# Representing Vague Places : Determining a Suitable Method

Mohammed Imaduddin Humayun and Angela Schwering

Institute for Geoinformatics, University of Muenster, Germany
humayun@uni-muenster.de, schwering@uni-muenster.de

**Abstract.** The representation of places with vague or ill-defined boundaries continues being an issue for information systems. Despite the presence of multiple representation methods, it is still unclear how to determine which approach is best suited for a particular task. This paper proposes a set of characteristics based on the application domain, conceptual, and logical levels and differentiates the approaches according to these characteristics. We demonstrate how they are matched with the task requirements and influence the choice of representation method.

## 1 Introduction

Representing 'place' poses challenges, more so when the extents or boundaries cannot be well-defined. Although humans are capable of interpreting what is being referenced in such cases, handling these in information systems is more complex. Some of the associated problems are discussed in literature under discipline of *spatial vagueness* and *uncertainty*, especially their philosophical and representational aspects. The former aspects address whether vagueness is intrinsic to the real world or just a feature of language[1], the different kinds of vagueness [2] and how to handle imperfection in geographic information [3]. The latter suggest models and theories to handle spatial vagueness, each with its distinct assumptions and properties. This has resulted in the development of various representation methods such as *probabilistic* [4,5], *fuzzy-set based* [6,7], *egg-yolk* model [8], *rough-sets* [9], and *supervaluation*[10], among others. We propose a methodology to distinguish between different representation methods based on their characteristics, which may then be matched with the application requirements in order to determine a suitable method.

No single representation can claim to be applicable for all cases. The methods differ in the way assumptions are made about space, the underlying formal models, applicability of data models and the kinds of reasoning they allow. Selecting the right one for a given task is a matter of fitness for purpose and requires that the method's capabilities are matched to the requirements. Requirements vary and can be specified in numerous ways. A consistent way of specifying these requirements is needed. Our approach is to specify these in terms of the model characteristics. The characteristics themselves may be defined at different levels similar to the levels of data abstraction in an information system [11].

1. At the *application domain* level, a subset of the reality to be represented is chosen with respect to a particular domain. We also specify what kind of reasoning is to be performed on a representation. It is also important to decide here whether vagueness is perceived to be intrinsic to the entity, or if different possible interpretations should be supported.
2. The *conceptual* level is the next and deals with how the vague place is conceptualized in an implementation independent fashion. Important concerns here are, how the vague referent can be *individuated*? (e.g. through use of objectifiable parameters), how it is *demarcated*? and its *identity* (e.g. temporal changes).
3. The *logical* level is the next and deals with more detailed specifics such as the data model of the data sources, or how the extents should be modelled.

In this paper, we identify a criteria set to determine suitable representation methods for vague places. Section 2 analyzes the requirements of an application task. Based on these requirements, section 3 develops a criteria set and differentiates vague representations in terms of these criteria. Section 4 gives an example how to choose the right representation method for a given task based on using our criteria set.

## 2 Analyzing Requirements for a Use Case

Lake Carnegie in Australia is ephemeral. Depending on the amount of precipitation the lake may or may not be filled with water (Fig. 1). Though a lake in vernacular terms, in dry seasons it is reduced to a muddy marsh[1]. This presents a problem, since it is now unclear where exactly the boundaries of the lake lie.

Suppose a user needs a representation of Lake Carnegie. We examine a few questions that need to be answered to arrive at a clear understanding of what needs to be represented.

1. *What is a 'lake'?* First, the semantics of the term 'lake' need to be clear. Is it a single contiguous body of water or does it include smaller scattered pools in the vicinity as well? Do the requirements dictate that water be present in the lake all year round? This is treated as the first step towards arriving at a solution.
2. *What is the purpose of representation?* Requirements for an ecologist differ from that of a cartographer. An ecologist is likely to be interested in the variation of the lake over time; a fuzzy spatial extent rather than precisely defined boundaries being of importance. A cartographer is more interested in the lake as a crisp object.
3. *What data sources are available?* The choice of representation is influenced by the data sources. A different method is needed for a representation built up from satellite imagery, than another which uses water level observations from sensors scattered through the lake.

---

[1] `http://www.nasa.gov/multimedia/imagegallery/image_feature_817.html`

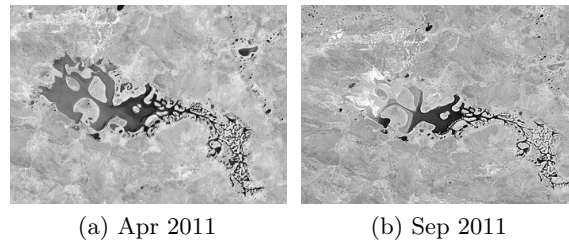(a) Apr 2011                    (b) Sep 2011

Fig. 1: Landsat images of Lake Carnegie, Australia
(*courtesy: U.S. Geological Survey*)

As one can infer, varying user needs and requirements must be met with a suitable method to represent the same place. From the myriad of possibilities, one needs a way to identify the right representation approach. Next, we briefly propose certain characteristics and explain how these characteristics can be used to distinguish among representation methods as the first step to this end.

## 3   Methods of Representation

From the different levels of abstraction, we identify characteristics which will serve as the criteria to differentiate between methods. Some commonly used methods for representing vague places are then briefly analyzed based on these.

### 3.1   Criteria for Differentiation

Starting from the different levels of abstraction, we propose the following characteristics for use in deciding upon the correct method to employ for representation of vague entities.

1. *Conceptualization of space* - The adopted perspective of vagueness (whether it is linguistic or ontic) affects the choice of the representational and semantic framework [12]. How the phenomenon is treated by the method forms a useful basis for differentiation. This also has implications on the kind of boundary of the phenomenon (crisp, graduated, indeterminate etc.).
2. *Formal model* - This differentiates between the methods based on whether the underlying model is *stochastic, fuzzy set based, three-valued logic* or other.
3. *Data model* - Certain methods handle only regions (egg-yolk and supervaluation) whereas others are well suited for points or grid based data structures (fuzzy sets for instance). Since sources of data differ according to the data model they use (raster versus vector data), it is important to consider how a method behaves with respect to it. This also has implications on the kinds of boundaries that can be defined, e.g. how is a crisp boundary generated in the *raster* data model?

4. *Reasoning* - This characteristic determines what kinds of reasoning can be performed with the representation methods. Reasoning covers metric, directional and topological operations performed on vague places. This is particularly important from the perspective of a task, since it limits what kind of analysis can be done on the vague place. Some representations provide a well-defined framework for reasoning, whereas others do not.

## 3.2   Analysis of Representation Methods

***Base representations -*** We coin this term to refer to those methods which abstract or crisp the vague place. Possible ways are to define the feature *a priori* according to some metric, or reduce it to a simple feature type (point, line or polygon), a minimum bounding rectangle (MBR) which covers the entire extent of the space where the entity is located, or through tessellations of space. These are usually in the form of vector data. Examples may be seen in VGI where a real world feature is outlined by contributors (from GPS tracks or tracing from aerial imagery), or in gazetteers where a feature is simply located by a representative point. Here vagueness is not preserved, and they are generally not classified under methods for vagueness representation. They are included here for the sake of completeness since they are often applied and prove adequate in some cases. The methods themselves do not provide any theory for reasoning.

***Probabilistic methods -*** These methods derive the membership value of an individual in a set through a statistically defined probability function. These are used mainly to handle *uncertainty*. The underlying stochastic model assumes that phenomena are crisp and knowable, with the result that no measurable way for metrics such as *precision* in the case of vagueness exist [13]. These methods are best suited for phenomena with measurable objective properties such as flow, temperature, or water level. Probabilistic interpretations have also been employed to determine where city centres lie, based on probability of sample points computed from trials using participant studies [5]. These are generally suited for point or field based data and allow for a variety of statistical reasoning techniques to be performed.

***Fuzzy-set methods -*** These are based on Fuzzy-set theory and ideal for modelling objects which have graduating or indeterminate boundaries [6,7,14]. The membership value $\alpha$ ( $0 \leq \alpha \leq 1$), of a point in the region is highest at the *core* of the region and decreases gradually as the *boundary* is approached. Determination of membership value itself is subjective and may not relate directly to the phenomenon itself. The model also allows for obtaining a crisp boundary by means of $\alpha$-cuts which are a way of obtaining crisp sets from a fuzzy set. This method is applicable in the case of both raster and vector data, where a feature or cell may be assigned a membership. Reasoning using fuzzy set operations such as intersection, union, complement etc. can also be performed [6].

***Egg-yolk method -*** A vague region is considered analogous to an egg – the *yolk* corresponds to the minimal extent, the *white* being the indeterminate region and its maximal extent. Any acceptable crisping must lie between these inner and outer subregions. This method allows performing qualitative spatial reasoning between vague regions or between a vague region and a crisp region under the framework of the regional connection calculus (RCC-5) [8]. Reasoning is possible on different possible configurations of two regions represented this way. However, the theory itself does not make any assertion as to how the crisp regions are obtained. Egg-yolk models are ideal when topological reasoning on vector data in the form of regions is to be performed.

***Rough sets -*** The basis for rough sets is the *indiscernibility relation* – where a collection of elements is indiscernible from another. Rough sets use a three-valued logic ($true, false, maybe$) to determine the membership of a point to a region as opposed to the binary notion of membership ($true, false$) in classical set theory. Similar to the egg-yolk model, a region may be represented by its determinate *lower approximation* and an indeterminate *upper approximation* [9,3]. Rough sets are ideal for reasoning on multi-resolution raster data, where a change in resolution results in indiscernibility.

***Supervaluation -*** The idea behind *supervaluation* is to account for the different possible interpretations of a vague predicate when multiple interpretations for a vague region exist. The *positive extension* is where all interpretations are true. Its inverse the *negative extension* is the region where no interpretation is true. The remaining regions constitute the *penumbra* [10]. Supervaluation enables use of classical logic to reason about vagueness, but computational applications are hampered by the fact that all admissible interpretations must be explicitly specified, which is difficult in practice [12]. These are applicable in data models where regions are primitives and allow for reasoning on vague regions where several boundaries may be associated with an object.

## 4    Determining a Suitable Representation - An Example

We take the example of Lake Carnegie and consider two different sets of requirements for representations.

- The cartographer imagines the lake to be a single contiguous body of water with a crisp boundary, though the reality is different. Satellite imagery is used as source of data. No reasoning needs to be performed.
- The ecologist views the lake as a non-crisp object defined by level of water. Water level observation data from sensors is available. The need is to generate a surface exhibiting water presence over a period of time.

In the first case, the space is conceptualized as a crisp body. Data from satellite imagery is a raster, from which the boundary needs to be derived. One obvious solution is to simply trace the outline from the image, resulting in a base

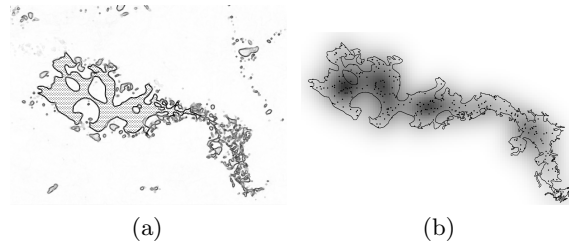(a)                                    (b)

Fig. 2: (a) Fuzzy set and (b) probabilistic representations of Lake Carnegie

representation. The suggested approach in this case is however to use fuzzy-set modelling with $\alpha$-cuts. By varying $\alpha$, different boundaries (each of which is crisp) can be obtained. One such representation by using fuzzy membership of pixel values and obtaining a crisp boundary is seen in Fig. 2a.

In the second case, the lake is conceptualized depending on its objective property (water level). Spatial distribution of the data source, sensors which provide observations, can be thought of as consisting of points (vector). Since the user needs to obtain interpolated values in order to obtain a lake surface, application of probabilistic methods is suitable here. A probabilistic representation simulated from randomly distributed sensors with arbitrary observations is shown in Fig. 2b, with outline of the lake from OpenStreetMap [2] for reference.

This is a trivial example, but the same principles apply in other cases as well. For example, in the Tell Us Where[3] project dataset, it would be possible to obtain a representation of places with noncrisp boundaries. This however has not been attempted here owing to sampled locations in the current dataset being insufficient in number to demonstrate our cause.

## 5   Conclusion

Various representation methods have been proposed in literature for the representation of places with vague boundaries. It is important to enable decision makers to adopt the right method based on their needs. The approach taken here uses multiple levels of abstraction to specify the requirements in a consistent manner. The levels allow identification of characteristics with which different methods can be analyzed. Differing requirements in the modelling of a vague region such as a lake can lead to different possible solutions as presented in the lake use case.

## 6   Acknowledgements

This research was carried out under the International Research Training Group on Semantic Integration of Geospatial Information (IRTG-SIGI) and is funded by the DFG (German Research Foundation), GRK 1498.

---

[2] http://www.openstreetmap.org
[3] http://telluswhere.net

# References

1. Varzi, A.C.: Vagueness in geography. Philosophy & Geography **4**(1) (2001) 49–65
2. Bennett, B.: Modes of concept definition and varieties of vagueness. Applied Ontology **1**(1) (2005) 17–26
3. Duckham, M., Mason, K., Stell, J., Worboys, M.: A formal approach to imperfection in geographic information. Computers, Environment and Urban Systems **25**(1) (2001) 89 – 103
4. Leung, Y., Yan, J.: A locational error model for spatial features. International Journal of Geographical Information Science **12**(6) (1998) 607–620
5. Montello, D.R., Goodchild, M.F., Gottsegen, J., Fohl, P.: Where's downtown?: Behavioral methods for determining referents of vague spatial queries. Spatial Cognition & Computation **3**(2) (2003) 185–204
6. Usery, E.L.: A conceptual framework and fuzzy set implementation for geographic features. In Burrough, P.A., Frank, A.U., eds.: Geographic Objects with Indeterminate Boundaries. Taylor and Francis, London (1996)
7. Wang, F., Hall, G.B.: Fuzzy representation of geographical boundaries in gis. International journal of geographical information systems **10**(5) (1996) 573–590
8. Cohn, A., Gotts, N.: The 'egg-yolk' representation of regions with indeterminate boundaries. Geographic Objects With Indeterminate Boundaries **2** (1996) 171–187
9. Worboys, M.: Imprecision in finite resolution spatial data. GeoInformatica **2**(3) (1998) 257–279
10. Kulik, L.: A geometric theory of vague boundaries based on supervaluation. In Montello, D., ed.: Spatial Information Theory. Volume 2205 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2001) 44–59
11. Worboys, M., Duckham, M.: GIS: A Computing Perspective. 2nd edn. CRC Press (2004)
12. Bennett, B.: Spatial vagueness. In Jeansoulin, R., Papini, O., Prade, H., Schockaert, S., eds.: Methods for Handling Imperfect Spatial Information. Springer (2011) 15–47
13. Duckham, M., Sharp, J.: Uncertainty and geographic information: Computational and critical convergence. In: Re-presenting GIS. John Wiley & Sons Ltd. (2005) 113–124
14. Fisher, P., Wood, J., Cheng, T.: Where is Helvellyn? Fuzziness of multi-scale landscape morphometry. Transactions of the Institute of British Geographers **29**(1) (2004) 106–128