Geographical Information Science Research (GISRUK) Conference,
pp 195-198, Sheffield, UK, April 2002,

CONTEXT-BASED CLASSIFICATION OF OBJECTS IN CARTOGRAPHIC DATA

Leo Mulhare, Diarmuid O'Donoghue, Adam C. Winstanley
Department of Computer Science,
National University of Ireland, Maynooth,
Maynooth, Co. Kildare, Ireland.
Tel: +353-1-708 3847
Fax: +353-1-708 3848
E-mail: {lmulhare, dod, adamw}@cs.may.ie

INTRODUCTION

The Ordnance Survey has traditionally recorded the large-scale topography of Britain as Cartesian co-ordinate-based point, line and text label features within the tile-based Land-Line® Database. Under their Digital National Framework™ (DNF™) project, this data has been re-engineered into a topologically structured format known as OS MasterMap™ [Ordnance Survey]. This required the modelling of the area features enclosed by the line data as polygon objects. This new polygon-enriched data can be provided seamlessly for pre-defined areas and by theme. Each feature is assigned a unique Topographic Identifier (TOID™) number, allowing for the easy updating of a data holding, and the association of any topographic feature with external information. Each point object is classified with a particular feature code, such as post-box or bench-mark; likewise, a line feature could be labelled as a building outline or a public road edge. The feature-coding of polygons is the most difficult requirement of the DNF, as it requires the inferring of information that is not present in the Land-Line data. Properly classified area features greatly add to the intelligence of the resulting OS MasterMap data, allowing a myriad of valuable analyses to be carried out. The OS has accomplished high quality polygon classification semi-automatically, largely by examining the feature codes of the lines that bound each polygon. Using novel feature-coding techniques, the accuracy can be further improved.

Work is continuing within our research group on the application of computer vision techniques to polygon classification through shape recognition [Keyes and Winstanley]. The results of several shape descriptors have been fused and good results have been achieved. By combining the results with other classification techniques, a more robust feature-coding tool can be developed. In this paper, the classification of polygons based on their context is described.

POLYGONAL CONTEXT

For the purposes of our research, the context of a particular polygon may be defined as a description of the classifications of the adjacent polygons and the topology between these polygons. When considering adjacency between polygons, there are two separate topological relationships to be considered. We define these as:

1. ***Line Adjacency***
   Two polygons are line-adjacent if they share a bordering line.
2. ***Point Adjacency***
   Two polygons are point-adjacent if they are not line-adjacent but they meet at 1 or more points.

Examples of the two types of adjacency can be seen in figure 1.

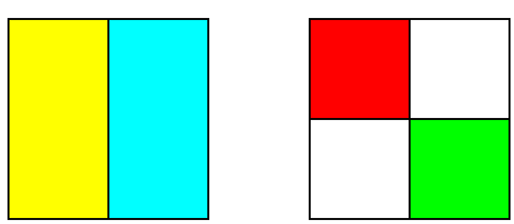

Figure 1. *The polygons on the left are line-adjacent, while the shaded polygons on the right are point-adjacent.*

Point adjacency is very common between land parcels and buildings. Its treatment as a special case of adjacency allows for a more precise description of such polygons. In addition, to avoid the problem of having one unmanageable polygon representing the road network for the whole of Britain, the OS have added polygon closing links at most road junctions. These create artificial point-adjacency topology at many junctions, making the road network amenable to context-based classification as well. Two approaches to polygon classification through context are being taken, the first of which is content vector matching.

CONTENT VECTOR MATCHING

A content vector is a list of attributes in some pre-defined order. The content vectors that are being used to model the context of a particular polygon record the number of line-adjacent neighbours of each feature code and the number of point-adjacent neighbours of each feature code for any given polygon. The context of a polygon is thus represented by a list of integer values. If these content vectors match for any pair of objects, the two polygons are considered analogous.

The use of analogy has been shown to be central to any systems that can learn and solve novel problems [Evans; Bohan & O'Donoghue]. An analogy allows us to map some known information from a source domain to a matching target domain [O'Donoghue & Winstanley]. For the purposes of polygon classification, feature codes are mapped between the domains. Using a large and representative corpus of very well classified polygon data, all the content vectors contained therein are recorded. Associated with each of these "template" content vectors is a record of the number of polygons of each feature code within the data set that matched that vector. This "training set" of templates can then be used as the source domain in attempting to classify a particular polygon, whose content vector is the target domain. If the content vector of an unclassified polygon P is found to match that of a particular template vector T, then the probability that P is of class C is:

$$\textbf{Probability P is of class C} = \frac{\textbf{\# polygons of class C matching T}}{\textbf{\# polygons matching T}}$$

If the feature code with the highest probability has a margin of error that the user is prepared to accept, the target polygon is assigned that code. The number of polygons in the training set that matched a particular content vector is also considered; the higher the number is the more accurate are the probability figures.

Once an input data set has been processed and the contexts of its polygons recorded, content vector matching is a very fast classification technique, as it merely requires comparisons between numbers. While this is a good classifier for polygons with many neighbours, a more fine-grained approach yields more accurate results. Such an approach is now described.

ANALOGICAL STRUCTURE MATCHING

Analogical structure matching [Gentner] is an extension of content vector matching. Content vectors are used to record the number of polygons of each feature code that are point-adjacent or line-adjacent to the central object. In addition, the context descriptions that are used in analogical structure matching record the adjacencies between these neighbouring polygons [Mulhare, O'Donoghue & Winstanley]. This more precise model of context allows classifications that are at least as accurate as content vector matching, and usually more so. The price that is paid for this accuracy is greater computational expense. As with content vector matching, a training set of data is used build a set of template contexts. If the structural context of an unclassified polygon matches one of these templates, probability can again be used to feature code the polygon in question.

Whereas the first technique's template matching required comparisons between a list of numbers of fixed length, structural matching is accomplished by identifying a mapping between each neighbouring object in the template and each of the unclassified polygon's neighbours. In searching for a match between polygons with *n* neighbours, a search space of up to *n!* possible matches has to be traversed to determine if an isomorphism exists between the two contexts. This exponential search space limits structural matching's usefulness to polygons with smaller numbers of neighbours. Each polygon's TOID is used to reference it uniquely during the matching process, and mappings from an object adjacent to a template to one adjacent to the target polygon are constrained by only allowing mappings between:

1. pairs of point-adjacent or pairs of line-adjacent objects
2. polygons with the same feature code

An example of a template context and an unclassified polygon context can be seen in figure 2. A full structural match can be established between the context of the template polygon (1) and the unclassified polygon (6), using the mappings (2→10, 3→8, 4→9, 5→7). This allows the feature code suggested by the template, “building”, to be mapped to the target polygon. In this case there is only one structural mapping between the two contexts, but in general there may be many; any of which are sufficient to establish an isomorphism.

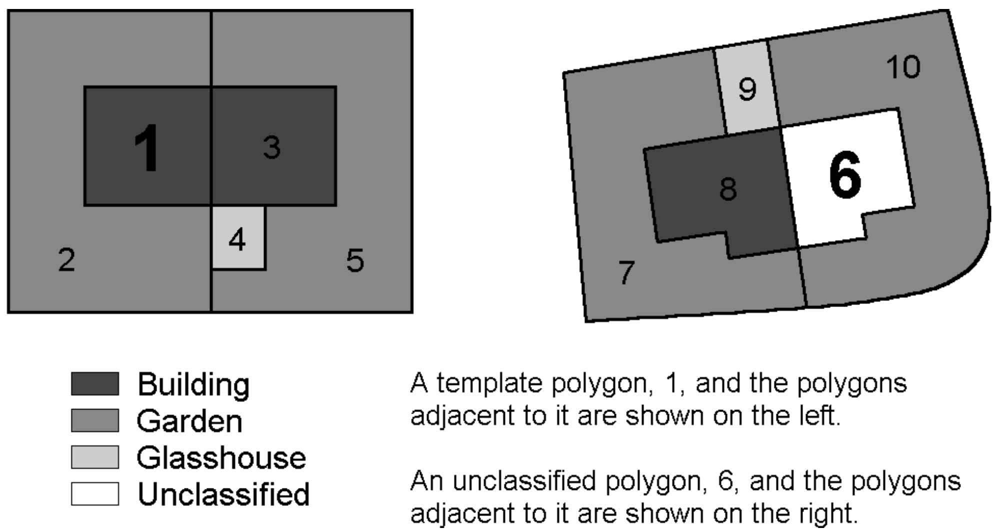

Figure 2. *A structurally-matching template context and unclassified polygon context*

CONCLUSION

Content vector matching offers a fast means of classifying polygons in cartographic data. Analogical structure matching allows for greater precision, but its computational expense limits its usefulness to area features with a smaller number of neighbours. In practice, a threshold for the number of adjacent neighbours is set, above which only content vector matching is used. Below this limit, structure matching is used in favour of content vector matching if the corresponding structural template suggests a classification with higher probability. As both techniques require all adjacent polygons to have been previously classified, their usefulness is limited to data sets where most objects have feature codes to begin with. In addition to feature coding currently unclassified polygons, they can identify misclassified features. Context-based classifiers complement shape-based recognition; fused together, they can produce a level of accuracy in object classification that no one technique could achieve on its own.

ACKNOWLEDGEMENTS

We would like to thank Ordnance Survey (Great Britain), for their help and the use of their DNF data sets.

REFERENCES

Bohan, A., & O’Donoghue, D. (2000). A Model for Geometric Analogies using Attribute Matching. *AICS 2000 – 11th Artificial Intelligence and Cognitive Science Conference*, NUI Galway, Ireland.

Evans, T.G. (1968). A Program for the Solution of a Class of Geometric-Analogy Intelligence Test Questions. In M. Minsky (Ed.), *Semantic Information Processing*. MIT Press.

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, **7**, 155-170.

Keyes, L., & Winstanley, A.C. (2001). Using Moment Invariants for classifying shapes on large-scale maps. *Computers, Environment and Urban Systems*, **25**(1), 119-130.

Mulhare, L., O'Donoghue, D., & Winstanley, A.C. (2001). Analogical Structure Matching on Cartographic Data. *AICS 2001 – 12th Artificial Intelligence and Cognitive Science Conference,* NUI Maynooth, Ireland.

O'Donoghue, D., & Winstanley, A.C. (2001). Finding Analogous Structures in Cartographic Data. *4th AGILE Conference on Geographic Information Science,* Czech Republic.

Ordnance Survey (Great Britain). The Digital National Framework. http://www.ordsvy.gov.uk/dnf

BIOGRAPHY

Leo Mulhare received his B.Sc. in Computer Science from NUI Maynooth in 2000. His final year dissertation involved the extraction of polygons from the vector line data of an object-oriented cartographic database. He is currently a second year M.Sc. research student within the Department of Computer Science at NUI Maynooth, investigating the classification of topographic objects through analogical reasoning.