

A FORMAL APPROACH TO IMPERFECTION IN GEOGRAPHIC INFORMATION

Full paper

Matt Duckham*, Keith Mason, John Stell, Mike Worboys

*Corresponding author
Department of Computer Science
University of Keele
Staffordshire, UK
ST5 5BG

Tel: +44 1782 584270; Fax: +44 1782 713082; Email: matt@cs.keele.ac.uk

ABSTRACT

Traditional computational models of geographic phenomena offer no room for imperfection. Underlying this tradition is the simplifying assumption that reality is certain, crisp, unambiguous, independent of context, and capable of quantitative representation. This paper reports on initial work which explicitly recognises that most geographic information is intrinsically imperfect. Based on an ontology of imperfection the paper explores a formal model of imperfect geographic information using multi-valued logic. The development of Java software able to assist with a geodemographic retail site assessment application is used to illustrate the utility of a formal approach.

1. INTRODUCTION

No observation of geographic phenomena will ever be perfect. Imperfection in some form or another is an endemic feature of geographic information (Goodchild, 1995). Traditional computational models of geographic phenomena offer no room for such imperfection. Underlying this tradition is the simplifying assumption that geographic reality is certain, crisp, unambiguous, independent of context and capable of quantitative representation. Over recent years, research challenging such simplifying assumptions has led to a much greater understanding of the complex multi-dimensional nature of imperfection in geographic information. Work so far has addressed a variety of different aspects of imperfection in geographic information, including vagueness (eg Erwig and Schneider, 1997), error (eg Heuvelink, 1998) and imprecision (eg Worboys, 1998a). However, this paper reports on initial work that attempts to recognise the multi-dimensional nature of imperfection more generally. Section 2 introduces an ontology of imperfection that can claim to offer a clear conceptual foundation for discourse about imperfect geographic information. Section 3 indicates how different types of imperfection in the ontology can assume different formal interpretations. Section 4 looks at the application of these different formal interpretations of imperfection to a geodemographics example. Section 5 outlines the development of prototype Java-based software, which aims to encapsulate the formal approach in a manner capable of supporting the spatial decision-making process for the geodemographics example. The paper concludes with a review of the research and an agenda for further work in section 6.

2. ONTOLOGY OF IMPERFECTION

While there may be consensus on the complex multi-dimensional nature of imperfection in geographic information, there is at best limited consensus in the literature on what these different dimensions actually are. Terms such as error, uncertainty, accuracy, precision, detail are subject to a range of different definitions and interpretations across the different disciplines that comprise GIS (Mowrer, 1999). The starting point for this paper, then, is to introduce a simple ontology of the key aspects of imperfection. It is not the intention to suggest that the ontology and definitions developed below are necessarily exclusive of other ontologies of imperfection. However, it is argued in section 3 that the ontology introduced here is congruent with formal interpretations of imperfection.

2.1 Error, imprecision and vagueness

Knowledge about reality is gained through observations. Observations are therefore first class objects in our account, rather than the underlying objects that are observed (Worboys, 1998a). Observations are imperfect in the sense that they can never fully or correctly reflect all aspects of reality. Imperfection is therefore the root of our ontology, as the concept refers generally to the inevitable deviations from perfection when observing reality. Imperfection can be thought of as comprising two distinct orthogonal concepts: *error* and *imprecision*. Error, or *inaccuracy*, concerns a lack of correlation of an observation with reality; imprecision concerns a lack of specificity in representation. Observations will usually be inaccurate and imprecise, but error and imprecision are orthogonal concepts since the level of accuracy of an observation is not implied by the level of precision, nor *vice versa*. Intuitively, the statement “York is in England” is at the same time more accurate and less precise than the statement “York is in Lancashire”. The general definitions of accuracy and precision above correspond closely to the more specialised statistical definitions of the terms in common usage (see Drummond, 1995).

Any observation of reality will be subject to imprecision: Veregin (1999) discusses some of the different causes and types of imprecision. *Granularity* is closely related, but not identical to imprecision. Granularity refers to the existence of clumps or grains in information, in the sense that individual elements in the grain cannot be distinguished or discerned apart. Granulation is therefore the result of distinct entities becoming indiscernible due to the imprecision in an observation. Observations or representations of coarser granularity offer less *detail*, for example where the clumping of information into pixels in remotely sensed images may prevent sub-pixel entities being distinguished (Fisher, 1997).

Vagueness, however, is a special type of imprecision which concerns the existence of indeterminate borderline cases. “Yorkshire is in England” is not a vague statement (both Yorkshire and England have clearly defined national or international boundaries), but is an imprecise statement. Although intuitively more precise, “Yorkshire is in the East of England” is a vague statement, since the concept of the East of England has borderline cases. Vagueness and imprecision are not equivalent, but every vague statement must also be imprecise, because of the lack of specificity at the boundary.

Difficulties surround deduction using vague assertions, exemplified by the *sorites paradox* (Cargile, 1969). The sorites paradox concerns the logical problem that arises when attempting to reason about the vague concept of a ‘heap’ (etymologically, ‘sorites’ derives from the Greek for ‘heap’, *soros*). Equally, the paradox can be illustrated using a geographically vague concept, such as “the South of England”. For example, we may be certain that Southampton, say, is in the South of England; similarly York is certainly not in the South of England. Crucially, since the concept of “the South of England” has no clear boundary, it seems reasonable to assert that a single step taken by a person walking along the road from Southampton to York cannot make the difference between that person being in the South of England and not being in the South of England. Given that no single step can make such a difference, it follows that a rambler starting at Southampton and walking to York could reasonably infer at each step that they were still in the South of England. Eventually, that rambler would arrive both at York and at a paradox: the original assertion that York is definitely not in the South of England is contradicted by our reasoning process during the walk from Southampton to York. The sorites paradox is no mere sophistry; it impacts directly on one of the foundations of classical logic, mathematical induction (Keefe and Smith, 1996). The illustrative example above is analogous to an inductive proof, with Southampton as a base case and each induction step actively paced out by our rambler. Fisher (2000) argues that many concepts in geography are vague. The sorites paradox shows that classical logic can be inadequate in dealing with such common concepts.

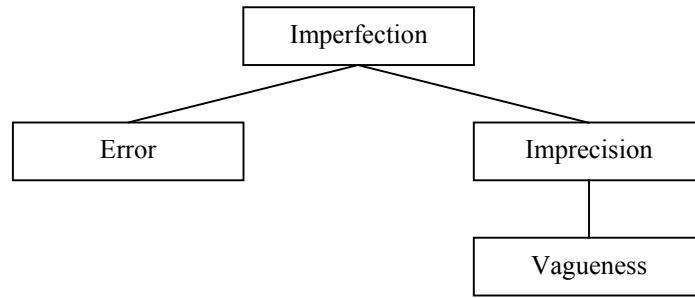


Figure 1: An ontology of imperfection

2.2 An ontology of imperfection

The four key concepts in our ontology are imperfection, error, imprecision and vagueness. Figure 1 organises these concepts into a hierarchy, with increasingly specialised concepts represented lower down the hierarchy. Error and imprecision are orthogonal sub-concepts of imperfection; vagueness is a specialised form of imprecision. Further refinements of this ontology are possible and, as suggested above, other ontologies are also conceivable. However, the following section explores the argument that an understanding of the interactions between different concepts in this ontology is congruent with formal interpretations of imperfection.

3. FORMAL MODEL

Formal models are a vital component in moving from conceptual to computational frameworks for imperfection. While the ontology described above offers a clear conceptual model of imperfection, translating that conceptual model into a practical computational setting demands a formal understanding of the ontology. The discussion above indicated some of the shortcomings of classical logic in dealing with vagueness in particular. In classical logic a subset of a set X is described by classifying each element of X as either *in* or *out* of the subset. Thus subsets of X correspond to functions $\alpha: X \rightarrow \Omega$, where Ω is the set of Boolean truth values $\{T, F\}$. Replacing Ω by the set of real numbers in the interval $[0,1]$, leads to the notion of a fuzzy subset of X . Several researchers have explored the use of fuzzy sets for example in addressing imprecision and inaccuracy in geographic information (Leung, 1987; Guesgen and Albrecht, 2000; Woodcock and Gopal, 2000). Fuzzy set theory has the advantage of very close ties with classical logic. However, practical and theoretical difficulties with fuzzy sets surround how to decide what fuzzy membership value to assign to a particular subset of X (Keefe and Smith, 1996). A related approach that does not depend on the assignment of membership values uses rough sets, introduced below.

3.1 Rough sets

In rough set theory Ω , above, is replaced by a set of three truth values $\{T, M, F\}$. Like fuzzy set theory, rough set theory is also beginning to enjoy some applications to spatial information (eg Worboys, 1998b; Ahlqvist et al., 2000). The approach used in the theory of rough sets is to define an equivalence relation. Formally, an equivalence relation can be represented by a surjective function $f: X \rightarrow Y$ from a set X to a set of equivalence classes Y . Since f is surjective, the set of equivalence classes Y will be a partition of X . By comparing different partitions Y_i of X induced by different functions $f_i: X \rightarrow Y_i$ the concept of a rough set emerges. For any subset Z of X , we define $\bar{S}(Z) = \bigcup \{Y_i \mid Y_i \cap Z \neq \emptyset\}$ and $\underline{S}(Z) = \bigcup \{Y_i \mid Y_i \subseteq Z\}$ to be the upper and lower approximations of Z in S respectively. In a rough set, the upper and lower approximations, $\bar{S}(Z)$ and $\underline{S}(Z)$, effectively define a boundary region where elements of Z are said to be *indistinguishable* in S and where Z is said to be *undefinable* (see Munakata, 1998, pp. 149-150).

It is worth noting that this approach proves inadequate when information at the coarser level of detail is derived from the combination of several higher detail sources. To deal with such situations it is necessary to remove the requirement that f be surjective. This leads to a generalisation of the theory of rough sets involving a four valued logic, rather than the three valued logic which appears in rough sets. The discussion below explores the different interpretations of rough sets for imperfect information.

3.2 Interpretations of rough sets

The mathematical structure of a function $\alpha: X \rightarrow \{T, M, F\}$ can be given three distinct interpretations. In a 'vague' interpretation, classifying $x \in X$ as M means that x is a borderline case. In an 'error-based' interpretation, classifying $x \in X$ as M means x is subject to inaccuracy and may be either as T or F, but only at some higher level of knowledge which is currently inaccessible. In an 'imprecise' or 'granular' interpretation, classifying $x \in X$ as M means that x has parts some of which can be classified as T and others as F, but only at some higher level of detail which is currently inaccessible. Consequently, M may have different interpretations and require different mathematical structures dependent on the type of imperfection, defined in the ontology in figure 1. Rough set theory can be used to side-step practical difficulties surrounding the assignment of membership values needed in fuzzy sets. However, theoretical difficulties do persist, as an acceptance of vagueness at the boundary between T and F can also be taken to imply vagueness at the boundary between T and M and between M and F, not dealt with in rough set theory. Such *higher-order vagueness* is not covered here, except to say that it should be considered when interpreting rough sets.

3.3 Imperfection in graphs

This paper deals primarily with the treatment of imperfection in sets. In spatial information, graphs are often a more important construct than sets, for example as in the representation of rail or road networks. The treatment of vagueness, error and granularity in graphs, as opposed to sets, may require different mathematical structures to represent them. Generalising from sets to graphs has led to a theory of graphs at different levels of detail (Stell, 1999). While not elaborated upon further here, it should be noted that the ideas introduced in this section could potentially be extended to apply to graphs as well as to sets.

4. GEODEMOGRAPHIC APPLICATION

The study of certain types of imperfection in geodemographic applications has a long history, for example through the use of fuzzy geodemographic classifications (eg Feng and Flowerdew, 1999). However, the approach to imperfection described above has the potential to unify a range of different research strands in geodemographic applications. Vague concepts are very common in the largely linguistic propositions used both in geodemographic classifications (eg consumer categories such as 'thriving' and 'striving') and in the original census count upon which they are based (eg socio-economic group classifications such as 'managers in small establishments'). The problem becomes especially acute when different geodemographic systems need to be reconciled and correspondence between the different system categories is limited, as in pan-European campaigns (Birkin, 1995). Issues of imprecision and granularity are also widespread in terms of the spatial granularity of boundary data, the temporal granularity of information collected, and attribute granularity, for example in the different levels of hierarchical classifications. The ecological fallacy and the modifiable areal unit problem (MAUP) are well known consequences of spatial granulation (Openshaw and Albanides, 1999). MAUP itself can be regarded an example of the wider difficulties associated with integrating data derived from different measurement frameworks (Chrisman, 1999). Error too is implicated in the choice, measurement, specification, and parameterisation of attributes used in the clustering algorithms to produce geodemographic classifications.

4.1 Retail site assessment

The geodemographic application used here is that of a retail site assessment support system, based in part on the work of Clarke et al. (1999). Clarke et al. describe a retail site assessment application that attempts to incorporate elements of cognition and intuition into the decision making process by using knowledge representation techniques in combination with (fuzzy) neural networks. The input to their system is in the form of *cognitive map* of the site assessment process, which aims to represent the partially subjective process of retail site assessment. Figure 2 gives a simplified hypothetical example of a cognitive map, where the likely turnover for a new store at a particular site is understood to be dependent on accessibility, presence of competition and customer category, which in turn is dependent on a variety of socio-economic indicators, such as social class, car ownership and age. A full cognitive map may contain a much greater number of more detailed concepts than shown in figure 2. Based on a variety of such maps from individuals across an organisation Clarke et al. describe the relatively complex process of combining, weighting, training and clustering this information necessary for their analysis to operate.

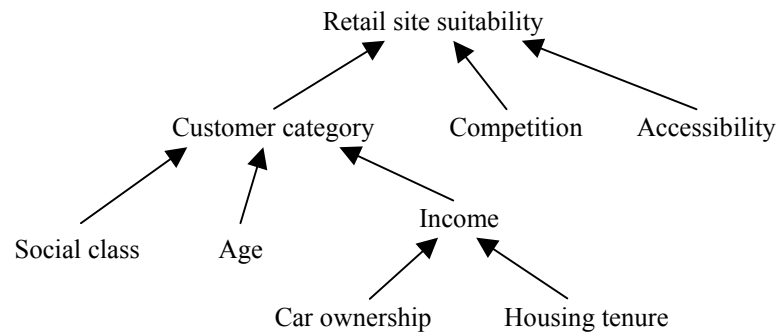


Figure 2: Example retail site assessment cognitive map (after Clarke et al. 1999)

An important observation is that much of the difficulty in handling cognitive maps stems from the presence of vagueness, imprecision and error in the concepts, processes and data used in the analysis. For example, the categories used in customer profiles are inherently vague; the census data upon which information about car ownership is inherently imprecise, as census data is aggregated to protect individuals' privacy; error is likely to be a factor in all of the mapped data. Consequently, this paper argues that retail site assessment is a useful testbed for the formal model discussed in section 3. The following section explores the practical application of the formal model to the development of software able to support a retail site assessment process related to that described by Clarke et al. (1999).

5. RETAIL SITE ASSESSMENT SOFTWARE

Based on the ontology of imperfection outlined in section 2, the formal model described in section 3 and the geodemographic application issues discussed above, prototype software able to support the process of retail site assessment was developed. This section outlines the key issues encountered during development in moving from the formal model to practical decision support software. At the moment the software is aimed at addressing vagueness and imprecision in the analysis. The incorporation of error handling into the software is amongst the work still in progress, touched on in section 6.

5.1 Vagueness in rough set retail site analysis

The retail site assessment software made use of a modified version of conventional rough set analysis. Conventional rough set analysis builds a rough set from a set of objects Z on the basis of equivalence classes induced from the attributes of the objects in Z . The upper and lower approximations of this rough set can then be used to infer rules which, given a number of assumptions, can be applied more generally. The technique is illustrated in figure 3, where upper and lower approximations for sites suitable for a luxury car showroom are inferred from four observations of different existing showrooms with two attributes, consumer income and retail site suitability. Analogy with existing stores has been common practice in retail site assessment for some time (Applebaum, 1966). In the figure 3, Z is the set of observations, R_{income} and $R_{site\ suitability}$ are partitions induced by the attributes 'Consumer income' and 'Site suitability' respectively. Clearly, a full rough set analysis would involve many more observations and attributes. The analysis is useful because general rules can be derived from the upper and lower approximations of the rough set. In the case of figure 3 below, it is possible to derive three naïve rules: if 'Consumer income' is 'High' then 'Site suitability' is 'Good'; if 'Consumer income' is 'Low' then 'Site suitability' is 'Poor'; if 'Consumer income' is 'Medium' then 'Site suitability' is undecidable. The derivation of such rules is related to inductive learning algorithms. Refer to Russell and Norvig (1995) and Munakata (1998) for general information on inductive learning algorithms, and Aspinall (1992) and Duckham and McCreadie (1999) for a discussion of the application of inductive learning to geographic information and the assumptions it entails.

Object	Consumer income	Luxury car showroom Site suitability
Store1	High	Good
Store2	Medium	Good
Store3	Medium	Poor
Store4	Low	Poor

$Z = \{\text{Store1, Store2, Store3, Store4}\}$

$R_{income} = \{Y1, Y2, Y3\}$ where $Y1 = \{\text{Store1}\}$, $Y2 = \{\text{Store2, Store3}\}$, $Y3 = \{\text{Store4}\}$

$R_{site\ suitability} = \{X1, X2\}$ where $X1 = \{\text{Store1, Store2}\}$, $X2 = \{\text{Store3, Store4}\}$

As a result, the upper and lower approximations for stores with Site suitability attribute Good are

$$\bar{S}(Z) = \{\text{Store1, Store2, Store3}\} \quad \underline{S}(Z) = \{\text{Store1}\}$$

Figure 3: Example rough set analysis

By using rough set analysis, the software is able to utilise existing vague knowledge about attributes that impact site suitability. It is important that vague concepts, such as ‘Site suitability’, can be used in the analysis as these concepts are often central to the intuitive, partially subjective process of retail site assessment. In order to allow information in the form of cognitive maps to be used, two extensions to the analysis are needed. First, rough set analysis is usually conducted upon aspatial categorical information. In the example above, the implicit assumption is that the categorisation of income refers to consumers in some way spatially connected with the store, ie living in the same area. The software developed here is able to deal with spatial regions explicitly by treating a spatial region as a category, membership of which is defined by the region’s spatial extent. Second, to enable cognitive maps such as that in figure 2 to be analysed, the software performs a recursive rough set analysis where attributes for a single rough set analysis may be based on the results of other rough set analyses. Figure 4 illustrates how the cognitive map in figure 2 would be dealt with by performing three rough set analyses: first derive information about the vague concept of ‘Income’ (in terms of, say, low, medium and high); second derive information about the vague concept of ‘Customer category’ based in part on the results of the first analysis; finally derive information about the vague concept of ‘Retail site suitability’ based in part on the results of the second analysis.

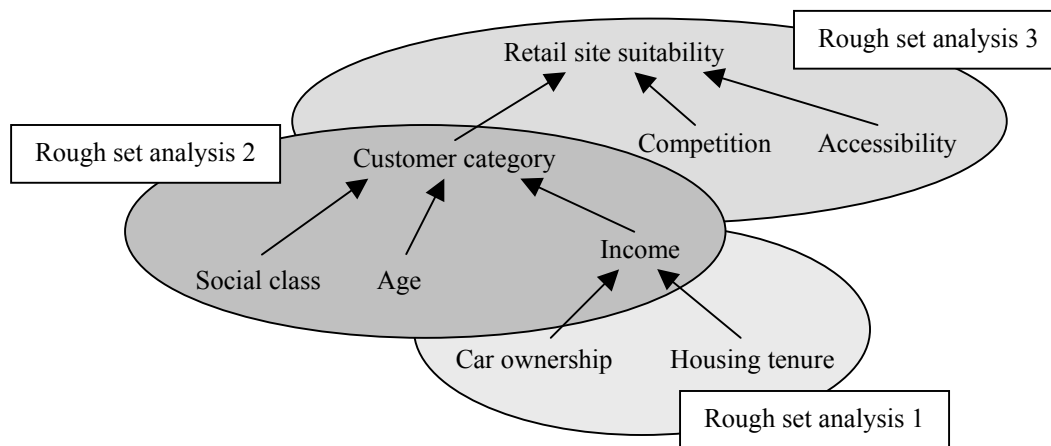


Figure 4: Recursive rough set analysis

5.2 Imprecision in aggregated census data

For reasons of privacy, most census data is publicly available for analysis only at an aggregated, relatively coarse level of spatial granularity. In contrast, the results of a retail site assessment may need to be at a much finer granularity, perhaps even down to the level of a location within a retail park. Unfortunately, the census

data commonly used as a basis for retail site analysis is sensitive to granularity changes: it is MAUP-susceptible. A range of geodemographic techniques have been developed to address MAUP, usually through the integration of census data with additional data, models or assumptions (Martin, 1995). It would have been possible to base the retail site assessment software on one of these existing techniques. However, the model allows a fresh perspective on this well trodden path. By thinking of MAUP-susceptible aggregated data as a rough set, it follows from the granular interpretation of rough sets in section 3.2 that every mapped areal unit may contain elements of the mapped phenomenon. Formally, for a set of mapped areas Z the lower approximation for the mapped phenomena in a census data set is the empty set, $\underline{S}(Z) = \emptyset$. At the same time the upper approximation is equal to Z itself, $\overline{S}(Z) = Z$. A rough set where $\underline{S}(Z) = \emptyset$ and $\overline{S}(Z) = Z$ is said to be *totally non-definable* (Munakata, 1998). Figure 5 below illustrates the problem: in moving from a census-like data set to a rough set, all the mapped areas are found to have value M for each mapped phenomena (in this case car ownership).

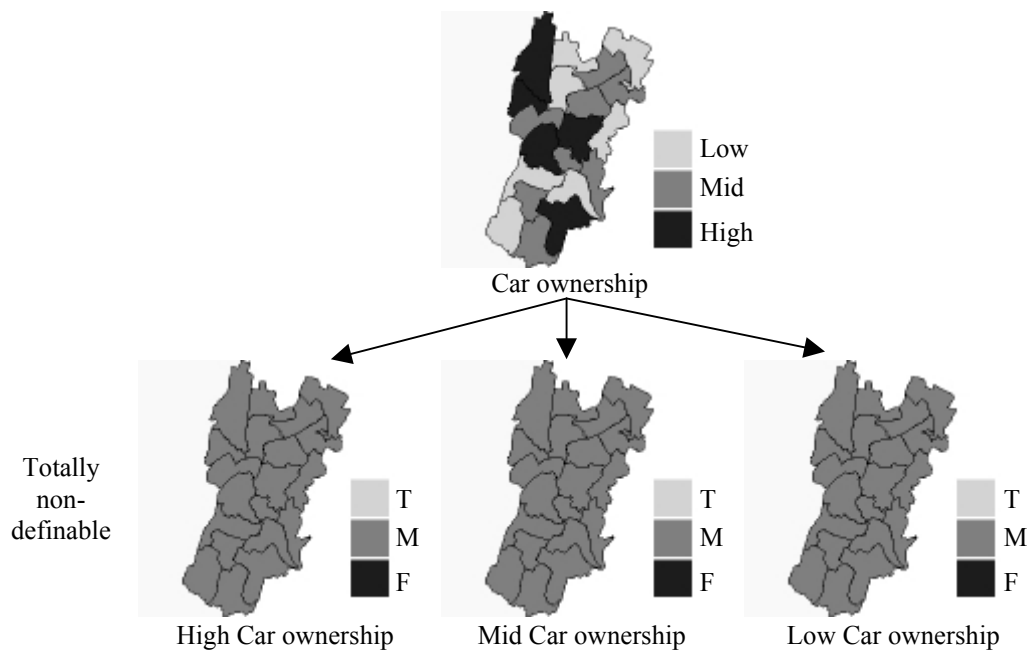


Figure 5: Totally non-definable data sets

Bearing the imprecise interpretation of rough sets in mind, the approach taken in developing the retail site assessment software was to treat totally non-definable data sets as a constraint on the analysis process. By formulating a series of increasingly restrictive assumptions, it is possible to derive an ordered series of realisations of the totally non-definable data set with which to constrain the analysis process. Building on figure 5, figure 6 gives an example of 4 such assumptions and the sets of maps derived from each. Starting from the bottom of figure 6, in figure 6d, the analysis is constrained by assuming that only those areal units which are actually as exhibiting high levels of car ownership, actually contain elements of high car ownership. This assumption is analogous to the *closed world assumption* common in database design, where “all and only” those relationships implied by the database are assumed to be true (Reiter, 1984). A less restrictive assumption, in figure 6c, constrains the analysis to both those areal units which are actually classified as exhibiting high car ownership, or are thematically *and* spatially adjacent to those units classified as high car ownership. Loosening the restriction further to include areal units that are thematically *or* spatially adjacent produces results illustrated in figure 6b, while the least restrictive totally non-decidable assumption presents no constraints to the analysis, shown in figure 6a and figure 5. The software tool described below performs repeated realisations of the retail site assessment, based on each of the assumptions outlined above. The results, when presented to the user, combine each of the repeated assessments, in an attempt to emphasise the compromise that has to be struck between greater specificity in the assessment results and greater dependence on assumptions.

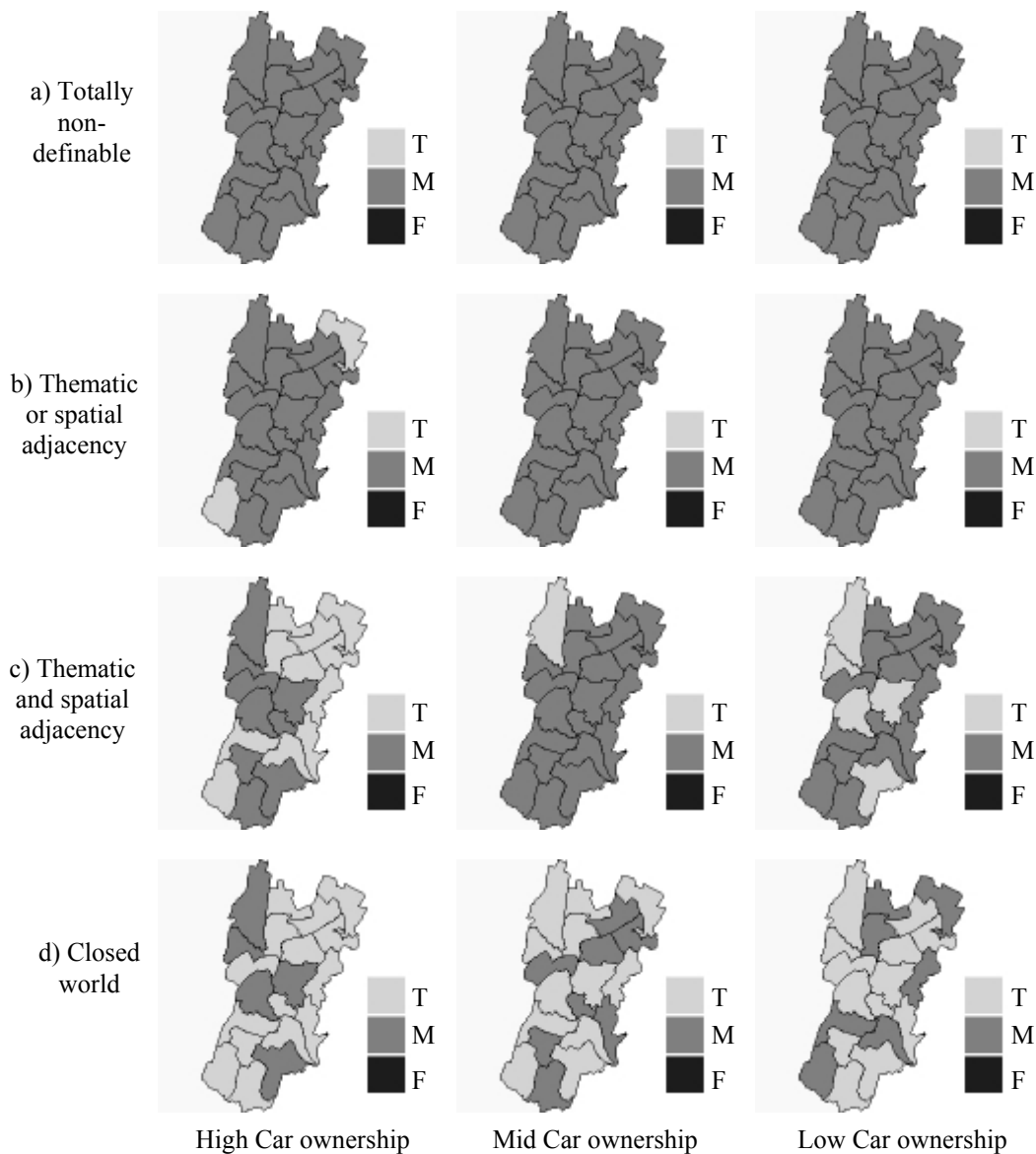


Figure 6: Totally non-definable data as a constraint on analysis

5.3 Java retail site assessment tool

The retail site assessment process described above was implemented using Java programming language. Figure 7 gives a snapshot of the tool interface, which illustrates the key stages of the analysis outlined below.

- The overall structure of the analysis is determined using the “Analysis” window. Users effectively construct their own conceptual map of the retail site assessment, adding concepts to the tree which reflect their individual intuitive understanding of the site assessment process.
- Mapped data can be associated with the leaves of the analysis tree. This mapped data may be in a variety of forms, including totally non-decidable information, such as the partially obscured car ownership data in the “CarOwnership” window in figure 7.
- Mapped data associated with the nodes of the analysis tree can be derived using the rough set analysis process. The tool needs to be ‘trained’ on a range of known locations that make up the set of observations Z , discussed in section 5.1. These observations are shown in tabular form in the “Observation” window.
- The results of the analysis for a particular node are shown in two forms. First, the “Classification rules” window gives a decision tree of rules inferred from the observations. Second, a map window displays the spatial expression of these rules, for example as in the “Income” map window.

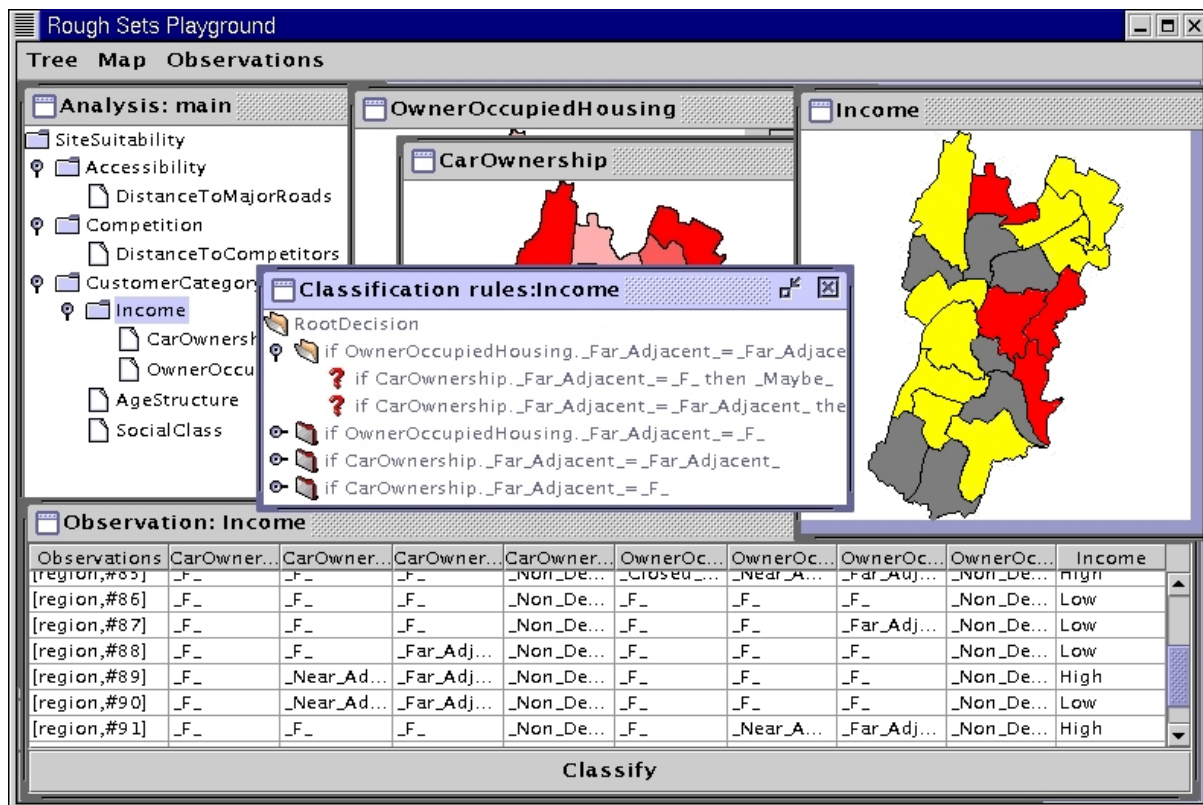


Figure 7: Rough set theory retail site assessment tool interface

The analysis tool already provides some encouraging indications. The integration of vague and imprecise information and concepts into the decision making process does seem possible using the approach outlined here. However, further work both on the tool and the approach generally is suggested by the experiences so far. The scope for such further work is the subject of the final section.

6. DISCUSSION AND FURTHER WORK

This paper describes how the combination of conceptual and computational models of imperfection can support the use and analysis of imperfect geographic information. The approach taken here represents an extension of previous work on imperfection, which has tended to emphasise either the conceptual or the computational. For example, Guptill and Morrison (1995) offer an extensive collection discussing of the key conceptual elements of spatial data quality, while from the converse perspective Hunter (1996) provides a comprehensive exploration of the computational aspects of uncertainty. The central contention of this paper is that both formal models, such as rough set theory, and conceptual models, such as the ontology of imperfection, are necessary to a more complete understanding of imperfection in geographic information. This balance between the conceptual and computational aspects of handling imperfect geographic information is crucial to the ultimate aim of developing practical tools for helping geographic information users to tackle complex and subtle data quality issues.

Undoubtedly some difficulties remain. Compared with other formal approaches to imperfection, such as fuzzy set theory, rough set theory is a relatively Spartan theory that can support only the most simple, least sophisticated models of imperfection. Amongst the further work highlighted below is the need for a wider range of formal models to be explored in addition to rough set theory. From a more practical perspective, the rough set analysis described in section 5 makes only limited use of the spatial relationships in the data. Topological relationships between different regions used in the analysis are not explicitly modelled, so the results of the site assessment within a particular spatial region will be largely independent of surrounding regions. Finally, the results of the analysis using totally non-definable data (section 5.2) can be difficult to visualise as these results comprise multiple mapped realisations of site suitability. Preliminary attempts were

made to combine the multiple realisations into a single “certainty map” and into an animated map series, but there is currently no clear mechanism for visualising the results that can be considered entirely satisfactory.

Despite these difficulties, this research suggests that further work may help realise important advantages for a variety of geographic applications in several ways. First, this work indicates that different formal interpretations can be used to address the different types of imperfection in our ontology. While the prototype application outlined in section 5 offers an integrated mechanism for handling imprecision and vagueness in geodemographic information, further work being planned aims to address imperfection more generally, including accuracy issues. Second, work currently underway will explore imperfection in different application areas, such as land use classification and natural resource management. The hope is that by providing a fresh perspective on long-standing problems, such as the MAUP discussed in section 4, the same general approach might profitably be used in a variety of application areas. Third, by using formal tools, imperfection can be tackled in data and in the operations performed upon the data. The prototype application in section 5 uses rough set theory to provide a mechanism to model imprecision and vagueness in geodemographical data and in the partially subjective reasoning process used during retail site assessment. Future work is planned aiming to extend these existing tools using a variety of logical formalisms, such as fuzzy set theory (Zadeh, 1965) and supervaluation theory (van Fraassen, 1969), in addition to rough set theory. Finally, the application as presented here is a prototype. Further work is planned to incorporate the site assessment tool within a full object-oriented GIS and to compare its performance with conventional retail site assessment techniques.

The underlying theme of this work is the potential application of information about imperfection to the difficult task of geographic information integration. By developing both the conceptual and computational mechanisms needed to support reasoning about imperfect geographic information, the ultimate aim is to facilitate much greater integration of information about vague geographic phenomena at different levels of precision and accuracy. This paper has focussed particularly on the application of information about imprecision and vagueness to the integration of geographic information in a geodemographic decision support context. Following on from the research reported in this paper, the long-term goal of this work is to provide a more general theoretical framework for the integration of imperfect geographic information and practical tools able to assist with integrating heterogeneous data sets across a range of application domains.

ACKNOWLEDGEMENTS

This project is supported by the EPSRC under grant GR/M 56685 “Managing vagueness, uncertainty and granularity in spatial information systems”, funded jointly with the School of Computer Studies, University of Leeds. The authors are grateful to the anonymous referee for some useful comments.

REFERENCES

- Ahlqvist, O., Keukelaar, J. and Oukbir, K. (2000) Rough classification and accuracy assessment, *International Journal of Geographical Information Science* 14, 475-496.
- Applebaum, W. (1966) Methods for determining store trading areas, market penetration and potential sales, *Journal of Marketing Research* 3, 127-141.
- Aspinall, R. (1992) An inductive modelling procedure based on Bayes theorem for analysis of pattern in spatial data, *International Journal of Geographical Information Systems* 6, 105-121.
- Cargile, J. (1969) The sorites paradox, *British Journal for the Philosophy of Science*, 20, 193-202.
- Clarke, I., Horita, M. and Mackaness, W. (1999) Capturing intuition in retail site assessment: towards the integration of 'hard' and 'soft' modelling approaches, paper presented at *Store, industrial and public facility location in a European context: theories and practices*.
- Birkin, M. (1995) Customer targeting, geodemographics and lifestyle approaches, in Longley, P. and Clarke, G. (Eds.), *GIS for business and service planning*, 104-149, New York, John Wiley.
- Chrisman, N. (1999) A transformational approach to GIS operations, *International Journal of Geographical Information Science* 13, 617-637.

- Duckham, M. and McCreddie, J. (1999) An intelligent, distributed, error-aware object-oriented GIS, in W. Shi, M. Goodchild and P. Fisher (Eds.) *Proceedings International Symposium on Spatial Data Quality*, pp.496-505.
- Drummond, J.E. (1995) Positional accuracy, in S. Guptill and J. Morrison (Eds.) *Elements of spatial data quality*, pp.31-58, Oxford, Elsevier Science.
- Erwig, M. and Schneider, M. (1997) Vague regions, in A.U. Frank and W. Kuhn (Eds), *Advances in spatial databases (SSD '97)*, Lecture Notes in Computer Science, 1262, 298-320, Berlin, Springer-Verlag.
- Heuvelink, G.B.M. (1998) *Error propagation in environmental modelling with GIS*, London, Taylor & Francis.
- Hunter, A. (1996) *Uncertainty in information systems*, London, McGraw Hill.
- Feng, Z. and Flowerdew, R. (1999) The use of fuzzy classification to improve geodemographic targeting, in B. Gittings (Ed.), *Innovations in GIS 6*, 133-144, London, Taylor & Francis.
- Fisher, P.F. (1997) The pixel: a snare and a delusion, *International Journal of Remote Sensing* 18, 679-685.
- Fisher, P.F. (2000) Sorites paradox and vague geographies, *Fuzzy Sets and Systems* 113, 7-18.
- Goodchild, M. (1995) Sharing imperfect data, in H. Onsrud and G. Rushton (Eds.), *Sharing geographic information*, 413-425, New Jersey, Rutgers.
- Guesgen, H.W. and Albrecht, J. (2000) Imprecise reasoning in geographic information systems, *Fuzzy Sets and Systems* 113, 121-131.
- Guptill, S. and Morrison, J. (Eds.) (1995) *Elements of spatial data quality*, Oxford, Elsevier Science.
- Keefe, R. and Smith, P. (1996) Theories of vagueness, in Keefe, R. and Smith, P. (Eds.), *Vagueness: a reader*, Cambridge, Massachusetts, MIT Press.
- Leung, Y. (1987) On the imprecision of boundaries, *Geographical Analysis* 19, 125-151.
- Martin, D. (1995) Censuses and the modelling of population in GIS, in Longley, P. and Clarke, G. (Eds.), *GIS for business and service planning*, 48-71, New York, John Wiley.
- Mowrer, T. (1999) Accuracy (re)assurance: selling uncertainty assessment to the uncertain, in K. Lowell and A. Jaton (Eds.), *Spatial accuracy assessment*, 3-10, Michigan, Ann Arbor.
- Munakata, T. (1998) *Fundamentals of the new AI*, Berlin, Springer.
- Openshaw, S. and Alvanides, S. (1999) Applying geocomputation to the analysis of spatial distributions, in P.A. Longley, M.F. Goodchild, D.J. Maguire and D.W. Rhind (Eds), *Geographical information systems*, 267-282, New York, Wiley.
- Reiter, R. (1984) Towards a logical reconstruction of relational database theory, in M. Brodie, J. Mylopoulos and J. Schmidt (Eds.), *On conceptual modelling*, 191-233, Michigan, Ann Arbor.
- Russell, S. and Norvig, P. (1995) *Artificial intelligence: a modern approach*, New Jersey, Prentice Hall.
- Stell, J.G. (1999) Granulation for graphs, in C. Freska and D.M. Mark (Eds), *Spatial information theory (COSIT '99)*, Lecture Notes in Computer Science, 1661, 417-432, New York, Springer-Verlag.
- van Fraassen, B.C. (1969) Presuppositions, supervaluations and free logic, in K. Lambert (Ed.), *The logical way of doing things*, 67-91, New Haven, Yale University Press (1969).
- Veregin, H. (1999) Data quality parameters, in P.A. Longley, M.F. Goodchild, D.J. Maguire and D.W. Rhind (Eds), *Geographical information systems*, 177-189, New York, Wiley.
- Woodcock, C.E. and Gopal, S. (1999) Fuzzy set theory and thematic maps: accuracy and area estimation, *International Journal of Geographical Information Science* 14, 153-172.
- Worboys, M.F. (1998a) Imprecision in finite resolution spatial data, *GeoInformatica*, 2, 257-280.
- Worboys, M.F. (1998b) Computation with imprecise geospatial data, *Computers, Environment and Urban Systems*, 22, 85-106.
- Zadeh, L.A. (1965) Fuzzy Sets, *Information and Control* 8, 338-353.