Automatic Detection of Ports For Map Generalisation

William Mackaness
The School of GeoSciences,
The University of Edinburgh,
Edinburgh EH8 9XP
william.mackaness@ed.ac.uk

Phil Bartie
Biological and Environmental Sciences
University of Stirling
Stirling FK9 4LA
phil.bartie@stir.ac.uk

Abstract[1]

The vision of automated cartography is built around the idea of a highly detailed representation of the world from which we can construct a continuum of dynamic outputs ranging to the very smallest scale. Deriving higher order objects from the grouping of lower order constituent parts can be framed as a pattern recognition task. We can devise a prototypical representation of a higher order object – represented as an ontology. The ontology can then be used as a basis for searching the database for higher order objects (eg cities, airports, ports) based on the proximity between their constituent parts. A city defined in terms of {suburbs, municipal buildings, transport networks}; airports in terms of {runways, passenger terminals, taxiways}, and ports in terms of {harbours, docks, container ports}). Such an approach provides a basis for extending the range of scale dependent representations of geographical concepts – in turn facilitating the automated creation of thematic maps. By taking account of the geographical composition of such higher order objects we can extend automated map generalisation over larger changes in scale. Such thinking chimes with ideas in Gestalt theory (in reaction to structuralist approaches) in which the geographical concept emerges from the juxtaposition of a particular set of finer scale concepts.

Introduction

Much research in map generalisation is structuralist in nature; that is to say it breaks the problem down into distinct and unrelated elements that unfortunately often extinguish the notion of geography. The considerable and sustained focus around the Douglas Peuker Algorithm is an excellent example of structuralist thinking with all its associated problems and limited application domain. As an antidote to such thinking, increasingly researchers are exploring ways of making explicit the metric and topological properties *between* collection of entities in order to construct higher order entities (for example that the concept of a city can be considered to be made up a dense collection of transport networks, municipal and industrial buildings and residential areas). This idea of automatically detecting entities from their 'sub entities' has been explored by (among others) Chaudhry et al (2009) in which they demonstrated the automatic detection of schools, retail parks and airports (defined through the detection of their functional elements). Thus 'playing fields', 'car parks', 'classrooms', 'sports facilities' might variously define the extent of a School and from a map generalisation perspective, we can envisage replacement of these finer scale objects with a single polygon (or its centroid) at smaller scales. Such an approach worked well for entities composed of contiguous objects (eg airports, retail parks, railway stations) but some higher order objects are much more complex – in their composition (how they are constituted), their geographical extent, and how they 'interact' with other phenomena. An intriguing example is the entity 'Port'.

[1] Mackaness, W.A. & Bartie, P. 2016 Automatic Detection of Ports For Map Generalisation, Workshop of the ICA Commission on Generalisation and Multiple Representation 'Automated generalisation for on-demand mapping', Helsinki, Finland, 14 June, 2016

How do we define a Port?

From a functional perspective we can think of a port that connects sea based networks with land based networks. They lie at the 'interface' between land and sea, often associated with coastal towns; the nature of 'goods being transferred' means there are a range of spaces that provide the function of storage and ship repair (Alderton 2008). They have enourmous strategic significance and are a defining quality of some cities (e.g. Rotterdam). From close inspection of various ports (for example 'Southampton' Figure 1) we can derive an ontology (Noy & McGuiness 2016) of 'port' (Figure 2).

Figure 1a & 1b: Two representations of the Port of Southampton

Variable	Contains the word	Variable	Contains the word
Amenity	boat	man-made	breakwater
Amenity	ferry	man-made	crane
Boundary	maritime	man-made	lighthouse
Building	industrial	man-made	pier
Building	warehouse	man-made	silo
Building	train station	man-made	storage
Craft	boat builder	military	naval base
Craft	sail	natural	bay
Emergency	life	natural coast	
Historic	ship	power	plant
Landuse	commercial	public transport	station
Landuse	industrial	railway	rail
Landuse	military	railway station	
Landuse	port	service	siding
leisure	marina	service	yard
Leisure	slipway	route	ferry
Seamark	Berth	waterway	canal
Seamark	Boatyard	waterway	dock
Seamark	slipway	waterway	boatyard

Table 1: Various ways of recording entities that describe Ports in OSM

The features we find on the map reflect the functions of a port. The ontology of a 'Port' is made up of 'sub entities'. OSM is rich in attribution (if a little suspect in positional quality) and researchers have been exploring its utility in the context of map generalisation (Touya 2012; Touya and Reimer 2015). Table 1 shows some of the entities variously 'littering' the OpenStreetMap database (Coleman 2013; Fairbairn & Al-Bakri 2013). 'Littering' in the sense that there is no explicit modelling of 'Port' in OSM, only of the entities that constitute it (ie its sub entities). We may not need to inspect the geometry of these sub entities. Instead we can simply search for their abundance and proximity (Figure 3). In other words if we find some ‘constellation’ or proximal arrangement of (some) these entities then we can justifiably call it a ‘Port’. It is on this basis that we can create a map at smaller scale (Figure 1b).

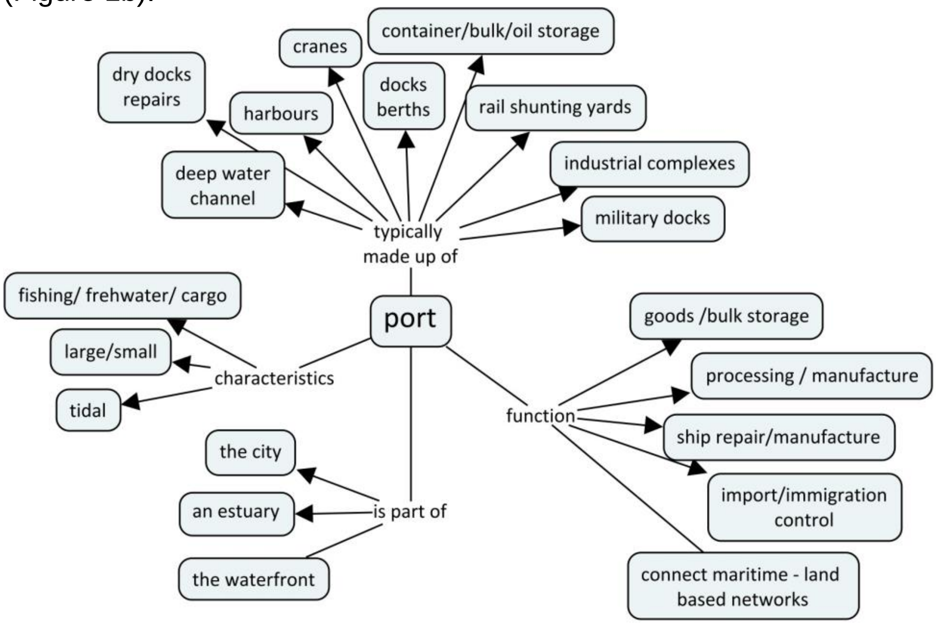

Figure 2: A port ontology derived from visual inspection of various UK Ports

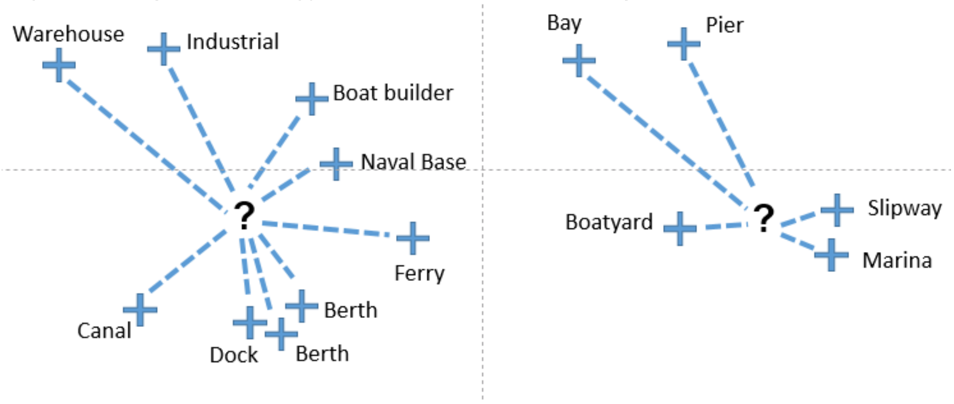

Figure 3: A constellation of sub entities – revealing (perhaps) the presence of a Port

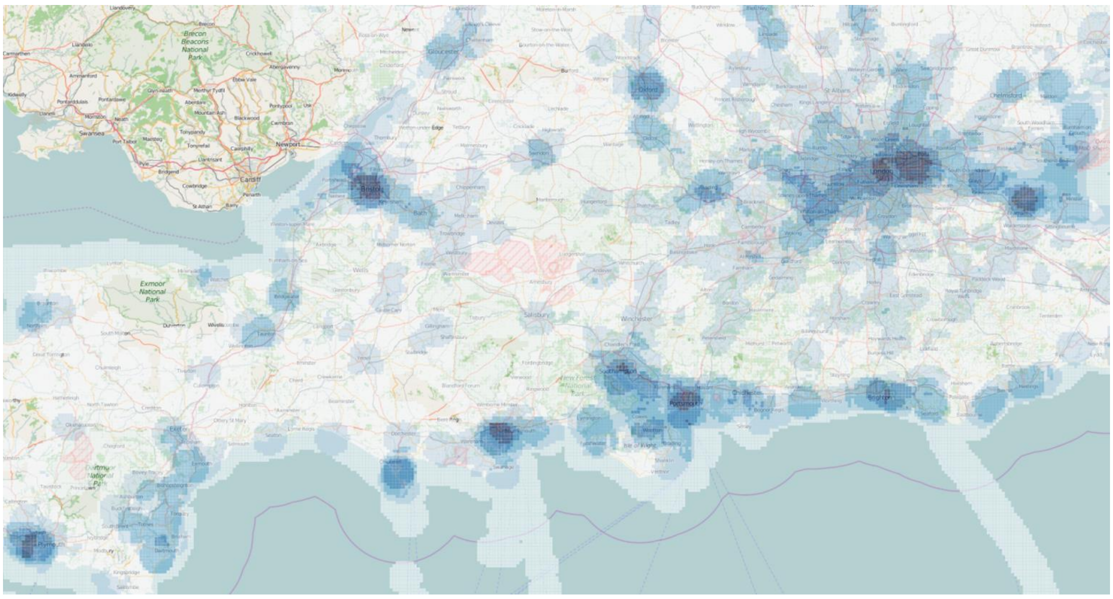

Figure 4: summation of distances of sub entities defining Ports (south English coast).

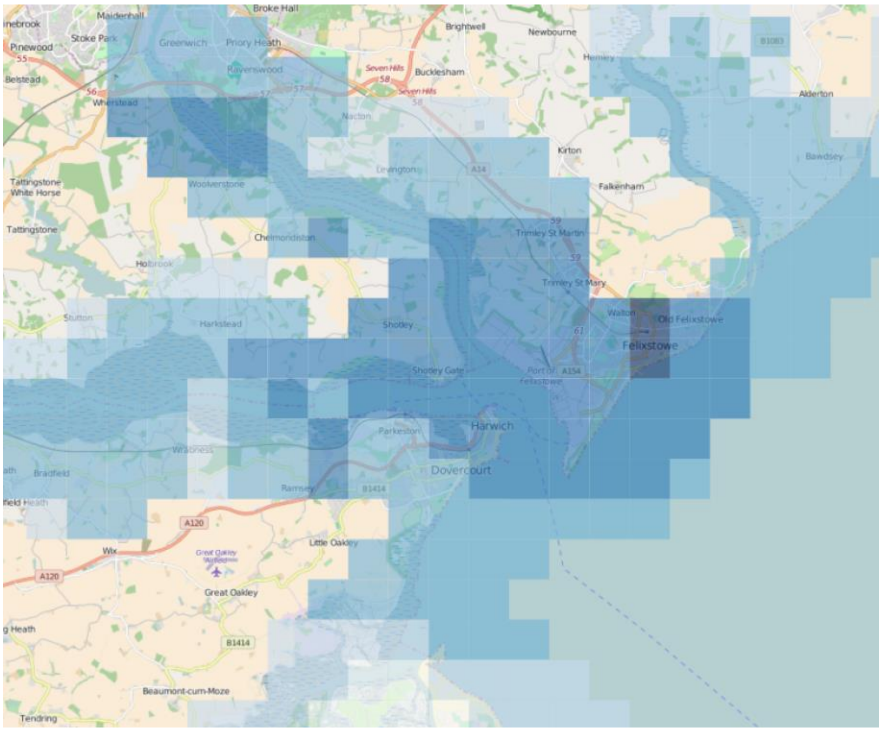

Figure 5: The Port of Felixstowe – a major fishing port on the west coast

Pattern Recognition Techniques

Pattern recognition techniques aim to classify data (patterns) based on either a priori knowledge or on statistical information extracted from the patterns. These patterns can be said to exist in a multidimensional space (the attribute space and the geometric space). Few ports will exhibit all the qualities listed in Table 1. Our first approach was to search for spatial clusterings of these entities. This was done by dividing England into a fine fishnet of 1km grid squares, and for each grid square calculating the distance to each of the entities listed in Table 1. Figure 4 (and Figure 5) shows the resulting pattern (darker blue shades reveal greater concentrations of the sub entities and a greater degree of confidence in classification).

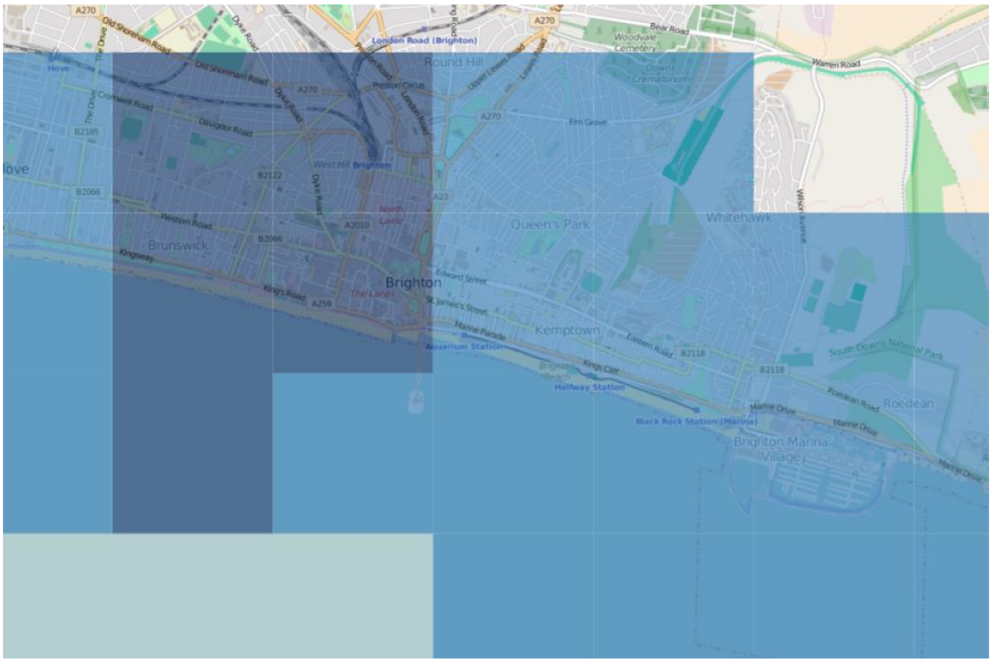

Figure 6: Errors of omission and commission: Brighton is not a Port

Machine Learning
But this approach did not always produce correct results (Figure 6). Such a simple model did not differentiate the importance among the entities (is a dock or a pier more important than a boatyard or a marina for example?). The question arose: can we use machine learning techniques and a small training dataset of known ‘Ports’ and ‘non ports’ to identify the optimal balance of entities constituting a typical port?

Three approaches were experimented with (Decision trees, Bayes classifier, and Neural Nets), all implemented within rapidminer (rapidminer.com) (Hofmann & Klinkenberg 2014). The training set comprised two significant ports (Plymouth and Southampton), and a load of places that were most definitely not ports (Figure 7).

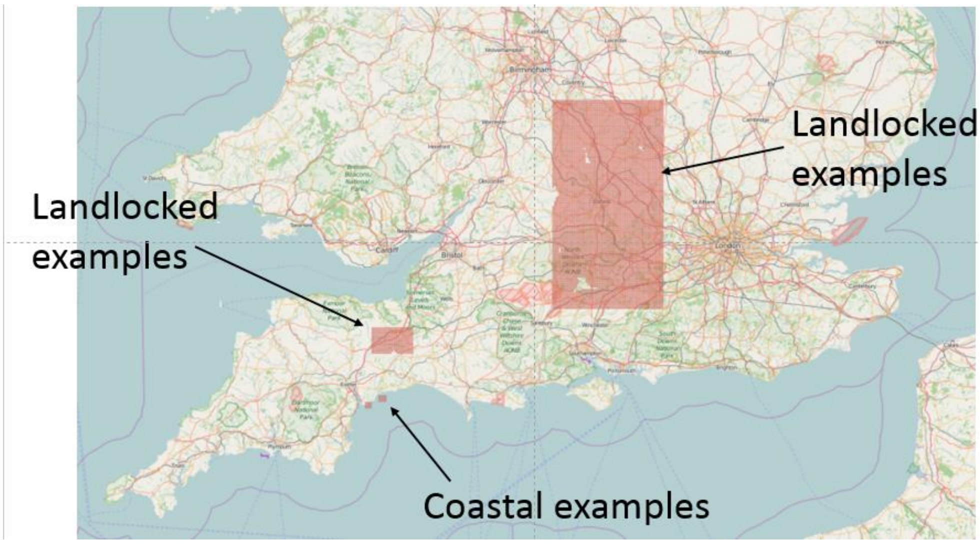

Figure 7: Training set used for the three types of machine learning approaches.

Results Using Decision Trees
Decision trees build classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed resulting in a tree with decision nodes and leaf nodes. A decision node has two or more branches; the topmost decision node in a tree which corresponds to the best predictor is called the root node.

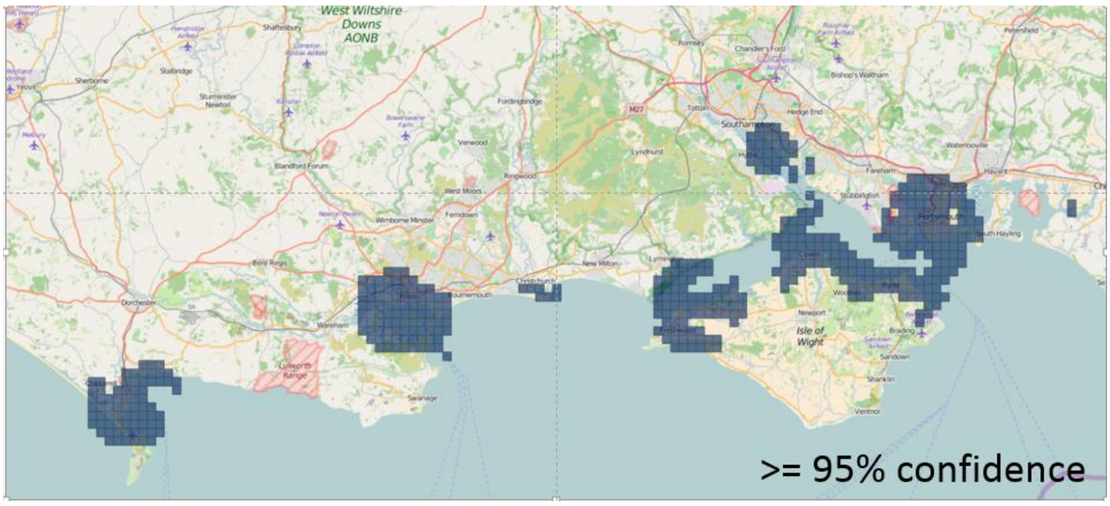

Figure 8: Port like structures (95% confidence level)

Decision trees can handle both categorical and numerical data. The results are presented (Figure 8) and the decision tree (Figure 9) reveals the significance of Ferry Route and Ferry amenity in the classification process. The decision tree is very much a reflection then, of the data captured in OSM. A different database may not have such records, in which case the decision tree would comprise other sub entities. Further experimentation and comparison is anticipated.

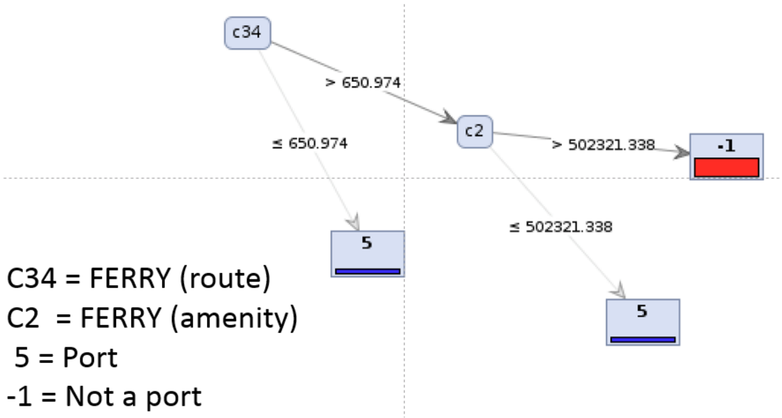

Figure 9: The decision tree associated with Figure 8, showing how ports and non ports are differentiated.

We have also experimented using a Baye's classifier and artificial neural networks. Space precludes inclusion of the results suffice to say that further training was used to improve the quality of the results. Time is required to refine both the tuning of the algorithm, the entities included, the choice of training ports and non-ports. Work is also required in the finer classification of ports (large and small for example), and the derivation of single points based on the region defined by contiguous cells. Some interesting results nevertheless.

Conclusion

Here we have presented the basis of a technique that will enable us to automatically identify higher order objects based on their sub entities. In addition to a simple summation of distances we have explored the use of three different machine learning methodologies – each resulting in some promising output. . Rather than explore the metric and topological properties (necessary for small changes in scale), it is argued that an ontological perspective (where a functional perspective has precedent over the geometric) should provide a basis for passing through Muller's conceptual cusps.

A map is a geographical characterization and map generalization needs to be articulated through a set of scale dependent geographical concepts (whether it be a lamppost, a house, a city, or a continent). The explicit modeling of the connectedness and compositional form of these concepts is surely the way forward. GIS has encouraged us to see the geometry as having primacy but this has stymied developments in map generalization over large changes in scale. From this perspective we argue that the geometry plays second fiddle to these ontological mappings when it comes to map generalization.

References

Alderton P (2008) Port Management and Operations. *Informa.*

Chaudhry, O. Z, Mackaness, W.A. & Regnauld, N., (2009) A Functional perspective on Map Generalisation, Computer Environment and Urban Systems: Special Issue on GeoInformation Generalisation and Multiple Representation, 33 (5), pp. 349-362.

Coleman, D. (2013). Potential Contributions and Challenges of VGI for Conventional Topographic Base-Mapping Programs. In Sui, D., Elwood, S and Goodchild, M. (Eds.), Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice (pp. 245–264). New York, London: Springer Science+Business Media Dordrecht

Fairbairn, D and Al-Bakri, M (2013) Using Geometric Properties to Evaluate Possible Integration of Authoritative and Volunteered Geospatial Data. ISPRS International Journal of GeoInformation 2(2): 349-370.

Gaffuri, J. (2011). Improving web mapping with generalization. Cartographica: The International Journal for Geographic Information and Geovisualization 46 (2), 83-91.

Hofmann,M & Klinkenberg, R (2013) "RapidMiner: Data Mining Use Cases and Business Analytics Applications (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series)," *CRC Press*

Noy, NF & McGuinness, DL (2016) Ontology Development 101: A Guide to Creating Your First Ontology

Ramm, F; Topf, J; Chilton, S (2010). OpenStreetMap: Using and Enhancing the Free Map of the World. UIT Cambridge.

Touya G (2012) What Is the Level of Detail of OpenStreetMap? Workshop on the role of VGI in Advancing Science: Quality & Credibility in Columbus (ohio) USA.

Touya, G. & Reimer A. (2015) Inferring the Scale of OpenStreetMap Features. OpenStreetMap in GIScience: Experiences, Research, and Applications edited by Jamal Jokar Arsanjani, Alexander Zipf, Peter Mooney, Marco Helbich.