DATA ENRICHMENT FOR ROAD GENERALIZATION THROUGH ANALYSIS OF MORPHOLOGY IN THE CARGEN PROJECT

Sandro Savino, Massimo Rumor, Matteo Zanon, Igor Lissandron

University of Padova, Department of Information Engineering,
via Gradenigo 6/B 35131 Padova, Italy

ABSTRACT

One of the most used operators during the generalization process is “selection”.
Data selection is used both in model generalization and in cartographic generalization, and can be driven using the knowledge stored in the attributes of a feature (e.g. delete all the huts), the information stored in its geometry (e.g. delete all the building smaller than 50sqm) or both.
Unfortunately sometime the information explicitly available is not enough and more data needs to be acquired using different sources, with a process called data enrichment.
In this paper we will show how through morphological analysis it's possible to extract further information from a feature and provide that extra data that allows to perform and drive data selection.
This paper will show how this concept has been used in the context of road generalization in the Italian CARGEN project.

INTRODUCTION

The automatic cartographic generalization is a field of research active since long time. Thanks to the progress done in years of studies and of contributions by many, some concrete solutions have been developed and at the present day some NMAs are already using [Lecordix 2006, Stoter 2007] or planning to use in their production workflow softwares that are able to perform cartographic generalization.
In Italy, since some years, a project named CARGEN (CARtographic GENeralization) is active. The project, funded by the Regione Veneto, is carried on by students and researchers at the Department of Information Engineering of the University of Padova, with the collaboration of the national geographic institute (IGMI, Istituto Geografico Militare Italiano). The objective of the CARGEN project is to study and develop informatic techniques to achieve automatic cartographic generalization. The project started his life in 2007 studying the generalization of the regional geodatabase in 1:5000 scale (DBT) to the IGM geodatabase in 1:25000 scale (IGM DB25) and recently its scope has been extended to include also the IGM maps at 1:50000 scale.
In this paper we will focus on some of the work that has been done on the generalization of the road network from the 1:5000 scale to the 1:25000 scale; while the explanation of the complete generalization process of the road network is beyond the scope of this article, we will show how morphology helped us to devise some original solutions in the context of road junctions and highways generalization.

WORKING WITH ROADS

Roads, and road networks in general, are usually represented as a graph composed by many edges; because of their extension, the number of edges composing a road network can be very high: one of the most used operators to generalize such a big quantity of simple features is selection. The selection operator, albeit being probably the easiest to explain, can be very complex to implement due to the problem of driving the selection in a clever way.
Usually selection rules can be derived from some specifications on the target map, or assumptions on the size of objects at the target scale; selection can be then driven by knowledge stored in the attributes of the features (e.g. delete all the huts), the information stored in their geometries (e.g. delete all the building smaller than 50sqm) or both.

Problems arise when the input data model doesn't convey enough information to perform a correct or complete selection: during cartographic generalization an input model that is not rich enough could hinder an automatic selection of the features to be generalized (e.g. to be deleted, collapsed or typified) from those that do not need to be generalized, while during model generalization it might impede to correctly select and classify the features within the target model taxonomy. While generalizing the road network during the CARGEN project, we faced both these situations: in the following sections we will show how we used the analysis of shape and form to enrich our data and drive the generalization of highways and of junctions in (ordinary) roads.

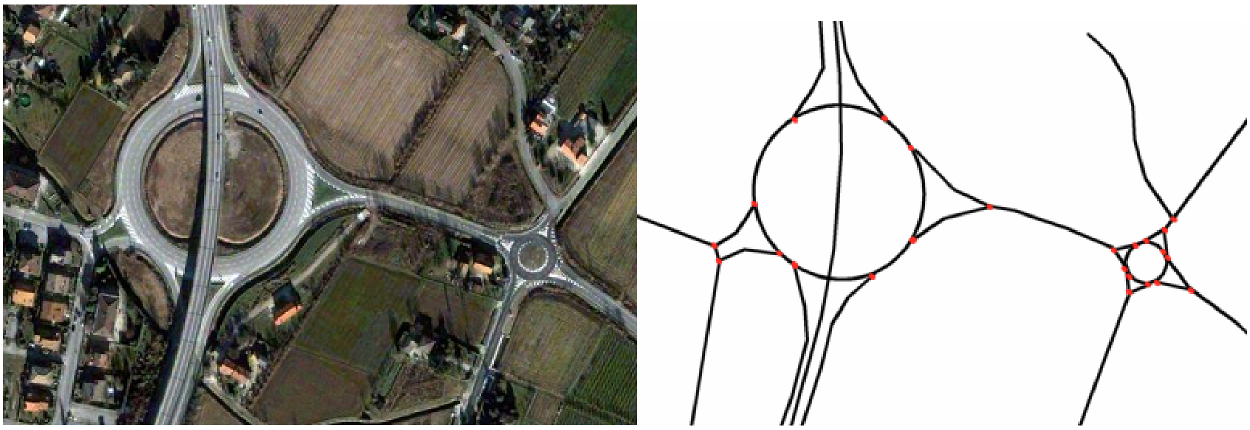


Figure 1: what a human eye sees in the situation depicted in figure is a roundabout that needs to be simplified to be represented at a smaller scale. What the computer can see is a set of connected edges of the road graph, not different to other parts of the network. The computer's "perception" could be improved if those edges were classified as crossing roads or slip roads. The purpose of data enrichment is to give the computer a better "insight" on the phenomena represented in a map.

THE PROBLEM

Developing an automatic process to generalize the road network from 1:5000 scale to 1:25000 scale in the CARGEN project, the generalization of road junctions and of highways were two big challenges.
The problem with the generalization of highways was a model generalization problem: in our output data model, the IGM DB25, there exists a specific object for the highway toll stations, the highway slip roads, the highway rest areas and the highway carriageways

whereas the input data model, the DBT 1:5000, had only a feature "highway", not further specialized, from which to derive all those objects.
The problem with road junctions was a typical cartographic generalization one: while input and data model matched, in the input data some road junctions were described with too many edges that at target scale would clutter the visualization. The IGM specifications for the DB25 stated specifically that every roundabout of radius smaller than 25 m should be collapsed to its centroid and, more generally, that road junction should be “simplified”. To fulfill these specifications it was necessary to recognize both the roundabouts and the “complex” road junctions (i.e. those needing to be simplified), even if the output data model didn’t explicitly requested this information.
Both to generalize highways and road junctions we had to enrich our data model; as there was no mean to extract the needed information from the semantic data of the input model, we resorted to use the geometric information, more specifically the shape and form.
Three factors influenced the decision to use morphology to enrich our data:

- the first was a matter of necessity, as we could not access to any other data source that could provide the needed information;
- the second was the desire/aim to develop an algorithm that was robust to semantic error: as a matter of fact input datasets may contain both semantic and geometric errors; while geometric errors are more easy to detect and to some extent to correct (e.g. with a topological validation), semantic errors are more subtle. Developing an algorithm that relies more on geometric data than on semantic data allows to generalize even when the input dataset contains some classification errors;
- the last was the challenge to mimic the human capability of reading a map: if the human eye could easily detect a complex road junction as well as tell the different parts of a highway, could we develop an algorithm with similar ability?

RELATED WORK

Morphology, intended as the study of shape and form, is an important topic in the field of generalization: one of the main aims when generalizing a map is in fact to maintain the form and shapes represented in the input map.
The shape and form of the features have been studied, measured and characterized (see for example [Agent 2000]) and researches have been done to understand the perception of shapes and forms [Wertheimer 1938], [Thomson and Richardson 1999], [Thomson and Brooks 2000].
All these information have been usually used to drive the generalization process; in the examples that we will present in this paper instead, we use morphology at an earlier stage, to reclassify features or to refine the existing classification in order to gather a better knowledge of what is represented on the map and operate a more conscious generalization.

The research on the analysis and generalization of road network has been very intense too, because of the main role played by roads in maps. Most of the authors working on road networks use the concept of strokes derived from the work of [Thomson and Richardson 1999] on perceptual grouping: in a road network represented by a graph, we can define a stroke as a chain of edges that are joined on the principle of good continuation.
The concept of strokes is much used both in the analysis and in the generalization of road networks; strokes are usually built on the basis of straightness, but in some cases other data can be taken in consideration, using information directly from the input model (e.g.

road names), or enriching the data calculating new metrics [Claramunt 2004], [Heinzle 2005].

Although strokes are a very important tool in generalization, strokes alone can not provide a complete solution to the generalization of road junctions and highways; some works that address more specifically these topics are those of [Mackaness and Mackechnie 1999, Thom 2005, Touya 2007].

In [Mackaness and Mackechnie 1999] the authors propose an interesting approach to generalize road junctions, using cluster analysis and graph theory. Their idea is to find the road junctions as the regions where the nodes of the road network are more dense. This is done clustering the vertices of the road network and applying a “granularity” threshold to create the clusters. Every cluster represents a road junction. The graph representing each junction is then created and contracted: the junction is simplified collapsing all the vertices of the cluster to the centroid and connecting all the edges to it. Changing the granularity threshold is possible to control the level of generalization of the junction, by collapsing more or less vertices. Although the algorithm proposed to detect and generalize road junctions produces viable results, the choice of the right granularity is still an open question; furthermore, as noted by the same authors, in some instances the results were not acceptable leading to what they defined “the collapsing star effect”.

In [Thom 2005], the problem of collapsing dual-carriageway is addressed with a four-steps algorithm that builds the strokes from the road sections, pairs the strokes, collapses each pair and connects the resulting line work with the remaining road network. The author notes that because the direction of the slip roads is almost tangential to the main roads, building the strokes only on the basis of straightness leads to unpredictable results. This problem is solved using the direction of the road (stored in the input data model) to develop a method of tracking one-way sections.

In [Touya 2005] the author describes a full and generic process to allow road network selection in model generalization. The author orchestrates many different algorithms in a process entailing four steps: data enrichment through structures and pattern recognition, rural selection based on assessing traffic by shortest path computing, street selection algorithm based on road block aggregation and structures typification. The classification of road junctions is achieved by classifying first simple road junctions analyzing, at every node of the road graph, the angles between the incident edges; complex road junctions can be then found as particular aggregation of simple ones. Unfortunately this classification process is not explained in details.

In this paper we will describe an original approach to classify road junctions and the sections composing a highway network using solely geometric information. Strokes are used, but they are built only on the basis of straightness. Road cycles are used to detect road junction and the distinction between slip roads and the carriageways of highways is done using a special metric (the bend ratio). The classification process is explained in further details in the following sections. All the algorithms have been programmed in Java, using Oracle Spatial for the spatial queries.

ROAD JUNCTIONS

When generalizing road junctions, the first problem to solve was to detect those that needed to be generalized. This was like asking: “what makes some road junctions so complex that they need to be simplified?” The answer that we found is "redundancy": the difference between a "simple" junction and a "complex" one is the presence of short edges (e.g. slip roads, access ramps) that create redundant connections in the graph.

Figure 2: simple and complex road junctions

Since a redundant edge in a graph create a cycle, our algorithm finds the junctions to generalize by looking for all the cycles in the road graph; of course, as the road graph is highly cyclic, we had to set a threshold: we empirically set it as 250 m of maximum perimeter length.
What the algorithm finds is a set of cycles of different sizes and shapes that may be isolated or adjacent to other cycles.
The most recognizable junctions are probably the roundabouts: testing the "roundness" of every cycle (perimeter to area ratio similar to 4π/p) we could easily find them; this however left many cycles still unclassified.
As it was clear by visual inspection of the results, some of the cycles found were part of more complex junctions. Then, in order to look at the broader picture, we merged together all the adjacent cycles and calculated how many points the boundary of the resulting merged cycle had in common with the road graph: we found out that the number of these points (called special nodes) and the type of junction represented by the merged cycles were related and so this could be a good way to classify them.
We built the strokes on the basis of the gestalt principle of "straightness" connecting the most straight chain of edges passing through the special nodes. Strokes could be built just “locally” on the road edges touching the road junction; from our experiments, the best results were obtained not considering any semantic information (e.g. road name or classification) of the edges: in some cases, in fact, the original classification changed right after the road junction, thus preventing the construction of longer strokes.

Figure 3: road junctions are detected using road cycles. Each road junction can be formed by more than one road cycle

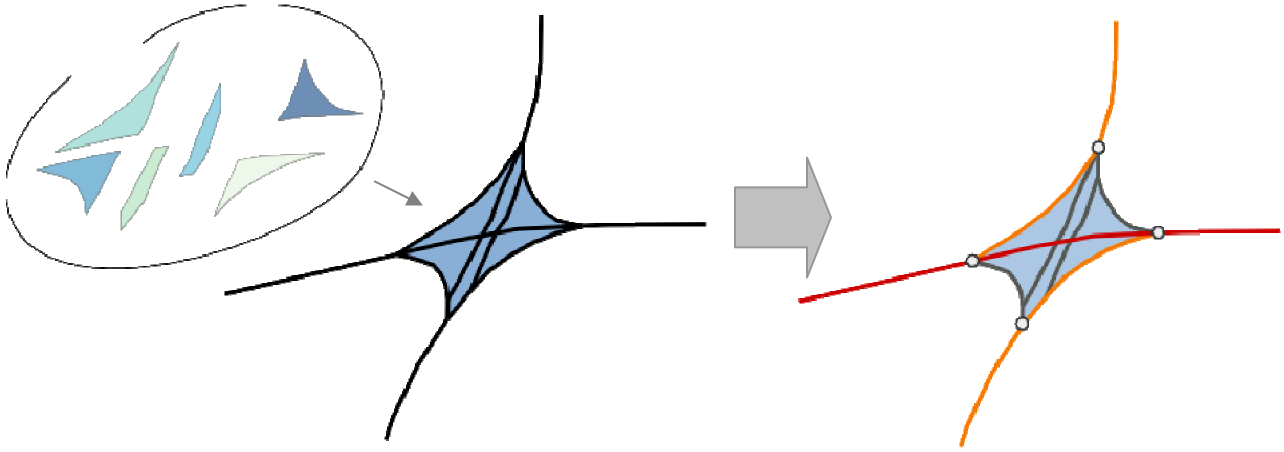


Figure 4: road cycles are merged together into a merged cycle. The points where the boundary of the merged cycle touches the road network are marked as special nodes (the gray-outlined white dots). Strokes are generated and on the basis of how many special nodes each of them crosses, they are classified as crossing roads (red), incoming roads (orange) and internal roads (gray).

Depending on how many special nodes a stroke was crossing (one or two), we classified the strokes as crossing roads (crossing 2 special nodes) and incoming roads (crossing only 1 special node). All the remaining strokes were classified as internal roads.
On the basis of the number of special nodes and the number and type of strokes of each junction, we could further classify the road junctions in:

- T-junctions
- Paired T-junctions
- Crossroads

Each road junction not falling in one of these classes is tagged as “unclassified junction”.

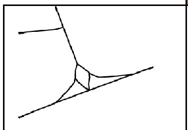
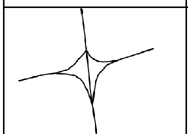
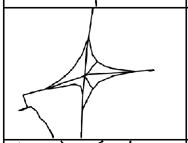
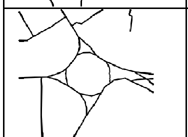
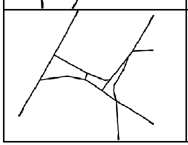
	Junction type	Number of special nodes	Number of crossing roads	Number of incoming roads
	T-Junction	3	1	1
	Paired T-Junction	4	1	2
	Crossroads	4	2	0
	Roundabout	(classified at earlier stage)		
	Unclassified	(any junction not falling in the criteria above)		

Table 1: the relation among the type and the number of special nodes, crossing roads and incoming roads in each road junction

Visually inspecting the results of the algorithm, we found that it performs in accordance with the expectations: roundabouts, T-junctions, paired T-junctions and crossroads are correctly detected and classified most of the time and what the algorithm tags as “unclassified junction” are usually junctions that are arguably difficult to classify, even for a human. In some cases, though, there are some false negatives: T-junctions, paired T-junctions and crossroads can end up in the “unclassified” group because of a single edge touching the boundary of the merged cycle, thus increasing the number of special nodes over the thresholds. False positives can also happen, in particular road cycles with three special points can be mistakenly classified as T-junctions. A concavity test is used to avoid this case: since a real T-junction should have slip roads to connect smoothly the crossing road with the incoming road, and slip roads by design have a concave shape, the merged cycle of a real T-junction should be contained by a triangle drawn on its three special nodes. Empiric tests revealed that it is sufficient to compare the area size of the triangle built on the three special nodes with that of the merged cycle to filter out false T-junctions.

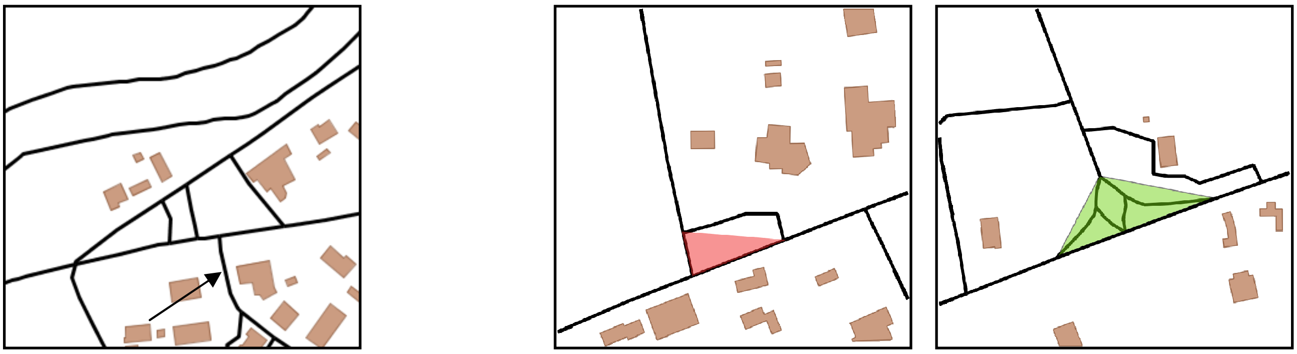


Figure 5: (left) a T-junction is treated as an "unclassified junction" because the edge indicated with an arrow increases the number of special nodes. (right) testing two candidate T-junctions: the first fails the concavity test (red triangle), the second passes (green triangle).

At the end of the process, using solely information retrieved from the geometry of the features, we were able to find the road junctions, classify each edge and also group the junctions in different types. On the basis of the number of special nodes, the number of incoming roads and the number of crossing roads we could devise an ad-hoc algorithm to generalize each type of junction: as a rule of thumb, all the internal roads are deleted, the crossing roads are preserved and the incoming roads are eventually edited to smooth their connection to the crossing roads. Roundabouts may be collapsed to a point or replaced with a circle depending on their size.

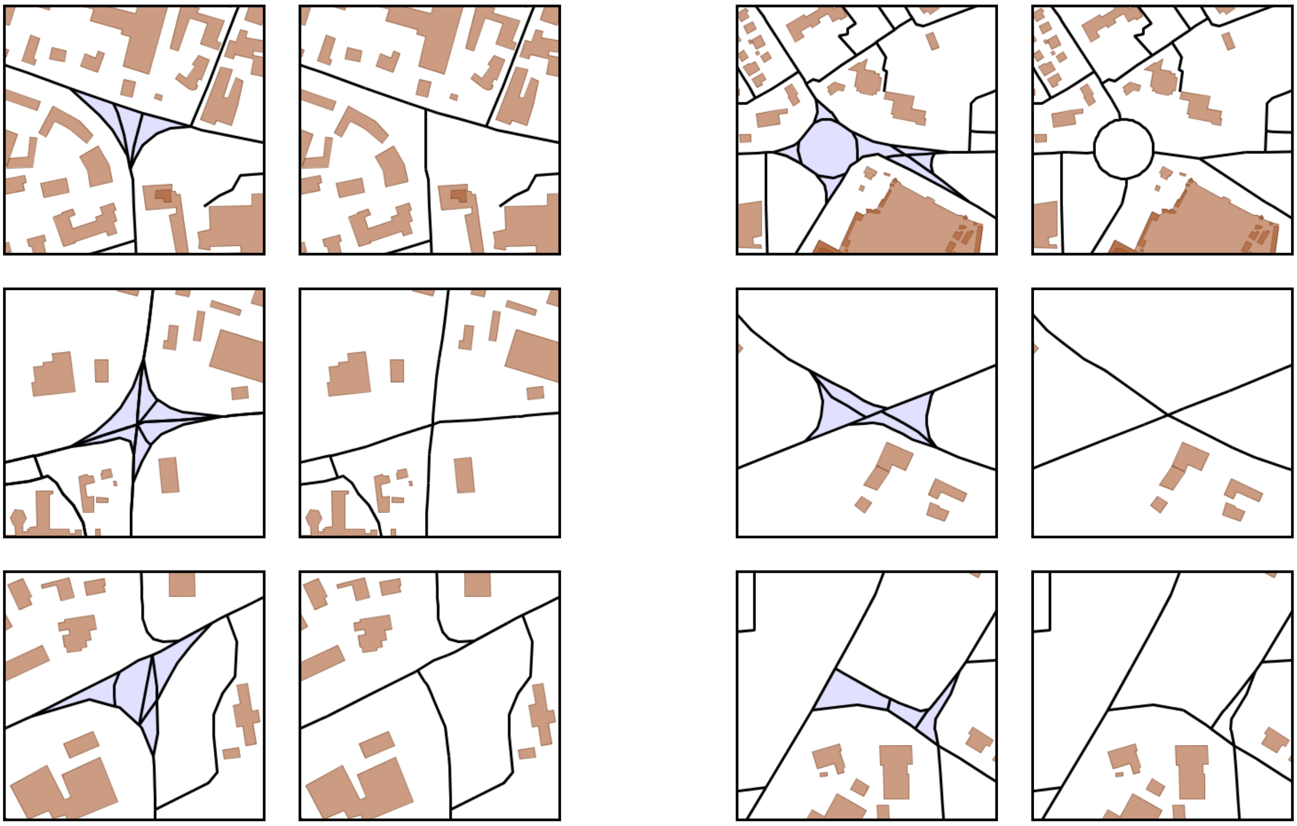


Figure 6: some results of the generalization of road junctions. From left to right, top to bottom: T-junction, roundabout, crossroad, paired T-junction and unclassified junctions. In the case of unclassified junctions the algorithm provides a "best effort" solution.

HIGHWAYS

In Italy highways are a special part of the road network: they run isolated from the ordinary roads and the connection to them must pass through a toll gate; the highway network can be considered then a sub-graph of the whole road network. The most relevant features in the highway network are the two carriageways: other features are connected to them, as rest areas, slip roads and toll plazas. In our input data model, all the edges belonging to the highway graph are only classified as "highway" and not further specialized.

Once again the study of form and shape of the edges composing the highway graph enabled us to tag each of them as part of different objects. If with road junctions the first step was to detect the road junctions, similarly with highways the first problem to solve was how to find in the highway graph the edges that form the main carriageways.

The first thought one has when thinking of a highway is something long, continuous and straight; from this very thought we moved our first step toward the solution: we found among all the edges the longest and the most straight and we classified it as "carriageway". This first edge was used as a "seeding edge": starting from it we grew the carriageway adding all the edges connected to it first in one direction and then in the opposite. This procedure went on until a fork was met.

A fork in the highway means either that there is a slip road joining or leaving the carriageway, either because the highway splits in two directions or because there is an exit. As noted by others [Thom 2005], because slip roads are by design close to tangential when joining or leaving their dual carriageway, straightness alone is not sufficient to create reliably strokes from dual carriageways. Because of the function of slip roads anyway, their

property of being tangential is required only locally, in close proximity to the junction with the carriageway: looking “further away”, the slip road changes its direction (e.g. to route the traffic to a rest area, a toll station, or another highway).
To construct the strokes from the carriageway we devised a metric, called bend ratio, that takes into account the way the direction of an edge varies. The bend ratio of an edge A composed by n vertices a_0, a_1, a_2, …. a_n is defined as

$$\text{bend ratio} = \frac{\sum_i L_i \, diff_i}{L}$$

where

- L is the length of the edge A
- L_i is the distance between two consecutive vertices a_{i-1} and a_i
- $diff_i$ is the difference between the angle of the segment from a_{i-1} and a_i and the angle of the segment from the first to the last vertices of the edge A

The value of the bend ratio increases the less the edge is rectilinear.

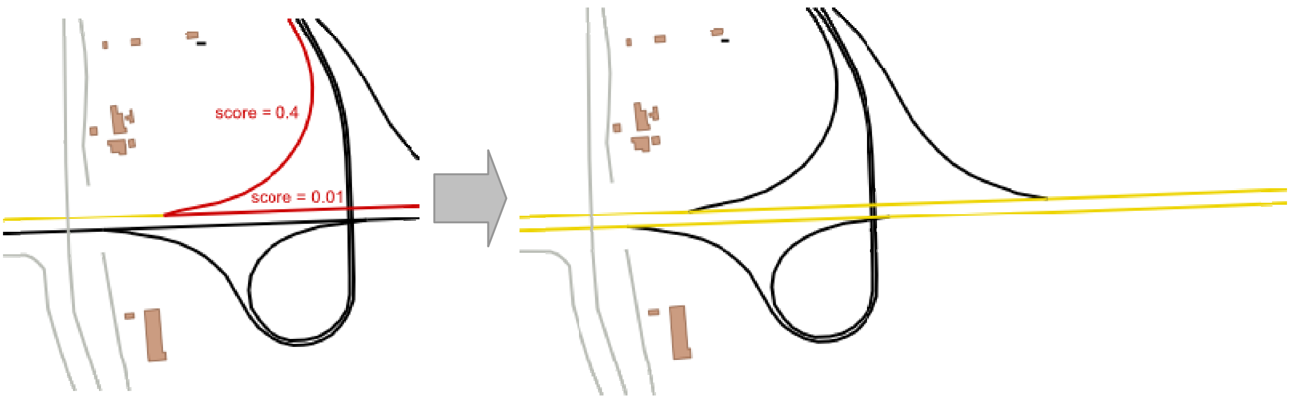


Figure 7: the score of the bend ratio is used to build the strokes from the carriageway

Using the bend ratio slip roads can be distinguished and the construction of the stroke continues along the carriageway; when a carriageway cannot be extended further, a new seed edge is searched among the edges not tagged and the process starts again.

Figure 8: classification of carriageways (left) and slip roads (right). In the pictures, in gray the network of ordinary roads, in black the highway network, in yellow a group of carriageway (left) and some slip roads (right).

At the end of the process all the carriageway and the slip roads connected to them are identified and classified, but there are still other objects to be recognized: rest areas and toll stations. Some rules derived from conditions that apply to them helped us in doing so:

- toll stations are the only edges of the highway graph allowed to be connected with the ordinary roads and are either at the end of a carriageway or connected to a slip road;
- rest areas are portions of the highway graph that are connected only to the same carriageway and can not be connected to normal roads

These restrictions apply in Italy, but other more general characteristic can be used to classify them: for example in a toll plaza the presence of the toll booths divides the carriageway in many lanes, while a rest area can be identified by the presence of a graph that has sharper bends -since it's intended for low speed traffic.

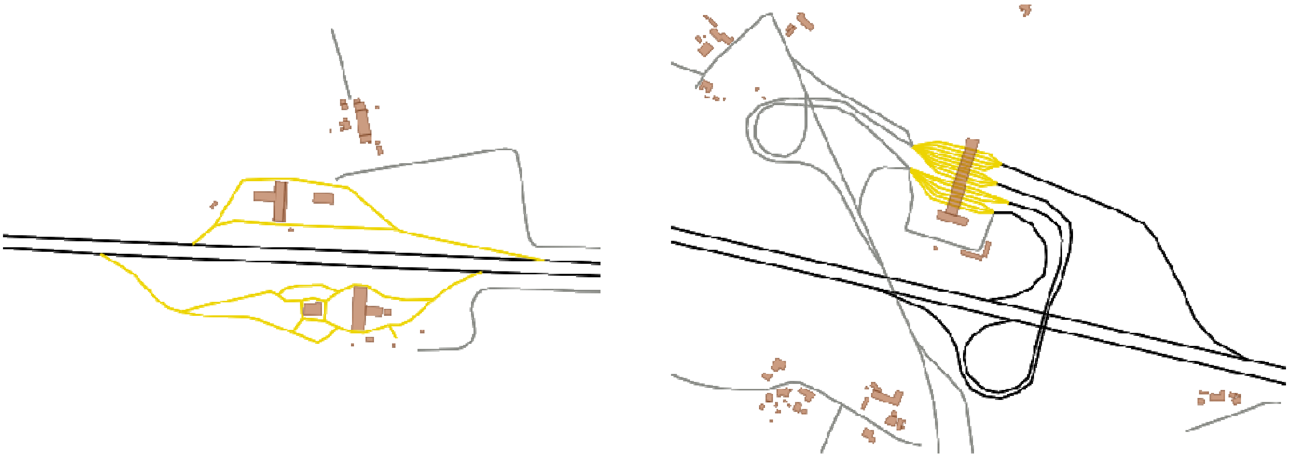


Figure 9: rest areas (left) and toll stations (right) can be recognized due to their particular shape and their connection with carriageway and the road network. In the pictures, in gray the network of ordinary roads, in black the highway network, in yellow a rest area (left) and a toll station (right).

Applying the conditions above, also toll plazas and rest areas have been classified.
Having set up a set of rules to identify the objects we are interested in, we can then be confident that one edge not matched by any of these does not need further attention and can be discarded: this is an important aspect to consider as it helps to avoid that errors in the input dataset are passed to the generalized one.

Figure 10: (left) an error in the classification of the input data is detected: the edge classified as highway (red) is too short to be a highway and, as a slip road, it is not connected to anything; it will then not considered part of the highway. (right) The classification of slip roads and carriageways is reliable also in complex cases.

Once all the objects that form a highway have finally been classified, it's possible to generalize them; this task is performed for each of them by a different algorithm: the multiple lanes at the toll plazas are simplified, carriageways are coupled and collapsed to a single line, resting areas are simplified and slip roads are tested in order to maintain the connection among all the different parts of the highway.

Figure 11: generalization of the highway: original data (top), generalized data (bottom).

CONCLUSIONS

In this paper we briefly presented how we solved some aspects of the generalization of the road network from the 1:5000 scale to the 1:25000 scale in the CARGEN project.
We described an original approach to classify road junctions and the sections composing a highway network with the purpose to enrich the input data model and drive the process of selection and generalization. In particular we showed how this was done using solely the information stored in the geometry of the features through the analysis of their form and shape.

REFERENCES

AGENT project (1999), *Selection of Basic Algorithms, AGENT project technical report*. Online documentation available at http://agent.ign.fr/deliverable/DD2.pdf

Chaudhry O., Mackaness W. (2005), *Rural and Urban Road Network Generalization Deriving 1:250,000 form OS MasterMap*, in proceedings of the International Cartography Association Conference, La Coruña (Spain), 2005.

Edwardes A.J., Mackaness W.A. (2000), *Intelligent Generalization of Urban Road Networks*. In proceedings of GIS Research UK Conference (GISRUK'2000), York, UK, 2000. pp. 81-85.

Heinzle F., Anders K.-H., Sester M. (2005), *Graph Based Approaches for Recognition of Patterns and Implicit Information in Road Networks*, in proceedings of the International Cartography Association Conference, La Coruña (Spain), 2005.

Jiang B., Claramunt C. (2004), *A Structural Approach to the Model Generalization of an Urban Street Network*, GeoInformatica, Vol. 8, Issue 2, June 2004

Jiang B., Harrie L. (2004) *Selection of Streets from a Network Using Self-Organizing Maps* Transactions in GIS, 2004, 8, 335 - 350

Marc van Kreveld, J. Peschier (1998), *On the Automated Generalization of Road Network Maps*, in proceedings of the Third International Conference on Geocomputation, 1998, University of Bristol, Bristol, United Kingdom

Mackaness, W., Beard, K. (1993), Use of Graph Theory to Support Map Generalisation. Cartography and Geographic Information System (CaGIS), 20(4), 1993, pp. 210-221

Mackaness, W., Edwards, G. (2002), *The Importance of Modelling Pattern and Structure in Automated Map Generalisation*. In Proceedings of Joint Workshop on Multi-scale Representations of Spatial Data, Ottawa, Canada (2002)

Mackaness W., Mackechnie G.A. (1999), *Automating the Detection and Simplification of Junctions in Road Networks*. GeoInformatica, Vol. 3, Issue 2, 1999. pp. 185-200.

Savino S., Rumor R., Lissandron I. (2009), *Road junction generalization in large scale geodatabases*. In A. Krek, M. Rumor, S. Zlatanova, E. M. Fendel (Eds), UDMS Annual 2009, Taylor & Francis

Thom S. (2005), *A Strategy for Collapsing OS Integrated Transport Network Dual Carriageways*, in proceedings of the 8th ICA Workshop on Generalization and Multiple Representation, La Coruña (Spain), 2005.

Thomson R., Richardson D. (1999), *The "Good Continuation" principle of Perceptual Organization applied to the Generalization of Road Networks*, in proceedings of the 19th ICC. ICA, Ottawa, Canada (1999).

Thomson R.C, Brooks R. (2000), Efficient Generalisation and Abstraction of Network Data Using Perceptual Grouping. In: Proceedings of the 5th International Conference on GeoComputation, Manchester (CD-Rom)

Touya G. (2007), *A Road Network Selection Process Based on Data Enrichment and Structure Detection* in proceedings of the 10th ICA Workshop, Moscow, 2 - 3 August 2007

Wertheimer M. (1938). *Laws of organization in perceptual forms*. First published as Untersuchungen zur Lehre von der Gestalt II, in Psycologische Forschung, 4, 301-350. Translation published in Ellis, W. (1938). A source book of Gestalt psychology (pp. 71-88)., 1923.

Zhang Q. (2004), *Modelling Structure and Patterns in Road Network Generalization*, in proceedings of ICA Workshop on Generalisation and Multiple Representation - Leicester - 20-21 August 2004