

**Intégration des biais de langage
à l'algorithme générer-et-tester**

Contributions à l'apprentissage disjonctif

FABIEN TORRE

Équipe Inférence et Apprentissage
Laboratoire de Recherche en Informatique
Université Paris-Sud, Orsay

LIP6, 8 juin 2000

PLAN DE L'EXPOSÉ

1. Apprentissage conjonctif en Programmation Logique Inductive

- (a) utilisation de biais
- (b) propriétés privées
- (c) relations naturelles
- (d) opérateurs idéaux pour ces relations naturelles

2. Apprentissage disjonctif en attribut-valeur

- (a) objectifs : optimiser la couverture des exemples positifs
- (b) algorithmes stochastiques du système GloBo
- (c) évaluation de la fiabilité

Conclusions et perspectives

PROGRAMMATION LOGIQUE INDUCTIVE

[Muggleton and Buntine, 1988, Muggleton, 1991]

- Exemples et hypothèses sont des formules du premier ordre.
- Espaces de recherche infinis.
- Implication logique : indécidable. [Schmidt-Schauss, 1988]

Définition de la θ -subsumption [Plotkin, 1970]

C θ -subsume D s'il existe une substitution θ telle que $C\theta \subseteq D$.

- Équivalente à l'implication logique pour des clauses non récursives. [Gottlob, 1987]
- θ -subsumption : NP-complet. [Kapur and Narendran, 1986]

BIAIS DE LANGAGE

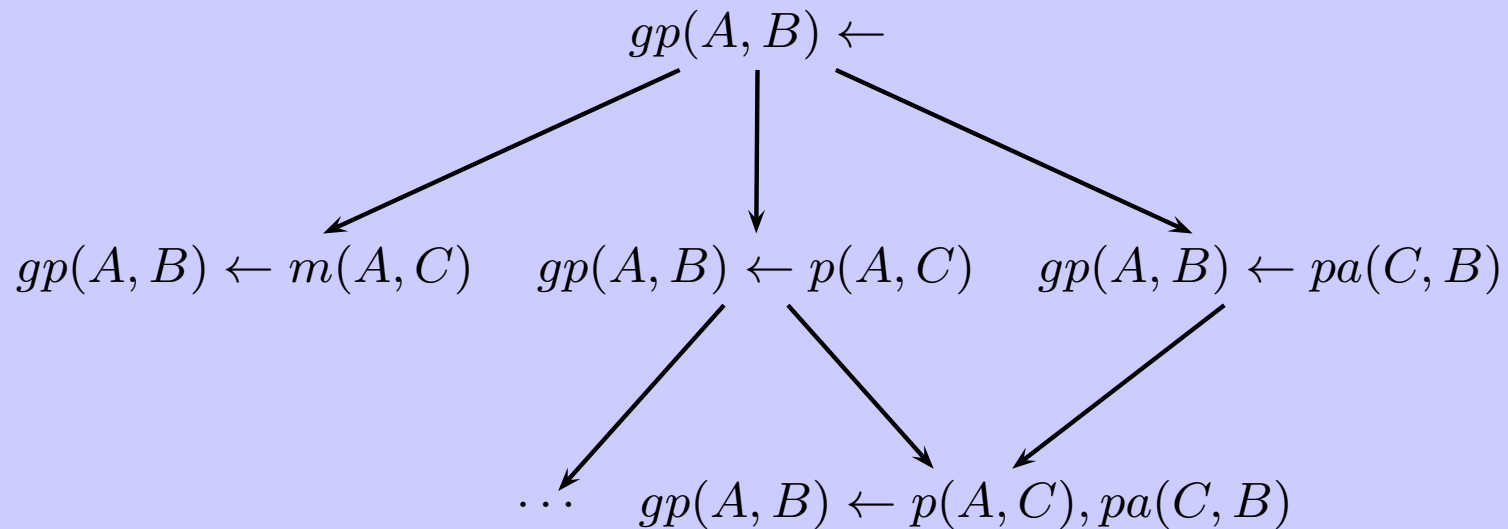
[Utgoff and Mitchell, 1982, Utgoff, 1986]

- Les biais sont *nécessaires* à l'apprentissage. [Mitchell, 1980]
- Qualité des résultats et efficacité.
- Biais de langage en PLI : contraintes sur la syntaxe des hypothèses. [Nédellec et al., 1996]
 - champ réduit, connexion,
 - borne sur la taille, le nombre de variables, la profondeur, etc.
- Ces biais ne rendent pas la recherche plus efficace : l'élagage dynamique est généralement impossible par rapport à ces contraintes.

ÉLAGAGE DYNAMIQUE

Opérateur d'ajout, P_1 : couverture de e^+ et P_2 : rejet de e^-

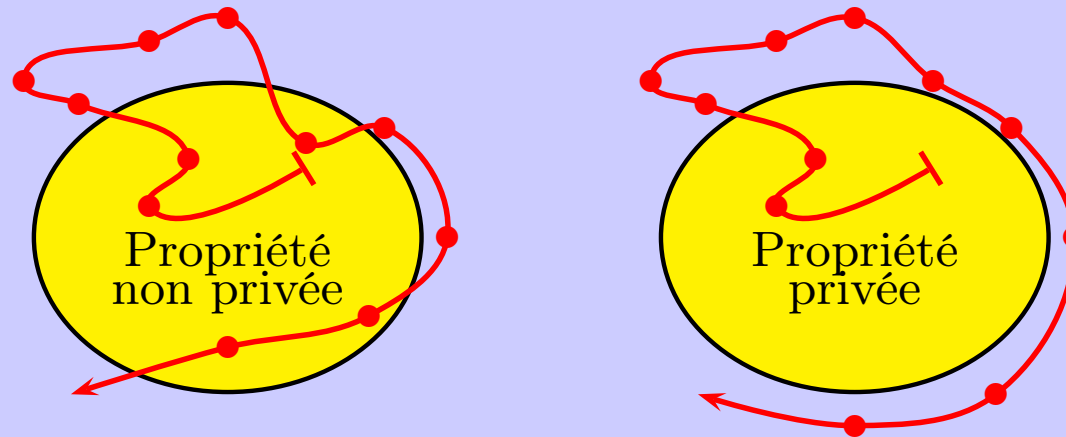
Respect d'une relation de généralité pour permettre un élagage par rapport à la couverture des exemples. [Mitchell, 1982]



P_3 : champ réduit

Objectif : élaguer dynamiquement par rapport aux biais de langage.

PROPRIÉTÉS PRIVÉES



Définition d'une propriété privée

Une propriété P est dite *privée* par rapport à une relation \mathcal{R} si et seulement si

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

\mathcal{R} = ajout, P = couverture de e^+ ou P = taille bornée.

Quelles sont les relations qui rendent une propriété privée ?

RELATIONS NATURELLES

Relation naturelle d'une propriété P : une relation, la plus large possible, pour laquelle P est privée.

- La relation naturelle d'une propriété est unique : toute relation qui la rend privée est incluse dans sa relation naturelle.
- Deux hypothèses C et D sont en relation naturelle pour une propriété $f(H) \mathcal{R} k$ si et seulement si $f(C) \mathcal{R} f(D)$.

$$H \models e^+ \dots\dots\dots C \models D$$

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

- Recherche d'opérateurs dans les relations naturelles.

OPÉRATEURS IDÉAUX

Définition de l'idéalité [van der Laag and Nienhuys-Cheng, 1994]

Un opérateur \mathcal{O} est *idéal* pour une relation de généralité \succeq s'il est à la fois localement fini, strict et complet.

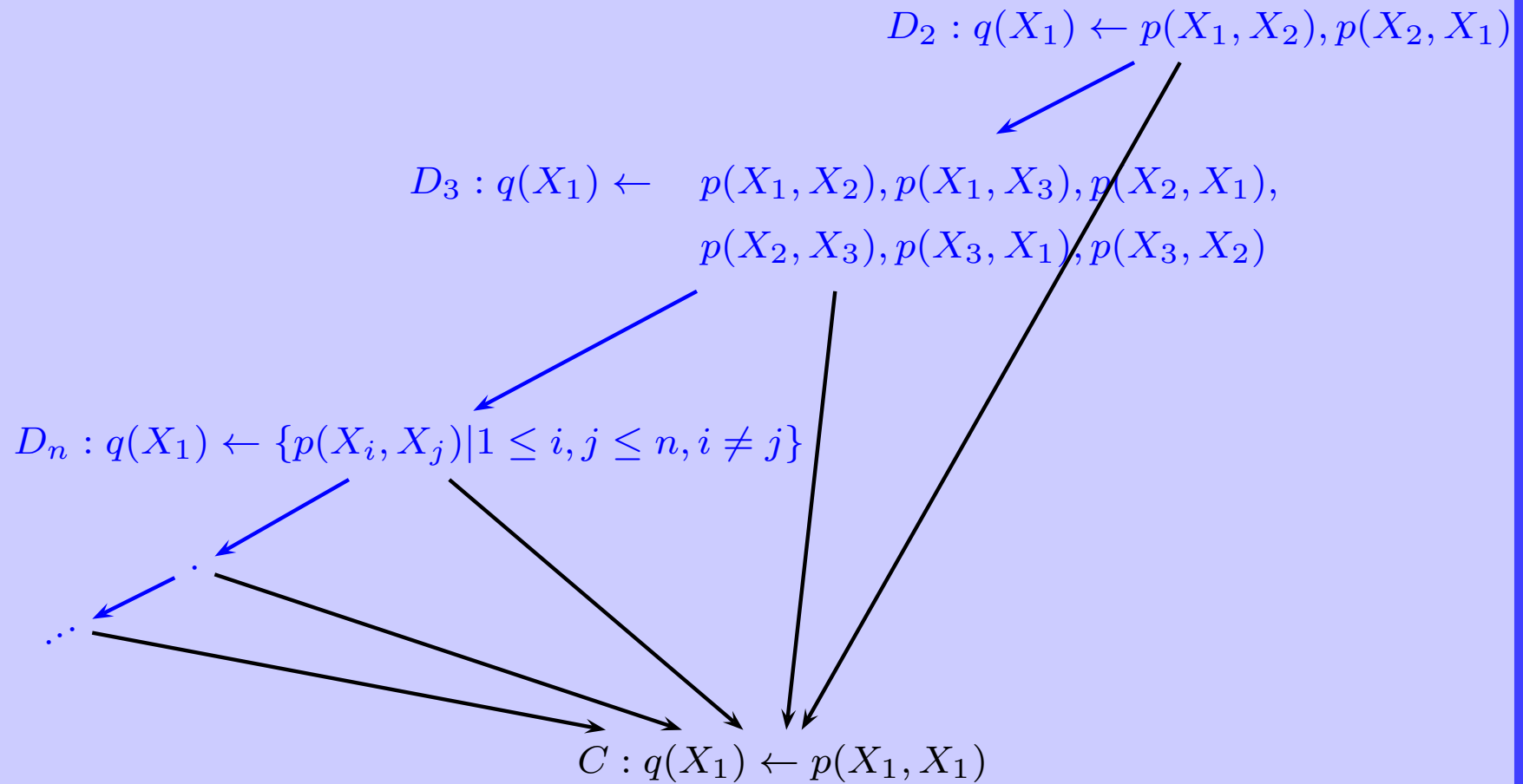
Localement fini : $\mathcal{O}(H)$ est calculable ;

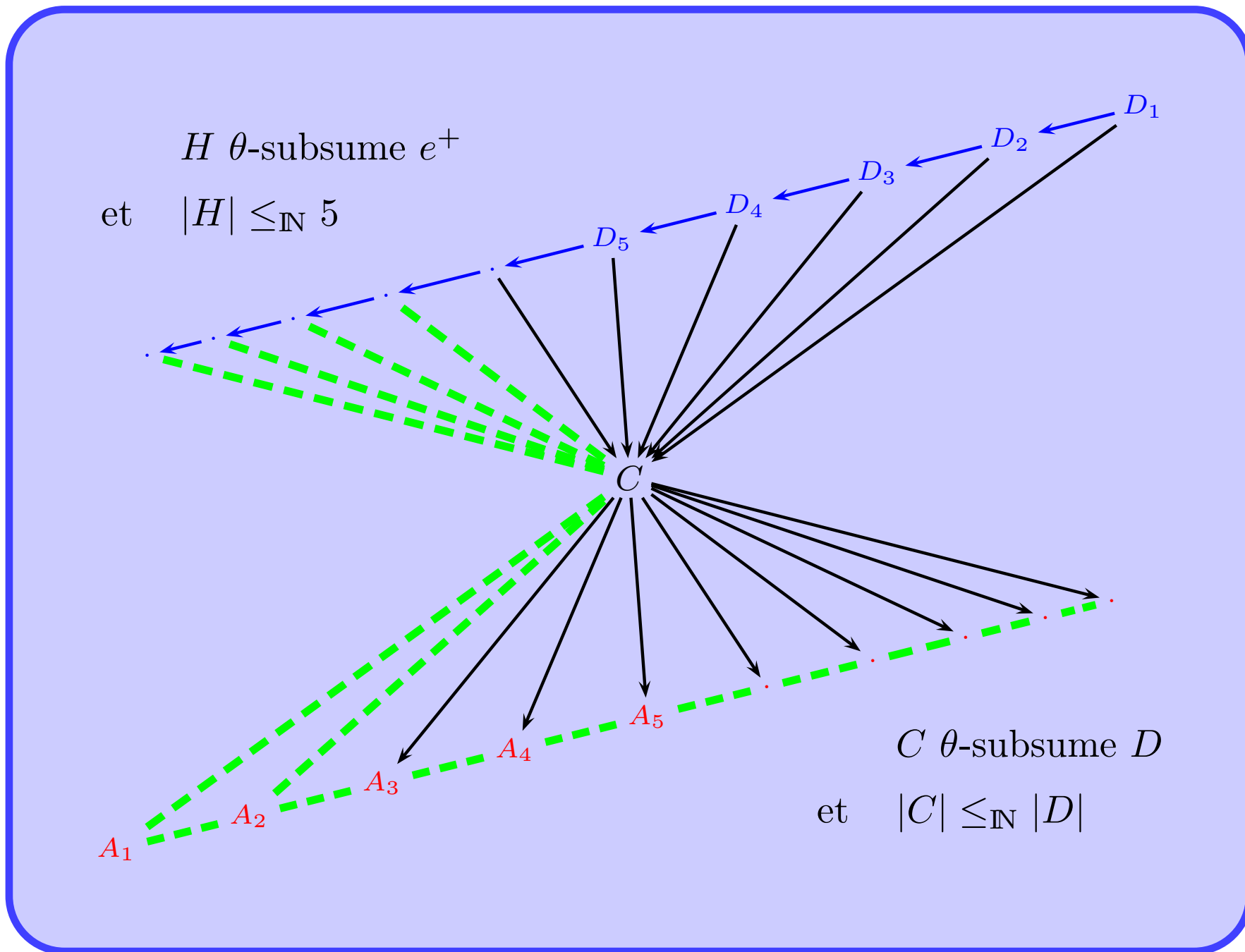
Strict : $\mathcal{O}(H)$ ne contient aucune clause \succeq -équivalente à H ;

Complet : si H et H' sont comparables pour \succeq , $H \in \mathcal{O}^*(H')$ ou $H' \in \mathcal{O}^*(H)$.

Les opérateurs idéaux n'existent pas sous implication logique, ni sous θ -subsumption.

Un opérateur idéal pour \succeq doit calculer les successeurs immédiats de chaque clause sous \succeq .





DÉCOUVERTE D'UN OPÉRATEUR IDÉAL

- Opérateur idéal pour $\succeq_{||}^{\theta}$
 - (2) Unifier deux variables X_1 et X_2 de C , telles que $C >^{\theta} C\{X_1/X_2\}$. Si la taille de la clause-résultat a diminué, ajouter un littéral ne contenant que des nouvelles variables.
- Résultats de non existence pour :
 - une borne sur la profondeur des hypothèses,
 - la limitation du nombre de variables utilisées.
- Piste : le biais MaxOcc (bornes sur l'occurrence de chaque symbole de prédicat dans les hypothèses).

BIAIS ET GÉNÉRER-ET-TESTER : BILAN

- Propriétés privées.
- Élagage dynamique par rapport aux biais de langage basé sur l'utilisation de nouveaux pré-ordres : les *relations naturelles*.
[Shapiro, 1981]
- Conditions nécessaires sur les propriétés pour permettre l'existence d'opérateurs idéaux.
- Découverte d'un opérateur idéal pour $\perp_{\parallel}^{\theta}$.
[Champesme et al., 1995, Esposito et al., 1996]
- Cadre parfait : optimalité. [De Raedt and Bruynooghe, 1993]

APPRENTISSAGE DISJONCTIF : PROBLÈME DU MORPION

- Caractériser les fins de jeu au morpion gagnantes pour les croix.
- La solution est la disjonction des huit manières de faire des lignes de croix :

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| × | ? | ? | ? | × | ? | ? | ? | × | × | ? | ? |
| × | ? | ? | ? | × | ? | ? | ? | × | ? | × | ? |
| × | ? | ? | ? | × | ? | ? | ? | × | ? | ? | × |
| ? | ? | × | × | × | × | ? | ? | ? | ? | ? | ? |
| ? | × | ? | ? | ? | ? | × | × | × | ? | ? | ? |
| × | ? | ? | ? | ? | ? | ? | ? | ? | × | × | × |

EXPÉRIMENTATIONS SUR LE PROBLÈME DU MORPION

- On dispose de 958 exemples (toutes les fins de jeu possibles).
- On utilise 70 % des exemples pour apprendre, les 30 % restants pour le test.

| | |
|---------|--------|
| default | 65.3 % |
| NewID | 84.0 % |
| CN2 | 98.1 % |
| MBRtalk | 88.4 % |

| | |
|--------|--------|
| IB1 | 98.1 % |
| IB3 | 82.0 % |
| IB3-CI | 99.1 % |

[Aha, 1991]

MOINDRE GÉNÉRALISÉ

| | | |
|---|---|---|
| × | × | × |
| × | | ○ |
| | ○ | ○ |

| | | |
|---|---|---|
| × | × | × |
| | ○ | |
| ○ | ○ | × |

moindre généralisé →

| | | |
|---|---|---|
| × | × | × |
| ? | ? | ? |
| ? | ○ | ? |

| | | |
|---|---|---|
| × | | ○ |
| × | ○ | × |
| × | | ○ |

moindre généralisé ↓

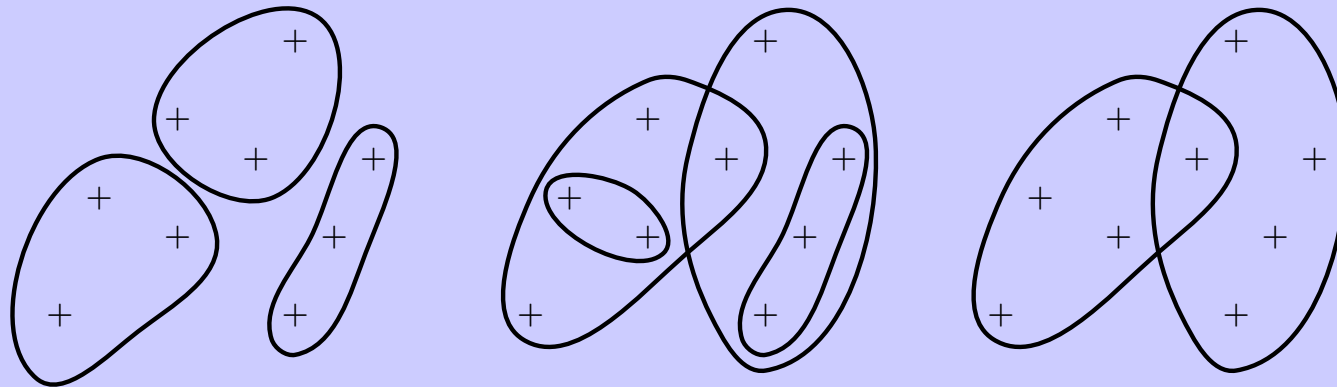
| | | |
|---|---|---|
| × | ? | ? |
| × | ? | ? |
| ? | ? | ○ |

subsume →

| | | |
|---|---|---|
| × | | × |
| × | × | |
| ○ | ○ | ○ |

APPRENTISSAGE SUPERVISÉ DISJONCTIF

Différentes couvertures des exemples positifs :



Difficile pour les approches par *couverture* et pour les méthodes de type *diviser pour régner*. [Boström, 1995]

Solution idéale : paquets maximale­ment corrects,
nombre minimal de paquets.

Problème NP-difficile.

ALGORITHME DE GLoBo : SÉLECTION STOCHASTIQUE

| | | | |
|--------|---|---|------------------------------------|
| graine | autres positifs | → | paquet maximalement correct |
| p_1 | p_5 p_8 p_2 $p_{14} \dots$ | → | $\{p_1, p_5, p_8, p_{14}, \dots\}$ |
| p_2 | p_4 p_3 p_1 $p_{12} \dots$ | → | $\{p_2, p_3, p_{12}, \dots\}$ |
| p_3 | p_7 p_4 p_{18} p_{12} \dots | → | $\{p_3, p_7, \dots\}$ |
| p_4 | p_{13} p_3 p_9 p_{11} \dots | → | $\{p_4, p_{13}, p_3, \dots\}$ |
| p_5 | p_8 p_2 p_1 $p_{14} \dots$ | → | $\{p_5, p_8, p_1, p_{14}, \dots\}$ |
| ⋮ | ⋯ | → | ⋯ |

PAQUETS VERROUILLÉS ET PAQUETS CONDAMNÉS

| | | |
|---|---|---|
| × | | |
| × | ○ | |
| × | | ○ |

| | | |
|---|---|---|
| × | ○ | ○ |
| × | × | ○ |
| × | | |

| | | |
|---|---|--|
| × | | |
| × | ○ | |
| × | ○ | |

| | | |
|---|---|---|
| × | ? | ? |
| × | ? | ? |
| × | ? | ? |

| | | |
|---|---|---|
| × | | |
| × | ○ | ○ |
| × | | |

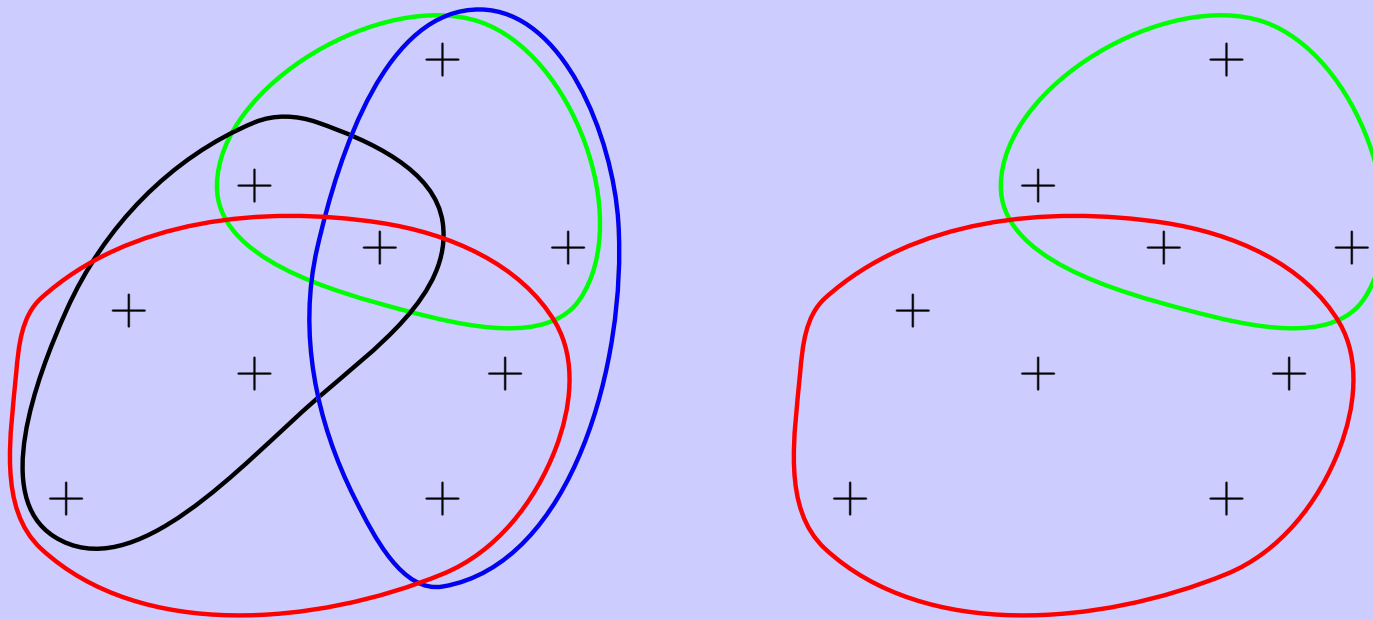
| | | |
|---|---|---|
| × | × | × |
| | ○ | |
| | | ○ |

| | | |
|---|---|---|
| × | ? | ? |
| ? | ○ | ? |
| ? | | ? |

ALGORITHME DE GLOBo : COUVERTURE MINIMALE

- Problème NP-difficile.
- Heuristique polynomiale : choisir le paquet qui couvre le plus d'éléments non encore couverts.

[Paschos, 1997]



COMPLEXITÉ DE GLOBo

- Construction d'un paquet maximale correct : $|A^+| - 1$ calculs de moindre généralisés et un test de couverture vis-à-vis de $|A^-|$ (au pire, $|A^-|$ tests de subsomption).
- Couverture minimale : $|A^+|^2$. [Paschos, 1997]
- GloBo : $\Theta(|A^+|^2)$ en calculs de moindre généralisé, et $\Theta((|A^+| + |A^-|)^3)$ en tests de subsomption.
- Pré-requis : unicité du moindre généralisé et la propriété de correction vérifie :

$$\text{Correct}(A \cup B) \Rightarrow \text{Correct}(A) \wedge \text{Correct}(B)$$

ESTIMATION DE LA PROBABILITÉ DE RÉUSSITE DE L'APPRENTISSAGE

- n nombre de sous-concepts à apprendre,
 s nombre moyen d'exemples positifs dans un sous-concept,
 α la probabilité d'être favorable à la graine,
 b la taille minimale d'un paquet verrouillé (hors graine).

| événement | probabilité |
|--------------------------------|----------------------------|
| verrouiller le paquet | α^b |
| échouer pour s exemples | $(1 - \alpha^b)^s$ |
| réussir pour n sous-concepts | $[1 - (1 - \alpha^b)^s]^n$ |

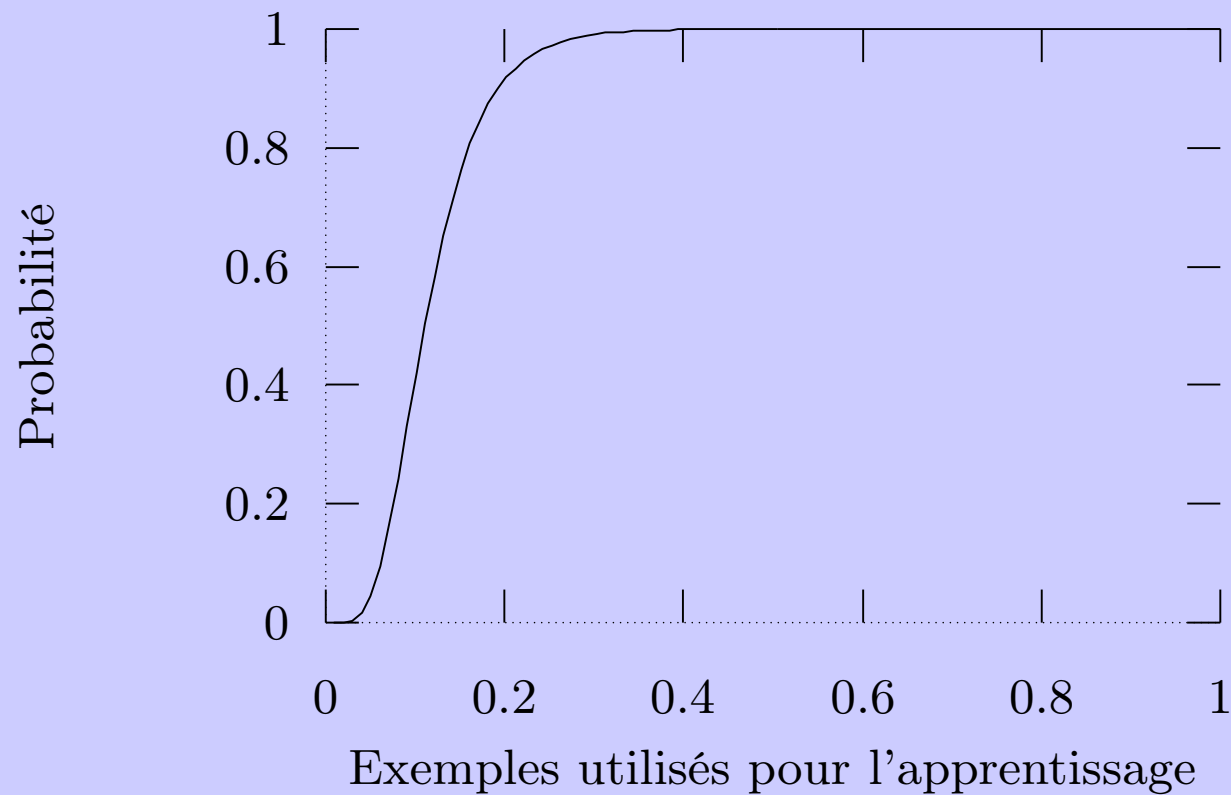
INTÉRÊTS DE LA PROBABILITÉ DE RÉUSSITE

$$\left[1 - (1 - \alpha^b)^s\right]^n$$

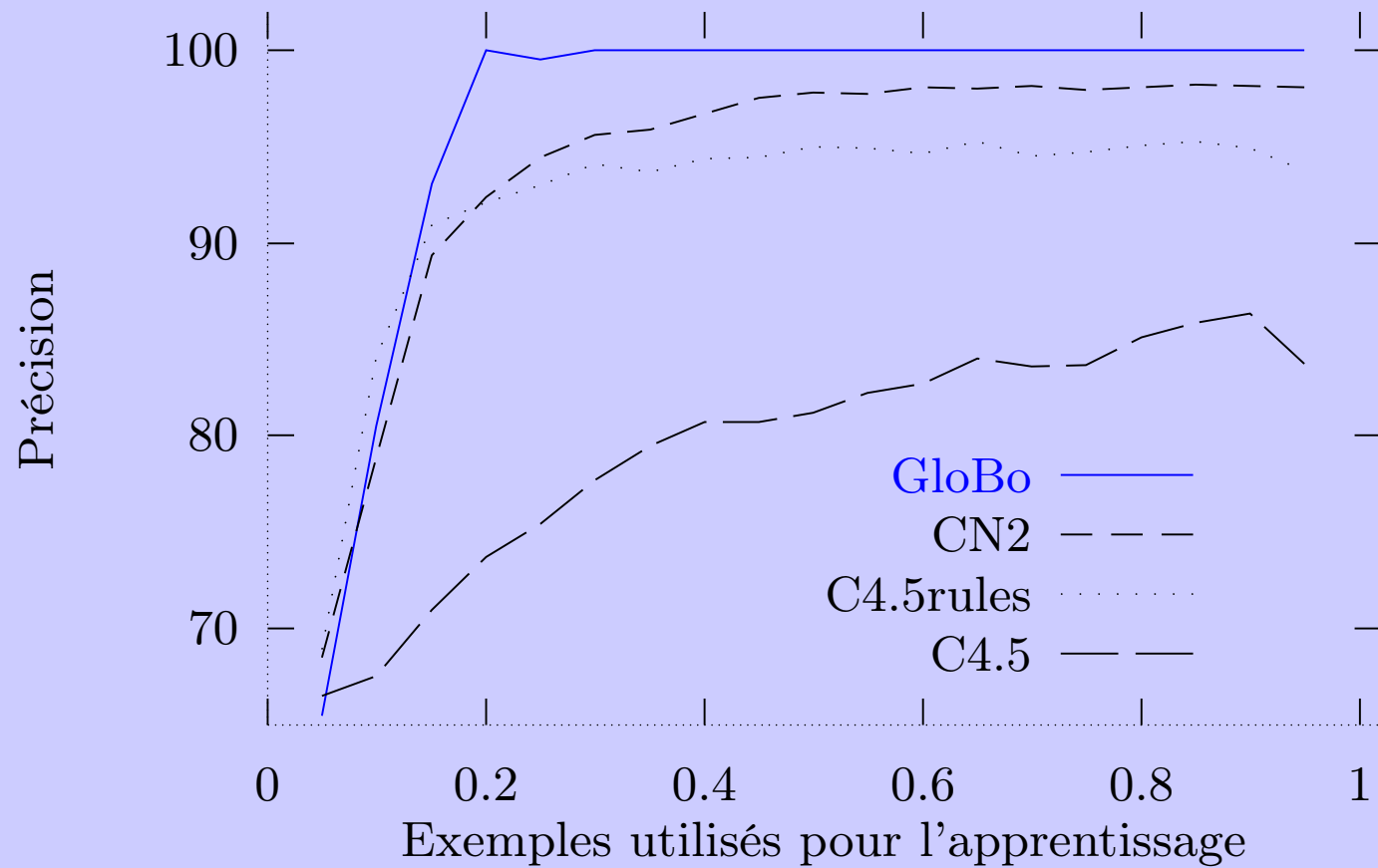
- Fiabilité a posteriori de la solution découverte.
- Notion de représentativité des exemples disponibles par rapport au concept à découvrir.
- No Free Lunch Theorem : [Wolpert and Macready, 1995]
[Schaffer, 1994]
caractérisation des cas d'échec.

PROBABILITÉ DE SUCCÈS DE GLOBO POUR LE MORPION

$n = 8$, $\alpha = \frac{1}{2}$, $b = 2$ et s vaut $\frac{1}{8}$ du nombre d'exemples disponibles.



RÉSULTATS DE GLOBo POUR LE MORPION



AUTRES EXPÉRIMENTATIONS

| Problème | Taille | Précision |
|--------------------------|--------|-----------|
| Mushroom (avec négation) | 3 | 100.0 % |
| Breast Cancer | 4 | 93.0 % |
| Pima Indians Diabetes | 24 | 68.8 % |
| Echocardiogram | 3 | 88.2 % |

PTE challenge : molécules cancérigènes.

[Srinivasan et al., 1997]

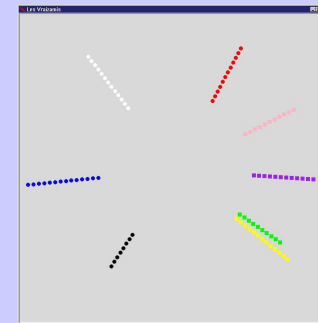
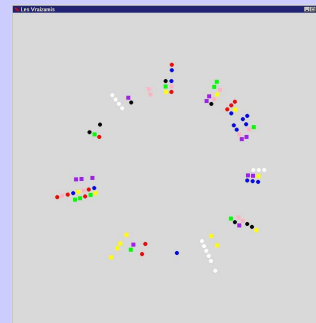
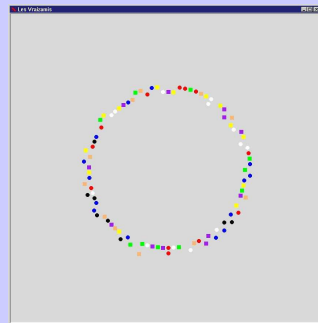
Résultats :

[Srinivasan et al., 1999]

- GloBo est premier des systèmes produisant une théorie compréhensible.
- GloBo est jugé optimal pour toute application pratique.

APPRENTISSAGE DISJONCTIF : BILAN

- Correction maximale et minimalité de la solution.
- Stochastique : pseudo-parallélisme et complexité raisonnable.
- Estimation du risque d'erreur de la procédure stochastique.
- Version générique : instanciations linéaire en temps, dédiée aux données bruitées, pour le non supervisé.
- Variation : les Vraizamis.



CONCLUSIONS

1. Apprentissage conjonctif en Programmation Logique Inductive

- intégration des biais de langage dans la relation de généralité,
- élagage intensif de l'espace de recherche.

2. Apprentissage disjonctif en attribut-valeur

- couverture minimale par des paquets maximale corrects,
- construction stochastique des paquets,
- approximation polynomiale d'une couverture minimale.

PERSPECTIVES

- Poursuite des applications à des problèmes réels et bruités :
 - application d'opérateurs de PLI à la fouille de textes, contes de Grimm (YVES KODRATOFF, équipe I&A) ;
 - projet INDANA, prévention des maladies cardio-vasculaires (ISABELLE COLOMBET, hôpital Broussais).
- Extension de GloBo et des Vraizamis :
 - à l'apprentissage non-supervisé (classification de textes) ;
 - à la Programmation Logique Inductive.

References

- [Aha, 1991] Aha, D. W. (1991). Incremental constructive induction: An instance-based approach. In *Proceedings of the Eighth International Workshop on Machine Learning*, pages 117–121. Morgan Kaufmann.
- [Boström, 1995] Boström, H. (1995). Covering vs. divide-and-conquer for top-down induction of logic programs. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 1194–1200.
- [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. In Bajcsy, R., editor, *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, 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.

[Gottlob, 1987] Gottlob, G. (1987). Subsumption and implication. *Information Processing Letters*, 24(2):109–111.

[Kapur and Narendran, 1986] Kapur, D. and Narendran, P. (1986). Np-completeness of the set unification and matching problems. In *Proceedings of 8th Conference on Automated Deduction*, volume 230, pages 489–495. Springer-Verlag.

- [Mitchell, 1980] Mitchell, T. M. (1980). The need for biases in learning generalizations. In *Readings in Machine Learning*, pages 184–191. Morgan Kaufmann. Published in 1991.
- [Mitchell, 1982] Mitchell, T. M. (1982). Generalization as search. *Artificial Intelligence*, 18:203–226.
- [Muggleton, 1991] Muggleton, S. (1991). Inductive logic programming. *New Generation Computing Journal*, 8(4):295–317.
- [Muggleton and Buntine, 1988] Muggleton, S. H. and Buntine, W. (1988). Machine invention of first-order predicates by inverting resolution. In *Proceedings 5th International Conference on Machine Learning*, pages 339–352, San Mateo, CA. Morgan Kaufmann.
- [Nédellec et al., 1996] Nédellec, C., Rouveirol, C., Adé, H., Bergadano, F., and Tausend, B. (1996). Declarative bias in ILP.

In De Raedt, L., editor, *Advances in Inductive Logic Programming*, pages 82–103. IOS Press.

[Paschos, 1997] Paschos, V. T. (1997). A survey of approximately optimal solutions to some covering and packing problems. *ACM Computing Surveys*, 29(2):171–209.

[Plotkin, 1970] Plotkin, G. (1970). A note on inductive generalization. In Meltzer, B. and Mitchie, D., editors, *Machine Intelligence*, volume 5, pages 153–165. Edinburgh University Press.

[Schaffer, 1994] Schaffer, C. (1994). A conservation law for generalization performance. In Cohen, W. W. and Hirsh, H., editors, *Proceedings 11th International Conference on Machine Learning*, pages 259–265. Morgan Kaufmann.

[Schmidt-Schauss, 1988] Schmidt-Schauss, M. (1988). Implication of clauses is undecidable. *TCS: Theoretical Computer Science*,

59(3):287–296.

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

[Srinivasan et al., 1999] Srinivasan, A., King, R., and Bristol, D. (1999). An assessment of submissions made to the predictive toxicology evaluation challenge. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, pages 270–276. Morgan Kaufmann.

[Srinivasan et al., 1997] Srinivasan, A., King, R. D., Muggleton, S. H., and Sternberg, M. J. E. (1997). The predictive toxicology evaluation challenge. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence*, pages 4–9. Morgan Kaufmann.

[Utgoff, 1986] Utgoff, P. E. (1986). Shift of bias for inductive

concept learning. In Michalski, R. S., Carbonell, J. G., and Mitchell, T. M., editors, *Machine Learning: An Artificial Intelligence Approach*, volume II, pages 107–148. Morgan Kaufmann.

[Utgoff and Mitchell, 1982] Utgoff, P. E. and Mitchell, T. M. (1982). Acquisition of appropriate bias for inductive concept learning. In Waltz, D., editor, *Proceedings of the National Conference on Artificial Intelligence*, pages 414–417. AAAI Press.

[van der Laag and Nienhuys-Cheng, 1994] van der Laag, P. R. J. 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.

[Wolpert and Macready, 1995] Wolpert, D. H. and Macready,

W. G. (1995). No free lunch theorems for search. Technical Report SFI-TR-95-02-010, The Santa Fe Institute.