

Transition de phase en induction supervisée

Antoine Cornuéjols

Laboratoire de Recherche en Informatique
Université de Paris-Sud, Orsay

Séminaire INPG 16 février 2006

Plan

- 1 Un point de vue sur l'apprentissage
 - L'induction supervisée
 - Analyse des conditions de l'induction
- 2 Transition de phase en induction
 - Gain d'information et transition de phase
 - Transition de phase en induction de programmes logiques
 - Transition de phase en inférence grammaticale
- 3 Analyse et bilan
 - Analyse et explication ?
 - Remèdes ?
 - Bilan

Les pionniers {1970-1985}

- Des *simulations cognitives*

- E.g. ARCH [Winston70], AM [Lenat,78], ACT* [Anderson, 80], Meta-Dendral [Mitchell, 78], ...

- Apprentissage de *connaissances «complexes»*

- Règles, réseaux sémantiques, procédures, ...

- Concepts liés aux *sciences cognitives*

- Connaissances procédurales vs. déclaratives
 - Mémoire épisodique vs. mémoire sémantique
 - Mécanismes de raisonnement : généralisation, déduction, abduction, ...
 - Modification incrémentale : ajouts, ajustements locaux, restructurations

- Des processus se déployant dans le temps
- Des connaissances complexes

Les pionniers {1970-1985}

- Des ***simulations cognitives***

- E.g. ARCH [Winston70], AM [Lenat,78],
ACT* [Anderson, 80], Meta-Dendral [Mitchell, 78], ...

- Apprentissage de ***connaissances «complexes»***

- Règles, réseaux sémantiques, procédures, ...

- Concepts liés aux ***sciences cognitives***

- Connaissances **procédurales** vs. **déclaratives**
- Mémoire **épisodique** vs. mémoire **sémantique**
- Mécanismes de raisonnement : **généralisation, déduction, abduction, ...**
- Modification incrémentale : **ajouts, ajustements locaux, restructurations**

- Des processus se déployant dans le temps
- Des connaissances complexes

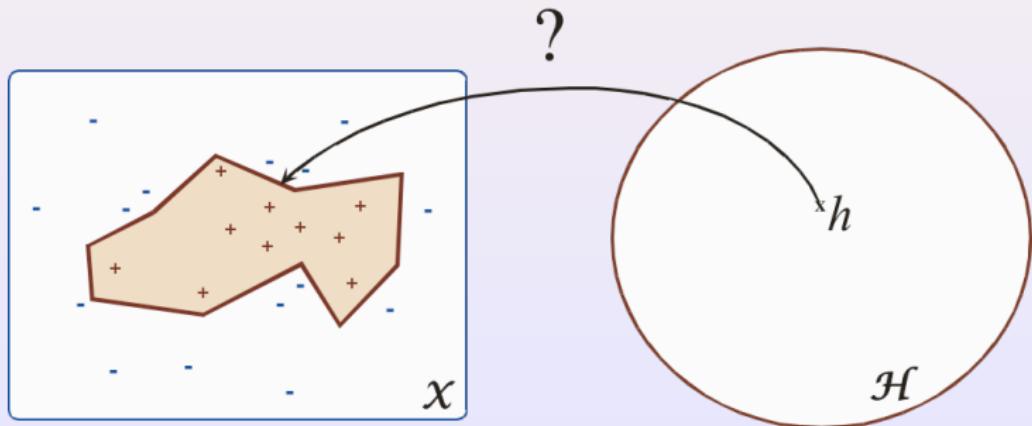


L'essence du problème

Définition [Mitchell, 82]

Apprentissage = Recherche dans un espace d'hypothèses

- Sous la contrainte des exemples d'apprentissage



Le paradigme

Définition (*Apprentissage*)

Apprentissage = Problème inverse mal posé

- À partir d'**observations**, trouver la loi ***f*** à laquelle obéissent ces observations

Le paradigme

Définition (*Apprentissage*)

Apprentissage = Problème inverse mal posé

- À partir d'**observations**, trouver la loi f à laquelle obéissent ces observations

Hypothèses

- Les observations sont des réalisations (i.i.d.) d'une variable aléatoire de loi f
- On cherche un estimateur \hat{h} aussi **proche** que possible de la loi f

Le paradigme

Apprentissage = Problème inverse mal posé

- ... chercher \hat{h} aussi proche que possible de la loi f

Proximité : Espérance de risque

$$R(\hat{h}) = \mathbb{E}_{D_{\mathcal{X} \times \mathcal{Y}}}[\hat{h}] = \int_{\mathcal{X} \times \mathcal{Y}} \underbrace{\ell(\hat{h}(\mathbf{x}), f(\mathbf{x}))}_{\text{coût pour une observation}} d\mathbf{x}d\mathbf{y}$$

Le paradigme

Apprentissage = Problème inverse mal posé

- ... chercher \hat{h} aussi proche que possible de la loi f

Proximité : Espérance de risque

$$R(\textcolor{green}{h}) = \mathbb{E}_{D_{\mathcal{X} \times \mathcal{Y}}}[\textcolor{green}{h}] = \int_{\mathcal{X} \times \mathcal{Y}} \underbrace{\ell(h(\mathbf{x}), f(\mathbf{x}))}_{\text{coût pour une observation}} d\mathbf{x}d\mathbf{y}$$

MRE

Choisir l'hypothèse \hat{h} telle que $\hat{h} = \text{ArgMin}_{h \in \mathcal{H}} [R_{\text{Emp}}(h)]$

$$R_{\text{Emp}}(\textcolor{green}{h}) = \frac{1}{m} \sum_{(\mathbf{x}_i, \textcolor{red}{u}_i) \in \mathcal{S}} \ell(\textcolor{green}{h}(\mathbf{x}_i), \textcolor{red}{u}_i)$$

La théorie statistique de l'apprentissage

Consistance du MRE

Conditions sous lesquelles le critère de MRE est correct ?

La théorie statistique de l'apprentissage

Consistance du MRE

Conditions sous lesquelles le critère de MRE est correct ?

→ *Diversité* de l'espace des hypothèses \mathcal{H} limitée

La théorie statistique de l'apprentissage

Consistance du MRE

Conditions sous lesquelles le critère de MRE est correct ?

→ *Diversité* de l'espace des hypothèses \mathcal{H} limitée

Qualité de l'estimation

$$|R(h) - R_{\text{Emp}}(h)| \leq_P \text{fct}(\text{diversité}_{\mathcal{H}}, m)$$

La théorie statistique de l'apprentissage : actualités

Théorie

- Estimations plus fines de la diversité (capacité)
- **Apprentissage actif** : modification de la distribution en apprentissage

La théorie statistique de l'apprentissage : actualités

Théorie

- Estimations plus fines de la diversité (capacité)
- Apprentissage actif : modification de la distribution en apprentissage

... de nouvelles méthodes

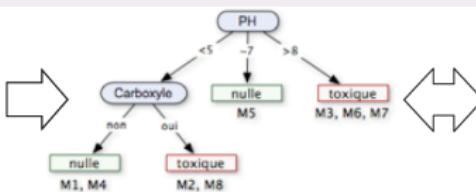
- Minimisation du ϕ -risque empirique
 - Méthodes de votes (**bagging**, **boosting**, ...)
 - Méthodes à noyaux (**SVM**, ...)

Le paradigme est adapté à ...

... l'analyse de données

BD \Rightarrow régularités / prédictions

	#Cycles	Masse	PH	Carboxyle	Activité
M1	1	faible	<5	non	nulle
M2	2	moyen	<5	oui	toxique
M3	0	moyen	>8	oui	toxique
M4	0	moyen	<5	non	nulle
M5	1	lourd	~7	non	nulle
M6	2	lourd	>8	non	toxique
M7	1	lourd	>8	non	toxique
M8	0	faible	<5	oui	toxique



- R1: Si pH > 8 alors toxique
 R2 : Si pH < 5 & Carboxyle = oui alors toxique

[CM02]

A. Cornuéjols and L. Miclet.

Apprentissage Artificiel. Concepts et Méthodes.
Eyrolles, 2002.



Le paradigme est adapté si ...

... peu de connaissances *a priori*

Seul critère : fidélité aux données

① Prise en compte de la structure de \mathcal{H} très pauvre

- Relations de généralité
- Niveaux d'abstraction
- ...

② Pas d'articulation à ce qui est déjà connu

- Incrémentalité / Révision de théorie / Transfert
- Critère de compréhensibilité

... monde supposé statique

Cadre i.i.d.

- Centralité du théorème central limite (et variantes)



Plan

- 1 Un point de vue sur l'apprentissage
 - L'induction supervisée
 - Analyse des conditions de l'induction
- 2 Transition de phase en induction
 - Gain d'information et transition de phase
 - Transition de phase en induction de programmes logiques
 - Transition de phase en inférence grammaticale
- 3 Analyse et bilan
 - Analyse et explication ?
 - Remèdes ?
 - Bilan

Transition de phase en induction

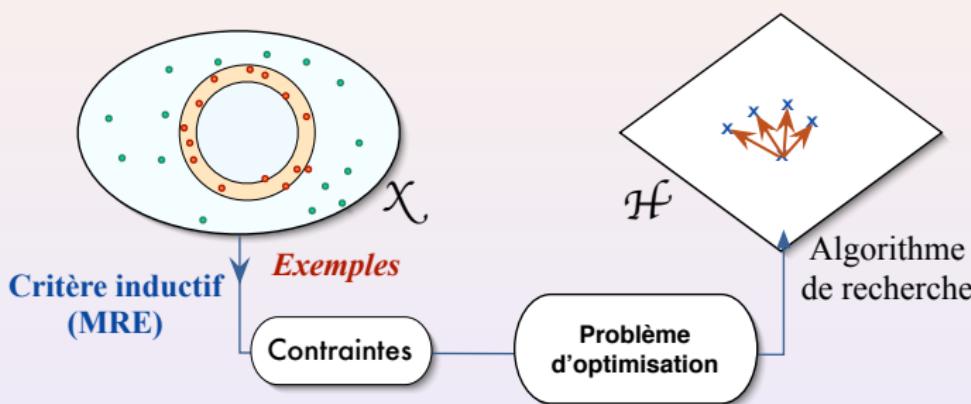
Collaborations

- Nicolas Baskiotis (*Doctorant*)
- Jérôme Maloberti (*Thèse*)
- Nicolas Pernot (*stage DEA*)
- Sandra Pinto (*stage DEA*)
- Raymond Ros (*Doctorant*)
- Michèle Sebag (DR, CNRS)

- Mario Botta
- Attilio Giordana
- Lorenza Saitta

Sous quelles conditions l'induction est-elle possible ?

Des conditions sur le gain d'information

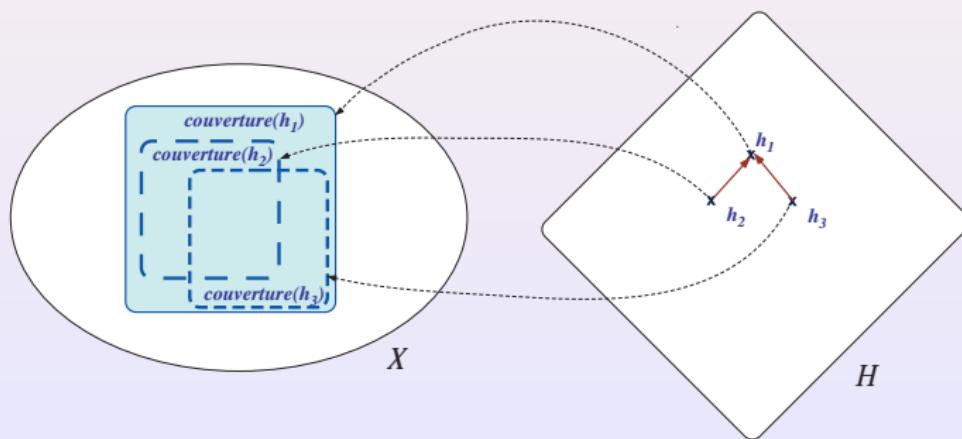


Les entrées doivent **se traduire en «différences» sur les hypothèses**

Des conditions sur le gain d'information

Gradient et taux de couverture

- Le gradient est lié aux variations du taux de couverture



Des conditions sur le gain d'information

Variations du taux de couverture

La mesure du taux de couverture apporte-t-elle de l'information ?

Des conditions sur le gain d'information

Variations du taux de couverture

La mesure du taux de couverture apporte-t-elle de l'information ?

Definition (Taux de couverture)

$$\tau(h) = P_{\mathcal{D}_X}(h)$$

Des conditions sur le gain d'information

Variations du taux de couverture

La mesure du taux de couverture apporte-t-elle de l'information ?

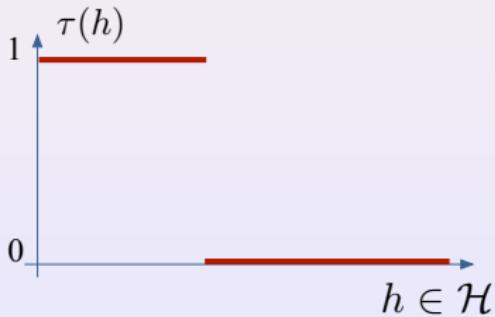
Definition (Taux de couverture)

$$\tau(h) = P_{\mathcal{D}_{\mathcal{X}}}(h)$$

*Étude des variations de $\tau(h)$
en fonction des variations de h (partie de \mathcal{X})*

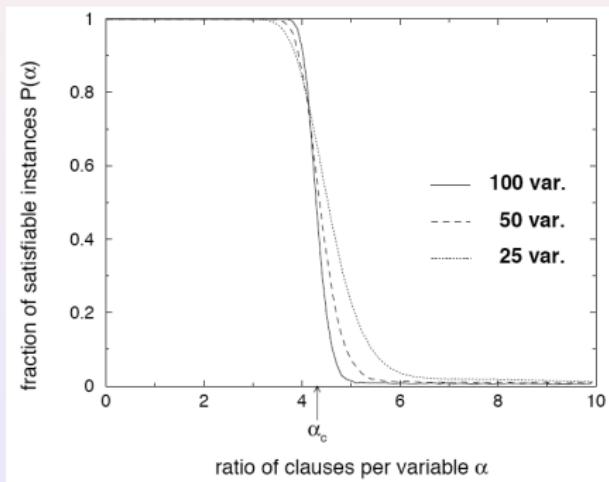
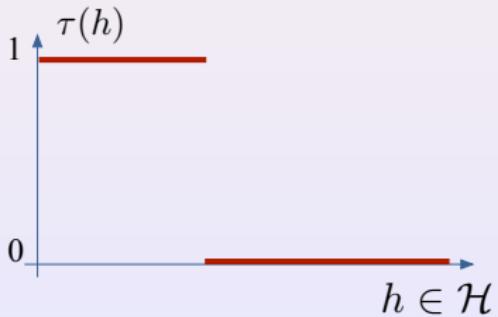
Des conditions sur le gain d'information

Un cas limite ... mais ...



Des conditions sur le gain d'information

Un cas limite ... mais ...



L'induction de connaissances structurelles

Programmation Logique Inductive (ILP)



active (d1)

lumo (d1, -1.246)

loop (d1, 4.23)

benzene (d1,[d1_6,d1_1,d1_2,d1_3,d1_4,d1_5])

atm (d1,d1_1,c,22,-0.117)

atm (d1,d1_2,c,22,-0.117)

atm (d1,d1_3,c,22,-0.117)

atm (d1,d1_4,c,195,-0.087)

atm (d1,d1_5,c,195,0.013)

bond (d1,d1_1,d1_2,7)

bond (d1,d1_2,d1_3,7)

bond (d1,d1_3,d1_4,7)

bond (d1,d1_4,d1_5,7)

bond (d1,d1_5,d1_6,7)

INDUCE (Dietterich & Michalski, 1983)
 SMART+ (Botta & Giordano, 1988, 1993)
 FOIL (Quinlan, 1990)
 PROGOL (Muggleton, 1994)
 STILL (Sebag, 1998)



Nonactive (d167)

lumo (d167, -1.246)

loop (d167, 4.23)

atm (d167,d167_1,n,22,-0.117)

atm (d167,d167_2,c,22,-0.117)

atm (d167,d167_3,n,22,-0.117)

atm (d167,d167_4,c,195,-0.087)

atm (d167,d167_5,n,195,0.013)

bond (d167,d167_1,d167_2,7)

bond (d167,d167_2,d167_3,7)

....



Relational Learner

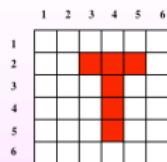


active(M) :- $\neg \text{chrg}(x_1, [-0.2]) \wedge \neg \text{type}(x_2, [N]) \wedge$
 $\neg \text{anm}(x_3, [22]) \wedge \neg \text{chrg}(x_3, [-0.6, -0.4]) \wedge$
 $\neg \text{type}(x_4, [H, N, O]) \wedge \text{bound}(x_2, x_3) \wedge \text{bound}(x_3, x_4) \wedge$
 $\text{atm}(M, x_1) \wedge \text{atm}(M, x_2) \wedge \text{atm}(M, x_3) \wedge \text{atm}(M, x_4)$

L'induction de connaissances structurelles

Programmation Logique Inductive (ILP)

Matching Problem



Formula φ

$$\varphi(x,y,z,u,v,w) = \text{red}(x) \wedge \text{red}(y) \wedge \text{red}(z) \wedge \text{red}(u) \wedge \text{red}(v) \wedge \text{red}(w) \wedge \text{west}(x,y) \wedge \text{adj}(x,y) \wedge \text{west}(y,z) \wedge \text{adj}(y,z) \wedge \text{north}(y,u) \wedge \text{adj}(y,u) \wedge \text{north}(u,v) \wedge \text{adj}(u,v) \wedge \text{north}(v,w) \wedge \text{adj}(v,w)$$

		1	2	3	4	5	6
1							
2							
3							
4							
5							
6							

		1	2	3	4	5	6
1							
2							
3							
4							
5							
6							

1 solution

Universe U

		1	2	3	4	5	6
1							
2							
3							
4							
5							
6							

		x	y	z	u	v	w
1	a ₁₁	a ₂₁	a ₃₁	a ₄₁	a ₅₁	a ₆₁	
2	a ₁₂	a ₂₂	a ₃₂	a ₄₂	a ₅₂	a ₆₂	
3	a ₁₃	a ₂₃	a ₃₃	a ₄₃	a ₅₃	a ₆₃	
4	a ₁₄	a ₂₄	a ₃₄	a ₄₄	a ₅₄	a ₆₄	
5	a ₁₅	a ₂₅	a ₃₅	a ₄₅	a ₅₅	a ₆₅	
6	a ₁₆	a ₂₆	a ₃₆	a ₄₆	a ₅₆	a ₆₆	



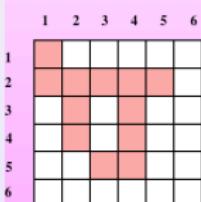
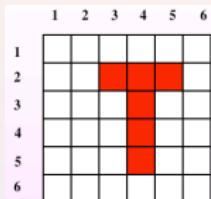
		1	2	3	4	5	6
1							
2							
3							
4							
5							
6							

		1	2	3	4	5	6
1							
2							
3							
4							
5							
6							

No solution

4 solutions

ILP et satisfaction de contraintes



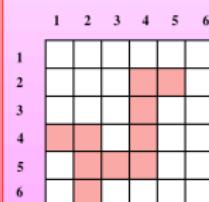
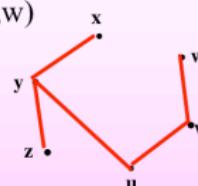
1 solution

red(x)		west(x,y)		north(x,y)		adj(x,y)	
x		x	y	x	y	x	y
a ₂₁		a ₂₁	a ₂₂	a ₁₁	a ₂₁	a ₁₁	a ₂₁
a ₂₁		a ₂₁	a ₂₃	a ₂₂	a ₃₂	a ₂₁	a ₂₂
a ₂₁		a ₂₁	a ₂₄	a ₂₂	a ₄₂	a ₂₂	a ₂₁
a ₂₁		a ₂₂	a ₃₂	a ₂₂	a ₂₃
a ₂₁		a ₃₂	a ₃₄	a ₃₂	a ₄₂	a ₂₂	a ₃₂
...		a ₃₂	a ₃₂
a ₅₄		a ₅₃	a ₅₄	a ₄₄	a ₅₄
				a ₅₃	a ₅₄		

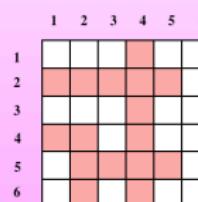
Universe U

Formula φ

$$\varphi(x, y, z, u, v, w) = \text{red}(x) \wedge \text{red}(y) \wedge \text{red}(z) \wedge \text{red}(u) \wedge \text{red}(v) \wedge \text{red}(w) \wedge \text{west}(x, y) \wedge \text{adj}(x, y) \wedge \text{west}(y, z) \wedge \text{adj}(y, z) \wedge \text{north}(y, u) \wedge \text{adj}(y, u) \wedge \text{north}(u, v) \wedge \text{adj}(u, v) \wedge \text{north}(v, w) \wedge \text{adj}(v, w)$$

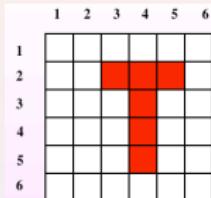


No solution



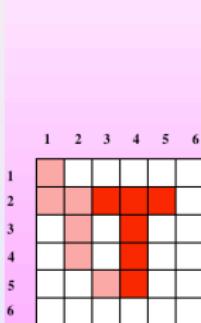
4 solutions

ILP et satisfaction de contraintes



Formula φ

$$\varphi(x, y, z, u, v, w) = \text{red}(x) \wedge \text{red}(y) \wedge \text{red}(z) \wedge \text{red}(u) \wedge \text{red}(v) \wedge \text{red}(w) \wedge \text{west}(x, y) \wedge \text{adj}(x, y) \wedge \text{west}(y, z) \wedge \text{adj}(y, z) \wedge \text{north}(y, u) \wedge \text{adj}(y, u) \wedge \text{north}(u, v) \wedge \text{adj}(u, v) \wedge \text{north}(v, w) \wedge \text{adj}(v, w)$$



red(x)

x	y
a ₂₁	a ₂₂
a ₂₁	a ₂₃
a ₂₂	a ₃₂
a ₂₁	a ₂₄
a ₂₂	a ₄₂
....	...
a ₃₂	a ₃₄
a ₃₂	a ₄₂
...	...
a ₅₃	a ₅₄

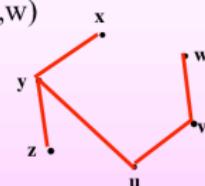
west(x,y)

x	y
a ₁₁	a ₂₁
a ₂₁	a ₁₁
a ₂₁	a ₂₂
a ₂₂	a ₂₁
a ₂₂	a ₃₂
a ₂₂	a ₂₄
....	...
a ₃₂	a ₃₄
a ₃₂	a ₄₂
...	...
a ₅₃	a ₅₄

north(x,y)

x	y
a ₁₁	a ₂₁
a ₂₁	a ₁₁
a ₂₁	a ₂₂
a ₂₂	a ₂₁
a ₂₂	a ₃₂
a ₂₂	a ₂₄
....	...
a ₃₂	a ₃₄
a ₃₂	a ₄₂
...	...
a ₅₃	a ₅₄

adj(x,y)

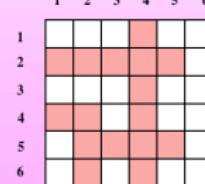
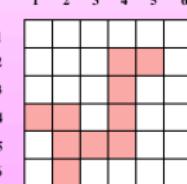


1 solution

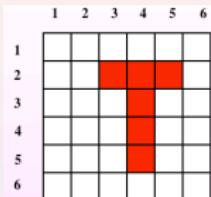
Universe U

No solution

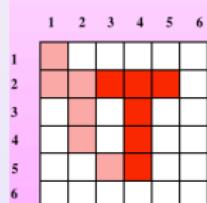
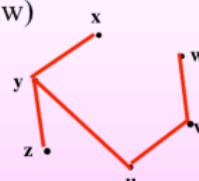
4 solutions



ILP et satisfaction de contraintes



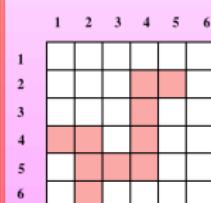
Formula φ

$$\varphi(x,y,z,u,v,w) = \text{red}(x) \wedge \text{red}(y) \wedge \text{red}(z) \wedge \text{red}(u) \wedge \text{red}(v) \wedge \text{red}(w) \wedge \text{west}(x,y) \wedge \text{adj}(x,y) \wedge \text{west}(y,z) \wedge \text{adj}(y,z) \wedge \text{north}(y,u) \wedge \text{adj}(y,u) \wedge \text{north}(u,v) \wedge \text{adj}(u,v) \wedge \text{north}(v,w) \wedge \text{adj}(v,w)$$


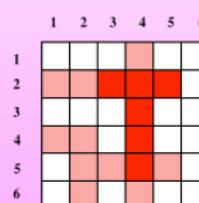
1 solution

red(x)		west(x,y)		north(x,y)		adj(x,y)	
x	y	x	y	x	y	x	y
a ₂₁	a ₂₂	a ₂₁	a ₂₃	a ₂₂	a ₃₂	a ₂₁	a ₁₁
a ₂₁	a ₂₃	a ₂₁	a ₂₄	a ₂₂	a ₄₂	a ₂₁	a ₂₂
a ₂₁	a ₂₄	a ₂₂	a ₃₂	a ₂₂	a ₂₁
....	a ₃₂	a ₃₄	a ₃₂	a ₄₂	a ₂₂	a ₂₃
a ₂₁	a ₃₂	a ₃₂	a ₃₄	a ₃₂	a ₄₂	a ₂₂	a ₃₂
...
a ₅₄	a ₅₄	a ₅₃	a ₅₄	a ₄₄	a ₅₄	a ₅₃	a ₅₄

Universe U



No solution



4 solutions

ILP et satisfaction de contraintes

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

Formula φ

$$\varphi(x,y,z,u,v,w) = \text{red}(x) \wedge \text{red}(y) \wedge \text{red}(z) \wedge \text{red}(u) \wedge \text{red}(v) \wedge \text{red}(w) \wedge \text{west}(x,y) \wedge \text{adj}(x,y) \wedge \text{west}(y,z) \wedge \text{adj}(y,z) \wedge \text{north}(y,u) \wedge \text{adj}(y,u) \wedge \text{north}(u,v) \wedge \text{adj}(u,v) \wedge \text{north}(v,w) \wedge \text{adj}(v,w)$$

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

$\text{red}(x)$

X	
a ₁₁	
a ₂₁	
....
a ₅₁	
...	...
a ₅₄	

$\text{west}(x,y)$

$\text{north}(x,y)$

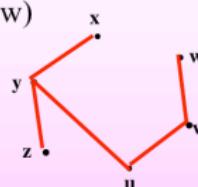
x	y
a ₂₁	a ₂₂
a ₂₁	a ₂₃
a ₂₂	a ₃₂
a ₂₁	a ₂₄
a ₂₂	a ₄₂
a ₂₂	a ₃₂
a ₃₂	a ₃₄
....
a ₃₂	a ₃₄
...	...
a ₅₃	a ₅₄
...	...
a ₄₄	a ₅₄
a ₅₃	a ₅₄

$\text{adj}(x,y)$

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

1 solution

Universe U

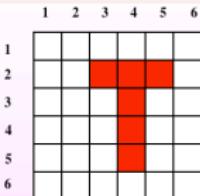


No solution

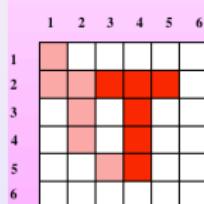
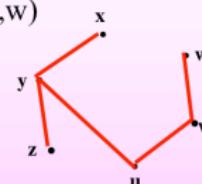
4 solutions

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

ILP et satisfaction de contraintes



Formula φ

$$\varphi(x,y,z,u,v,w) = \text{red}(x) \wedge \text{red}(y) \wedge \text{red}(z) \wedge \text{red}(u) \wedge \text{red}(v) \wedge \text{red}(w) \wedge \text{west}(x,y) \wedge \text{adj}(x,y) \wedge \text{west}(y,z) \wedge \text{adj}(y,z) \wedge \text{north}(y,u) \wedge \text{adj}(y,u) \wedge \text{north}(u,v) \wedge \text{adj}(u,v) \wedge \text{north}(v,w) \wedge \text{adj}(v,w)$$


$\text{red}(x)$

west(x,y)	
x	y
a ₂₁	a ₂₂
a ₂₁	a ₂₃
a ₂₁	a ₂₄
a ₂₂	a ₄₂
a ₂₂	a ₃₂
a ₃₂	a ₃₄
a ₃₂	a ₄₂
a ₅₃	a ₅₄
a ₅₃	a ₅₄

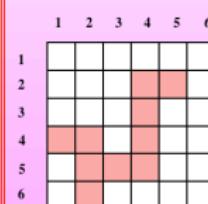
1 solution

Universe U

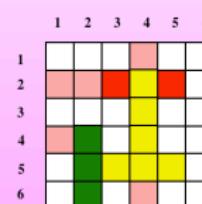
$\text{north}(x,y)$

$\text{north}(x,y)$	
x	y
a ₁₁	a ₂₁
a ₂₂	a ₃₂
a ₂₁	a ₂₄
a ₂₂	a ₃₂
a ₂₂	a ₄₂
a ₂₂	a ₂₁
a ₂₂	a ₂₃
a ₂₂	a ₃₂
a ₂₂	a ₃₄
a ₄₄	a ₅₄
a ₅₃	a ₅₄

$\text{adj}(x,y)$

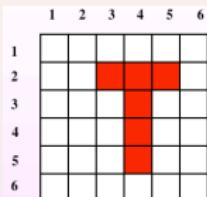


No solution



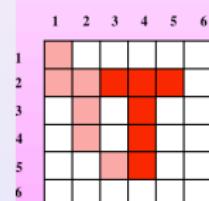
4 solutions

ILP et satisfaction de contraintes



Formula φ

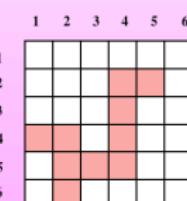
$$\varphi(x, y, z, u, v, w) = \text{red}(x) \wedge \text{red}(y) \wedge \text{red}(z) \wedge \text{red}(u) \wedge \text{red}(v) \wedge \text{red}(w) \wedge \text{west}(x, y) \wedge \text{adj}(x, y) \wedge \text{west}(y, z) \wedge \text{adj}(y, z) \wedge \text{north}(y, u) \wedge \text{adj}(y, u) \wedge \text{north}(u, v) \wedge \text{adj}(u, v) \wedge \text{north}(v, w) \wedge \text{adj}(v, w)$$



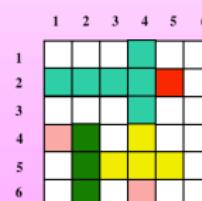
1 solution

red(x)		west(x,y)		north(x,y)		adj(x,y)	
x	y	x	y	x	y	x	y
a ₂₁	a ₂₂	a ₁₁	a ₂₁	a ₂₂	a ₁₁	a ₁₁	a ₂₁
a ₂₁	a ₂₃	a ₂₂	a ₃₂	a ₂₁	a ₁₁	a ₂₁	a ₂₂
a ₂₁	a ₂₄	a ₂₂	a ₄₂	a ₂₂	a ₂₁	a ₂₂	a ₃₂
....	a ₂₂	a ₃₂	a ₂₂	a ₂₁	a ₂₂	a ₄₂
a ₃₂	a ₃₄	a ₃₂	a ₄₂	a ₂₂	a ₂₃	a ₂₂	a ₃₂
...
a ₅₄	a ₅₄	a ₄₄	a ₅₄	a ₅₃	a ₅₄	a ₅₃	a ₅₄

Universe U



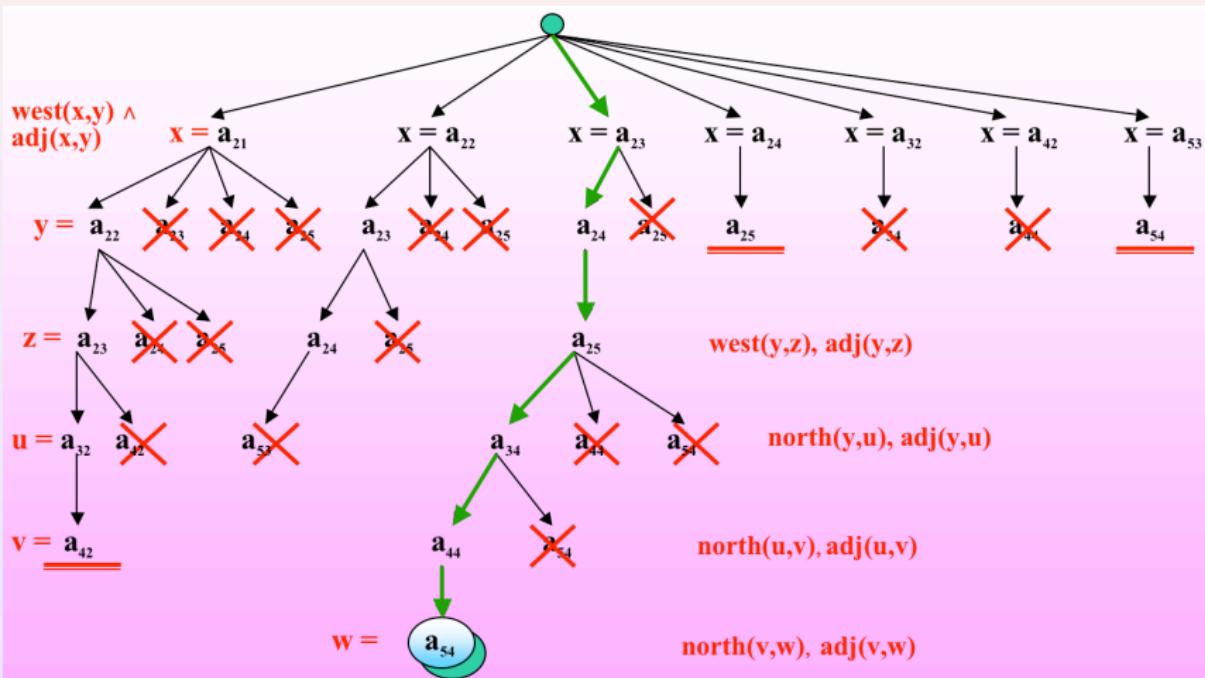
No solution



4 solutions

ILP et satisfaction de contraintes

Complexité de la recherche



Principe des expériences

Étude en cas moyen

Génération aléatoire de problèmes

Hypothèses

- n : nombre de *variables* dans l'hypothèse \underline{h} testée,
- m : nombre de *symboles de prédicats* dans \underline{h} ,

Exemples

- L : nombre total de *constantes* dans l'exemple \underline{e} ,
- N : nombre de *littéraux* construits sur chaque symbole de prédicat dans \underline{e} .

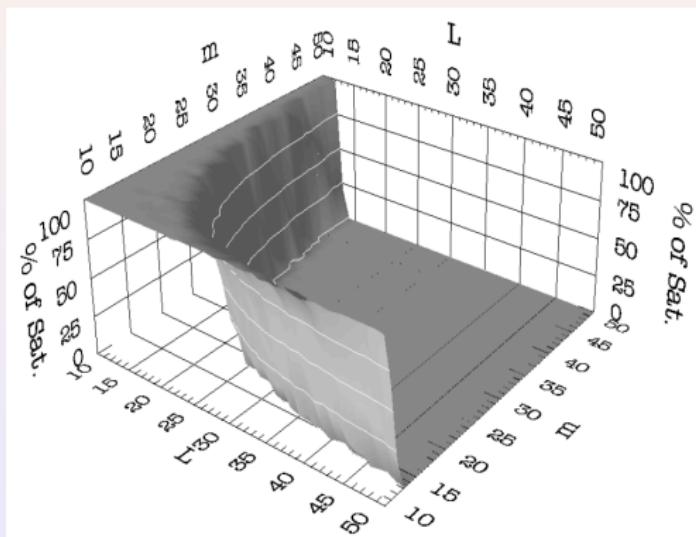
→ *Étude dans le cas moyen*

Des conditions sur le gain d'information

Distribution uniforme suivant des paramètres de contrôle : le cas de l'ILP

Paramètres de contrôle :

- n : nombre de *variables* dans l'hypothèse h testée,
- m : nombre de *symboles de prédictats* dans h ,
- L : nombre total de *constantes* dans l'exemple e ,
- N : nombre de *littéraux* construits sur chaque symbole de prédictat dans e .



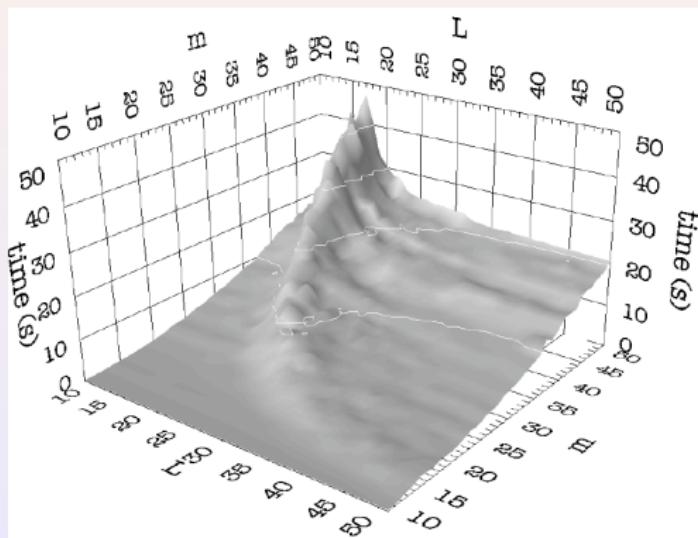
[bot03] Botta, M., A. Giordana, L. Saitta, and M. Sebag
Relational learning as search in a critical region.
Journal of Machine Learning Research, 4, 431-463, 2003.

Coût du test de couverture

Distribution uniforme suivant des paramètres de contrôle : le cas de l'ILP

Paramètres de contrôle :

- n : nombre de *variables* dans l'hypothèse h testée,
- m : nombre de *symboles de prédictats* dans h ,
- L : nombre total de *constantes* dans l'exemple e ,
- N : nombre de *littéraux* construits sur chaque symbole de prédictat dans e .

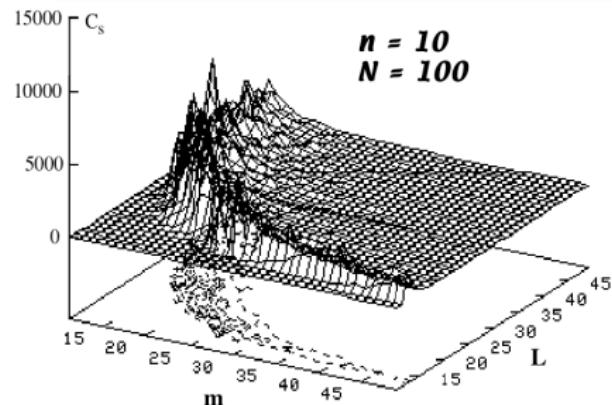
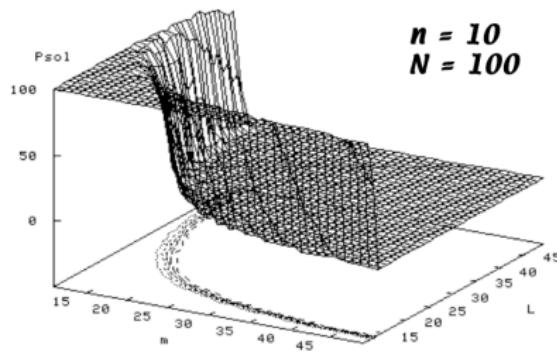


ILP et satisfaction de contraintes

Probabilité de couverture et complexité du test

L = Number of constant in the universe

m = Number of (binary) predicates in a formula



$n = 4, 6, 10, 12, 14$
 $N = 30, 50, 100, 130$

100 problems
for each pair (m, L)

Set of 900,000
matching problems

Conséquences sur l'apprentissage de concept

Paramètre :

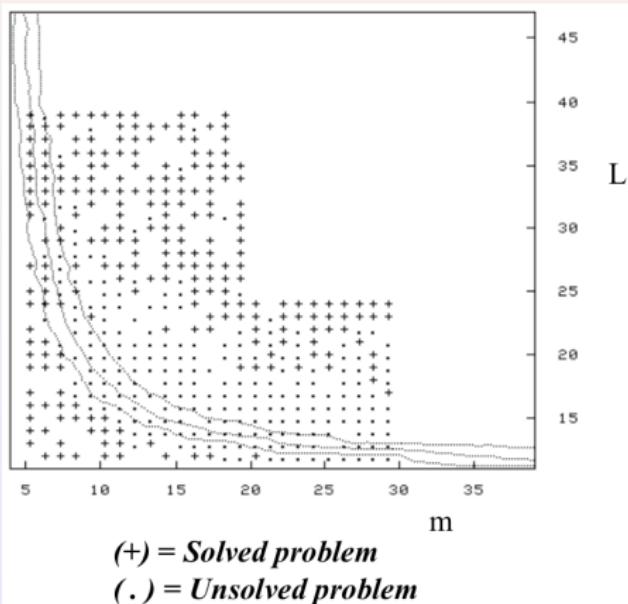
- $n = 4$
- $N = 100$

Contours :

- $P_{sol} = 0.9$
- $P_{sol} = 0.5$
- $P_{sol} = 0.1$

Succès :

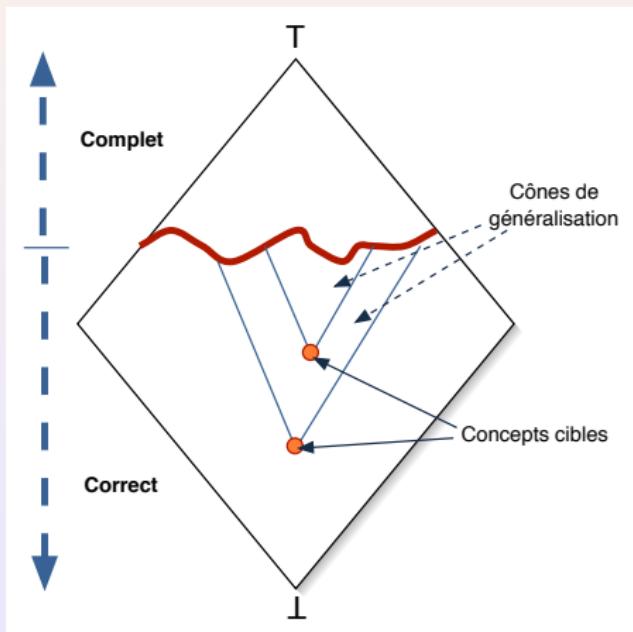
- + : précision > 80%
- . : précision < 80%



Conséquences sur l'apprentissage de concept

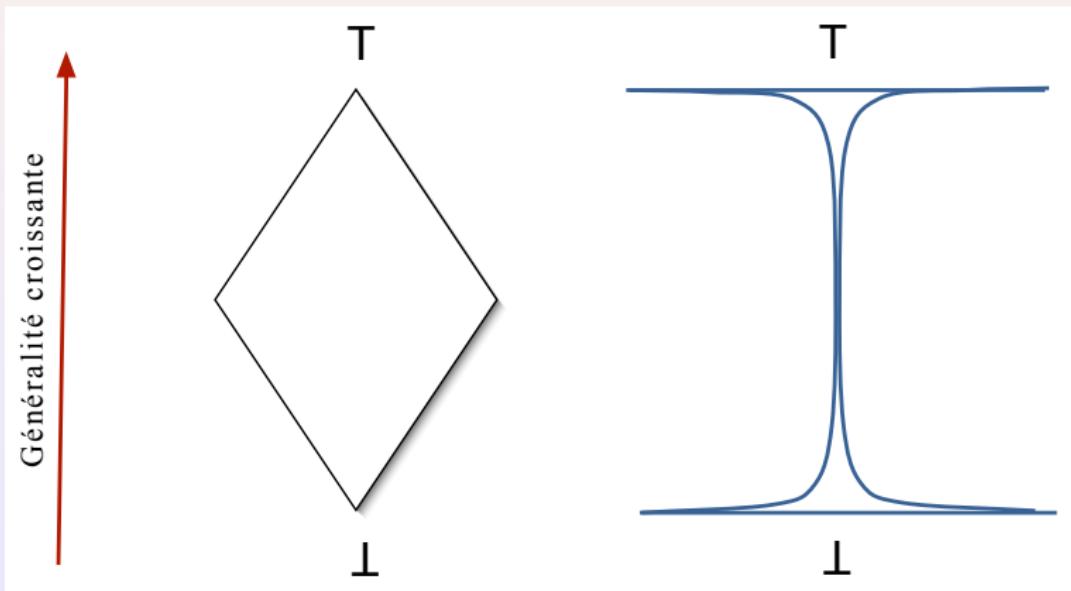
	Probl.	#Nof Clauses	Complexity	Class.Rate (LS) [%]	Class. Rate (TS) [%]	CPU Time [sec]	Avg. #N of Models	
NO	LP ₁ *	10	<7-13>	8.9	88	50	398.2	1.7
	LP ₂ *	11	<6-11>	8.6	92	53	624.7	2.0
	LP ₃ *	15	<7-11>	8.9	98.5	52	513.9	4.9
	LP ₄ *	1	6	100	100	43.3	2.1	
	LP ₅ *	1	6	99.9	100	132.6	1.25	
Approx	LP ₆	12	<1-12>	6	81	58	825.4	10.7
	LP ₇	1	6	100	96	73.4	34.6	
	LP ₈	6	<1-11>	5	98.5	75.3	723.8	1.4
YES	LP ₉	1	9	100	99.6	620.1	1.0	
	LP ₁₀	1	6	100	99.6	36.9	4.2	
	LP ₁₁	1	6	100	99.6	72.2	9.1	

Analyse



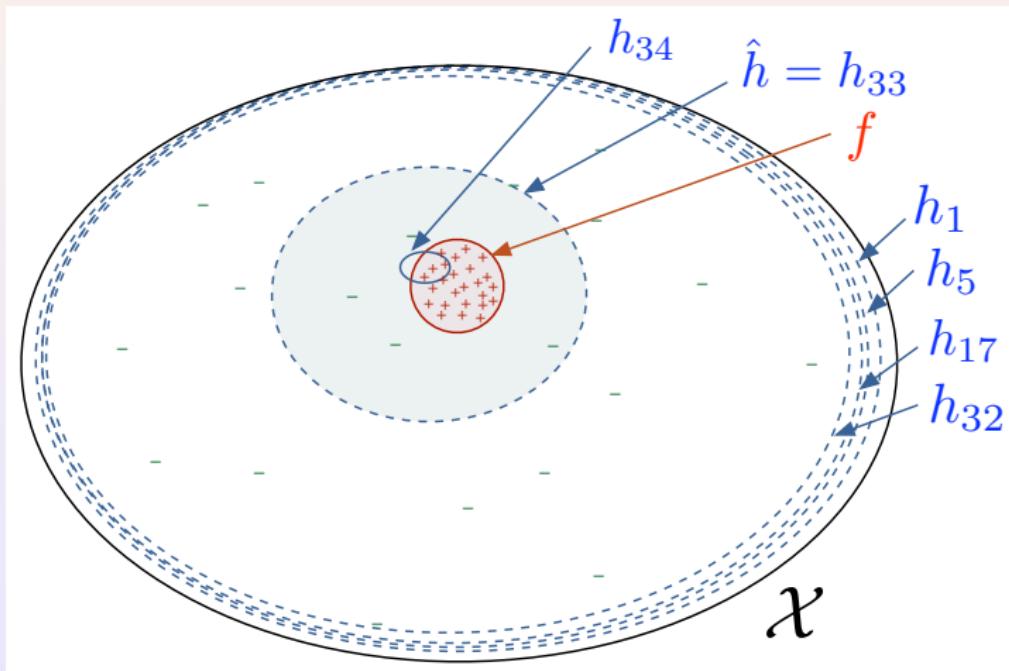
Des conditions sur le gain d'information

Transition de phase et espace des versions



Des conditions sur le gain d'information

Une distribution uniforme ... suspecte ?



Des conditions sur le gain d'information

Bilan

- \exists **transition de phase** dans les variations de taux de couverture
 - observé en ILP
- **Impact considérable** sur les performances
- **Non prévu** par l'analyse statistique

Ce phénomène dépend de :

- ➊ $\mathcal{L}_{\mathcal{H}}$: langage des hypothèses
- ➋ $\mathcal{L}_{\mathcal{X}}$: langage des exemples

Des conditions sur le gain d'information

Questions ouvertes

1. *Quels sont les langages affectés ?*

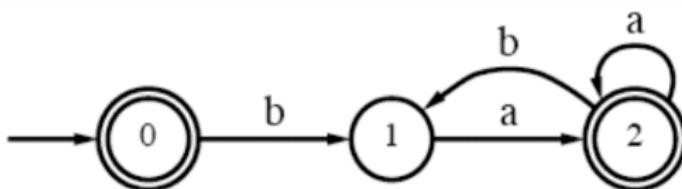
2. *Peut-on contourner le problème ?*

Inférence grammaticale : rappels

- **Entrée** : *chaînes* sur un alphabet Σ , de longueur ℓ
- **Sortie** : *Automate fini* (langage régulier)
 - DFA : *Deterministic Finite Automata*
 - NFA : *Non deterministic Finite Automata*

Échantillon positif : $S^+ = \{ba, baa, baba, \lambda\}$

FSA (DFA)
couvrant S^+



Gain d'information en **inférence grammaticale**

Distribution uniforme avec paramètres de contrôle

Paramètres de contrôle :

- **Q** états
- **B** d'arcs sortants / état
- **L** lettres / arc
- Fraction **a** $\in [0, 1]$ d'états acceptants
- Taille $|\Sigma|$ de l'alphabet
- Longueur **ℓ** des exemples testés.

[Pin01] S. Pinto.

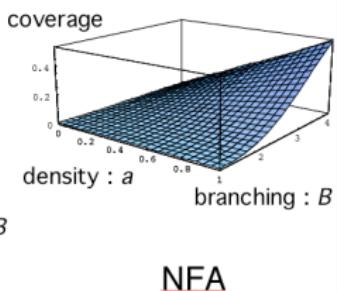
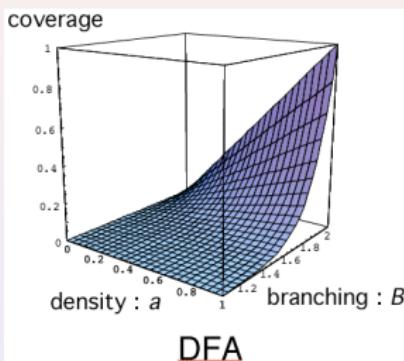
*Etude du phénomène de transition de phase dans l'induction supervisée.
Rapport de DEA (LRI, Univ. Paris-Sud, Orsay), 2001.*

Gain d'information en inférence grammaticale

Distribution uniforme avec paramètres de contrôle

Paramètres de contrôle :

- **Q** états
- **B** d'arcs sortants / état
- **L** lettres / arc
- Fraction **a** ∈ [0, 1] d'états acceptants
- Taille **|Σ|** de l'alphabet
- Longueur **ℓ** des exemples testés.



$$P(\text{accept}) = \begin{cases} a \cdot \left(\frac{B \cdot L}{|\Sigma|}\right)^\ell & \text{pour un DFA} \\ a \cdot [1 - (1 - \frac{L}{|\Sigma|})^B]^\ell & \text{pour un NFA} \end{cases}$$

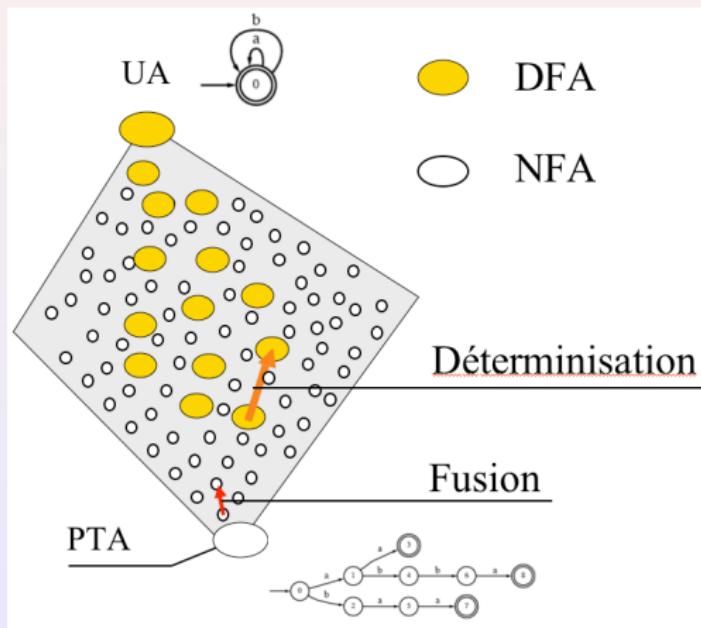
[Pin01] S. Pinto.

Etude du phénomène de transition de phase dans l'induction supervisée.

Rapport de DEA (LRI, Univ. Paris-Sud, Orsay), 2001.

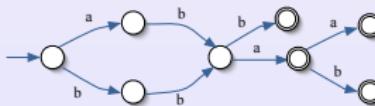
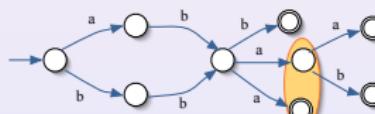
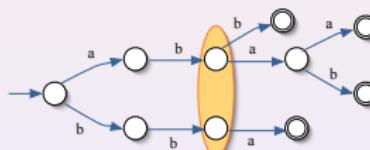
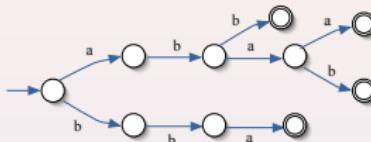
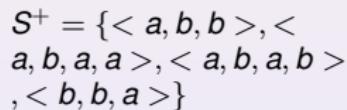
Gain d'information en inférence grammaticale

Principe des algorithmes d'apprentissage



Gain d'information en inférence grammaticale

Distribution uniforme avec paramètres de contrôle



Gain d'information en inférence grammaticale

Étude sur l'espace d'hypothèses effectivement exploré

Protocole expérimental

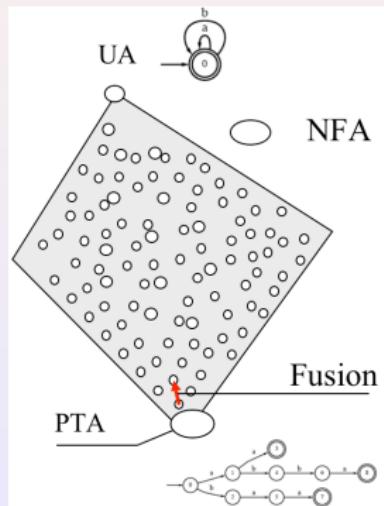
- ① Génération aléatoire d'un **échantillon d'apprentissage** : $|S^+| (= 200)$ chaînes de taille ℓ
- ② Construction du **PTA** pour chaque échantillon S^+
- ③ **Calcul de chemins de généralisation partant du PTA :**
 - **Fusions aléatoires**
 - **Couverture calculée pour chaque automate engendré** (sur un *ensemble test* : 1000 chaînes aléatoires \notin ens. d'apprentissage)

Expériences

- $|\Sigma| = \{2, 4, 8\}$
- $\ell = \{4, 8, 16, 32\}$
- $50 \text{ PTAs} \times 20 \text{ trajectoires aléatoires} = 1000 \text{ trajectoires}$
 $(\approx 270\,000 \text{ automates})$

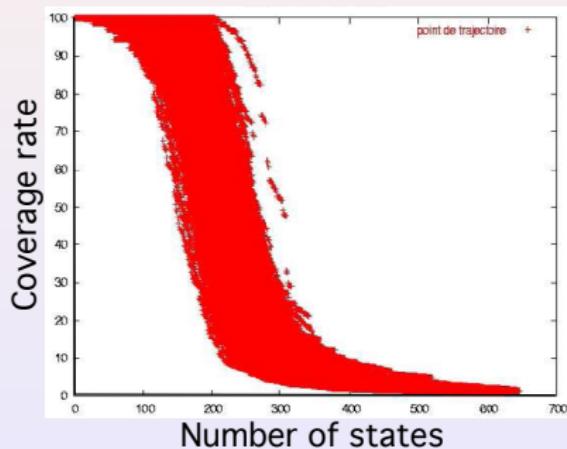
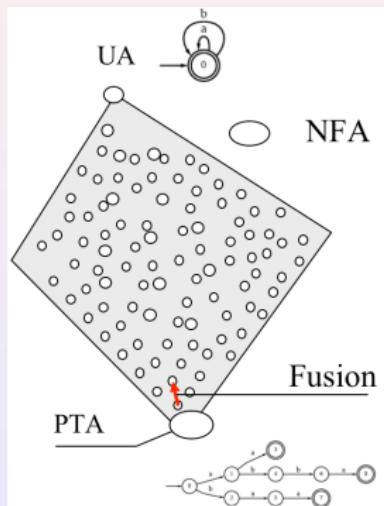
Gain d'information en inférence grammaticale

Cas non-déterministe : NFA



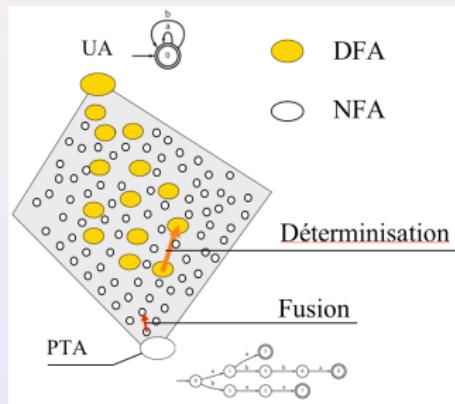
Gain d'information en inférence grammaticale

Cas non-déterministe : NFA



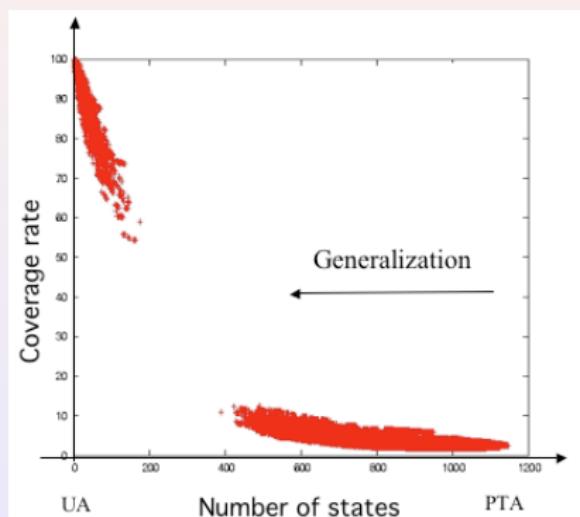
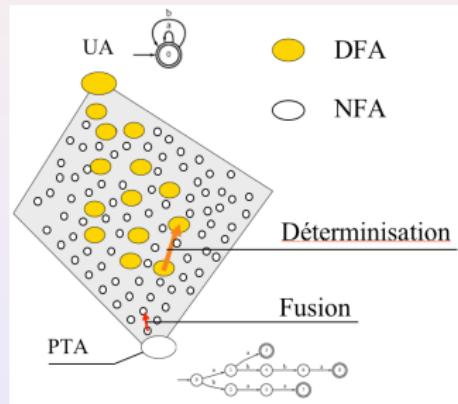
Gain d'information en inférence grammaticale

Cas déterministe : DFA



Gain d'information en inférence grammaticale

Cas déterministe : DFA



$$|\Sigma| = 8 ; \ell = 8$$

Gain d'information en inférence grammaticale

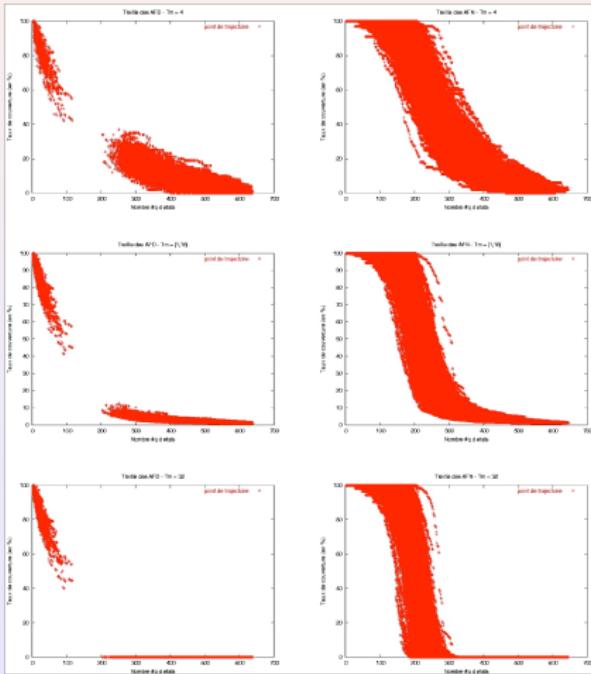
Variété de situations

$$|\Sigma| = 4$$

$$\ell = 16$$

$$|S^+| = 100$$

Test sur 1000 chaînes
de tailles : 4, 16 et 32



Gain d'information en inférence grammaticale

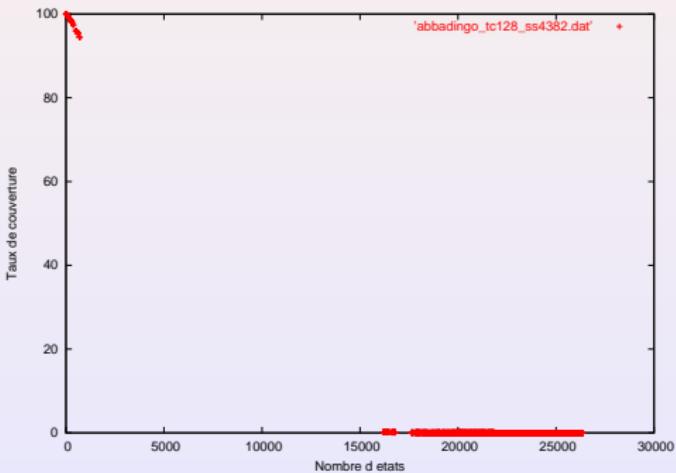
Le défi *Abbadingo*

$$|\Sigma| = 2$$

$$\ell = 17$$

$$|S^+| = 4382$$

Test sur 1000 chaînes
de taille : 17



Analyse pour les DFA

Algorithmes avec heuristiques de recherche

- Par fusion d'états
- Jusqu'à la couverture d'exemples négatifs

RPNI [OG92][Lan92]

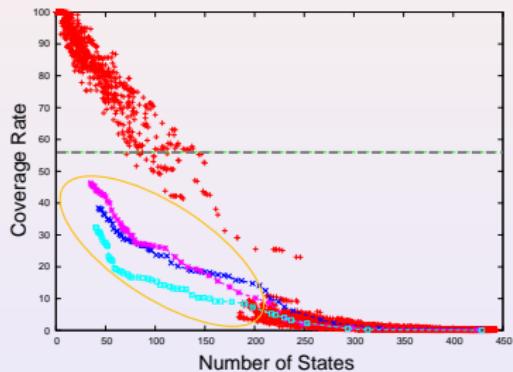
- Choix des états *en largeur d'abord*

EDSM [Lan et al.98][Lan98]

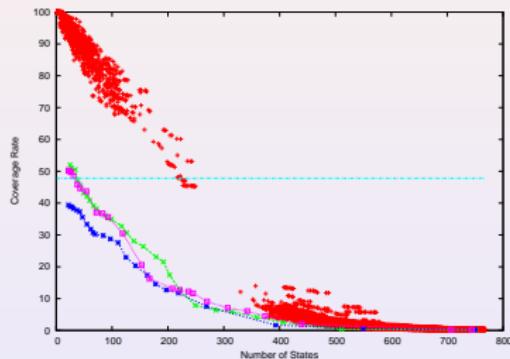
- Choix des états *dont la fusion conduit au maximum de fusions pour déterminisation*

Analyse pour les DFA

Algorithmes avec stratégies de recherche : **résultats**



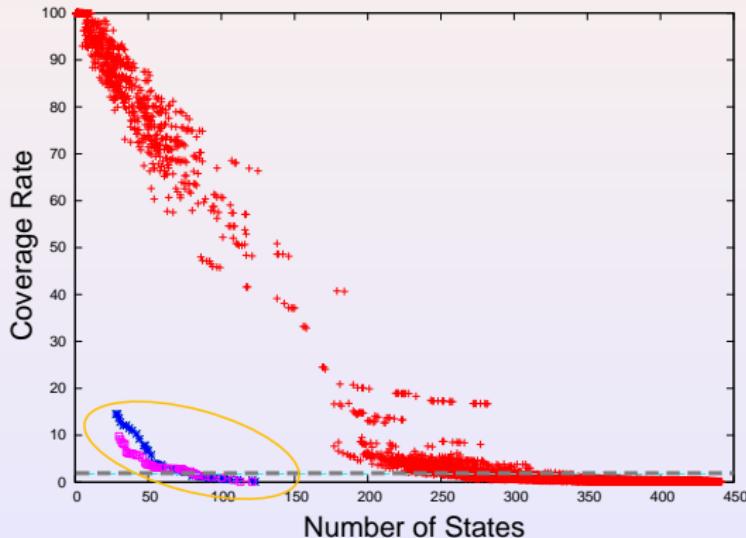
RPNI



EDSM

Analyse pour les DFA

Algorithmes avec stratégies de recherche : **résultats**



Analyse pour les DFA

Algorithmes avec stratégies de recherche : **résultats**

Target automata

Learned automata

Algo.	<i>Target automata</i>			<i>Learned automata</i>		
	Q_c	$ucov_c$	Q_f	$ucov_f$	%cov+	%cov-
RB	15	5.97	10.38	33.81	60.93	34.69
RB	25	4.88	12.77	40.35	62.68	37.87
RB	50	4.2	14.23	45.38	66.14	42.23
RB	100	3.39	13.13	30.35	42.81	28.69
RPNI	15	5.95	5.14	22.9	57.51	26.99
RPNI	25	4.7	7.56	23.07	56.38	25.98
RPNI	50	3.87	14.08	23.45	51.89	24.42
RPNI	100	3.12	26.41	23.151	50.12	24.40

Plan

- 1 Un point de vue sur l'apprentissage
 - L'induction supervisée
 - Analyse des conditions de l'induction
- 2 Transition de phase en induction
 - Gain d'information et transition de phase
 - Transition de phase en induction de programmes logiques
 - Transition de phase en inférence grammaticale
- 3 Analyse et bilan
 - Analyse et explication ?
 - Remèdes ?
 - Bilan

Analyse pour les DFA

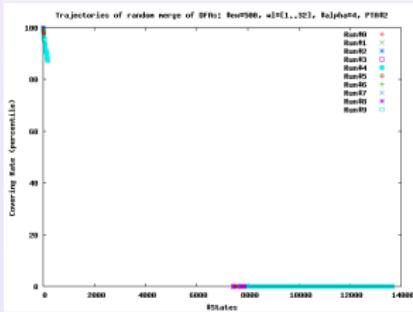
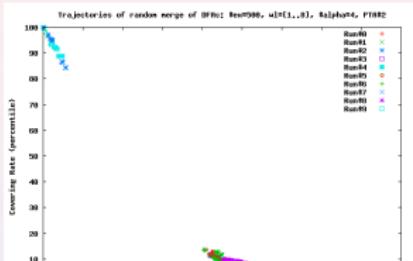
Un saut de couverture spectaculaire

$$|\Sigma| = 4$$

$$\ell \in [1, \dots, 8] \text{ ou } [1, \dots, 32]$$

$$|S^+| = 500$$

8 trajectoires aléatoires



Analyse pour les DFA

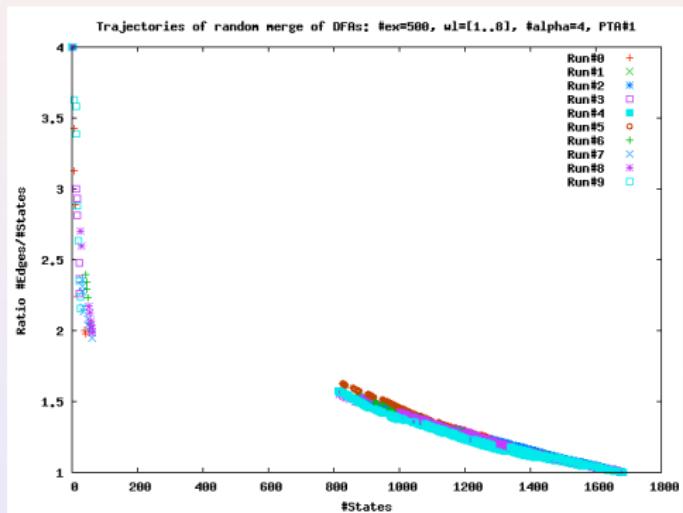
Évolution du nombre d'arcs / états

$$|\Sigma| = 4$$

$$\ell \in [1, \dots, 8]$$

$$|S^+| = 500$$

8 trajectoires aléatoires



Analyse pour les DFA

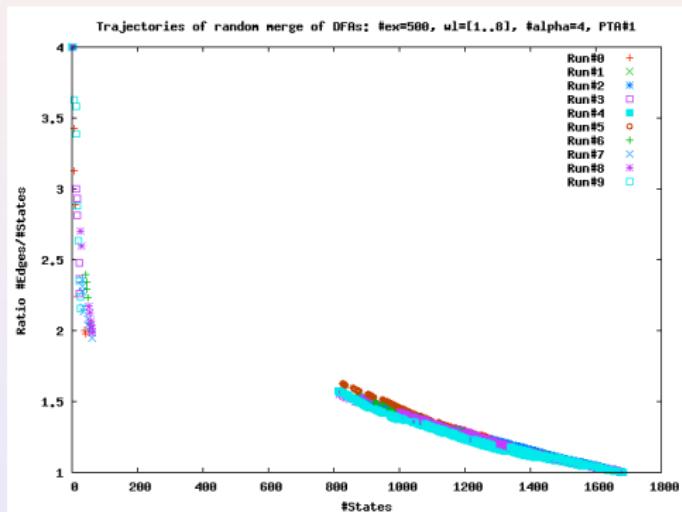
Évolution du nombre d'arcs / états

$$|\Sigma| = 4$$

$$\ell \in [1, \dots, 8]$$

$$|S^+| = 500$$

8 trajectoires aléatoires



Pas de saut !!

Analyse pour les DFA

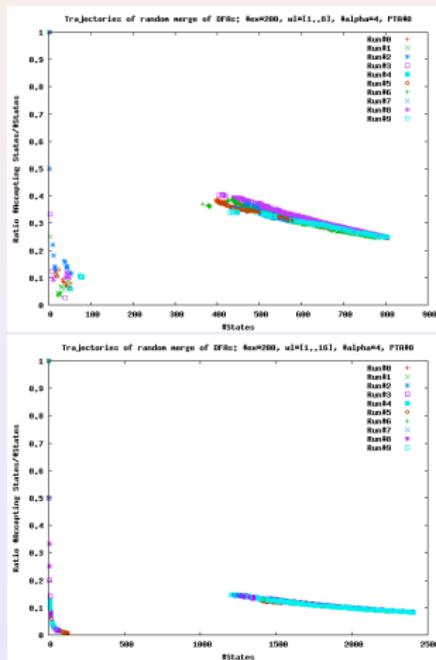
Évolution du taux d'états acceptants

$$|\Sigma| = 4$$

$\ell \in [1, \dots, 8] \text{ ou } [1, \dots, 16]$

$$|S^+| = 200$$

8 trajectoires aléatoires



Analyse pour les DFA

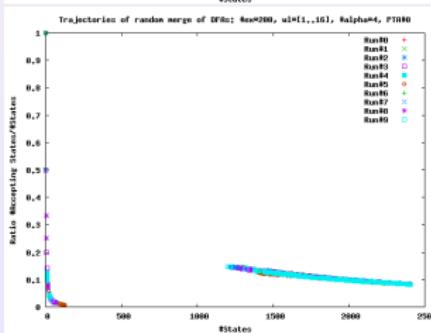
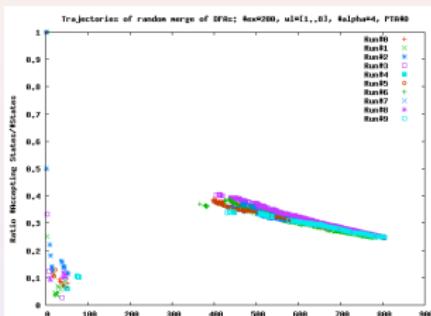
Évolution du taux d'états acceptants

$$|\Sigma| = 4$$

$\ell \in [1, \dots, 8] \text{ ou } [1, \dots, 16]$

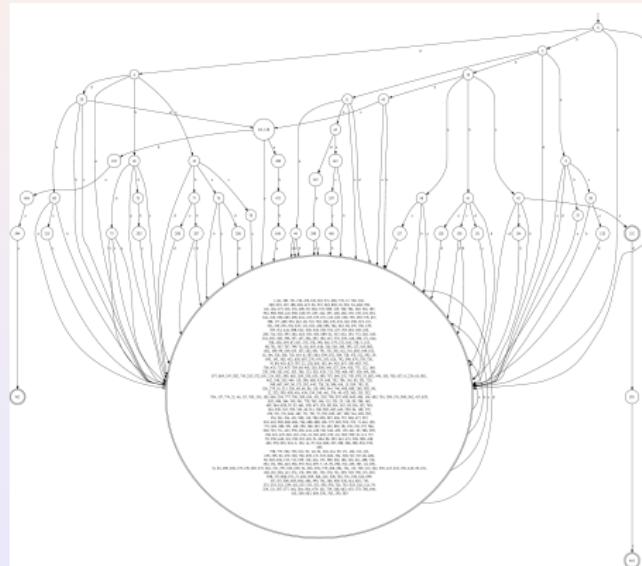
$$|S^+| = 200$$

8 trajectoires aléatoires



Analyse pour les DFA

Saut du taux d'états acceptants



Les états acceptants sont davantage fusionnés !!

Analyse pour les DFA

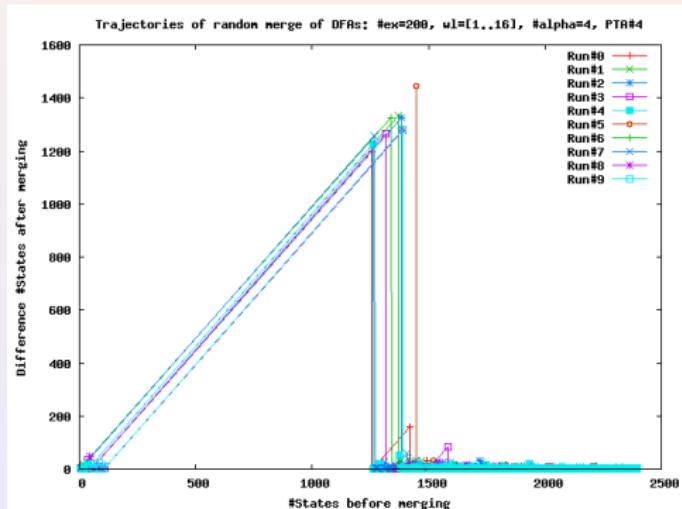
L'explication : avalanches de fusions pour déterminisation

$$|\Sigma| = 4$$

$$\ell \in [1,..,16]$$

$$|S^+| = 200$$

8 trajectoires aléatoires



Analyse pour les DFA

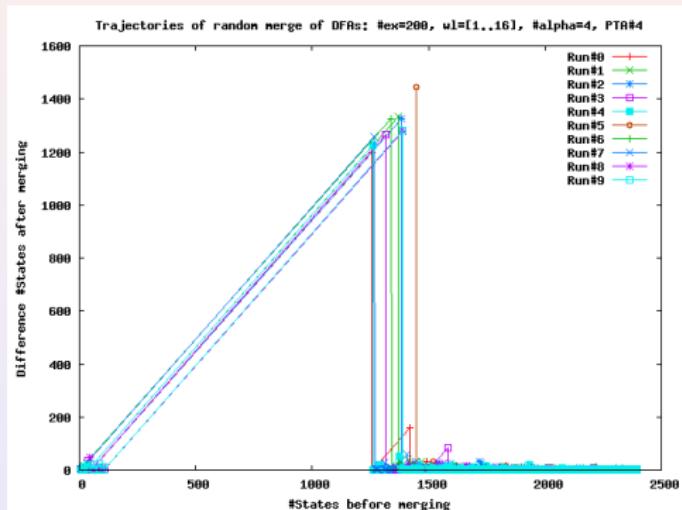
L'explication : avalanches de fusions pour déterminisation

$$|\Sigma| = 4$$

$$\ell \in [1, \dots, 16]$$

$$|S^+| = 200$$

8 trajectoires aléatoires



Phénomène d'avalanche (ou de réactions en chaînes)

Analyse pour les DFA

Des fusions pour déterminisation particulières ?

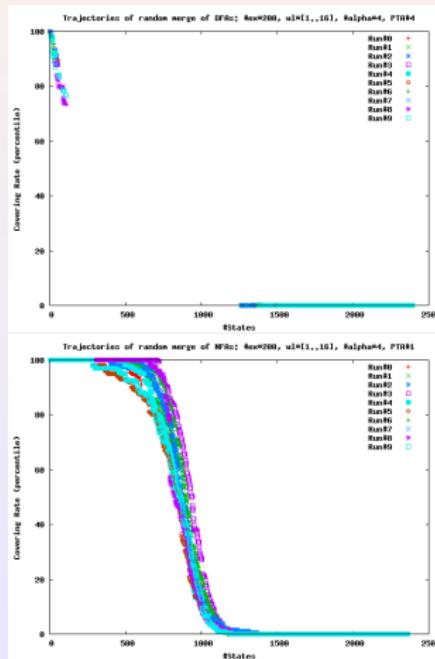
Comparaison entre DFA et NFA

$$|\Sigma| = 4$$

$$\ell \in [1, \dots, 16]$$

$$|S^+| = 200$$

8 trajectoires aléatoires



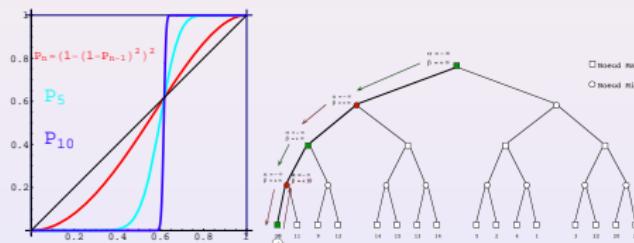
Analyse pour les NFA

Quelle transition de phase ?

L'exemple de l'algorithme de Min-Max

Probabilité que la **position racine** soit **gagnante** en fonction de la probabilité qu'une **position feuille** soit **gagnante**

$$P_n = 1 - (1 - P_{n-1}^b)^b$$



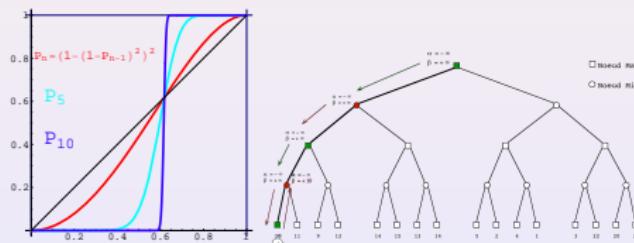
Analyse pour les NFA

Quelle transition de phase ?

L'exemple de l'algorithme de Min-Max

Probabilité que la **position racine** soit **gagnante** en fonction de la probabilité qu'une **position feuille** soit **gagnante**

$$P_n = 1 - (1 - P_{n-1}^b)^b$$

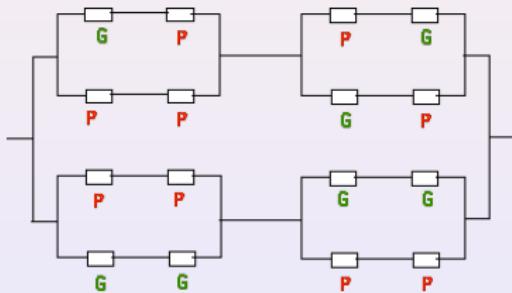


En informatique :

La récurrence peut conduire à un phénomène de TP

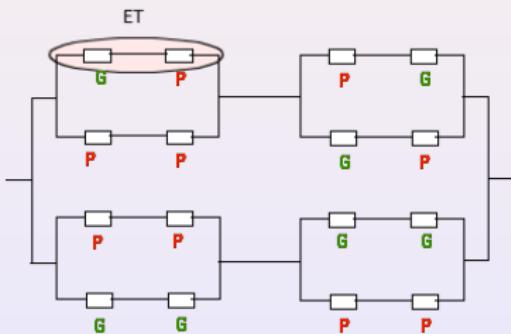
Analyse pour les NFA

Structures auto-similaires



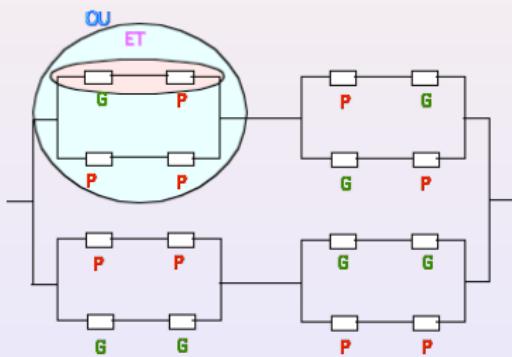
Analyse pour les NFA

Structures auto-similaires



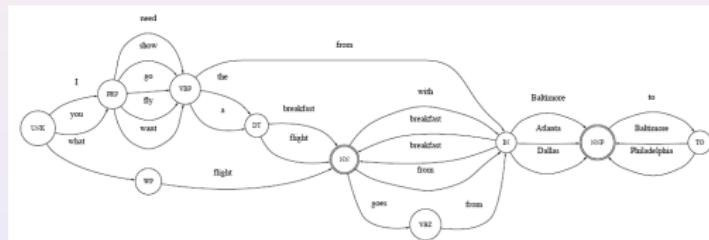
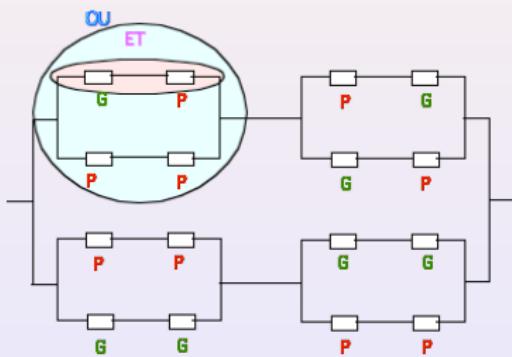
Analyse pour les NFA

Structures auto-similaires



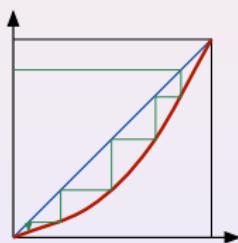
Analyse pour les NFA

Structures auto-similaires

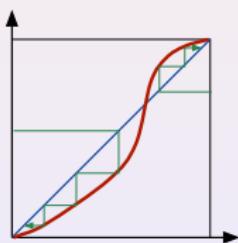


Analyse pour les NFA

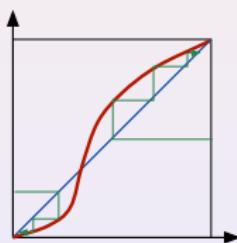
Une explication à envisager



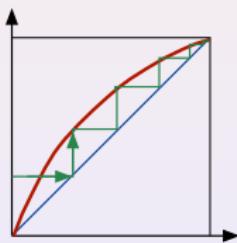
$\tau \approx 0$



$\tau < 1/2$



$\tau > 1/2$

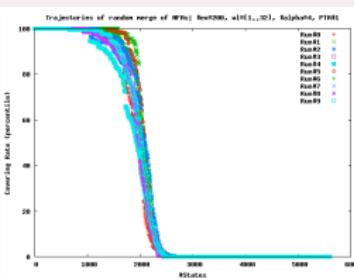
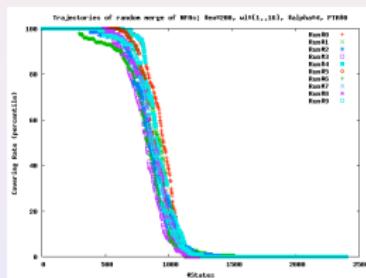
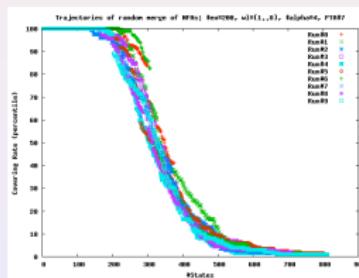


$\tau \approx 1$

Analyse pour les NFA

Transition de phase ou pas ?

La modélisation en réseau récurrent s'applique-t-elle ?



$I = 8$

$I = 16$

$I = 32$

Autre modélisation analytique en cours ...

Analyse pour les DFA

Remèdes ?

Lutte contre l'avalanche de déterminisations

- *Heuristiques de choix des états à fusionner*

Autres opérateurs de généralisation

- *Fusion des nœuds feuilles*
- *Ajout d'arcs*
- *Ajout d'états acceptants*

Autres types d'apprentissage : choix des exemples

- *Apprentissage actif*
- *Apprentissage guidé*

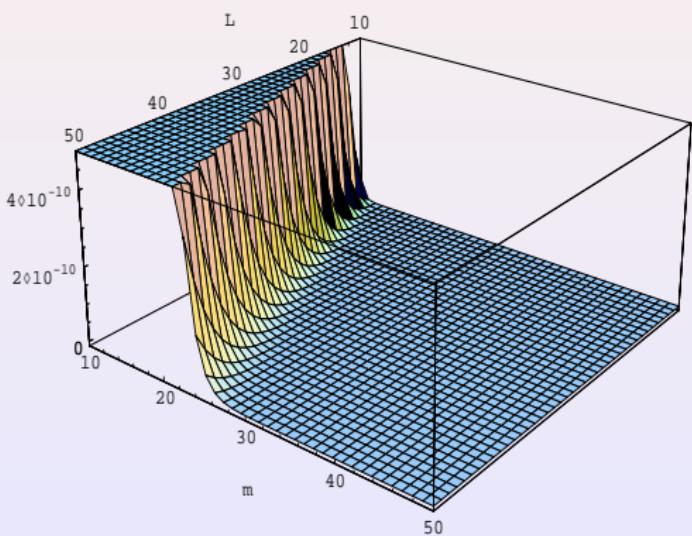
Transition de phase ou pas ?

Le cas de la programmation logique inductive

$$h(\mathbf{x}) = \bigwedge_{i=1}^n p_i(x_i, y_i)$$

Probabilité de couverture :

$$(1 - (\text{une probabilité})^L)^m$$



Rappel :

- m : nombre de *symboles de prédictas* dans h
- L : nombre total de *constantes* dans l'exemple e

Analyse

Alors ... pour quels langages ?

- Programmation logique inductive
- Inférence d'automates à états finis
- ... ?

Importance du concept de récurrence

- structure (e.g. inférence grammaticale)
- test de couverture

Encore du domaine de la recherche

Gain d'information et induction

Conclusions générales

Analyse statistique de l'induction

- ① Ne prend en compte que la capacité de \mathcal{H}
- ② Indépendamment de toute structure sur \mathcal{H}

- Étude plus fine de l'induction
- prenant en compte le **gradient du gain d'information**
 $(\frac{\partial \tau_s(h)}{\partial h})$
- et l'**espace effectivement exploré** : les opérateurs

Gain d'information et induction

Conclusions générales

Analyse statistique de l'induction

- ① Ne prend en compte que la capacité de \mathcal{H}
- ② Indépendamment de toute structure sur \mathcal{H}

- Étude plus fine de l'induction
- prenant en compte le **gradient du gain d'information**
 $(\frac{\partial \tau_s(h)}{\partial h})$
- et l'**espace effectivement exploré** : les opérateurs

Tout aussi fondamental pour l'étude de l'induction !

Plan

4

Annexe

- Changement de référentiel et analogie
- Changement de référentiel et effet tunnel

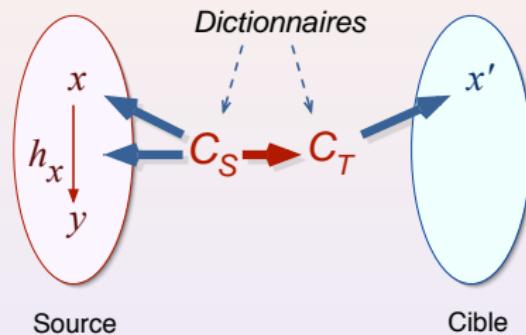
Table des Appendices

4

Annexe

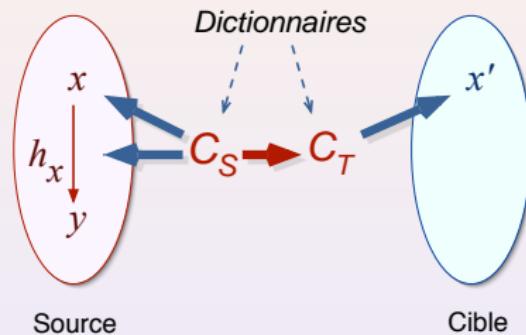
- Changement de référentiel et analogie
- Changement de référentiel et effet tunnel

Changement de référentiel et analogie



$$\text{Coût}((x \rightarrow y), x') = K(C_S) + K(C_T|C_S) + K(x|C_S) + K(h_x|C_S) + K(x'|C_T)$$

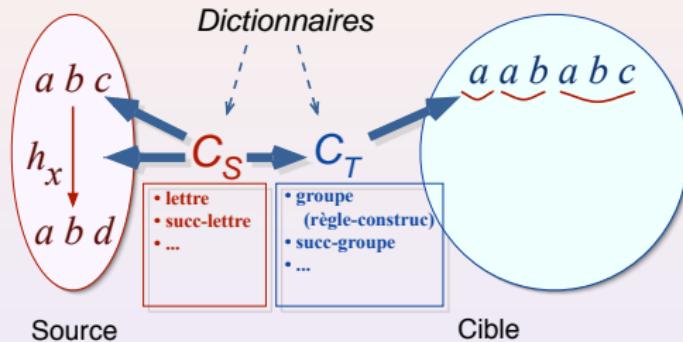
Changement de référentiel et analogie



$$\text{Coût}((x \rightarrow y), x') = K(C_S) + K(C_T|C_S) + K(x|C_S) + K(h_x|C_S) + K(x'|C_T)$$

Induction = cas particulier de l'analogie

Changement de référentiel et analogie



[Cor96a] A. Cornuéjols.

Analogie, principe d'économie et complexité algorithmique.

Journées Francophones d'Apprentissage (JFA-96), Sète, France, 1996, pp.233-247.

A. Cornuéjols and J. Ales-Bianchetti.

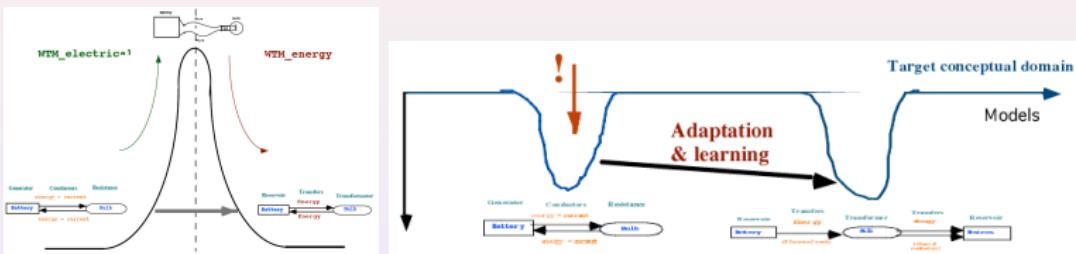
[CAB98] *Analogy and Induction : which (missing) link ?*

Workshop Advances in Analogy research : Integration of theory and data from cognitive, computational and neural sciences, Sofia, Bulgaria, 1998. New Bulgarian University Series (Eds. K. Holyoak, D. Gentner and B. Kokinov), pp. 365-372.

Changement de référentiel

Transferts entre domaines conceptuels

L'effet tunnel cognitif



Apprentissage du concept d'énergie chez des lycéens

[CTC00] A. Cornuéjols, A. Tibergien and G. Collet

A new mechanism for transfer between conceptual domains in scientific discovery and education
Foundations of Science, vol.5, No.2, (2000), 129-155.

[ECCS'97], [MBR'98], [AISB'99], [CAP'99], [Book-chapter,02]