Fuzziness in Spatial Databases and GIS

Jörg Verstraete, Rita De Caluwe

Abstract— **This paper explains why and how spatial databases and geographic information systems can benefit from an appropriate representation of imprecise and/or uncertain information. Not only are the imprecission and uncertainty inherent to the data to be modeled, they also appear in both various queries and the result of various queries.**

Keywords— **spatial database, GIS, fuzzy, uncertainty, imprecision, fuzziness**

I. Introduction

FUZZY SET THEORY (introduced in 1965 by L.A. Zadeh [1]) and the related fuzzy logic are extensions of the traditional set theory and binary logic. They have provided new, more suitable means to model imprecission and uncertainty. The data to be dealt with in real life, commonly are only known approximately and many propositions are not strictly true or false.

Up to now, these theories have been used successfully for the modeling of numerous applications, ranging from data-analysis and planning, to decision support systems and process control. Applications of fuzzy logic are probably best known in the latter field. Of course, during the past decades, the theories have been further developed and refined [2]and their applications have become more and more widespread [3].

A new emerging and challenging application, can be found in the field of Geographic Information Systems (GIS), at the heart of which are spatial databases. Due to the nature of the geographic information, there are particular problems and limitations, which are the subject of this PhD-research.

II. Traditional Spatial Databases and GIS

A. Data: Features, Objects and Layers

It is a misconception that a geographic information system merely contains maps [4]. In a GIS, *features* and information about these features are stored. A *feature* can range from names, locations, dates to numerical data (measurements, statistics, . . .) and maps. The notion *feature* corresponds to what in the spatial database is called an *object* (note that is does not imply that a spatial database has to be an object-oriented database). An object may have four types of attributes: spatial (*where* an object is), temporal (*when* an object is), thematic (*what an object is*) and scale (*how* an object is).

It must be possible to store the data in a spatial database, as well as perform operations on the objects (or attributes) of data, hence *spatial queries* must be supported, which will be addressed in the next section.

Traditionally, features are stored in *layers*. A layer is a group of features, or a usable subset of a dataset, generally containing objects of certain classes (e.g. layers containing roads, layers containing mountains, . . .). In order to solve a query, data of several layers are often needed. Multiple layers can be combined: they are superimposed in such a way that the resultant set would contain the features of the different layers. Superimposition of the required layers is called an *overlay*, and is a technique, additional to the difference and intersection operations, used to solve queries.

B. Queries

The main purpose of a GIS is the querying and analysis of the contained data. An example of a simple query (which involves a single layer) is “What is the name of the mountain at a given particular point on the map?” To obtain the result, it would suffice to look up the longitude and latitude for the selected point, and determine the name of the mountain by looking at the *mountain* layer.

A more complex query (involving several layers, thus requiring the overlay-technique) is “Display all skiable mountains within ten kilometers of an airport”. This query obviously requires several operations on the spatial data. To start the overlay-process, we could look at the mountain layer, and eliminate all non-skiable mountains (which could be contained in a different layer). The next step would then be to overlay this with a layer containing all the airports, and to draw circles with a ten kilometer radius around each airport. The answer of the query, contained in the resultant layer, can then be returned to the user. In general, a GIS should be able to support the handling of five different types of queries [4]:

- Location queries: What is at a given position? Instead of this position being given by coordinates (latitude and longitude), it could also be indicated by the name of a mountain, a zipcode, the name of a building, or a set of constraints that pinpoint a location.
- Condition queries: What locations satisfy a given set of conditions? A set of conditions is listed, and GIS is expected to return every point that satisfies these conditions (e.g. one could query a GIS for the names of mountains which exceed a given altitude and are located in Europe.)
- Trends: What has changed since a given point in time ? Only GIS that consider the notion of time can handle this type of query, unfortunately not all GIS allow the temporal aspect to be taken into consideration. In the case of any GIS however, the incorporation of the notion of time would be very interesting. For example, when considering the soil data set of a mountain range, having several successive sets of the same geographic area would let us determine the rate

---

J. Verstraete and R. De Caluwe are with the Department of Telecommunications and Information Processing, Ghent University (RUG), Gent, Belgium. E-mail: {jorg.verstraete, rita.decaluwe} @telin.rug.ac.be .

of erosion.

- Patterns: What spatial patterns exist? This more complex and sophisticated kind of query could for example be used to find the origin of a certain pollutant, given information about how this type of pollutant disperses.
- Simulation: These types of query are useful for planning and prediction about future evolutions. They allow the user to pose questions regarding "What happens if ...".

III. FUZZINESS IN SPATIAL DATABASES AND GIS

Imprecission and uncertainty can appear both in the data to be modeled in the GIS as in the query itself. The traditional modeling of these semantic notions is however very basic (few or no information about the accuracy of certain values is considered). The occurrence of imprecission can be illustrated by reconsidering the previous example "Display all skiable mountains within ten kilometers of an airport".

The first issue is the meaning of the term "within ten kilometers". While this could be a useful stipulation, it is possible that is not intended in a clear cut way. Locations on a distance of 10.1 km might still be acceptable, 10.2 km might be less acceptable, but e.g. 11 km might be unacceptable. Besides, should ten kilometers be considered in a straight line, or along real, existing roads ? It might also be desirable for the GIS to be queried using linguistic terms such as *walking distance*, in order to provide a more natural way of querying.

The next issue is the term *airport*. Do we consider a small airfield, only meant for private, recreational flights or should only the large commercial airports be taken into account, or are both alternatives valid ? [1]

The term *skiable* depends heavily on the capacities of the person posing the question. Some slopes might be considered skiable for an experienced athlete, but at the same time not at all skiable for the occasional tourist. While this term could be interpreted by having the user specify what he/she considers skiable, this is not straightforward.

An example of uncertainty is e.g. when somebody remembers a hotel in a ski-resort, but is uncertain about the city in which it is located. It could be in city A or B.

These examples illustrate that – although it is unlikely that every problem will be solved – an appropriate representation of the imprecission and uncertainty would allow for a more realistic modeling of numerous data.

IV. RESEARCH OBJECTIVES

The purpose of this PhD-work is to define a formal model to work with imprecise and uncertain spatial data, and to develop and evaluate techniques to store and manipulate such data in a prototype implementation.

Fuzzy set theory will be used to model the imprecission and uncertainty concerning the information. Every element of a fuzzy set has associated with it a number (in the range 0 to 1) which represents the degree to which it belongs to the set. This number is the *membership degree.* As fuzzy set theory is an extension of the traditional set theory, crisp data can also be modeled using this theory. The operators of the traditional set theory have been generalized to be applicable on fuzzy sets, yet the definition of some specific spatial operators will also need to be generalized in order for them to be applicable on fuzzy sets.

Fuzzy propositional logic – which is an extension of the traditional propositional logic – will provide the basis for the definition of appropriate fuzzy querying methods.

Conceptually, several methods for representing fuzzy spatial data are possible. A first intuitive representation is by means of *contourlines* [5,6], simply put : a set of lines connecting points ,which have the same membershipdegree for a specific feature, are used to represent the data. While this is a good representation from a human point of view, it is not that interesting a representation to work with due to its increasing complexity when operations have to be applied on them. Currently, a representation based on a triangulated irregular network (TIN for short) is being investigated. A TIN is a network of triangles, the vertex points of which are all datapoints (these are points for which there is knowledge about a certain feature, the information about points in between will be obtained through interpolation). A grid of triangles form a TIN if the following property is satisfied: three vertex points define a triangle if and only if the circle defined by them contains no other vertex points. Except for degenerate cases (four points that are located at the corners of the same square, ...), a unique TIN can be constructed. Additionally, there are rules that uniquely define the TIN in these degenerate cases. Commonly the method of *Delaunay-triangulation* is used to construct triangulated irregular networks. These triangulated irregular networks have a relatively simple structure, yet are very flexible to use.

In order to implement these representations for a real application, appropriate data structures need to be defined. These data structures can not only be used to represent fuzzy data in the spatial database itself, but could also be useful to store intermediate data (needed while processing a query) or to store the result of the query (even if this result does not need to be persistent).

At a first stage, all these problems are approached in a formal way, but we will also thrive to provide prototype-implementations suitable to test the developed techniques. These will most likely be the implementations of several stand-alone sub-problems than a full GIS-implementation.

REFERENCES

[1] Lotfi Zadeh *"Fuzzy Sets"*, Information and Control, textbf1 3 (1965) 338-353.

[2] Didier Dubois, Henri Prade *" Fundamentals of Fuzzy Sets"* Kluwer Academic Publishers, 2000.

[3] Hans-Jürgen Zimmerman *Practical Applications of Fuzzy Technologies*, Kluwer Academic Publishers, 1999.

[4] Ashley Morris, *Why Spatial Databases Need Fuzziness*, proc. of Nafips 2001, pp. 2446-2451.

[5] Eliseo Clementini, Paoline Di Felice, *An algebraic model for spatial objects with undetermined boundaries*, GISDATA Specialist Meeting - revised version, 1994.

[6] Nicholas Mark Gotts, Anthony G. Cohn, *A mereological approach to representing spatial vagueness*, Working Papers, Ninth International Workshop on Qualitative Reasoning, 1995, pp. 246-255.

---

[1] In the English language, there is a difference between airport, airfield and airstrip, but the example is only meant as an illustration.