Proceedings of the 6th AGILE
April 24th-26th, 2003 – Lyon, France

PIECEWISE REVISION FOR GEOGRAPHIC INFORMATION

Mahat Khelfallah[1], **Robert Jeansoulin**[1], **Geoffrey Edwards**[2]

[1] Laboratoire des Sciences de l'Information et des Systèmes (LSIS), Université de Provence (Aix-Marseille I), 39 Rue Joliot Curie, 13453 Marseille Cedex 13, France.

[2] Centre de Recherche en Géomatique (CRG), Pavillon Casault, Université Laval, Québec, G1K 7P4, Canada

1. INTRODUCTION

How to reason with uncertain data?

This general question is often asked in geomatic applications where data, acquired by observations on the real world, present many imperfections. To reduce these imperfections, we can act directly by "checking" all or a part of data, i.e., by repeating observations with better quality (at least on samples) : it's costly if the sample is important. The second approach consists on "validating" data in respect with a model representing a behaviour the data must satisfy : for instance, integrity constraints (e.g. : being in an interval of values), or binary relationships (e.g. : topological relationships). Such approach needs two steps :

consistency checking, generally done by a logic program (e.g. : Prolog),

modification of inconsistent elements (data or rules, ...). This operation is called "revision" and aims to restore the consistency.

Belief revision has motivated many research works [3,6,7,10,11,12,16] in the field of artificial intelligence, especially after Alchourrón, Gärdenfors et Makinson [1] have stated the famous AGM postulates. However, the proposed methods have generally a big computation complexity, the reason for which their use in real applications is very limited, a fortiori, in geomatic applications which handle very important quantity of data. The solution to this problem is to find reasoning algorithms as efficient as is allowed by actual research in the domains of artificial intelligence and geographical information systems.

Our work is in the framework of geographic information revision. We are interested in finding efficient revision methods which take into account the characteristics of geographic information, in particular the property data containment. To do this, we think that this property can be exploited to treat locally sets of information, by decomposing the initial set of data into several parts. These parts are not completely independent and they mutually overlap. The decomposition method depends strongly on information structure as well as the notion of containment.

This paper is organised as follows. In the second section, a brief definition of belief revision is given. We show that the current methods are generally costly in terms of computing time and are often unusable for real-world applications.

In the third section, the characteristics of geographic information, namely imperfection, size of data and data containment, are presented. In the fourth section, we detailed the notion of "locality" as the property of the containment of error influence, and therefore, the possibility to reason locally on data. In the fifth section, we show how those characteristics are exploited in a method based on piecewise revision. A related work follows and a conclusion is given at the end.

2. BELIEF REVISION

Many researches have been done on the topics of belief change [1,3,6,7,10,11,12,16]. Change represents an evolution of the world in the case of update, or an evolution in our perception of the represented world. This change is the consequence of the addition of one or possibly many new information considered as more reliable (than the information we have). The new information will be used to "correct" the last one.

The addition of the new information to the belief base (or set), which was initially consistent, often leads to inconsistencies. Revision aims to restore the consistency by removing some beliefs from the initial base. The identification of such beliefs is not easy, and the choice between different obtained sets (of beliefs to remove) depends on the undertaken revision strategy. There exists many strategies, for instance those based on minimal change as in [1], or based on priority on beliefs as in [3] (where the choice to remove beliefs is dependent on the order that was established on beliefs. The order can be total [3], partial [5], or local [4]).

The construction of revision operators is generally very cautious [11], especially in real-world geographical applications [15,16].

3. GEOGRAPHIC INFORMATION CHARACTERISTICS

Geographic information has properties which can be exploited for reasoning in real-world applications. The characteristics we are interested in in this paper, are :

5. the geographic data imperfection, which justifies theoretic use of revision ;
6. the important volume of data, which makes difficult the effective treatment on them;
7. the "local containment" of quality influence, which could make possible to have efficient algorithms.

3.1 Imprecision and imperfection of data

Geographic information is generally imprecise and imperfect because it comes from different sources, and at different granularities, and with different points of view. According to Worboys [14], imprecision is an incomplete specification or representation of the object, whereas imperfection is the consequence of limitations on information collect means, measure means, etc. Moreover, the data can be incomplete over space and time.

The limitations on data quality are important and have different aspects : imprecision on various attributes (geographic, thematic), uncertainty on affectation in some class, incompleteness on collected observations (spatial or temporal coverage). These aspects are studied in other strands of the REV!GIS project, but we focus here on the logical consistency.

On the other hand, sharing the same spatial reference leads to several obligatory relationships between data : topology, distance, ... whose satisfaction is also measured with uncertainty. But geographical space has many of such relationships and the data validation according to these relationships is an important challenge in the geomatics domain.

3.2 Data size

Manipulating geographic information is synonymous with big size of data to handle. This makes difficult, and in certain cases impossible, the use of different algorithms.

3.3 Local containment of quality problems

The notion of « *local containment* » of the quality expresses the following phenomenon, and we assume that it is widespread in geomatic applications :

> « if some bad quality data has no influence beyond a certain limit within a data subset, then it will have no influence outside the data subset ».

There are two implicit properties in this phenomenon :

- "dampening of error effect", in other words its influence decreases according to the distance ;

- "connexity in error effect", i.e., there is not (important) possible "jump" in the influence of this error, unless other errors make impossible their observation.

These properties do not hold in general constraints propagation systems, but we think that they do in mostly all geographical applications.

This containment represents a limitation on the quality effects in the sense that an error cannot have, indefinitely, an influence on other information, but within a relative "local effect area" which depends on the kind of data. For instance, in meteorology, it's limited to continental scale, in cadaster application, to district scale, ... It depends also on information topology and constraints on this topology.

4. ERROR CONTAINMENT AND LOCAL REASONING

4.1 Definition and illustration of error containment

We take as an example a flooding application [13]. It consists in finding water heights in flooded parcels from initial estimates on water levels and observations on flows between parcels. Data are represented by constraints and the given solution is based on constraints propagation. The propagation was performed from upstream to downstream of the flooded valley.

The first hypothesis (ignored here) is that the "effect" of the phenomenon in this application is limited by the propagation time of the flooding front. Beyond this limit (a few kms), data become almost independent for the time period during the analysis (from one to a few hours). This size is illustrated below (Fig. 2): an error on the height of the water for a parcel in the upper part will not yield a conflict with an estimation in the lower part of the figure, while keeping fully consistent with all intermediate estimations !

Fig. 2 delineation of the flooded area over the aerial picture, and trhree sampled parcels

The second hypothesis is that the error (coming from a bad observation or a lack of observation) can cause with other observations, an inconsistency which can be locally detected : the area which contains the involved information set is spatially limited (few hundreds meters to one km). In other words, a measure error on a parcel situated in the upstream has not influence on the consistency of the downstream values.

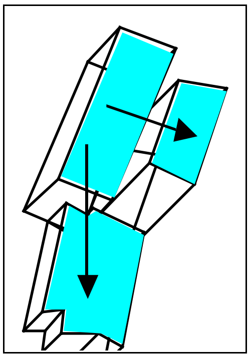

Fig. 3 hydraulic consistency checking: the down arrow is consistent with the estimated heights, but not the right arrow (no “upstream” flow).

In this example, the term "consistency" means that the hypotheses on water height of a parcel have to be consistent with water height of its neighbours according to flow direction observed between them. Fig. 3 shows the estimations and the flows, and this illustrates the expected constraint between these informations.In the considered application, the confidence in an information about flows is very high : if we have to change values, we will change interval bounds ones. For every detected inconsistency, many bounds may be changed : this is the choice a "good" revision operator has to do.

4.2 Counter-examples and further axioms on containment

Fig. 4 shows how two local models checked independently can lead to an inconsistency if they are merged. Horizontal boxes represent the interval of estimated water height for each parcel. Flows, not represented in the figure, are oriented from right to left. The error (not known) on the most left parcel does not cause any inconsistency in the model L (left). The model R (right) is consistent also. But merging the two models shows the inconsistency by propagation.

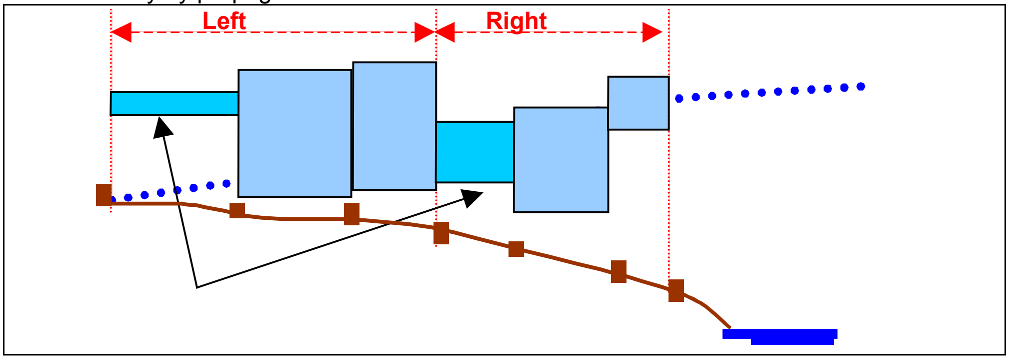

Fig. 4 flow follows the real water slope (unknown): the models L (left) and R (right) are both individually consistent, but their union is inconsistent. The cause is the conflict between the parcels pointed by the arrows.

This example proves that knowledge fusion is more complex than simple set theoretic union. It proves also that without any more restrictive hypothesis on data, an error influence can be discovered only in the other end of the data set : there will be no containment, only global model will be satisfactory.

So local models may be useless !!

But we can learn some lessons from this counter-example : if the error influence is "hidden" in the propagation through many parcels (n° 2 and n° 3 in the figure), it's because of the high uncertainty persistence, clearly higher than for parcels n°1 and n° 4 which cause the conflict. This is the kind of situation which normally does not occur all along the complete dataset : the error distribution must respect some spatial containment axioms, because the observation process behind it is not random.

4.3 Local containment and local reasoning

Containment property can be exploited in local reasoning which is reasoning on possibly many parts of the problem, instead of the entire problem.

The motivation of using such method is to reduce the data set to manipulate and hence to reduce computing time of applied algorithms.

In other cases, the goal is to make applicable algorithms to real-world problems.

We are interested in one sort of reasoning which is revision. We think that information about local containment of data can be used to develop a piecewise revision method, which handles locally parts of the entire problem.

5. PIECEWISE REVISION

Generally, the addition of a new information to a belief base causes inconsistencies. Revision restore the consistency by removing some beliefs in the base. In practical terms, revision has the following steps :

- Adding the new information : if the new information is consistent with the initial base, it is not a revision case but a simple case of adding information.
- Checking inconsistencies : this step consists on applying, according to the problem representation, a method for checking consistency. For instance, if the

problem is described by constraints, algorithms based on constraints propagation are used. One the other hand, if the used presentation is propositional calculus or first order logic, SAT algorithms are appropriate.
- Identifying beliefs involved in detected inconsistencies : this step can be merged with the previous one. Information used in the detection of inconsistencies would be retracted.
- Selecting beliefs set to be retracted : the choice of the beliefs to remove depends on the revision strategy, which is based generally either on minimal change principle [1,11,12,16] or priority [3,5].

These steps deal with all the application data whereas they can handle only parts of the problem, in such a way to reduce computation time which is often very big by decomposing the initial problem on many overlapping parts or pieces. These overlaps represent links existing between different parts of the problem.

One decomposition criterion (proposed by Amir et McIlraith [2]) is to minimize the following parameters :
- the size of each component,
- the number of shared symbols/constraints between two components (according to the representation),
- components number.

This decomposition takes into account information structure and information related to the geographic data containment.

At present, piecewise revision is not well formalised. However, its steps are defined as follows :
- to divide the initial problem into several sub-problems or pieces. This devision takes into account information structure and properties,
- to revise each piece of the problem by applying a revision algorithm, and then
- to merge the obtained results.

It's clear that the last step is the hardiest to develop because of the dependance existing between the different pieces of the problem.

6. RELATED WORK

There are many papers in the literature in the framework of partition-based and local reasoning. Among research works, we cite Amir and McIlraith's ones [2] on partition-based reasoning. They proposed an algorithm for reasoning on problem partitions in the framework of propositional calculus and first order logic, based on messages passing. These messages represent information related to shared symbols between the problem pieces. They also proposed a systematic algorithm for partitioning the initial problem into different pieces.

Benferhat [4] has worked in local reasoning for stratified belief bases. However, the method they proposed is not based on decomposition strategy as the previous one.

In the domain of CSPs, local reasoning arouses interest of researchers the last ten years. We can cite Dechter and Rish [8,9] who worked on constraints network decomposition.

7. CONCLUSION

In this paper, we make the postulate that the « containment property » of the quality of data holds for most applications, because of some *regularity* in the observation process. We show the necessity of discovering and exploiting this « containment property » in the framework of geographical applications. This is motivated by the big complexity, theoretic as well as practical one, of revision algorithms and the huge size of data to handle in geographical applications.

We describe a solution to this problem which exploits the containment property of geographic data to decompose the initial problem into sub-problems or pieces which we can treat locally. The decomposition strategy being related to the information structure and its effects.

Many domains have to be explored to further results. Belief revision, reasoning, CSPs domains are necessary to find formal tools we can use in our method. On the other hand, the geographical information field is needed for representing and exploiting the quality informations and their spatial behaviour.

8. ACKNOWLEDGEMENT

This work is presently supported by the European Communities, under the IST-1999-14189 project *REV!GIS* of the « Future & Emerging Technologies » programme. Also with the funding of the Ministère de la Recherche du Québec.

9. REFERENCES

[1] C. Alchourrón, P. Gärdenfors, D. Makinson. On the logic of theory change : Partial meet functions for contraction and revision. Journal of Symbolic Logic, 50:510-530, 1985.

[2] E. Amir, S. McIlraith. Partition-based logical reasoning for first-order and propositional theories. A paraître dans Artificial Intelligence, 2002.

[3] S. Benferhat, D. Dubois, H. Prade, M.A. Williams. A practical approach to revising prioritized knowledge bases. In Studia Logica, 70(1), pp 105-130, 2002.

[4] S. Benferhat, L. Garcia. Handling locally stratified inconsistent knowledge bases. Studia Logica, 70:77-104, 2002.

[5] S. Benferhat, S. Lagrue, O. Papini. Reasoning with partially ordered information in a possibilistic logic framework. In Proc. of the 9th Intl Conf on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'02), pp 1047-1052, 2002.

[6] M. Dalal. Investigations into a theory of knowledge base revision : preliminary report. In Proc of the 7th National Conference on Artificial Intelligence (AAAI'88), pp 475-479, 1988.

[7] A. Darwiche and J. Pearl. On the logic of iterated revision. Artificial Intelligence, 89:1-29, 1997.

[8] R. Dechter and J. Pearl. Tree clustering for constraint networks. Artificial Intelligence, 38: 353-366, 1989.

[9] R. Dechter, I. Rish. Directional resolution : the Davis-Putnam procedure, revisited. In Proc. of the 4th Intl Conf. on Principles of Knowledge Representation and Reasoning (KR'94), 134-145, 1994.

[10] L. Giordano, V. Gliozzi and N. Olivetti. Iterated belief revision and conditional logic. Studia Logica, 70:23-47, 2002.

[11] B. Nebel. Base revision operations and schemes : semantics, representation and complexity. In Proc. of the 11th European Conference on Artificial Intelligence (ECAI'94), 1994.

[12] O. Papini. A complete revision function in propositional calculus. In Proc. of the European Conference on Artificial Intelligence (ECAI'92), 339-343, 1992.

[13] D. Raclot and C. Puech. Photographies aériennes et inondations : Globalisation d'informations floues par un système de contraintes pour définir les niveaux d'eau en zone inondée. Marne la Vallée. Revue Internationale de Géomatique, 8:191-206, 1998.

[14] M. Worboys. Representing and reasoning with (imperfect) geospatial information. In Proc. of the 1st International Conference on Geographic Information Science (GIScience'00), 2000.

[15] E. Würbel, R. Jeansoulin and O. Papini. Revision : an application in the framework of GIS. In Proc. of the 7th International Conference on Principles of Knowledge Representation and Reasoning (KR'00), 2000.

[16] E. Würbel. A divide and revise method for geographical information. In ECSQARU'01 Workshop on Spatio-Temporal Reasoning and Geographic Information Systems, 2001.