

SPATIAL DATA UNCERTAINTY MANAGEMENT

D. Klimešová, E. Ocelíková*

Czech University of Life Sciences, Prague, Faculty of Economics and Management, Kamýcká 129, 165 21
Prague 6, Czech Republic - klimesova@pef.czu.cz
Czech Academy of Sciences, Institute of Information Theory and Automation, Pod vodárenskou věží 4, 182 00
Prague 8, CR - klimes@utia.cas.cz

*Technical University Košice, Faculty of Electrical Engineering and Informatics, Department of Cybernetics
and Artificial Intelligence, Letná 9, 042 00 Košice, Slovak Republic - ocelike@ccsun.tuke.sk

Commission IV, WG IV/2

KEY WORDS: Knowledge, Classification, Knowledge Management, Contextual Modelling, Temporal Modelling, Decision Support

ABSTRACT:

The paper deals with the relations between knowledge management, uncertainty and the context evaluation on the background of the new possibilities of information technologies that can help us to carry out the knowledge management strategies. The paper discusses the problem of wide context including to compensate and decrease the uncertainty of data on one hand and to increase accuracy of classification or segmentation and efficiency of further analytical processes to increase the information value of decision support. The technique of fuzzy measurements and a fuzzy approach in general is addressed also as a way to catch the uncertainties of analytical process (classification, segmentation, ...) and to transmit them to the other processing stages.

1. INTRODUCTION

The ways of managing and distributing data and particularly data sources has rapidly changed. Data are collected and processed and during the last couple of years the data flows in and between organizations have extremely increased. In the connection with these facts also the data management tools and techniques are continually changed [10].

Knowledge management (KM) is an essential process improving competitive advantage. KM tries to compensate the loss of stable procedural knowledge, the loss of customer-related or project-related experiences and know-how or the loss of the middle management information analysis and routing services.

The area in Computer Science that is most influenced by the concept of knowledge is Artificial Intelligence and Knowledge Based Systems together with Geographic Information Systems (GIS) [6]. Since the knowledge is specified independently from the application domain, reuse of the knowledge is enabled for different domains and applications [5].

The modelling process is dependent on the subjective interpretation of the knowledge engineer. Therefore this process is faulty and an evaluation of the model with respect to reality is indispensable for the creation of an adequate model. Since this control knowledge is specified independently from the application domain, reuse of this strategically knowledge is enabled for different domains and applications. Besides knowledge modelling also knowledge representation is an important field of research in computer science and AI.

Besides knowledge modelling and knowledge representation is also an important field of the building process of knowledge database. The modelling process is a cyclic process that may lead to the refinement, modification, or completion of the

already constructed model or guide further acquisition of knowledge – contextual understanding.

2 UNCERTAINTY

2.1 Multi-source Data

The Spatial variability, the dimensionality of data and the complexity of objects structure hierarchy are rapidly growing and consequently with these aspects increase the uncertainty entering into the processing. A great number of existing databases offer a variety of data sets covering different thematic aspects like topographic information, cadastral data, statistical data, digital maps, aerial and satellite images including temporal data.

GIS works with the combination of data sets which may have a very different uncertainty structure. There is the question of correlation and the necessity to consider the covariances together with variances.

The problem of scale exists – the loss of knowledge about the variability within the map unit, the quality of estimations and the unknown model of overall uncertainty.

GIS is proposed as multisource technology as opened system with wide spectra of interconnections. But on the other hand It is quite another situation than in case of the closed system, where the user has full control over all steps of processing from data input to presentation of results. In frame of open interoperable system with access to web sources with a great number of existing databases the user control gets completely lost. User needs an appropriate uncertainty model for this purpose, integrated in GIS [4].

Uncertainty arises from imperfect understanding of the events and processes. From the philosophical point of view the

uncertainty is quite natural part of our life and the surrounding world and the fact of uncertainty is very stimulating for the research on the field of *defining, measuring, modelling and visualizing*. Moreover, the uncertainty opens the space for further questions and the answers to this question can help us to do better decisions [7].

Data are not perfect from many reasons: incomplete data, precision of measurements, discreet description of connective phenomena or inherent part reflecting our understanding of things [11]. To reduce uncertainty of data it is mainly the question of the proof of recognized quality assurance. Users often take the pragmatic approach to the cost versus accuracy. Sometimes, without the relevance testing, the resolution of data is used for the whole set of different task. Then the problem of over-defined and under-defined objects brings the difficulties [8].

2.2 Uncertainty Modelling

Especially uncertainty of a geographic object can be modelled through uncertainty of its geo-spatial, temporal and thematic attributes. Uncertainty of relations takes into consideration spatial, temporal and spatial-temporal relations. To add suitable attribute or to spread the net of relations reduce the uncertainty of the object.

The special case is to model objects uncertainty using spatial-temporal approach to the objects and incorporate spatial-temporal relationships. The dynamics of object is very powerful tool to obtain exact results about the object and phenomenon behaviour to support further decision [16].

The decision making process is always associated with some level of uncertainty which can rise from: definition of the problem, data used, sequence of operations used to obtain result and understanding of result.

In case when the concept is not precisely defined, the characteristics are not precisely measured and insufficient estimation of parameters is in disposal than the techniques of fuzzy logic may be applicable. This way allows computing and evaluating results when the object or phenomenon state can be sufficiently described using linguistic terms that can be consequently quantified in some fashion.

Temporal aspects

GIS packages mostly lack the ability to perform temporal analysis of spatial data in general. The temporal analysis is applied for selected applications with specially arranged conditions. The fact is that decision support procedure is highly dependent on the ability to represent and analyse temporal processes in a GIS. Moreover, the widespread application of GIS in urban, regional and environmental planning, in marketing and logistics requires the integration of temporal data in order to perform analyses or to deal with temporally variable data.

Another motivation for the development of temporal GIS techniques is that the great amounts of data have been collected the research projects have been solved and important results have been achieved, which cannot be revaluated and further reused. The development of appropriate methods and tools as well as their implementation is affected by the lack of standards for the temporal data representation and analysis. It means the

set of tasks where no single criterion is sufficient to arrive at a decision with the comfortable level of certainty.

2.3 Multi-criteria decision problems

In many cases only the combination of many information sources and eventually to combine expert knowledge can produce requested information. Very similar situation except GIS is also in medical tasks connected with diagnosis determination.

Parameters estimation

The use of modelling functions strongly depends up exact definition of parameters as follows:

- specification of one or more target locations,
- specification of the neighbourhood that surrounds the targets,
- specification of the way spatial elements are interconnected,
- set of rules that specify the allowed movement along these interconnections,
- specification of a function to be performed on the "things" found surrounding the targets in the specified neighbourhoods,
- set of resources (such as a finished manufactured product),
- one or more locations where the resources reside (warehouse or factory location),
- objective to deliver the resources to a set of destinations (customers),
- set of constraints that places limits on how the objective can be met (speed of travel, time spent delivering the products, etc.)

The different requirements can be described by the set of factors and coefficients, but these factors are often connected to the critical characteristics coming from the selected area and surrounding objects that can influence the estimation quality in the frame of classification procedure. It is possible to define priority function **PF** with the aid of priority functions of the selected set of factors and weighting coefficients as follows:

$$F = \sum_{i=1}^n w_i f_i, \quad \text{where} \quad \sum_i w_i = 1 \quad (1)$$

The great part of parameters can be successfully put more precisely with the aid of expert knowledge. GIS applications are frequently used in producing new information by combining information from different sources, by spatial analysis of existing data and by implementation of additional information coming from previously processing and analysis, expert knowledge, objects dynamics and trends.

Usually the objective in applications involving contextual modelling is to locate the area or areas where the given criteria apply and eventually calculate the measure of exposure to hazard in case of infections, diseases and pests, find the optimal routes and produce different complex scenarios. The powerful tool is the way of buffering where the expert knowledge can help us to set the ranges and find the areas with defined ways of protection.

The evaluation procedure consists of the two steps: to set up parameters and determine their importance at first and provide the sensitivity analysis to demonstrate the effect of selected

parameters, to define weighting measures and to decrease uncertainties resulting from the importance determination.

3. CONTEXT AND FUZZY APPROACH

3.1 Context Understanding

It is possible consider context as follows.

- context as the reflection of object or phenomena using different interpretation through the system of cognition: perception, conception and interpretation,
- context as the reflection of selected facts is concerned with validity of statements and the system of argumentation,
- context as the reflection concerning validity of statement using knowledge generating system – knowledge based context

Using context it is possible to derive new quality of information that can be used to support decision. The integral part of control GIS [9] is the modelling where the information layers from real, artificial and virtual world are composed together to select optimal solution or verify given hypothesis or assumptions.

The contextual design of spatial data and further development of geo-information technologies, image processing techniques and the possibilities of object history modelling together with the geographical networks environment will provide quite new and considerably wider possibilities of using GIS.

Good message is that the increase of interactions and correlations implies decrease of differences between objects. From the point of uncertainty and contextual modelling the wide context (temporal, spatial, local, objective, attribute oriented, ...) influences in the sense of decreasing overall uncertainty of the results – the similar effect like the use of sequence of images in case of noise suppression – image reconstruction.

3.2 Fuzzy Approach

GIS architecture is open to incorporate new requirements of knowledge-based analysis and modelling, namely in connection with web designed spatial databases and temporal oriented approaches. The example of context implementation is composed classifier where the fuzzy elements are implemented. Composed classifier is composition of component classifiers, which predictions are connecting by combining classifier, unlike others simple classifiers. There are several architectures for possible combination of classifiers.

The analysis of geographical data tends to tasks that are solved using defined sequences of operations and algorithms where the

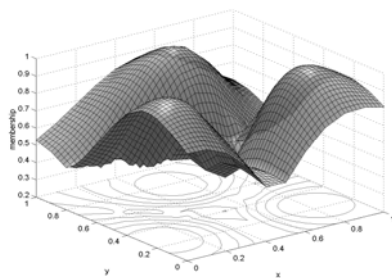


Figure 1. Outcome of the Fuzzy C- means

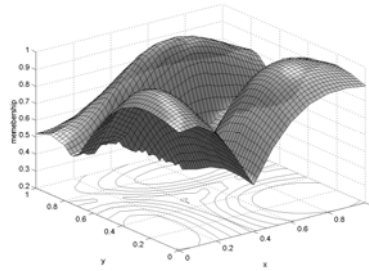


Figure 2. Outcome of the Gustafson-Kessel Algorithm

set of conditions is provided per parts. The amount of cumulative functions allows to capture the uncertainty and imprecision. The solution of zoning and tolerance extents where the fuzzy approaches are used the slivers are generated and together with overlay function arise the complexity of results.

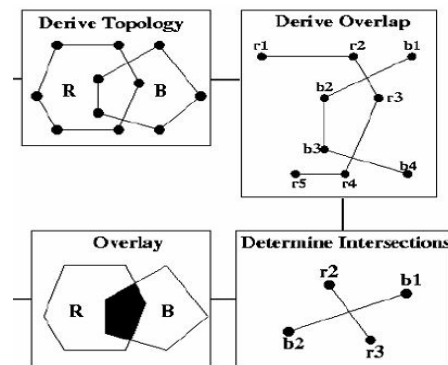


Figure 3. Overlay function and its consequences

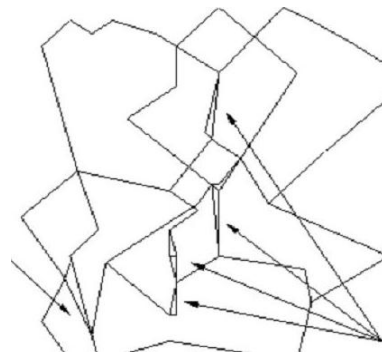


Figure 4. Overlay function and its consequences

4. CONCLUSIONS

In this paper, the problem of uncertainty of multisource data and multicriterial analysis is addressed and the use of wide spatial and temporal context and expert knowledge integration to decrease the uncertainty is discussed. The running development of information technologies, image processing techniques and knowledge-based databases, together with the geographical networks environment, will provide quite new and considerably wider possibilities of using GIS. Our decisions are becoming increasingly dependent on understanding of complex relations and phenomena in the world around and GIS technology is able to incorporate new requirements.

The main goal has been to show selected aspects of this process and compare the increasing possibilities of the sources with the difficulties of data contextual structuring, object dynamics implementation and expert knowledge integration. The estimation of uncertainty is important: the positional accuracy as well as the semantic one using simulation, sensitivity analysis or statistical characterization.

REFERENCES

- [1] Aamodt, A. and Nygard, M., 1995. Different roles and mutual dependencies of data, information and knowledge. *Data & Knowledge Engineering*, 16, 191-222.
- [2] Benedikt J., Reinberg S., Riedl L., 2002. A GIS application to enhance cell-based information modeling. *Information Sciences* 142 (2002): 151-160.
- [3] Bernbom, G., 2001. *Information Alchemy: The Art and Science of Knowledge Management*, EDUCAUSE Leadership Series #3. San Francisco: Jossey-Bass. Graham, Ricci.
- [4] Bolloju N., 1996. Formalization of qualitative models using fuzzy logic. *Decision support systems* 17(1996): 275-289.
- [5] Cornelis, B., and Brunet, S., 2000. "A policy-maker point of view on uncertainties in spatial decisions". "Spatial data quality" (Shi W., Fisher P., and Goodchild M., Eds), Chapter 12, pp. 168-185.
- [6] Fensel, D., Decker, S., Erdmann, M., and Studer, R., 1998. Ontobroker: Transforming the WWW into a Knowledge Base. In Proceedings of the 11th Workshop on Knowledge Acquisition *Modeling and Management*, Banff, Canada, April 18-23.
- [7] Fuller R., 2000. In: *Introduction to Neuro-Fuzzy systems. Advances in soft computing*, Physica-Verlag Heidelberg. 289 pages.
- [8] Klimešová D., 2006. Study on Geo-information Modelling, 5 (2006), *WSEAS Transaction on Systems*, pp. 1108-1114.
- [9] Klimešová D., 2004. Geo-information management. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35 (2004), 1, pp. 101-106.
- [10] O'Leary, D. 1998. Knowledge Management Systems: Converting and Connecting. *IEEE Intelligent Systems*, May/June 1998, pp. 30-33.
- [11] Parent C., Spaccapietra S., and Zimanyi E., 2000. MurMur: database management of multiple representations. Proceedings of AAAI-2000 Workshop on Spatial and Temporal Granularity, Austin, Texas.
- [12] Peuquet, D.J., 2002. *Representations of Space and Time*. The Guilford Press.
- [13] Power, D., J., 2002. *Decision Support Systems: Concepts and Resources for Managers*, Quorum Books Published 2002.
- [14] Yao T., Journel A. G., 1998. Automatic modeling of (cross) covariance tables using fast Fourier transform. *Mathematical Geology*, 30(6): 589-615.
- [15] Zerger A. 2003. Examining GIS decision utility for natural hazard risk modeling. *Environmental modeling & software*, 17 287-294.
- [16] Zhang J., Goodchild M. 2002. *Uncertainty in geographical information*. Taylor & Francis, London, pp. 127-130.

ACKNOWLEDGEMENTS

The Project Information and knowledge support of strategic control - MSM 6046070904 and grant project VEGA 1/0386/08 supports this work.