13th Workshop of the ICA on Generalization and Multiple Representation, Zurich, 12-13 September 2010

The Generalization of the Canadian Landmass: A Federal Perspective

Daniel Pilon[1], Alexandre Beaulieu[1] and Nouri Sabo[1,2]

1 : Centre for Topographic Information, Natural Resources Canada
dpilon@nrcan.gc.ca, abeaulie@nrcan.gc.ca, and nsabo@nrcan.gc.ca

2 : Centre for Research in Geomatics, Laval University, Québec, Canada

1 Introduction

To meet increasing and diversified user needs for geographic information, national mapping agencies (NMAs) must produce and maintain maps and geographic data at multiple scales. A few years ago, the aim of these NMAs was mainly to meet the needs of specialists. But technological development in recent years and particularly the advent of the Internet have promoted the democratization of geospatial data. Now, NMAs must continue to meet not only the needs of professionals, but also those of ordinary citizens. In this context, the need for precise and up-to-date multi-scale spatial data has become more significant. Unfortunately, because of a lack of resources and effective tools, many NMAs are unable to keep their spatial data up-to-date at all scales.

In Canada, this problem is even more important given the extent of the territory to be mapped (about 10,000,000 km^2) and the number of scales to be maintained (1:50,000, 1:250,000, 1:1,000,000 and 1:7,500,000).For example, to cover the country, 13,200 maps at the scale of 1:50,000 or 1,000 maps at the scale of 1:250,000 are maintained. In addition, since 2007, data have been distributed free of charge with an unrestricted use licence. To facilitate and accelerate the production and updating of spatial data at different scales, the Mapping Information Branch (MIB) at Natural Resources Canada (NRCan) has initiated an automatic map generalization project. This project will develop methods and tools for generating data at 1:250,000 from those at 1:50,000.

The purpose of this article is to present a generalization approach developed in this project. This approach will enrich existing generalization algorithms by incorporating integrity constraints, measures and behaviour patterns into them in order to create meta-algorithms. This approach improves the result of traditional generalization algorithms by minimizing conflicts that are generally created when applying those algorithms alone. The developed approach is inspired by SGOs (Self-Generalizing Objects), which are objects that integrate several elements: spatial integrity constraints, behaviour patterns, geometric patterns (forms common to several map objects) and treatment patterns (Sabo 2007). Unlike SGOs, which enrich each map object, the approach that we developed directly enriches generalization algorithms. We will first present the context of this project: the various Canadian mapping programs and the problems with updating data. We will then present the developed approach and the meta-algorithm implementing it (Sherbend) before concluding.

2 The various Canadian mapping programs

To better understand the mapping context in Canada, it is helpful to look back at the two major programs of Canadian base topographic mapping at scales 1:50,000 and 1:250,000, which are the initial acquisition program and the initial updating program. They both cover paper maps and digital maps.

The 1:250,000 mapping program began in 1947 and ended in 1970. In general, two acquisition methods were used, namely, data acquisition directly from aerial photographs, and the derivation and manual generalization of 1:50,000 maps by photomechanical processes (when the 1:50,000 maps were available). During the 1980s, all datasets were digitally converted by scanning and vectorization.

The 1:50,000 mapping program really began in 1949 and will end in 2011 (as of April 2010, there are still 600 maps to produce). In general, three acquisition methods are used: first, stereo compilation from aerial photographs (9,500 maps); second, stereo digitizing from aerial photographs (1,500 maps); and third, monoscopic acquisition from orthorectified satellite imagery (1,200 maps). During the 1990s, all paper maps were digitally converted by scanning and vectorization.

2.1 Problems updating data

Since the first maps were created, updating has been an issue. Various programs have emerged and most of these programs have continued to update maps at 1:50,000 and at 1:250,000 as independent silos for both the analog and digital components. The 1:250,000 map updating program ended in the early 1990s. Since 2000, all of the MIB’s efforts have focused solely on updating the 1:50,000 geographic models, generally using traditional information sources (aerial photographs and satellite imagery). In addition, new partners (Figure 1), namely federal agencies and provinces, have recently been added. They are now responsible for updating specific data layers (roads, hydrography, limits etc.).

As of April 2010, the MIB has a very good northern inventory of updated and accurate 1:50,000 geographic models. However, the data at 1:250,000 are desynchronized in comparison to the 1:50,000 geographic models. There are three different types of desynchronization:

- Data catalogue desynchronization: 1:50,000 and 1:250,000 models are based on different data dictionaries. The semantics for the names of the classes, the names of the attributes and the names of the topological constraints are different;
- Content desynchronization: On average, 1:250,000 models are over 30 years old and only a portion of the 1:250,000 models were derived analogically from 1:50,000 maps;
- Accuracy desynchronization: The 1:50,000 geographic models are accurate (the accuracy is generally better than 30m) whereas the 1:250,000 models are often not precise (imprecision can reach 500m).

The 1:250,000 product is widely downloaded from the MIB distribution site (www.georatis.rncan.gc.ca). This product is in demand in the northern regions of the country for reasons of efficiency (Jolicoeur 2008) because of the vastness of the area (2,500,000 km^2) and low population density (.028 inhabitants per km^2). It is crucial for the MIB to realign those 1:50,000 and 1:250,000 products (scales). The best way to achieve this goal is by the generalization of the 1:50,000 geographic models.

3 Generalization project and generalization approach

In 2008, MIB began a generalization project. The two main objectives of the project are to derive the 1:250,000 geographic models from the 1:50,000 geographic models and to derive 1:250,000 cartographic models from the corresponding geographic models (Figure 1). The 1:250,000 geographic models will be used to create a vector positional product, whereas the 1:250,000 cartographic models will be used to create a raster representation product with cartographic symbology. The intent of the project is to automate as much of the generalization process as possible. Modifications of the mapping specifications are also conceivable to accommodate a higher level of automation

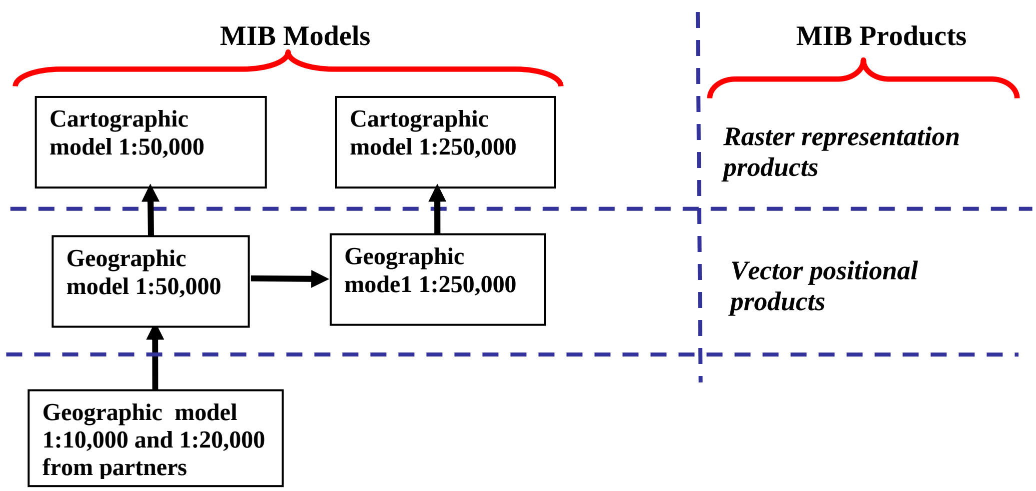

Figure 1: Model used and product distributed at the Mapping Information Branch (MIB)

Traditional generalization methods based on the sequential application of algorithms have proven to be limited (Sabo 2007, Scholl et al. 1996). This is why we decided to develop a generalization approach inspired by "Self-Generalizing Objects" (SGOs) (Sabo 2007). SGOs are objects created from a map's features, which encapsulate several components: 1) integrity constraints to ensure data consistency; 2) treatment patterns which are a set of algorithms to generalize the map's features; 3) behaviour patterns which are the decisional component of SGOs (generalization knowledge held by cartographers); and 4) geometric patterns that are common to several forms of map objects. The geometric patterns are particularly useful in the generalization of buildings. Such integration enables both the facilitation of the generalization process and the reuse of the mapping knowledge. Unlike most approaches that encode generalization knowledge in the mapping system, the SGO-based approach clearly separates mapping knowledge from the generalization system. Thus, in this approach, the generalization knowledge can be stored in a database and can then be automatically transformed into behaviour patterns. Such a separation enables the generalization system to be easily adapted to data with different characteristics.

Contrary to the SGO approach, where mapping knowledge is directly associated with map objects (because an SGO is associated with each map object or group of map objects), our approach implements additional knowledge into existing generalization algorithms to improve their functioning and make generalization easier. In this way, a behaviour pattern and integrity constraints are associated with an existing algorithm to create a meta-algorithm. Spatial integrity constraints preserve certain characteristics of map objects after their generalization with the aid of the meta-algorithm (e.g., a constraint for avoiding self-crossings caused by some simplification algorithms). As for the behaviour pattern, it coordinates the use of the algorithm and the satisfaction of the constraints used by the meta-algorithm. Behaviour patterns are based on the cartographer generalization knowledge.

Generalization algorithms are generally based on mapping knowledge, but this knowledge is often insufficient to adequately support the generalization process and minimize conflicts. This is why they are frequently used in often complex systems. The implementation of additional knowledge into existing algorithms improves generalization quality and minimizes conflicts without resorting to complex systems. Therefore, such an approach would develop autonomous meta-algorithms that can be used in cartographic data production systems that do not have sophisticated generalization tools.

3.1 Meta-algorithm prototype

To test the approach, a meta-algorithm prototype based on the algorithm presented by Wang and Muller (1998) was developed using Python. When the meta-algorithm was implemented, a methodology based on three key steps was adopted.

1. Formalization of generalization knowledge and creation of the behaviour pattern
2. Choice of the algorithm to be implemented and adaptation of this algorithm to support the approach
3. Choice and implementation of integrity constraints to be preserved

The different cartographic generalization rules formerly used by cartographers to analogically derive 1:250,000 maps from 1:50,000 maps were collated. With the help of a senior cartographer, the various generalization scenarios were prepared for each class of information (a scenario can involve one or more classes). The scenarios are written in a pseudo-code language where each generalization statement is defined in terms of a predefined dictionary of measures, constraints and generalization operators. The scenarios define the steps and necessary orchestration required to generalize the classes from 1:50,000 to 1:250,000 geographic models. The scenarios are the input for the behaviour patterns. When the generalization rules are formalized, one must identified the meta-algorithms to develop.

Even though line simplification is a well-known and well-documented subject (Lang 1969; Douglas et Peucker 1973; Jenk 1989; Dougenik 1980; Wang et Müller 1998; etc.), it was the cause of considerable difficulties. For example, the classic Douglas and Peucker (1973) algorithm is not cartographically well adapted for the line simplification of natural features (hydrography, vegetation, wetland, contours and so on) and the results were unsatisfactory. The algorithm that generates the most interesting results for these types of features is the Wang algorithm (Wang and Muller 1998). Despite good results, after evaluation of the results of this algorithm used in a commercial implementation, there were still enough weaknesses remaining (topological problems and poor graphical rendering in dense areas) to motivate the MIB to develop its own implementation in the form of a meta-algorithm (Sherbend). Thus, a new version of this algorithm was created. Sherbend improved several aspects, namely, the simplification of multiple consecutives similar bends, per feature tolerance simplification, constraint enforcement (self-intersection constraint, line-line crossing constraint and sidedness constraint) and improved capacity to resolve line simplification conflicts in dense areas. A behaviour pattern for line simplification was implemented in Sherbend. This pattern orchestrates simplification through an iterative trial-and-error strategy that allows the behaviour pattern to converge rapidly toward a solution. Sherbend implemented at different levels three of the four components of SGOs (generalization algorithm, behaviour pattern, and integrity constraints). However, coupling among the various concepts is too large. This excessively large coupling makes external parameterization and code reusability more difficult. One of the challenges would be to reduce the coupling.

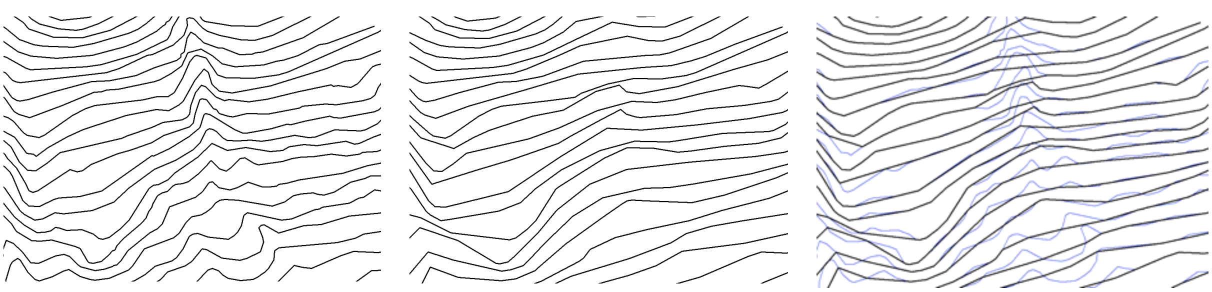

Figure: 2.a Figure: 2.b Figure: 2.c

Figure 2.a: Original contours from 1:50,000; Figure 2.b: Contours simplified using SherBend with a tolerance of 125m (0.5 mm at the scale of 1:250,000); Figure2.c: Combination of figure 2.a and 2.b; Figure 2.a, 2.b and 2.c are displayed at 1:15,000

3.2 Prototype tests

In order to test the possibility of producing and automating map generalization, 16 1:50,000 geographic models were selected and generalized into one 1:250,000 geographic model. The chosen area is representative of the inhabited part of Northern Canada (northern Saskatchewan). The model contains six entity classes. A new 1:250,000 data catalogue containing six entity classes (hydrography, contours, wetland, esker, vegetation and rock in water) was derived from the 1:50,000 data catalogue. Manual preparation and edition (edge match and flow direction on the hydrography network) was necessary to ensure that the 1:50,000 primary models met their own specifications at 100%. When the primary models fully met the specifications, automatic generalization was possible.

Following line simplification (Sherbend meta-algorithm), the six entity classes were generalized to create a geographic model using a sequential approach based on classic generalization operators (selection, amalgamation, collapse, exaggeration, etc.). Generalization of hydrographic networks is the main challenge in the generalization of the Canadian landmass. Over 60% of the occurrences of entities are related to hydrography and, in Northern Canada, this ratio is over 85% (if we exclude contours). In relation to hydrographic networks, it is worth mentioning that the hydrography was processed by watershed and a feature selection of the rivers based on the Horton order (Mazur 1990) was implemented. It produced very good results as demonstrated by Thompson and Brooks, 2007. The same approach should also give the same predictable result for all the hydrography in Canada.

The team's intention being to produce a sample map, during prototype development, we evaluated the possibility of implementing the various concepts elaborated in the generalization scenarios in FME from Safe Software. FME is a sound business choice for the MIB since it is already our basic tool for the manipulation of geospatial data. After assessment of the software, the major gaps in the FME were identified and a contract was signed with the firm to add the necessary tools to develop the prototype.

The generalized 1:250,000 geographic model produced from the prototype was directly printed with the appropriate cartographic symbolization and no cartographic displacement. A group of senior technicians inspected the content qualitatively and compared the printed output geographic model against the official 1:250,000 map, which was analogically derived by photomechanical process from the 1:50,000 maps. In general the result was satisfactory. The most important difference between the two maps was related to the density of information on the map. Most of the time, the density difference was caused by non-compliance with the minimal dimensions and tolerances of the cartographic specifications when analogically deriving the 1:250,000 paper maps. For the MIB, there are two significant conclusions to be drawn from the prototype results. Firstly, implementing an automatic generalization process using meta-algorithm is foreseeable, and secondly, the vast majority of entities in the geographic model will not require cartographic displacement; with the exception of some highly dense areas such as the

MacKenzie Delta, the Thousand Islands etc and for some “less important classes” such as rocks in water, rapids and waterfalls that require displacement for representation purposes. It is important to note that these conclusions apply only to the datasets that contain only natural elements (no man-made features); i.e., 75% of the Canadian landmass datasets.

4 Conclusion and perspectives

In 2010 the MIB has decided that the results of the prototype are encouraging enough to establish a new production process for the generalization of the 1:250,000 geographic model. The MIB would like to obtain results quickly; production will therefore start processing datasets from Northern Canada and will include the same feature classes as those contained in the prototype. As the production system develops, other natural features will gradually be incorporated into the production process. The new 1:250,000 generalized datasets will enable the MIB to create a new 1:250,000 vector positional product and a new 1:250,000 representation product (raster product with cartographic symbology). Both products are the natural continuity of the two 1:50,000 existing products (vector and raster).

For Sherbend, decoupling the various SGO components (treatment patterns, behaviour pattern, geometric patterns) will decrease the complexity of the Sherbend simplification algorithm, improve its parameterization and increase the reusability of these components. The software should also be released under an open-source licence.

For the northern generalization process, create new meta-algorithms which will solve specific generalization problems such rock in water, rapids in river etc.

Lastly, in our opinion, it would be essential to increasingly integrate more SGO concepts into our generalization process in order to deal with generalization problems associated with man-made features that are more complex to generalize in our case.

References

Dougenik J. (1980). Whirpool: A geometric processor for polygone coverage data. *Proceedings of AUTO-CARTO,* Vol. 4, p. 304-311.

Douglas D.H. and Peucker T.K. (1973), Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, *The Canadian Cartographer,* Vol. 10(2), p. 112-122.

Jenk G.F. (1989). Geographic logic in line generalization,. *Cartographica*, Vol. 26(1), p. 27-43.

Jolicoeur, P. (2008). The Canadian Northern Geospatial Infracstructure, *Geomatica*, Vol. 62, No. 2 pp. 125-138.

Lang T. (1969). Rules for the robot droughtsmen, *The Geographical Magazine*, Vol. 42(1), p. 50-51.Mazur E. Ruzak, Castner H. W. (1990). Horton's ordering scheme and the generalization of river networks, Cartographic Journal,, vol. 27, no2, pp. 104-112

Sabo M.N. (2007). Intégration des algorithmes de généralisation et des patrons géométriques pour la création des objets auto-généralisants (SGO) afin d'améliorer la généralisation cartographique à la volée, *PhD thesis, pp. 224*, Laval University, Canada.

Scholl M., Voisard A., Peloux J.P., Raynal L. et Rigaux P. (1996). *SGBD Géographiques Spécificit,. Paris, International Thomson Publishing*, France, *1996, 185 p.*

Tompson, R., Brooks, R., (2007). Generalization of Geographical Networks, in *Generalization of Geographic Information: Cartographic Modelling and Applications*, edited by: Mackaness, Ruas, Tiina Sarjakoski, Italy, Elsevier, pp.255-267.

Wang Z., Muller J.-C. (1998). Line Generalization Based on an Analysis of Shape Characteristics, *Cartography and Geographic Information Systems,* 25(1), 3-15.