Natural Ideal Operators in Inductive Logic Programming

Fabien Torre & Céline Rouveirol

Inference and Learning group

Laboratoire de Recherche en Informatique

Université Paris-Sud, Orsay

Outline of the presentation

- Classical operators do not allow a dynamic pruning with respect to both example coverage and language bias.
- Definition of new quasi-orders, *natural relations*, that allow so.
- Do ideal operators [van der Laag and Nienhuys-Cheng, 1994] exist for natural relations ?
- Conclusion, Perspectives.

Inductive Logic Programming: Definite settings

Examples and hypotheses are definite clauses.

[Muggleton and Raedt, 1994]: Given E^+ (positive training examples) and E^- (negative training examples) for a target concept, and a background knowledge B,
find a hypothesis H such that

$$\forall e^+ \in E^+ : \quad B \cup H \models e^+ \quad (H \text{ is complete}) \ ,$$
$$\forall e^- \in E^- : \quad B \cup H \not\models e^- \quad (H \text{ is consistent}) \ .$$

An illustrative example: the *grand-father* concept

pa(A,B) $\leftarrow$ f(A,B) *(pa=parent, f=father, m=mother)*

pa(A,B) $\leftarrow$ m(A,B)

gf(abraham,bart) $\leftarrow$ f(abraham,homer),f(homer,bart)

gf(grampa-bouvier,bart) $\leftarrow$ f(grampa-bouvier,marge),m(marge,bart)

$\leftarrow$ gf(mona,bart),m(mona,homer),f(homer,bart)

$\leftarrow$ gf(jackie,bart),m(jackie,marge),m(marge,bart)

gf(A,B) $\leftarrow$ f(A,C),pa(C,B)

Refinement operator & Pruning

[Mitchell, 1982]: The search should respect a generality order to allow for pruning (with respect to example coverage).

Example: adding literal operator.

$$
\begin{tikzcd}
 & & gf(A,B) \leftarrow \arrow[dll] \arrow[d] \arrow[drr] & & \\
gf(A,B) \leftarrow m(A,C) \arrow[d, dashed] & & gf(A,B) \leftarrow f(A,C) \arrow[dl] \arrow[dr] & & gf(A,B) \leftarrow pa(C,B) \arrow[dl] \\
gf(A,B) \leftarrow m(A,C), pa(C,B) & \cdots & & gf(A,B) \leftarrow f(A,C), pa(C,B) &
\end{tikzcd}
$$

Language bias

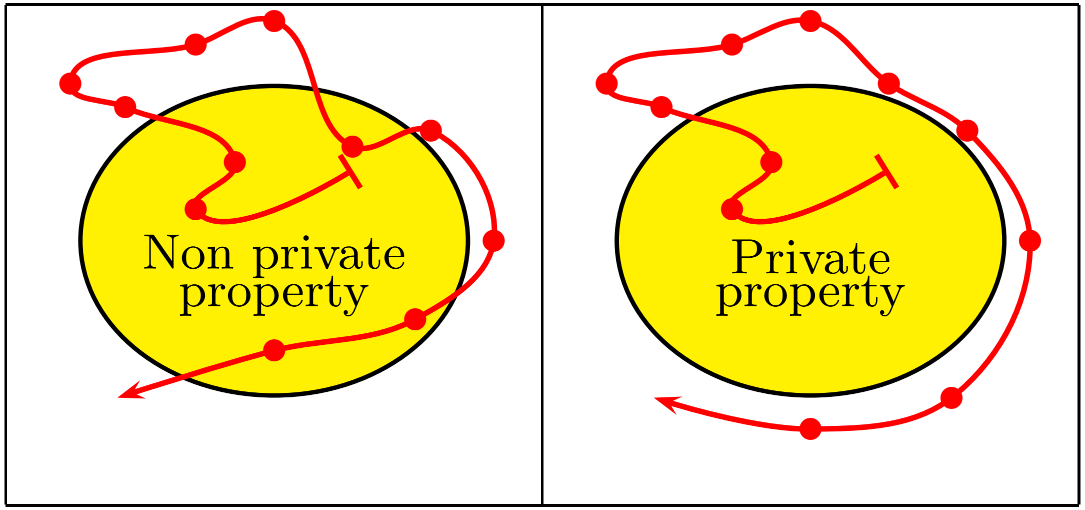
- [Mitchell, 1991]: Bias is *necessary* for learning (quality of learning results and efficiency).
- Language bias: constraints on the hypotheses syntax.
  - range-restriction, connection,
  - bound on the size, on the number of variables, on the depth of terms, etc.
- Those biases do not make the search more efficient, as dynamic pruning with respect to those constraints is in general not possible.

Example (fd)

$$\begin{tikzcd}
 & & gf(A,B) \leftarrow \arrow[dll] \arrow[d] \arrow[drr] & & \\
gf(A,B) \leftarrow m(A,C) & & gf(A,B) \leftarrow f(A,C) \arrow[dl] \arrow[dr] & & gf(A,B) \leftarrow pa(C,B) \arrow[dl] \\
 & \cdots & & gf(A,B) \leftarrow f(A,C), pa(C,B) &
\end{tikzcd}$$

Language bias: Range restriction.

Private properties



Definition 1 (private property) *A property P is* private *with respect to a relation $\mathcal{R}$ iff*

$$H \mathcal{R} H' \wedge \overline{P(H)} \Rightarrow \overline{P(H')}$$

$\mathcal{R}$ = adding literal, P = coverage of e^+ or P = size.

Natural relations

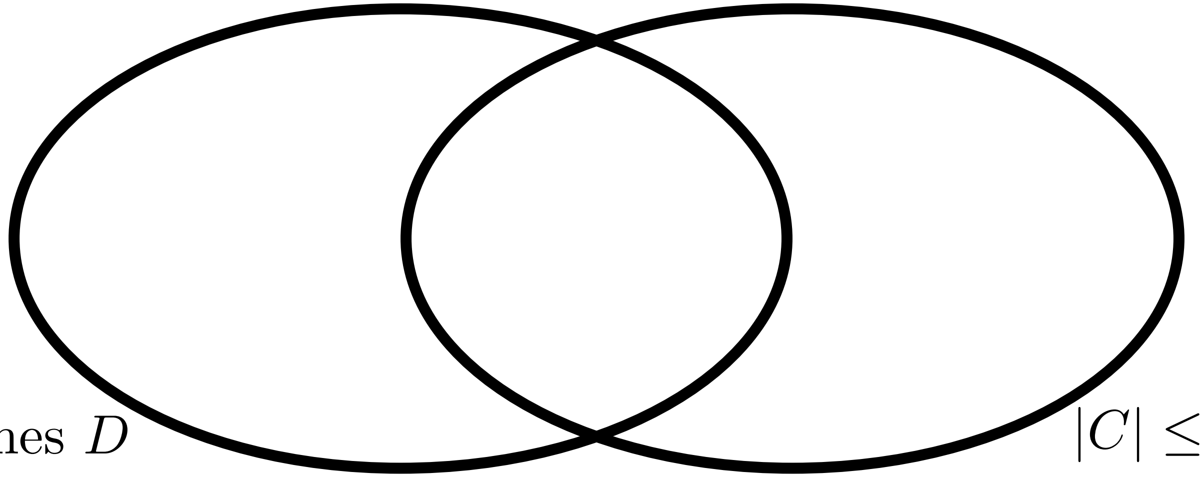
Natural relation of a property P: a largest relation for which P is private (a relation that contains the most of different links).

- The natural relation of a property P is unique and any relation that makes P private is included in this natural relation.
- Two hypotheses C and D are in natural relation for a property $f(H)\,\mathcal{R}\,k$ iff $f(C)\,\mathcal{R}\,f(D)$.

$$H \models e^+ \;\ldots\ldots\ldots\ldots\ldots\ldots \; C \models D$$

$$|H| \leq_{\mathbb{N}} 5 \;\ldots\ldots\ldots\ldots \; |C| \leq_{\mathbb{N}} |D|$$

Conjunctions of properties



Ideal operators

[van der Laag and Nienhuys-Cheng, 1994]

Definition 2 (ideality) *An operator is* ideal *if it is locally finite, proper and complete.*

Locally finite: $\mathcal{O}(H)$ is computable ;

Proper: $\mathcal{O}(H)$ does not contain any clause equivalent to H ;

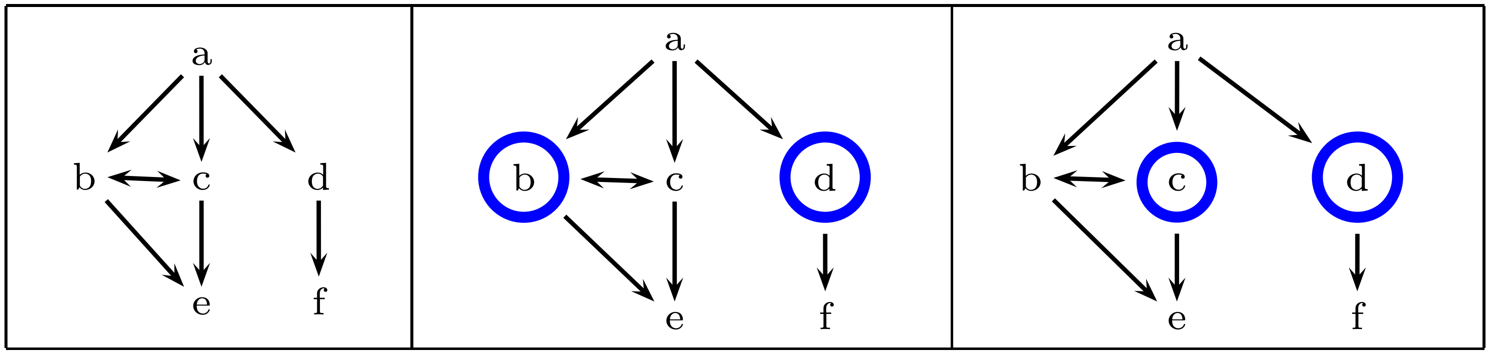
Complete: $\mathcal{O}^*(H)$ contains all clauses comparable to H.

Ideal operators do not exist for θ-subsumption or logical implication.

Covers

Definition 3 (cover) *C covers D iff $C > D$ and there exists no E such that $C > E > D$. C is an upward cover of D, D a downward cover of C.*

Definition 4 (cover set) *A downward (resp. upward) cover set of a clause C is a maximal set of uncomparable downward (resp. upward) covers of C.*

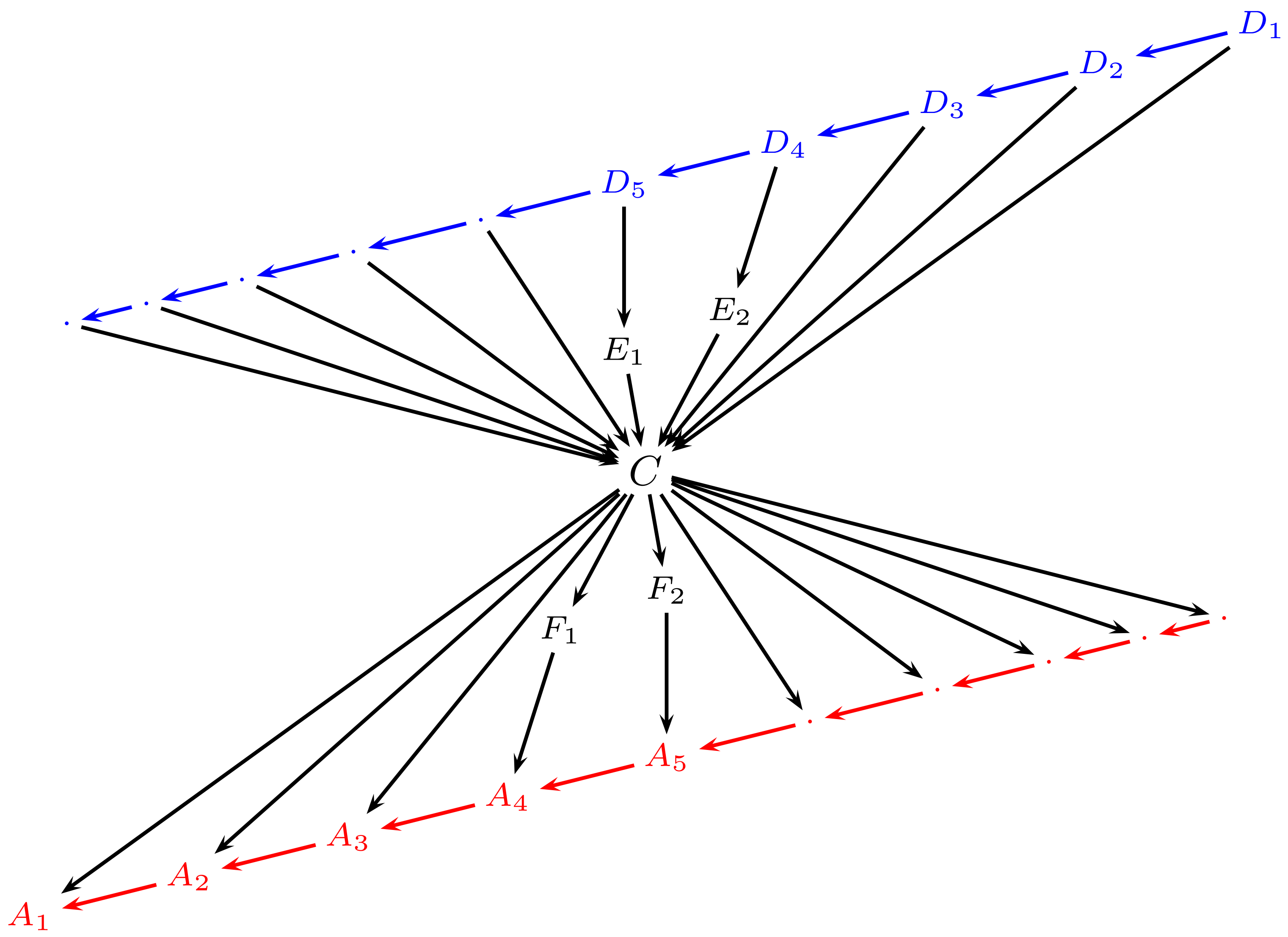


Non-existence

An ideal operator computes at least a cover set.

Then, there are two possible problems.

1. The cover set is not defined, there is an uncovered infinite chain, no complete and computable operator exists.
2. The cover set is infinite, a complete operator cannot be computable.


D_1
D_2
D_3
D_4
D_5
E_1
E_2
C
F_1
F_2
A_1
A_2
A_3
A_4
A_5

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

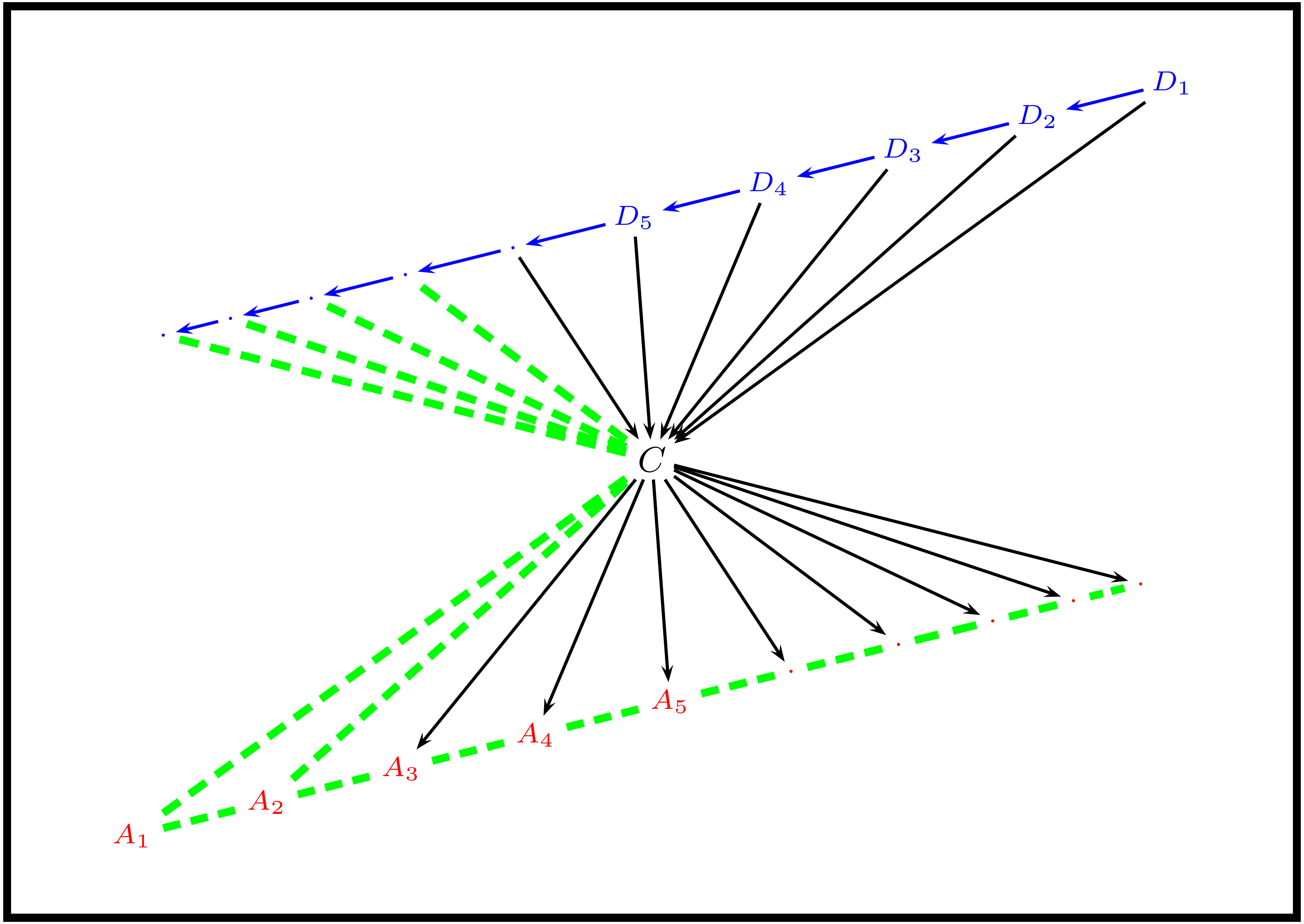
$$D_2 : q(X_1) \leftarrow p(X_1, X_2), p(X_2, X_1)$$

$$D_3 : q(X_1) \leftarrow \quad p(X_1, X_2), p(X_1, X_3), p(X_2, X_1), p(X_2, X_3), p(X_3, X_1), p(X_3, X_2)$$

$$D_n : q(X_1) \leftarrow \{p(X_i, X_j) | 1 \leq i, j \leq n, i \neq j\}$$

$$\cdot \qquad \cdots$$

$$C : q(X_1) \leftarrow p(X_1, X_1)$$


D_1
D_2
D_3
D_4
D_5
C
A_1
A_2
A_3
A_4
A_5

An ideal operator: $\rho^{\theta}_{||}(C)$

1. Add a literal with new variables to C.
2. Unify two variables X_1 and X_2 of C, such that $C >^{\theta} C\{X_1/X_2\}$. If the size of the result decreases, add literals with new variables.
3. Apply previous operations (1, 2) on clauses equivalent to C (θ-equivalent and same size).
4. Apply operation 1 on subsets of C which are equivalent to C, one new literal at least must use a predicate symbol which does not appear in C.

Conclusion

- Dynamic pruning with respect to langage bias based on the use of new quasi orders: *natural relations*.
- Existence of ideal operators for unrestricted search spaces ordered by natural relations.

Related works:
- [Shapiro, 1981]
- [Champesme et al., 1995, Esposito et al., 1996]

Perspectives: consider other families of operators for spaces ordered by natural relations (*optimal* [De Raedt and Bruynooghe, 1993]).

References

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

[De Raedt and Bruynooghe, 1993] De Raedt, L. and Bruynooghe, M. (1993). A theory of clausal discovery. pages 1058–1063. Morgan Kaufmann.

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

[Mitchell, 1982] Mitchell, T. M. (1982). Generalization as search. *Artificial Intelligence*, 18:203–226.

[Mitchell, 1991] Mitchell, T. M. (1991). The need for biases in learning generalizations. In *Readings in Machine Learning*. Morgan Kaufmann.

[Muggleton and Raedt, 1994] Muggleton, S. and Raedt, L. D. (1994). Inductive logic programming: Theory and methods. *Journal of Logic Programming*, 19:629–679.

[Shapiro, 1981] Shapiro, E. Y. (1981). Inductive inference of theories from facts. Technical Report 192, Yale University Department of Computer Science.

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