IEEE IGARSS Conf., Vol. 6, pp 3730 – 3732, July 21-25, Toulouse, France, 2003, ISBN 0-7803-7929-2.

Applications of Cartographic Structure Matching

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

Introduction

In the first part of this paper we describe the Cartographic Structure-Matching (CSM) algorithm that identifies multi-object structures from topological data. We describe the significant differences between CSM, and traditional structure matching developed for the purposes of cognitive modelling. CSM was initially developed as a tool for categorising topological data based on the context of some unclassified object. (Traditional techniques examine only an objects content, like size and shape). CSM looks at the categorisation of objects adjacent to the unclassified object, and uses this as a basis for classification. CSM is particularly adept at classifying objects whose shape does not uniquely identify it. The remainder of this paper describes a number of more recent applications of CSM.

Classification Error Detection

Error detection is central to the quality assurance needs of national ordnance survey offices. Specific classification errors are identified by explicitly defining an illegal context, 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.

Quality Estimation/ frequency Distribution

Previous results indicate there is an exponential distribution in the frequency with which different contexts occur within a map. However, individual map segments may vary, perhaps using more urban related contexts. When updating a map (segment) we may compare the distribution of context before and after update - any significant discrepancy may indicate an error in the updating process.

Rejoining Segmented Objects

Topological data is a two-dimensional (2D) representation of three-dimensional (3D) information. Occlusions frequently segment objects, like bridges occluding the underlying river. CSM can identify such contexts, and introduce an occluded object segment.

Composite-Object Identification

Topological data is stored as individual land parcels. Introducing hierarchical structure based thematically related collections of objects. Such collections are generally adjacent, and thus CSM is ideally suited to identifying such structures. For example, a road plus adjacent

Conclusion

CSM is a powerful and flexible means of processing topological data, based on the analogical reasoning models developed by cognitive scientists.

IEEE IGARSS Conf., Vol. 6, pp 3730 – 3732, July 21-25, Toulouse, France, 2003, ISBN 0-7803-7929-2.

Applications of Cartographic Structure Matching

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

Introduction

AW - GIS Intro. Object shape vs. object context.

Analogy, Cognitive Modelling and Structure Matching

Structure matching algorithms were developed to model how people solve analogy problems (Gentner 1983; Falkenhainer *et al* 1989; Keane *et al* 1994; Veale *et al* 1999). An analogy is a comparison between two collections of information, called the *source* and *target* - each being a collection of objects and relations between them. Solving analogy problems involves aligning (or overlaying) the source domain over the target. This alignment highlights the crucial *inter-domain mapping*. Each object in the target has one (and only one) exact counterpart in the source domain. The key insight is that it is the *structure* or topology of the two domains that is important, rather than the properties of the aligned objects. So, a square in the target may be mapped to a circle in the source - as dictated by the overall structure of both domains. Algorithms developed to solve such problems focus on identifying the isomorphism between structures of the two domains. This Structure Mapping problem is known to be NP-Complete (Veale *et al,* 1999), meaning it is essentially unsolvable for large problems. However, algorithms exist that can solve analogy problems efficiently.

Figure 1a - A Problem domain with one unclassified object

Figure 1b - *Template* used to match against the problem domain

Figure 1 - A simplified Geometric Analogy

Geometric analogies are a particular subclass of analogy problem, often found in IQ problem sets. Evans' model (1967) solved a simple class of geometric analogy problems - those involving plain attribute-free objects. However, many geometric analogies rely on extensive use of attributes like colour, pattern, shape and orientation (Bohan and O'Donoghue, 2000). Most such problems can be solved using the Local Attribute Matching algorithm (for details on the alternative Global Attribute Matching algorithm, see Bohan and O'Donoghue, 2000). However, cartographic structure matching problems are very straightforward. They required a mapping (between the problem and the known domain) that involves pairing objects with identical categorisation. This makes the structure matching process very straightforward.

We tailored this general Geometric analogy algorithm to cartography, producing Cartographic Structure Matching (Winstanley, O'Donoghue and Keyes, 2000). This was initially developed to categorize partly categorized cartographic data. By forming an analogy between a problem domain and a known valid domain, we can infer the category of an unclassified polygon. Cartographic Structure Matching is characterised by a number features.

- Two identical (isomorphic) structures
- 1-to-1 mapping between objects in the two domains
- Shape attribute does Not affect the objects that participate in the mapping
- Mapped objects have identical categorisation
- Analogical inferences relate to feature assignment (categorisation).

IEEE IGARSS Conf., Vol. 6, pp 3730 – 3732, July 21-25, Toulouse, France, 2003, ISBN 0-7803-7929-2.

Structure matching focuses exclusively on the topology of the presented information. Thus, it is *Rotation, Translation, Scale* and *Shape* invariant. This makes it ideal for a wide variety of topographic applications that are based on the underlying topology of the data. Applications include classification, error detection, error correction, composite object formation, and

We define a *domain* as one central polygon plus all objects adjacent to it. The domain stores the categories of all objects, plus all adjacency relations among domain members. As can be seen from the following diagram, even a small domain contains a great deal of structural information. Each adjacency relation is highlighted in Figure 2, and plays the central role in the Cartographic Structure Matching algorithm. Thus, a map is covered by a large number of overlapping domain descriptions. Cartographic Structure Matching involves identifying the *inter-domain mapping* between a problem domain and a pre-stored domain description.

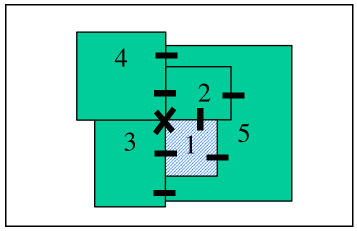


Figure 2 - A *domain*: its central object, adjacent objects, and adjacency relations.

A *content vector* summarises the contents of each domain, recording the frequency that each classification occurs in that domain description. These are used to identify the appropriate domain description, against which the problem will be compared.

AW - Content Vector indexing of prototypes.

In the following sections of this paper we describe some applications of Cartographic Structure Matching.

Structure-Based Classification.

The first use of Cartographic Structure Matching we discuss is classifying partly categorised topographic data. Object-based classification of topographic data can fail to give a definitive classification for some topographic objects. Context based categorisation can be used as either an alternative strategy for classification, or as a complementary strategy to object based classification.

Before categorisation can proceed, we first require a set of validated domain descriptions. These (*templates*) are easily collected form a correctly categorized map, wherein all domain descriptions are recorded. For example, consider a template describing a central building, a conjoined building and two adjacent parcels of land (see Figure 2). Any identical structure wherein the object at position X is unidentified, can be classified as a building.

Generate a domain description centred on each unclassified polygon. Next we generate the *content vector* for that domain, using that to identify the appropriate source domain.

We then use the cartographic structure-matching algorithm to identify inter-domain mapping. This aligns the template with the domain description. Thus, the unclassified polygon is given the classification of the aligned object in the template. (This is the analogical learning process, which is the hallmark of all analogical reasoning activities). For the majority of applications we have analysed, we have focused on templates that uniquely identify the category of the central polygon.

One interesting use of context based classification relies not on the contents of a template - as much as what is excluded from the template. On a correctly classified topographic database, we might wish to identify "*cul de sac*" roads (aka dead end roads). These can be identified with a template (in fact, this can be identified directly from the content vector alone), whose crucial facet is that the road in question has only one single other road in its domain description. Thus, categorisation does not necessarily rely on detailed template matching.

Some templates can have a number of alternative polygons at its centre. These intersecting templates are identified when two templates differ only in the category of the central polygon. By analysing the frequency with which template occur, we can calculate the classification probability associated with each category. Work on this topic is still ongoing, and is closely related to the occurrence frequency of different templates.

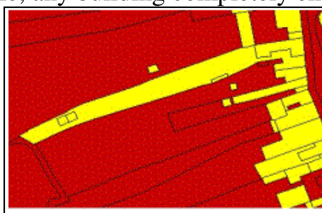
IEEE IGARSS Conf., Vol. 6, pp 3730 – 3732, July 21-25, Toulouse, France, 2003, ISBN 0-7803-7929-2.

Frequency of Occurrence of Domain Descriptions

We analysed the domain descriptions that were extracted from topographic data supplied by the UK Ordnance Survey. We plotted each domain against the frequency with which that domain occurred with the remained of the map. Although the map segment that was analysed contained only tens of thousands of polygons, its was found there was an exponential distribution in the occurrence of domain descriptions (Mulhare *et al*, 2001). Thus there were a relatively small number of domains that occurred very frequently across the whole map.

Detecting Classification Errors

Detecting errors involves a very similar process to that of category assignment. However, for this application the domain descriptions represent "invalid" domains. Typically, this process is only required when rectifying known errors that occurred during the (frequently manual) classification process. For example, any building completely enclosing another building might be identified as a classification error.



LEO
Nice picture of a road junction that is mis-classified as a building.

Figure - "Building within building" error

Figure - Detecting classification errors

Other instances of classification error that can be detected include:

- Road-junction mis-classified as a building
- Road that is unattached to another road.
- River that is unattached to another river
- Railway that is unattached to another railway

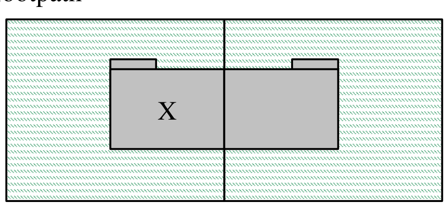
While these features may not always represent errors, they do represent topographic situations requiring further attention. For example, isolated polygons that are categorized as "river" may be isolated from the remainder of the river by an obscuring bridge. Later in this paper we describe how cartographic structure matching can help to re-join partitioned objects.

Composite Object Identification

To make maps truly useful, we must support activities that rise above the level of individual polygons. Many activities involve reasoning about collections of individual polygons, and treating these as units. We may want to unify a dwelling with its enclosing land parcel, thereby creating a "homestead" object.

Composite object identification can be achieved with Cartographic Structure Matching. We create a suitable domain description, against which candidate domains are matched. In practice, we may use the content vector as a basis for identifying potential objects, which may (or may not) form a suitable composite object. Examples of the contents of some useful composite objects include:

- Building plus surrounding land
- Road plus roadside footpath



Delete this one?

IEEE IGARSS Conf., Vol. 6, pp 3730 – 3732, July 21-25, Toulouse, France, 2003, ISBN 0-7803-7929-2.

AW/DOD - Quality estimation in maps (Exponential) frequency of prototype/temlpate occurrence.

One potential application involves managing the process of updating map tiles. Firstly, record the frequency of use of each template before updating the tile. Secondly, record template usage after the map segment has been updated. We can estimate the accuracy of the updated map, by comparing these statistics. Any significant change in template usage may highlight a potential problem with the classification process.

Rejoin Segmented Objects

Topographic data details the topography of a region from a vertical viewpoint. Because of the three dimensional nature of the real world, topographic data is necessarily incomplete. Road and rail bridges generally obscure other roads, rivers and railways. This givens us road, rivers, and railway segments that are (apparently) disconnected from the remainder of that real world object. Again cartographic structure matching offers a strategy for uniting these segmented objects.

Solving this problem involves an iterative application of template matching. Firstly, we use the content vector to find an isolated river segment (say). We then identify neighbouring polygons that also border a river. This identifies the structure that is characteristic of a segmented object. Further development will focus on generating an additional polygon to rejoin the partitioned object. This new polygon will be given the identity of the segmented object, although it will be explicitly tagged as being obscured. This application is still in development. Applications include:

- River segmented by a rail/road bridge.
- Road segmented by a rail/road bridge.
- Railway segmented by a rail/road bridge.

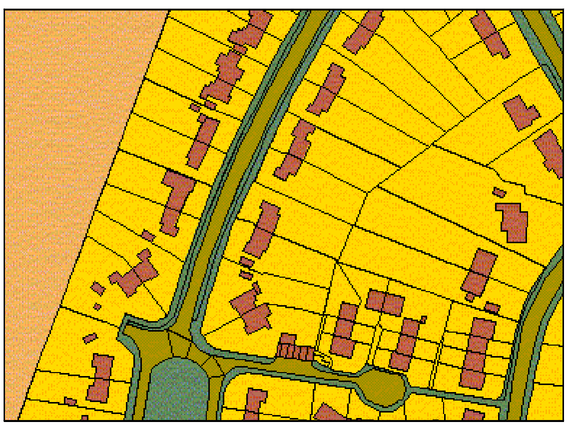


Figure 3a - Repeated iustances of the composite "dwelling" object

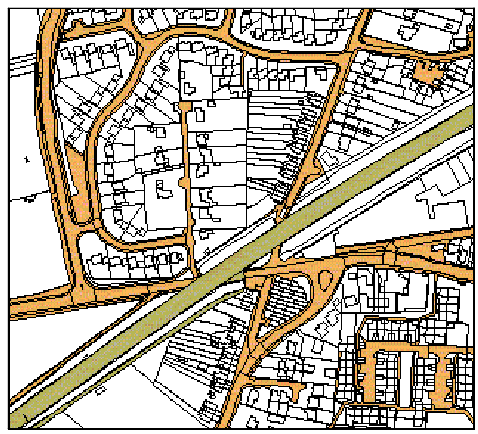


Figure 3b - Two roads are split by the obscuring Railway polygon

Conclusion

Cartographic Structure Matching is a powerful algorithm for categorising topographic data. In this paper we examine a number of related activities that can easily be achieved by this algorithm. We highlight that the strength of Cartographic Structure Matching is its ability to integrate *content* and *context*, within a cognitive framework. Cartographic Structure Matching is a very powerful algorithm, characterised by its wide range of application.

References

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.

Evans, T. G. *"A program for the solution of a class of geometric-analogy intelligence test questions",* In "Semantic Information Processing", (Ed.) M. Minsky, MIT Press, 1968.

Falkenhainer, B. Forbus, Gentner, D. *"The Structure Mapping Engine: Algorithm and Examples",* Artificial Intelligence, 41, 1-63, 1989.

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

Holyoak, K.J. Thagard, P. "*Analogical Mapping by Constraint Satisfaction*", Cognitive Science, 13, 295-355, 1989.

Keane, M. T. Ledgeway, T., Duff, S. *"Constraints on Analogical Mapping: A comparison of Three Models",* Cognitive Science, 18, 387-438, 1994.

Mulhare L, O'Donoghue D., Winstanley A., *"Analogical Structure Matching on Cartographic Data",* AICS-2001 - 12th Artificial Intelligence and Cognitive Science, NUI Maynooth, Ireland, pp 43-53, Sept, 2001, ISBN 0-901519-48-0.

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).

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.