Natural Ideal Operators in Inductive Logic Programming

Fabien Torre and Céline Rouveirol

Inference and Learning Group,
Laboratoire de Recherche en Informatique
Bâtiment 490, Université Paris-Sud
91405 - Orsay Cedex (France)
{fabien,celine}@lri.fr

Abstract. We present in this paper the original notion of *natural relation*, a quasi order that extends the idea of generality order: it allows the sound and dynamic pruning of hypotheses that do not satisfy a property, be it completeness or correctness with respect to the training examples, or hypothesis language restriction.
Natural relations for conjunctions of such properties are characterized. Learning operators that satisfy these complex natural relations allow pruning with respect to this set of properties to take place before inappropriate hypotheses are generated.
Once the natural relation is defined that optimally prunes the search space with respect to a set of properties, we discuss the existence of ideal operators for the search space ordered by this natural relation. We have adapted the results from [vdLNC94a] on the non-existence of ideal operators to those complex natural relations. We prove those non-existence conditions do not apply to some of those natural relations, thus overcoming the previous negative results about ideal operators for space ordered by θ-subsumption only.

1 Introduction

A search problem consists of a set of states (the search space), a set of operators, an initial state and a goal state. The goal state may not be explicitly described, but may rather be specified through a set of properties it must satisfy.

After [Mit82], "*generalization problem is essentially a search problem*". In the framework of definite semantics, the search space is a set of definite clauses, the initial state is given by a positive example of the target concept and the operators are learning operators that alter a hypothesis clause for the target concept into a set of new and possibly better hypotheses. The learning goal is classically defined in Inductive Logic Programming (ILP) as follows [MR94]: given a set E^+ of positive examples and a set E^- of negative examples for the target concept, a background knowledge B, find a hypothesis H such that

$$\begin{array}{lll} \forall e^+ \in E^+ : & B \cup H \models e^+ & (H \text{ is complete}) \ , \\ \forall e^- \in E^- : & B \cup H \not\models e^- & (H \text{ is correct}) \ . \end{array}$$

In ILP, the cost of exhaustively exploring the search space in order to find a hypothesis that satisfies the completeness and correctness criteria is prohibitive, and several techniques have been developed to prune the search space, either statically (before search) or dynamically (during search).

On the one hand, a well-known pruning technique [Mit82] is to exploit the generality ordering on the learning search space. For instance, given that the search proceeds bottom up with respect to a generality ordering, and that a given hypothesis H in the search space covers a negative example, it is not necessary to develop any of its generalizations, as none of them will ever meet the correctness criterion anymore.

On the other hand, pruning techniques exploiting additional constraints on the expected target concept definition known as *learning bias*, have also been extensively studied in ILP (see [NRA+96] for a survey). In particular, *language bias* allows constraint setting on the syntax of the target concept definition. Handling language bias, by pruning hypotheses which are irrelevant with respect to a specific learning problem, adapts the learning process to the problem at hand, and enhances both the quality and the efficiency of learning.

However, in learning systems that search a hypothesis space ordered by a generality relation, the handling of language bias may be expensive: hypotheses generated by the learning operator may not always satisfy the language bias, which therefore have to be tested at each learning step. Even worse, it may well be that a hypothesis does not satisfy a language bias, and that some of its descendants through the learning operator satisfy it: even if a hypothesis fails against the language bias test, its descendants nevertheless have to be generated and tested.

In this paper, we study a smoother and more efficient way to integrate language bias handling in learning. As opposed to systems that only use a generality ordering to explore and prune the search space, the idea is to also take into account the language bias to order the search space. In that aim, we propose an extended definition for the learning task: given a set of properties $\mathcal{P} = \{P_1, \ldots, P_n\}$, find a hypothesis H such that $P_1(H) \wedge \ldots \wedge P_n(H)$. The P_i necessarily include at least completeness or correctness with respect to the examples of the target concept. To deal with this new definition, new quasi-orders called *natural relations* are designed that allow optimal pruning with respect to training example coverage *and* a subset $\mathcal{P}_{ord}$ of $\mathcal{P}$. This pruning is dynamic : when a generated hypothesis H does not satisfy $\mathcal{P}_{ord}$, all the descendants of H can be safely pruned. Roughly speaking, this amounts to pruning the search space with respect to both a generality relation and some language bias, as done previously with respect to a generality order only. This saves the cost of generating and testing inappropriate hypotheses.

On a more theoretical basis, ILP related works have studied different quasi-orderings for First Order Logic (FOL) search spaces: θ-subsumption [Plo70], generalized subsumption [Bun88], T-implication [IA95], or logical implication [NCdW96], and have formalized learning operators as *refinement operators* that go through a quasi-ordered space of clauses [Sha81, Nib93, vdL95]. Given a

quasi-ordered space, [vdLNC94a] has defined the notion of *ideal* operator. Intuitively, the ideality property for an operator ensures that given any hypothesis H of the search space, the set of alterations of H is finite, that none of such alterations of H is equivalent to H, and that the operator is complete. Ideality seems reasonable to expect from a learning operator, but it has been proved that no ideal operator exists for unrestricted search spaces ordered by θ-subsumption [vdLNC94a] or by logical implication [vdLNC94b]. Given our natural relations, we then check whether some ideal operators may exist for the search space ordered by a natural relation, by adapting the conditions introduced in [vdLNC94a, vdLNC94b]. We find a number of favorable cases, in which those non-existence conditions do not apply. Finally, we exhibit an ideal operator for the search space ordered by a given natural relation.

The paper is organized as follows. In section 2, we give general definitions necessary to introduce our framework. In section 3, we define $\geq_{\mathcal{P}}$, the *natural relation* for a property set $\mathcal{P}$. We discuss in section 4 the necessary conditions for the existence of ideal learning operators for sets ordered by a natural relation and, in section 5. As an illustration, in a specific case for which these conditions are met, a new ideal operator is proposed. We conclude by situating this work with respect to previous approaches in ILP and by describing its perspectives. For the sake of brevity, proofs have been omitted from this paper, but full proofs for all propositions can be found in [TR97].

2 Definitions

2.1 Refinement Operator

In this paper, we consider a *refinement operator* as a binary relation on the search space. An operator $\mathcal{O}$ is then represented by the set of pairs (H, H'), such that $H' \in \mathcal{O}(H)$. Typically, the search space is a set of clauses S ordered by a generality relation $\geq$. A binary relation $\mathcal{O}$ is an *upward refinement operator* for $\langle S, \geq \rangle$ iff $\forall C, D \in S : C \mathcal{O} D \Rightarrow C \geq D$. In other words, $\mathcal{O}$ is a refinement operator iff

$$\forall C \in S : \{D \in S \mid C \mathcal{O} D\} \subseteq \{D \in S \mid C \geq D\} \ .$$

In the same way, a binary relation $\mathcal{O}$ is a *downward refinement operator* for $\langle S, \geq \rangle$ iff $\forall C, D \in S : C \mathcal{O} D \Rightarrow D \geq C$. We note $\mathcal{O}(C)$ the set $\{D \in S \mid C \mathcal{O} D\}$. Then, we will say that an operator $\mathcal{O}$ *satisfies* a relation $\mathcal{R}$ iff $\mathcal{O} \subseteq \mathcal{R}$. [vdLNC94a] has introduced a class of theoretically interesting operators, *ideal* operators.

2.2 Ideality

We present here a summary of definitions and results on ideal operators from [vdLNC94a, vdLNC94b, vdL95].

A refinement operator $\mathcal{O}$ is said *ideal* for $\langle S, \geq \rangle$ if it is locally finite, complete and proper for $\langle S, \geq \rangle$.

1. A refinement operator $\mathcal{O}$ is said *locally finite* for $\langle S, \geq \rangle$ iff $\forall C \in S : \mathcal{O}(C)$ is finite and computable, i.e., for any hypothesis H of $\langle S, \geq \rangle$, the set of refinements of H through $\mathcal{O}$ is finite and computable, which is the least we can expect for any practical implementation of $\mathcal{O}$.
2. A downward refinement operator $\mathcal{O}$ is said *complete* for $\langle S, \geq \rangle$ iff $\forall C, D \in S : C \geq D \Rightarrow \exists E \in S : E \in O^*(C) \wedge E \sim D$, where $\mathcal{O}^*$ denotes the transitive closure of $\mathcal{O}$, and $\sim$ denotes the equivalence relation of $\geq$. An upward refinement operator $\mathcal{O}$ is said *complete* for $\langle S, \geq \rangle$ iff $\forall C, D \in S : D \geq C \Rightarrow \exists E \in S : E \in O^*(C) \wedge E \sim D$. This means that, given any two comparable hypotheses C and D of $\langle S, \geq \rangle$, there exists a refinement chain from C to D. In other words, if the target concept definition belongs to $\langle S, \geq \rangle$, a complete learning operator will reach it.
3. A refinement operator $\mathcal{O}$ is said *proper* for $\langle S, \geq \rangle$ iff $\forall C, D \in S : D \in \mathcal{O}(\mathcal{C}) \Rightarrow \mathcal{C} \not\sim \mathcal{D}$. This last property ensures that set of refinements of H does not contain any hypothesis equivalent to H. Intuitively, at each application of $\mathcal{O}$, the rank of hypotheses with respect to the quasi-order increases.

There are two situations in which ideal operators do not exist: there exist *uncovered infinite chains* or infinite *cover set* in the search space.

The notion of *cover* is used to represent the immediate successors or predecessors of a hypothesis with respect to $\geq$.

Definition 2.1 (cover). C covers D iff $C > D$ and there is no E such that $C > E > D$. We call C an *upward cover* of D and D a *downward cover* of C.

Definition 2.2 (cover sets). A *downward* (resp. *upward*) *cover set* of a clause C in $\langle S, \geq \rangle$ is a maximal set of non-equivalent downward (resp. upward) covers of C.

Example 2.3 (downward cover sets). Let us consider a set $\{a, b, c, d, e\}$ ordered as shown on figure 1. a has two possible downward cover sets: $\{b, d\}$ or $\{c, d\}$ (b and c are equivalent and so, they cannot appear in the same cover set).

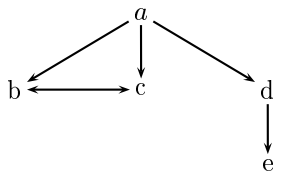

Figure 1. Example of quasi-ordered set.

It has been proved that a cover set of a hypothesis has to be included in its set of refinements through an ideal operator. As a consequence, two problems may raise:

– the cover set cannot be computed: given two hypotheses C and D, such that D is more general than C, there is an infinite number of hypotheses in between C and D, therefore completeness is lost[1];
– the cover set is infinite, which implies, that the production of an ideal operator would be infinite also and hence, the operator uncomputable.

The first case is formalized through the notion of *uncovered infinite chain*.

Definition 2.4 (uncovered infinite chains). An *uncovered infinite (strictly) descending chain* for $\langle S, \geq \rangle$ is an infinite set $\{D_i\}_{i \geq 1}$ where $D_i \in S$ such that $D_1 > D_2 > \ldots > D_n > D_{n+1} > \ldots > C$ and C has no upward cover E in S covered by all D_i. An *uncovered infinite (strictly) ascending chain* for $\langle S, \geq \rangle$ is an infinite set $\{A_i\}_{i \geq 1}, A_i \in S$ such that $C > \ldots > A_{n+1} > A_n > \ldots > A_2 > A_1$ and C has no downward cover F in S covering all A_i.

Example 2.5 (uncovered infinite descending chain). Figure 2 illustrates the descending chain definition on an example: under θ-subsumption, $\{D_i\}_{i \geq 2}$ is an uncovered infinite descending chain of C with:

$$\begin{cases} C & : q(X_1) \leftarrow p(X_1, X_1) \ , \\ D_n & : q(X_1) \leftarrow \{p(X_i, X_j) | 1 \leq i, j \leq n, i \neq j\} \ . \end{cases}$$

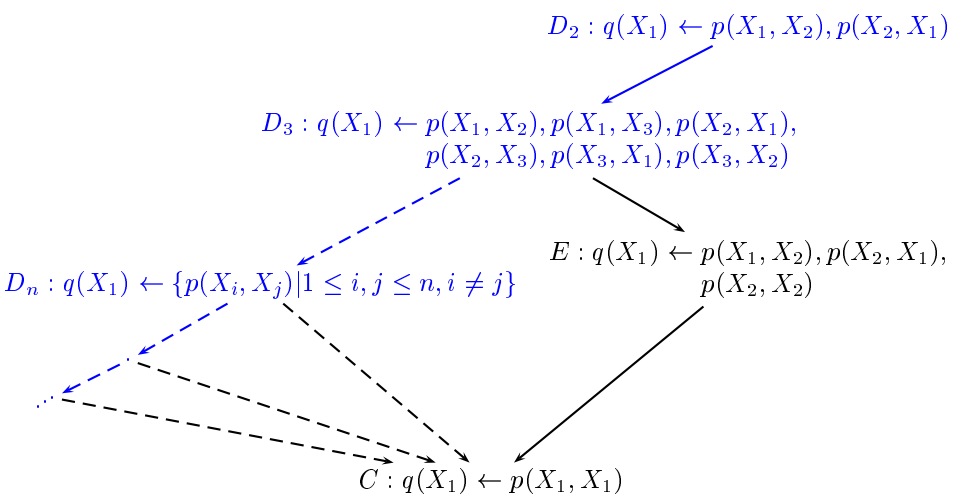

Figure 2. Uncovered infinite descending chain.

[1] One may think, for instance,of the set $\mathbb{R}$ of reals: for a given real, it is impossible to build a real immediately greater.

2.3 Properties

As quoted in the introduction, we indifferently include in the set of expected properties for the target concept completeness and correctness with respect to training examples and language bias constraints. As quoted in the introduction, we can find in ML and ILP literature a large range of language restrictions that a target concept definition should satisfy. We detail in the following the ones we have focused our study on. This list is largely representative of language biases classically used in ILP. Besides, as we will see in next section, it can be considered as indicative, as our framework is general enough to integrate new properties.

- A hypothesis must, or must not cover a given example.
- One may impose a (upper or lower) bound on
  - number of existential variables (variables not occurring in the head of the clause),
  - number of literals in the body of the clause,
  - depth [MF90]. The depth of a clause is the maximal depth of its terms. Variables and constants have depth zero. A term $f(\ldots, t_i, \ldots)$ has depth one plus the maximal depth of t_i.
  - level [DMR92]. The level of a clause is the maximum of the level of its variables. Variables of the head have level zero. The level of an existential variable V is one plus the minimal level of variables appearing in the first literal containing V.
- One may impose a hypothesis to be
  - range-restricted (variables of the head must appear in the body),
  - connected [Qui90] (level is defined for all variables of the hypothesis),
  - reduced for θ-subsumption (C is reduced iff $\nexists \theta : C\theta \subset C$).

In the above list, some of the language biases have been specifically developed for FOL languages. On the one hand, those language biases have an interest as they allow to express meaningful information about the expected form of the concept. For instance, the range-restriction property states that a meaningful definition for the concept "X is the grandfather of Y", should contain constraints on X and Y. On the other hand, those language biases define subsets of FOL for which the coverage test of hypotheses with respect to examples is (relatively) efficient. Finally, any hypothesis that fails against a language bias does not have to be checked against examples, which may save a lot of computation efforts.

Now the use of bias in ILP is motivated, we show in next section that the only way to prune the search space with respect to a given property is to explore it with an operator that satisfies a relation "induced" by this property, namely, its natural relation.

3 Natural Relations

Idestam-Almquist [IA95] says "*Implication is the most natural and straightforward basis for generalization in inductive learning, since the concept of induction can be defined as the inverse of logical entailment*". We give here a more formal justification of why logical implication is the most natural relation to order a search space of hypotheses when solutions are defined with respect to at least completeness or correctness criteria. We extend the notion of natural relation to other properties.

3.1 Private property

The aim is to explore a small search space without risking to miss a solution. It is safe to stop the refinement of a given H that does not satisfy the expected properties (and therefore to prune the search space), iff no descendant of H will ever satisfy those properties. This intuition is illustrated on figure 3, and formalized in the following definition.

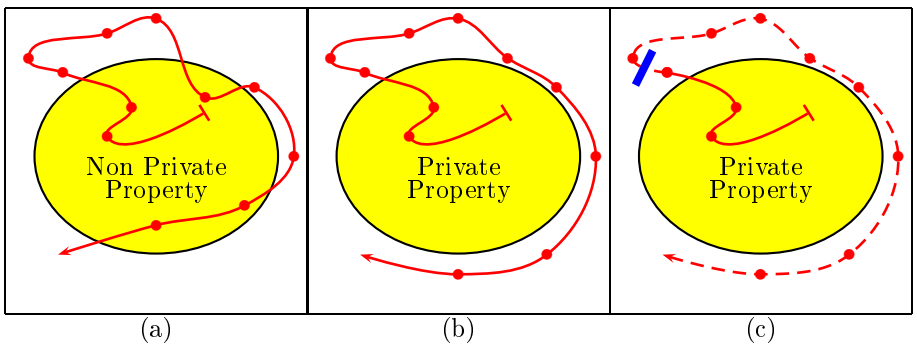

Figure 3. Private and non private properties. Figure (a) illustrates that safe pruning is not possible with respect to non private property. In case (b), pruning is sound and (c) shows how pruning operates.

Definition 3.1 (private property). A property P is said *private* with respect to the relation $\mathcal{R}$ iff

$$\forall H, H' \in S: \quad \forall \left(H \mathcal{R} H' \wedge \overline{P(H)} \Rightarrow \overline{P(H')} \right) \ ^{2}.$$

Let us assume the property P is expressed by a FOL formula. The parameters of the property are the free variables in this formula.

[2] If F is a formula, then $\forall F$ denotes the universal closure of F, which is the closed formula obtained by adding a universal quantifier for every variable having a free occurrence in F [Llo87]. The semantic of $\overline{F}$ is the negation of F.

Example 3.2 (length of clause). Let us consider the property that bounds the length of clause to k literals, expressed as $|H| \leq k$. The property LENGTH_k is private with respect to the relation $\mathcal{R}$ iff

$$\forall H, H' \in S : \quad \forall k \in \mathbb{N} : (H \mathcal{R} H' \wedge |H| > k \Rightarrow |H'| > k) \quad .$$

A relation $\mathcal{R}$ that satisfies this property is, for instance, the one defined by: $H \mathcal{R} H'$ iff there exists a literal L such that $H' = H \cup \{L\}$.

3.2 Natural relation

A property is private for many relations. For instance, the empty and identity relations make any property private and are, as a consequence, of little interest, as far as the pruning of the search space is concerned. Indeed, an operator satisfying the identity relation applied to a given hypothesis H, generates at most H itself.We will consider in the following that a natural relation for a given property is one of the largest relations that make this property private.

Definition 3.3 (natural relation). $\mathcal{R}$ is a *natural relation* for the property P iff

1. P is private with respect to the relation $\mathcal{R}$;
2. if a relation $\mathcal{R}'$ exists such that P is private with respect to the relation $\mathcal{R}'$ and $\mathcal{R} \Rightarrow \mathcal{R}'$, then $\mathcal{R} = \mathcal{R}'$.

Let us try to give a more precise characterization of this natural relation. We will see in the following that the natural relation of a property is indeed unique. Two hypotheses are "naturally" related if, for all possible instantiations of the property parameters, either the first hypothesis satisfies the property or the second one does not.

Proposition 3.4. *A property P has a single natural relation, denoted $\geq_P$. $\geq_P$ is defined by*

$$\forall C, D \in S : \quad C \geq_P D \Leftrightarrow \forall \left(P(C) \vee \overline{P(D)} \right) \quad .$$

Therefore, a property P is private with respect to the relation $\mathcal{R}$ iff $\mathcal{R} \Rightarrow \geq_P$. This result justifies why we have chosen the largest relation as natural relation: a relation makes a property private iff this relation is included in the natural relation of the property. Therefore, safe and dynamic pruning of the search space with respect to a given property can only be achieved through an operator which satisfies its natural relation.

Remark 3.5. Given the definition of downward refinement operator (section 2.1), if $\mathcal{O}$ is downward with respect to $\geq_P$ then P is private with respect to $\mathcal{O}$.

We assume, in the remainder of this paper, that any property can be expressed as $f(H) \mathcal{R} k$ where

- f is a function from the search space S to a domain $\mathcal{D}_f$;
- $\mathcal{R}$ is a quasi-order on $\mathcal{D}_f \times \mathcal{D}_f$;
- $k \in \mathcal{D}_f$ (k represents the parameter of the property P).

Table 1 shows our properties expressed in this framework. For each property of this table, the dual property can be considered[3].

Table 1. Properties expressed in our framework.

COVER$_e$	$H \models e$
RANGE-RESTRICTED$_b$	$\mathrm{RR}(H) = b$
CONNECTED$_b$	$\mathrm{Connected}(H) = b$
REDUCED$_b$	$\mathrm{Reduced}(H) = b$
LENGTH$_n$	$\lvert H \rvert \leq_{\mathbb{N}} n$
VARS$_n$	$\mathrm{Nv}(H) \leq_{\mathbb{N}} n$
DEPTH$_d$	$\mathrm{Depth}(H) \leq_{\mathbb{N}} d$
LEVEL$_l$	$\mathrm{Level}(H) \leq_{\mathbb{N}} l$

This assumption on the form of the property allows the simplification of the expression of natural relation.

Proposition 3.6. *Let P be a property defined by $\forall C \in S : P(C) \Leftrightarrow f(C)\ \mathcal{R}\ k$. The natural relation of P is also defined by*

$$\forall C, D \in S : \quad C \geq_P D \Leftrightarrow f(C)\ \mathcal{R}\ f(D)$$

Example 3.7. Let us consider the property COVER$_e$ stating that a hypothesis H covers an example e, $H \models e$. Function f is here identity, $\mathcal{R}$ is identified to $\models$, the parameter of the function is e. By direct application of the above proposition, the natural relation for COVER$_e$ is defined by $\forall C, D \in S : C \geq_{\mathrm{Cover}} D \Leftrightarrow C \models D$.

Thus, saying that an operator makes COVER$_e$ private is equivalent to imposing that this operator satisfies a generality ordering. This justifies that a refinement operator must satisfy a generality ordering. The notion of natural relation extends that of generality order to all properties a target concept must satisfy. Table 2 lists the natural relations for all the target concept properties we address.

Now that the natural relation for a single property has been characterized, we explore how to compute the natural relation of its dual property and then, and the natural relation of a conjunction of properties, as there are usually more than one imposed on the target concept definition.

[3] For a given property P, we call the dual P, the property $\overline{P}$.

Table 2. Natural relations.

Property P	Relation $\forall C, D \in S : C \geq_P D \Leftrightarrow$
COVER$_e$	$C \models D$
RANGE-RESTRICTED	$\mathrm{RR}(C) = \mathrm{RR}(D)$
CONNECTED	$\mathrm{Connected}(C) = \mathrm{Connected}(D)$
REDUCED	$\mathrm{Reduced}(C) = \mathrm{Reduced}(D)$
LENGTH$_n$	$\lvert C \rvert \leq_{\mathbb{N}} \lvert D \rvert$
VARS$_n$	$\mathrm{Nv}(C) \leq_{\mathbb{N}} \mathrm{Nv}(D)$
DEPTH$_p$	$\mathrm{Depth}(C) \leq_{\mathbb{N}} \mathrm{Depth}(D)$
LEVEL$_l$	$\mathrm{Level}(C) \leq_{\mathbb{N}} \mathrm{Level}(D)$

Proposition 3.8 (dual property). *Let P be a property and, $\geq_P$, its natural relation, $\forall C, D \in S : \quad D \geq_P C \Leftrightarrow C \geq_{\overline{P}} D$.*

The natural relation of a dual property is the inverse relation of the natural relation. After remark 3.5, a downward operator with respect to $\geq_P$ makes this property private and allows the dynamic pruning of hypotheses that do not satisfy this property. Therefore, the corresponding upward operator deals with the dual property.

Example 3.9. The natural relation for the dual of the property COVER$_e$ $\forall C, D \in S : C \geq_{\overline{\mathrm{Cover}}} D \Leftrightarrow D \models C$.

Proposition 3.10 (conjunction of properties). *Let P_1 and P_2 be two properties, $\geq_{P_1}$ and $\geq_{P_2}$, their respective natural relation,*

$$\forall C, D \in S : \quad C \geq_{P_1 \wedge P_2} D \Leftrightarrow C \geq_{P_1} D \wedge C \geq_{P_2} D \ .$$

Example 3.11. By the expression of the completeness criterion as a conjunction of COVER$_{e_i}$ for e_i ranging over the set of positive examples of the target concept, the natural relation for completeness is logical entailment, as expected.

4 Ideal Operators for naturally ordered sets

In the previous section, we have defined the notion of natural relation for a property. This natural relation characterizes the set of operators that allow dynamic pruning of the search space with respect to this property. As quoted in the introduction, there is no ideal operator for unrestricted search space ordered by θ-subsumption or logical implication. The open question is now whether some of those complex natural relations are a sound basis for ideal operators, i.e., whether we may overcome the previous negative results about the existence of ideal operator for unrestricted search space ordered by a natural relation.

Notation 4.1. We denote by $\geq_f^{\mathcal{G}}$ the natural relation obtained by combining the generality ordering $\mathcal{G}$ with the property represented as $f(H)\,\mathcal{R}\,k$.

For instance, $\geq_{||}^{\theta}$ denotes the conjunction of θ-subsumption and natural relation of LENGTH_k. In the following, we restrict the natural relation for correctness/completeness (logical implication) to θ-subsumption, which is equivalent to logical implication for most of the practical problems (see [Got87]).

When looking for an ideal operator for a conjunction of properties P_a and P_b, a simple case is when their respective natural relations are included (i.e., when one property implies the other).

Proposition 4.2. *Let $\geq_a$ and $\geq_b$ be quasi-orders such that $\forall C, D \in S : C \geq_a D \Rightarrow C \geq_b D$. An operator is ideal for $\langle S, \geq_a \cap \geq_b \rangle$ iff it is an ideal operator for $\langle S, \geq_a \rangle$.*

There are two direct applications for this proposition. Firstly, if there is no ideal operator for $\langle S, \geq_a \rangle$, then looking for an ideal operator for $\langle S, \geq_a \cap \geq_b \rangle$ is pointless. Secondly, if we find an ideal operator for $\langle S, \geq_a \cap \geq_c \rangle$ for a given $\geq_c$, then this one is ideal for $\langle S, \geq_a \cap \geq_b \cap \geq_c \rangle$ too.

In both cases, we do not have to consider $\geq_b$, since it has no incidence on the construction of ideal operators for $\langle S, \geq_a \cap \geq_b \rangle$.

Example 4.3. Since $C \geq^{\theta} D \Rightarrow \text{Depth}(C) \leq_{\mathbb{N}} \text{Depth}(D)$, there exists no ideal operator for $\langle S, \geq_{depth}^{\theta} \rangle$.

Now that this simple case is solved, let us adapt the non-existence conditions for ideal operators (section 2.2) to natural relations defined in the previous section. By doing so, we aim at finding a general method for discriminating favorable cases, i.e., natural relations for which an ideal operator may exist, from others. Theorem 4.7 and proposition 4.8 are the sufficient non-existence conditions of an ideal operator (existence of an uncovered infinite chain and infinite cover sets).

The first result is a general observation about uncovered infinite chains for θ-subsumption.

Lemma 4.4. *Let $\{D_i\}_{i\geq 1}$ be an uncovered infinite chain of a clause C for θ-subsumption. In $\{D_i\}_{i\geq 1}$, the number of variables increases infinitely.*

Remark 4.5. This lemma points out that some simple quantities infinitely increase in an uncovered infinite chain, that chain being ascending or descending: number of variables, length, and number of occurrences of at least one predicate symbol. Intuitively, the only way to have an uncovered infinite chain is to add to a clause new literals with new variables.

We now focus on relations defined by conjunctions of two relations with one of them being the natural relation $\geq_f$ of a property expressed as $f(C)\,\mathcal{R}\,k$ where f is a function from the search space to a countable (possibly infinite) set.

Lemma 4.6. *Let $\{D_i\}_{i\geq 1}$ be an uncovered infinite chain of C for $\langle S, \geq_? \cap \geq_f \rangle$. There exists a n such that, for all i great enough, $f(D_i) = n$.*

We are now able to identify quasi-ordered sets which do not contain uncovered infinite chains.

Theorem 4.7. *If $\langle S, \geq_f^\theta \rangle$ contains uncovered infinite chains then there is a subset of S in which hypotheses have the same value by f and their number of variables is not bounded.*

Unfortunately, the absence of such chains does not necessarily imply the existence of ideal operators: a quasi-ordered set without uncovered infinite chain may contain some clauses which have an infinite cover set (definition 2.2). We give here a sufficient condition to have an infinite cover set.

Proposition 4.8 (infinite cover sets). *If, for a given clause C of S (ordered by $\geq_f^\theta$), we have an infinite number of clauses D which are incomparable under θ-subsumption and such that $C >^\theta D$ and there is no E such that $C >^\theta E >^\theta D$, and $f(C) = f(D)$, then the downward cover set of C is infinite.*

5 Natural Operators

We are now able to characterize natural relations for which we may expect to construct an ideal operator. The next step is to identify such relations among the ones we have considered before. Finally, we may then exhibit one ideal operator.

Notation 5.1. The notation for operators associated to natural orderings is similar to that of natural orderings (notation 4.1). G denotes the generality ordering the natural relation is based on, and f refers to the property denoted as $f(H)\mathcal{R}k$ (see table 1 for the possible values of f). $\rho_f^{\mathcal{G}}$ denotes a downward refinement operator, $\lambda_f^{\mathcal{G}}$ denotes a upward operator with respect to $\langle S, \geq_f^{\mathcal{G}} \rangle$. The operator $\rho_f^{\mathcal{G}}$ searches through a space ordered by $\geq_f^{\mathcal{G}}$. Stating that $\rho_f^{\mathcal{G}}$ is ideal means that $\rho_f^{\mathcal{G}}$ is ideal for $\langle S, \geq_f^{\mathcal{G}} \rangle$.

After example 4.3, there exists no ideal operator for $\langle S, \geq_{depth}^\theta \rangle$. As θ-subsumption ordering does not imply any of the other natural relations of table 2, we therefore look for ideal operators for conjunctions of θ-subsumption and one such relation.

Remark 4.5 provides guidelines on how to abort uncovered infinite chains for θ-subsumption: breaking those chains amounts to bounding the number of variables (with VARS_v), the number of literals (with LENGTH_n) or the number of occurrences of predicate symbols.

No language bias allows the bounding of the latter quantity. This leads us to introduce the new following language bias: for each predicate symbol p of a given

alphabet $\mathcal{A}$, and a given clause C, we define $\mathrm{Occ}(p, C)$ as the number of occurrences of the predicate symbol p in clause C. This language bias, called MAX-OCC, consists in bounding the following value: $\mu(C) = \max\{\mathrm{Occ}(p, C)\}_{p \in \mathcal{A}}$. This bound can be global for all predicate symbols in $\mathcal{A}$, or specific bounds can be defined for each predicate symbol. For instance, one may specify that at most three benzene rings should appear in the body of a clause describing an active molecule in the mutagenesis application.

Given these three properties, we will now apply theoretical results of the previous section in order to identify relevant combinations to break uncovered infinite chains of θ-subsumption.

Proposition 5.2. *Neither $\langle S, \geq^\theta_{||} \rangle$, nor $\langle S, \geq^\theta_{Nv} \rangle$, nor $\langle S, \geq^\theta_{\mu} \rangle$ contain uncovered infinite chains.*

Theorem 4.7 cannot be applied for range-restriction, connection, reduction, depth and level: for these properties, the number of variables may increase infinitely. So, the existence of uncovered infinite chains is possible and, indeed, the uncovered infinite chain of example 2.5 is also an uncovered infinite chain for those properties.

Now that we may ensure that no uncovered infinite chain may occur neither for $\langle S, \geq^\theta_{||} \rangle$, nor for $\langle S, \geq^\theta_{Nv} \rangle$ and $\langle S, \geq^\theta_{\mu} \rangle$, we still have to consider the possible existence of infinite cover sets.

Proposition 5.3. *Some clauses in $\langle S, \geq^\theta_{Nv} \rangle$ have an infinite downward cover set.*

By exploiting the previous results, we may now build ideal operators for all three orders $\geq^\theta_{||}$, $\geq^\theta_{Nv}$, $\geq^\theta_{\mu}$. In the following, we will only provide the characterization of operator ρ^θ_μ. A similar technique may give us $\rho^\theta_{||}$ (which is similar to Shapiro's operator ρ_0), $\lambda^\theta_{||}$, λ^θ_{Nv}, λ^θ_μ and others operators associated to more complex combination of properties (see [TR97] for a full description of those ideal operators).

We introduce terms or literals which are *most general* (see [Sha81] for a similar approach).

Definition 5.4. t is a *most general term* with respect to a clause C and a variable X iff $C >^\theta C\{X/t\}$ and there is no u such that $u >^\theta t$ and $C >^\theta C\{X/u\}$; L is a *most general literal* with respect to a clause C, iff $C >^\theta C \cup \{L\}$.

One may notice that a most general literal with respect to a clause C does not unify with any literal in C (adding L to C that unifies with a literal in C would produce a clause which is θ-equivalent).

We may now introduce as an illustration the downward refinement operator ρ^θ_μ, based on θ-subsumption and MAX-OCC.

Definition 5.5 (operator ρ^θ_μ). Let C be a clause of S. Then, $D \in \rho^\theta_\mu(C)$ when exactly one of the following holds:

1. $D = C\{X/t\}$ where X is a variable of C and t is a most general term with respect to C and X.
2. $D = C\{X_1/X_2\}$ and $|C| = |D|$, where X_1 and X_2 are variables of C.
3. $D = C \cup \{p\}$, where p is a predicate symbol of arity zero and p does not occur in C.
4. $D = C \cup \{p(\ldots, Y_i, \ldots)\}$, where p is a predicate symbol (which is not a symbol without argument) such that $\mathrm{Occ}(p, C) = 0$.
5. $D = C \cup \{p(\ldots, Y_i, \ldots)\}$, where p is a predicate symbol (which is not a symbol without argument) such that $\mathrm{Occ}(p, C) = \mu(C) > 0$, and Y_i are new variables.
6. $D = C \cup \{L\}$, where L has a predicate symbol p such that $0 < \mathrm{Occ}(p, C) < \mu(C)$ and L is a most general literal with respect to C.

This operator is ideal [TR97]. Let us informally illustrate some of its features on the clause C: $q \leftarrow p(X_1, X_2), p(X_2, X_1)$. C is known as a clause with an uncovered infinite ascending chain under θ-subsumption (see [vdLNC94a] for further details). First in this chain is the clause D_1:

$$q \leftarrow p(X_1, X_2), p(X_2, X_1), p(Y_1, Y_2), p(Y_2, Y_3), p(Y_3, Y_1)$$

Along this chain, $D_{i+1} >^\theta D_i$ (by definition of uncovered infinite ascending chain) and $\mu(D_{i+1}) \geq_{\mathbb{N}} \mu(D_i)$ (see remark 4.5). Therefore, these clauses are incomparable for $\geq^\theta_\mu$: uncovered infinite chains are broken. D_i are still more specific than C and should be computable from C by our ideal operator. Assuming that p is the only predicate symbol and the constant a is the only function symbol, ρ^θ_μ applied on C will compute the following clauses:

$$\begin{array}{ll} q \leftarrow p(a, X_2), p(X_2, a) & \text{(applying subcase 1)} \\ q \leftarrow p(X_1, a), p(a, X_1) & \text{(applying subcase 1)} \\ q \leftarrow p(X_1, X_2), p(X_2, X_1), p(Y_1, Y_2) & \text{(applying subcase 5)} \end{array}$$

Further applications of ρ^θ_μ on the third refinement will yield the clause D_1.

To sum up, the search space ordered by $\geq^\theta_\mu$ does not have uncovered infinite chains any more and our operator is complete. Note that, by construction, ρ^θ_μ satisfies the natural relation of Max-Occ, and the search space can be indifferently pruned with respect to Max-Occor coverage of positive examples.

Remark 5.6. These operators exactly compute the cover set with respect to the associated natural relation. This does not prove ideality [vdL95], but they are as efficient as possible, since the minimal number of refinements is computed: in order to have an ideal operator $\mathcal{O}$, one must have for every hypotheses of S the cover set of H included in the set of refinements of H, as we compute here the exact cover set of H.

6 Conclusion and Future Research

In this paper, we have introduced new relations, called *natural relations*, that allow the optimal pruning of the search space; we have adapted non-existence

conditions of ideal operators to these natural relations. Moreover, we have designed a new language bias, MAX-OCC, which breaks uncovered infinite chains for θ-subsumption. Finally, we have proposed a new ideal operator that ideal for unrestricted search space ordered by combination of θ-subsumption and our natural relation.

The approach of [Sha81] is similar to our. It introduces a language bias, *size*, the aim of which is to make the refinement operator computable. Therefore, strong restrictions are set on *size*: *size* is valued in $\mathbb{N}$, and for a given $n \in \mathbb{N}$, the set of hypotheses such that $size(H) = n$ is finite. As opposed, we do not set any restriction on the language bias used, except that the property should be expressed as $f(H)\,\mathcal{R}\,k$.

We have adopted the framework introduced in [vdLNC94a], as [CBS95] ("empirical ordering") and [ELMS96] (quasi-ordering on DATALOG clauses). Our approach differs in that we have extended generality orderings in order to take into account the necessity to prune the search space with respect to language biases. This leads us to exhibit ideal operators for unrestricted search spaces.

Future work will extend our results for θ-subsumption to logical entailment. We also plan to overcome negative results of this paper by considering more complex conjunctions of properties.

Ideality has been defined initially to guarantee sound exploration of the search space. Our claim in that paper is that operators that dynamically prune their search space with respect to a set of properties may be ideal. We expect that this pruning capacity will yield efficient and adapted learning procedures. By any means, we will consider our natural relations with respect to other class of operators, such as optimal operators.

Finally, our aim is to produce an interactive learning system which provides most interesting operators given a learning task expressed as the set of properties which must be satisfied by the target concept. Such a system is in the line of generic learning architectures such as $\mathcal{HAIKU}$ [NRA+96], in which a prelimary version of private bias handling has been implemented.

Acknowledgments. This work is partially supported by the Esprit LTR Project n°20237 (ILP^2). The authors wish to thank C. Nédellec, M.E. Goncalves and Q. Elhaik for their helpful and constructive comments on previous versions of this paper.

References

Bun88. W. Buntine. Generalized subsumption and its application to induction and redundancy. *Artificial Intelligence*, 36:375–399, 1988.

CBS95. M. Champesme, P. Brézellec, and H. Soldano. Empirically conservative search space reductions. In L. De Raedt, editor, *Proceedings of the 5th International Workshop on Inductive Logic Programming*, pages 387–402. Department of Computer Science, Katholieke Universiteit Leuven, 1995.

DMR92. S. Džeroski, S. Muggleton, and S. Russel. PAC-Learnability of Determinate Logic Programs. In *Proceedings of 5th ACM Workshop on Computational Learning Theory*, pages 128–135. ACM Press, 1992.

ELMS96. F. Esposito, A. Laterza, D. Malerba, and G. Semeraro. Refinement of datalog programs. In *Proceedings of the MLnet Familiarization Workshop on Data Mining with Inductive Logic Programming (ILP for KDD)*, pages 73–94, July 1996.

Got87. Georg Gottlob. Subsumption and implication. *Information Processing Letters*, 24(2):109–111, January 1987.

IA95. P. Idestam-Almquist. Generalization of clauses under implication. *Journal of Artificial Intelligence Research*, 3:467–489, 1995.

Llo87. J. W. Lloyd. *Foundations of Logic Programming*. Springer, Berlin, 2 edition, 1987.

MF90. S. Muggleton and C. Feng. Efficient induction of logic programs. In *Proceedings of the 1st Conference on Algorithmic Learning Theory*, pages 368–381. Ohmsma, Tokyo, Japan, 1990.

Mit82. T. M. Mitchell. Generalization as search. *Artificial Intelligence*, 18:203–226, 1982.

MR94. S. Muggleton and L. De Raedt. Inductive logic programming: Theory and methods. *Journal of Logic Programming*, 19:629–679, 1994.

NCdW96. S.H. Nienhuys-Cheng and R. de Wolf. Least generalizations and greatest specializations of sets of clauses. *Journal of Artificial Intelligence Research*, 4:341–363, 1996.

Nib93. T. Niblett. A note on refinement operators. In P. B. Brazdil, editor, *Proceedings of the 6th European Conference on Machine Learning*, volume 667 of *Lecture Notes in Artificial Intelligence*, pages 329–335. Springer-Verlag, April 1993.

NRA+96. C . Nédellec, C. Rouveirol, H. Adé, F. Bergadano, and B. Tausend. Declarative bias in ILP. In L. De Raedt, editor, *Advances in Inductive Logic Programming*, pages 82–103. IOS Press, 1996.

Plo70. G. Plotkin. A note on inductive generalization. In *Machine Intelligence*, volume 5. Edinburgh University Press, 1970.

Qui90. J. R. Quinlan. Learning logical definitions from relations. *Machine Learning*, 5(3):239–266, 1990.

Sha81. E. Y. Shapiro. Inductive inference of theories from facts. Technical Report 192, Yale University Department of Computer Science, February 1981.

TR97. F. Torre and C. Rouveirol. Natural ideal operators in inductive logic programming. Technical report, Laboratoire de Recherche en Informatique, Université Paris Sud, 1997. To appear.

vdL95. P.R.J. van der Laag. *An Analysis of Refinement Operators in Inductive Logic Programming*. PhD thesis, Erasmus Universiteit, Rotterdam, the Netherlands, 1995.

vdLNC94a. P.R.J van der Laag and S.H. Nienhuys-Cheng. Existence and nonexistence of complete refinement operators. In F. Bergadano and L. de Raedt, editors, *Proceedings of the 7th European Conference on Machine Learning*, volume 784 of *Lecture Notes in Artificial Intelligence*, pages 307–322. Springer-Verlag, April 1994.

vdLNC94b. P.R.J. van der Laag and S.H. Nienhuys-Cheng. A note on ideal refinement operators in ILP. In S. Wrobel, editor, *Proceedings of the 4th International Workshop on Inductive Logic Programming*, volume 237 of *GMD-Studien*, pages 247–262. Gesellschaft für Mathematik und Datenverarbeitung MBH, September 1994.