ROSA: An Algebra for
Rough Spatial Objects in Databases

Markus Schneider* & Alejandro Pauly

University of Florida
Department of Computer & Information Science & Engineering
Gainesville, FL 32611, USA
{mschneid, apauly}@cise.ufl.edu

Abstract. A fundamental data modeling problem in geographical information systems and spatial database systems refers to an appropriate treatment of the *vagueness* or *indeterminacy* features of spatial objects. Geographical applications often have to deal with spatial objects that cannot be adequately described by the determinate, crisp concepts exclusively available in these systems since these objects have an intrinsically indeterminate and vague nature. The goal of this paper is to show that rough set theory can be leveraged in an elegant manner to seamlessly model this kind of spatial data. Our approach introduces novel *rough spatial data types* for *rough points*, *rough lines*, and *rough regions* that can be employed as attribute types in database schemas. These data types are part of a data model called *ROSA* (*ROugh Spatial Algebra*). Their formal framework is based on already existing, general, exact models of crisp spatial data types, which simplifies the definition of the rough spatial model. In addition, we obtain *executable specifications* for the operations on rough spatial objects; these can be immediately used as implementations. This paper gives a formal definition of the three rough spatial data types as well as some basic operations.

Keywords. Spatial database; spatial vagueness; rough spatial data type

1 Introduction

Geographical information systems (GIS) and spatial database systems (SDBS) are currently confronted with two main data modeling problems. Beside an appropriate integration of the temporal aspect, the feature of *spatial vagueness* or *spatial indeterminacy* is inherent to many geometric and geographic data [1]. The current mapping of spatial phenomena of the real world to exclusively crisp, i.e., precisely determined, spatial objects has turned out to be an insufficient abstraction process for many geographic applications. So far, applications based on indeterminate spatial data cannot be supported by current GIS and SDBS.

These systems assume, often contrary to reality, that the positions of points, the locations and routes of lines, and the extent and hence the boundary of

regions are precisely determined and universally recognized. Examples are especially man-made spatial objects (e.g., monuments, highways, buildings) and immaterial spatial objects (e.g., countries, districts, land parcels with their political, administrative, and cadastral boundaries). We denote this kind of entities as *crisp* or *determinate spatial objects*.

However, to an increasing degree, there are many geometric applications in which positions of points are not exactly known, the locations and routes of lines are unclear, and regions do not have sharp boundaries, or their boundaries cannot be precisely determined. Examples are social or natural phenomena (e.g., terrorists' refuges and escape routes, population density, unemployment rate, soil quality, vegetation, oceans, oil fields, biotopes, deserts). We denote this kind of entities as *rough* or *indeterminate spatial objects*.

In GIS and SDBS[1], *spatial data types* (see [2] for a survey) like *point*, *line*, or *region* provide fundamental abstractions for modeling the structure of geometric entities, their relationships, properties, and operations. This paper presents an object model for defining *rough spatial data types* for *rough points*, *rough lines*, and *rough regions*. We use the term *rough* for the characterization of these types since we leverage concepts of *rough set theory* [3] as a formal framework. The types are part of a novel data model called *ROSA* (*Rough Spatial Algebra*). The model rests on "traditional" (i.e., exact) modeling techniques and extends, rather than replaces, the current theory of SDBS and GIS. Further, moving from an exact to a rough domain does not necessarily invalidate conventional (computational) geometry for executing spatial operations; it is merely an extension. Hence, current exact object models can be considered as special cases of our rough spatial object model. We show in this paper that all rough spatial data types and some main rough spatial operations can be defined generically, i.e., without type-specific definitions. Since our rough spatial data types and operations are based on their crisp counterparts and can be expressed by them, we obtain *executable specifications* that can be directly used as an implementation. In this paper, we do not aim at developing a type system with a "complete" set of operations and predicates. The goal is more to demonstrate the power, simplicity, and expressiveness of our rough spatial data model.

Section 2 discusses related work. Section 3 informally introduces the concept of rough spatial objects and motivates it by giving some application examples. Section 4 gives a generic definition of rough spatial data types and rough spatial operations. Finally, Section 5 draws some conclusions and addresses future work.

2 Related Work

The problem of spatial vagueness, indeterminacy, uncertainty and imprecision with their many different nuances and diversities has been a research topic in GIS (but not SDBS) for a long time. As a trend, GIS clearly advocate *fuzzy set theory* [4] as an appropriate formal framework for solving this problem. Despite

---

[1] It is important to understand that SDBS deal exclusively with vector data in contrast to image databases that exclusively handle raster data.

many concepts and ideas, unfortunately, no overall and satisfying solution has been found so far. In particular, the implementation of fuzzy concepts turns out to be rather difficult in the spatial domain.

We regard *rough set theory* [3] as an appropriate, alternative option and compromise for representing spatial vagueness. This approach is based on representing a set X by a pair of determinate lower and upper approximations. From the lower approximation, we know that its elements belong definitely to X. The upper approximation is the set of elements that possibly belong to X. In this paper, we demonstrate how this concept can be seamlessly transferred to a model of rough spatial objects. The use of rough sets for representing vague spatial data has been proposed before in a few publications [5–7]. These approaches have in common that they are based on image data but not on vector data.

A benefit of the fact that the lower and upper approximations of a set are crisp implies for the transferal to the spatial domain that we can leverage the existing definitions, techniques, data structures, and algorithms of exact spatial object models. As an example, in the past, we have developed the *ROSE* (*Robust Spatial Extension*) *Algebra* [8, 2], which provides crisp spatial data types like *point*, *line*, and *region* together with a comprehensive collection of spatial operations and predicates. Such an algebra (type system) can be taken to define our Rough Spatial Algebra ROSA and to obtain an executable specification of rough spatial operations.

3 What are Rough Spatial Objects?

As indicated before, our concept of rough spatial objects necessitates a general, underlying crisp spatial object model which incorporates the determinate spatial data types *point*, *line*, and *region*. These data types must be defined in a way so that they are closed under (appropriately defined) *geometric union*, *geometric intersection*, *geometric difference*, and *geometric complement* operations. Such crisp type systems have, e.g., been formally defined in [9, 10, 8, 2], and we will leverage them in this paper. Informally, these models represent a *point* object as a finite set of individual points, a *line* object as a finite set of disjoint *blocks* where each block represents a finite set of curves, and a *region* object as a finite set of disjoint, connected areal components called *faces* possibly with disjoint holes (see Figure 1). Examples of *point* objects are collections of lighthouses, collections of junctions, and collections of landmarks. Examples of *line* objects are streets, railways, and waterways. Examples of *regions* are districts, land parcels, and parks.

The central idea of *rough spatial objects* is to represent them by a lower approximation, which specifies those object parts that definitely belong to the rough spatial object, and an upper approximation, which, in addition, specifies those object parts that possibly partially or completely belong to the rough spatial object. As an illustrating example, we consider a homeland security scenario to introduce our concept for dealing with spatial vagueness and to demonstrate its usability. Secret services (should) have knowledge of the whereabouts of ter-
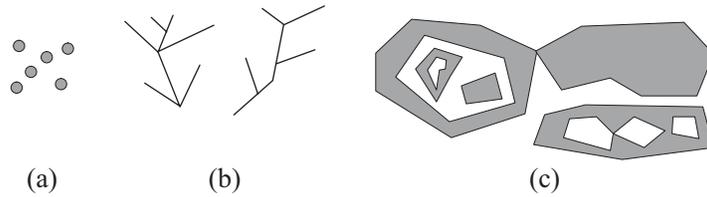
Fig. 1. Examples of a crisp point object (a), a crisp line object (b), and a crisp region object (c). Each collection of components forms a single crisp spatial object.

rorists. For each terrorist, some of their refuges are precisely known, some are not and only conjectures. We can model these locations as a *rough point* object where the precisely known locations are called the *lower point approximation* and all possible locations are called the *upper point approximation*. Secret services are also interested in the routes a terrorist takes to move from one refuge to another. These routes can be modeled as *rough line* objects. Some paths have definitely been identified as terrorist routes. They form the *lower line approximation* of the rough line object. All possible paths that terrorists can have taken yield its *upper line approximation*. Knowledge about areas of terroristic activities is also important for secret services. From some areas it is well known that a terrorist operates in them; we call them the *lower region approximation*. From other areas we can only assume that they are the target of terroristic activity. We summarize all possible areas in the *upper region approximation*. Figure 2 gives some examples. Grey shaded areas, straight lines, and grey points indicate lower approximations; areas with white interiors, dashed lines, and white points refer to vague parts and form the upper approximations together with the determinate parts.

Based on this scenario and taking into account spatial vagueness, we are able to pose interesting queries. We can ask for the locations where any two terrorists have taken the same refuge. We can determine those terrorists that operated in the same area. We can compute the locations where routes taken by different terrorists crossed each other. Many further queries are possible. Vague concepts offer a greater flexibility for modeling properties of spatial phenomena in the real world than determinate concepts do. Still, vague concepts comprise the modeling power of determinate concepts as a special case.
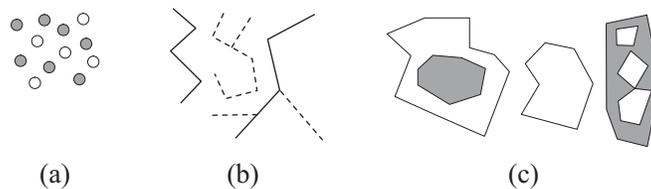

Fig. 2. Examples of a rough point object (a), a rough line object (b), and a rough region object (c). Each collection of components forms a single rough object.

In this sense, many scenarios can be found that could make meaningful use of the concept of rough spatial objects. They all have in common that a rough spatial object (e.g., a rough line) is described by a pair of two crisp spatial objects (e.g., two crisp lines) where the first object is topologically contained in or inside of the second object. The first crisp spatial object, called the *lower object approximation*, describes the determinate parts of the rough spatial object, i.e., the parts that definitely and always belongs to the rough object. The second crisp spatial object, called the *upper object approximation*, describes the possible parts of the rough spatial object, i.e., the parts that definitely or perhaps belong to the rough object.

4 A Generic Definition of Rough Spatial Data Types and Rough Spatial Set Operations

Based on the motivation in the previous section, in Section 4.1, we first introduce some needed concepts from crisp (i.e., determinate) spatial type systems. Afterwards, we give a formal definition of rough spatial data types and rough spatial set operations (Section 4.2). An interesting observation is that these definitions can be given in a generic manner, i.e., type-specific considerations are unnecessary.

4.1 Crisp Spatial Data Types and Crisp Spatial Set Operations

Since rough sets are based on the concept of sets, the first issue is how crisp spatial objects can be modeled as sets. In this paper, we are exclusively dealing with two-dimensional spatial objects in the plane. Hence, crisp spatial objects are formally defined as point sets and subsets of the Euclidean space $\mathbb{R}^2$ that have to satisfy certain topological constraints in order to be well defined for geographic applications. The definition of the crisp spatial data types *point*, *line*, and *region* [10] is based on point set theory and point set topology [11]. We obtain:

(i) $point = \{P \subset \mathbb{R}^2 \mid P \text{ is finite}\}$
(ii) $line = \{L \subset \mathbb{R}^2 \mid$ (a) $L = \bigcup_{i=1}^{n} f_i([0,1])$ with $n \in \mathbb{N}_0$
　　　　　　　　　　　(b) $\forall\, 1 \leq i \leq n : f_i : [0,1] \to \mathbb{R}^2$ is a continuous mapping
　　　　　　　　　　　(c) $\forall\, 1 \leq i \leq n : |f_i([0,1])| > 1\}$
(iii) $region = \{R \subset \mathbb{R}^2 \mid$ (a) R is regular closed
　　　　　　　　　　　(b) R is bounded
　　　　　　　　　　　(c) The number of connected sets of R is finite$\}$

A set is *regular closed* if it is equal to the closure of its interior [11]. The above definition gives an *unstructured* view of crisp spatial objects. Another equivalent definition leads to a *structured* view [10], which we only describe informally here. The unstructured and structured views for the data type *point* coincide. A *line* object is assembled from a finite number of connected components called *blocks*;

each block contains a finite number of curves. A *region* object consists of a finite number of disjoint *faces*; each face possibly contains a finite number of disjoint holes. An illustration of these informal descriptions is shown in Figure 1.

For $\alpha \in \{point, line, region\}$, each type α is closed under the *geometric set operations union* ($\oplus : \alpha \times \alpha \to \alpha$), *intersection* ($\otimes : \alpha \times \alpha \to \alpha$), and *difference* ($\ominus : \alpha \times \alpha \to \alpha$) [10]. The partial order $(\alpha, \subseteq)$ is a distributive lattice $(\alpha, \oplus, \otimes)$. But it is not a complemented lattice, and consequently not a Boolean algebra, since the spatial data types are not closed under the operation *complement* ($\sim$). That is, $\forall v \in \alpha : \sim v \notin \alpha$. The identity of $\otimes$ is denoted by $\mathbf{1}$, which corresponds to $\mathbb{R}^2$. The identity of $\oplus$ is presented by $\mathbf{0}$, which corresponds to the empty spatial object (empty point set). The geometric set operations of the type *point* are equal to the standard set operations. The operation $\oplus$ is equal to the set operation $\cup$ for the types *line* and *region*. The geometric operations $\otimes$ and $\ominus$ require a *regularization* step so that they cannot produce geometric anomalies.

4.2 Rough Spatial Data Types and Rough Spatial Set Operations

Syntactically, the extension of a crisp spatial data type to a corresponding rough spatial type is given by a type constructor ρ as follows:

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

That is, each rough spatial data type is represented as a pair of corresponding crisp spatial data types. For example, for $\alpha = point$ we obtain $\rho(point) = point \times point$, which we also name *rpoint*. Accordingly, the data types *rline* and *rregion* are defined. For a rough spatial object $R = (\underline{R}, \overline{R}) \in \rho(\alpha)$, we call $\underline{R} \in \alpha$ the *lower object approximation* of R, and $\overline{R} \in \alpha$ denotes the *upper object approximation* of R.

Semantically, the lower (minimal, guaranteed) object approximation represents the determinate, crisp part of R, i.e., the area which *definitely* belongs to R. The upper (maximally possible, speculative) object approximation describes the *potential* spatial extent of R. Hence, we know that $\mathbf{1} - \overline{R} = \mathbb{R}^2 - \overline{R}$ does *definitely not* belong to R. Therefore, the area of *spatial determinacy* is $\underline{R} \cup (\mathbb{R}^2 - \overline{R})$. The *area of spatial vagueness* extends over $\overline{R} - \underline{R}$. It represents the indeterminate, vague part of R, i.e., the area for which we cannot say with any certainty whether it or parts of it belong to R or not. *Maybe* it or parts of it belong to R, *maybe* this is not the case. We could also say that this is *unknown* or *unclear*. Note that, in general, $\overline{R} - \underline{R} \notin \alpha$ and $\overline{R} \ominus \underline{R} \in \alpha$ hold. To enable the intended semantics described above, we require:

$$\forall\, \alpha \in \{point, line, region\}\ \forall\, R = (\underline{R}, \overline{R}) \in \rho(\alpha) : \underline{R} \subseteq \overline{R}$$

In the spatial domain, subset relationships are expressed by so-called *topological relationships* [10], which characterize the relative position of spatial objects to each other. Examples of such relationships are *disjoint*, *meet*, and *overlap*. As [10] shows, the subset relationship depends on the combination $\alpha \times \alpha$ of spatial data types considered and corresponds to several topological relationships. In

an abbreviated form, on the basis of so-called *clustered* topological relationships (indicated by a subscript 'c'), the above requirement can be rewritten as:

$$\forall\, \alpha \in \{point, line, region\}\ \forall\, R = (\underline{R}, \overline{R}) \in \rho(\alpha):$$
$$equal_c(\underline{R}, \overline{R})\ \lor\ inside_c(\underline{R}, \overline{R})\ \lor\ coveredBy_c(\underline{R}, \overline{R})$$

For $\alpha = point$, we obtain three (unclustered) topological relationships, for $\alpha = line$ ten relationships, and for $\alpha = region$ five relationships.

Let $points: \rho(\alpha) \to \mathbb{R}^2$ be an auxiliary function that yields the (unknown) point set of a rough spatial object $R = (\underline{R}, \overline{R}) \in \rho(\alpha)$. We can conclude that

$$\underline{R} \subseteq points(R) \subseteq \overline{R}$$

If $\underline{R} = \overline{R}$, R is either the empty rough spatial object $\mathbf{0} = (\varnothing, \varnothing)$ or corresponds to the crisp spatial object $\underline{R}$. The 1-element is $\mathbf{1} = (\mathbb{R}^2, \mathbb{R}^2)$ and corresponds to the Euclidean plane. Even if we do not know the exact point set of R, we assume and require that $points(R)$ is not arbitrary but compatible to α, i.e.,

$$points(R) \in \alpha \quad \text{and} \quad points(R) \ominus \underline{R} \in \alpha$$

Using the characteristic function χ deciding about the existence or non-existence of an element in a set, we obtain $\chi(p) = 1$ for all $p \in \underline{R}$, $\chi(p) = 0$ for all $p \in \mathbb{R}^2 - \overline{R}$, $\chi(p) = 1\ \lor\ \chi(p) = 0$ for all $p \in \overline{R} - \underline{R}$, and $\chi(p) = 1$ for all $p \in points(R) \in \alpha$. Note the deliberate use of set-theoretic operations. In particular, possible common boundary points of $\underline{R}$ and $\overline{R} \ominus \underline{R}$ are mapped to 1.

Let $R = (\underline{R}, \overline{R}), S = (\underline{S}, \overline{S}) \in \rho(\alpha)$. We then define the geometric set operations **union**, **intersection**, and **difference** on rough spatial objects as follows:

$$R \text{ \textbf{union} } S = (\underline{R} \oplus \underline{S}, \overline{R} \oplus \overline{S})$$
$$R \text{ \textbf{intersection} } S = (\underline{R} \otimes \underline{S}, \overline{R} \otimes \overline{S})$$
$$R \text{ \textbf{difference} } S = (\underline{R} \ominus \overline{S}, \overline{R} \ominus \underline{S})$$

The equalities and subset relationships used in the following are derived from [3]. For the operation **union**, we define the lower object approximation in a more pessimistic and stricter way since $\underline{R} \oplus \underline{S} \subseteq \overline{R \oplus S}$. For the upper object approximation, there is no difference since $\overline{R} \oplus \overline{S} = \overline{R \oplus S}$. For the operation **intersection**, we are rather optimistic for the upper object approximation since $\overline{R} \otimes \overline{S} \supseteq \overline{R \otimes S}$. For the lower object approximation, $\underline{R} \otimes \underline{S} = \underline{R \otimes S}$ holds. For the operation **difference**, for the lower object approximation, we must subtract both the intersecting determinate and indeterminate parts of S from $\underline{R}$. Only the result of this computation can be definitely part of the difference of R and S. For computing the upper object approximation, we only have to subtract the definite parts of S from $\overline{R}$ since they can definitely not belong to R. The particular reason why all operations make use of $\underline{R}$, $\overline{R}$, $\underline{S}$, and $\overline{S}$ is that these approximations are known and can be used in an implementation.

5   Conclusions

In this paper, we have made a first step towards a simple but expressive data model of points, lines, and regions that is capable of describing many different aspects of spatial vagueness. It is based on rough set theory and a canonical extension of determinate spatial data models. This facilitates the treatment of rough and exact objects in one single model. Since our approach is based on exact spatial modeling concepts, it allows us to build upon existing work and simplifies many definitions. In particular, we can leverage already existing implementations of crisp spatial type systems (like the ROSE Algebra) to realize rough spatial objects with only minimal effort by executable specifications.

The data model presented in this paper is part of ROSA, our Rough Spatial Algebra. So far, ROSA is incomplete. We plan to supplement it by further spatial operations, topological predicates, directional predicates, numerical operations, and more. A later step refers to the implementation of ROSA on the basis of our ROSE Algebra and to the embedding of ROSA into a database query language.

References

1. Burrough, P.A., Frank, A.U., eds.: Geographic Objects with Indeterminate Boundaries. GISDATA Series, vol. 2. Taylor & Francis (1996)
2. Schneider, M.: Spatial Data Types for Database Systems - Finite Resolution Geometry for Geographic Information Systems. Volume LNCS 1288. Springer-Verlag (1997)
3. Pawlak, Z.: Rough Sets. Int. Journal of Computer and Information Sciences **11**(5) (1982) 341–356
4. Zadeh, L.A.: Fuzzy Sets. Information and Control **8** (1965) 338–353
5. Ahlqvist, O., Keukelaar, J., Oukbir, K.: Rough Classification and Accuracy Assessment. Int. Journal of Geographical Information Sciences **14**(5) (2004) 475–496
6. Beaubouef, T., Ladner, R., Petry, F.: Rough Set Spatial Data Modeling for Data Mining. Int. Journal of Intelligent Systems **19** (2004) 567–584
7. Worboys, M.: Computation with Imprecise Geospatial Data. Computational, Environmental and Urban Systems **22**(2) (1998) 85–106
8. Güting, R.H., Schneider, M.: Realm-Based Spatial Data Types: The ROSE Algebra. VLDB Journal **4** (1995) 100–143
9. Clementini, E., Di Felice, P.: A Model for Representing Topological Relationships between Complex Geometric Features in Spatial Databases. Information Systems **90**(1-4) (1996) 121–136
10. Schneider, M., Behr, T.: Topological Relationships between Complex Spatial Objects. ACM Trans. on Database Systems **31**(1) (2006) 39–81
11. Gaal, S.: Point Set Topology. Academic Press (1964)