Towards Handling Uncertainty of Excavation Data into a GIS

Cyril de Runz[1,2] – Eric Desjardin[1] – Frederic Piantoni[2] – Michel Herbin[3]

[1]CReSTIC-SIC, IUT de Reims Châlons Charleville
[2]HABITER, Université de Reims Champagne Ardenne
[3]Antenne CReSTIC-Châlons, IUT de Reims Châlons Charleville
{cyril.de-runz; eric.desjardin; frederic.piantoni; michel.herbin}@univ-reims.fr

Abstract

Research on data quality through the analysis of accuracy and uncertainty has become a major development in GIS science. Fisher's taxonomy provides solutions for the processing of uncertainty in spatial data representation. However, excavation data differs from geographical data, owing to the specific features of archeology such as incompleteness, temporal and functional aspects. Therefore, archaeological excavation representation requires a GIS tool accounting for its specific needs. This article presents a new taxonomy taking into account the specificity of excavation representation. The authors examine the major theories dealing with the representation of uncertain data, and their use. Finally, they explore the relevance of this taxonomy to knowledge representation in a GIS, and analyze the impact of uncertainty representations according to time and to the shapes of objects in Reims Roman streets.

Keywords

Uncertainty, Taxonomy, GIS

1. Introduction

The analysis of spatial data quality according to accuracy and uncertainty is definitely a major issue in GIS science research. In fact the problem of representation and modelling of spatiotemporally uncertain data has been studied the middle of the 1990s (Burrough and Frank 1996). Thus, it is today a recognized field in GIS, there is at least one international conference a year on the subject (cf. ISSDQ and Spatial Accuracy). There are many approaches for the management of uncertain geographical data in a GIS. One of these consists of studying the nature of the uncertainty and choosing a formalism accordingly. We propose to apply this approach to archaeological data.

The management of archaeological knowledge is complex and a GIS is often resorted to (Conolly and Lake 2006). Indeed, archaeological data can be spatiotemporal, multimodal but also vague, uncertain, ambiguous and patchy. Unlike geographical data, excavation data concerns time periods (Rodier and Saligny 2007), mostly estimated and not defined by the observation period. Therefore, the excavation context requires a specific approach to uncertainty.

In this article, we are presenting a new taxonomy of uncertainty in GIS, specific to excavation data. The use of uncertainty taxonomies is classical in usual GIS. In spatial context, the method developed by Fisher (Fisher 2005; Fisher *et al.* 2005) links each type of uncertainty to a formalism of representation for uncertain data. For example, Fisher proposes to associate error to probabilities and also vagueness to fuzzy sets. Our taxonomy is derived from Fisher's one. To consider the specificity of excavation data uncertainty, we have defined some vague categories.

Moreover, we explore the relevance of this taxonomy to knowledge representation in a GIS for the SIGRem project. The goal of this project is to store, in a spatiotemporal GIS, all the excavation data about Roman objects found in Reims. On the road to reconstructing the past, we analyze the impact of spatial, temporal and shape uncertain representations of objects (Reims Roman streets) into some SIG requests.

First, we will describe the nature of uncertainty in excavation data, then we will expose our new taxonomy of uncertainty specific to archaeological knowledge, and finally we will examine the relevance of this taxonomy by using it in a spatiotemporal complex query.

2. Description of the nature of uncertainty in excavation data

We observe four major kinds of uncertainty: error, vagueness, ambiguity and incompleteness. The aim of this section is to present all of them.

When an archaeological object is well-defined but there is no evidence of the validity of the object, then the object is liable to error. For instance, the following questions induce some errors: "Are we or are we not in the object?" "Is the description true?" and "Was the object present at a given date?". We may also have come to the conclusion that "the object is a wall with a probability of 90%".

Vagueness in data means an object is not clearly defined, i.e. its attributes are fuzzy. This can be frequently experienced with semantic code estimations (the First Century), values obtained using observation tools with poor resolution, or descriptions like yes-maybe-no. Two categories of vagueness can be defined. The first one is imprecision; in this case we include the degrees of relevance of computational and cognitive values (e.g. semantic code estimations) or of some deviations of measurement value from the true value (e.g. values depending on the resolution of observation tools). The second one is the approximation, where the object definition could be either "yes", "no" or "maybe", or "inside", "outside" or "maybe".

A definition of an object is ambiguous when there is doubt as to how a phenomenon should be classified because of differing perceptions of it. There are two major kinds of ambiguity: discord and non-specificity. In the case of vague data, as first century above, with semantic code estimations (the first century), values obtained using observation tools with poor resolution, or descriptions like yes-maybe-no, there can be discord in the definition. The case of an object which is described as a wall and also as a part of an amphitheater is non-specific.

We call incompleteness the fact that information is partially or not defined. The lack is when some information is null in the databases. This may occur when an archeological object cannot be dated. Lacunarity then refers to parts of bigger objects which have not been primarily identified in the database: fragments of street or fragments of wall.

Using those descriptions of the different natures of uncertainty in excavation context, in the next section we propose a new taxonomy explaining the choice of formalism according to the nature of uncertainty.

3. A taxonomy of uncertainty for archaeological knowledge

There are many approaches for managing uncertainty. Fisher's is certainly one of the most frequently used in the spatial context. We suggest a taxonomy, derived from Fisher's, specifically adapted to the needs of archaeological research. Our approach links each kind of uncertainty to one of the five classical uncertainty representation theories. Those theories are the probability theory, the fuzzy set theory (Zadeh 1965), the possibility theory (Zadeh 1978; Dubois and Prade 1988), the rough set theory (Pawlak 1991) and the theory of Evidence also called Dempster-Shafer theory or belief functions theory (Dempster 1968; Shafer 1976). Our taxonomy results from a specific approach for archaeological GIS, and presents a unified view of uncertainty for both time and space.

This taxonomy is based on practical works in the field of GIS (Burrough and Frank 1996, Davidson *et al.* 1994). The link probabilities and error is natural: the chance and the risk are really close to probability in a semantic view. Fuzzy sets (as possibilities) are associated to non-specificity because the complementary notions of necessity and possibility allow us to make soft decision. The theory of Evidence proposes a formalism which softly extends probability in order to obtain a conflict index which characterizes the confidence in the decision. Fuzzy set proposes a granularity between 0 and 1 associated to the membership of an element to a concept and it is easy in this approach to represent vagueness.

Our taxonomy is based on Fisher's (Fisher 2005; Fisher *et al.* 2005). The taxonomy of Fisher (*Fig. 1*) associates the probability theory to error, the fuzzy set theory to vagueness, the theory of evidence to discord and the fuzzy set theory to non-specificity.

The granularity of our approach, which is illustrated in *Fig. 2*, is more detailed for the vagueness. For each of the three kinds of vagueness we propose to use a different formalism. Indeed, we represent imprecise data by fuzzy sets and approximate data by rough sets. The granularity of the fuzzy set theory is essential in order to represent deviations and gradual membership; thus it is a good formalism for modeling imprecision. The sets of properties in the rough set theory may represent the notion of "A is surely a member of B", "A is

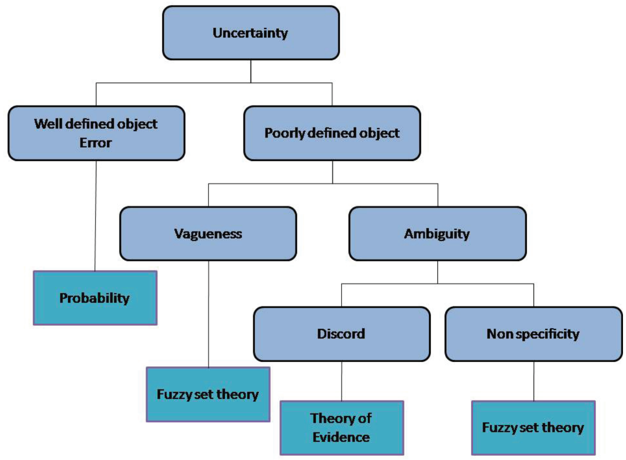


Fig. 1. Taxonomy of Fisher for uncertain spatial objects.

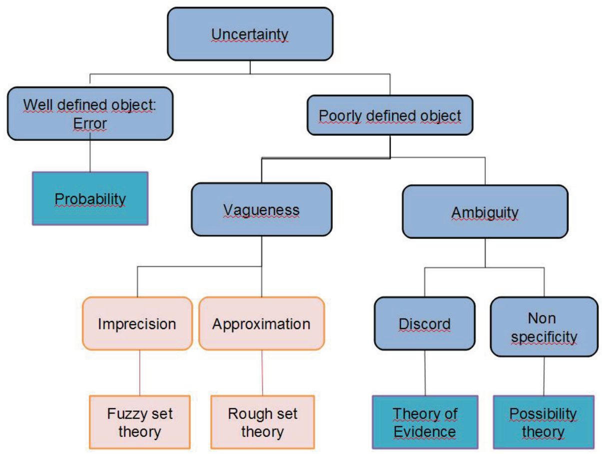


Fig. 2. An uncertainty taxonomy for excavation knowledge.

possibly member of B" and "A is not member of B" and thus its use makes the consideration of approximate information easy.

The lack of data could be represented using any theory. The lacunar aspect of data should be solved during the exploitation.

We use this taxonomy in the SIGRem project for the choice of datum representation in accordance with its quality. Exploiting the results of our choice, we have defined in (de Runz *et al.* 2007) some complex spatiotemporal queries that will be examined in the next section.

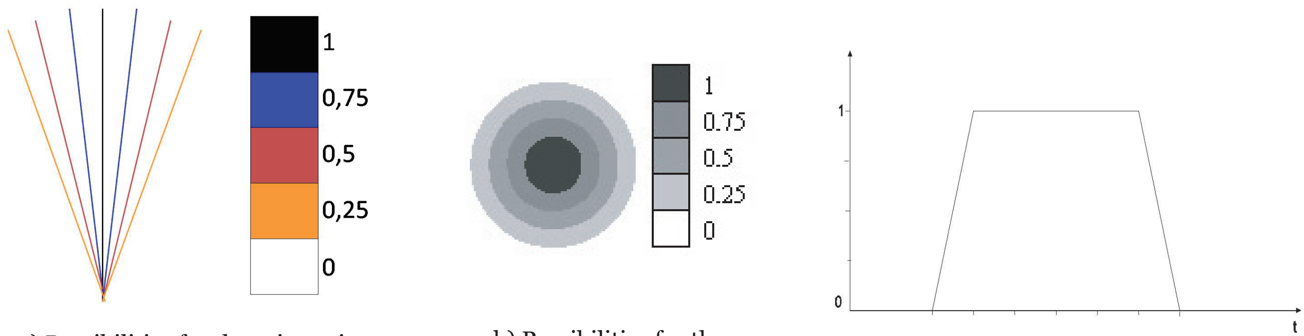
4. A new query for the SIGRem project with shape criterion

The goal of the SIGRem project is the development of a GIS for Roman excavation data in Reims**.** This development is organized in two activities. The first one is the storage and exploitation of excavation data about Durocortorum (Reims during the Roman period). The second one aims at producing some new kinds of support for archaeologist's needs.

In this project, the BDRues database stores Roman street excavation information. Data in this database have three main features: location, orientation and period of activity of objects (see de Runz 2007). Moreover, we know that the streets in Durocortorum were linear. In this section we propose to represent this information according to their uncertainty and to use those representations in a query to obtain simulated street positions for a given period. The goal of this query is to try to reduce the lacunar part of BDRues information.

Due to tool precision, the BDRues data are imprecise. Due to the fact that we only have information on fragments of streets, the BDRues data are lacunar and due to the evaluation process of activity periods, the BDRues data are imprecise. Thus we propose to use, for locations and orientations, some possibilistic representations and some convex and normalized fuzzy sets for dates. Some models are presented in *Fig. 3*. Fuzzy set is defined using a membership function taking values in [0,1] on the domain of a concept definition. The membership values correspond to the degree of the membership of a domain element of the studied concept.

According to the fact that possibility distributions are also fuzzy sets, the query we propose is organized



a) Possibilities for the orientation fuzzy set : fOrien_ep

b) Possibilities for the location fuzzy set : fLoc_ep

c) Fuzzy period of activity fDate_ep

Fig. 3. Fuzzy models for the representation of an object ep in BDRues.

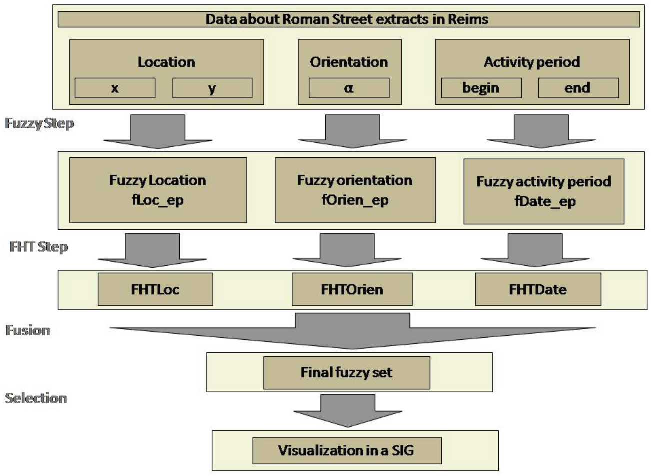
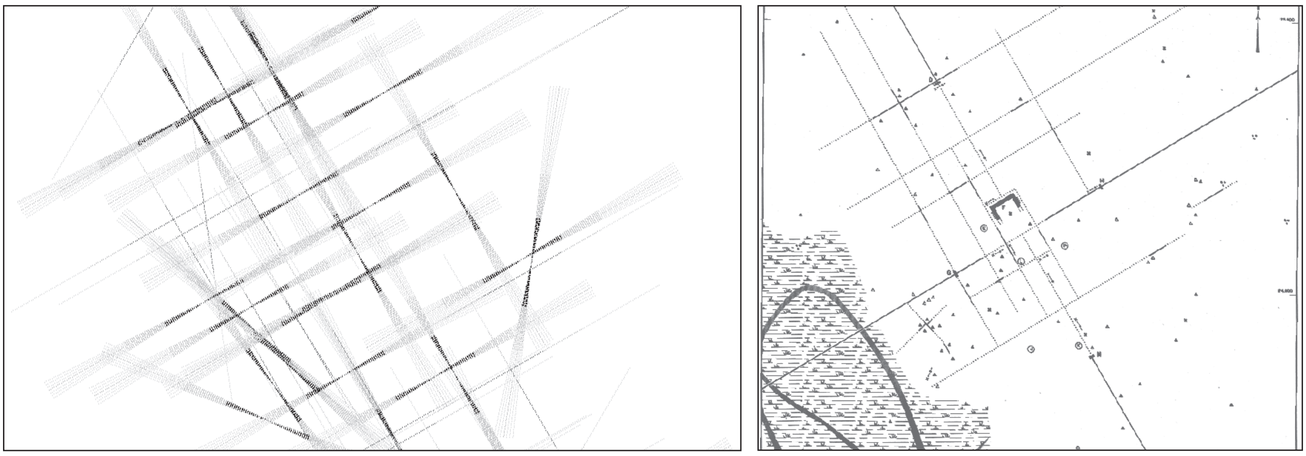


Fig. 4. The structure of a query with shape criterion.



a) Map resulting of the query

b) Map from experts

Fig. 5. Maps for the 3rd Century.

in four steps (*Fig. 4*). The first one is the fuzzy step where objects are represented by fuzzy sets. The second step uses a classical method for pattern detection – the fuzzy Hough Transform (FHT) introduced in (Han *et al.* 1993) – on each feature. The third step performs the fusion of the information resulting from the previous step. The last step is the selection of information from the fusion.

The second step uses the Fuzzy Hough Transform, an evolution of the Hough transform (Hough 1962; Duda and Hart 1972). The goal is to detect simple shapes such as lines in a 2D data set. In those algorithms, each object votes for cells in an accumulator corresponding to all shapes going through it and its neighborhood. The cells with the highest values represent the detected shapes. The vote value of each object depends on the value of the fuzzy membership function defined on its neighborhood.

We build an accumulator for each component (location, orientation and activity period) using this principle. We thus obtain three Fuzzy Hough Transform accumulators: FHTLoc for location, FHTOrien for orientation and FHTDate for activity period.

The third step consists in merging the three accumulators into a final fuzzy set. To do it, we normalize each accumulator and use a weighted mean. Depending of the level of confidence wished by the user, we will only select the street positions for which the membership value is high enough.

In our approach, we use an α–cut on the final fuzzy set for the selection of data and then visualize the results in a GIS. An α-cut consists in the selection of all domain values (elements in our application possible streets) that corresponding membership function values are higher or equal than α (Zadeh 1965). The use of α-cut allows us to visualize multiple propositions for each street if the index values of their possibilities of presence are higher than α.

The query for the third century gives the map in *Fig. 5a*. The comparison between the results and the map by experts (as *Fig. 5b*), allows us to conclude that our query is pertinent. Indeed, simulated streets and streets defined by experts are most of the time similar. The confidence on streets propositions is higher than 0.95 (an α-cut is used with α=0.95.

5. Conclusion

In conclusion, modeling uncertainty is fundamental in archaeology. In this paper, we explained the nature of uncertainty in the context of excavation data. We also propose a taxonomy in which kinds of uncertainty are associated to theories of representation. This organization helps, according to the quality of the data, to choose the *ad-hoc* formalism to represent data in GIS.

We exploit this taxonomy for the choice of representation formalism in the SIGRem project. Using the chosen formalism, we expose a complex query to help reconstructing the past. This query has four steps and uses a classical method of shape detection and a merging function. The results, which are visualized in a GIS, are pertinent. Thus this kind of query provides a new tool for archaeologists.

Acknowledgements

The authors would like to thank SRA (Regional Service in Archaeology) Champagne-Ardenne and INRAP (National Institute of Research in Preventive Archaeology) in Reims for the access to their data, and Dominique Pargny for his contribution to the SIGRem project.

Bibliography

Burrough, Peter and Andrew Frank (eds.) (1996). *Geographic Objects with Indeterminate Boundaries*. Taylor & Francis.

Conolly, James and Mark Lake (2006). *Geographic Information System in Archaeology*. Cambridge University Press.

Davidson, Donald, Sid Theocharopoulos and Ron Bloksma (1994) A land evaluation project in Greece using GIS and based on Boolean fuzzy set methodologies. *International Journal of Geographical Information Systems 8,* 369–84.

De Runz, Cyril, Eric Desjardin, Frédéric Piantoni and Michel Herbin (2007). Using fuzzy logic to manage uncertain multi-modal data in an archaeological GIS. *ISSDQ'07*.

Dempster, Arthur (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society* 30, 205–247.

Dubois, Didier and Henri Prade (1988). *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press.

Duda, Richard and Peter Hart (1972). Use of the Hough transform to detect lines and curves in pictures. *Communication of ACM* 15, 11–15.

Fisher, Peter (2005). Models of uncertainty in spatial data. In: Paul Longley, Michael Goodchild, David Maguire and David Rhind (eds.) *Geographical Information Systems. Principles, Techniques, Management and Application*. Wiley, 191–205.

Fisher, Peter, Alexis Comber and Richard Wadsworth (2005). Nature de l'incertitude pour les données spatiales. In: Rodolphe Devillers and Robert Jeansoulin (eds.) *Qualité de l'information géographique*. Hermes, 49–64.

Han, Joon and Laszlo Koczy and Timothy Poston (1994). Fuzzy Hough Transform. *Pattern Recognition Letters* 15, 649–648.

Hough, Paul (1962). *Method and means for recognizing complex pattern*. US Patent 3 069 654.

Pawlak, Zdsislaw (1991). *Rough Sets. Theoretical Aspects of Reasoning about Data*. Kluwer.

Shafer, Glenn (1976). *A Mathematical Theory of Evidence*. Princeton University Press.

Rodier, Xavier and Laure Saligny (2007). Modélisation des objets urbains pour l'étude des dynamiques urbaines dans la longue durée. *SAGEO'07*.

Zadeh, Lofti (1965). Fuzzy Sets. *Information Control* 8, 338–353

(1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems* 1, 3–28.