THE INCERTITUDE DISPOSAL METHODS OF THE SPATIAL DATA

M. H. Liu [a,] *, Z. W. Yuan [b,]

[a] Sino-Korea Chongqing GIS Research Center, College of Computer Science, Chongqing
[b] University of Posts and Telecommunications, Chongqing, 400065, P.R. China
Email:liumh@cqupt.edu.cn

KEY WORDS: Spatial Data, GIS, Incertitude Disposal

ABSTRACT:

In the process of constructing and applying the GIS, the data is the object of the GIS operation and the basis of the science decision-making and analysing so that the data quality is determined the success or failure of the whole system application. Any indeterminations in the data layers can diffuse through analysing and combining with other error source. It makes risks to decision-making. The theory and method of incertitude analysis in the GIS and RS is one of the GIS study focus at present, because there is not a mature and systemic theory and method to disposing the incertitude in the GIS. In this article, we study mostly the incertitude of the spatial data, the method of disposing the incertitude in the GIS and RS. In this paper, the inadequacies of the current incertitude control methods for spatial data were pointed out. We argue that some strategies can be taken to manipulate incertitude according the spacial data incertitude resource. Robust algorithm can effectively avoid incertitude and improve the spacial data qualities in the process of spacial data, and spatial data incertitude can be control by studying robust algorithm.

1. INTRODUCTION

Thanks to the advancements in Geographic Information systems (GIS) technology and the recent development in telemunication facilities, spatial databases have become easily accessible and the uses of spatial databases have greatly increased. Quite often, users of spatial data rely on GIS to store, manipulate, analyze, and display their spatial data. Maps are the most common form of output from GIS to represent spatial data for a wide variety of purposes, includmg decision making and policy formulation. However, users of spatial data should concern the quality of data which in turn affects the confidence of decision or the success of policies.

Spatial scientists and geographers do realize the important of quality of spatial data. they have been developing techniques and frameworks to visualize the quality of data through various cartographic means. However, most efforts either emphasize on the techniques of representing data quality information, Very few researchers investigate how to control quality of data by devising better algorithm in GIS, the environment in which spatial data are captured, stored, analyzed, and represented. In this paper, we argue that some strategies can be taken to Dispose incertitude according the spacial data incertitude resource . Robust algorithm can effectively avoid incertitude and improve the spacial data qualities in the process of spacial data. A lot of cuccessful strategy to manipulate incertitude lie in the realization of robust algorithm. In the following section, we will point out the inadequacies of the current incertitude control methods for spatial data. We will also provide a brief description of some strategies be taken to manipulate incertitude according the spacial data incertitude resource. Then, we will identify some aspects of spatial data incertitude that can be control by studying robust algorithm.

2. DERIVING SPACIAL DATA INCERTITUDE IN GIS

Fisher (1995)proposed the concept model of incertitude in spatial data (Klir and Yuan,1995) (Figure 1). He thought the essence of incertitude is how to define the object classes and single object to be tested .On his opinion, Incertitude include Objects well define and Objects maldefine, So we should take some responsing measures to control spatial data quality. Author argue that too many incertitude don't cause only by weather Objects be well defined or not.For example, the area data of forest and cultivated land are very different between Department of Agriculture and Department of forest.Though Objects maldefine is one of the reason, the other reason maybe concern to Department benefit protection. Grain for Green Project in china, the centre governoment will invest according the eara converting the land for forestry and pasture. Of course, Department of Agriculture maybe focus more on reserving natural farmland, while Department of forestry more on further speeding up environmental protection plans. So there are incertitude when we use the data come from different resource in GIS. Such incertitude maybe come from Objects maldefine, maybe Objects be well define, for Department benefit protection reason or another, incertitude be introduced in the name of Objects maldefine. On this conditions, how can we judge weather such incertitude comes from Objects well define or Objects maldefine? So author argue that this is another incertitude factor.

An land use comprehensive planning revisal example is taken to show the possibility of spacial data incertitude resource. In the program of land use comprehensive planning revisal,we must collect a lot of spacial data, some spacial data incertitudes were found and listed in following.

* Corresponding author: M. H. Liu, Email:liumh@cqupt.edu.cn, Tel: 023-62468627

1. Attribute discrepancies. typical eara unit of measure is hectare,but the territorial recorder is mitre square, and we traditionally use Mu.If we measure on digital map,maybe the typical eara is millimetre.when we exchange from one to another unit of measure,there are discrepancies. Attribute discrepancies can come from data aggregate, When data are aggregated spatially, a smoothing or generalization process is underway (David et al., 1996)Attribute discrepancies is very much related to time resolutions. We found that a lot of villages were ingathered,but the spacial did not refresh. (Figure 2)

2. The positional discrepancies. The positional discrepancies derived from overlaying different data sources vary among regions. Since GIS spatial data come from different sources, of different formats, and of different degrees of accuracy.

3. Completeness of database. Definitely, the completeness of database is very much related to how recent the database has been updated.

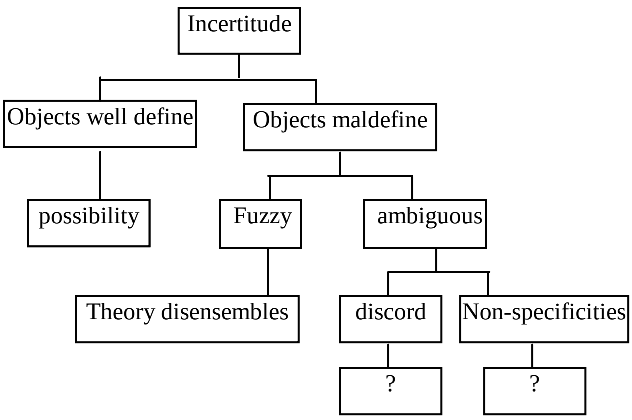

Figure 1. Taxono

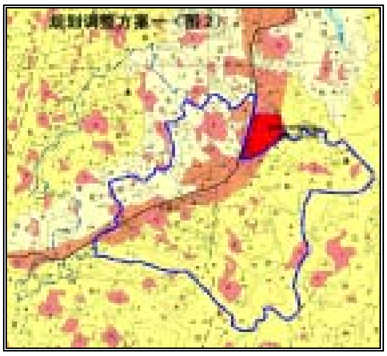

Figure 2. Map of land use comprehensive planning revisal

3. THE INCERTITUDE DISPOSAL STRATEGY OF THE SPATIAL DATA

3.1 Some Strategies can be Taken to Dispose Incertitude According the Spacial data Incertitude Resource

David W.S.Wong and C.Victor Wu had proposed a strategy to manipulate incertitude (David et al., 1996)(Figure3). In the process to manipulate incertitude,the type of appliction, decision and utilization must be considered at first . Then Select the data, hardware and software, processes, and type of product. We may test which significant form of incertitude affect the product and how to comment? What qualities of product is needed? The qualities of product is acceptable?at last we make decision weather we absorb incertitude residual or reduce incertitude existence.

Author argue that when Consider problem from another angle, some strategies can be taken to Dispose incertitude according the spacial data incertitude resource.

For Attribute discrepancies, we can compare the attribute data from different sources or different formats with prior knowledge about which data sources are more reliable. Comparing the database in question with the more reliable data sources can yield data quality information. Quite often, spatial decision and analysis rely upon spatially aggregated data. Therefore, Error is introduced during the generalization process. In general, highly aggregated data are less reliable than disaggregated data, an approach to evaluate aggregation error primarily for vector format data in GIS is to compares data of different scales or resolutions.

As far as the positional discrepancies, To evaluate the positional accuracy, users can import the data into a GIS package using an aerial ortho photo, SPOT image or similar type of remotely sensed data of the same area as the background. Superimposing the different data sources can tell users which type of data is more accurate and reliable in certain areas.

For example, a TIFF image generated from an aerial photo can be put in the background. Substantial positional discrepancies between the different spacial data can be recognized.

Conceptually, these discrepancies can be captured in terms of some statistics to reflect the discrepancies. These statistics can further be manipulated, stored, and also displayed together with the spatial data to reflect the positional accuracy of geographical features in the spatial database.

For completeness of database ,GIS can be used to compare different data sources and derive completeness information for spatial databases. Theoretically, when spatial databases are imported to GIS in which topological relationshps are stored or extracted, different data sources can be compared to derive data quality information about the logical consistency of spatial data. However, topological structure and relationship in geographcal features are modeled and captured implicitly in most GIS. Some systems have particular features or functions that may assist the evaluation or testing of logical consistency in the database. Since the process to assess logical consistency in GIS is relatively system-dependent, it is not feasible to derive a general framework to be implemented in GIS to derive logical consistency information.

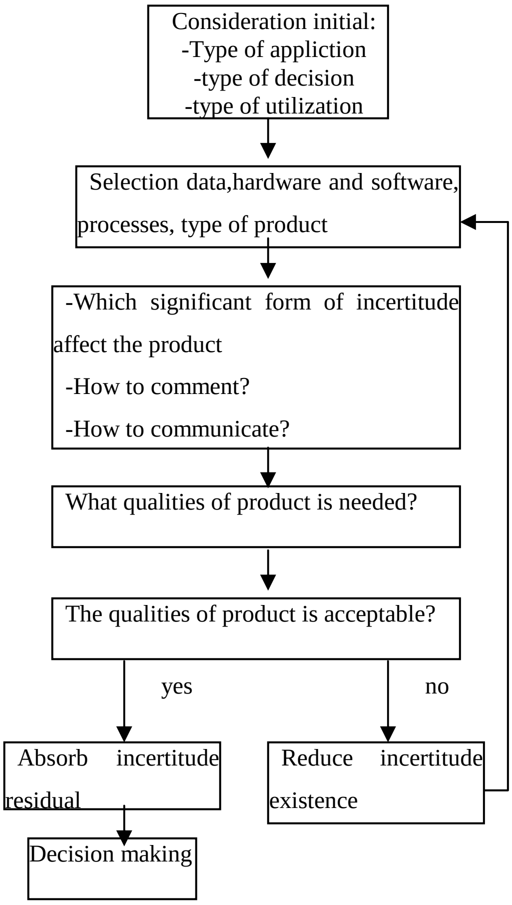

Figure3. A strategy to manipulate incertitude in GIS.

3.2 The data process methods play an important role during incertitude manipulate. Among these, robust algorithm is very important in all factors which cause spatial data . In the following , the relationship between the data process methods and spatial data qualities is be exemplified.

For example, Spacial object reconginion and classification play an important role in romote sensing gragh processing.

When interpretation of remote sensing image, we need the geometric properties feature and Topological associations of spacial objects as well as the spectral feature. In conventional methods of image analysis, remotely sensed data are processed mostly on a pixel-by-pixel or cell-by-cell basis, but little information about shapes of objects except simple measures such as sizes and form factors of objects are employed(WKPratt,1991;Serra,1988).As the spatial resolution of remotely sensed data increases, it has become more possible, desirable, and necessary, to identify the spatial characteristics, or shapes, of objects. This information, as well as conventional spectral intensity values, are required for the automated interpretation of remotely sensed images.

Many people have explored on shape-based Recognition. They present an approach to the recognition of complex shaped objects. In the process of shape-based Recognition of linar objects such as road or river,one of important task is to find fork pixel(junction)(see Fig4).Obviously, there are incertitudes on how to localizate the junction in shape-based Recognition of remote sensing. The junction may be 2、3、9、5,from different people's view of different direction of the line. An efficient approach for the detection of junction in gray-level images was be proposed in the following.

Detection of junction in Gray-level Images is divided into three steps. Firstly, digital image pre-processing is advised so as to get the lines extracted from the original image. Four different measures are proposed: the original image is enhenced and binarized, noise is removed, and skeletoned which kept the lines connectivity. Secondly, the neighborhood pixes -track algorithm is taken. For this purpose, the start forked pixel is detected , neighbor-track is taken to find the last forked pixel point and end pixel point, and the localization of junction is involves. Examples are provided based on experiments with synthetic and real images.

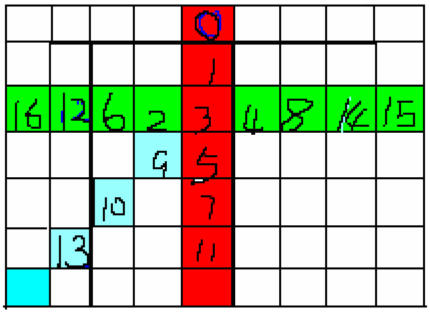

Figure 4. Localization the junction on shape-based Recognition (Illustrate as the picture above, three fractures intersect at pixel 3)

Basic stages

Three stage can be adopted to detect junction in gray-level images(see tab1).

1． started forked pixel detection. As show in fig.4, suppose we track the pixel from pixel 0, A pixel is called as a forked pixel if there are more than one pixels of nonzero graylevel except preprocessed pixel and the pixel itself, in the 3*3 area centered at the given pixel. The started pixel set formed. It is P1（P={P| 0,1,}. pixels (2、3、4) is the addiable pixels. Then add the addiable pixels (2、3、4)to privous pixel set P.The new pixel set P2 ({P| 0,1,2,3,4}）formed.

2．search for next forked pixel or end pixels. Started from any a addiable pixel of P1, (2、3、4)within set P2 , search for all the addiable pixels(new addiable pixels be called)，which there are one or more than one pixels of nonzero graylevel except preprocessed pixel and the pixel itself within 3*3 neighborhood area, add the new addiable pixels to P2{P| 0,1,2,3,4}）,and P3{P| 0,1,2,3,4}）is formed. Repeated this program until there is no addiable pixel can be added.At this moment, the addiable amount centered the pixel are all equal 1,the last addiable pixels are forked pixel or end pixels. △,devoted addiable pixel amount。

3 localization of junction. We define the started value of coordination is on the left-up cornor. There are two methods to localizate of junction.

Method 1:use the value of forked or end pixels to calcaulate the coordination （xi ,yi）of intersect pixel(or junction). this measure is given by

$$X = \frac{1}{n}\sum_{i=1}^{n} x_i$$
$$Y = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad (1)$$

Where X ,Y represent localization of junction.

Method 2: use the pixels coordirate with greatest value of total addiable pixel amount to calcaulate the coordination （xi ,yi） of intersect pixel(or junction). Because the total addiable pixel amount means that the probablility for a pixel is the most to become a intersect pixel .this measure is given by (1).

The detection method for line junctions described in this paper was tested using real images containing L,T,Y,and X line junctions. Test show that the response to a given junction is unique and localization is accurate.

From above analysis,we can concluded that robust algorithm can effectively avoid incertitude and improve the spacial data qualities in the process of spacial data. a lot of cuccessful strategy to manipulate incertitude lie in the realization of robust algorithm.It is necessary to speed time on studing good algorithm of spacial data so as to control spacial data qualities.

3.3 The Basic Principle of Spacial Data Incertitude Measure

Three basic principles of uncertainty were developed to guide the use of uncertainty measures in different situations(David Harmanec,; Klir,1988,1995). Of course ,it can also be used as the basic principle of spacial data incertitude measure. These principles are: a principle of minimum uncertainty, a principle of maximum uncertainty, and a principle of uncertainty invariance.

The principle of minimum uncertainty is an arbitration principle. It guides the selection of meaningful alternatives from possible solutions of problems in which some of the initial information is inevitably lost but in different solutions it is lost in varying degrees. The principle states that we should accept only those solutions with the least loss of information. This principle is applicable, for example, in simplification and conflict resolution problems. For some development of this principle see (C. Joslyn and G. J. Klir,1992).

The principle of maximum uncertainty is applicable in situations in which we need to go beyond conclusions entailed by verified premises. The principle states that any conclusion we make should maximize the relevant uncertainty within constraints given by the verified premises. In other words, the principle guides us to utilize all the available information but at the same time fully recognize our ignorance. This principle is useful, for example, when we need to reconstruct an overall system from the knowledge of (some) subsystems. The principle is widely used within classical probability framework (Christensen,1985; Paris,1994), but has yet to be developed in a more general setting.

The last principle, the principle of uncertainty invariance, is of relatively recent origin (Klir,1989,1990). Its purpose is to guide meaningful transformations between various theories of uncertainty. The principle postulates that the amount of uncertainty should be preserved in each transformation of uncertainty from one mathematical framework to another.

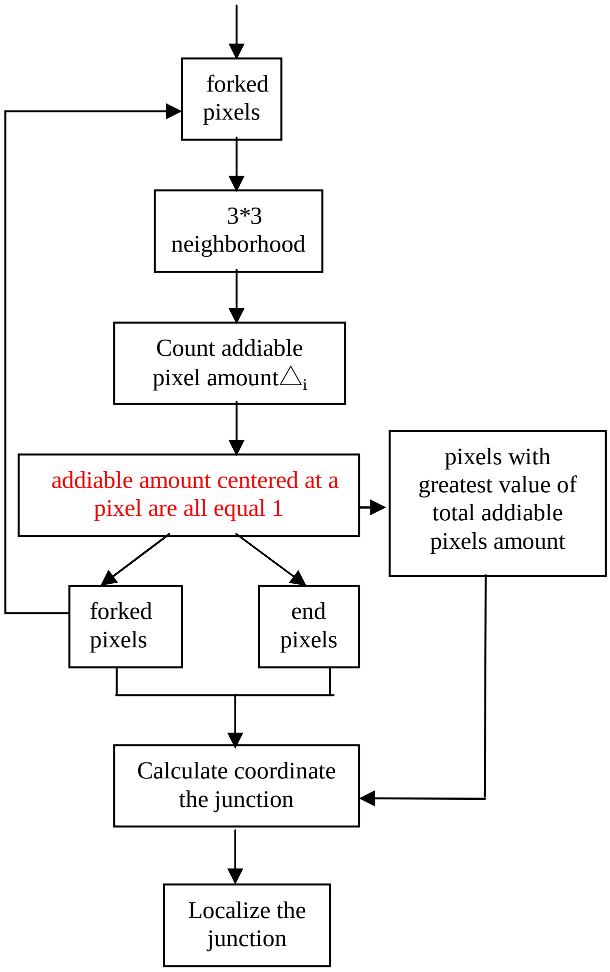

Figure 5. The pixels-track algorithm

Though important, the principles of uncertainty are not the only situations a measure of uncertainty can be used. Some other examples where the measures of uncertainty were used include measuring of closeness of possibility distribution(Higashi and Klir,1983) and investigating dynamics of combination of evidence.(Harmanec.,1997)

4. CONCLUSION AND DISCUSSION

In this article, we study mostly the incertitude of the spatial data, and the method of disposing the incertitude in the GIS and RS. In this paper, the inadequacies of the current incertitude control methods for spatial data were pointed out. Some strategies can be taken to manipulate incertitude according the spacial data incertitude resource . Robust algorithm can effectively avoid incertitude and improve the spacial data qualities in the process of spacial data, and spatial data incertitude can be control by studying robust algorithm. a principle of minimum uncertainty, a principle of maximum uncertainty, and a principle of uncertainty invariance can also be used as the basic principle of spacial data incertitude measure.

REFERENCES

C. Joslyn and G. J. Klir. Minimal information loss possibilistic approximations of random sets. In Proceedings of IEEE International Conference on Fuzzy Systems, pages 1081-1088, San Diego, 1992.

David W. S. Wong; C. Victor Wu，Spatial Metadata and GIS for Decision Support Proceedings of the 29th Annual Hawaii International Conference on System Sciences - 1996.

David Harmanec,summary measures of uncertainty, http://ippserv.rug.ac.be/documentation/summary_measures/.

D. Harmanec. A note on uncertainty, Dempster rule of combination, and conflict. *International Journal of General Systems,* 26(1-2):63-72, 1997.

G.J. Klir and B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice Hall, Upper Saddle River, NJ, 1995： 268.

G. J. Klir. The role of uncertainty principles in inductive systems modelling. Kybernetes, 17(2):24-34, 1988.

G. J. Klir. Probability-possibility conversion. In Proceedings of Third IFSA Congress, pp 408-411, Seattle, 1989.

G. J. Klir. A principle of uncertainty and information invariance. *International Journal of General Systems*, 17(2-3):249-275, 1990.

G. J. Klir. Principles of uncertainty: What are they? Why do we need them? *Fuzzy* Sets and S*ystems*, 74(1):15-31, 1995.

J. B. Paris. The uncertain reasoner's companion - a mathematical perspective. *Cambridge University Press*, 1994.

J.Serra, Image Analysis and Mathematical Morphology, *New york: Academic Press*, 1988.

M. Higashi and G. J. Klir. On the notion of distance representing information closeness: possibility and probability distributions. *International Journal of General Systems,* 9(2):103-115, 1983.

R. Christensen. Entropy minimax multivariate statistical modelling I: Theory. *International Journal of General Systems,* 11:231-277, 1985.

R. Christensen. Entropy minimax multivariate statistical modeling II: Applications. *International Journal of General Systems*, 12(3):227-305, 1985.

WKPratt, Digital Image Processing, New York: John Wiley & SonsJnc., 1991.