Handling Imprecise Data in Geographic Databases

D

Cyril de Runz
LIASD, University of Paris 8, France & CReSTIC, University of Reims, France

Herman Akdag
LIASD, University of Paris 8, France

Asma Zoghlami
LIASD, University of Paris 8, France

INTRODUCTION

The representation of imperfect data and its exploitation in Information Systems is a major theme of the artificial intelligence domain (Sapir, Shmilovici, & Rokach, 2008; Yang, We, Yu, & Yang, 2009). Several studies focused on proposing new conceptual data modeling approaches to store imperfect information (Ma & Yan, 2010).

For the geographic information, several approaches were proposed into the literature. Perceptory is a well-known approach that extends, by adding pictograms called PictograF, UML (the standard for modeling Information System) in order to design and build Geographic Information Systems (GIS).

In this article, we introduce an approach in order to design and implement a fuzzy spatiotemporal database storing imprecise geographic data. It is based upon the F-Perceptory approach, presented in (Zoghlami, 2013) which is an extension to handle fuzziness in the Perceptory modeling method (Bédard, Larrivée, Proulx, & Nadeau, 2004).

BACKGROUND

Imprecision in Geographic Data

Even though the use of GIS is common, the consideration of imprecision from the design of the information system to the data exploitation, is still a current issue and is the subject of this article. There is imprecision whenever the exact value of the truth status of a proposition of interest is not established uniquely, i.e., whenever its truth status is equivocal (Smets, 1995).

Indeed, by observing and modeling reality, the building of each dataset includes some imperfections. The data integration also produces other imperfections. In spatial science, a principal issue is how to deal with boundaries: it is hard to precisely and accurately delineate frontiers (Fisher, 1999).

Therefore, handling spatially imperfect data is essential. There are various methods for dealing with imperfection. The main mathematical theories are: probability theory; fuzzy set and possibility theory (Zadeh, 1965); rough set theory; theory of evidence.

According to the literature (Fisher, 1999), the fuzzy set theory is a good choice for dealing with imprecision.

Fuzzy Set Theory: Main Principles

Imprecision should be considered in the modeling of the information. As the sorites paradox makes it evident that probabilities are not adapted to imprecision, Zadeh (1965) introduced the fuzzy set theory. Indeed, the fuzzy set theory defines the notion of partial and valued membership of a value to a class. A fuzzy set A is characterized by a membership μA function taking values in [0, 1]. For each domain value x, a membership degree $\mu A(x)$, defined in [0, 1], is proposed. Therefore, concepts like young, old, etc. may be easily modeled by fuzzy sets.

An α-cut $A\alpha$, for all $\alpha > 0$, is the set of the domain values (the set of x) having a membership degree higher or equal to α ($\mu A(x) \geq \alpha$). By convention, A0 is the set of x such as $\mu A(x) > 0$.

DOI: 10.4018/978-1-4666-5888-2.ch172

A fuzzy set A is connected if, and only if, for all α in [0; 1] Aα is connected. Aα is connected if for all nonempty sets B and C, such as Aα, is their union, and there exists at least one point of B adhering to C or one point of C adhering to B. On R, Aα is connected if, and only if, it is an interval. In other words, a fuzzy set A is connected if, and only if, for all α in [0; 1], Aα is not composed with separate sets.

The use of connected α-cuts allows us to store different values of the imprecise data in the form of a multivalued set. Their use enables to draw the boundaries between a very low confidence membership (the 0-cut), a rather low confidence membership, a moderately low confidence membership, a low confidence membership, etc., which may also be interpreted as a range of values between almost impossible and very possible (Figure 1).

Modelling (Geographical) Information Systems

Among existing tools for modeling Information Systems, the Unified Modeling Language (UML) (OMG, 2000) is considered as a standard. The relational model for a database can be obtained from UML (Elmasri & Navathe, 2011).

However, the data represented in UML models were far from reflecting the real-world situations due to uncertainty, imprecision, etc. To respond to this new requirement, an extension called fuzzy UML based on the fuzzy sets has been introduced in (Ma & Yan, 2010) in order to enable the conceptual modeling of imprecise data. Thus, different levels of imprecision have been mainly introduced into the UML class model. However, fuzzy UML does not consider imprecise spatial and temporal data. Therefore, it cannot be used for Geographic Information Systems that are distinguished by the fact that the data is considered either according to its spatial or temporal component or according to its spatiotemporal component.

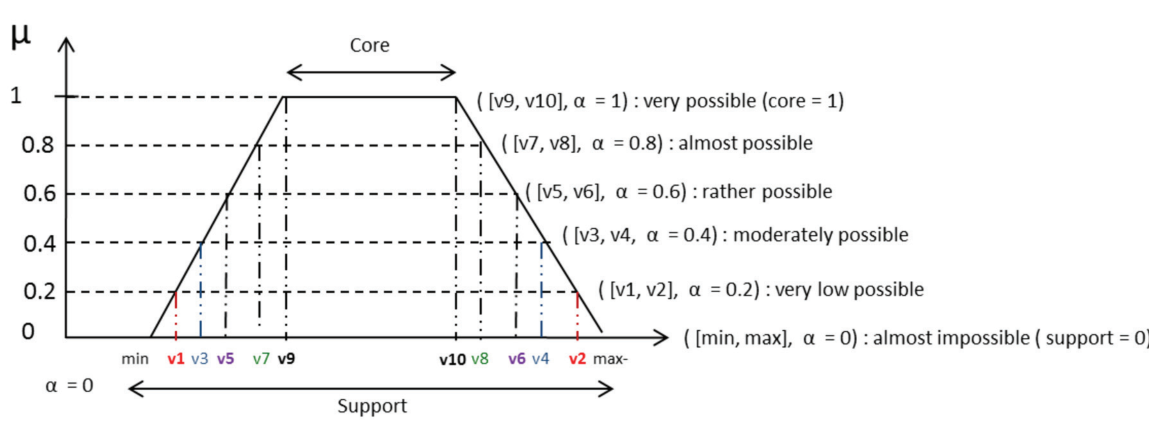
The geographic information field proposed several methods for the design of spatiotemporal information systems. Some of these methods result from the adaptation of non-specific methods by the spatialization and the temporalization of conceptual models, like the Perceptory modeling method which is a spatiotemporal extension of the UML data model method (Bédard, Larrivée, Proulx, & Nadeau, 2004). The other methods have their own tools for the design of Geographic Information Systems. The most known among them are MADS (Parent et al., 1997), CONGOO (Pantazis, 1994).

MODELING IMPRECISE INFORMATION

Issues

We distinguish two types of data imprecision. The first type corresponds to the fuzzy object. A fuzzy object corresponds to an object for object we have a core (restrictive) definition and a set of soft (less restrictive) definitions, including the more restrictive definitions.

Figure 1. Interpretation examples on connected α-cuts



Related Content