When nonmonotonicity comes from distances

Nicholas Asher and Jérôme Lang

IRIT, Université Paul Sabatier
118, route de Narbonne
F-31062 Toulouse Cedex, France
E-mail: asher,lang@irit.fr

Abstract. We propose a methodology for defining nonmonotonic inference relations, involving distances (in a non-topological meaning) between "labels", each label being associated with a logical theory. Our framework is motivated by applications to spatial reasoning (reasoning by proximity) and taxonomic reasoning, though it also applies to temporal reasoning (degradation of persistence). We propose several ways of defining nonmonotonic inference relations from distances.

1 Introduction

In the literature of nonmonotonic reasoning, and, analogously, belief revision, it has been widely recognized that most nonmonotonic inference relations have or impose an underlying *ordering structure* on worlds, or sets of worlds (e.g. [14], [15], [11]). Now, ordering relations are also used (in a more syntactical way - they are called then *priorities*) in syntax-based default reasoning (e.g. [17] [3] [1] [16]); and more generally, orderings and priorities are exploited in other nonmonotonic formalisms, in belief revision (e.g. [10]'s epistemic entrenchment relations), in updating formalisms, in logic programming, etc.

One major question which is often unanswered is *where do these orderings and priorities come from?* Two answers are given in the literature:

- orderings come from the user. In this case, they usually represent uncertainty: some statements are more reliable than others.
- orderings derive from a knowledge base consisting in a set of default rules $\alpha_i \mid\!\sim \beta_i$ (or equivalently from a set of conditional statements $\alpha_i \rightarrow \beta_i$) which determines a nonmonotonic inference relation. In Rational Closure [15] and in System-Z [18], defaults are automatically ranked according to their *specificity* and this ranking induces a rational inference relation.

In this paper, we propose an alternative way to induce orderings, and consequently an alternative way to derive inference relations. Here, orderings come directly from *pseudo-distances*[1]. These pseudo-distances can apply to time points, points of space, sources, situations, classes, and other types of things. To induce nonmonotonic inferences from distances, we "reason by *proximity*, which is a sort of *extrapolation* process. Consider for instance a time-stamped knowledge base, i.e. a set of formulas, each formula being indexed by the time point when it was put in the knowledge base: it is clear that the more recent the information, the more certain we are that it is still holding, and thus the more priority it should be assigned (provided that the given formulas encode fluents[2] which generally tend to persist). As a second example, consider a knowledge base consisting in observations at different points of a given region. Suppose that we do not know whether a given formula holds at

[1] We avoid using the word *distance* which evokes the topological definition; our pseudo-distances are not necessarily defined on a completely ordered scale, they are not necessarily symmetrical, nor necessarily satisfy the triangle inequality.

[2] A fluent is a proposition whose truth value changes with time.

a point x. To find out, we may consider points *close* to x at which information is known; the closer y is to x, the more pertinent is the information at y to x. We then extrapolate for x. The general principle behind these examples is that distances between "labels" (time points, points in the space etc.) play the role of priorities: the closer the labels, the higher the priority. Thus, from a labelled (temporal, spatial etc.) knowledge base we will define a set of nonmonotonic inference relations–one for each label. Our examples will focus on temporal, spatial, and taxonomic reasoning. We first develop the basic formalism and supply it with a modal semantics. Then we propose three different ways of defining nonmonotonic inference relations from pseudo-distances: *projection* (which is intuitive and easy to compute, but which sometimes gives unintended results), *radial accumulation* (which gives a rational inference relation when the distance is complete) and prioritized syntax-based entailment.

2 Labelled pseudo-metric structures

2.1 Labels and pseudo-distances

The general structure we have in mind is the following. We have a collection of knowledge items, i.e. formulas labelled by something indicating the location of the observation, or when it was put in the knowledge base, etc. This set of "labels" will be equipped with a kind of metric.[3] From now on, $\mathcal{L}$ denotes a propositional language whose formulas are denoted by greek letters φ, ψ etc. $\top$ and $\bot$ denote respectively tautology and contradiction, $\vdash$ classical entailment and $Cn(S)$ the set of all logical consequences of the set of formulas S.

Definition 1 (label structure) *A label structure is a pair $\langle X, d\rangle$ where X (the label set) is a nonempty set and d (the pseudo-distance on X) is a mapping from $X \times X$ to a (completely or not) ordered set U_d (the distance scale) with a minimal element denoted 0, verifying $\forall y \in X, d(x,y) = 0$ iff $x = y$.*

From d we induce the collection of pre-ordering relations (i.e. reflexive and transitive relations) $\leq_x, x \in X$, defined by: $\forall x, y, z \in X$, $y \leq_x z$ iff $d(x,y) \leq d(x,z)$. Intuitively, $y \leq_x z$ means that y is closer to x than z. If U_d is completely ordered, then all $\leq_x$ are complete preorderings. In this case, d and $\langle X, d\rangle$ are both said to be *complete*. $y <_x z$ will be an abbreviation for $y \leq_x z$ and not $(z \leq_x y)$. Note that we have $\forall y \in X, x \leq_x y$ and $\forall y \in X, y \leq_x x \Leftrightarrow x = y$.

As illustrations, we will explain each new definition with respect to *spatial* structures. However, the reader should keep in mind that the framework is much more general. As a simple example of a label structure, we consider the discretized portion of 2-dimensional space $\{(x,y),\ x = 0,1,2,3,\ y = 0,1,2,3\}$ equipped with the distance $d((x,y),(x',y')) = |x - x'| + |y - y'|$. The distance scale is obviously $U_d = \{0, 1, ..., 6\}$.

Definition 2 (labelled formula, labelled knowledge base) *Let $\langle X, d\rangle$ be a label structure. If φ is a formula of $\mathcal{L}$ and Y is a non-empty subset of X then $Y : \varphi$ is a labelled formula, meaning that φ holds at every label of Y. When Y is a singleton we will note $y : \varphi$ instead of $\{y\} : \varphi$.*

A labelled knowledge base LKB is a finite set of labelled formulas.

The theory at label x induced by LKB, denoted by $Cn_x(LKB)$ is the logical closure of the set of all formulas $Y_i : \varphi_i$ in LKB such that $x \in Y_i$. The logical closure of LKB, denoted $Cn(LKB)$ is the collection of closed theories $\{Cn_x(LKB), x \in X\}$.

[3] So far no internal operation on the set of labels is needed,and so our framework has thus far very little to do with [9]'s labelled deductive systems.

Note that, in practice, LKB may contain punctual *observations* $y_i : \varphi_i$ as well as *constraints* holding at every label $X : \varphi_j$.

We say that a labelled knowledge base is consistent iff it is consistent at every label, i.e. if all theories $Cn_x(LKB)$ are consistent. In this paper, we restrict our attention to consistent labelled knowledge bases only.[4] H being a consistent labelled theory, $H(x)(= Cn_x(LKB))$ will denote the theory at label x; thus $H = \{H(x)|x \in X\}$.

Example: let $\langle X, d\rangle$ be the discrete spatial structure as defined above, and $LKB = \{X : \neg(r \wedge s);\ \{(0,1),(0,2)\} : r \vee s;\ (1,0) : \neg s; (1,3) : \neg r;\ (3,0) : \neg r \wedge \neg s;\ \{(2,2),(3,2)\} : r\}$ (where r and s stand respectively for raining and snowing). LKB is depicted by Figure 1. The induced labelled theory contains, for example, $\neg s$ at label $(3,2)$.

x y	0	1	2	3
0		$r \vee s$	$r \vee s$	
1	$\neg s$			$\neg r$
2			r	
3	$\neg r \wedge \neg s$		r	

and everywhere: $\neg(r \wedge s)$

Figure 1: a spatial example of a labelled knowledge base

2.2 Examples of labelled pseudo-metric structures

- $X = [t_{min}, t_{max}]$ is a linear time scale (discrete or continuous) and labels are time points; $d(x,y) = |x - y|$.
- X is still a time scale but $d(x,y) = (sign(x-y), |x-y|)$, with $(s,\alpha) \leq (s',\alpha')$ iff $s = s'$ and $\alpha \leq \alpha'$, or $\alpha = 0$. This non-complete, non-symmetrical pseudo-distance considers past and future separately.
- X is a region of a n-dimensioned Euclidean space, and labels are points (of the Euclidean plan, of space, ...); d is the Euclidean distance of X.
- X is still a region of a n-dimensioned Euclidean space, but $d(x,y)$ is the vector **xy**, with the ordering $\mathbf{u} \leq \mathbf{v}$ iff $\mathbf{u} = k.\mathbf{v}$ with $0 \leq k \leq 1$.
- X is a set of *sources*; a source being closer to another if it generally behaves the same way, give the same information, has the same opinion etc.). The actual situation may be considered as a distinguished source s_0.
- $X = 2^S$ where S is a set of *hypotheses*; d can be the symmetric difference between subsets of S (the ordering on U_d being set inclusion), or the cardinality of the symmetric difference. A label, i.e. a set of hypotheses, corresponds to a "situation" or a "configuration". Thus, applying our methodology here consists roughly in completing the context of an environment by adding what is true in the "closest" environments.
- X is a taxonomic graph (the labels being classes or types of situations), d is a distance between classes - given with the graph. Applying our methodology here means that if a property of an object is not known, one extrapolates from the most similar classes (the closest w.r.t d). This gives rise to a kind of reasoning by analogy.

2.3 A modal view of pseudo-metric structures

It is possible to give our labelled structures a modal interpretation.[5] Since the knowledge at a label is generally not complete, each label will be associated not with a single world but with a set of worlds:

[4] This restriction to consistent knowledge bases assumes that we get first rid of punctual inconsistencies (by any method - there are many in the literature). A discussion of this point is beyond the scope of this paper.

[5] This section is not necessary for understanding most of the rest of the paper.

Definition 3 (pseudo-metric model) *A pseudo-metric model for a pseudo-metric structure $\langle X, d\rangle$ is a 4-uple $M = \langle W, m, C, \mathcal{O}\rangle$ where W is a set of worlds, m a classical meaning function on worlds (i.e. $\forall w \in W$, $m(w) \subseteq \mathcal{L}$); C is an equivalence relation on W, and $\mathcal{O} = \{\leq_w, w \in W\}$ is a collection of reflexive and transitive relations verifying $\forall w' \in W, w \leq_w w'$.*

The relation C defines clusters of worlds (one cluster for each label). Now, a consistent labelled theory whose atomic formulas and subformulas define a sublanguage of $\mathcal{L}$, $\mathcal{L}_H$ L induces a "biggest" labelled pseudo-metric model defined as the collection of all maximal consistent sets of $\mathcal{L}_H$ formulas that make all the formulas labelled at x in H true, for all $x \in X$. Note that several duplicate worlds (i.e. mapped identically by m) may appear in different clusters, since they may be possible at different labels. We will assume that these are distinguished by indices. Lastly, the preordering $\leq_w$ is the transposition of $\leq_x$. Formally:

Definition 4 *The pseudo-metric model M_H associated with the consistent labelled theory H on $\langle X, d\rangle$ is the tuple $M = \langle W*, m, C, \mathcal{O}*\rangle$ where: $W* = \bigcup\{D(x) : x \in X\}$, $D(x) = \{w_x | w_x$ is a maximal consistent set of $\mathcal{L}_H$ formulas that includes $H(x)\}$ (duplicate worlds being distinguished by the label x), and $\mathcal{O}* = \{\leq_w \mid \leq_w$ is reflexive and transitive on W^* and if $w \in C(x)$, $w' \in C(y)$ and $w'' \in C(z)$ then $w' \leq_w w''$ iff $y \leq_x z\}$.*

In the example of figure 1, for example, $C((0,0)) = \{r\bar{s}, \bar{r}s, \bar{r}\bar{s}\}$, $C((0,1)) = \{r\bar{s}, \bar{r}s\}$, etc.
We have the following property: w being any world in $C(x)$, $\varphi \in H(x) \Leftrightarrow w \models \Box\varphi$ (φ is known at x). Since this is true for each $w \in C(x)$ we will note $C(x) \models \Box\varphi$. $\Box$ is from S5 since C is an equivalence relation. The structure (W, m, C) enables us to draw *monotonic* conclusions. $\mathcal{O}$ will then enable nonmonotonic conclusions; namely we will define a nonmonotonic inference relation $\alpha \mid\!\sim_{H,x} \beta$ for each $x \in X$. To a class c of nonmonotonic inference relations induced by distances there will be a conditional operator $\Rightarrow^c$ such that $\alpha \mid\!\sim^c_{H,x} \beta$ iff $M_H, C(x) \models \alpha \Rightarrow^c \beta$.

3 Extrapolation by projection

We now propose to draw nonmonotonic conclusions at a given label, using the following intuitive principle: our background knowledge being represented by H, then φ nonmonotonically entails ψ at label x iff at the closest labels where φ is known to hold and the truth value of ψ is determined, then this truth value is TRUE. To guarantee that these *closest labels* exist, we have to add an extra assumption:

Definition 5 (smooth labelled theories) *A consistent labelled theory H is smooth iff $\forall x \in X, \forall \varphi \in \mathcal{L}$, for any infinite $\leq_x$-descending chain $\{z_i\}_{i\geq 1}$ such that $\forall i, \varphi \in H(z_i)$, then $\exists \bar{z} \in X$ such that $\bar{z} = \lim_{i\to\infty} z_i$ and $\varphi \in H(\bar{z})$.*

We will only consider subsequently smooth labelled theories. This constraint is easily fulfilled, of course if X is finite or if H associates a non-trivial theory to only a finite number of labels. It is also fulfilled also if $\langle X, d\rangle$ is a topological metric space and $\forall\varphi$, the set of labels where φ is true is topologically closed.[6] In the rest of the Section, H is a smooth consistent labelled theory.

Definition 6
$Known_H(\varphi) = \{x \in X | \varphi \in H(x)\}$; $Det_H(x, \varphi) = Known_H(\varphi) \cup Known_H(\neg\varphi)$

Thus, $Known_H(\varphi)$ is the set of labels where φ is known to be true (w.r.t. H) and $Det_H(\varphi)$ is the set of labels where φ has a determined truth value. Now, ψ is inferred

[6] Note that this last condition is fulfilled if H is the logical closure of a labelled knowledge base $LKB = \{Y_i : \varphi_i, i = 1...n\}$ such that all Y_i's are topologically closed.

nonmonotonically from φ given H iff φ implies logically ψ [7] or ψ holds at all closest labels where φ is known to be true and ψ is determined:

Definition 7 (projection inference relation)
$\varphi \mid\!\sim^{PROJ}_{H,x} \psi$ *iff* $\varphi \vdash \psi$ *or* $Min(\leq_x, Known(\varphi) \cap Det_H(\psi)) \subseteq Known(\psi)$

In the simple case where $\varphi = \top$ (which is frequent in practice) the definition reduces to $\top \mid\!\sim^{PROJ}_{H,x} \psi$ iff $Min(\leq_x, Det_H(\psi)) \subseteq Known(\psi)$.

We can see a similarity between the definition of projection entailment and preferential entailment where one looks if ψ holds in all preferred φ-worlds; here we are dealing with *clusters of worlds* instead of worlds, and one looks if ψ holds in all closest (preferred) "φ-clusters" where ψ has a determined truth value.[8]

An immediate and important property of $\mid\!\sim^{PROJ}_{H,x}$ is that it preserves what is already known in the labelled theory H:

Proposition 1 *if* $\psi \in H(x)$ *then* $\mid\!\sim^{PROJ}_{H,x} \psi$

Proposition 2 *Let* $\Rightarrow^P$ *be the conditional operator defined by* $M, w \models \varphi \Rightarrow^P \psi$ *iff* $\forall w'$ $\leq_w$*-minimal such that* $M, w' \models \Box\varphi \wedge (\Box\psi \vee \Box\neg\psi)$, *then* $w' \models \psi$. *Then* $\varphi \mid\!\sim^{PROJ}_{H,x} \psi$ *iff* $M_H, f(x) \models \varphi \Rightarrow^P \psi$.

The Example in Figure 1 and *PROJ*:

• at label $(1,1)$, nothing else than the constraint $\neg(s \wedge r)$ is known for sure; the closest label where s is determined is $(1,0)$ (at distance 1), therefore $\mid\!\sim^{PROJ}_{H,(1,1)} \neg s$. Now, the closest labels where r is determined are $(1,3)$ and $(2,2)$ (at distance 2) but they disagree about its truth value. Therefore, neither r nor $\neg r$ is inferred (from $\top$) at $(1,1)$.

• at label $(1,2)$ we get $\mid\!\sim^{PROJ}_{H,(1,2)} r \vee s$ (the closest labels where $r \vee s$ is determined are $(0,2)$ and $(2,2)$) and $\mid\!\sim^{PROJ}_{H,(1,2)} \neg s$ (the closest label where s is determined is $(2,2)$). However we cannot infer r nor $\neg r$ since the closest labels where r is determined, i.e., $(1,3)$ and $(2,2)$, disagree on r.

This last example shows that $\mid\!\sim^{PROJ}_{H,x}$ does not satisfy the (And) property since $\mid\!\sim^{PROJ}_{H,(1,2)} r \vee s$, $\mid\!\sim^{PROJ}_{H,(1,2)} \neg s$ and however $\not\mid\!\sim^{PROJ}_{H,(1,2)} r$ [9] and thus is not even a *basic* consequence relation in the sense of [11]. However, projection entailment is rather easy to compute and we think it might be adequate in some simple cases. In the next section we propose another definition which is generally more satisfactory.

4 Extrapolation by radial accumulation in the complete case

We assume $\langle X, d \rangle$ is complete (the general case is not considered in this paper for lack of space) and propose now a second inference relation consisting intuitively in gathering together the pieces of information attached to the labels inside *spheres* around x; one considers first the smallest sphere $\{x\}$, and then larger and larger spheres, until we can prove one of the formulas we look for.

---

[7] this condition is needed to ensure that $\bot \mid\!\sim^{PROJ}_{H,x} \psi$, and thus to guarantee supraclassicality [11].

[8] The alternative definition $\varphi \mid\!\sim^{PROJ}_{H,x} \psi$ iff $\varphi \vdash \psi$ or $Min(\leq_x, Known(\varphi)) \subseteq Known(\psi)$ would be too cautious: take for instance $\varphi = \top$, then $Min(\leq_x, Known(\top)) = \{x\}$ and thus $\mid\!\sim^{PROJ}_{H,x} \psi$ iff $\psi \in H(x)$ (i.e., $\mid\!\sim^{PROJ}_{H,x}$ would be reduced to its monotonic part).

[9] More intuitively, although there is a close label where $r \vee s$ is known and another close label where $\neg s$ is known, there is no close label where both are known together; in other terms, projection does not allow for using in a proof formulas known at different labels

Definition 8 (spheres and theories around a label) *Let H be a labelled theory, x a label, φ a formula and $\rho \in U$.*
$S(x,\rho) = \{y \in X | d(x,y) \leq \rho\}$
$Th(x,\rho,H) = Cn(\bigcup_{y \in S(x,\rho)} H(y))$

To avoid problems with infinite descending chains of spheres, we add:

Definition 9 *A labelled theory H is s-smooth iff for any descending chain $\{\rho_i\}_{i \geq 1}$ such that $\forall i,\ \varphi \in Th(x,\rho_i,H)$, then $\varphi \in Th(x,\bar{\rho},H)$ where $\bar{\rho} = \lim_{i\to\infty} \rho_i$*

Proposition 3 *if $\mathcal{L}$ is finite then H is s-smooth iff it is smooth.*

Let H be a s-smooth labelled theory.

Definition 10 $Rad(x,\varphi,H) = Min\{\rho \in U, \varphi \in Th(x,\rho,H)\}$

It is worth noticing that $\varphi \mapsto Rad(x,\varphi,H)$ is a kind of necessity function [10] (see for instance [7]), which is expressed by the following proposition:

Proposition 4 *(i)* $Rad(x,\top,H) = 0$
(ii) $Rad(x,\bot,H) > 0$
(iii) $Rad(x,\varphi \wedge \psi,H) = max(Rad(x,\varphi,H), Rad(x,\psi,H))$

The proof is immediate. Note that (ii) comes from the consistency of $H(x)$.

Definition 11 (radial inference relations) [11]
$\varphi \mid\!\sim^{RAD}_{H,x} \psi$ *iff* $\varphi \vdash \psi$ *or* $Rad(x,\varphi \to \psi,H) < Rad(x,\varphi \to \neg\psi,H)$

For $\varphi = \top$ we get $\mid\!\sim^{RAD}_{H,x} \psi$ iff $\vdash \psi$ or $Rad(x,\psi,H) < Rad(x,\bot,H)$, thus ψ is nonmonotonically inferred at x iff the smallest sphere around x proving ψ is contained in the smallest inconsistent sphere around x.

Proposition 5 $\mid\!\sim^{RAD}_{H,x}$ *is a comparative inference relation*[12].

The proof just consists in proving that the relation $<_{x,H}$ defined on $\mathcal{L} \times \mathcal{L}$ by $\alpha \leq_{x,H} \beta$ iff $Rad(x,\alpha) \geq Rad(x,\beta)$ is an expectation ordering ([11]) [13]. Then, since the definition given for $\mid\!\sim^{RAD}_{H,x}$ coincides with Gärdenfors and Makinson's definition, hence the result.
Note that $\mid\!\sim^{RAD}_{H,x}$ also preserves H too:

Proposition 6 *if $\psi \in H(x)$ then $\mid\!\sim^{RAD}_{H,x} \psi$*

Previous results on connections between comparative inference relations and conditional logics [2][8][4] imply that the conditional operator $\Rightarrow^R$ associated to $\mid\!\sim^{RAD}$ is the $\Rightarrow$ of Lewis' conditional logic VA (whose semantics is expressed with systems of nested spheres), or alternatively, expressed with ranked clusters of worlds, of the logic CO [2]. The underlying complete preordering relation R on the set of worlds is defined by: vRw iff $Min\{\rho, v \not\models Th(x,\rho,H)\} \leq Min\{\rho, w \not\models Th(x,\rho,H)\}$.

---

[10] A necessity function is a mapping from $\mathcal{L}$ to $[0,1]$ such that $N(\top) = 1$, $N(\bot) < N(\top)$ and $N(\varphi \wedge \psi) = min(N(\varphi), N(\psi))$. The only difference betwen N and Rad is that the scale $([0,1], \leq)$ has been replaced by $([0,+\infty], \geq)$ or any other complete ordered scale.

[11] equivalently, $\varphi \mid\!\sim^{RAD}_{H,x} \psi$ iff $\varphi \vdash \psi$ or $Rad(x,\varphi \to \psi,H) < Rad(x,\neg\varphi,H)$.

[12] A comparative inference relation [11] is a rational inference relation [15] which satisfies moreover Consistency Preservation).

[13] this is intuitively obvious knowing the links between expectation orderings and necessity functions [7], and Rad being a necessity-like function.

The Example of Figure 1 and *RAD*:

- Consider the label $(1,1)$. Nothing other than the constraint $\neg(r \wedge s)$ is known for sure, so $\rho = 0$ does not give anything interesting. The labels at the distance $\rho = 1$ are $(0,1)$, $(2,1)$, $(1,0)$ and $(2,0)$. Thus we have $Th((1,1),1,H) = Cn(\{\neg s, r \vee s\})$. As to $Th((1,1),2,H)$, it is inconsistent. Thus we have $\mid\sim^{RAD}_{H,(1,1)} r \wedge \neg s$.

- Consider the label $(1,2)$. We have $Th((1,2),1,H) = Cn(\{r \vee s, \neg r, \neg s\})$ which is inconsistent, therefore radial accumlation does not enable any non-trivial inference at $(1,2)$. Remember that at this label the projection inference relation was less cautious. The results of the application of radial accumulation to the example is depicted on Figure 2.

x y	0	1	2	3
0	r	r	$r \vee s$	s
1	$\neg s$	r		$\neg r$
2	$\neg r \wedge \neg s$	r	r	
3	$\neg r \wedge \neg s$	$\neg s$	r	r

Figure 2: radial accumulation applied to the spatial example (cf Fig.1)

A generalisation of *RAD* to the general (non-complete) case would define a preferential inference relation. We omit a discussion of this for lack of space.

5 Extrapolation by syntax-based entailment

Syntax-based approaches to nonmonotonic reasoning (see [17]) consider each formula of the knowledge base as an independent piece of information; they often make use of priorities ([3], [17], [1], [16]) without specifying where these priorities come from. Our methods for defining priorities based on distances can be combined with the usual methods for defining a syntax-based inference relation, especially the definition proposed in [17], [3], and *lexicographic entailment*, proposed in [1] and [16]. As argued in [1], these inference relations avoid the *drowning effect*, in contrast to radial accumulation. In order to apply distance based priorities in this case, however, we have to assume that the number of possible distance values is finite, due to the fact that there must be a finite number of priority levels.

6 Selected applications

6.1 Temporal reasoning

Although our first motivation was directed towards applications in spatial reasoning, our approach also has some relevance to temporal reasoning, in particular the persistence problem. Roughly, the persistence problem consists in extrapolating the truth value of fluents at some time points where it is unknown. The intuitive idea of our approach in the temporal domain is that our belief in the persistence of properties gradually grows weaker and weaker as time continues. Traditional approaches to persistence (see [19] for a critical survey) do not use this gradual notion; thus in the case where a fluent φ is known to be true at t_0 and false at a later time point t_1 (nothing else being known), skeptical approaches do not conclude anything about φ within (t_0, t_1), which is too cautious, and chronological approches such as [20] conclude that φ changes its truth value at the last possible time point (i.e. at t_1). In contrast, our approach extrapolates the truth value of φ to TRUE(resp. FALSE) at time-points close to t_0 (resp. t_1). Now, it is clear that our basic approach, used alone, cannot handle examples with laws about explicit changes (updates, actions, ...) such as the Yale Shooting Problem. The latter cannot be even represented in our language; our approach applies only to problems of pure persistence without causality. The next step toward a treatment of "realistic" temporal (whose technical details are beyond the scope of this paper) would consist in integrating distance-based persistence in the minimization criterion to select the preferred models in a given framework for handling changes with preferential entailment (for instance Sandewall's approach [19]).

Besides traditional approaches, we know two gradual approaches to persistence. Dean and Kanazawa's [5] probabilistic projection extrapolate a probabilistic persistence function, and it may be considered as the probabilistic counterpart of the application of our work to temporal reasoning. It is discussed in [6], where a more qualitative, possibilistic approach is proposed. Such approaches have the advantage of potentially solving another problem that affects the application of our distance based inference relations to temporal reasoning (as well as that of many others) All of the latter treat the persistence of all fluents in the same way. While we could separate totally persistent fluents such as *dead* from usual fluents (simply in the closure phase of the labelled knowledge base: the labelled theory must be such that if we know *dead* at t_0, then we know *dead* at any later time point), we cannot express the fact that fluents persist in different ways: for instance, *married* usually persists longer than *asleep*; some fluents are periodic, for instance $red-traffic-light$; some fluents may even be chaotic, i.e. they do not tend at all to persist.

6.2 Spatial reasoning (reasoning by proximity)

Our original motivation was to treat the persistence of properties over spatial regions. In the persistence of properties over spatial boundaries without known boundaries, agronomists and others have used, though not formally defined or studied, pseudo-metric based systems of inference. They have not attempted a formal study of persistence principles of the kind we have pursued here. The only work we know on this subject is [13]: the authors point out that some notions used in temporal reasoning transfer to spatial reasoning and they define then *spatial persistence*; they use distances for relaxing a spatial constraint satisfaction problem (using a fuzzy model). But their framework is not oriented towards deduction; it is oriented towards finding a solution. Nevertheless, the underlying ideas are related to ours.

For us, spatial reasoning corresponds to *nonmonotonic reasoning by proximity*: we extrapolate by taking account of the closest points of the space. Here is a plausible example of such reasoning having to do with weather. For example, if a pilot is making a trip from A to B and he asks for weather at three stations along his route and the weather briefer tells him: Station 1 rain, Station 2 no report, Station 3 rain, the pilot and the weather briefer will extrapolate concerning station 2's weather from the reports given at stations 1 and 3. If station 2 is between stations 1 and 3, then both RAD and PROJ will allow us to infer RAIN at station 2. To be sure this is not the only sort of information that the pilot might consider; he might have special knowledge of the terrain at station 2 that would override or defeat this inference—for example, the fact that station 2 is in a deep desert valley where the weather is usually much better than at stations 1 and 3).

In spatial reasoning as in temporal reasoning, we should not treat the peristence of spatial propositions completely uniformly; some tend to "persist through space" more than others. Further, in addition to our relations $PROJ$ and RAD, we would like to investigate more empirical, more quantitative but perhaps more realistic inference relations for space. Finally, we should exploit geometrical reasoning in performing the logical closure giving H within our distance based account of persistence; for instance, if we want to delimit an object which we know to be, say, convex, we can deduce many facts from a few punctual observations. Object boundaries will obviously affect persistence of spatial propositions.

6.3 Taxonomic reasoning

Distance based nonmonotonic inference relations also have a natural application in taxonomic reasoning, though we know of no work in AI other than our own in this area. Fields like classification theory, however, have used distance based reasoning, though again the formal study of inference relations based on distances have not been studied. One simple application of *PROJ* and *RAD* in taxonomic domain concerns a classification task. Consider, for instance, the problem of classifying a particular mushroom as being of a given type. Our "space" is in this case a space of features rather than physical distances. We propose two different ways to define such distances.

Feature distances

The label set X is now a conceptual net with nodes labelled by concepts $C_1, ..., C_n$ or by individuals $a_1, ..., a_m$. With each node is also associated a list of features. The labelled theory consists in specifying which features are associated to which nodes. The features give rise to a number of natural distance metrics between nodes. One, for instance, comes from the Jaccard Index (let $F_1, ...F_p$ denote the different features): $d(x,y) = \frac{|\{F_i : F_i \in H(x) \cap H(y)\}|}{|\{F_i : F_i \in H(x) \cup H(y)\}|} - 1$

With these preliminaries in mind, we return to the categorization example. We are interested in categorizing a particular object a_0 with features $F_1, ..., F_n$. In our example, a_0 might be a particular mushroom whose type we are trying to identify. Types of mushrooms (cepes, girolles, amanitas, etc.) also are nodes in our feature space. We will suppose that each type node is exclusive so (so for instance $\neg(cepe \wedge girolle)$ holds at all nodes. Suppose now using the Jaccard measure that the closest nodes to a_0 at which *cepe* is either true or false make *cepe* true. Then by the projection inference relation, we conclude $\mid\!\sim^{PROJ}_{H,a_0} cepe$, i.e. *cepe* is true at a_0. Another way to employ the distance based inference relations in taxonomic reasoning is to extrapolate particular properties about a_0 that are not known from the original inspection. Continuing our example above, if *cepe* is the closest node to a_0 at which the property "edible" is determined, then $\mid\!\sim^{PROJ}_{H,a_0} edible$.

Distances based on counting individuals

Taxonomic reasoning is usually thought to verify a specificity property like the following. If A strictly implies B and A's by default are not C's and B's are by default C's, then given that we know that something is an A and a B, then we by default infer that it is not C. To capture inheritance, we need a different kind of distance metric than the one constructed using the Jaccard Index. We fix a model M, captueing the" real world" extensional (i.e. cardinality and membership) relations between individuals and types in our conceptual space. Then a metric appropriate to the verification of the specificity property with both the projective and the radial methods is familiar from conditional probability:

$$d(C_i, C_j) = \frac{|\{x | M \models C_i(x)\}|}{|\{x | M \models C_i(x) \wedge C_j(x)\}|} - 1 = \frac{1}{Prob(C_j | C_i)} - 1$$

Note that we have $d(C_i, C_j)$ is minimal (and equals 0) iff $C_i \subseteq C_j$ and maximal (and equals $= \infty$) iff $C_i \cap C_j = \emptyset$.

As an example of specificity reasoning, consider the network of nodes in which we have Penguin (p), Bird (b), Flying-animals (fa) and Non-flying-animals (nfa) as nodes, with an assignment of extensions to these types in a model as one would expect (i.e. that model the real world or at least our expectations about the real world). As all penguins are birds, we have $d(p,b) = 0$. If the model is as we expect, then $d(p, nfa) < d(p, fa)$ (or equivalently $Prob(nfa|p) > Prob(fa|p)$), and $d(b, fa) < d(b, nfa)$. Now we would like to reason about what happens if we know that something is a penguin. Using any of the proposed inference relations and the previous distance, we get that $\mid\!\sim_{H,p} \neg fly$ since the closest node to p where fly is determined is nfa where fly is false.

It may be that nonmonotonic reasoning in general and taxonomic reasoning in particular uses a variety of measures are used; we select the appropriate measure because it is suited to a particular task. Hence, the distances based on the Jaccard index and on conditional probability might be suited to different taxonomic reasoning tasks.

7 Conclusion

We have proposed a general principle for defining several nonmonotonic inference relations from distances with applications to temporal, spatial and taxonomic reasoning. This methodology is an alternative way to define already known nonmonotonic inference relations (e.g., the comparative inference relation): here our premises do not include a set of conditional assertions or default rules as in most approaches, but a set of temporally or spatially indexed data (among other possibilities). It should be possible to use our approach to show how many formalisms already studied from a theoretical point of view apply in this way to spatially and temporally indexed data. As to complexity results for our nonmonotonic inference relations, some results can be directly taken from [12] and [17], and for the projection inference relation, the associated decision problem is obviously in Δ_2^P).

References

1. Salem Benferhat, Claudette Cayrol, Didier Dubois, Jérôme Lang and Henri Prade, *Inconsistency management and prioritized syntax-based entailment*, Proceedings of IJCAI'93,640-645.
2. Craig Boutilier, *Conditional logics for default reasoning*, PhD thesis, University of Toronto, 1992.
3. Gerd Brewka, *Preferred subtheories: an extended logical framework for default reasoning*, Proceedings of IJCAI'89, 1043-1048.
4. Gabriella Crocco and Philippe Lamarre, *On the connection between nonmonotonic inference relations and conditional logics*, Proceedings of KR'92, 565-571.
5. Thomas Dean and Keiji Kanazawa, *A model for reasoning about persistence and causation*, Computational Intelligence, 5(3): 142-150.
6. Dimiter Driankov and Jérôme Lang, *Possibilistic decreasing persistence*, Proceedings of Uncertainty in AI'93, 469-476.
7. Didier Dubois and Henri Prade, *Epistemic entrenchment and possibilistic logic*, Artificial Intelligence 50:223-239, 1991.
8. Luis Fariñas del Cerro, Andreas Herzig and Jérôme Lang, *From ordering-based nonmonotonic reasoning to conditional logics*, Artificial Intelligence 66, 375-303, 1994.
9. Dov Gabbay, *Labelled Deductive Systems*, Technical Report, Centrum für Informations und Sprachverarbeitung, Universität München.
10. Peter Gärdenfors and David Makinson, *Revision of knowledge systems using epistemic entrenchment*, Proceedings of TARK'88.
11. Peter Gärdenfors and David Makinson, *Nonmonotonic reasoning based on expectations*, Artificial Intelligence.
12. Georg Gottlob, *Complexity results for nonmonotonic logics*, Journal of Logic and Computation 2 (3), 397-425, 1992.
13. Hans Werner Guesguen and Joachim Hertzberg, *Spatial persistence*, Proceedings of the Ijcai Workshop on Temporal and Spatial Reasoning (Chambéry, 1993), 11-34.
14. Sarit Kraus, Daniel Lehmann and Menahem Magidor, *Nonmonotonic reasoning, preferential models and cumulative logics*. Artificial Intelligence 44 (1990), 167-207.
15. Daniel Lehmann and Menahem Magidor, *What does a conditional knowledge base entail?*, Artificial Intelligence 55 (1992), 1-60.
16. Daniel Lehmann, *Another perspective on default reasoning*, Technical Report 92-12, Hebrew University of Jerusalem, 1992.

17. Bernhard Nebel, *Belief revision and default reasoning: syntax-based approaches*, Proceedings of KR'91, 417-428.
18. Judea Pearl, *System Z: a natural ordering of defaults with tractable applications to default reasoning*, Proceedings of TARK'90, 121-135.
19. Erik Sandewall, *Features and fluents: a systematic approach to the representation of knowledge about dynamical systems*, Technical Report LiTH-IDA-R-92-30, Dept of Computer and Information Science, Linköping University, 1992.
20. Yoav Shoham, *Chronological ignorance: experiments in nonmonotonic temporal reasoning*, Artificial Intelligence 36 (1988), pp. 279-331.