Fusing uncertain structured spatial information*

Florence Dupin de Saint-Cyr, Robert Jeansoulin, Henri Prade
IRIT, CNRS, Univ. Toulouse III, 31062 Toulouse Cedex 09, France.
Institut Gaspard Monge, CNRS, Univ. Paris Est, 77453 Marne-la-Vallée, France.

Abstract

Spatial information associates properties to labeled areas. Space is partitioned into (elementary) parcels, and union of parcels constitute areas. Properties may have various level of generality, giving birth to a *taxonomy* of properties for a given universe of discourse. Thus, the set of properties pertaining to a conceptual taxonomy, as the set of areas and parcels, are structured by a natural partial order. We refer to such structures as ontologies. In fusion problems, information coming from distinct sources may be expressed in terms of different conceptual and/or spatial ontologies, and may be pervaded with uncertainty. Dealing with several conceptual (or spatial) ontologies in a fusion perspective presupposes that these ontologies be aligned. This paper introduces a basic representation format called *attributive formula*, which is a pair made of a property and a set of parcels (to which the property applies), possibly associated with a certainty level. Uncertain attributive formulas are processed in a possibilistic logic manner, augmented with a two-sorted characterization: the property may be true *everywhere* in an area, or at least true *somewhere* in the area. The fusion process combines the factual information encoded by the attributive formulas provided by the different sources together with the logical encoding of the conceptual and spatial ontologies (obtained after alignment). Then, inconsistency encountered in the fusion process may be handled by taking advantage of the existence of different fusion modes, or by relaxing when necessary a closed world-like assumption stating by default that what is true somewhere in an area may be also true everywhere in it (if nothing else is known). A landscape analysis toy example illustrates the approach.

Key words: spatial information, ontology, uncertainty, possibilistic logic, fusion.

*This is a draft version, the article was published in: International Conference on Scalable Uncertainty Management (SUM 2008), Naples (Italy), October 2008, Springer, LNAI 5291, p. 174-188. This work was funded by the Midi-Pyrénées and Provence-Alpes-Côte d'Azur Regional Councils (Inter-Regional Project n^o 05013992 "GEOFUSE"). A preliminary version is electronically available [6].

1 Introduction

The management of multiple sources raises many fusion problems due to the uncertainty and the heterogeneity of the information. Geographical information has all these problems [4, 2, 12, 18], its specific aspect being to deal with geographical-space areas, called *parcels*, on which we need to operate union and intersection.

A popular representation is the so called "field model", $f{:}(x, y) \to f(x, y)$, with a Cartesian coordinate domain as space, and real numbers as target domain. Though widely used in geophysics, meteorology, etc. and in most applications that involve imagery, terrain or any "gridded data", it is much too limited in many situations that deal with non quantitative data. Landscape analysis is one such situation. Spatial information may involve a mix of numeric and symbolic attributes, using different vocabularies more or less structured, but rarely unstructured. The sources may use different space partitions. Moreover, there may exist several kinds of dependencies, and spatial fusion must keep consistent with all of them. A previous paper [7] started an informal discussion of these problems. The present paper provides a logical framework for handling spatial information and ontological information. Another step is made by handling the merging of spatial information in the general setting of logical information fusion. Lastly, both numeric and symbolic information may be pervaded by several forms of uncertainty and imprecision [14]. This is why we allow for "uncertain attributive formulas" linking parcels to a property associated with a certainty degree: it expresses that for *any* parcel of a given set, we are sure at least at this degree that a property is true.

Hence, dealing with spatial data requires relatively powerful representation languages, as discussed in [15]. Ontology is often used for representing a structured vocabulary [12], and the fusion of ontology-based geospatial information must face the problem of heterogeneous vocabularies [10]. This paper deals with terminology integration and discusses the merging of information provided by different sources using multiple space partitions, and expressed with more or less precise labels from the same ontology resulting from a preliminary alignment.

Following Papini *et al.* [23], we use a logical framework for processing ontological information, and "attributive formulas" to link sets of parcels to property or attribute statements. We need a simple type of ontology that can be logically expressed by three and only three conditions: 1) a label may be a sub-label of another label, 2) a label is the reunion of its sub-labels, 3) labels referring to the most specific classes are mutually exclusive two by two. This representation allows us to express both ontological information and attributive formulas. Besides, the spatial extent on which an attributive formula applies may vary within a parcel: it means that we must distinguish between statements true everywhere, or only somewhere in a parcel.

The paper is organized as follows. Section 2 discusses representation needs, proposes a logical formalism for representing geographic information in ontologies, and introduces the notion of an *attributive formula* as a reified formula that links space and labels. Section 3 details the fusion process that helps to merge heterogeneous descriptions of the same space. In Section 4, "uncertain attributive formulas" are defined, and we introduce the explicit precision of the "somewhere" or "everywhere" reading associated to an attributive formula. Section 5 shows how to integrate possibilistic principles in the context of "attributive formulas". It is illustrated on a landscape information fusion example.

2 Geographic ontologies and attributive formulas

In *geographic information* we can distinguish the *geo* part, the *info* part, and the association that links them (the *what*, the *there* and the *is*, of Quine[20]). Hence, three aspects should be considered for representing geographic information:
1) the *(attributed) space*: one single space for all applications, but many different ways to split it into parts. *Parcels* have a spatial extent, and it is assumed that after intersecting all parcels from the different splittings, the most elementary parcels form a finite partition of the space. This is called a *partonomy structure*.
2) the *(attribute) properties*: many *property domains*, more or less independent, can serve different purposes. A *taxonomy structure* can represent a hierarchy of properties, reflecting some partial order. A consistent fusion of partial orders may help to detect, and to remove errors when mixing such structures.
3) the *attribution*: it results from an observation process, where the associations are often multiple, and largely pervaded by uncertainty for space and properties.

A similar, but not formalized, approach was proposed in [17]: *an ontology is built on three main concepts: (1) a partonomy of physical objects of which the attributes represent most of the relevant information, (2) a simple taxonomy of informational objects, (3) a relation between the informational objects and those physical objects they inform about.* In order to have a representation model more appropriate than the "field model", we use a logical "attributive formalism" to represent "property-parcel" information. Beside the attributive link, there are two other basic links: *property-property* (from the knowledge encoded in a property taxonomy), and *parcel-parcel* (from a partonomy). The logical representation is satisfactory for encoding such qualitative links too. The ontology representation we use is simpler than the ones offered by description logics since we remain propositional. The ontological relations are not uncertain here.

2.1 A logical encoding of an ontology of information

In fusion problems, it is advantageous to encode taxonomies in a logical manner, which makes the information merging easier. Let $\{\langle set\ of\ nodes\rangle, \subseteq\}$ be a *poset* structure that we name *ontology* [22], where nodes are concepts, and $\subseteq$ encodes specialization/subsumption relations: these relations are represented graphically by edges where arrow direction refers to generalization. Let $\mathscr{L}$ be a propositional logical language built on a vocabulary $\mathscr{V}$ with connectives $\wedge$, $\vee$, $\rightarrow$ ("and", "or", material implication).

Definition 1 (poset definition of an ontology) *An* ontology *is a directed acyclic graph (dag) $G = (X, U)$. $X \subseteq \mathscr{L}$ is a set of formulas (one per concept, or node); U is a set of directed arcs (φ, ψ) denoting that φ is a subclass of ψ. An ontology admits one single source, $\bot$, and one single sink $\top$.*

Definition 2 (levels in an ontology) Levels *are defined inductively: L_0 is the set of formulas that have no predecessor (it contains only the contradiction $\bot$) L_i is the set of formulas that have no predecessor in $G \setminus (L_0 \cup \ldots L_{i-1})$, etc. $\Gamma^+(x)$ and $\Gamma^-(x)$ are the sets of successors and predecessors of x.*

Level L_1 nodes are called leaves (i.e., formulas φ s.t. the edge $(\bot, \varphi) \in U$). Moreover, we impose: (a) G: to be a lattice, (b) all the sub-classes of a class: to appear in the ontology, (c) all the leaves: to be mutually exclusive two by two.

Proposition 1 *Providing that:*
(1) we add the appropriate formulas and arcs that turn a dag into a lattice;
(2) we add to each not-leave formula φ, a sub-formula "other elements of φ";
(3) we split leaves, wherever necessary, to make them mutually exclusive;
then, we can insure properties (a), (b) and (c) because the operations (1), (2) and (3) can always be done in the finite case.

Hence, an ontology will be encoded in the following way.

Definition 3 (logical encoding of an ontology) *Any dag $G = (X, U)$ representing an ontology can be associated to a set L_G of formulas that hold:*

1.	$\forall(\varphi, \psi) \in U$,	*it holds that*	$\varphi \rightarrow \psi$.
2.	$\forall \varphi \in X \setminus \{L_1 \cup L_0\}$,	*it holds that*	$\varphi \rightarrow \bigvee_{\varphi_i \in \Gamma^-(\varphi)} \varphi_i$.
3.	$\forall \varphi, \psi \in L_1$,	*it holds that*	$\varphi \wedge \psi \rightarrow \bot$.
4.	$\forall(\varphi, \psi) \in X \times X$, *s.t.* $\varphi \vdash \psi$,	*it exists a directed path from* φ *to* ψ *in* G.	

Rule 1 expresses that an inclusion relation holds between two classes, 2 is a kind of closed world assumption version of property (b), 3 expresses property (c), 4 expresses completeness, as follows: if all the inclusion relations are known in the ontology, hence all corresponding paths must exist in G. From this, it follows that: $\forall \varphi \in X$, $\varphi \rightarrow \bigwedge_{\varphi_i \in \Gamma^+(\varphi)} \varphi_i$. and $\forall \varphi \in X$, $\varphi \rightarrow \top$. Given any pair of formulas $(\varphi, \psi) \in X \times X$, the logical encoding of the ontology $G = (X, U)$ allows us to decide if $\{\varphi \wedge \psi\} \cup L_G$ is consistent or not; and if $\varphi \cup L_G \vdash \psi$ or not. Taxonomy 1 of Figure 2 provides a toy example of such an ontology, where e.g. $L_0 = \{\bot\}$, $L_1 = \{conifer, wetland, agriculture\}$.

2.2 Attributive formulas

Since we need to express binary links, our representational language is built on ordered pairs of formulas of $\mathscr{L}_i \times \mathscr{L}_s$, here denoted (φ, p). Such formulas should be understood as formulas of $\mathscr{L}_i$ reified by association with a set of parcels described by a formula of $\mathscr{L}_s$. In other words, to each formula is attached a set of parcels, where this formula applies. More precisely, (φ, p) expresses that φ

is true for *each* elementary parcels satisfying p. Another understanding would view (φ, p) as the material implication $\neg p \vee \varphi$ in the language based on the union of the two vocabularies $\mathscr{V}_i$ and $\mathscr{V}_s$. Alternatively, in a first order logic language view, this may be also understood as $\forall x, p(x) \rightarrow \varphi(x)$, here $p(x)$ means that the parcel x satisfies p, equating formula p with the union of elementary parcels x_0 satisfying p. A pair (φ, p) will be called an *attributive formula*.

Definition 4 (attributive formula) *An* attributive formula f*, denoted by a pair* (φ, p)*, is a propositional language formula based on the vocabulary* $\mathscr{V}_i \cup \mathscr{V}_s$ *where the logical equivalence* $f \equiv \neg p \vee \varphi$ *holds and* p *contains only variables of the vocabulary* $\mathscr{V}_s$ $(p \in \mathscr{L}_s)$ *and* φ *contains only variables of* $\mathscr{V}_i$ $(\varphi \in \mathscr{L}_i)$*.*

The intuitive meaning of $f = (\varphi, p)$ is that *for the set of elementary parcels that satisfy* p*, the formula* φ *is true.* Observe that there exist formulas built on the vocabulary $\mathscr{V}_i \cup \mathscr{V}_s$ which cannot be put under the attributive form, e.g., $a \wedge p_1$ where a is a literal of $\mathscr{V}_i$ and p_1 a literal of $\mathscr{V}_s$. The introduction of connectives $\wedge$, $\vee$ and $\neg$ does make sense, since any pair (φ, p) is a classical formula. From the above definition of (φ, p) as being equivalent to $\neg p \vee \varphi$, several inference rules straightforwardly follow from classical logic:

Proposition 2 (inference rules on attributive formulas)
1. $(\neg\varphi \vee \varphi', p), (\varphi \vee \varphi'', p') \vdash (\varphi' \vee \varphi'', p \wedge p')$
2. $(\varphi, p), (\varphi', p) \vdash (\varphi \wedge \varphi', p)$*;* *3.* $(\varphi, p), (\varphi, p') \vdash (\varphi, p \vee p')$
4. if $p' \vdash p$ *then* $(\varphi, p) \vdash (\varphi, p')$*;* *5. if* $\varphi \vdash \varphi'$ *then* $(\varphi, p) \vdash (\varphi', p)$

From these rules, we can deduce the converse of 2: $(\varphi \wedge \varphi', p) \vdash (\varphi, p), (\varphi', p)$ and that $(\varphi, p), (\psi, p') \vdash (\varphi \vee \psi, p \vee p')$ and $(\varphi, p), (\psi, p') \vdash (\varphi \wedge \psi, p \wedge p')$. Thus, reification allows us to keep potential inconsistency *local*, namely restricted to a subset of parcels rather than pervading the whole knowledge base.

2.3 Taxonomy of properties and partonomy of parcels

The previous formalization of an ontology can be applied both to parcels, which gives birth to partonomies, and to properties for describing conceptual taxonomies. The properties associated to parcels can be labels taken from a *vocabulary.* It might seem more suitable to develop first on parcels, before developing on properties that we will attribute to parcels. But, in fact we agree with [13] who says that *"the taxonomic basis of single-resource classifications precludes their direct placement in a spatially based ecological hierarchy (partonomy)."*

Taxonomies divide and organize items into hierarchies of *kind-of* relations [21]. *"They work well for arranging entities possessing distinct, identifiable characteristics [...] (soils, vegetation, etc.).* But, this strict and rigid identification is also a limitation, as announced in [13]: *Applying taxonomic classifications to characterize ecological patterns over space proves difficult."* A *taxonomy* is an ontology, hence a lattice where the nodes are labeled on a given *vocabulary*, and where the partial order entails a relation, named *sort-of* or *is-a*, with the following peculiarities in practice: (i) Any level can exist without antecedent; (ii) If *a sort-of b*, then *a* may be unique. Let's name *taxon* a node of this graph.

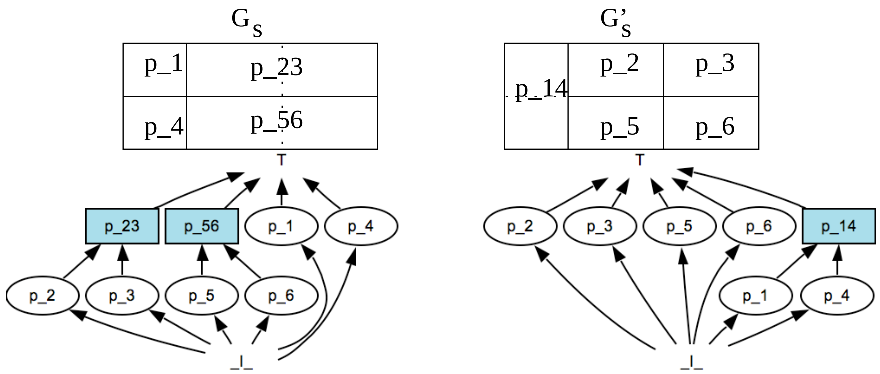

Figure 1: Two partonomies (space ontologies) for the same set of parcels.

Partonomies reflect *part-of* relations based on space or proximity [21]. [13] says: *"Recognition of patterns at different spatial resolutions is fundamental to partonomies. Fortunately, there is a natural tendency for humans to perceive and subdivide the environment on the basis of part-whole relationships [5]. [...] most patterns or structures originate from ecological processes that are inherently spatial and thus partonomic in nature."* In a partition of a territory, particular subsets of parcels may have names, hence any partition equipped with the set inclusion relation, can be represented by partonomy. We further assume that all these partonomies share the same set of elementary parcels. Fig.1 exhibits two partonomies Gs and Gs' and their common elementary parcels $p_1, \ldots, p_6$.

A *partonomy* [1] is an ontology, hence a lattice, where the nodes are labeled by the elementary parcels, and the partial order entails a part-of relation which, in practice, has the following properties: (i) A class exists if and only if all its sub-classes exist; (ii) Only leaves can exist without antecedent; (iii) If a *part-of* b, then b *made-of* a, and it exists c (in the parcel vocabulary), complement of a in b. The union of two taxons can always exist, but it is not the case for two elements of a partonomy (partons), because taxons are classes, but partons are individuals that must exist when used by an operator. Figure 1 represents two partitions G_s and G'_s of the same space, leading to two partonomies where elementary parcels are identified by ovals.

3 Fusion of properties as an ontology alignment problem

Because the vocabulary is often insufficient for describing any subset of objects in a non-ambiguous way; or conversely because there may be no proper set of objects that satisfy a given set of properties and only them, only many-to-many relationships are really useful for representing geographic information. For a many-to-many relationship between the parcels of a given subset P_i of the

partonomy, and the properties of a given list L_j of excerpts from the taxonomy, we need classically to build three database relations:
- R_s that distributes the subset P_i over its parcels;
- R_p that distributes the subset L_j over its properties;
- R_a made of the attributive formulas: pairs from $R_s \times R_p$ (learning samples).

What interests us is to discover if some additional knowledge emerges from the fusion of two information sources (R_{s1}, R_{p1}, R_{a1}) and (R_{s2}, R_{p2}, R_{a2}). The fusion of partonomies is not a problem, if we accept to ignore data matching issues, and that the geometric intersection between parcels of R_{s1} and R_{s2}, become leaves of the fusion R_s. The fusion of taxonomies is more difficult (many papers in FCA, semantic web, database integration), and it converges now to the notion of *ontology alignment* (see: Euzenat and Shvaiko [11]). We can distinguish several aspects: (a) the construction of $R_a = R_{a1} + R_{a2}$ (concatenation), (b) the structural alignment that will identify th number of nodes for candidate attributive formulas, and their partial order (classical FCA); (c) the labeling of this nodes that may unify them possibly on either R_{p1} or R_{p2}, or may need to form a new label by coupling (sign &) concepts from both R_{p1} and R_{p2}; (d) the decision to keep or discard these candidates nodes, according to one or several criteria (this aspect is skipped here, but similar to the discussion of section 5).

Let's now illustrate the problem with a landscape analysis example. Fig. 2 exhibits two concurrent taxonomies about land cover, as often, when experts from different disciplines try to build a domain ontology that reflects their own knowledge. Here, taxonomy 1, seems broader than taxonomy 2, which focuses on moor lands (shrubs, heath, and grass that can be natural or cultivated). We also notice that taxonomy 1 accepts multi-heritage, while the second does not.

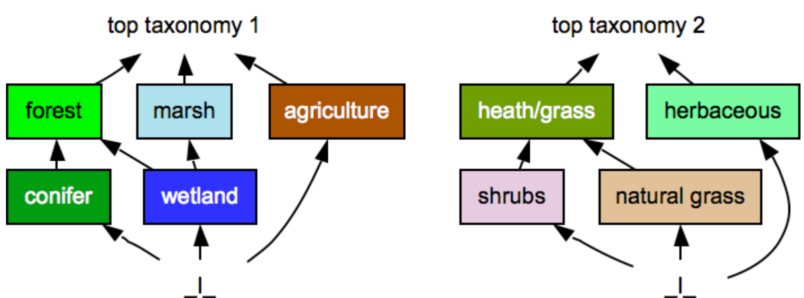

Figure 2: an example of two taxonomies

One solution is to combine the two taxonomies with the assumption that they are totally disjoint, and that only one type of information is possible at one parcel (full mutual exclusion: Fig.3). This first solution means that for each parcel, we must choose only one label, from either taxonomy. This is a much too strong constraint, e.g.: *Agriculture* and *Herbaceous* are not necessarily incompatible.

A second approach is to consider every association as equally possible, under

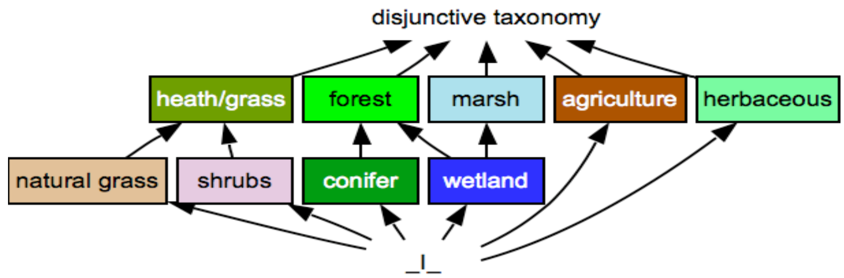

Figure 3: mutual exclusion taxonomy (solution 1).

the only constraint to preserve both original partial orders (Fig.4). It doesn't impose anything: consequently, it doesn't provide any additional knowledge.

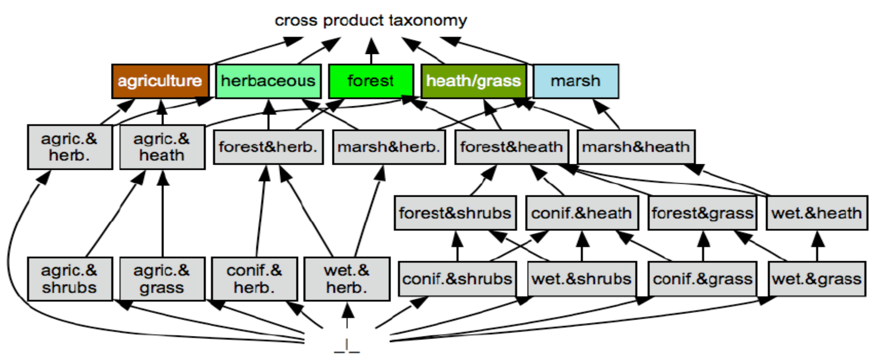

Figure 4: corresponding cross product taxonomy (solution 2).

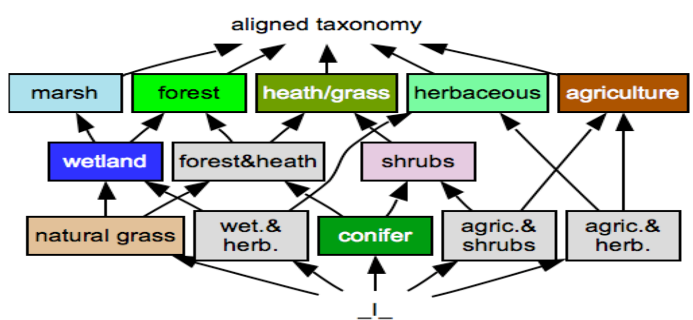

Figure 5: corresponding aligned taxonomy (solution 3).

The third solution is to use the relation R_a, built for each p with all the attributive formulas (φ_i^1, p) expressed in taxonomy 1, together with all the $({\varphi^2}_j, p)$ expressed in taxonomy 2. Using a FCA algorithm [16], we can compute the taxonomy of Fig. 5: this is the most informative solution, which filters only the concepts that fit with the actual observations. The principle of the algorithm is to 'learn', among several partial orders compatible with both taxonomies, the minimal which complies with the given set of observations. Of course, this data-mining technique, if used with different observations, may lead to different "learned taxonomies", but a stability can be obtained with reliable enough samples.

4 Representing uncertain geographical information

Our attributive language is now extended in a possibilistic logic manner, by allowing uncertainty on properties. Let us recall that a standard propositional possibilistic formula [8] is a pair made of a logical proposition (Boolean), associated with a certainty level. The semantic counterpart of a possibilistic formula (φ, α) is a constraint $N(\varphi) \geq \alpha$ expressing that α is a lower bound on the necessity measure N [9] of logical formula φ. Possibilistic logic has been proved to be sound and complete with respect to a semantics expressed in terms of the greatest possibility distribution π underlying N ($N(\varphi) = 1 - \sup_{\omega \models \neg\varphi} \pi(\omega)$). This distribution rank-orders interpretations according to their plausibility [8].

Note that a possibilistic formula (φ, α) can be viewed at the meta level as being only true or false, since either $N(\varphi) \geq \alpha$ or $N(\varphi) < \alpha$. This allows us to introduce possibilistic formula instead of propositional formula inside our attributive pair, and leads to the following definition.

Definition 5 (uncertain attributive formula) *An* uncertain attributive formula *is a pair* $((\varphi, \alpha), p)$ *meaning that for the set of elementary parcels that satisfy* p*, the formula* φ *is certain at least at level* α*.*

The inference rules of possibilistic logic [8] straightforwardly extend into the following rules for reasoning with uncertain attributive formulas:

Proposition 3 (inference rules on uncertain attributive formulas)
1. $((\neg\varphi \vee \varphi', \alpha), p), ((\varphi \vee \varphi'', \beta), p') \vdash ((\varphi' \vee \varphi'', \min(\alpha, \beta)), p \wedge p')$
2. $((\varphi, \alpha), p), ((\varphi', \beta), p) \vdash ((\varphi \wedge \varphi', \min(\alpha, \beta)), p)$
3.A. $((\varphi, \alpha), p), ((\varphi, \beta), p') \vdash ((\varphi, \min(\alpha, \beta)), p \vee p')$
3.B. $((\varphi, \alpha), p), ((\varphi, \beta), p') \vdash ((\varphi, \max(\alpha, \beta)), p \wedge p')$
4. if $p \vdash p'$ *then* $((\varphi, \alpha), p') \vdash ((\varphi, \alpha), p)$*; 5. if* $\varphi \vdash \varphi'$ *then* $((\varphi, \alpha), p) \vdash ((\varphi', \alpha), p)$

Rules 3.B. and 3.A. correspond respectively to the fact that either i) we locate ourselves in the parcels that satisfy both p and p', and then the certainty level of the formula φ can reach the maximal upper bound of the certainty levels known in p or in p', or ii) we consider any parcel in the union of the models of

p and p' and then the certainty level is only guaranteed to be greater than the minimum of α and β. Note that this formalism allows us to express a greater uncertainty about a rather specific label than about a more general label, as in:

Example 1 *In order to express that parcel p_1 has either "Conifer" or "Wetland" and more plausibly "Conifer", we use the two uncertain attributive formulas: $((Conifer, \alpha_1), p_1)$ and $((Wetland \vee Conifer, \alpha_2), p_1)$ where $\alpha_1 \leq \alpha_2$. At the semantic level, this is represented by the possibility distribution π_1 for p_1:*

$$\pi_1(\omega) = \begin{cases} 1 & \textit{if } \omega \models Conifer, \\ 1 - \alpha_1 < 1 & \textit{if } \omega \models Wetland \wedge \neg Conifer, \\ 1 - \alpha_2 & \textit{otherwise.} \end{cases}$$

Suppose that parcel p_2 has almost certainly Forest and more plausibly Conifer, knowing that Conifer are Forest $((\neg Conifer \vee Forest, 1), \top)$. Then for p_2:

$$\pi_2(\omega) = \begin{cases} 0 & \textit{if } \omega \models Conifer \wedge \neg Forest, \\ 1 - \alpha_2 & \textit{if } \omega \models \neg Conifer \wedge \neg Forest, \\ 1 - \alpha_1 & \textit{if } \omega \models \neg Conifer \wedge Forest, \\ 1 & \textit{if } \omega \models Conifer \wedge Forest, \end{cases}$$

This distribution can be syntactically encoded by the three formulas $((\neg Conifer \vee Forest, 1), p_2)$, $((Forest, \alpha_2), p_2)$ and $((Conifer, \alpha_1), p_2)$, with $\alpha_2 \geq \alpha_1$.

Fusion operations. The syntactic counterpart of the pointwise combination of two possibility distributions π_1 and π_2 into a distribution $\pi_1 \oplus \pi_2$ by any monotonic combination operator $\oplus$ such that $1 \oplus 1 = 1$, can be easily computed. Namely, if Σ_1 is associated with π_1 and Σ_2 with π_2, a possibilistic base that is semantically equivalent to $\pi_1 \oplus \pi_2$ can be computed as [3]:

$$\Sigma_{1\oplus 2} = \left| \begin{array}{lll} & \{(\varphi_i, 1 - (1 - \alpha_i) \oplus 1) & \text{s.t. } (\varphi_i, \alpha_i) \in \Sigma_1\}, \\ \cup & \{(\psi_j, 1 - 1 \oplus (1 - \beta_j)) & \text{s.t. } (\psi_j, \beta_j) \in \Sigma_2\}, \\ \cup & \{(\varphi_i \vee \psi_j, 1 - (1 - \alpha_i) \oplus (1 - \beta_j)) & \text{s.t. } (\varphi_i, \alpha_i) \in \Sigma_1, (\psi_j, \beta_j) \in \Sigma_2\}. \end{array} \right.$$

For $\oplus = \min$, we get $\pi_{\Sigma_1 \cup \Sigma_2} = \min(\pi_1, \pi_2)$ as expected. For $\oplus = \max$, we get $\Sigma_{\max(\pi_1, \pi_2)} = \{(\varphi_i \vee \psi_j, \min(\alpha_i, \beta_j))$ s.t. $(\varphi_i, \alpha_i) \in \Sigma_1$, and $(\psi_j, \beta_j) \in \Sigma_2\}$.

Localization of attributive knowledge. Still, attributive information itself may have two different intended meanings, namely when stating (φ, p) one may want to express that:

- *everywhere* in each parcel satisfying p, φ holds as true, denoted by (φ, p, e). Then, for instance, $(Agriculture, p, e)$ cannot be consistent with $(Forest, p, e)$ since "Agriculture" and "Forest" are mutually exclusive in taxonomy 1.
- *somewhere* in each parcel satisfying p, φ holds as true, denoted by (φ, p, s). Then, replacing e by s in this example is no longer inconsistent, since in each parcel there may exist "Agricultural" parts and "Forest" parts.

Note that these two meanings differ from the case where two exclusive labels such as "Water" and "Grass" might be attributed to the same parcel because

they are intimately mixed, as in a "Swamp". This latter case should be handled by adding a new appropriate label in the ontology. More formally, for a given parcel p in the partonomy, if p is:
-not a leave, (φ, p, s) means: $\forall p', p' \vdash p, (\varphi, p', s)$ holds;
-a leave, but made of parts o, (φ, p, s) means that $\exists o \in p, \varphi(o)$.

Thus, it is clear that inference rules that hold for "everywhere", not necessarily hold for "somewhere". Indeed, the rule 2.2 $(\varphi, p), (\psi, p) \vdash (\varphi \wedge \psi, p)$ is no longer valid since $\exists o \in p, \varphi(o)$ and $\exists o' \in p, \psi(o')$ doesn't entail $\exists o'' \in p, \varphi(o'') \wedge \psi(o'')$. More generally, here are the rules that hold for the "somewhere" reading:

Proposition 4 (inference rules on attributive formulas)
1'. $(\neg\varphi \vee \varphi', p \wedge p',\ e), (\varphi \vee \varphi'', p',\ s) \vdash (\varphi' \vee \varphi'', p \wedge p',\ s)$
2'. $(\varphi, p,\ s), (\varphi', p,\ e) \vdash (\varphi \wedge \varphi', p,\ s)$; *3'.* $(\varphi, p,\ s), (\varphi, p',\ s) \vdash (\varphi, p \vee p',\ s)$
4'. if $p' \vdash p$ *then* $(\varphi, p,\ s) \vdash (\varphi, p',\ s)$; *5'. if* $\varphi \vdash \varphi'$ *then* $(\varphi, p,\ s) \vdash (\varphi', p,\ s)$
where $(\varphi, p,\ s)$ *stands* $\forall p', p' \vdash p\ \exists o \in p', \varphi(o)$, *and* $(\varphi, p,\ e)$ *for* $\forall o \in p, \varphi(o)$.
Moreover, between "somewhere" and "everywhere" formulas, we have:
6'. $\neg(\varphi, p, s) \equiv (\neg\varphi, p, e)$

Taxonomy information and attributive information *should be handled separately*, because they refer to different types of information, *and, more importantly*, because taxonomy distinctions expressed by mutual exclusiveness of taxons do not mean that they cannot be simultaneously true in a given area: the taxonomy-formula $(a \leftrightarrow \neg b)$, with $a, b \in \mathscr{V}_i$ coming from the same taxonomy, differs from the attributive-formula $(a \leftrightarrow \neg b, \top)$, applied to every parcel (with the *everywhere* reading), since it may happen that for a parcel p, we have $(a, p) \wedge (b, p)$ (with a *somewhere* reading). The latter may mean that p contains at least two distinct parts, and that $\exists o \in p, \varphi(o) \wedge \exists o' \in p, \psi(o')$.

However, subsumption properties can be added to attributive formulas without any problem. Indeed $\varphi \vdash \psi$ means $\forall o, \varphi(o) \to \psi(o)$, and if we have (φ, p), implicitly meaning that $\exists o \in p, \varphi(o)$, then we obtain $\exists o \in p, \psi(o)$, i.e., (ψ, p). Thus we can write the subsumption property as $(\varphi \to \psi, \top)$.

5 Information fusion: general discussion on an example

Generally speaking, fusing *consistent* knowledge bases merely amounts to apply logical inference to the union of the knowledge bases. In presence of inconsistency, another combination process should be defined and used. In this section, we develop an example, represented in the language of section 4, on two sources using the same taxonomy (possibly aligned: section 3), but different partonomies.

Possibilistic information fusion easily extends to attributive formulas: each given (φ, p) is equivalent to the conjunction of the (φ, p_i), where the p_i's are the leaves of the partonomy, such that $p_i \models p$. Using finite partonomies, it is always possible to refine them by taking the non-empty intersection of pairs of leaves, and possibilistic information fusion takes place for each p_i.

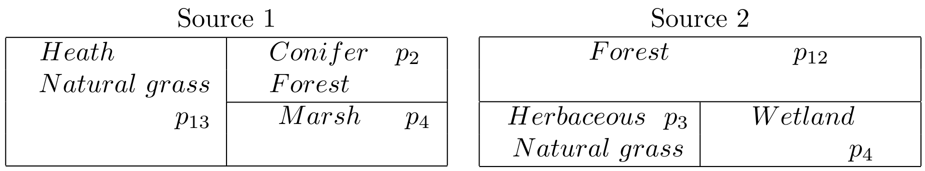

Figure 6: The information given by the sources (inspired from [19]).

Let us detail the example of Fig.6: two sources report observations about an area which is partitioned in four elementary parcels, after refinement: p_1, p_2, p_3, p_4, using the aligned taxonomy from Fig.5. Clearly, we have four possible logical readings of two labels a and b associated with an area covered by two elementary parcels p_1 and p_2:

i. $(a \wedge b,\, p_1 \vee p_2)$: means that both a and b apply to each of p_1 and p_2.

ii. $(a \wedge b,\, p_1) \vee (a \wedge b,\, p_2)$: both a and b apply to p_1 or both apply to p_2.

iii. $(a \vee b,\, p_1 \vee p_2)$: a applies to each of p_1, p_2 or b applies to each of p_1, p_2.

iii. $(a \vee b,\, p_1) \vee (a \vee b,\, p_2)$: we don't know what of a or b applies to what of p_1 or p_2. This may be particularized by adding the mutual exclusiveness constraint $\neg(a, p_1 \vee p_2) \wedge \neg(b, p_1 \vee p_2)$: that a label cannot apply to both parcels.

When a and b are mutually exclusive the everywhere meaning is impossible (if we admit that sources provide consistent information).

Another ambiguity is about if the "closed world assumption" (CWA) holds or not, e.g.: if a source says that p_i contains *Conifer* and *Agriculture*, does it exclude that p_i would also contain *Marsh* ? It would be indeed excluded by applying CWA. Also, CWA may help to induce "everywhere" information from "somewhere" information. Indeed, if we know that all formulas attached to p are $\varphi_1, \ldots \varphi_n$ with a somewhere meaning: $(\varphi_1, p,\, \mathrm{s}) \wedge \ldots \wedge (\varphi_n, p,\, \mathrm{s}))$, then CWA entails that if there were another ψ that holds somewhere in p, it would have been already said, hence we can jump to the conclusion that $(\bigvee_{i=1,n} \varphi_i, p,$ e).

Let's consider the non ambiguous reading i. of the example with the formulas:

Spatial formulas	*Property formulas*	
1. $p_1 \to p_{12}$,	14. Natgrass → Wetland,	
2. $p_1 \to p_{13}$,	15. Natgrass → ForHeath,	22. AgriHerb → Herbac,
3. $p_2 \to p_{12}$,	16. WetHerb → Wetland,	23. AgriHerb → Agric,
4. $p_3 \to p_{13}$,	17. WetHerb → Herbac,	24. Wetland → Marsh,
5. $p_{12} \vee p_{13} \vee p_4$,	18. Conifer → ForHeath,	25. Wetland → Forest,
6. $p_{12} \to p_1 \vee p_2$,	19. Conifer → Shrubs,	26. ForHeath → Forest,
7. $p_{13} \to p_1 \vee p_3$,	20. AgriShrub → Shrubs,	27. ForHeath → Heath,
8. $p_1 \wedge p_2 \to \bot$,	21. AgriShrub → Agric,	28. Shrubs → Heath,
9. + 5 mut. excl.		29. + 20 mutual excl.

Under the CWA:

Source 1

49. (Heath, p_{13}, s)
50. (Natural grass, p_{13}, s)
51. (Conifer, p_2, s)
52. (Forest, p_2, s)
53. (Marsh, p_4, s)

Source 2

54. (Forest, p_{12}, s)
55. (Herbaceous, p_3, s)
56. (Natural grass, p_3, s)
57. (Wetland, p_4, s)

Let's project on: $p_1 \equiv p_{12} \wedge p_{13}$, using formula 7: $p_{13} \to p_1 \vee p_3$, and inference rule 4' (with $p_1 \vdash p_{13}$ and $p_3 \vdash p_{13}$). Idem with p_{12}. We obtain:

Source 1

58. (Heath, p_1, s)
59. (Heath, p_3, s)
60. (Natural grass, p_1, s)
61. (Natural grass, p_3, s)

Source 2

62. (Forest, p_1, s)
63. (Forest, p_2, s)

With the closed world assumption, we deduce:

Source 1

64. (Heath ∨ Natural grass, p_1, e)
65. (Heath ∨ Natural grass, p_3, e)
66. (Conifer ∨ Forest, p_2, e)
67. (Marsh, p_4, e)

Source 2

68. (Forest, p_1, e)
69. (Forest, p_2, e)
70. (Herbaceous ∨ Natural grass, p_3, e)
71. (Wetland, p_4, e)

Now we can proceed with the fusion step, in the conjunctive mode. We

obtain:
parcel p_1: (64) and (68) yields (Conifer, p_1, e), which contradicts (60).
parcel p_2: the conjunction of (51), (52), (66), (63) is consistent, and yields (Woods, p_2, e) $\wedge$ (Conifer, p_2, s).
parcel p_3: the conjunction of (59), (61), (65), (55), (56),(70) consistently yields (Herbaceous $\vee$ Naturalgrass, p_3, e) $\wedge$ (Herbaceous, p_3, s) $\wedge$ (Naturalgrass, p_3, s)
parcel p_4: (67) and (71) yields (Rivers, p_4, e).

Conclusion:

$\perp$	$(Conifer, s)$ $(Forest, e)$
$(Herbaceous, s)$ $(Naturalgrass, s)$	$(Wetland, e)$

Sources 1 and 2 are conflicting on p_1: we can perform a disjunction of their formulas on this parcel. This conflict may come from the application of CWA to each source prior the fusion: the induction from (Forest, p_1, s) to (Forest, p_1, e) is perhaps too adventurous. We can check that (Forest, p_1, s) would yield (Conifer $\vee$ Natural grass, p_1, e) $\wedge$ (Conifer, p_1, s) $\wedge$ (Natural grass, p_1, s).

The treatment of this kind of fusion problem in [19] and [16] distinguishes between pessimistic and optimistic fusion modes. Our approach uses i) a pure logical representation setting (with an explicit distinction between conjunction and disjunction of labels), ii) distinguishes between somewhere and everywhere statements, iii) allows to express CWA (or not), iv) applies the general setting of logic-based information fusion. Our fusion result may also be more precise, thanks to a greater expressivity power of the representation framework.

Our logical framework also allows us to have a possibilistic handling of uncertainty, and then a variety of combination operations, which may depend on the level of conflict between the sources, or on their relative priority [3], can be encoded. The uncertainty setting enables us to enrich the reading of the example. Consider the information given by source 1 on p_2, namely "Conifer, Forest". As discussed in section 4.2, such an information may express that p_2 is covered by Forest, and plausibly by Conifer. With the "everywhere" reading, this can be syntactically encoded by the possibilistic formulas $((Forest, 1), p_2, e)$ and $((Conifer, \alpha), p_2, e)$, with $\alpha < 1$, together with the ontology information $((\neg Conifer \vee Forest, 1), \top)$. Similarly, the information given by source 2 on p_2 can be encoded as $((Forest, 1), p_2, e)$. Here, there is no inconsistency, hence ((Forest, 1), p_2, e) $\wedge$ ((Conifer, α), p_2, e).

Imagine that, now, source 2 says $((Forest, 1), p_2, e)$ and $((Wetland, \beta), p_2, e)$. The two sources are now partially inconsistent on p_2, and it can be checked that the level of possibilistic inconsistency of the information provided by the two sources, about p_2, is $Inc = \min(\alpha, \beta)$.

Different fusion modes can be used. One may use a renormalized conjunction [3]: the syntactic counterpart of this operator yields, if we assume $\alpha > \beta$, $((Woods, 1), p_2, e) \wedge ((Conifer, \alpha), p_2, e)$. Or one may choose a disjunctive attitude ($\oplus = \max$), one gets $((Forest, 1), p_2, e) \wedge ((Conifer \vee Wetland, \beta), p_2, e)$.

In case we again combine the two previous results obtained with the above fusion modes, by a product-based conjunction ($\oplus$ =product), one would obtain $((Woods, 1), p_2, e) \wedge ((Orchards \vee Wetland, 1 - (1 - \alpha)(1 - \beta)), p_2, e) \wedge ((Conifer, \alpha), p_2, e)$. This is a more refined result, since it keeps track of the conflict, and of a preference for the more certain information $((Conifer, \alpha), p_2, e)$ since $\alpha > \beta$. Observe however that $1 - (1 - \alpha)(1 - \beta)) > \alpha$, which makes the statement $Conifer \vee Wetland$ more certain.

6 Conclusion

After having identified representational needs (use of two vocabularies referring respectively to parcels and to properties, references to ontologies, uncertainty) when dealing with spatial information and restating ontology alignment procedures, a general logical setting has been proposed. It offers a non-ambiguous representation, propagates uncertainty in a possibilistic manner, and provides also the basis for handling multiple source information fusion. Moreover, we have seen that it is often important to explicitly distinguish between the cases where a property holds everywhere or somewhere into a parcel. An issue of interest for further research would be to allow for uncertain or default inheritance in ontologies. Note that, since subsumption relations can be easily added to the pieces of attributive spatial information, it would be possible to make some of these relations uncertain in our framework.

References

[1] C. Rosse B. Smith. The role of foundational relations in the alignment of biomedical ontologies. In M. Fieschi et al., editor, *MedInfo*, Amsterdam, 2004. IOS Pres, IMIA.

[2] S. Balley, C. Parent, and S. Spaccapietra. Modelling geographic data with multiple representations. *Int. J. of Geographical Information Science*, 18(4):327 – 352, 2004.

[3] S. Benferhat, D. Dubois, and H. Prade. A computational model for belief change and fusing ordered belief bases. In Mary-Anne Williams and Hans Rott, editors, *Frontiers in Belief Revision*, pages 109–134. Kluwer Academic Publishers, 2001.

[4] I. Bloch and A. Hunter (Eds). Fusion: General Concepts and Characteristics. *International Journal of Intelligent Systems*, 16(10):1107–1134, oct 2001.

[5] B. Tversky D. Mark, B. Smith. Ontology and geographic objects: An empirical study of cognitive categorization. In *COSIT'99. Lecture Notes in Computer Science*, volume 1661, pages 283–298, Stade, Germany, Aug. 1999. Springer.

[6] F. Dupin de Saint-Cyr, R. Jeansoulin, and H. Prade. Spatial information fusion: Coping with uncertainty in conceptual structures. In *ICCS Supplement*, pages 66–74, 2008. See http://ceur-ws.org/Vol-354/p36.pdf.

[7] F. Dupin de Saint Cyr and H. Prade. Multiple-source data fusion problems in spatial information systems. In *11th Int. Conf. on Inf. Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'06)*, pages 2189–2196, 2006.

[8] D. Dubois, J. Lang, and H. Prade. Possibilistic logic. In D.M. Gabbay, C.J. Hogger, and J.A. Robinson, editors, *Handbook of logic in Artificial Intelligence and logic programming*, volume 3, pages 439–513. Clarendon Press - Oxford, 1994.

[9] D. Dubois and H. Prade. *Possibility Theory*. Plenum Press, 1988.

[10] M. Duckham and M. Worboys. An algebraic approach to automated information fusion. *Intl. Journal of Geographic Information Systems*, 19(5):537–557, 2005.

[11] Jérôme Euzenat and Pavel Shvaiko. *Ontology matching*. Springer-Verlag, Heidelberg (DE), 2007.

[12] F. Fonseca, M. Egenhofer, P. Agouris, and G. Cmara. Using ontologies for integrated geographic information systems. *Transactions in GIS*, 6(3):231–257, 2002.

[13] A. Sorokine G. Nowacki. The limitations of applying single-resource taxonomies to ecological partonomies. In *Ecological Interpretations and Principles: Soil info. for a changing world. NCSS conf.*, pages 17–20, Plymouth MA, Aug. 2003.

[14] M. Goodchild and R. Jeansoulin. *Data Quality in Geographic Information : from Error to Uncertainty*. Hermés, Paris, 1998. 192 pages.

[15] M.F. Goodchild, M. Yuan, and T.J. Cova. Towards a general theory of geographic representation in gis. *Int. J. of Geogr. Information Science*, 21(3):239–260, 2007.

[16] R. Jeansoulin, T.T. Pham, and V. Phan-Luong. A quality-aware theme fusion for spatial information. In *Int. Conf. on Formal Concept Analysis (FCA'07)*, 2007.

[17] R. Klischewski. How to 'rightsize' an ontology: a case of ontology-based web information management to improve the service for handicapped persons. In *15th Int. Workshop on Database and Expert Systems Applications*, pages 158–162, 2004.

[18] F. Petry, M. Cobb, L. Wen, and H. Yang. Design of system for managing fuzzy relationships for integration of spatial data in querying. *Fuzzy Sets and Systems*, 140(1):51–73, November 2003.

[19] T. Trung Pham. *Fusion de l'information géographique hiérarchisée.* PhD thesis, Université de Provence, septembre 2005.

[20] W.V.O. Quine. *From a Logical Point of View*, chapter On What There Is, pages 1–19. Harper and Row, New York, 1953.

[21] B. Smith. Mereotopology: A theory of parts and boundaries. *Data and Knowledge Engineering*, 20:287–303, 1996.

[22] Steffen Staab and Rudi Studer (eds). *Handbook on Ontologies.* Springer, 2004.

[23] E. Wurbel, O. Papini, and R. Jeansoulin. Revision: an application in the framework of GIS. In *7th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'00)*, pages 505–516, 2000.