PROBABILISTIC ALGORITHMIC APPROACH AND GEOMETRY FOR CHECKING AND VERIFICATION ERROR IN GEORGRAPHICAL DATA

Kone Forokoro[1] & Xie Zhong[1, 2]
[1, 2] Faculty of Information Engineering, China University of Geosciences, Wuhan, China.
[2] Wuhan Zondy Cyber T & S Co, Ltd, Wuhan, China.
[1]**Email:** forocone@yahoo.fr

ABSTRACT
Geographical data, like any other data, are prone to errors. The management of spatial integrity constraints in geographic information system is made by special procedures and declarative mechanisms at priori to avoid registration errors of measurement. In this article we propose not to define special procedures or to preclude the presence of measurement errors in a geographic database, but to introduce an approach that allows reconstructing the actual spatial object from the object obtained in the field with errors of measurement. This geographical object is modeled in triangular meshing in the context of algorithmic geometry. So we use this formality of Bayesian networks for pattern recognition. Thus, constraint checking will submit to a learning process provided with explanatory information for the decision. The study is applied to a spatial planning.

Keys words: *Geographical Data, Measurement errors, Geographic information system (GIS), Triangular Meshing, Bayesian networks, spatial planning, Pattern recognition*

1. INTRODUCTION

Needless to say, any operations performed on data with errors (and that includes all data!) will also produce results that have a built-in error.
Checking the integrity constraints in a Geographic Information System (GIS) is to ensure data consistency with the world that the database models. This audit should lead to a consistent database.[1,2,3]
In a geographic database, we distinguish the following statements of consistency:

- Initial state before constraint checking
- Consistent state with the presence of isolated or redundant elements.
- Consistent state after removal or correction of these isolated or redundant elements.

We are interested in this article to develop an approach for the analysis of a GIS in order to detect the presence of measurement errors and suggest an approach to correcting the erroneous elements.
Indeed because of the presence of measurement error, some information stored in the database is different from their meaning in the real world. Error correction enables to ensure these data conformity to reality.
In general, to restore the consistency of the database, the recovery of the spatial object may use various techniques including pattern. In this regard, maintaining the database in a consistent state is guaranteed by the execution of the two learning tasks and decision systems for pattern recognition.
The decision theory aims to study the mathematical techniques used to make a decision in the presence of several alternatives. Different mathematical frameworks are possible to address this problem. We have for example Dumpster's, hidden Markov chains fuzzy set theory, probability theory or the use of Bayesian Networks.
We have chosen to use Bayesian Networks as this theory seems well suited to the spatial planning field. Indeed, the use of Bayesian Networks permits to model through the network structure the knowledge of an "expert". Moreover, they are more suitable to carry out actions during reasoning, which allows the introduction of perception strategies in the process of meshing process space.
This article is divided into five sections. The first paragraph is a brief introduction in which we present the problem of consistency of a database. In the second paragraph, we define integrity constraints in a database and particularly in a geographic database. We talk in general terms the process of pattern recognition and propose the formalism of Bayesian networks for the reconstruction of a spatial object. The third section presents the triangular meshing in a GIS. We use the Bayesian probabilistic method to reconstitute the modeling of a triangular mesh field for checking consistency. The fourth paragraph is about the presentation of results in the context of a simple example. We conclude our contribution at the fifth paragraph with an evaluation of studies

2. INTEGRITY CONSTRAINTS PATTERN RECOGNITION

2-1 Integrity Constraints

In databases, whether geographical or not, we must ensure that the stored information is consistent that is to say according to the reality it represents. Integrity constraints in geographic databases relate to the geometry and topology .They should be checked with the construction of the database or a subset of the database. For example, a polygon must be always closed, a segment should have two distinct points, a network must be connected so do not contain isolated branches, etc.

In our minds, to preserve the consistency of the geographic database will involve both validate the topology and verify accuracy.

2-2- Pattern Recognition

Pattern recognition can be defined as the set of computing techniques of representation and decision enabling machines to simulate behavior "sensitive". In pattern recognition, Bayesian networks allow to represent joint probability distributions of a set of variables using a set of prior knowledge on the relationship between these variables. A Bayesian network is a graphical model that encodes probabilistic relationships among variables of interest. When used in conjunction with statistical techniques, the graphical model has several advantages for data analysis. One, because the model encodes dependencies among all variables, it readily handles situations where some data entries are missing. Two, a Bayesian network can be used to learn causal relationships, and hence can be used to gain understanding about a problem domain and to predict the consequences of intervention. Three, because the model has both a causal and probabilistic semantics, it is an ideal representation for combining prior knowledge (which often comes in causal form) and data. Four,Bayesian statistical methods in conjunction with Bayesian networks offer an efficient and principled approach for avoiding the overt ting of data. Propagation in a Bayesian network is to propagate the network variables influence variables updated. Bayes' theorem (equation below) is the basis of the mechanism of propagation.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

3. MESHING AND CHECKING THE CONSISTENCY

3-1 Triangular Meshing

In our application, the field is consisted of triangles bounded by three segments; each segment is defined by its two end points, as well as by its two triangles that it bounds. For each end we define the coordinates x, y and altitude z. It is a representation called Wired and is expressed in relational form of three relations: R1 (# triangle # Segment1, segment2 #, # segment3) R2 (# segment, # point1, # point2, triangle1 #, # triangle2) R3 (# item, x, y, z)

3-2- Triangle Patterns

Patterns (or groups) of the triangle are the basic elements of our representation [4]. It is important to differentiate in the one hand the real pattern, that is to say triangles representing plots in the GIS, and in the other hand representation must be consistent with the data in the geographic database.

We use an encoding described in that combines two of the three segments of the triangle orientation discredited in eight sectors with respect to a reference line.

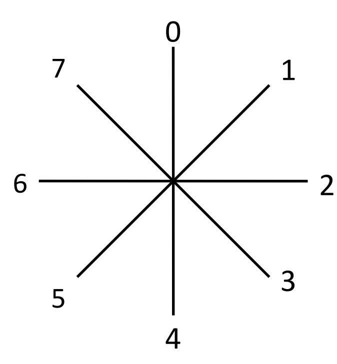
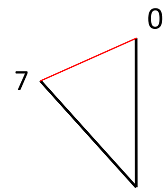


Figure 1: Coding junctions of the triangle

We deduce the pattern triangles from these junctions by connecting any two segments above. For example, joining segments [0] and [7], we get the triangle below:



Figiure 2 : An element of the pattern triangle

In fact, we obtain four pattern triangles that are as follow:
- Each pattern includes a rectangle triangle;
- Each pattern includes three triangles;
- The positioning segment determines the membership to a pattern

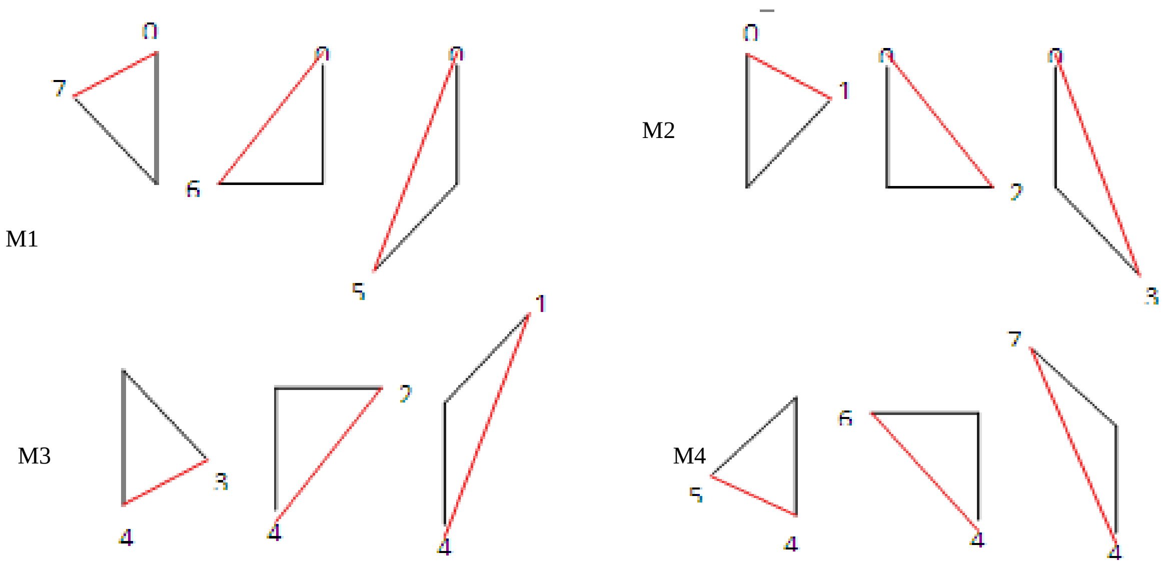


Figure 3: Pattern triangles: two segments of the triangle describes a clear direction

To these four patterns, we add the other pattern for triangles that cannot be classified. A triangle that has its two positioning segments [0] and [5], [0] and [6] or [0] and [7] belongs to the pattern m1. More generally, a triangle will be represented by the pattern "mi" when its two segments have defined positioning by one of the triangles of "mi."
In our application, we want to reconstruct a meshing of triangles from a point cloud. This networking already allows us in a first time to detect some errors on the objects points, lines or triangles of the geographic database. Previously detected errors will be corrected later during a guided field trip.

3-3- Description of the networking algorithm:
Let Ω be any domain representing the border of a certain domain and F (Ω) a set of edges discrediting its border. The vertex of these edges are a set of points denoted S (Ω).
Classically, to network a domain Ω we firstly construct a network bordering of Ω relying solely on the points S (Ω),then by adding the points inside the bordering network, constructing a new network that, after optimization give the final network of Ω. The inner points are defined in such a way that an inner network has a satisfactory length.
Nevertheless we limit our study at a polygonal scope Ω and the points S (Ω) are already known (points in the geographic database) and constitute the nodes of the polygon. The inner points I (Ω) are also present in the database; they are points obtained during the data capture. This allows us to get the diagram below:

- Creating the bordering meshing
  - Creating of a meshing T_0 (Ω) into which you insert the points S(Ω) or more precisely the edges that constitute the polygonal contour of the land.
- Creating the meshing T(Ω) of Ω.
  - Initialize the meshing T (Ω) by T_0 (Ω).

- loop insertion of points I(Ω :

choice of internal points T(Ω) according to a criterion of adjacency patterns triangle (Using the Bayesian network), inserting the selected points in T(Ω), Iteration if T(Ω) modified and all the points I(Ω) were not selected,

- Optimization T(Ω).
-

4- EXAMPLE:

We use MapInfo and Map Basic programming language to customize applications and develop interfaces. Here, we give a simplified example of a land model from [5,6]. This model represents the spatial object that we want to rebuild.

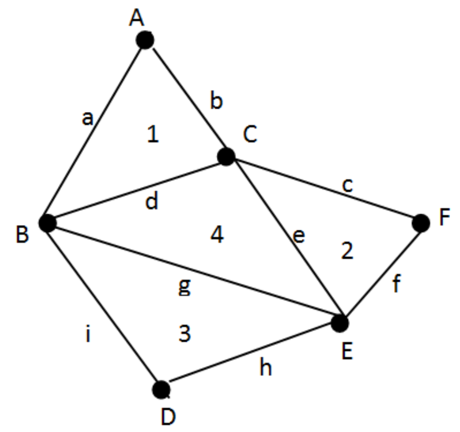


Figure 4: true value

We assume that in reality errors have slipped into the database and that we have stored information subject to errors.

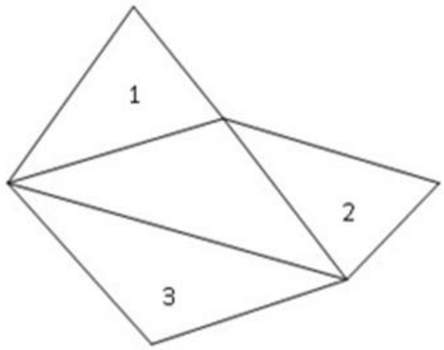


Figure5: triangle in database

Here are some measurement errors that contain the database:

- point A is unknown
- point C has changed
- the triangle 4 is not recognized

By applying the network algorithm of the cloud of points in the database, that is to say, the points with errors, we obtain the following meshing:

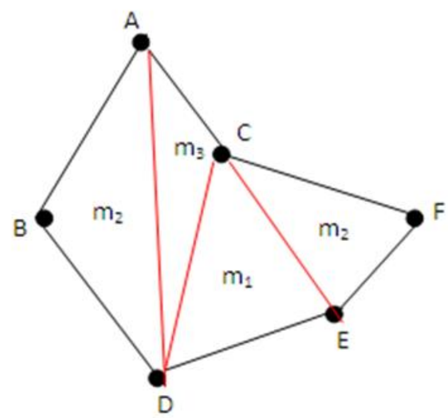


Figure 6: Geographical object reconstructed

Steps of the triangulation meshing

- The mesh in the West-east direction makes it possible to choose like first pattern for triangle, the pattern "m2".
- To the right of "m2", the only possible patterns are the patterns "m1" and "m3". The selected pattern is "m3 considering the orientation of the edges of the zone of interest of the image.
- The East of the pattern "m3", the point E makes it possible to obtain the pattern "m1".
- The last point not yet inserted in the mesh is the point F, and it provides the pattern "m2".

Steps of the detection of errors:

- The first pattern detected "m2" informs the segment I which was changed into H in triangle 3.
- The pattern m3 made it possible to inform the point C (which have a coordinate X unknown factor) and consequently all the segments of the base which have as an end the point C. In particular segment D.
- The pattern m1 in the third zone of interest of the image makes it possible to go up the point E whose coordinates were located out of the framework and to rectify it.
- The item I which does not appear as end in any relation segment is a wandering point and he is proposed for the correction. In short, with each reason for triangle detected in a zone, we associate various actions to check the coherence of the objects in this zone.

Correction of the errors:

The objects comprising of possible errors detected in a database are listed in a table in order to guide the manual corrections carried out by operators: it is the visit of ground.

5- CONCLUSION

In this article, we used the formalism of the networks Bayesians' to rebuild a real geographical object starting from an initial object contained in a SIG. The rebuilding of the object makes it possible to detect possible errors and a sequence of the constraints of integrity enables us to propose corrections with the technicians of the land register, with an aim of directing their visit. Our principal contribution was to define a methodology generally used in vision by computer to check the constraints of integrity in a GIS. We then try to visualize the errors graphically detected by our approach, thus facilitating the work of the operators. Our study related to the spatial planning, in particular on the modeling of the land in triangular meshing.

6- REFERENCE

[1]. Serge Milles , Jean Lagofun. Topographie et Topométrie Moderne, volume 1 - Techniques de Mesure et de Representation, Eyrolles (2000) , ISBN: 2212022875 , 526P .[online: http://www.ebookee.net/Topographie-et-topom-trie-moderne-volume-1-Techniques-de-mesure-et-de-r-pr-sentation_382415.html].

[2]. Serge Milles Jean Lagofun ,Topographie et topométrie moderne, volume 2 : Calculs , Eyrolles (2000) , ISBN: 2212022891 , 332P [online: http://www.book123.net/______Topographie-et-topom%C3%A9trie-moderne-volume-2-Calculs.html].

[3]. Harold Reynolds, An Introduction To Geographical Information Systems (GIS), December 18, 1997 [online http://www.badpets.net/IntroGIS/GIS_Intro.pdf].

[4]. Christopher Gold , Spatially Ordered Networks and Topographic Reconstructions , Lawrence town N.S., Canada BOS 1MO, Int. J. Geographical Information Systems, 1987, Vol. 1, NO. 2, 137 148 . [online:http://www.voronoi.com/wiki/images/0/0d/Spatially_ordered_networks%26topographic_reconstructions1987.pdf].

[5]. Wise, Stephen, GIS Basics(2002), London: Taylor & Francis Routledge **lcc:** G70.212.W57 2002eb**, ddc:** 910.285.

[6]. J. Ronald Eastman , Guide to GIS and Image Processing Volume 2, May 2001. Clark Labs ,Clark University 950 . [online: http://gisak.vsb.cz/~pen63/DPZ_pro_geodety/Guide2.pdf].