A Spatial Structuration Heuristic for Integrated Automated Map Generalisation with Attribute and Geometry

Didier G. Leibovici, Jerry Swan, Suchith Anand and Mike Jackson

Centre for Geospatial Science, University of Nottingham, NG7 2RD, UK
Tel. +44(0)115 846 8408 Fax +44(0)115 951 5249
Didier.leibovici@nottingham.ac.uk, www.nottingham.ac.uk/cgs

KEYWORDS: map generalization, classification, competing algorithm, entropy, thematic maps

1. Introduction

Map Generalization is the process by which coarse scale maps are to be derived from fine scale maps, balancing the amount of real-world information with visual confusion, Robinson et al. (1995). This requires the use of operations such as simplification, selection, displacement and amalgamation of features that are performed subsequent to scale change. Recently, focusing on the attribute values of the geometrical objects, some research has been on thematic maps, Liu et al. (2003), Steiniger and Weibel (2005) and Revell (2007), with application in demographic maps, soil maps, land cover and land use maps. In ecological, public health or epidemiological monitoring, presenting maps to decision makers needs often to be "simplified" for better communication or privacy concerns, Waller and Gotway (2004), Curtis et al. (2006).
For these situations, algorithms need to consider the ontology associated with the theme and/or statistical clustering methods as well as geometrical transformations. In Leibovici et al. (2008) two approaches were considered: performing sequentially geometrical and attribute generalisation algorithms, or integrating the two types of transformations in a combined optimisation approach. This paper extends the research on the second approach.
A general generic framework algorithm is described to allow competing optimization at each iteration step. A similar approach is done in Neun (2008) using web services and multi-criteria selection and collaboration filtering for orders of transformations, Foerster et al. (2008). The herein approach is nonetheless seen as operating at a micro-level or atomic level, therefore proposing a real integration between the two types of generalization. The paper focuses particularly on the spatial structuration goal based on a spatial entropy approach. Emphasising on some potential extensions, this generic approach allows also direct comparisons of the sequential and combined approaches, and the possibility of testing families of operators in competition with more traditional ones.

2. Integrated Optimisation approach

This map generalisation approach aims to iteratively transform locally the objects of the map, in order to give a "simpler" visual perception, while "preserving" most of the information. At each iteration, an *admissible transformation* is selected among the collections of geometric transformations and attribute transformations, increasing the most the impression of "spatial structuration" in the map. Within this *minimax* strategy, only the maximisation is really part of the optimisation; the minimisation part is left to the term of *admissible*: initial settings of transformations. A full *minimax* could be done either by looking for the transformation with a maximum heuristic among all families of transformations but minimum among its own family, or by adding a second heuristic.

2.1 Map characteristics

A map $\boldsymbol{M}$ is described by characteristics related to the objects composing the map: the geometry characteristic ***geo***, and the attribute characteristic ***att***. Each characteristic is referring to some object properties and is measured by one or more characteristic descriptions or variables.

- For ***geo*** the characteristic descriptions, $\boldsymbol{c_{geo}}$, can be ***(#vertices**, **circularity**, **orientation ....)*** as here, the objects are seen as polygons, $\boldsymbol{c_{geo}}$ can refer to one or to the whole vector of variables.
- For ***att*** the variables are the attribute variables ***(v1, v2, ...)***, but functions of these can also be

used.

The ***circularity*** ratio, defined as ***4π Area /Perimeter²***, measures the compactness of the shape, with a maximum of one for a circle, and a minimum limit of zero for a narrow and infinitely long shape. The ***orientation*** of a polygon is not clearly defined here, but could be for example the collection of the angles of the segments (making the polygon) and any functional derived from it. This characteristic description could be useful when considering schematisation where one would like to "align" as much as possible the polygons.

2.2 Characteristics classes

Either just by their different values occurring within the map or by a grouping associated to them, the characteristic descriptions define for each characteristic a system of equivalent classes. c_{geo} has kc_{geo},classes, and c_{att} has kc_{att} classes.

These classes illustrate the distributions of the characteristic descriptions (variables).For each characteristic, each of the ***n*** objects in the map belongs to a particular class. During the generalisation process, we have the option of keeping records of the changes of its membership, and/or transforming the object for its characteristic variable values to match the one representing the class it belongs to, *e.g.* for a numerical variable using the mean or median of the object values within the class, for the ***#vertices*** variable, transforming "as much as possible" the objects so their number of vertices (***#vertices)*** gets closer the min or the median for the entire class of polygons, *i.e.* reducing the variation within the initial classes.

Notice it would be possible to split the list of variables attached to a characteristic to define either a multidimensional membership function, or sub-characteristics class memberships, by then enhancing or refining the descriptions and differentiations of the objects of the map.

2.3 Micro Transformations

For attribute and geometric characteristics a set of transformations operating directly on the characteristic classes and/or on the map objects allow the map to evolve to a generalised display. Geometric transformations come from usual approach of operators in generalisation, Jones and Ware (2005), whilst attribute transformations are derived from the general statistical clustering literature, with or without spatial constraint, Leibovici et al. (2008).

2.4 Heuristic for Spatial Structuration

The Shannon entropy ***H()***, of a distribution describing the spread and organisation of the information, has already been investigated in GIS, *e.g.* Bjorke (1996). As the entropy increases the organisation or structuration diminishes and become more uniform. This is illustrated by the fact that for ***k*** classes a uniform distribution will have maximum entropy and a more spiky distribution will have lower entropy, becoming zero for one class getting the entire sample. Using the same principle, the objective function (***Of***) or heuristic value to maximise is, for a characteristic ***i***, for a transformation T_i operating according to the characteristic description c_i on the map ***M***:

$$Of_i(T_i,(c_i,M)) = -H(p_k) = \sum_{k=1}^{n_{c_i}} p_k \log(p_k) \qquad (1)$$

where p_k are proportions of the classes, ($p_k = 0$,the contribution to the entropy is ***0***). To take into account the spatial domain, Li and Claramunt (2006) introduced some spatial weights in this formula, in our approach the focus is on the distribution itself. As we are looking at spatial organisation, the distribution of cooccurrences of classes is used instead, and p_{mk} refers then with a multi-index ***mk*** to the proportion of collocations within a given distance ***d*** of the classes expressed in the multi-index. Figure 1 gives an explicit formulation of the collocation entropy, using the table of spatial cooccurrences, here of order ***3***. Distance of collocation and order of collocation are then parameters of this spatial entropy calculation. In order to normalise the entropy, regarding to the number of classes, one can use the ratio to the uniform case, that is, $p_k = 1/nc_i$, or $p_{mk} = 1/nc_i^3$ for collocation entropy of order ***3***, which gives an entropy of $\log(nc_i)$ or collocation entropy of $\log(nc_i^3)$. With normalised spatial entropy using collocation distribution, the formula (1) becomes:

$$Of_i(T_i,(c_i,M)) = -H_N^{C,d}(p_k) = log(n_{c_i}^3)^{-1} \sum_{mk} p_{mk} \log(p_{mk}) \qquad (2)$$

For the attribute characteristic looking for more spiky distribution of cooccurrences of the classes will clearly transform the map towards a "simpler" perception. For the geometric characteristic this is not always obvious and depends on the characteristic description used. For example with the number of vertices (***#vertices***) as characteristic description, having collocations of polygons with the same number of vertices will not necessarily provide a simple perception, but if some polygons are already simplified (a small number of vertices) the collocations with them will force the algorithm to simplify the neighbourhood polygons as well.

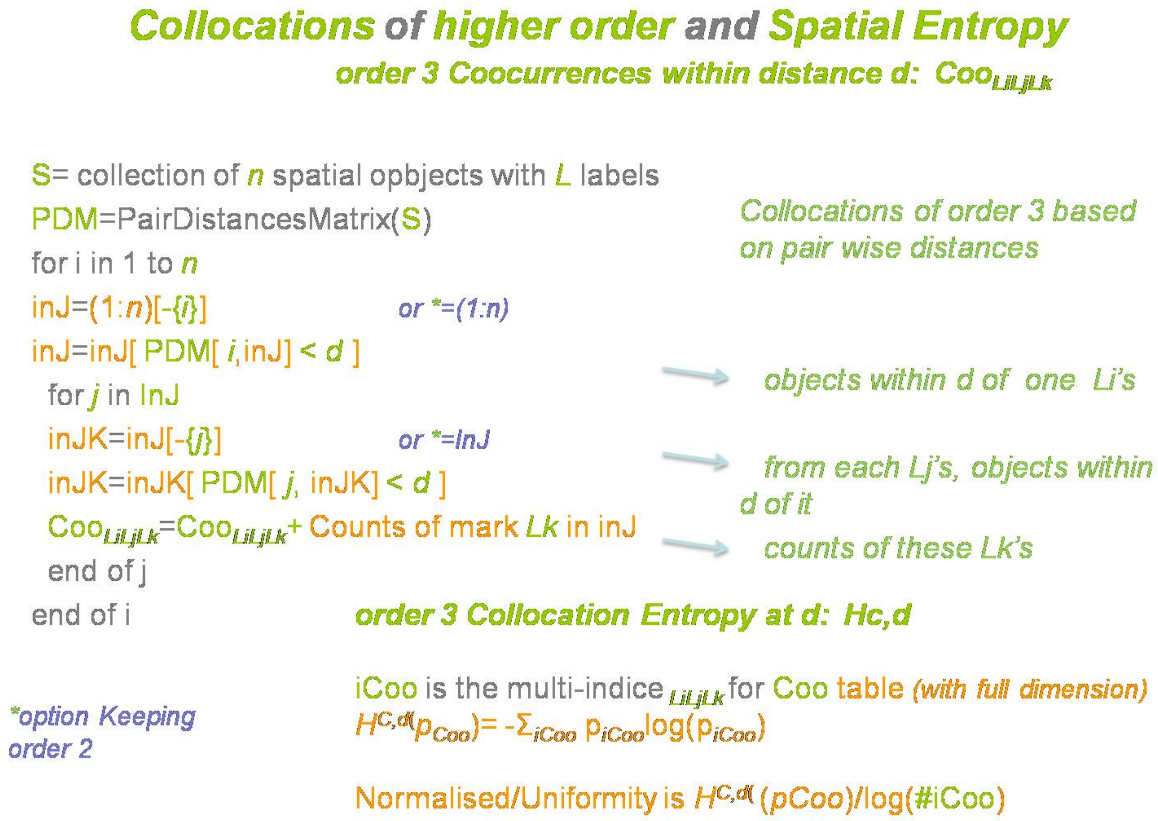

Figure 1. Spatial Collocation Entropy and Algorithm to compute cooccurences table for collocations of order ***3***

2.5 Combined Generalisation Algorithm

For each characteristic seen through their characteristic description, a list of transformation operators has to be defined before running of the algorithm. These transformations are said admissible in term of minimal or micro-transformations simplifying the map with underlined minimum loss of information. Applying a particular transformation may have impacts on the other characteristic descriptions; therefore the optimisation nonetheless combines the characteristics by looking for the best transformation among the whole list of operators according to the heuristic given in equation (2) but also needs to combine these heuristics. This is done either with a multidimensional optimisation:

$$\widetilde{T} = arg \max_{\substack{i \in \{att,geo\} \\ T_{att} \in \mathfrak{F}_{att} \\ T_{geo} \in \mathfrak{F}_{geo}}} (Of_{geo}(T_i,(c_i,M)), Of_{att}(T_i,(c_i,M))) \qquad (3)$$

where $\mathfrak{F}_{att}$ and $\mathfrak{F}_{geo}$ are families or lists of operators associated to the mentioned characteristic. With a simple summation of the heuristics as they are always positive and they have been normalised, one gets:

$$\tilde{T} = arg \max_{\substack{i\in\{att,geo\} \\ T_{att}\in\hat{\mathfrak{F}}_{att} \\ T_{geo}\in\hat{\mathfrak{F}}_{geo}}} (Of_{geo}(T_i,(c_i,M)) + Of_{att}(T_i,(c_i,M)) \qquad (4)$$

Notice the multidimensional optimisation implies the use of a pseudo-order, here in dimension 2; for more than 2 dimensions this may be preferred as the sum.

The use of one micro local transformation may lead to a non significant change in the heuristics, so a strategy could be to increase the number of transformations (composition of operators) until a change of the objective function occurs. The objective function will then look like:

$$\tilde{T} = arg \max_{\substack{i_1,\ldots,i_{n_s} \\ n_s minimal \\ \in\{att,geo\} \\ T_{att}\in\hat{\mathfrak{F}}_{att} \\ T_{geo}\in\hat{\mathfrak{F}}_{geo}}} \sum_{m_c\in\{att,geo\}} Of_{m_c}(T_{i_{n_s}} \circ \ldots T_{i_2} \circ T_{i_1},(c_i,M)) \qquad (5)$$

where the minimality of $\boldsymbol{n_s}$ means it is the smallest number such as the increase of the whole objective function is "significant" (increase or threshold).. In other words, if for $\boldsymbol{n_s = 1}$ there is no best $\boldsymbol{T}$ one looks for the best $T_{i_2} \circ T_{i_1}$ so $\boldsymbol{n_s}$ is set to ***2***, and so forth. The composition order of the $\boldsymbol{n_s}$ operators may play an important role in the optimisation

3. Further Discussion

The generic framework, developed with java using geotools libraries, will be available as a web processing service. The flexibility of this framework allows exploring computationally and algorithmically some properties of the integrated combined approach:

- direct comparison of, and with, sequential approaches,
- multi-competing optimisation
- constraints setting
- schematisation operators

Concrete examples will be demonstrating the potentials expressed in this abstract, in the domain of demographic and public health mapping.

References

Bjorke, J.T (1996) Framework for Entropy-based Map Evaluation. Cartography and Geographic Information Science, 23(2), p.78-95(18).

Curtis, A Mills, JW, Leitner, M (2006) Keeping an eye on privacy issues with geospatial data. Nature 2006, 441:150.

Foerster, T Burghardt, D Neun, M Regnauld, N Swan, J and Weibel, R (2008): Towards and Interoperable Web Generalization Services Framework – Current work in progress. Proceedings of the 11th ICA workshop on Automated generalization and Multiple representation.

Jones, C.B and Ware, J.M (2005) International Journal of Geographical Information Science: Map Generalization Special Issue (Taylor and Francis), Volume 19, Number13 8-9, pages 859-997.

Leibovici, D.G Swan, J Anand, S and Jackson, M (2008) Combining Attribute with Geometry for Automated Generalization and Schematisation. AutoCARTO, USA, 6-10 Septembre 2008.

Li, X and Claramunt, C (2006) A spatial-based decision tree for classification of geographical information , Transactions in GIS, 10(3), Blackwell Publishing, pp. 451-467.

Liu, Y Molenaar, M Ai, T and Liu, Y (2003) Categorical Database generalization aided by Data Model. Proceedings of the 21st International Cartographic Conference 201, 10-16.

Neun, M (2008) Automated processing for map generalization using web services. GeospatialWeb Services workshop University of Nottingham 2008,

http://www.nottingham.ac.uk/geography/geowebservices/
Revell, P (2007) Generic Tools For Generalising Ordnance Survey Base Scale Landcover Data International Cartographic Association 10th conference on Generalization and Multiple Representation, 2-3 August 2007, Moscow.
Robinson, A.H Morrison, J.L Muehrcke, P.C Kimerling, A.J and Guptill, S.C (1995) Elements of Cartography, 6th Edition, New York: John Wiley & Sons.
Steiniger, S and Weibel, R (2005) A Conceptual Framework For Automated Generalization and its Application to Geologic and Soil Maps. Proceedings 22nd International Cartographic Conference, La Coru-a (Spain), 11-16 July 2005.
Waller, L. A., and Gotway, C. A. (2004). Applied spatial statistics for public health data. Wiley series in probability and statistics. Hoboken, N.J., John Wiley & Sons.

Biography

Dr Didier G. Leibovici, is a Research Fellow in geospatial modelling and analysis, with previous posts as statistician in epidemiological/medical imaging research and as geomatician for landscape changes in agro- ecology. Interests refer to interoperability and conflation models with cross-scales for integrated modelling applications within an interoperable framework with chaining web services.

Dr Jerry Swan, is a Research Fellow in Geocomputation. His approach includes optimization, graph theory, symbolic computation, knowledge representation, machine learning, semantics and ontologies. Jerry is particularly involved in the Persitent Test Bed project commissioned by OGC, AGILE and EUROSDR.

Dr Suchith Anand, is an Ordnance Survey Research Fellow at the Centre for Geospatial Science, University of Nottingham. His research interests are in open source GIS, automated map generalization, geohydroinformatics, mobileGIS, location based services, optimization techniques and asset management.

Pr. Mike Jackson is Director of the Centre for Geospatial Science. Prior to this he worked in industry at QinetiQ, Hutchison 3G, Laser Scan in various geospatial specialist and executive roles and in research for NERC. Mike is non-executive director of the Open Geospatial Consortium, and has research interests in combining new technologies such as positioning, pervasive computing and location based services for geo-informatics applications