Applications of Cartographic Structure Matching

Diarmuid O'Donoghue, Adam Winstanley, Leo Mulhare, and Laura Keyes
Department of Computer Science
National University of Ireland, Maynooth,
Co. Kildare, Ireland.
{adam.winstanley; diarmuid.odonoghue}@may.ie

Abstract — **Automatic categorization of large-scale topographic vector data into roads, buildings and similar classes typically examines each object description in isolation. We describe a Cartographic Structure Matching (CSM) algorithm that automatically classifies objects in topographic maps by examining the context of the object. Matching clusters of objects against known templates serves to categorize ambiguous polygons by including context in the categorization process. We describe a number of applications that emerged from the core structure-matching algorithm, addressing problems of error detection, rejoining partitioned objects, composite object identification and data quality estimation.**

Keywords- topographic data; classification; analogical inference algorithms; automated processing.

I. INTRODUCTION

We describe the structure-matching algorithm known as Cartographic Structure Matching (CSM). We illustrate how matching the context of a problem object against a known template solution serves to classify ambiguously shaped polygons [1]. We then examine four novel applications that are based on this core structure-matching algorithm. They address the problems of: error identification, composite object identification, obscured object identification, and assist with quality control activities. The only significant difference between these applications relates to the templates that are available to the matching algorithm. We also describe how the algorithm identifies the required template.

The database used in this project consisted of 46,000 polygons, encompassing a large urban area with some rural potions. Polygons had previously been classified into 21 mutually exclusive categories, including; *road, building, made-land, unmade-land* in addition to *unclassified*. The initial objective of this project was to identify a classification for "unclassified" polygons based on the classes of their neighboring polygons. Frequently, these represent polygons whose categorization is difficult to determine, using other techniques.

II. STRUCTURE MATCHING

Before examining CSM, we outline how structure-matching algorithms play a central role in modeling people's understanding of analogical comparison [2]. An analogy is a structured comparison between two concepts, one familiar and one incomplete. The familiar concept acts as a template, supplying missing knowledge to the problem domain. Analogy is a very well developed field of study, having produced many computational models of how analogies are formed [3,4]. Computational models of analogy rely on structure alignment between both domains. Predicate calculus assertions capture the structure of each domain. CSM evolved from these cognitive models.

A. *Identifying a Locality Collection*

A *locality* is a collection of topographic objects consisting of one central polygon, plus all polygons adjacent to it. We represent the essential structure of a locality using predicate calculus assertions that represent adjacency information occurring within each locality. A *line-adjacent* relationship is generated where two polygons share a line boundary. A *point-adjacent* relationship is asserted between polygons that are not line-adjacent but meet at one or more vertices. All adjacency relations referencing the problem polygon, or one of its immediate neighbours, presents all the required structural information related to a locality. When performing classification, it is the central object that is the polygon whose classification we are testing or determining. The locality description derived is crucial for the structure-matching operation that allows the class of the central polygon to be inferred.

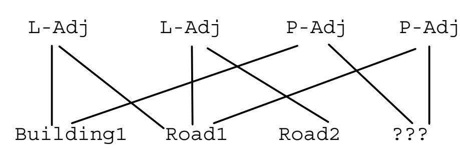

Figure 1. A Problem Locality

The domain in figure 1 represents a (simplified) problem structure involving four objects, where all except the central polygon are classified. We "align" this description with the template of figure 2, and can infer the class of the problem polygon based on the identified alignment.

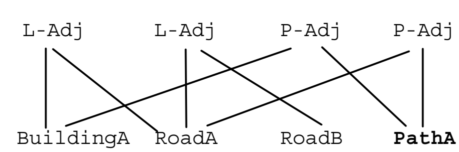

Figure 2. A Template Solution

Sponsored by the Enterprise Ireland, British Council, and Ordnance Survey (Southampton), UK

Computationally, this relies on a structural alignment algorithm, solving this variant of the Largest Common Sub-graph problem [5]. A number of constraints serve to simplify this potentially difficult problem.

- Identify a one-to-one mapping between objects in the two domains.
- Mapped relations must be of the same type.
- Mapped objects must have identical categorizations [6].
- The category of the central polygon is taken directly from the template.

Without breaking the one-to-one restriction, the only possible maximal inter-domain mapping, generates the following analogous object-pairs:

(Building1, BuildingA), (Road1, RoadA),

(Road2, RoadB), (PathA, ???)

From this it can be inferred that the unclassified object is a path.

B. Content Vector Indexing

Of course, when presented with a problem locality, we must first identify an appropriate template. Templates are indexed using a list of types of objects within a template, as well as the quantities of each. In some circumstances the context vector uniquely identifies a single template. However, if a small set of possible templates is identified, each undergoes detailed structure-matching with the presented problem. This often serves to identify a single matching template.

C. A Simple Example

The central polygon in figure 3 represents a road intersection. However, it is also a similar size and shape to the surrounding buildings. Previous tools [7, 8] have used object shape as a classifier. Such a tool might, when examining this polygon in isolation, misclassify it as a building. However, including its context in the categorization process using structure matching resolves this ambiguity. Thus, CSM avoids the misclassification often made by object-based classifiers.

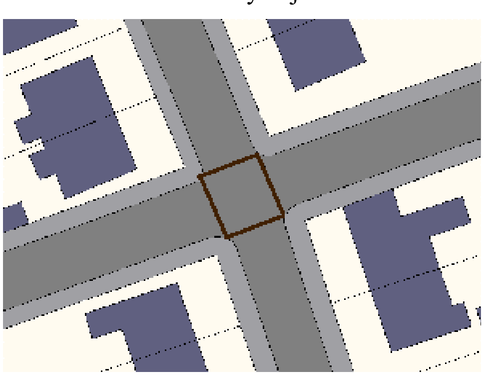

Figure 3. Example of Road/Building shape ambiguity

III. APPLICATIONS OF CSM

We now describe a number of developed and emerging applications based on the central CSM algorithm. These applications require little (if any) modification to the basic algorithm, other than supplying it with a different set of templates. It is these templates that support the differing inferences that CSM generates.

A. Detecting Classification Errors

Error detection is central to the quality assurance needs of national mapping organisations. CSM identifies specific classification errors by explicitly defining and identifying an illegal locality - like an isolated section of road. Detecting specific errors is perhaps of greatest use when there is a known problem with the data gathering or data entry processes that must be quickly rectified. Figure 4 shows an example of the illegal context “building within a building” caused by the miss-classification of a land parcel as a building. Other examples include isolated segments of rail and rivers. Currently, this module only identifies the errors, passing the relevant polygons back to the classification process.

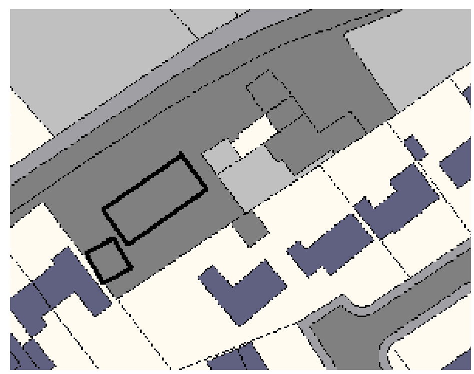

Figure 4. Example of a “Building within a Building” error.

B. Rejoining Segmented Objects

Topological data is a two-dimensional (2D) representation of three-dimensional (3D) information. Occlusions frequently result in segmented objects, such as a bridge crossing a river thereby creating two distinct river polygons. CSM can identify such contexts, a necessary precursor of introducing an “occluded object” segment to rejoin these polygons.

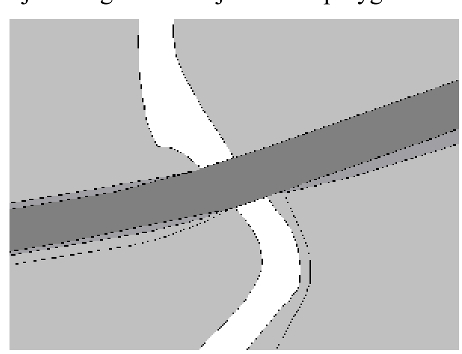

Figure 5. A River partitioned by a Road

Figure 5 depicts a road apparently segmenting the river passing beneath it. CSM identifies this structure, not by examining any one polygon in isolation, but based on the overall structure of the assembly. Roads, railways, rivers and paths are often subject to this segmentation problem.

C. *Identifying Composite Objects*

Topographical data is often stored as individual mutually exclusive objects (or polygons) as, for example, in Ordnance Survey Mastermap data sets. This ignores the inherently hierarchical nature of what is being modeled. The objects applications are concerned with are often complex entities made up of several individual objects. For example, a university is made of typically several large buildings surrounded by ancillary structures and facilities.

CSM can identify thematically related collections of objects, and thereby supports the addition of hierarchical structure to the map. Such collections are generally adjacent, and thus CSM is ideally suited to identifying them. A typical example is the identification of a semi-detached house, as in Figure 6, consisting of a dwelling and the adjacent land parcel. In the test data set this template was matched 1224 times, all of which were semi-detached houses.

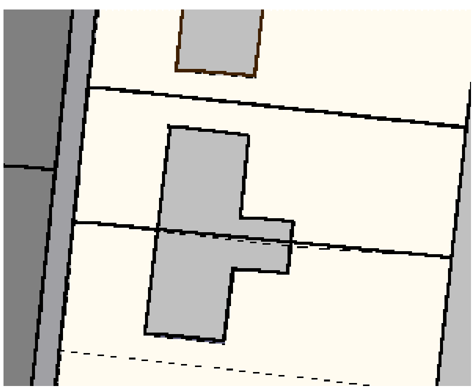

Figure 6. A semi-detached house composite object

A more complex example involves propagating the identity of one object to its neighbours. Figure 7 shows a building that is identified as a school in the annotation theme. This information may also be propagated to the remainder of this locality. However, this application is still in development.

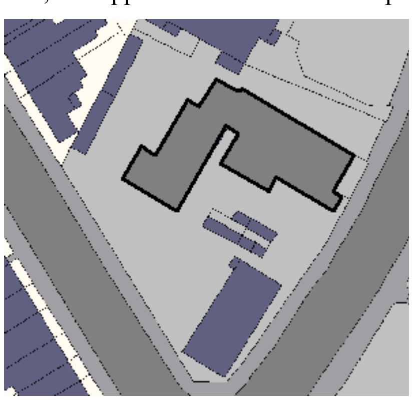

Figure 7. A composite object

D. *Quality Estimation by Template Usage Frequency*

Quality assurance can also make use of the frequency with which various templates occur within a map segment. When updating a map (segment), any unexpected changes in the frequency distribution may highlight systematic classification errors. (This application is still in development).

IV. CONCLUSION

Previous techniques for topographic data classification examine an object's description in isolation, focusing on shape features like area and boundary length. This approach however, is unreliable when categorizing objects whose shape does not uniquely identify it. We described the Cartographic Structure-Matching (CSM) algorithm that classifies topographic objects, by examining the object within the context of its immediately neighboring objects. CSM looks at the categories of objects adjacent to some unclassified object, and uses this as a basis for classification. We then described a number of applications built on the central structure-matching engine. These applications include error identification, composite object identification, obscured object identification, and quality control.

ACKNOWLEDGMENT

We would like to thank Glen Hart and The Ordnance Survey, Southampton, UK for supplying the data used in this project. This work was partly supported by an Enterprise Ireland/British Council Research Visits Scheme Grant (BC/2002/015).

REFERENCES

[1] Winstanley A., O'Donoghue D., and Keyes L., "Topographical object recognition through structural mapping", 1st International Conference on Geographic Information Science - GIScience 2000, Savannah, Georgia, USA, October 28-31, 2000.

[2] Gentner, D. "Structure-mapping: a theoretical framework for analogy", Cognitive Science, vol. 7, pp. 155-170, 1983.

[3] Keane, M. T. Brayshaw, M. "Indirect analogical mapping: a computational model of analogy", in Third European Working Session on Machine Learning. Ed. D. Sleeman, London Pitman, 1988.

[4] Falkenhainer, B. Forbus, Gentner, D. "The structure mapping engine: algorithm and examples", Artificial Intelligence, 41, 1-63, 1989.

[5] Veale T, O'Donoghue D, and Keane M. "Computability as a limiting cognitive constraint", in Cultural, Psychological and Typological Issues in Cognitive Linguistics, Ed. M. Hiraga, C. Sinha and S. Wilcox., John Benjamins Publ. Amsterdam/Philadelphia. pp 129-155. (1999).

[6] Bohan A. O'Donoghue "A Model for Geometric Analogies using Attribute Matching", AICS-2000 11th Artificial Intelligence and Cognitive Science Conference, Aug. 23-25, NUI Galway, Ireland, 2000.

[7] A.C. Winstanley and L Keyes, "Applying computer vision techniques to topographic objects"", International Archives of Photogrammetry and Remote Sensing, 33 (B3), 480-487, July 2000.

[8] L Keyes and A.C. Winstanley, "Using moment invariants for classifying shapes on large-scale maps", Computers, Environment and Urban Systems, Vol. 25, pp. 119-130, 2001.