description and Structure Recognition [I]

Representative Self-Similar Segments

Example: Ithaca Island

Preliminary segmentation
Shape Description and Structure Recognition [II]

Segmentation Refinement

Final Segmentation and Clustering

Example: Ithaca Island