ICA Workshop on Generalisation and Multiple representation; 20-21 August 2004 - Leicester

Data Enrichment for Adaptive Generalisation

Moritz Neun, Robert Weibel, Dirk Burghardt
Department of Geography, University of Zürich, Switzerland, *neun@geo.unizh.ch*

Abstract

This paper presents the research framework for a new project funded by the Swiss National Science Foundation, called DEGEN (Data Enrichment for automated map GENeralisation) which supports two PhD projects. With a focus on thematic cartography (GIS maps, web mapping, location-based services) and the data types prevailing in these map types (polygonal subdivisons, points of interest, etc.), the project aims at developing methods for enriching cartographic databases as well as representing multiscale data of several levels of details (LODs) in multiresolution databases (MRDB). The key working hypothesis is that more adaptive real-time generalisation can be achieved by integrating generalisation methods with MRDB, and particularly by exploiting so-called horizontal and vertical relations. Horizontal relations describe patterns at a particular LOD (e.g. clusters, alignments), while vertical relations are described by links that associate homologous objects or patterns across two or more LODs. The paper reviews the pertinent literature, identifies the research issues and defines the research objectives for the project.

Introduction

Today maps and visualisations are no longer limited to printed maps. Interactive and dynamic media such as web mapping and mobile devices require near-real time generalisation. Therefore feasible and performant algorithms are needed.

Many generalisation algorithms are only suitable for specific problems. The term ”adaptive generalisation” expresses the intention to use the right algorithm for a given generalisation task. The selection of the most sophisticated algorithms - particularly when several algorithms are integrated into a comprehensive workflow - is a complex decision making process and depends critically on the availability of auxiliary information to inform the generalisation process. Typically spatial databases do not contain much information that could support this decision making process. Therefore data enrichment is necessary to equip the ”raw” spatial data with additional information about the objects and their relationships. Such auxiliary information is used for a variety of purposes within the overall generalisation process.

- *Characterisation*: Map generalisation seeks to maintain important map objects, patterns, and relationships, while suppressing unimportant ones. Hence, the spatial and semantic characteristics of map objects have to be detected in order to obtain priority orderings among map objects; meaningful groups of objects have to be formed (e.g., clusters of buildings, or objects aligned in a particular arrangement); and spatial and semantic relationships have to be detected (e.g., adjacency and proximity relations, hierarchical relations) before sensible and informed decisions can be made about generalisation. It is important to note that such rich information is typically not available in spatial databases that are available for mapping, as these data-bases are usually designed to serve multiple purposes, and because comprehensive anticipatory object characterisation would be far too costly.

- *Conflict detection*: Generalisation is only warranted if cartographic principles (i.e., cartographic constraints) are violated. Conflicts are detected by comparing the results of the characterisation step with thresholds imposed by cartographic principles (e.g., minimal size or distance thresholds for the target scale).

- *Algorithm and parameter selection*: Depending on the character of the data to be generalized and the conflicts detected for the target scale, appropriate algorithms have to be selected to remedy the conflicts. Auxiliary data relating to the character of the data should help not only to select the most appropriate generalisation algorithm(s) from a range of potential candidates, but also to set appropriate values for the parameters that control these algorithms.

- *Evaluation*: Since the overall generalisation process may involve many algorithms, it is critical that the success of individual and overall generalisation operations is evaluated, particularly in systems with a capability for self-evaluation such as multi-agent systems (MAS) or evolutionary systems. Once again, the same methods can be used to generate the necessary auxiliary information as for continuous characterisation, conflict evaluation, and algorithm selection and parameterisation.

Combining vertical and horizontal relations

Current data enrichment methods mainly focus on *horizontal relationships*, which exist on the same level of detail (LOD). Horizontal relationships represent common structural properties (e.g. neighborhood, pattern, alignments). They are used in a broad range of generalisation algorithms (see *Detecting common patterns*). Maps often are collections of many different patterns. They can also represent *vertical* relationships between homologous objects or object groups. These links can be based on attribute but especially on geometric relations.

Multiresolution Databases (MRDB) offer the opportunity to represent *vertical relationships* (i.e. relationships between map objects across scales). Generalisation from a MRDB then allows the exploitation of *horizontal* and *vertical relationships* between objects or object groups. Not only the analysis of the geometric relationships but also, even more, the analysis of semantic relationships benefit from this vertical extension. Vertical relationships often represent hierarchic partonomic (part-of) relations, which exist between the objects and their composite object on a higher level of detail. [Ruas, 2000] introduced so called meso agents for the grouping of objects. They can be very useful for detecting and preserving pattern emanating from one LOD over large changes in scale.

Usage of enriched data

Due to the sometimes costly algorithms for analysis and generalisation the use of partly pre-computed data in a MRDB offers great potential as [Cecconi, 2003] showed. Also the GEMURE project [Bernier et al., 2004] aims at combining cartographic generalisation and MRDB systems. Generalisation algorithms for a complete palette of generalisation operations are controlled by a variety of parameters were shown by [Galanda, 2003] using a self-evaluating agent system [Barrault et al., 2001]. The agent concept may be extended using the new enriched data available from the MRDB.

Most thematic maps (e.g. vegetation or soil maps, zoning maps) rely on polygonal data [Follger et al., 2003]. Therefore our investigation is focused mainly on this data type. Generalisation algorithms for polygonal data need to collectively implement the complete palette of generalisation operations which can be envisioned for this purpose: Simplification and smoothing of polygon boundaries, polygon elimination, polygon enlargement, polygon aggregation, polygon displacement, and partial collapse of polygons.

Although research on this class of algorithms is not as rich as for topographic map generalisation, a number of researchers (e.g. [de Berg et al., 1998], [Saalfeld, 1999], [Bader and Weibel, 1997], [Jones et al., 1995] and [Müller and Wang, 1992]) have contributed to extending the tool set for polygonal map generalisation. In our investigation, we concentrate on algorithms for vector data.

It is interesting to note that most of these existing generalisation algorithms rely on auxiliary geometric data structures, due to the contextual nature of the generalisation operations that they attempt to implement. Geometric data structures that are useful to represent contextual topological and/or proximity relations include the Delaunay triangulation ([DeLucia and Black, 1987], [Jones et al., 1995]), the Voronoi diagram ([Gold, 1999], [Hangouët, 1998]), minimum spanning trees and other graph structures ([Mackaness and Beard, 1993], [Regnauld, 2003]), skeletons [Bader and Weibel, 1997], and hierarchical partitioning schemes ([Ruas, 1995]). Since they represent spatial relationships not originally represented in the ’raw’ polygonal data these geometric data structures can also be seen as a form of data enrichment.

Data analysis for adaptive generalisation

In the process of data enrichment for adaptive generalisation the available data in a first step has to be analysed for generating the enriching information. According to [Beard, 1988], [Beard, 1991] and [Ruas and Lagrange, 1995] we can distinguish the characterisation of geometric primitives and the modelling of spatial and semantic relations. In a later step the representation and storage of this newly acquired information will be discussed.

Geometric primitives and object properties

Geometric properties have been rather thoroughly investigated for many different purposes and object types. Shape measures for line sinuosity include measures based on distance relations between vertices of boundaries ([McMaster, 1986], [Dutton, 1999]), based on the analysis of inflection points ([Plazanet et al., 1995], [Plazanet, 1996]), or based on epsilon bands ([Mustière, 1995]). The techniques proposed by [Plazanet, 1996], and [Mustière, 1995] can also be used to isolate and characterise individual meaningful shapes, such as individual bends of a line.

Object properties (attributes) support the generalisation process of objects according to basic schemata. Schema generalisation operators such as classifcation, aggregation and association need the consideration of spatial and semantic relations between objects [Ruas and Lagrange, 1995].

Spatial and semantic similarity measurement

A basic way of describing spatial relations is the usage of a topological model with connectivity, containment or adjacency [Egenhofer and Franzosa, 1991]. In [Beard, 1988] and [Beard, 1991] also the importance of spatial distances (proximal relations) is emphasized. Finally, as a third category, directional relations [Clementini et al., 1997] are of interest. These spatial properties can describe a similarity between spatial objects. Combining both spatial and semantic similarity measures, e.g. spatial and semantic features of an object for measuring its containment in a specific area, offers better matching and grouping results.

[Tversky, 1977] introduced the notion of similarity expressing an intermediate state where objects are neither ”identical” nor ”different”. The single properties or abstract attributes (such as quality or complexity) of each object define the degree of similarity between the objects. He also describes asymmetry in semantic relationships (e.g. a hospital is a building but a building is not in every case a hospital). [Rodriguez and Egenhofer, 2004] extended this similarity measurement with relevance factors by weighting the attributes. They propose the Matching-Distance Similarity Measure (MSDM) to extend the similarity measurement of [Tversky, 1977] by a semantic distance calculation and contextual relations through the weighting. [Yaolin et al., 2002] propose a generalisation procedure combining set theory, classification hierarchy and attribute structure to measure the semantic similarity among object types in a classification hierarchy. These procedures are included into a data model for the generalisation of categorial databases (with classification and aggregation hierarchy). Both [Rodriguez and Egenhofer, 2004] and also [Yaolin et al., 2002] use in their models hierarchical structures for the representation and measurement of similarity distance. These structures are capable of representing semantic and topological relations, however, they lack the combination with strictly geometric properties (e.g. proximal and directional relations).

Detecting common patterns

Maps usually contain a number of patterns, which consist within an object or between objects that are represented with sufficient regularity [Mackaness and Edwards, 2002]. The objects can have equal or similar attributes or other object properties such as size or shape which correspond. Cartographically meaningful patterns include clusters of polygons (that may potentially be aggregated to a single polygon in generalisation) or alignments of polygons (e.g., a series of polygons aligned in a chain). Detection of clusters can to some extent build on spatial clustering algorithms ([Han et al., 2001]) and more particularly graph based algorithms ([Regnauld, 1998], [Anders, 2003]), with the limitation that these algorithms are typically designed for point clusters rather than polygons and hence tend to work best for small polygons (e.g., buildings) that can be represented by their centroid. Voronoi diagrams ([Hangouët, 1998]) or polygon buffering may have some potential for alternative solutions for irregular polygons. The detection of alignments or small polygons has been studied by [Christophe and Ruas, 2002] for the case of buildings in topographic maps.

While a considerable body of literature exists on shape characterisation methods that may be exploited for cartographic data enrichment there still exists a missing link between the knowledge that can be obtained by cartographic pattern recognition and description on the one hand (structural knowledge) and the knowledge embedded in the generalisation algorithms that can be used to transform the source data to the target map (procedural knowledge). For a complete automation of the generalisation process, it must be possible to infer from the results of the data characterisation process the appropriate generalisation algorithms to modify the source data such that they are properly generalized for the target scale. For example, once a polygon’s size has been identified as being too small to be legible on the target scale (i.e., a cartographic conflict has been detected) two options exist for the choice of generalisation algorithm: the polygon can either be removed or enlarged. Hence, it must be decided, based on the results of the characterisation process, which of the two algorithms is used (and if enlargment is used, by what factor the polygon is enlarged). Research on building these links between structural and procedural knowledge is still scarce. Examples include work using inductive machine learning in the context of road generalisation algorithms ([Plazanet et al., 1998], [Mustière and Zucker, 2002]) as well as empirical tuning of algorithm and parameter selection for building generalisation in a multi-agent system [Regnauld, ress].

Data enrichment using vertical and horizontal relationships

Another weak area relates to the lack of research on actually representing and storing extracted relations and patterns in an cartographic database. To our knowledge, no relevant research exists in this domain for the types of data that we will be concerned with (i.e. polygonal data for thematic mapping), except for limited research on multiscale topographic databases. Patterns, representing structural knowledge, and algorithms, representing procedural knowledge, usually apply to horizontal relations. The preservation e.g. of important patterns over large scales has also a strong vertical implication. Hence, additionally to combining the structural and procedural knowledge horizontally they can also benefit from vertical links between different scales in a MRDB.

Objects in neighboring LODs, and the structural and procedural knoweledge about them, can be represented in a *vertical relationship*. This may provide hints to guide the generalisation algorithms. Using a MRDB to represent these vertical relationships, as the work by [Cecconi, 2003] suggests, offers a considerable potential, particularly in time-critical mapping applications such as web mapping or LBS.

Enriching a MRDB with geometric and semantic information can serve the following purposes:

- *Description*: Store additional descriptive, geometric and semantic (structural) information and procedural knowledge that can inform adaptive generalisation.
- *Acceleration*: Pre-computed information for generalisation algorithms is stored in the MRDB (e.g., pre-computed tolerance values for line simplification may be stored to obtain real-time performance).
- *Substitution*: Generalisation algorithms that are too time-consuming to compute in a time-critical context (e.g., object aggregation) may be substituted by storing pre-computed versions).

Research on MRDB for map generalisation is relatively sparse and has so far exclusively concentrated on databases for topographic mapping. Examples include the work by [Abraham, 1989], [Jones et al., 1996], and [Harrie and Hellström, 1999].

An issue that has received particular attention is the task of building a MRDB by matching homologous map objects of separate databases representing different LODs and linking them across scales. [Flewelling, 1999] introduced a method for measuring similarities between spatial datasets. [Timpf and Frank, 1997] proposed acyclic graphs for representing cartographic objects in a multi-scale data structure. In [Timpf, 1998] hierarchical structures for aggregation, generalisation and filtering are introduced. [Sester et al., 1998] presents techniques to match and link objects of different spatial data sets by integration and aggregation. [Sheeren et al., 2004] introduce an approach to deal with representation differences in the matching and data integration process for preserving object consistency.

Research objectives and outlook

The main working hypothesis of this paper and the reported project is that integrating map generalisation and MRDB offers the potential to represent and exploit vertical as well as horizontal relationships, an approach that has not been reported in the literature so far. The specific focus of this project is on problems of thematic cartography and hence the particular data types prevailing in thematic maps (polygonal subdivisions, irregular polygons, points of interest, etc.)

Our project of adaptive generalisation aims at combining structural and procedural knowledge for LOD matching and representing it in a MRDB. In order to do so, we address the following main objectives:

- delivery of a set of methods (or measures) for the recognition and description of complex cartographic patterns (e.g., clusters and alignments of irregular polygons); use examples from polygon and/or linear feature generalisation, e.g. topograpic constraints (slope orientation) or also topologic constraints through different datasets;

- integration of structural knowledge (obtained in cartographic pattern recognition) and procedural knowledge (represented in generalisation algorithms);
- combination of the *horizontal* structural and procedural knowledge with *vertical* links between different scales in a MRDB; introduce hierarchical structures for representing those *vertical* links; supply methods for the horizontal (pattern) and vertical matching of objects

The overall objective of this project is therefore to develop - with a focus on data enrichment for thematic cartography - better methods for complex pattern and structure recognition and to better integrate structural and procedural information into generalization procedures as well as cartographic, multiscale databases.

References

[Abraham, 1989] Abraham, I. M. (1989). *Automated Cartographic Line Generalization and Scale-Independent Databases*. PhD thesis, University of Glamorgan.

[Anders, 2003] Anders, K. H. (2003). A hierarchical graph-clustering approach to find groups of objects. In *Proceedings 5th Workshop on Progress in Automated Map Generalization*, Paris, France. Accessed: 05/10/2003.

[Bader and Weibel, 1997] Bader, M. and Weibel, R. (1997). Detecting and resolving size and proximity conflicts in the generalization of polygonal maps. In *Proceedings of the 16th ICA/ACI Conference*, pages 1525–1532, Stockholm.

[Barrault et al., 2001] Barrault, M., Regnauld, N., Duchêne, C., Haire, K., Baeijs, C., Demazeau, Y., Hardy, P., Mackaness, W., Ruas, A., and Weibel, R. (2001). Integrating multi-agent, object-oriented and algorithmic techniques for improved automated map generalization. In *Proceedings of the 19th ICA/ACI Conference*, pages 2110–2116, Beijing.

[Beard, 1988] Beard, K. (1988). *Multiple representations from a detailed database: a scheme for automated generalization*. PhD thesis, University of Wisconsin-Madison.

[Beard, 1991] Beard, K. (1991). Constraints on rule formation. In Buttenfield, B. and McMaster, R., editors, *Map Generalization: Making Rules for Knowledge Representation*, pages 121–135. Longman, London.

[Bernier et al., 2004] Bernier, E., Bédard, Y., and Moulin, B. (2004). Des cartes à la carte? il suffit de demander. *GéoInfo, Bulletin du Plan géomatique du gouvernement du Québec*.

[Cecconi, 2003] Cecconi, A. (2003). *Integration of Cartographic Generalization and Multi-Scale Databases for Enhanced Web Mapping*. PhD thesis, Department of Geography, University of Zurich.

[Christophe and Ruas, 2002] Christophe, S. and Ruas, A. (2002). Detecting building alignments for generalisation purposes. Symposium on Geospatial Theory, Processing and Applications, Ottawa.

[Clementini et al., 1997] Clementini, E., Di Felice, P., and Hernández, D. (1997). Qualitative representation of positional information. *Artificial Intelligence*, 95:317–356.

[de Berg et al., 1998] de Berg, M., van Kreveld, M., and Schirra, S. (1998). Topologically correct subdivision simplification using the bandwidth criterion. *Cartography and Geographic Information Systems*, 25(4):243–257.

[DeLucia and Black, 1987] DeLucia, A. A. and Black, R. T. (1987). A comprehensice approach to automatic feature generalization. *Proceedings of the International Cartographic Conference, Morelia, Mexico*, pages 169–192.

[Dutton, 1999] Dutton, G. (1999). Sinuosity and point selection in digital line generalization. *Cartography and Geographic Information Systems*, (20):33–53.

[Egenhofer and Franzosa, 1991] Egenhofer, M. J. and Franzosa, R. (1991). Point-set topological spatial relationships. *International Journal of Geographic Information Systems*, (5).

[Flewelling, 1999] Flewelling, D. (1999). Measuring similarities of spatial datasets. *URISA Journal*, 11(1):45–52.

[Follger et al., 2003] Follger, D.-M., Richter, A., and Scharfe, W. (2003). Automationsgestützte Generalisierung von flächenhaft dargestellten Reliefeinheiten - Geomorphographische Übersichtskarte der Bundesrepublik Deutschland 1:1 Million. *Kartographische Nachrichten*, 3:126–133.

[Galanda, 2003] Galanda, M. (2003). *Automated Polygon Generalization in a Multi Agent System*. PhD thesis, Department of Geography, University of Zurich.

[Gold, 1999] Gold, C. M. (1999). Crust and anti-crust: A one-step boundary and skeleton extraction algorithm. In *ACM Symposium on Computational Geometry (Miamy, Florida)*.

[Han et al., 2001] Han, J., Kamber, M., and Tung, A. K. H. (2001). Spatial clustering methods in data mining: A survey. In Miller, H. and Han, J., editors, *In Geographic Data Mining and Knowledge Discovery*, pages 187–217. Taylor&Francis.

[Hangouët, 1998] Hangouët, J. (1998). *Approche et méthodes poul l'automatisation de la généralisation cartographique; application en bord de ville*. PhD thesis, Marne la Valée.

[Harrie and Hellström, 1999] Harrie, L. and Hellström, A. K. (1999). A protortype system for propagating updates between cartographic data sets. *The Cartographic Journal*, 36(2):133–140.

[Jones et al., 1995] Jones, C. B., Bundy, G. L., and Ware, J. M. (1995). Map generalization with a triangulated data structure. *Cartography and Geographic Information Systems*, 22(4):317–331.

[Jones et al., 1996] Jones, C. B., Kidner, D. B., Luo, L. Q., Bundy, G. L., and Ware, J. M. (1996). Database design for a multi-scale spatial information system. *International Journal of Geographic Information Systems*, 10(8):901–920.

[Mackaness and Edwards, 2002] Mackaness, W. and Edwards, G. (2002). The importance of modelling pattern and structure in automated map generalisation. Joint Workshop on Multi-Scale Representations of Spatial Data, Ottawa. CCRS - Canada Centre for Remote Sensing.

[Mackaness and Beard, 1993] Mackaness, W. A. and Beard, M. K. (1993). Use of graph theory to support map generalization. *Cartography and Geographic Information System, Vol. 20, No. 4*, pages 210–221.

[McMaster, 1986] McMaster, R. (1986). A statistical analysis of mathematical measures for linear simplification. *The American Cartographer*, 13(2):103–116.

[Müller and Wang, 1992] Müller, J. C. and Wang, Z. (1992). Area-patch generalization: A competitive approach. *The Cartographic Journal*, 29(2):137–144.

[Mustière, 1995] Mustière, S. (1995). Mesures de la qualité de la généralisation du linéaire. Rapport de Stage de DESS, Université Paris I / Ecole Nationale Supérieure de Géographie.

[Mustière and Zucker, 2002] Mustière, S. and Zucker (2002). Généralisation cartographique et apprentissage automatique á partir d'exemples. In Ruas, A., editor, *Généralisation et représentation multiple*, pages 353–368. Hermes, Paris.

[Plazanet, 1996] Plazanet, C. (1996). *Enrichissement des bases de données géographiques: analyse de la géométrie des objets linéaires pour la généralisation cartographique (application au routes)*. PhD thesis, Université de Marne-la-Vallée.

[Plazanet et al., 1995] Plazanet, C., Affholder, J.-G., and Fritsch, E. (1995). The importance of geometric modeling in linear feature generalization. *Cartography and Geographic Information Systems*, 22(4):291–305.

[Plazanet et al., 1998] Plazanet, C., Bigolin, N. M., and Ruas, A. (1998). Experiments with learning techniques for spatial model enrichment and line generalization. *GeoInformatica*, 2(4):315–333.

[Regnauld, 1998] Regnauld, N. (1998). *Généralization du bâti: Strucuture spatiale de type graphe et répresentation cartographique*. PhD thesis, Université de Provence, Aix-Marseille 1.

[Regnauld, 2003] Regnauld, N. (2003). Algorithms for the amalgamation of topographic data. *In: Proceedings of the 21st International Cartographic Conference, Durban.*

[Regnauld, ress] Regnauld, N. (in press). Constraint based mechanism applied to automated map generalisation using agent based modelling techniques. *Computers, Environment and Urban Systems*.

[Rodriguez and Egenhofer, 2004] Rodriguez, A. and Egenhofer, M. (2004). Comparing geospatial entity classes: an asymmetric and contextdependent similarity measure. *Geographical Information Science*, 18(3):229–256.

[Ruas, 1995] Ruas, A. (1995). Multiple paradigms for automated map generalization: Geometry, topology, hierarchical partitioning and local triangulation. In *Proceedings AutoCarto 12*, pages 69–78, Charlotte, USA.

[Ruas, 2000] Ruas, A. (2000). The roles of meso objects for generalisation. In *Proceedings 9th Symposium on Spatial Data Handling*, pages 3b50–3b63, Beijing, China.

[Ruas and Lagrange, 1995] Ruas, A. and Lagrange, J. P. (1995). Data and knowledge modelling for generalization. In Müller, J. C., Lagrange, J. P., and Weibel, R., editors, *GIS and Generalization. Methodology and Practice. GISDATA 1*, pages 73–90. Taylor&Francis.

[Saalfeld, 1999] Saalfeld, A. (1999). Topologically consistent line simplification with the douglas-peucker algorithm. *Cartography and Geographic Information Science*, 26(1):7–18.

[Sester et al., 1998] Sester, M., Anders, K. H., and Walter, V. (1998). Linking objects of different spatial data sets by integration and aggregation. *GeoInformatica*, 2(4):335–358.

[Sheeren et al., 2004] Sheeren, D., Mustière, S., and Zucker, J.-D. (2004). Consistency assessment between multiple representations of geographical databases: a specification-based approach.

[Timpf, 1998] Timpf, S. (1998). *Hierarchical Structures in Map Series*. PhD thesis, Technical University of Vienna, Austria.

[Timpf and Frank, 1997] Timpf, S. and Frank, A. (1997). Cartographic objects in a multiscale data structure. In Craglia, M. and Couclelis, H., editors, *Geographic Information Research - Bridging the Atlantic*, pages 224–234. Taylor&Francis.

[Tversky, 1977] Tversky, A. (1977). Features of similarity. *Psychological Review*, 84:327–352.

[Yaolin et al., 2002] Yaolin, L., Molenaar, M., and Kraak, M. J. (2002). Categorical database generalization aided by data structure. In Wise, S., Brindley, P., Kim, Y. H., and Openshaw, C., editors, *Proceeding GIS research UK 10th Annual Conference*, pages 345–348. University of Sheffield, Sheffield.