A Qualitative Notion of Spatial Proximity

Hans W. Guesgen
Department of Computer Science
University of Auckland
Private Bag 92019
Auckland, New Zealand
hans@cs.auckland.ac.nz

Jochen Albrecht
Department of Geography
University of Auckland
Private Bag 92019
Auckland, New Zealand
j.albrecht@auckland.ac.nz

Abstract

The concept of space underlying geographic information systems is basically Euclidean, requiring all subjects to adhere to the same view of space. This makes most attempts to deal with imprecise or uncertain geographic information difficult or sometimes even impossible. In this paper, we describe a way of incorporating imprecise qualitative spatial reasoning with quantitative reasoning in geographic information systems that is not restricted to Euclidean geometry. The idea is to use fuzzy sets to model qualitative spatial relations among objects, like *The downtown shopping mall is close to the harbor*. The membership function of such a fuzzy set defines a fuzzy distance operator, for which a new algorithm is introduced in this paper.

Introduction

Although geographic information systems (GIS) have been around for quite a while (Coppock & Rhind 1991), there has been little change in the functionality of the systems. In spite of their name, geographic information systems have so far been mostly *geometric* in nature, ignoring the thematic and temporal dimensions of geographic features (Molenaar 1996; Sinton 1978; Usery 1996). Various attempts to overcome these limitations are documented in a number of disciplines. (Frank 1992; Goodchild 1992; Gupta, Weymouth, & Jain 1991; Herring 1991; 1992; Raper & Maguire 1992) deal with extensions of the data model, while Allen's work forms the basis for numerous temporal logic endeavors that deal with dynamic aspects of geographic information (Egenhofer & Golledge 1997; Frank 1994; Peuquet 1994). Applications of fuzzy techniques are most commonly found in remote sensing literature but (Altmann 1994; Brimicombe 1997; Molenaar 1996; Plewe 1997) provide examples that the inherent fuzziness of geographic features becomes increasingly acknowledged in geographic information science as well.

The way in which GIS perform spatial reasoning, i.e., the extraction of new information from stored spatial data, has been quantitative in nature. On the other hand, humans often prefer a qualitative analysis over a quantitative one, as this is more adequate in many cases from the cognitive point of view (Clementini, Di Felice, & Hernández 1997). In the following, we will look at a new way of expressing spatial proximity in a qualitative way.

The notion of qualitative reasoning stems originally from artificial intelligence and has been applied to spatial phenomena for about a decade (Freksa 1990; Guesgen 1989; Hernández 1991; Mukerjee & Joe 1990). It has become well established with conference series such as COSIT and the European research initiative SPACENET. The notion of proximity as introduced in this paper builds on well-established concepts from qualitative spatial reasoning, in particular the application of fuzzy sets. In the following, we will briefly review these concepts.

Quantitative vs. Qualitative Reasoning

We assume the reader to be familiar with the traditional form of spatial reasoning as it underlies every GIS-based analysis. In essence, it uses map overlay to compute new maps from existing ones. For example, if we want to find a suitable location for a new city dump given certain constraining factors like *The dump must be at least 500 meters from any water*, we would select all areas of water, buffer these areas by 500 meters, and overlay the resulting map with maps corresponding to other constraints.

This form of quantitative spatial reasoning delivers precise results, but is often too rigid and therefore not applicable to scenarios like the city dump scenario. The reason is that quantitative statements like *All locations that are more than 500 meters from water* may eventually result in an empty map, as they restrict the search space too dramatically by excluding any areas, for example, which are 490 meters from water. Such an area, however, might be the best choice available and therefore perfectly acceptable.

This problem can be solved by using qualitative spatial statements rather than quantitative ones. Instead of *All locations that are more than 500 meters from water*, we would employ the restriction *All locations that are far from water*. The system would then analyze the qualitative relation *far from* that is used in this restriction and would find the areas that best match this restriction and that are compatible with the other restrictions.

To achieve this goal, we interpret spatial relations among objects as restrictions on linguistic variables which represent spatial information about the objects.[1] Consider, for example, the position x_A of some object A in the city. A qualitative approach would specify x_A in terms of qualitative values like *near church*, *at harbor*, *downtown*, etc. This approach can be translated directly into an approach using linguistic variables.

Informally, a linguistic variable is a variable whose values are words or phrases in a natural or artificial language. The values of a linguistic variable are called linguistic values. For example, the position of A can be represented by a linguistic variable x_A whose linguistic values are from the domain $L(x_A) = \{downtown, near\ church, at\ harbor, \ldots\}$. To express spatial information, we introduce restrictions on the values of the linguistic variables that represent these relations. For example, if A is either downtown or at the harbor, we restrict the value of x_A to $R(x_A) = \{downtown, at\ harbor\}$.

Spatial relations between objects can be represented by restrictions on composite linguistic variables. For example, the spatial relation between two objects A and B can be represented by introducing a binary composite variable (x_A, x_B), the values of which are from the domain $L(x_A, x_B) = L(x_A) \times L(x_B)$, and a restriction $R(x_A, x_B) \subseteq L(x_A) \times L(x_B)$ on the values of (x_A, x_B). In other terms, a spatial relation is a relation on linguistic variables representing spatial information.

Linguistic variables provide us with a convenient means to express qualitative spatial relations. However, they alone aren't sufficient to integrate qualitative and quantitative spatial reasoning. Only when combined with fuzzy sets, they allow us to add quantitative aspects to the qualitative ones. The next section will discuss this issue.

Fuzzy Sets

Fuzzy spatial reasoning is a method for handling different types of uncertainty inherent in almost all spatial data. Most often, it is employed for dealing with classification errors (Chrisman 1991; Goodchild & Gopal 1989) and the imprecision of boundaries (Leung 1987; Plewe 1997), although as early as 1985, (Robinson, Thongs, & Blaze 1985) introduced a representation language based on fuzzy logic to process natural language queries on geographic data. A fuzzy subset $\tilde{R}$ of a domain D is a set of ordered pairs, $\langle d, \mu_{\tilde{R}}(d)\rangle$, where $d \in D$ and $\mu_{\tilde{R}} : D \rightarrow [0,1]$ is the membership function of $\tilde{R}$. The membership function replaces the characteristic function of a classical subset $R \subseteq D$, which maps the set D to $\{0,1\}$ and thereby indicating whether an element belongs to R (indicated by 1) or not (indicated by 0). If the range of $\mu_{\tilde{R}}$ is $\{0,1\}$, $\tilde{R}$ is nonfuzzy and $\mu_{\tilde{R}}(d)$ is identical with the characteristic function of a nonfuzzy set.

Fuzzy sets can be used to associate quantitative information with qualitative one. Consider, for example, a linguistic value like *downtown*. We can associate this qualitative value with a fuzzy set that characterizes for each coordinate on some given street map to which extend this coordinate represents some location downtown. Assuming that D represents the possible coordinate (usually a set of character-digit combinations), *downtown* may be represented by a fuzzy set such as the following:

$$\tilde{R} = \{\langle M5, 1\rangle, \langle M4, 0.8\rangle, \langle L5, 0.8\rangle, \ldots \langle L4, 0.7\rangle, \ldots\}$$

In other words, each location on the city map is considered to be more or less downtown. If its membership value equals 1, the location is definitely downtown. If it equals 0, then it isn't downtown at all. If it doesn't cause any confusion, we denote a fuzzy set as follows:

$$\tilde{R} = \sum_{d \in D} \langle d, \mu_{\tilde{R}}(d)\rangle$$

In general, the fuzzy set corresponding to a spatial linguistic value may be a continuous rather than a countable or even finite set. For example, the spatial linguistic value *illuminated*, which specifies that an object is near some light source, may be associated with a fuzzy set $\tilde{R}$ in the domain of real numbers, $\mathbb{R}$. An element $d \in \mathbb{R}$ then indicates the distance of the object to the light source. If the distance is 0, then the object is definitely considered to be illuminated. The greater (the square of) the distance to the light source, the less we consider the object to be illuminated. Since $\tilde{R}$ is a continuous set, we denote it as follows, assuming that $\mu_{\tilde{R}}(d) = 1/(1 + d^2)$:

$$\tilde{R} = \int_0^\infty \left\langle d, \frac{1}{1 + d^2} \right\rangle$$

Using the inference rules of fuzzy set theory, we are now able to deal with a new type of GIS query. Instead of producing map overlays in the traditional way, we compute the intersection, union, and negation of fuzzy sets. This allows queries like the following:

- Name two neighboring places, at least one of which offers good job opportunities while the other one has superb recreational properties.
- Name combinations of work and residence locations offer good shopping opportunities on an acceptable commuting route.

Distance and Proximity Measures

There are numerous studies over a wide range of different data domains, including geographical space, of

[1]This step is just a shift in terminology rather than an introduction to a new reasoning method. The motivation for this step is the attempt to stay within the terminology used in fuzzy set theory.

how humans make subjective judgements regarding distances. According to (Gahegan 1995), the human perception of closeness or proximity is influenced by the following:

1. In the absence of other objects, humans reason about proximity in a geometric fashion. Furthermore, the relationship between distance and proximity can be approximated by a simple linear relationship.
2. When other objects of the same type are introduced, proximity is judged in part by relative distance, i.e., the distance between a primary object and a reference object.
3. Distance is affected by the size of the area being considered, i.e., the frame of reference.

Proximity measures in spatial reasoning must behave in a way that follows the human perception of proximity. Otherwise, the result of the GIS is counterintuitive and therefore unreliable.

The most intuitive form of reasoning about proximity is based on the absolute (or physical) distance between objects. Absolute distance is the major factor that affects proximity. It is commonly defined by a symmetric Euclidean distance matrix, in which an entry $\delta(A, B)$ specifies the distance between an object A with coordinates (x_A, y_A) and B with coordinates (x_B, y_B):

$$\delta(A, B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$

Euclidean distance can be used to calculate a degree of proximity between any primary object C with coordinates (x_C, y_C) and a given reference object A with coordinates (x_A, y_A). This degree can then be used to define a fuzzy set $\tilde{P}_A$ of primary objects that are in the proximity of A:[2]

$$\tilde{P}_A = \int_{x_C, y_C \in \mathbb{R}} \left\langle (x_C, y_C), \frac{1}{1 + \delta(A, C)^2} \right\rangle$$

In (Guesgen & Poon 1997), it is shown how this approach can be extended to queries referring to a class of reference objects rather than a particular object, like *a place close to a waste dump* as opposed to *a place close to the town hall.* This type of query is commonly found in applications like resource planning and allocation.

Proximity without Euclidean Distance

So far we have assumed that the Euclidean distance is used as the basis for calculating fuzzy membership grades. This assumption is unnecessary, as any other distance measure can serve the same purpose. More than that, we can define proximity without any distance measure at all by using the notion of fuzzy sets. Given a set of primary objects $\{B_1, B_2, \ldots, B_n\}$ and a reference object A, the proximity of A is a fuzzy set $\tilde{P}_A$ over $\{B_1, B_2, \ldots, B_n\}$:

$$\tilde{P}_A = \sum_{i \in \{1,\ldots,n\}} \langle B_i, \mu_{\tilde{P}_A}(B_i) \rangle$$

For example, given a reference object A and a set of primary objects $\{B, C, D, E, F, G, H\}$, then the proximity of A may be defined as follows:

$$\tilde{P}_A = \langle G, 1\rangle + \langle B, 0.8\rangle + \langle E, 0.8\rangle + \langle C, 0.6\rangle + \langle F, 0.4\rangle + \langle D, 0.4\rangle + \langle H, 0.4\rangle$$

$\tilde{P}_A$ defines a set of neighborhoods, each neighborhood containing the primary objects that have a fuzzy membership grade of at least α, where $\alpha \in \{0.4, 0.6, 0.8, 1\}$.[3] Figure 1 shows the neighborhoods for the example graphically.

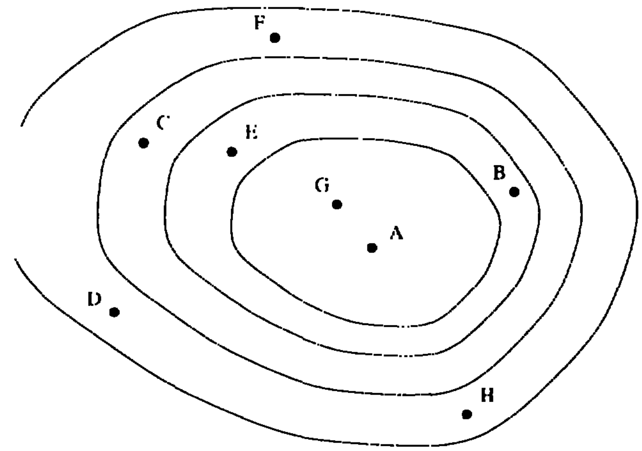

Figure 1: Proximity illustrated as set of neighborhoods.

Since membership grades are ranging over real numbers (in the interal [0,1]), we can perform a simple form of reasoning about the proximity of objects without additional requirements:

If

B is closer to A than C and
C is closer to A than D

then

$$\mu_{\tilde{P}_A}(B) < \mu_{\tilde{P}_A}(C) \text{ and } \mu_{\tilde{P}_A}(C) < \mu_{\tilde{P}_A}(D)$$

which implies

$$\mu_{\tilde{P}_A}(B) < \mu_{\tilde{P}_A}(D)$$

which means that

B is closer to A than D

This form of reasoning stays within the scope of a single reference object. Even more useful, however, might be reasoning involving several reference objects. In this

[2]The membership function used here is very similar to the nearness function introduced by Worboys (Worboys 1996).

[3]In the fuzzy set literature, these neighborhoods are usually called α-level sets.

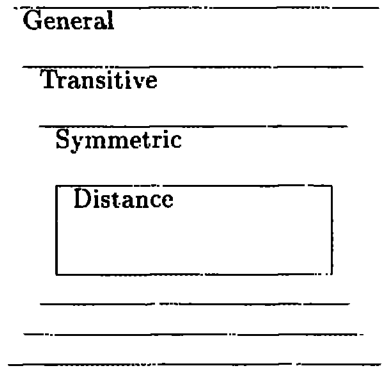

Figure 2: Classes of proximities.

case, we might want to ensure that the reasoning process does not lead to contradictions. In particular, we might want to require that, for any three objects A, B, and C, the proximities of these objects are consistent with each other. One way to guarantee such consistency is by imposing the following transitivity rule on any objects A, B, and C:

If

$$\mu_{\tilde{P}_A}(B) < \mu_{\tilde{P}_A}(C) \text{ and } \mu_{\tilde{P}_B}(C) < \mu_{\tilde{P}_B}(A)$$

then

$$\mu_{\tilde{P}_C}(B) < \mu_{\tilde{P}_C}(A)$$

If the proximities are defined by using a distance measure, then this rule holds for any three objects A, B, and C. In other cases, the transitivity rule might not be applicable in general. For example, determining the distance of a location on the basis of what it costs to make a long-distance phone call to that location might differ from reference point to reference point, as different countries and phone companies have different rates.

A sufficient (but not necessary) condition for the applicability of the transitivity rule is the symmetry of proximity:

$$\text{For all objects } A, B\text{: } \mu_{\tilde{P}_A}(B) = \mu_{\tilde{P}_B}(A)$$

If proximity is symmetric, we can easily show that the right-hand side of the transitivity rule holds whenever the left-hand side is true:

$$\mu_{\tilde{P}_C}(B) = \mu_{\tilde{P}_B}(C) < \mu_{\tilde{P}_B}(A) = \\ \mu_{\tilde{P}_A}(B) < \mu_{\tilde{P}_A}(C) = \mu_{\tilde{P}_C}(A)$$

Figure 2 summarizes how proximities might be related with each other by introducing a hierarchy of classes. The largest class in this hierarchy represents the most general case, in which we cannot make any assumptions about the relationships among proximities. The next class requires the transitivity rule to be applicable. This can be achieved by using symmetric proximities, which leads us to an even tighter class in the hierarchy. Within this class, we can find the proximities that are defined on the basis of a distance measure.

Summary

Qualitative reasoning is reasoning in terms of linguistic values, whereas quantitative reasoning is reasoning based on numerical values such as measurements. Both qualitative and quantitative reasoning are used by humans to deduce new information from given one. However, it is believed that there is a preference towards qualitative reasoning, and that often some sort of translation takes place from quantitative to qualitative information.

In this paper, we introduced a scheme of representing qualitative spatial information by associating qualitative relations with fuzzy sets. We extended the concept of absolute distance to non-metric notions of proximity. We proved that qualitative spatial reasoning is possible in any form of space, and that symmetry of proximity is a sufficient condition for reasoning with several reference objects to be consistent.

Acknowledgments

This paper is based on research work previously published by Jonathan Histed, Ute Lörch, David Poon, and the authors. We are grateful for the support that we received from the University of Auckland Research Committee under various grants (most recently under grant XXXXX/9343/3414084).

References

Altmann, D. 1994. Fuzzy set-theoretic approaches for handling imprecision in spatial analysis. *International Journal of Geographical Information Systems* 8(3):271–289.

Brimicombe, A. 1997. A universal translator of linguistic hedges for the handling of uncertainty and fitness-of-use in GIS. In Kemp, Z., ed., *Innovations in GIS 4*. London, England: Taylor and Francis. 115–126.

Chrisman, N. 1991. The error component of spatial data. In et al., M., ed., *Geographical Information Systems: Principles and Applications*, volume 1. Harlow, England: Longman. 165–174.

Clementini, E.; Di Felice, P.; and Hernández, D. 1997. Qualitative representation of positional information. *Artificial Intelligence* 95(2):317–356.

Coppock, J., and Rhind, D. 1991. The history of GIS. In Maguire, D.; Goodchild, M.; and Rhind, D., eds., *Geographical Information Systems: Principles and Applications*. Essex, England: Longman Scientific & Technical. 21–43.

Egenhofer, M., and Golledge, R. 1997. *Spatial and Temporal Reasoning in Geographic Information Systems*. New York: Oxford University Press.

Frank, A. 1992. Spatial concepts, geometric data models, and geometric data structures. *Computers and Geosciences* 18:409–417.

Frank, A. 1994. Qualitative temporal reasoning in GIS-ordered time scales. In *Proc. SDH-94*, 410–431.

Freksa, C. 1990. Qualitative spatial reasoning. In *Proc. Workshop RAUM*, 21–36.

Gahegan, M. 1995. Proximity operators for qualitative spatial reasoning. In Frank, A., and Kuhn, W., eds., *Spatial Information Theory: A Theoretical Basis for GIS*, Lecture Notes in Computer Science 988. Berlin, Germany: Springer.

Goodchild, M., and Gopal, S., eds. 1989. *Accuracy of Spatial Databases*. London, England: Taylor and Francis.

Goodchild, M. 1992. Geographic data modeling. *Computers and Geosciences* 18:401–408.

Guesgen, H., and Poon, D. 1997. Computing proximity in geographic information systems. In *Proc. IJCAI-97 Workshop on Spatial and Temporal Reasoning*.

Guesgen, H. 1989. Spatial reasoning based on Allen's temporal logic. Technical Report TR-89-049, ICSI, Berkeley, California.

Gupta, A.; Weymouth, T.; and Jain, R. 1991. An extended object-oriented data model for large image bases. In Günther, O., and Schek, H.-J., eds., *Advances in Spatial Databases*, Lecture Notes in Computer Science 525. Berlin, Germany: Springer. 45–61.

Hernández, D. 1991. Relative representation of spatial knowledge: The 2-D case. In Mark, D., and Frank, A., eds., *Cognitive and Linguistic Aspects of Geographic Space*. Dordrecht, The Netherlands: Kluwer. 373–385.

Herring, J. 1991. The mathematical modeling of spatial and non-spatial information in geographic information systems. In Mark, D., and Frank, A., eds., *Cognitive and Linguistic Aspects of Geographic Space*. Dordrecht, The Netherlands: Kluwer. 313–350.

Herring, J. 1992. TIGRIS: A data model for an object-oriented geographic information system. *Computers and Geosciences* 18:443–452.

Leung, Y. 1987. On the imprecision of boundaries. *Geographical Analysis* 19(1):125–151.

Molenaar, M. 1996. The extensional uncertainty of spatial objects. In *Proc. SDH-96*, 9B.1–13.

Mukerjee, A., and Joe, G. 1990. A qualitative model for space. In *Proc. AAAI-90*, 721–727.

Peuquet, D. 1994. It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of American Geographers* 84(3):441–461.

Plewe, B. 1997. A representation-oriented taxonomy of gradation. In Hirtle, S., and Frank, A., eds., *Spatial Information Theory: A Theoretical Basis for GIS*. Lecture Notes in Computer Science 1329. New York: Springer. 121–136.

Raper, J., and Maguire, D. 1992. Design models and functionality in GIS. *Computers and Geosciences* 18:387-394.

Robinson, V.; Thongs, D.; and Blaze, M. 1985. Machine acquisition and representation of natural language concepts for geographic information retrieval. In *Proc. Pittsburgh Modeling and Simulation Conference*.

Sinton, E. 1978. The inherent structure of information as a constraint to analysis: Mapped thematic data as a case study. In Dutton, G., ed., *Harvard Papers on Geographic Information Systems*, volume 6. Reading, Massachusetts: Addison-Wesley.

Usery, L. 1996. A conceptual framework and fuzzy set implementation for geographic features. In Burrough, P., and Frank, A., eds., *Geographic Objects with Indeterminate Boundaries*. GISDATA Series No. 2. London, England: Taylor and Francis. 71–85.

Worboys, M. 1996. Metrics and topologies for geographic space. In *Proc. SDH-96*, 7A.1–11.