A Logic-based Model for uncertainty reduction: parallel processing

Omar Doukari\textsuperscript{a}, Robert Jeansoulin\textsuperscript{b}

\textsuperscript{a}Géode lab., 31100 Toulouse, France (omar.doukari@univ-tlse2.fr)
\textsuperscript{b}Informatique lab., Institut Gaspard Monge, 77454 Paris, France
on leave at Embassy of France, Washington DC, USA (robert.jeansoulin@univ-mlv.fr)

Abstract:

Geographic Information Systems (GIS) often use incomplete and uncertain information leading to inconsistency. Thus, a clear definition of information revision is required. In logic, most of the usual belief revision operations are characterized by a high computational complexity, which implies actual intractability even for rather small data sets. On the other hand, GIS deal with large amounts of data, what means that revision is practically impossible. This challenge can be addressed only if a local belief revision strategy can be defined. In this paper, a new model for spatial information representation is defined, called G-structure model, and a local belief revision is designed, in order to answer, piece-by-piece, the problem of large data. This model complies with the principle of minimal change, as in any revision operation, by assuming the hypothesis that the size (spatial extent) of minimal inconsistencies is spatially bounded.

Keywords: Spatial Information, Uncertainty, Belief Revision, Local Processing.

1 Introduction

Geographic Information Systems (GIS) often deal with incomplete and uncertain information, Devillers and Jeansoulin [2006]. Data come from different sources, data quality is variable and uneven, they represent beliefs rather than facts, contradictions appear when mapped together and conflictual data require belief revision operations.

A Knowledge Representation operator is a rational agent that tries to interpret such a mix of perceptions and beliefs. As agents often face incomplete, uncertain, and inaccurate information, they need a revision operation in order to manage possible belief changes in the presence of new, or supposedly better information. Considering a certain "epistemic state" of an agent, represented by its reasoning process with his beliefs, then a belief revision consists in modifying the initial epistemic state in order to maintain or restore the consistency, while preserving the new -or better- information and respecting the principle of minimal change, Alchourrón et al. [1985]. It means minimizing the amount of beliefs to be changed for accepting in a consistent way the new information.

Unfortunately, in the general case, the theoretical complexity of revision is high, Khelfallah and Benhamou [2005]; Würbel et al. [2000]. More precisely, it belongs to the $\Pi_2$ class in the framework of propositional logic, Nebel [1998].

On the other hand, geographic information is characterized by large data and at first glance it seems to be no hope of performing revision in the context of GIS. Several authors, Khelfallah and Benhamou [2005]; Würbel et al. [2000], worked out revision op-
erations to take advantage of the spatial representation of knowledge, and to define heuristics speeding up the algorithms. Despite performance gains, limitations remain in the volume of knowledge handled by revision operations or the principle of minimal change is violated.

In this paper, we answer this issue of revising large spatial data sets by defining a representation model for spatial information and local belief revision, called the G-structure (Geographic structure) model. It respects the principle of minimal change by adding a condition on the size of minimal conflicts. Indeed, spatial data inconsistencies are supposed to be local, Doukari and Jeannoulin [2007], hence, a condition on the size of minimal conflicts seems quite realistic, and global revision can be made of series of local revisions, each one at a more tractable size.

After defining spatial knowledge revision in Section 2, section 3 is devoted to define the G-structure model. Section 4, based on a Reiter revision approach, Würbel et al. [2000], defines the local revision based on G-structures. Finally, in Section 5, we apply this operation on a real experiment, with real data. It consists of the flooded valley problem.

2 Preliminaries

Let \( L \) be a propositional language defined over some finite set of propositional variables \( \mathcal{V} \) and the usual connectors (\( \neg, \lor, \land, \rightarrow, \leftrightarrow \)). A literal is a variable \( v \in \mathcal{V} \) or its negation \( \neg v \). A clause is a disjunction (\( \lor \)) of literals and \( \mathcal{V}(\alpha) \) denotes the set of variables involved in the clause \( \alpha \). Similarly for a set \( X \) of clauses \( \mathcal{V}(X) = \bigcup_{\alpha \in X} \mathcal{V}(\alpha) \). Therefore, we use propositional logic to represent our spatial knowledge.

2.1 Revision of Spatial Knowledge

Data included in GIS are derived from many sources exploited through various methods and technologies. One of the great advantages of GIS is to integrate into a single base, data generated from different sources (\( S_1; S_2 \), etc.), but associated with the same portion of geographic space. The data sources used can be primary or secondary ones. In the former case, data are directly measured (by sampling the land or by interpreting images), whereas in the latter case, the data are obtained from sources such as existing maps, other databases, attribute tables, etc.

In our case, we consider a geographic space, denoted by \( \xi \), segmented into parcels. On each parcel, a set of attributes is defined, and each attribute is represented by a variable which is defined over a sampled domain. We have two data sources \( S_1 \) and \( S_2 \); giving us two sets of constraints related to these variables. These constraints can be expressed by a CSP\(^1\) problem, Jégou and Terrioux [2003], from which we generate: the set of variable domain description clauses, and the set of clauses representing the constraints related to the attributes defined on the parcels. For more details, see Würbel et al. [2000].

In the rest of the paper, we shall denote by \( S_1 \cup S_2 \) the set of clauses related to \( \xi \). If \( B \) is a subspace of \( \xi \), then \( \text{Var}(B) \) denotes the set of variables defined in \( B \), and \( SB(B) \) denotes the subset of clauses related to \( B \), i.e., \( SB(B) = \{ c | c \in S_1 \cup S_2, \text{ and } \mathcal{V}(c) \subseteq \text{Var}(B) \} \).

The "revision problem" for spatial knowledge can be expressed as follows. Let \( S_1 \) be a finite set of clauses representing initial knowledge related to \( \xi \), and \( S_2 \) be a second set of clauses related to \( \xi \), more reliable than \( S_1 \). Furthermore, \( S_1 \) and \( S_2 \) are consistent.

\(^1\)Constraints Satisfaction Problem.
If the union $S_1 \cup S_2$ is consistent, then the revision of $S_1$ by $S_2$ consists in adding $S_2$ to $S_1$. Otherwise, we need to keep $S_2$ and removing the least possible information from $S_1$ in order to maintain consistency. This can be interpreted by the identification of minimal subsets $R$ of $S_1$ such that $(S_1 \setminus R) \cup S_2$ is consistent, Papini [1992]. The approach that naturally comes into mind consists of three steps. First, to calculate the set of all minimal inconsistent subsets (MISs) of $S_1 \cup S_2$. We define formally a MIS as follows:

**Definition 1** Let $M$ be an inconsistent subset of $S_1 \cup S_2$. $M$ is a MIS in $S_1 \cup S_2$ iff $\forall M' \subset M$, $M'$ is consistent. We denote by $I(S_1 \cup S_2)$ the set of all MISs in $S_1 \cup S_2$.

Second, to construct the minimal sets $S$, called minimal hitting sets (MHSs) of $I(S_1 \cup S_2)$, such that: (i) $\forall M \in I(S_1 \cup S_2)$, $S \cap M \neq \emptyset$, and (ii) $\forall S' \subset S$, $\exists M \in I(S_1 \cup S_2)$ such that $S' \cap M = \emptyset$. The REM revision approach introduced in Würbel et al. [2000], presents a detailed description of the computation of MHSs. REM$(S_1 \cup S_2)$ computes the set of MHSs of $I(S_1 \cup S_2)$ denoted by $N(I(S_1 \cup S_2))$.

Finally, it remains only to choose which set $S$ to pick to restore consistency. Generally, in this last step, the domain’s experts who choose one MHS among the subset of MHSs that do not intersect with $S_2$ (i.e., $S \cap S_2 = \emptyset$).

### 3 The Geographic structure model

The idea behind the G-structure model is that an inconsistency has a finite influence on other information and this influence is limited spatially by a certain distance. This intuition is introduced for the first time in Doukari and Jeansoulin [2007] through a new property called the “containment property”.

To guarantee minimal change, tractability, and by taking into account the locality of inconsistencies intuition, we will define a new model called G-structure model which split up the geographic space into a set of G-cores and each G-core has a set of parcels surrounding it called its G-covering. This concept of G-covering expresses the notion of local inconsistency and allows some degree of overlapping between the different subspaces.

Before defining these two concepts (G-core and G-covering of G-core), we should order the parcels composing $\xi$ by the following spatial adjacency relation using the external connection relation between parcels.

**Definition 2** Let $p_i, p_j$ be two parcels in $\xi$. We define the relation of spatial adjacency $R^m$ ($m \geq 0$) as follows: (i) $R^m(p_i, p_j)$ iff $\exists p_1, p_2, \ldots, p_m \in \xi$ s.t., $R^0(p_i, p_1)$, $R^0(p_1, p_2)$, $\ldots$, $R^0(p_{m-1}, p_m)$, and $R^0(p_m, p_j)$; and (ii) $R^0(p_i, p_j)$ iff $p_i$ and $p_j$ are “externally connected”. If $R^m(p_i, p_j)$, then $p_i$ and $p_j$ are called “neighbors”.

Distance between two parcels $p_i, p_j$, denoted by $dist(p_i, p_j)$, can be defined as follows.

**Definition 3** Let $p_i, p_j$ be two parcels of $\xi$. The distance between them is such that :

$$
\text{dist}(p_i, p_j) = \begin{cases} 
0 & \text{if } p_i = p_j \\
\min \{m | R^m(p_i, p_j)\} + 1 & \text{if } p_i, p_j \text{ are neighbors} \\
\infty & \text{otherwise.}
\end{cases}
$$

A G-core of $\xi$, a Set of G-cores of $\xi$, and the G-covering whose thickness is equal to $k$ for a G-core $B$ are given by Definitions 4 and 5 below :

**Definition 4** Let $p$ be an arbitrary parcel in $\xi$. A G-core $B$ with radius $r$ ($r \geq 0$) and center $p$ is s.t., $B = \{ p_i \in \xi | \text{dist}(p_i, p) \leq r \}$. $\{B_1, \ldots, B_n\}$ is called a Set of G-cores of $\xi$, such that $B_1, \ldots, B_n$ are G-cores with radius $r$ and different centers, iff $\{B_1, \ldots, B_n\}$ is a partition of $\xi$. 

Definition 5 A G-covering with thickness $k$, of a G-core $B$ with radius $r$ and center $p$ ($r < k$), denoted by $GCov_k(B)$, is s.t., $GCov_k(B) = \{ p_i \in \xi | r < dist(p_i, p) \leq k \}$.

Informally, the G-covering of a G-core $B$ represents the influence area of the MISs depending of $B$ (i.e., the MISs containning at least one clause in $SB(B)$).

In order to parametrize a G-structure by a particular thickness (as we will see later), we require the definition of the size of a MIS. Informally, if $C$ is a MIS, then its size is equal to the maximal distance between the parcels of the smallest subspace of $\xi$ containing all the clauses in $C$.

Definition 6 Let $C \in \mathcal{I}(S1 \cup S2)$ be a MIS. The size of $C$, denoted by $size(C)$, is s.t., $size(C) = \max\{dist(p_i, p_j) | p_i, p_j \in \xi, \mathcal{V}(C) \cap \mathcal{V}(\{p_i\}) \neq \emptyset, \text{and } \mathcal{V}(C) \cap \mathcal{V}(\{p_j\}) \neq \emptyset \}$.

The only assumption made by the G-structure model is that the maximal size of existing MISs in $\mathcal{I}(S1 \cup S2)$ is known. Hence, when we construct a G-structure $G$ on $\xi$, we only require that the thickness of G-coverings of G-cores (value of $k$) should be (at least) equal to the maximal size of MISs existing in $\mathcal{I}(S1 \cup S2)$. Informally, a G-structure is a set of structures such that each one of them is composed of two parts. The first part called subspace part which is a G-core plus its G-covering, and the second one is the subbase part which is a belief subbase attached to the subspace part.

Definition 7 Let $\xi$ be a geographic space. The set $G = \{(B_1, GCov_k(B_1), SB(B_1 \cup GCov_k(B_1))), ..., (B_n, GCov_k(B_n), SB(B_n \cup GCov_k(B_n)))\}$ is a G-structure of $\xi$ defined as follows: (i) $\{B_1, ..., B_n\}$ is a set of G-cores of $\xi$, (ii) $GCov(G) = \{GCov_k(B_1), ..., GCov_k(B_n)\}$ is a corresponding set of G-coverings with thickness $k$ s.t., $\forall M \in \mathcal{I}(S1 \cup S2), Size(M) \leq k$.

Clearly, the only area in which the set of MISs depending of a given G-core $B_i$ of a G-structure $G$, can be intersected by the set of MISs which do not depend of $B_i$ is the G-covering of $B_i$. This holds, of course, provided that the maximal size of MISs in $\mathcal{I}(S1 \cup S2)$ is limited by the thickness of G-coverings of G-cores of $G$. Consequently, if the MISs depending of $B_i$ do not intersect any MIS depending of the G-covering of $B_i$, then these MISs do not intersect any MIS existing outside $B_i$. As we will see later, this last consequence allows us to compute global MHSs (i.e., the MHSs of $\mathcal{I}(S1 \cup S2)$) by computing only a simple union of local MHSs (i.e., the MHSs of MISs depending of each G-core) without any need of minimality check.

To simplify the notation, in the following, we shall not specify neither the radius of G-cores, nor the thickness of the G-coverings. So, if $B_i$ is a G-core, we just put $GCov(B_i)$ to denote its G-covering.

4 G-STRUCTURE BASED REVISION

Let $G = \{(B_1, GCov_k(B_1), SB(B_1 \cup GCov_k(B_1))), ..., (B_n, GCov_k(B_n), SB(B_n \cup GCov_k(B_n)))\}$ be a G-structure of $\xi$. In this section, we present a thorough study of MISs and their processing locally in each structure of $G$.

4.1 Independent MISs

We have two kinds of independence between MISs: spatial independence, and informational independence. The former is a special case of the later such that all spatial independences are informational independences.

Space Independent MISs. In this case, there do not exist MISs depending of the G-core $B_i$ and, in the same time, depending of its G-covering $GCov(B_i)$. Hence, we
say that \( I(SB(B_i)) \) are space independent MISs. The computation of the MHSs of \( I(SB(B_i)) \), can be made independently of the remaining MISs. After this computation, we can ignore \((B_i, GCov(B_i), SB(B_i ∪ GCov(B_i)))\), since there will not be any MIS depending of the G-core \( B_i \). To verify if such an independence exists, and to compute the MHSs of \( I(SB(B_i)) \), we define a new procedure called \( REM-Local \). It has the same theoretical complexity than the \( REM \) procedure, but it processes less clauses.

**Definition 8 (\( REM-Local \) Algorithm)** The procedure \( REM-Local \) (\( C_1 \): set of clauses, \( C_2 \): set of clauses) is a slightly modified version of \( REM(C_1 \times C_2 \times E) \) such that it calculates the MHSs of the MISs in \( C_1 \) which contain at least one clause in \( C_2 \).

**Information Independent MISs.** In this case, there exist MISs depending of the G-core \( B_i \), in the same time, depending of its G-covering \( GCov(B_i) \). But the intersection between these MISs and the MISs in \( I(SB(B_i)) \) is empty. Hence, we say that \( I(SB(B_i)) \) is information independent MISs. The computation of the MHSs of \( I(SB(B_i)) \), can be made independently of the remaining MISs. Furthermore, after this computation, we ignore \((B_i, GCov(B_i), SB(B_i ∪ GCov(B_i)))\), because the MISs which are still depending of the G-core \( B_i \) are considered in another structure of \( G \).

**Processing Independent MISs.** The detection of the independence property, spatial and/or informational, between two sets of MISs, make the computation of the global MHSs of the union of the two sets of MISs easier. Thus, to calculate the MHSs of the two sets of MISs, we calculate them for each set of MISs independently, then we concatenate the resulting MHSs of the two sets of MISs (i.e., we build the union of the resulting MHSs), without any need of minimality check. Formally, we have the following result.

**Proposition 1** Let \( F \) be a set of MISs and \( E \) be the set of MHSs of \( F \). If \( F_1 = \bigcup_{i=1}^{n} F_i \) s.t., \( ∀F_i ∈ F_i, ∀F_j ∈ F_j, i ≠ j \) we have: \( F ∩ F' = ∅ \) \( (F_1, ..., F_n) \) are information independent two by two), then \( E = \{e_1 ∪ ... ∪ e_n|(e_1, ..., e_n) ∈ E_1 × ... × E_n \) s.t., \( E_1, ..., E_n \) are the sets of MHSs of \( F_1, ..., F_n \), respectively.

### 4.2 Dependent MISs

In this case, there exist MISs depending of \( B_i \), in the same time, depending of \( GCov(B_i) \). Furthermore, the intersection between these MISs and the MISs in \( I(SB(B_i)) \) is not empty. Hence, the MISs in \( I(SB(B_i)) \) are called dependent MISs.

**Processing Dependent MISs.** To process dependent MISs, we define a simple approach. This approach consists in computing the local MHSs of the MISs depending of the different G-cores of the G-structure \( G \). We process the structures of \( G \) one after one such that, when we come to treat the \( i^{th} \) structure (i.e., \((B_i, GCov(B_i), SB(B_i ∪ GCov(B_i)))\)) of \( G \), we proceed as follows. First, we compute the MHSs of the MISs in \( I(SB(B_i)) \) which contain at least one clause in \( SB(B_i) \). Second, we discard this structure from the list of structures of \( G \) to process, in order to avoid reprocessing the treated set of MISs. Then, we concatenate the computed sets of local MHSs with the set of MHSs obtained in the previous iterations. Finally, we check minimality to keep only the MHSs among the resulting ones. Clearly, the construction of global MHSs is made incrementally, at the same time than the computation of the sets of local MHSs related to the different structures of \( G \). We have proved the following result which validates this approach.

**Proposition 2** Let \( F \) be a set of MISs and \( \{F_1, ..., F_n\} \) is a partition of \( F \). If \( E_1, E_2, ..., \) and \( E_n \), are the sets of MHSs of \( F_1, F_2, ..., \) and \( F_n \), respectively. Then \( E \), the set of MHSs of \( F \), is s.t., \( E = \min E \{e_1 ∪ ... ∪ e_n|(e_1, ..., e_n) ∈ E_1 × ... × E_n\} \).

\(^{2}\text{Minimality in the sense of set inclusion.}\)
Algorithm 1 G-structure Revision

Require: $G$ - a G-structure of $\xi$

1: [First step for filtering independent MISs to improve the efficiency : omitted here]
2: $Conf \leftarrow$ the clauses involved in the treated MISs in the first step.
3: $GMH \leftarrow$ the set of MHSs treated in the first step.
4: $T_{\text{max}}$ max geographical extension of MISes of the collections $I(S_1 \cup S_2)$.
5: for all $(Bi, GCov(Bi), SB(Bi \cup GCov(Bi)))$ in $G$ do
6:     $TMH \leftarrow \emptyset$
7:     $MHB \leftarrow \text{REM} - \text{Local}(SB(GCov_{T_{\text{max}}}(B) \cup B) \setminus Conf, SB(B))$
8:     for all $gmh \in GMH$ do
9:         $TMH \leftarrow TMH \cup (gmh \cdot MHB)$
10:    end for
11:    if $(\exists tmh, tmh' \in TMH)$ and $(tmh \subset tmh')$ then
12:        to remove $tmh'$
13:    end if
14:    $GMH \leftarrow TMH$
15:    $G \leftarrow G \setminus \{(Bi, GCov(Bi), SB(Bi \cup GCov(Bi)))\}$
16: end for
17: return $GMH$

4.3 G-structure Revision Algorithm

From the previous sections, we define the G-structure Revision algorithm to calculate the MHSs of $I(S_1 \cup S_2)$. Firstly, it searches all structures in $G$ whose MISs are (information and/or space) independent (omitted part in Algorithm 1 for the sake of simplicity). Thereafter, processing the remaining structures of $G$ is made with respect to the dependence case of MISs. Finally, we concatenate the local MHSs resulting from the two previous steps to constitute the MHSs of $I(S_1 \cup S_2)$.

In Algorithm 1 the operator $\cdot$ is used in order to build all possible unions of a set with the elements of a collection of sets.

5 Real Application: The Flooded Valley Problem

We implement the G-structure Revision on a real experiment, with real data. It consists of an extensively studied application: the assessment of water heights in the flooded valley of the Herault river (southern France, 1994), Raclot and Puech [1998]. The flood of the valley was studied, using two information sources: (i) $S_1$ : assessments of water levels in the flooded parcels, using a priori knowledge of the vegetation height. (ii) $S_2$ : graph of hydraulic relations observed from an aerial picture. Quality of $S_2$ is more reliable than $S_1$. Hence, we would revise $S_1$ by $S_2$.

The structural splitting of the space adopted by Algorithm 1 allows us to decompose the revision problem into tasks that can be executed concurrently. Unified Parallel C (UPC) is an explicit parallel language that provides the facilities for direct user specification of program parallelism and control of data distribution and access, El-Ghazawi et al. [2003]. Therefore, we implement our algorithm in UPC so that it can be run locally and parallelly in a Single Program, Multiple Data (SPMD) mode, since any G-structure is composed of a set of structures with different data. The results are given in Table 1.

Clearly, we ameliorate computing time whenever we process each structure in parallel by an independent processor, even if at the local level, we execute the same program. This
is due to the reduced amount of information contained in each structure. With 0-radius G-cores, and 0-thickness G-coverings: i.e., each parcel is a G-core and its G-covering simultaneously. Thus, the sub-bases of this $G$-structure are disjoint. Then, no minimal hitting set is generated, because $S_2$ (hydraulic relations) is barely ignored. It confirms that $S_1$ is consistent, but this isn’t a revision solution. For ($r = k = 1$), the number of MHS is less than the number found with ($r = 0$, $k = 1$). Thus, the number of MIS we were able to process is also less. This is explained by the fact that we consider independent structures (that is, $r = k$). Indeed, some MIS, which lie between these two structures, are missed. The only cases that allow overlapping are ($r = 0$, $k = 1$), ($r = 1$, $k = 2$), ($r = 2$, $k = 3$).

In this application, the estimated $T_{\text{max}}$ value is 3, according to the spatial distribution of uncertainty errors. Therefore, if $T_{\text{max}} = 3$ is a postulate, the result obtained for ($r = 2$, $k = 3$) is the same as the result that we would have obtained with a global revision. The flooding problem has been "revised" within about 4 minutes, what would have been out of reach with a global approach.

### 6 Conclusion

Since in real applications, especially geographic ones, we process a huge amount of information, implementation of global revision operators is difficult and may be impossible.

To solve this problem, in this paper, we introduced a new model for spatial information representation and local belief revision. Then, based on this new model, we proposed a local revision strategy which answers to the problem induced by the large amount of
data and allows us to respect the principle of minimal change whenever inconsistencies are local. Finally, we implemented our approach using UPC language so that it can be run in parallel on a real experiment, with real data, the flooded valley problem, which we succeed to process, correctly and completely, though the global revision always failed.

The preliminary results obtained are quite encouraging. Indeed, even if the flooding problem was previously intractable by the global (or traditional) revision operation, our results show that the application falls into a tractable case.

There are two ways to improve our results. First, the algorithm generating the minimal hitting sets uses a procedure which generates arbitrary inconsistent subsets which are not necessary minimals. We think that this can be dramatically improved by the use of recent results in the domain of SAT solvers. Second, the parallelism introduced in this paper consists only in treating independently the structures of $G$-structure. It could be also integrated in calculating the minimal hitting sets of each structure.

It should be noticed that it is not always possible to make an assumption on the maximal size of MISs. Actually, this question is still dependent on the application and its domain.

REFERENCES


