Vague Spatial Data Types

Alejandro Pauly and Markus Schneider
CISE Department, University of Florida

SYNONYMS

Data types for uncertain, indeterminate, or imprecise spatial objects.

DEFINITION

Naturally occurring phenomena in space often (if not always) cannot be precisely defined because of the intrinsic uncertainty of their features. The location of animal refuges might not be precisely known, and the path of rivers might be uncertain due to water volume fluctuations and changing land characteristics. The extension of lakes can also change and thus have uncertain areas. All these are examples of *vague spatial objects*. The animal refuge locations can be modeled as a *vague point* object where the precisely known locations are called the *kernel point* object and the assumed locations are denoted as the *conjecture point* object. The river paths can be modeled as *vague line* objects. Some segments or parts of the path, called *kernel line* objects, can be definitely identified since they are always part of the river. Other paths can only be assumed, and these are denoted as *conjecture line* objects. Knowledge about the extension of lakes can be modeled similarly with *vague regions* formed by *kernel* and *conjecture* parts. Figure 1 illustrates the examples above. Dark shaded areas, straight lines, and black points indicate kernel parts; areas with light gray interiors, dashed lines, and hollow points refer to conjecture parts.

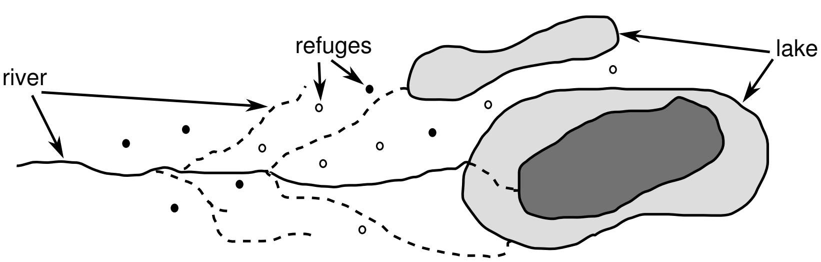

Figure 1: Examples of a (complex) vague point object representing the animal refuges, a (complex) vague line object as a river, and a (complex) vague region object representing a lake.

As another example, consider a homeland security scenario in which secret services (should) have knowledge of the whereabouts of terrorists. For each terrorist, some of their refuges are precisely known, some are not and only conjectures. These locations can be modeled as a *vague* *point* object where the precisely known locations are represented by the kernel part of the object and the assumed locations are denoted as its conjecture part. Secret services are also interested in the routes a terrorist takes to move from one refuge to another. These routes can be modeled as *vague line* objects. Some routes, represented by the kernel part of the object, have been identified. Other routes can only be assumed to be taken by a terrorist; they are denoted as the conjecture part of the object. Knowledge about areas of terrorist activities is also important for secret services. From some areas it is well known that a terrorist operates in them. These areas are denoted as the kernel parts. From other areas it can only be assumed that they are the target of terrorist activity, and they are denoted as the conjecture parts. Figure 1 gives some examples. Grey shaded areas, straight lines, and gray points indicate kernel parts; areas with white interiors, dashed lines, and white points refer to conjecture parts.

The definition of vague points, vague lines, and vague regions leverages the data types *point* for crisp points, *line* for crisp lines, and *region* for crisp regions. All crisp spatial data types $\alpha \in \{point, line, region\}$ are assumed to have a complex inner structure as it has been defined in [4]. In particular, this means that a *point* object includes a finite number of single points, a *line* object is assembled from a finite number of curves, and a *region* object consists of a finite number of disjoint faces possibly containing a finite number of disjoint holes. Further, these types must be closed under the geometric set operations *union* ($\oplus : \alpha \times \alpha \rightarrow \alpha$), *intersection* ($\otimes : \alpha \times \alpha \rightarrow \alpha$), *difference* ($\ominus : \alpha \times \alpha \rightarrow \alpha$), and *complement* ($\sim: \alpha \rightarrow \alpha$). Each type α together with the operations $\oplus$ and $\otimes$ forms a boolean algebra. The identity of $\otimes$ is denoted by **1**, which corresponds to $\mathbb{R}^2$. The identity of $\oplus$ is presented by **0**, which corresponds to the empty spatial object (empty point set).

A vague spatial object is defined by a pair of two *disjoint* or *meeting* crisp complex spatial objects [5]. The extension of a crisp spatial data type to a corresponding vague type is given by a type constructor v as follows:

$$v(\alpha) \quad = \quad \alpha \times \alpha \qquad\qquad \forall\, \alpha \in \{point, line, region\}$$

This means that for $\alpha = region$ the type $v(region) = region \times region$, which is also named *vregion* is defined. Accordingly, $v(line) = line \times line$ and $v(point) = point \times point$ define *vline* and *vpoint* respectively. For a vague spatial object $A = (A_k, A_c) \in v(\alpha)$, the first crisp spatial object A_k, called the *kernel part*, describes the determinate component of A, that is, the component that definitely and always belongs to the vague object. The second crisp spatial object A_c, called the *conjecture part*, describes the vague component of A, that is, the component for which it cannot be said with any certainty whether it or subparts of it belong to the vague object or not. *Maybe* the conjecture part or subparts of it belong to the vague object, *maybe* this is not the case. Since the kernel part and the conjecture part of the *same* vague spatial object may not share interior points a restriction is imposed to assure that the interior point sets[1] do not intersect, formally:

$$\forall\, \alpha \in \{point, line, region\}$$
$$\forall A = (A_k, A_c) \in v(\alpha) : A_k^\circ \cap A_c^\circ = \emptyset$$

Hence, A_k can be regarded as a lower (minimal, guaranteed) approximation of A and $(A_k \oplus A_c)$ can be considered as an upper (maximally possible, speculative) approximation of A.

[1] x° is used to denote the interior point set of crisp spatial object x

HISTORICAL BACKGROUND

Spatial vagueness has to be seen in contrast to spatial uncertainty resulting from either a lack of knowledge about the position and shape of an object (*positional* uncertainty) or the inability of measuring such an object precisely (*measurement* uncertainty). Much literature has been published on dealing with positional and measurement uncertainty; it mainly proposes probabilistic models. Spatial vagueness is an intrinsic feature of a spatial object for which it cannot be said whether certain components belong to the spatial object or not. The design goal for dealing with spatial vagueness in VASA is to base the definition of vague spatial data types and their operations on already existing definitions of exact spatial objects. This so-called exact model approach is also followed in the definition of broad-boundary regions [1] and the egg-yolk approach [2] as it is detailed in [this same chapter]. A generalization of the ideas from the broad-boundary approach can be found in the original definition of *vague regions* [3]. This definition proposes a data type for vague regions that is closed under the union, intersection, difference, and complement operations. The components of VASA are based on the original vague regions concept which which is generalized in order to deal with vague points and vague lines.

SCIENTIFIC FUNDAMENTALS

One of the major objectives of exact model based design is to make use of the formalisms introduced by the underlying models upon which the design is based. This allows the new design to relay the major responsibilities of robustness and correctness to the underlying model. A side effect of this type of design is the centralization of the mathematical definitions that form the core of both the underlying model and the new model. The result is a modular design that enables more robust and less error prone specifications.

In the next section the proper definitions of vague sptial operations are formalized. Further details, specifically in what relates to topological predicates between vague spatial objects can be found in [6, 7].

Vague Spatial Operations

The three vague geometric set operations **union**, **intersection**, and **difference** have all the same signature $v(\alpha) \times v(\alpha) \to v(\alpha)$. In addition, the operation **complement** is defined with the signature $v(\alpha) \to v(\alpha)$. All of these operations are defined in a type-independent and thus generic manner. In order to define them for two vague spatial objects u and w, it is helpful to consider meaningful relationships between the kernel part, the conjecture part, and the outside part of u and w. For each operation a table is given where a column/row labeled by k, c, or o denotes the kernel part, conjecture part, or outside part of u/w. Each entry of the table denotes a possible combination, i.e., intersection, of kernel parts, conjecture parts, and outside parts of both objects, and the label in each entry specifies whether the corresponding intersection belongs to the kernel part, conjecture part, or outside part of the operation's result object.

The *union* (Table 1) of a kernel part with any other part is a kernel part since the union of two vague spatial objects asks for membership in either object and since membership is already assured by the given kernel part. Likewise, the union of two conjecture parts or the union of a conjecture

union	k	c	o
k	k	k	k
c	k	c	c
o	k	c	o

intersection	k	c	o
k	k	c	o
c	c	c	o
o	o	o	o

difference	k	c	o
k	o	c	k
c	o	c	c
o	o	o	o

complement	k	c	o
	o	c	k

Table 1: Components resulting from intersecting kernel parts, conjecture parts, and outside parts of two vague spatial objects with each other for the four vague geometric set operations.

part with the outside should be a conjecture part, and only the parts which belong to the outside of both objects contribute to the outside of the union.

The outside of the *intersection* (Table 1) is given by either region's outside because intersection requires membership in both regions. The kernel part of the intersection only contains components which definitely belong to the kernel parts of both objects, and intersections of conjecture parts with each other or with kernel parts make up the conjecture part of the intersection.

Obviously, the *complement* (Table 1) of the kernel part should be the outside, and vice versa. With respect to the conjecture part, anything inside the vague part of an object might or might not belong to the object. Hence, it cannot be said with certainty that the complement of the vague part is the outside. Neither can be said that the complement belongs to the kernel part. Thus, the only reasonable conclusion is to define the complement of the conjecture part to be the conjecture part itself.

The definition of *difference* (Table 1) between u and w can be derived from the definition of complement since it is equal to the intersection of u with the complement of v. That is, removing a kernel part means intersection with the outside which always leads to outside, and removing anything from the outside leaves the outside part unaffected. Similarly, removing a conjecture part means intersection with the conjecture part and thus results in a conjecture part for kernel parts and conjecture parts, and removing the outside of w (i.e., nothing) does not affect any part of u.

Motivated by the intended semantics for the four operations described above, the formal definitions are provided. An interesting aspect is that these definitions can be based solely on already known crisp geometric set operations on well-understood exact spatial objects. Hence, *executable specifications* can be defined for the vague geometric set operations. This means, once having the implementation of a crisp spatial algebra available, it can directly *execute* the vague geometric set operations without being forced to design and implement new algorithms for them.

Let $u, w \in v(\alpha)$, and let u_k and w_k denote their kernel parts and u_c and w_c their conjecture parts:

$$
\begin{aligned}
u \textbf{ union } w &:= (u_k \oplus w_k, (u_c \oplus w_c) \ominus (w_k \oplus w_k)) \\
u \textbf{ intersection } w &:= (u_k \otimes w_k, (u_c \otimes w_c) \oplus (u_k \otimes w_c) \oplus (u_c \otimes w_k)) \\
u \textbf{ difference } w &:= (u_k \otimes (\sim w_k), (u_c \otimes w_c) \oplus (u_k \otimes w_c) \oplus (u_c \otimes (\sim w_k))) \\
\textbf{complement } u &:= (\sim u_k, u_c)
\end{aligned}
$$

Spatial operations that result in unique numeric values can be applied to vague spatial objects generally by transforming the result into ranges of values. That is, the operations are specified as executions of their crisp versions returning a lower bound result and an upper bound result. These values depend on whether the conjecture parts are considered in the computation or not.

To compute the minimum distance between two vague spatial objects, define the operations *vague-min-mindistance* : $v(\alpha) \times v(\beta) \rightarrow$ *real* and *vague-max-mindistance* : $v(\alpha) \times v(\beta) \rightarrow$ *real* are defined. Both operations can be applied to pairs of vague spatial objects of possibly distinct types (that is, $(\alpha = \beta \lor \alpha \neq \beta)$ and $\alpha, \beta \in \{point, line, region\}$). The first operation considers all kernel and conjecture parts, thus returning the minimum possible distance between both objects. The second operation, only considers kernel parts, thus returning the maximum possible minimum distance between both objects that is, the maximum value that the minimum distance will actually be. An illustration of the maximum minimum distance and minimum minimum distance between two vague regions is shown in Figure 2. Formally, let $mindistance : \alpha \times \beta \rightarrow real$ be the minimum distance operation defined for crisp spatial objects:

$$\begin{aligned}
\textit{vague-min-mindistance}(u,w) &:= \textit{mindistance}((u_k \oplus u_c),(w_k \oplus w_c)) \\
\textit{vague-max-mindistance}(u,w) &:= \textit{mindistance}(u_k,(w_k)
\end{aligned}$$

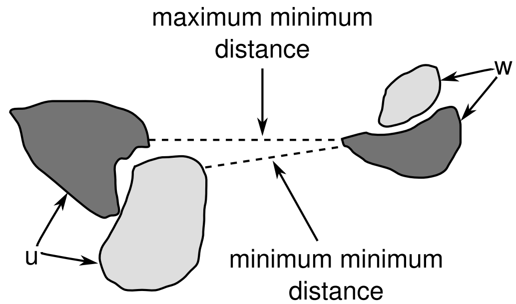

Figure 2: An example illustrating the maximum and minimum minimum distances between two vague spatial regions. The dark shaded areas conform the kernel parts of the objects and the light shaded areas represent the conjecture parts.

Unary numeric operations are used to express properties of a vague spatial object. The operations $min\text{-}length : vline \rightarrow real$ and $max\text{-}length : vline \rightarrow real$ are defined to compute the range of the length of a vague line object. The operations $min\text{-}area : vregion \rightarrow real$ and $max\text{-}area : vregion \rightarrow real$ are used to compute the area of a vague region. Inversely to the distance operation, the minimum length (area) of a vague line (region) is computed by taking into consideration all parts, including the conjecture part of the object. The maximum length (area) is computed by only considering the kernel part of the object. Formally, let $length : line \rightarrow real$ and $area : region \rightarrow real$ refer to the operations that compute the length and area of a crisp line and region respectively. Also consider $a \in vline$ and $b \in vregion$:

$$\begin{aligned}
\textit{min-length}(a) &:= \textit{length}(a_k) \\
\textit{max-length}(a) &:= \textit{length}(a_k \oplus a_c) \\
\textit{min-area}(b) &:= \textit{area}(b_k) \\
\textit{max-area}(b) &:= \textit{area}(b_k \oplus b_c)
\end{aligned}$$

The definitions provided above serve as a sample of the operations that can be defined for vague spatial objects as an executable specification of operations on the underlying crisp spatial objects.

KEY APPLICATIONS

Generally, because many GIS applications largely deal with naturally occurring spatial phenomena that often contain implicit uncertain features, they will all benefit from data models that include considerations for dealing with spatial vagueness. Experts from a wide range of domains such as biology and agriculture can begin to take into account the inexact data that can make a difference in their decision making process. The following three applications are just examples of the wide range of domains that can benefit from dealing with vague spatial data.

- **Ecology**: Ecologists require an abundance of data related to the distribution and interactions of living organisms in their environment. The vast majority of these data suffers from indeterminacy stemming not only from its implicit nature but also from the inability to process exact observations. As a result many of the data is inferred or approximated from actual observations and thus must be treated as uncertain.
- **Military**: Military operations are often designed on the basis of intelligence collected on-site or remotely via technological media. It is also often the case that the intelligence is vague because only pieces of information can be collected or because the sources are not trustful (amongst other reasons).
- **Soil Sciences**: Soil variability is often a problem that must be taken into account when treating soil related data. Fine grained modeling of soil data turns out to be very costly, due to its high variability and inhomogeneity. As a result, in many applications within soil studies, it is enough to approximate the composition of soil to vague models in order to generate the necessary information.

FUTURE DIRECTIONS

As new data models are generated, the main element upon which their popularity depends is the availability of appropriate data sets. VASA provides a data model appropriate for dealing with uncertainty of spatial data. To motivate its future use, it is necessary to collect data in the appropriate format so that it can be fully exploited by the data model.

CROSS REFERENCES

Broad-Boundary Regions
Egg-Yolk Approach
Fuzzy and Rough Set Approaches for Uncertainty in Spatial Data

RECOMMENDED READING

[1] E. Clementini and P. Di Felice. *An Algebraic Model for Spatial Objects with Indeterminate Boundaries*, In *Geographic Objects with Indeterminate Boundaries*, Taylor & Francis, 1996, pp. 153–169.

[2] A. G. Cohn and N. M. Gotts. *The 'Egg-Yolk' Representation of Regions with Indeterminate Boundaries*, In *Geographic Objects with Indeterminate Boundaries*, Taylor & Francis, 1996, pp. 171–187.

[3] M. Erwig and M. Schneider. Vague Regions. *5th Int. Symp. on Advances in Spatial Databases*, pages 298–320. Springer-Verlag, 1997.

[4] M. Schneider and T. Behr. Topological Relationships between Complex Spatial Objects. *ACM Trans. on Database Systems (TODS)*, 31:39–81, 2006.

[5] A. Pauly and M. Schneider. Vague Spatial Data Types, Set Operations, and Predicates. *8th East-European Conf. on Advances in Databases and Information Systems*, pp. 379–392, 2004

[6] A. Pauly and M. Schneider. Topological Predicates between Vague Spatial Objects. *9th Int. Symp. on Spatial and Temporal Databases*, pp. 418–432, 2005.

[7] A. Pauly and M. Schneider. Topological Reasoning for Identifying a Complete Set of Topological Predicates between Vague Spatial Objects. *19th Florida Artificial Intelligence Research Society (FLAIRS) Conference*, pp. 731–736, 2006.